TRAINING MANUAL

DBT Sponsored Training Program

On

Cheminformatics- Tools & Applications

February 19 – 22, 2013

INDIAN INSTITUTE OF SPICES RESEARCH

(Indian Council of Agricultural Research)
Kozhikode – 673012, KERALA

Published by

M. Anandaraj
Director

Organized by

Dr. Santhosh J Eapen, Co-ordinator
Dr. D. Prasath, Co-cordinator

Compiled & Edited by

Santhosh J Eapen
Rosana Babu
Dhanya K. P.
Faisal Moossa Athikkavil
Anoop T. A.

The manual is an in-house publication intended for training purposes only and is not for public circulation.

Copyright © 2013 IISR. All rights reserved. Reproduction and redistribution prohibited without approval.

List of Resource Person

Resource Person	Qualification	Designation
Dr. S. Balaji	M. Sc Bioinformatics (Bharathiar Univ.), M. Phil Bioinformatics (Bharathidasan Univ.), Ph. D. Bio/Chem.-informatics (Mangalore Univ.)	Associate Professor, **Manipal Institute of Technology,** Manipal- Udupi District, Karnataka-576104 E-mail: biobalagi@gmail.com
Fayaz S. M.	M. Tech Bioinformatics (SASTRA Univ.) Ph. D. Student in Computational Neuroscience	Research Scholar, Department of Biotechnology, **National Institute of Technology,** NIT Campus P.O, Calicut. E-mail: fayazbioinfo@gmail.com
Pradeep H.	M. Sc Biochemistry (Mysore Univ.) Ph. D. Student in Computational Neuroscience	Research Scholar, Department of Biotechnology, **National Institute of Technology,** NIT Campus P.O, Calicut. E-mail: pradeeph1157@gmail.com
Mahesh Kumar Teli	M. Tech Biotechnology (SRM Univ.) Ph. D. Student in Computational Drug Designing	Research Scholar, Department of Biotechnology, **National Institute of Technology,** NIT Campus P.O, Calicut. E-mail: maheshkumar.teli@gmail.com
Rosana Babu	M. Sc Bioinformatics (Bharathiar Univ.) Ph. D. Student in Bioscience/Chem.-informatics (Mangalore Univ.)	Research Scholar Bioinformatics **Indian Institute of Spices Research** Kozhikode- 673012 E-mail: rosana.babu@gmail.com

CONTENTS

Serial No.	Topics	Page No.
1.	Cheminformatics- A brief introduction	1
2.	Homology Modeling	11
3.	Molecular docking & Drug designing	26
4.	Quantitative structure activity relationships (QSAR/sar)	39
5.	APPENDIX - I	47

DBT Sponsored Training Program on Cheminformatics – Tools and Applications

Bioinformatics Centre, Indian Institute of Spices Research, Kozhikode- 673012, Kerala, India.
In collaboration with Department of Biotechnology, National Institute of Technology, Kozhikode

February 19 - 22, 2013
Venue: DISC, IISR

Tuesday, 19 Feb. 2012

Time	Session	Facilitator
9.30 – 10.00	Registration	
10.00 – 11.00	**Inaugural session**	
	Welcome and workshop back ground	Dr. Santhosh J. Eapen, Course Coordinator
	Inauguration	**Dr. M. Anandaraj, Director, IISR**
	Introduction of participants	
	Group photo	
11.00 – 11.15	Tea/Coffee	
11.15 – 11.30	**Pre-evaluation**	**Dr. P. Rajeev**, IISR
11.30 – 1.00	**Cheminformatics- A brief introduction** (L&P)	**Dr. S. Balaji**, MIT, Manipal
1.00 – 1.30	Lunch	
1.30 – 3.00	Cheminformatics- A brief introduction (L&P)	-do-
3.00 – 3.15	Tea/Coffee	
3.15 – 4.30	Cheminformatics- A brief introduction (L&P)	-do-

Wednesday, 20 Feb. 2012

Time	Session	Facilitator
9.30 – 11.00	**Small molecule modeling & Homology modeling** (L)	**Fayaz S. M.**, NIT, Calicut
11.00 – 11.30	Tea/Coffee	
11.30 – 1.00	Small molecule modeling & Homology modeling (L)	-do-
1.00 – 1.30	Lunch	
1.30 – 3.00	Small molecule modeling & Homology modeling (P)	Fayaz S.M. & DISC staff
3.00 – 3.15	Tea/Coffee	
3.15 – 4.30	Small molecule modeling & Homology modeling (P)	-do-

Thursday, 21 Feb. 2012

Time	Session	Facilitator
9.30 – 11.00	**Molecular docking & Drug designing** (L)	**Pradeep H.**, NIT, Calicut
11.00 – 11.30	Tea/Coffee break	
11.30 – 1.00	Molecular docking & Drug designing (L)	-do-
1.00 – 1.30	Lunch break	
1.30 – 3.00	Molecular docking & Drug designing (P)	Pradeep H. & DISC staff
3.00 – 3.15	Tea/Coffee break	
3.15 – 4.30	Molecular docking & Drug designing (P)	-do-

Friday, 22 Feb. 2012

Time	Session	Facilitator
9.00 – 10.45	**Quantitative Structure Activity Relationships** (*QSARs*) (L)	**Mahesh Kumar Teli**, NIT, Calicut
10.45 – 11.00	Tea/Coffee break	
11.00 – 12.00	Quantitative Structure Activity Relationships (*QSARs*) (L)	-do-
12.00 – 13.00	Quantitative Structure Activity Relationships (*QSARs*) (P)	Mahesh Kumar Teli & DISC staff
13.00 – 13.30	Lunch break	
13.30 – 15.00	Quantitative Structure Activity Relationships (*QSARs*) (P)	-do-
15.00 – 15.15	**Post evaluation, Feed back**	
15.15 – 15.30	Tea/Coffee break	
15.30 – 16.00	**Valedictory**	**Dr. Rajanikant G.K.**, NIT, Calicut

Chapter 1

Cheminformatics – A brief introduction

Dr. S. Balaji

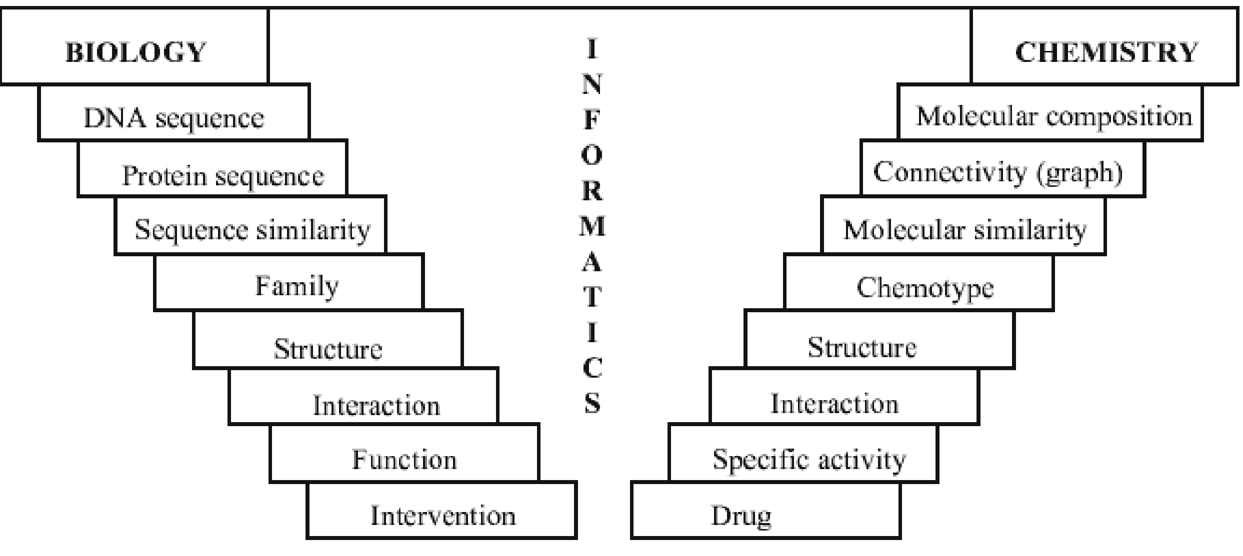
The term **chemoinformatics** (which is synonymously used with **cheminformatics**) was introduced in the literature by Brown in 1998. It is defined as the combination of all the information resources that a scientist needs to optimize the properties of a ligand to become a drug" (Brown 1998).

On the other hand, the term *'chemical informatics'* was used much earlier and generally understood as the application of 'information technology to chemistry', thus lacking a specific drug discovery focus. In addition, the *chemometrics* field focuses on the application of statistical methods to chemical data in order to derive predictive models or descriptors.

The overview of cheminformatics studies is mentioned as follows

- Chemical data collection, analysis and management
- Data representation and communication
- Database design and organization
- Chemical structure and property prediction (including drug-likeness)
- Molecular similarity and diversity analysis
- Compound or library design and optimization
- Database mining
- Compound classification and selection
- Qualitative and quantitative structure-activity or property relationships
- Information theory applied to chemical problems
- Statistical models and descriptors in chemistry
- Prediction of in vivo compound characteristics

The intercept of biology and chemistry in terms of informatics space is shown. The interface of biology and informatics (bioinformatics), and chemistry with informatics (cheminformatics) and their similarities can also be inferred.



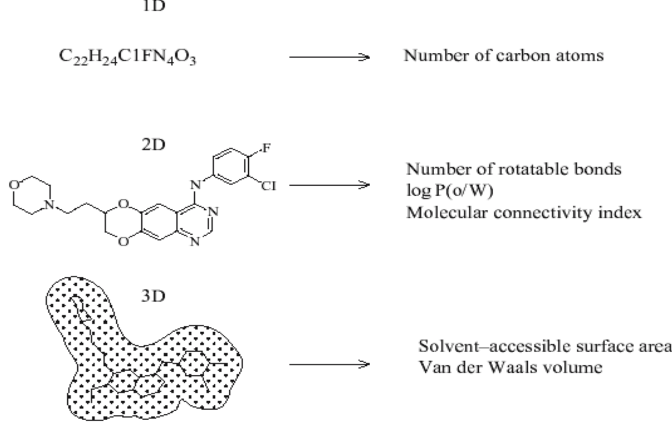
From chemoinformatics:theory, practice, &products (2007)

1.1. Molecular descriptors and chemical spaces

- The majority of cheminformatics methods depend on the generation of chemical reference spaces into which molecular data sets are projected and where analysis or design is carried out.
- The definition of chemical spaces critically depends on the use of computational descriptors of molecular structure, physical or chemical properties, or pharmacophores.
- Essentially, any comparison of molecular characteristics that goes beyond simple structural comparison requires the calculation of property values and the application of mathematical models.
- In chemical space design, each chosen descriptor adds a dimension to the reference space
- At least hundreds of molecular descriptors have been designed for chemical applications (Todeschini and Consonni, 2000)

Descriptor category	Examples
Physical properties	MW, logP
Atom and bond counts	Number of nitrogen atoms, aromatic rings, rotatable bonds, etc.
Pharmacophore features	Number of hydrogen bond acceptors, Sum of van der Waal surface areas of basic atoms
Charge descriptors	Total positive partial charge, dipole movement from partial charges
Connectivity and shape descriptors	Kier and Hall molecular shape indices
Surface area and volume	Solvent-accessible surface area

- Descriptors are frequently divided into 1D, 2D, or 3D descriptors, dependent on the dimensionality of the molecular representation from which they can be calculated
- The design and complexity of different types of descriptors often varies dramatically.
- Among very simple descriptors, for example, 2D structural fragments have high predictive value in many applications because they implicitly account for diverse molecular properties (such as complexity, polarity, hydrophobicity etc.).

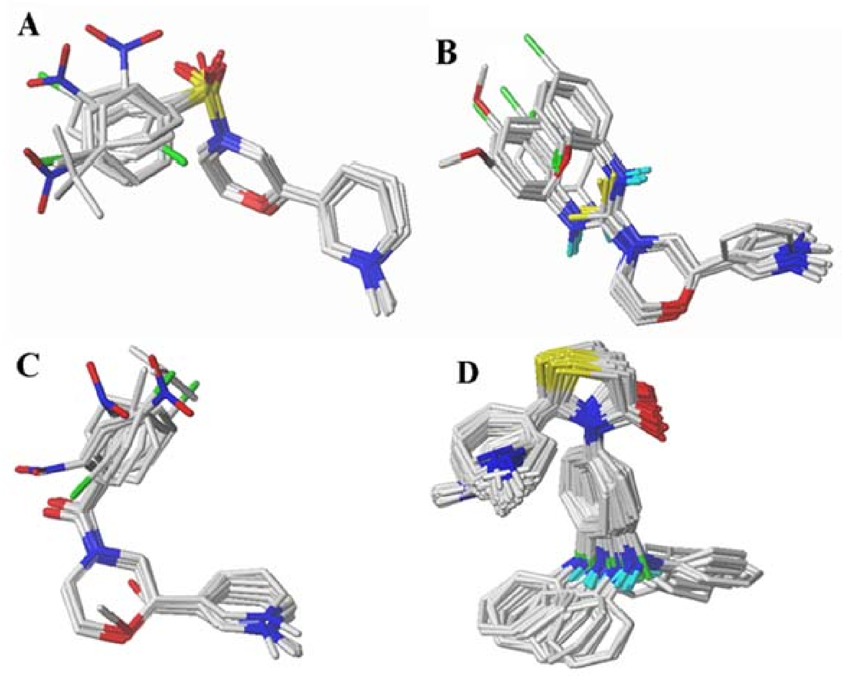


(Adapted from Bajorath 2002)

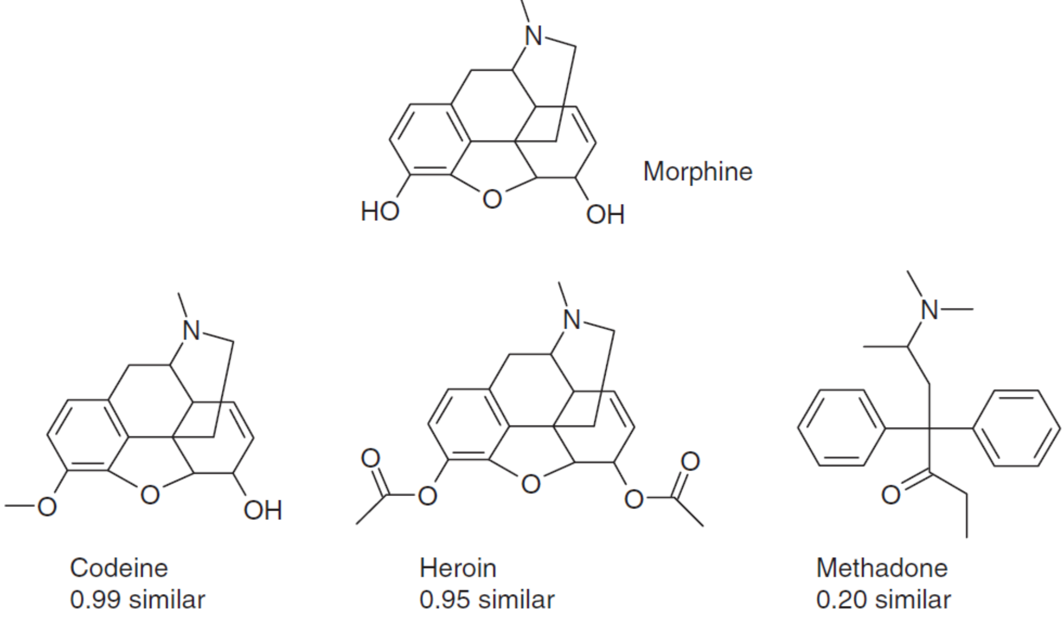
- These indices are also 2D descriptors because they require a molecular drawing (or graph) in order to determine the number of bonds.
- Other descriptor designs can become increasingly complex. Contributions of different types of descriptors can also be combined into composite formulations for example, descriptors combining molecular surface and charge information such as charged partial surface area (CPSA) descriptors (Stanton and Jurs 1990).

Chemical spaces and molecular similarity.

- There are no generally preferred descriptor spaces for cheminformatics applications
  - it is usually required to generate reference spaces for specific applications
  - on a case-by-case basis, either intuitively
  - based on experience or
  - by applying machine learning techniques to automate and optimize descriptor selection for a given problem.
- However, descriptors are ultimately selected for chemical space design
- *N* descriptors always produce an *n* dimensional reference space, into which compound sets can be mapped.
- In meaningful chemical space representations, similar compounds should map to similar regions or their intermolecular distance should be small.
- This represents a basic interpretation of the similarity concept.



Balaji*et al.,* 2012, (*communicated*)



1.2. Distance functions

Hamming distance	$HD = \sum_{i=1}^{n} x_i \oplus y_i,$
Euclidean distance	$ED = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$
Average distance	$AD = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} D_{ij}}{n(n-1)}$

- Here *ni* and *nj* are the number of descriptor values for molecules *i* and *j*, respectively,
- And *nij* is the number of common values.
- *Dij* is the distance between molecules *i* and *j*
- *D* the average distance, and *n* the total number of molecules.
- It should be noted that the general understanding of molecular similarity goes beyond simple structural similarity and extends to biological activity, in accord with the so-called Similar Property Principle (Johnson and Maggiora 1990)
- postulating that molecules having similar structures and properties should also exhibit similar activity

1.3. Similarity coefficients

Tanimoto coefficient	$Tc = n_{ij}/(n_i + n_j - n_{ij})$
Dice coefficient	$Dc = 2n_{ij}/(n_i + n_j)$
Cosine coefficient	$Cc = n_{ij}/(n_i n_j)^{1/2}$

- Thus, molecules that are located closely together in chemical reference space are often considered to be functionally related, which is one of the hallmarks of molecular similarity analysis.

Molecular similarity, dissimilarity, and diversity.

- As discussed, similar molecules can be identified by application of distance functions and analysis of nearest neighbors in chemical space.
- Diversity analysis, on the other hand, attempts to select different compounds from a given population
- This can also be accomplished using distance functions by only selecting compounds that are at least a pre-defined minimum distance away from others or – in diversity design – by trying to maximize average inter-compound distances.
- An alternative approach to diversity selection and design is to divide the descriptor axes into evenly spaced value intervals, a process called "binning", which produces *n*-dimensional subsections of chemical space.
- Molecular diversity is a global concept, which is applicable to the analysis of large compound distributions but not to the study of pair-wise molecular relationships. This is in contrast to molecular similarity analysis, which explores pair-wise relationships, the exploration of which is more local in nature. For example, one tries to find compounds similar to a given reference molecule or study the compound population within a limited region of chemical space.

- From this point of view, the inverse of molecular similarity is not diversity, but rather "dissimilarity", which is local in nature (addressing the question which molecule in a collection is most dissimilar from a given compound or set of compounds).
- Like similarity, dissimilarity calculations can focus on the exploration of pair-wise compound relationships (e.g., distances in chemical space).
- When similarity metrics are applied, the dissimilarity *d* between two molecules *i* and *j* is thus defined as, for example:

$$d_{ij} = 1 - Cc(i,j) \quad \text{or} \quad 1 - Tc(i,j)$$

- Dissimilarity analysis plays a major role in compound selection.
- Typical tasks include
  - the selection of a maximally dissimilar subset of compounds from a large set or
  - the identification of compounds that are dissimilar to an existing collection.

1.4. Drug-likeness

These play a major role in compound acquisition in the pharmaceutical industry.

- In pharmaceutical research, much emphasis has been put in recent years on the concept of "drug-likeness", which is based on the structural features and/or molecular properties.
- In fact, the rule-of-five is often cited in the context of drug-likeness. Lead-likeness
- Another recent trend has been to distinguish drug-likeness from lead-likeness (Rishton 2003).

This idea takes into consideration that virtual screening might rarely be capable of identifying mature drug candidates. Therefore, it should rather focus on the detection of “leads”, i.e. molecules that can be chemically optimized to ultimately become drugs.

Drug-like	Lead-like
MW < 500	MW < 350
ClogP < 5	ClogP < 3.0
Hydrogen bond donors < 5	Chemically stable
Hydrogen bond acceptors < 10	
Number of rotatable bonds ≤ 10	
PSA ≤ 140 $Å^2$	

1.5. Success stories

A successful example of the metamorphosis of a lead-like hit into a candidate for clinical development is the kinase inhibitor- **sorafenib**.

- A striking example of hits with more drug-like properties as starting point is the discovery of BAY 59-7939, a factor Xa inhibitor, in clinical trials.
- A total of only 4 minor modifications, such as removal or attachment of a halide, caused an almost 30,000-fold increase in activity, although the MW was increased by only 14Da.

References

1. Bajorath J, Nat. Rev. Drug Discov., 2002, 1, 882.
2. Bunin BA, Bajorath J, Siesel B, Morales G (EDs)chemoinformatics:theory, practice, &products (2007), springer
3. Johnson MA and Maggiora GM, in Concepts and Applications of Molecular Similarity, John Wiley & Sons, New York, 1990.
4. Rishton GM, Drug Discov Today.,2003, 8:86–96.
5. StantonDT andJurs PC, Anal. Chem., 1990, 62, 2323. Todeschini R and Consonni V, Handbook of Molecular Descriptors, Wiley-VCH, Weinheim, 2000.

Chapter 2

Homology Modeling

Mr. Fayaz S. M. & Rosana Babu

Introduction

Computational protein structure prediction is a process in which the three dimensional structure of proteins is predicted/built using *in silico* methods, based on principles of known protein structures derived through experimental techniques like X-Ray crystallography and NMR Spectroscopy and also from physical energy functions. There are mainly three methods available for *in silico* protein structure prediction. They are:

1. **Homology Modeling** – This method is used to predict structure of proteins from the already known protein structures that have more than 30% similarity to the unknown protein sequence.

2. **Threading/Fold Recognition** – This method is used to predict protein structures based on known protein folds of similar proteins.

3. ***Ab-initio* method** – This method is used to predict protein structures when structural information of similar proteins is not at all available. Protein structures are built from scratch using energy functions.

In this session, we would predict protein structures based on homology modeling method. Homology modeling has been a key component in structural bioinformatics for the prediction of three-dimensional structure of proteins from their sequences. This has become possible due to the availability of large number of experimentally determined protein structures in the protein structure databases. Still, many proteins whose structures are not yet identified are present and their structures are not available in these databases. The structures of such proteins can be built through homology modeling using the most similar protein structures available.

Though many tools and servers are available for homology modeling, the main steps in this process are almost the same as described below.

2.1. Methodology

a) Selection of Target Sequence (Optional)

This step depends on our necessity. The protein sequence that needs to be modeled is known as the "target sequence". Selection of the appropriate length of the target protein is important. Unwanted or unnecessary protein sequences may interfere with the process of template recognition. Sometimes, complete protein structure may not be required. In such cases, selecting the required protein sequences (domains) makes the process of modeling easy.

b) Recognition of Template Protein

The protein structure that is used for modeling our protein sequence (target protein) is known as "template protein". The target protein is paired with all the protein sequences of known protein structures, available in the protein structural databases, using simple sequence alignment programs. The protein having maximum identity with the target sequence is considered as the template protein.

c) Preparation of Template Protein (Optional)

This step also depends on our necessity. The template protein obtained from the protein structural database may contain many unwanted components like multiple chains, water molecules, ligands, etc. Therefore, preparation of the template protein is required for modeling.

d) Sequence alignment

The target and template protein sequences are aligned using a sequence alignment algorithm. This is a very important step in the process of protein modeling. The use of proper alignment algorithm is essential for obtaining a good modeled protein. The alignment compares the proteins and shows the identical regions in the proteins.

e) Secondary Structure Prediction

The secondary structure of the proteins is predicted using secondary structure prediction tools. The secondary structures of target and template proteins are compared and analyzed.

f) Building Homology Models

Using template protein, the homology models of the target protein are built. These homology models are visualized to check their 3D structure. If any loops are present in the structure, then they can be optimized further through loop modeling.

g) Loop Modeling

The loops present in the modeled protein are optimized using loop modeling. This is done to increase the accuracy of the structure.

h) Model Optimization

The modeled protein need to be optimized to its near native structure using energy minimization.

i) Model Validation

To assess the accuracy of the modeled protein, it is validated using various protein validation tools and servers. This validation shows whether our modeled protein is satisfactory or not.

2.2. Practical Tutorial

In this tutorial we are going to model a protein sequence through homology modeling using Prime module of Schrodinger software. This protein sequence has been arbitrarily chosen and modified to provide a case that best demonstrates various features of homology modeling. This modified protein sequence is considered as the target or query sequence.

[1] Open Maestro.

[2] Go to **Maestro** > **Change directory** > choose the folder **Homology Modeling Practical**.

[3] Choose **Applications** > **Prime** > **Structure Prediction** in the main window. The Structure Prediction panel opens.

[4] Click **From File** > select **Query.fasta** > click **Open**. The sequence is displayed in the Prime sequence viewer.

[5] Click **Next** to proceed to the next step. The next step is finding homologs or template proteins.

[6] We will use online web tools to identify the template protein(s). So, for time being, minimize the maestro window. Now, open your web browser and go to **Blast** tool web page. Choose **Protein Blast**.

[7] Paste the query sequence in the space provided. Change the search set database setting to **Protein Data Bank (pdb)**.

[8] Click **BLAST**.

[9] From the results, we could see that the protein with PDB Id 2CUN is having highest similarity. We will consider this as our template protein.

[10] Go to PDB and download the protein with PDB Id 2CUN in to the **Homology Modeling Practical** folder. Now, we will carry out the preparation of template protein.

[11] Open another maestro. Go to **Workflows** > choose **Protein preparation Wizard**.

[12] Select **Assign bond orders, Add hydrogens, Treat disulfides, Delete waters, Fill loops**, etc.

[13] Click **Preprocess**.

[14] Select all the displayed molecules except **chain A** > Click **Delete Selected**.

[15] Select **Exhaustive sampling** > click **Optimize**.

[16] Click **Minimize**.

[17] Go to **Project** > **Show Table**. Save the Protein in the **Homology Modeling Practical** folder as **Temp.pdb**. Close the maestro that has been used for template preparation.

[18] Now, open the maestro that is being used for Prime structure prediction.

[19] Under **Find Homologs,** click on **Import** > choose **Temp.pdb** > click **open**. The template protein will be imported in to the structure prediction panel.

[20] Under Homologs table, click on **Temp.pdb.** The template protein is now shown in the sequence viewer. The target and template protein sequences are aligned to each other.

[21] Click **Next** to proceed to the next step. The next step is secondary structure prediction.

[22] Click **Run SSP**. When the SSP job finishes, the secondary structure predictions of the query are displayed in the sequence viewer.

[23] Click **Next** to proceed to the Build Structure step.

[24] Click **Build**. When the job finishes, the model is displayed in the Workspace superimposed on the template.

[25] When the job finishes, click **Add to Project Table**. The Project Table panel opens with the selected structure as an entry. The title of the structure is **Query**.

[26] Close the Structure Prediction panel.

[27] Click the **In** column in the Project Table panel for this entry. The predicted structure is displayed in the Workspace.

To improve the structure most efficiently, we need to refine the areas of the structure that are likely to be problematic. In general terms, this means refining loops and re-predicting side-chain conformations.

[28] From the main window choose **Applications** > **Prime** > **Refinement**.

[29] Select **Refine loops** in the Task.

[30] Click **Load from Workspace**.

[31] Click the check box for loop6 in the Run column.

[32] Click **Start**. The Refinement - Start dialog box opens.

[33] Choose **Append new entries as a new group** from the “Incorporate” option menu.

[34] Enter **Loop Refinement** in the Name text box.

[35] Click **Start** to launch the job. The Monitor panel is displayed and lists the log file. When the job finishes, the predicted structure is incorporated into the Project Table and is displayed in the Workspace.

[36] This modeled protein need to be optimized. For this, it is energy minimized using protein preparation wizard. Save the minimized protein structure as **Query_min.pdb** in the **Homology Modeling Practical** folder.

[37] Finally, the protein model needs to be validated. This is done using Structural Analysis and Verification Server (SAVS).

[38] Open the web browser and go to the link **http://nihserver.mbi.ucla.edu/SAVES/**

[39] Click **Choose File** > select **Query_min.pdb** from **Homology Modeling Practical** folder > click **Open** > Click **Upload Files.**

[40] Check the boxes of all the programs > click **Run all programs.**

[41] Analyze the Ramachandran plot. We see that majority of the residues are in the Allowed region. This shows that we have built a decent structure for the query protein sequence.

[42] Also, open the template and query proteins in the maestro workspace > go to **Tools** > run **Protein Structure Alignment** option. Check the RMSD value. We see that this value is low and it shows that our modeled protein is aligning with the template protein to a greater extent and so, our modeled structure is accurate.

For queries and comments, mail to **fayazbioinfo@gmail.com**

2.3. Other Software's for Homology Modeling

1. Swiss-PdB Viewer

Aim

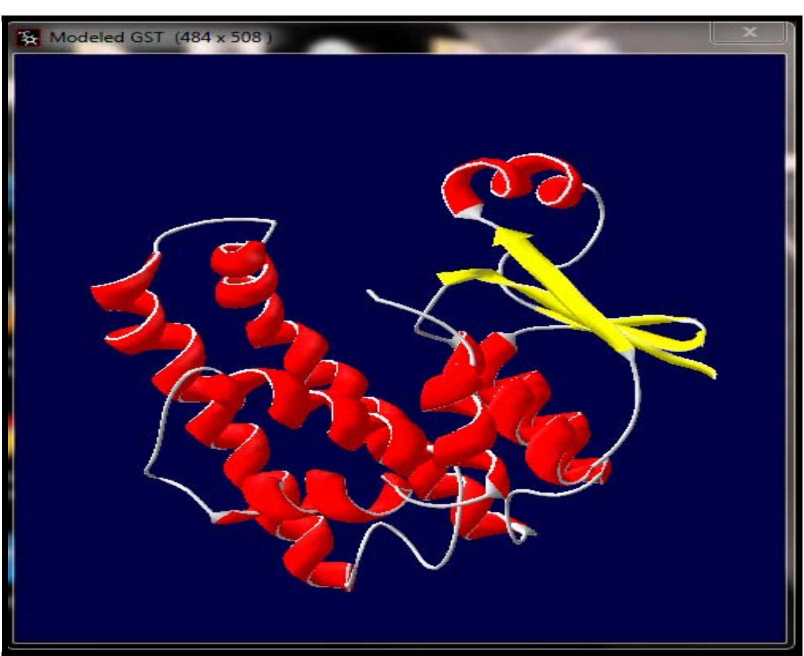
To do homology modeling of protein is using Swiss-PdB Viewer.

Procedure

1. Download protein sequence for *Meloidogyne incognita* Glutathione- S Transferases (GST) from NCBI in fasta format and save it in notepad
2. Open the Swiss-PdB Viewer and choose 'Load raw sequence to mode' item from the Swiss-Model menu.
3. Template for *M. incognita* GST can be obtained by choosing the "find appropriate ExPDB templates" item of the "Swiss Model" menu. This automatically launched a web browser, which contained the sequence in the FASTA format.
4. Click the submit button. A page containing the best available templates was shown of which 1zL9 was chosen.
5. Download the PDB template files from RCSB PDB http://www.rcsb.org/pdb/, and open in SwissPdB Viewer.

6. Perform magic fit and then the iterative magic fit for the raw sequence and the template. This results in structural overlay of the query on to the template.
7. Click on ‘submit modeling request’ from the Swiss-model menu.
8. A Swiss-PdB Viewer project file was obtained with our model aligned onto templates, ready for comparison. The modeled protein comes back at the e-mail address provided. Download it and
9. Open the modeled protein in Swiss-PdB Viewer, select Ribbon view from “Load Protien” in Preferences menu, and select “Render in 3D” from Display menu.

Output:



2. Modeller9v1

MODELLER is used for homology or comparative modeling of protein three-dimensional structures. The user provides an alignment of a sequence to be modeled with known related structures and MODELLER automatically calculates a model containing all non-hydrogen atoms. There are 5 modeling examples that the user can follow:

1. **Basic Modeling**. Model a sequence with high identity to a template. This exercise introduces the use of MODELLER in a simple case where the template selection and target-template alignments are not a problem.
2. **Advanced Modeling**. Model a sequence based on multiple templates and bound to a ligand.
   This exercise introduces the use of multiple templates, ligands and loop refinement in the process of model building with MODELLER.
3. **Iterative Modeling**. Increase the accuracy of the modeling exercise by iterating the 4 step process.
   This exercise introduces the concept of MOULDING to improve the accuracy of comparative models.
4. **Difficult Modeling**. Model a sequence based on a low identity to a template. This exercise uses resources external to MODELLER in order to select a template for a difficult case of protein structure prediction.
5. **Modeling with cryo-EM.** Model a sequence using both template and cryo-EM data. This exercise assesses the quality of generated models and loops by rigid fitting into cryo-EM maps, and improves them with flexible EM fitting.

Procedure for Basic Modeling

Example: Modeling lactate dehydrogenase from *Trichomonas vaginalis* based on a single template.

The individual modeling steps of this example are explained below.

1. **Download the target protein sequence from NCBI with accession** AAC72735.1**.**
2. **Target sequence preparation**
   - Modeller uses a special form of the PIR format where information about sequence numbering and chain codes are written into the 'description' line between the PIR protein tag and the protein alignment entry. Jalview will attempt to parse any PIR entries

conforming to the Modeller/PIR format, in order to extract the sequence start and end numbering and (possibly) a PDB file reference. The description line information is always stored in the sequence description string - so no information is lost if this parsing process fails.

- Prepare the target sequence in PIR format using (http://pir.georgetown.edu/pirwww/search/multialn.shtml) which is readable format for MODELLER and saved as **.ALI** file.

>P1;*Primary_Sequence_ID sequence or structureX*:*pdb-reference if available*:*start residue*:*start chain code*:*end residue*:*end chain code*:. *description text*

- The first field is either sequence or structureX, depending upon the presence of a PDB database ID for the sequence. If the protein has no PDB reference, then the chain code is not specified, unless one already existed when the sequence was imported into Jalview.

3. Searching for structures related to TvLDH

First, it is necessary to put the target TvLDH sequence into the PIR format readable by MODELLER (file "TvLDH.ali").

```
File: TvLDH.ali
>P1;TvLDH
sequence:TvLDH::::::0.00: 0.00
MSEAAHVLITGAAGQIGYILSHWIASGELYGDRQVYLHLLDIPPAMNRLTALTMELEDCAFPHLAGFVATTDPKA
AFKDIDCAFLVASMPLKPGQVRADLISSNSVIFKNTGEYLSKWAKPSVKVLVIGNPDNTNCEIAMLHAKNLKPEN
FSSLSMLDQNRAYYEVASKLGVDVKDVHDIIVWGNHGESMVADLTQATFTKEGKTQKVVDVLDHDYVFDTFFKKI
GHRAWDILEHRGFTSAASPTKAAIQHMKAWLFGTAPGEVLSMGIPVPEGNPYGIKPGVVFSFPCNVDKEGKIHVV
EGFKVNDWLREKLDFTEKDLFHEKEIALNHLAQGG*
```

The first line contains the sequence code, in the format ">*P1;code*". The second line with ten fields separated by colons generally contains information about the structure file, if applicable. Only two of these fields are used for sequences, "*sequence*" (indicating that the file contains a sequence without known structure) and "*TvLDH*" (the model file name). The rest of the file contains the sequence of TvLDH, with "*" marking its end. The standard one-letter amino acid codes are used.

A search for potentially related sequences of known structure can be performed by the **profile.build**() command of MODELLER. Run “ mod9.10 build_profile.py”

1. Selecting a template

The output of the "build_profile.py" script is written to the "build_profile.log" file. Errors and warnings in log files can be found by searching for the "*_E>*" and "*_W>*" strings, respectively. MODELLER also writes the profile in text format to the "build_profile.prf" file. An extract (omitting the aligned sequences) of the output file can be seen next. The first 6 commented lines indicate the input parameters used in MODELLER to build the profile. Subsequent lines correspond to the detected similarities by profile.build().

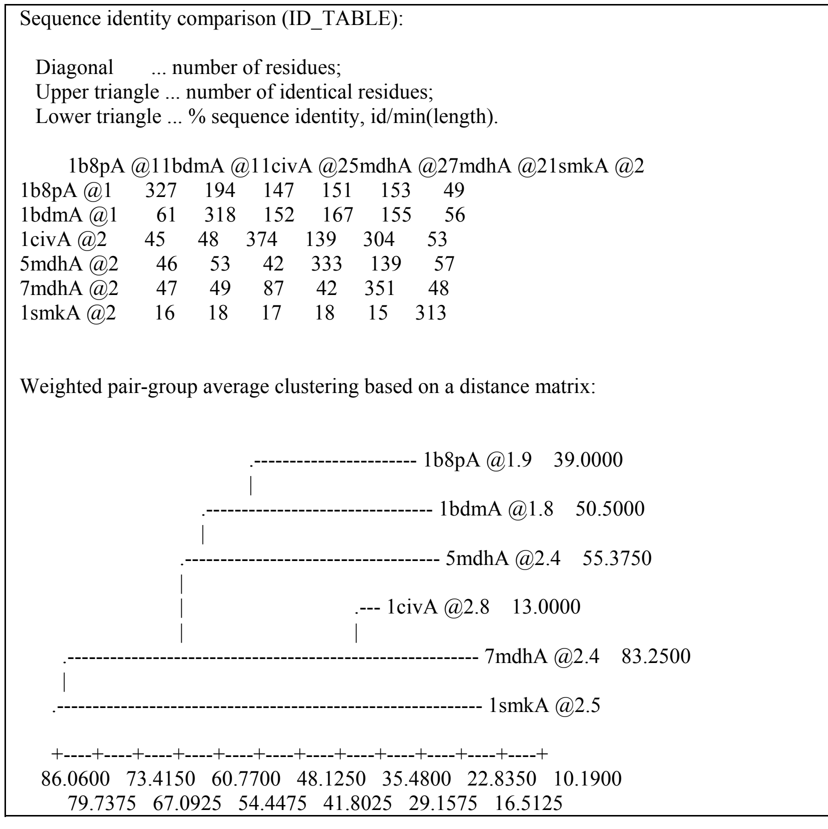
```
# Number of sequences:     30
# Length of profile  :    335
# N_PROF_ITERATIONS  :      1
# GAP_PENALTIES_1D   :   -900.0   -50.0
# MATRIX_OFFSET      :    0.0
# RR_FILE            : ${MODINSTALL8v1}/modlib//as1.sim.mat
   1 TvLDH     S    0  335    1  335    0    0    0   0.   0.0
   2 1a5z      X    1  312   75  242   63  229  164  28.   0.83E-08
   3 1b8pA     X    1  327    7  331    6  325  316  42.   0.0
   4 1bdmA     X    1  318    1  325    1  310  309  45.   0.0
   5 1t2dA     X    1  315    5  256    4  250  238  25.   0.66E-04
   6 1civA     X    1  374    6  334   33  358  325  35.   0.0
   7 2cmd      X    1  312    7  320    3  303  289  27.   0.16E-05
   8 1o6zA     X    1  303    7  320    3  287  278  26.   0.27E-05
   9 1ur5A     X    1  299   13  191    9  171  158  31.   0.25E-02
  10 1guzA     X    1  305   13  301    8  280  265  25.   0.28E-08
  11 1gv0A     X    1  301   13  323    8  289  274  26.   0.28E-04
  12 1hyeA     X    1  307    7  191    3  183  173  29.   0.14E-07
  13 1i0zA     X    1  332   85  300   94  304  207  25.   0.66E-05
  14 1i10A     X    1  331   85  295   93  298  196  26.   0.86E-05
  15 1ldnA     X    1  316   78  298   73  301  214  26.   0.19E-03
  16 6ldh      X    1  329   47  301   56  302  244  23.   0.17E-02
  17 2ldx      X    1  331   66  306   67  306  227  26.   0.25E-04
  18 5ldh      X    1  333   85  300   94  304  207  26.   0.30E-05
  19 9ldtA     X    1  331   85  301   93  304  207  26.   0.10E-05
  20 1llc      X    1  321   64  239   53  234  164  26.   0.20E-03
  21 1lldA     X    1  313   13  242    9  233  216  31.   0.31E-07
  22 5mdhA     X    1  333    2  332    1  331  328  44.   0.0
  23 7mdhA     X    1  351    6  334   14  339  325  34.   0.0
  24 1mldA     X    1  313    5  198    1  189  183  26.   0.13E-05
  25 1oc4A     X    1  315    5  191    4  186  174  28.   0.18E-04
  26 1ojuA     X    1  294   78  320   68  285  218  28.   0.43E-05
  27 1pzgA     X    1  327   74  191   71  190  114  30.   0.16E-06
  28 1smkA     X    1  313    7  202    4  198  188  34.   0.0
  29 1sovA     X    1  316   81  256   76  248  160  27.   0.93E-03
  30 1y6jA     X    1  289   77  191   58  167  109  33.   0.32E-05
```

The most important columns in the **profile.build()** output are the second, tenth, eleventh and twelfth columns. The second column reports the code of the PDB sequence that was compared with the target sequence. The PDB code in each line is the representative of a group of PDB sequences that share 95% or more sequence identity to each other and have less than 30 residues or 30% sequence length difference. The eleventh column reports the percentage sequence identities between TvLDH and a PDB sequence normalized by the lengths of the alignment (indicated in the tenth column). In general, a sequence identity value above approximately 25%

indicates a potential template unless the alignment is short (i.e., less than 100 residues). A better measure of the significance of the alignment is given in the twelfth column by the e-value of the alignment. In this example, six PDB sequences show very significant similarities to the query sequence with e-values equal to 0. Hits with maximum identity are 1bdm:A, 5mdh:A, 1b8p:A, 1civ:A, 7mdh:A, and 1smk:A

To select the most appropriate template for our query sequence over the six similar structures, we will use the **alignment.compare_structures()** command to assess the structural and sequence similarity between the possible templates (Run "compare.py").

Finally, the **id_table** command writes a file with pairwise sequence distances that can be used directly as the input to the **dendrogram** command (or the clustering programs in the PHYLIP package). **dendrogram** calculates a clustering tree from the input matrix of pairwise distances, which helps visualizing differences among the template candidates. Excerpts from the log file are shown below (file "compare.log").



```
Sequence identity comparison (ID_TABLE):

   Diagonal       ... number of residues;
   Upper triangle ... number of identical residues;
   Lower triangle ... % sequence identity, id/min(length).

         1b8pA @11bdmA @11civA @25mdhA @27mdhA @21smkA @2
1b8pA @1     327    194    147    151    153     49
1bdmA @1      61    318    152    167    155     56
1civA @2      45     48    374    139    304     53
5mdhA @2      46     53     42    333    139     57
7mdhA @2      47     49     87     42    351     48
1smkA @2      16     18     17     18     15    313


Weighted pair-group average clustering based on a distance matrix:


                                .---------------------- 1b8pA @1.9    39.0000
                                |
                           .------------------------------- 1bdmA @1.8    50.5000
                           |
                        .----------------------------------- 5mdhA @2.4    55.3750
                        |
                        |                         .--- 1civA @2.8    13.0000
                        |                         |
      .-------------------------------------------------------- 7mdhA @2.4    83.2500
      |
     .---------------------------------------------------------- 1smkA @2.5

     +----+----+----+----+----+----+----+----+----+----+----+----+
   86.0600   73.4150   60.7700   48.1250   35.4800   22.8350   10.1900
       79.7375   67.0925   54.4475   41.8025   29.1575   16.5125
```

The comparison above shows that 1civ:A and 7mdh:A are almost identical, both sequentially and structurally. However, 7mdh:A has a better crystallographic resolution (2.4Å versus 2.8Å), eliminating 1civ:A. A second group of structures (5mdh:A, 1bdm:A, and 1b8p:A) share some

similarities. From this group, 5mdh:A has the poorest resolution leaving for consideration only 1bdm:A and 1b8p:A. 1smk:A is the most diverse structure of the whole set of possible templates. However, it is the one with the lowest sequence identity (34%) to the query sequence. We finally pick 1bdm:A over 1b8p:A and 7mdh:A because of its better crystallographic R-factor (16.9%) and higher overall sequence identity to the query sequence (45%).

2. Aligning TvLDF with the template

Run “align2d.py”

The **align2d()** command is executed to align the two sequences. Finally, the alignment is written out in two formats, PIR ("TvLDH-1bdmA.ali") and PAP ("TvLDH-1bdmA.pap"). The PIR format is used by MODELLER in the subsequent model building stage, while the PAP alignment format is easier to inspect visually.

```
>P1;Q93Z60_ARATH
sequence:Q93Z60_ARATH:1:.:118:.:.
----MASTALSSAIVSTSFLRRQQTPISLRSLPFANT-QSLFGLKS-STARGGRVTAMATYKVKFITPEGEQ
EVECEEDVYVLDAAEEAGLDLPYSCRAGSCSSCAGKVVSGSIDQSD------QSFLD-D------------
---------------------*
>P1;PDB|1FER|_
structureX:1FER:1:.:105:.:.
--------------------------------------------------AFVVTDNCIKCKY---TDCV
EV-CPVDCFY----EGPNFLVIHPDECIDCALCEPECPAQAIFSEDEVPEDMQEFIQLNAELAEVWPNITEK
KDPLPDAEDWDGVKGKLQHLE*
```

3. Model building

Once a target-template alignment is constructed, MODELLER calculates a 3D model of the target completely automatically, using its **automodel** class. The following script will generate five similar models of TvLDH based on the 1bdm:A template structure and the alignment in file "TvLDH-1bdmA.ali" (Run "model-single.py").

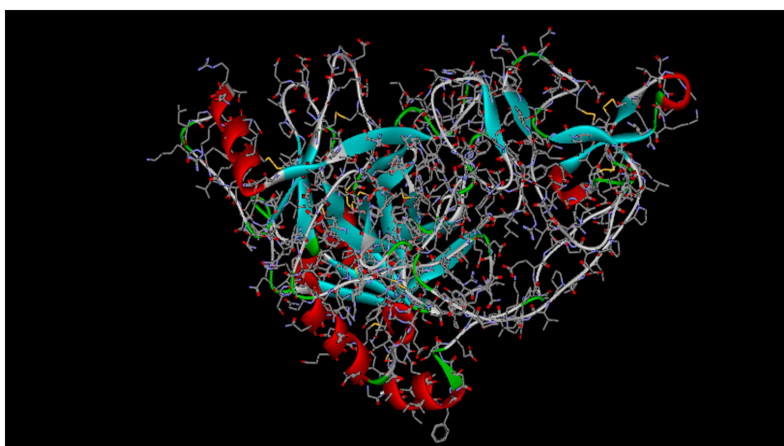
The most important output file is "model-single.log", which reports warnings, errors and other useful information including the input restraints used for modeling that remain violated in the final model. The last few lines from this log file are shown below.

```
>> Summary of successfully produced models:
Filename                    molpdf     DOPE score    GA341 score
----------------------------------------------------------------------
TvLDH.B99990001.pdb        1763.56104  -38079.76172       1.00000
TvLDH.B99990002.pdb        1560.93396  -38515.98047       1.00000
TvLDH.B99990003.pdb        1712.44104  -37984.30859       1.00000
TvLDH.B99990004.pdb        1720.70801  -37869.91406       1.00000
TvLDH.B99990005.pdb        1840.91772  -38052.00781       1.00000
```

As you can see, the log file gives a summary of all the models built. For each model, it lists the file name, which contains the coordinates of the model in PDB format. The models can be viewed by any program that reads the PDB format, such as Chimera. The log also shows the score(s) of each model.

4. Model evaluation

If several models are calculated for the same target, the "best" model can be selected in several ways. For example, you could pick the model with the lowest value of the MODELLER objective function or the DOPE assessment score, or with the highest GA341 assessment score, all of which are reported in the log file, above.
The file "evaluate_model.py" evaluates an input model with the DOPE potential



Secondary structure of modeled protein

References

1. Syam B. Nair, Fayaz SM, Rajanikant G. K. In silico prediction of novel inhibitors of the DNA binding activity of FoxG1. Med Chem. 2012;8(6):1155-62.
2. User manual Prime, version 2.1, Schrödinger, LLC, New York, NY, 2008.
3. Hooft RW, Sander C, Vriend G (1997) Objectively judging the quality of a protein structure from a Ramachandran plot. Comput Appl Biosci 13:425-430.

Chapter 3

Molecular docking & Drug designing

Mr. Pradeep H. & Rosana Babu

Introduction

Molecular docking is an important computational technique in structural biology and computer-aided drug design. The major objective of molecular docking is to evaluate the feasible binding geometries of a putative ligand with a target protein of known three-dimensional structure. The binding geometries are often known as binding poses, which include in principle, both the position of the ligand relative to the receptor and conformational state of the ligand and the receptor. There are three basic tasks any docking procedure must accomplish: (1) characterization of the binding site, (2) positioning of the ligand into the binding site (orienting), and (3) evaluating the strength of interaction for a specific ligand-receptor complex ("scoring").

Docking can be used to perform virtual screening on large libraries of compounds, rank the results, and propose structural hypotheses of how the ligands inhibit the target, which is invaluable in lead optimization. The docking is further useful as second step of the pharmacophoric screening for the identification of molecules that can fit into the pharmacophore (since, docking can exclude false hits having matching pharmacophoric sites but sterically hindered by some of the residues in the active site).

The major types of docking are-

- **Lock and Key or Rigid Docking** – In rigid docking, both the internal geometry of the receptor and ligand is kept fixed during docking.
- **Induced fit or Flexible Docking -** In this model, the ligand is kept flexible and the energy for different conformations of the ligand fitting into the protein is calculated. Though more time consuming, this method can evaluate many different possible conformations which make it more reliable.

Molecular docking can be thought of as a problem of *"lock-and-key"*, where one is interested in finding the correct relative orientation of the *"key"* which will open up the *"lock"* (where on the surface of the lock is the key hole, which direction to turn the key after it is inserted, etc.). Here, the protein can be thought of as the "lock" and the ligand can be thought of as a "key". Molecular docking may be defined as an optimization problem, which would describe the "best-fit" orientation of a ligand that binds to a particular protein of interest. However, since both the ligand and the protein are flexible, a *"hand-in-glove"* analogy is more appropriate than *"lock-and-key"*. During the course of the process, the ligand and the protein adjust their conformation to achieve an overall "best-fit" and this kind of conformational adjustments resulting in the overall binding is referred to as "induced-fit".

The focus of molecular docking is to computationally simulate the molecular recognition process. The aim of molecular docking is to achieve an optimized conformation for both the protein and ligand and relative orientation between protein and ligand such that the free energy of the overall system is minimized.

3.1. Mechanics of docking

To perform a docking screen, the first requirement is a structure of the protein of interest. Usually the structure has been determined using a biophysical technique such as x-ray crystallography, or less often, NMR spectroscopy. This protein structure and a database of potential ligands serve as inputs to a docking program. The success of a docking program depends on two components: the search algorithm and the scoring function.

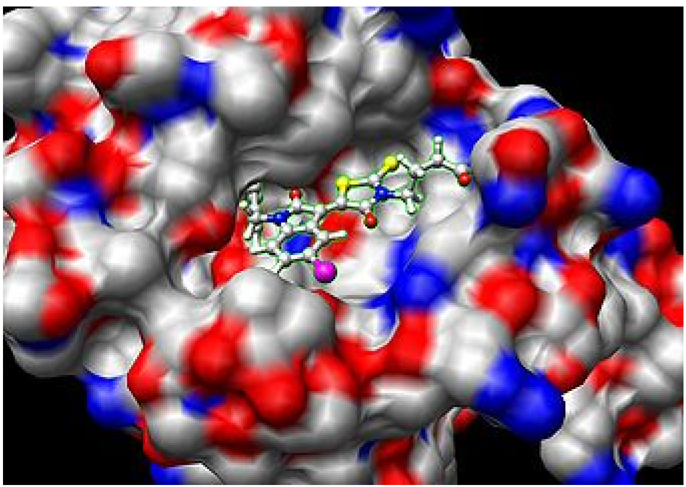


Fig 1. Small molecule **docked to a** protein**.**

a) Search algorithm

The search space in theory consists of all possible orientations and conformations of the protein paired with the ligand. However in practice with current computational resources, it is impossible to exhaustively explore the search space—this would involve enumerating all possible distortions of each molecule (molecules are dynamic and exist in an ensemble of conformational states) and all possible rotational and translational orientations of the ligand relative to the protein at a given level of granularity. Most docking programs in use account for a flexible ligand, and several attempt to model a flexible protein receptor. Each "snapshot" of the pair is referred to as a pose.

A variety of conformational search strategies have been applied to the ligand and to the receptor. These include:

- systematic or stochastic torsional searches about rotatable bonds
- molecular dynamics simulations
- genetic algorithms to "evolve" new low energy conformations

b) Ligand flexibility

Conformations of the ligand may be generated in the absence of the receptor and subsequently docked or conformations may be generated on-the-fly in the presence of the receptor binding cavity, or with full rotational flexibility of every dihedral angle using fragment based docking. Force field energy evaluation are most often used to select energetically reasonable conformations, but knowledge-based methods have also been used.

c) Receptor flexibility

Computational capacity has increased dramatically over the last decade making possible the use of more sophisticated and computationally intensive methods in computer-assisted drug design. However, dealing with receptor flexibility in docking methodologies is still a thorny issue. The main reason behind this difficulty is the large number of degrees of freedom that have to be considered in this kind of calculations. Neglecting it, however, leads to poor docking results in terms of binding pose prediction.

Multiple static structures experimentally determined for the same protein in different conformations are often used to emulate receptor flexibility. Alternatively rotamer libraries of amino acid side chains that surround the binding cavity may be searched to generate alternate but energetically reasonable protein conformations.

d) Scoring function

The scoring function takes a pose as input and returns a number indicating the likelihood that the pose represents a favorable binding interaction. Most scoring functions are physics-based molecular mechanics force fields that estimate the energy of the pose; a low (negative) energy indicates a stable system and thus a likely binding interaction. An alternative approach is to derive a statistical potential for interactions from a large database of protein-ligand complexes, such as the Protein Data Bank, and evaluate the fit of the pose according to this inferred potential. There are a large number of structures from X-ray crystallography for complexes between proteins and high affinity ligands, but comparatively fewer for low affinity ligands as

the later complexes tend to be less stable and therefore more difficult to crystallize. Scoring functions trained with this data can dock high affinity ligands correctly, but they will also give plausible docked conformations for ligands that do not bind. This gives a large number of false positive hits, i.e., ligands predicted to bind to the protein that actually doesn't when placed together in a test tube. One way to reduce the number of false positives is to recalculate the energy of the top scoring poses using (potentially) more accurate but computationally more intensive techniques such as Generalized Born or Poisson-Boltzmann methods.

3.2. Major steps employed in docking

a) Protein structure preparation

Protein preparation process involves. (1) Obtaining the 3D-structures of protein (either Protein Data Bank structures or homology models) (2) Addition of H-atoms to the proteins, including the protons necessary to define the correct ionization and tautomeric states of amino acid residues such as Asp, Ser, Glu, Arg and His. (3) Correcting the missing side chains or atoms of the residues. (4) Minimization of the protein structure, using the suitable force field to alleviate steric clashes that may exist in the structures [1].

Finally, prepared protein structures will be employed in the generation of grids for the subsequent docking calculations.

b) Ligand preparation

Preparing the ligand involves ensuring (1) that its atoms are assigned with the correct atom types. (2) The generation of all possible ionization states at pH 7.0±2.0 with ionizer. (3) Generation of tautomers and chiralities for all the ionization states [2].

c) Docking methodology

The docking study mainly depends on two components: the search algorithm and, the scoring function.

(1) The search algorithm is a tool which extensively searches all possible conformations

of the ligand (the paired ligand and protein) in a 3D-space (the binding site of interest). Further, the program will calculate the energy for each rotation made of each and every rotatable bond it can find. Each energy value calculated will be presented as a "binding pose".

The different types of algorithms that can be used for docking analysis are given below.

- Molecular dynamics
- Monte Carlo methods
- Fragment-based methods
- Point complementary methods
- Distance geometry methods

(2) The scoring function is a process which accepts a binding energy and designates a number or score to indicate the strength and likelihood of the binding interaction. Finally, this score will be applied to select the best conformation of the ligand [3, 4].

- Empirical scoring function of Igemdock

  Fitness = vdW + Hbond + Elec
- Binding Energy

$$\Delta G_{bind} = \Delta G_{vdw} + \Delta G_{hbond} + \Delta G_{elect} + \Delta G_{conform} + \Delta G_{tor} + \Delta G\ s$$

3.3. Applications of Molecular Docking

A binding interaction between a small molecule ligand and an enzyme protein may result in activation or inhibition of the enzyme. If the protein is a receptor, ligand binding may result in agonism or antagonism. Docking is most commonly used in the field of drug design — most drugs are small organic molecules, and docking may be applied to:

- **Hit identification** – docking combined with a scoring function can be used to quickly screen large databases of potential drugs in silico to identify molecules that are likely to bind to protein target of interest.

- **Lead optimization** – docking can be used to predict in where and in which relative orientation a ligand binds to a protein (also referred to as the binding mode or pose). This information may in turn be used to design more potent and selective analogs.
- **Bioremediation** – Protein ligand docking can also be used to predict pollutants that can be degraded by enzymes.

3.4. Practical Tutorial

Tutorial: Flexible Docking with Glide

Any meaningful docking calculations require a careful preparation of receptor and ligand structures before the docking programs can do their work. With Schrodinger Suite of programs, the bulk of receptor preparation is carried out with the **Protein Preparation Wizard** while the ligand preparation is handled by **Ligand Preparation Wizard**. This tutorial illustrates docking of the inhibitors into cyclin dependent kinase 2 (CDK2) protein.

1. Importing Molecules into the Maestro

1. Start the Maestro molecular visualization environment. **Maestro** serves as the interface to various modeling programs, such as the **Protein Preparation Wizard** and **Glide**. Maestro offers quick access to most of these programs via the **Applications** menu. In addition, common task, such as preparation of structures for docking are accessible via the **Workflows** menu.
2. From the **Project** menu, select **Get PDB**. Enter 1E9H for the PDB ID and hit **Download** to import in the 2.5 angstrom structure of CDK2 with bound inhibitor.
3. Maestro will perform the initial check for you, and display the results in color. Residues in grey are OK. If residues in cyan are missing their neighboring residues, which are not uncommon in case of flexible loops; for rigorous docking studies one would attempt to model these loops. Residues in orange are non-standard residues that oftentimes need to

be dealt with before docking. Rotate the structure by dragging while pressing the middle (wheel) button. Notice that inhibitor in the protein is in orange color.

2. Preparing Protein Structure for Docking

1. In order to be used as a receptor for docking, protein structures should be processed. Some of the typical operations include (a) addition of hydrogen atoms, (b) assignment atomic charges, (c) elimination of water molecules that are not involved in ligand binding, (d) replacement of selenocysteins with cysteins, and (e) Fixing the missing side chains or loops. Schrodinger Suite automates this work, and we can carry out the necessary preparation by launching the **Protein Preparation Wizard** from the **Workflows** menu.

2. Examine the binding pocket around the ligand and notice that few water molecules are interacting with the protein complex. Water molecules which are not interacting with ligand have to be deleted.

3. Hit **Preprocess**, The preprocessing should take less than a minute to complete.

4. Under the **Review and Modify** tab, you can observe that CDK2 is composed of four chains, 4 water molecules, and the structure also contains two inhibitors. Eliminate one set of chains by selecting and deleting C & D chains.

5. Generate alternative tautomeric states of protein residues by pressing the **Generate States** button.

6. Under the **Refine** tab, keep the options **Sampling water orientations** checked and hit **Optimize...**, then **Start**. The optimization will take about a minute. Notice that histidines are now explicitly labeled as HIS, HID, or HIE, and that some side chains carry a Flip label.

7. Under the **Refine** tab, hit **Minimize...** then **Start** to minimize the structure. The minimization will take nearly 10 minutes.

3. Preparation of the Ligands to be docked

1. The ligand should also be prepared with the same force field that the receptor has been prepared. Import the ligand file (ligand.mae which can be seen in your directory) into maestro.
2. In **Maestro**, under **Applications**, choose **LigPrep....** to open the tool that prepares ligands. In the **use structures from** line, hit **Workspace (included entries)....**
3. Use default options for the ionization state, desalt, or generate tautomers in this case.
4. Hit **Start ...** and give a meaningful **Name** for your job (e.g. Ligand_CDK). The ligand preparation will take less than a minute, and the prepared ligand is now in the file with .maegz extension.

4. Generating the Protein Grid

1. Glide is a grid-based docking program. So you need to define a grid on which you expect the ligand binds (usually the experimental binding site). Under **Applications**, select **Glide**, then **Receptor Grid Generation....** Doing so, a new window titled **Receptor Grid Generation** will pop up
2. Make sure that **Pick to identify ligand** is checked, and click on any atom in the ligand. Notice that green boxes surround selected ligand atoms.
3. Click on the **Site** tab and briefly examine the rectangular region for which the grid is generated.
4. Then hit **Start**, give a meaningful name (e.g. Grid_CDK) for your job, and hit **Start** one more time to start the grid generation. The grid generation takes about 5-10 minutes.

5. Docking with Glide

1. Under **Applications**, click **Glide**, then **Ligand Docking...** A new window titled **Ligand Docking** opens.

2. Hit **Settings** tab, hit **Browse** and select the grid file you generated earlier (Grid_CDK). Select **Dock Flexible**.
3. Three types of score functions are used in glide:

   HTVS: High Throughput Virtual Screening that is a very fast evaluation method for very large libraries.

   SP: Standard-Precision that is a quick way to evaluate the poses. SP is also used for high-throughput virtual screening of big libraries.

   XP: Extra-Precision that is a refinement tool used only for very small libraries or a good subset of big library that has been pre-evaluated with HTVS or SP score, because it needs more computational time.

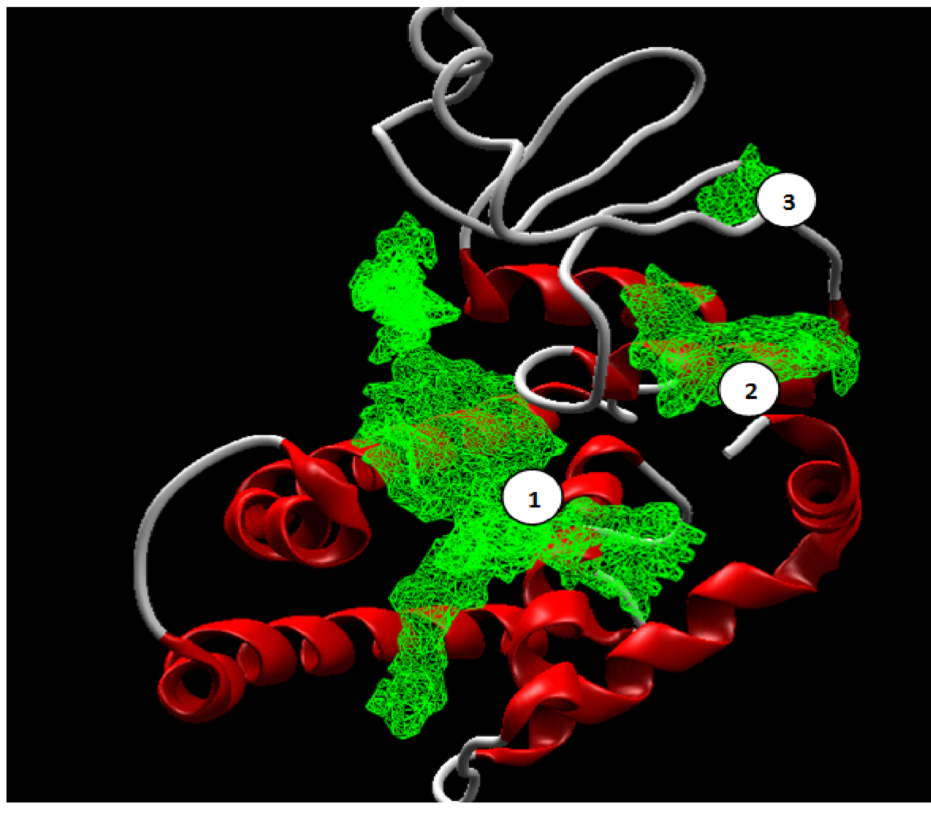
   Here we will use Extra-Precision scoring function, check **XP** in the **Precision** section of **settings** tab**.**
4. Under **Ligands** tab, hit **Browse** and select the prepared ligand file.
5. Hit **Start** and give a meaningful name (e.g. Glide_CDK_Flexible) to your job. Hit **Start** once more to start the docking calculation. The Extra-Precision flexible docking will complete in a few minutes.
6. Under **Project**, select **Show Table**. A new window with results opens. The set of entries can be seen with your receptor (without the ligand) and the docked ligand poses. You can use **Ctrl** key on the keyboard to visualize multiple entries.
7. It is possible to visualize the ligand interactions with the receptor in details. To do that, select receptor together with one of poses in the **Project Table** and click on **Ligand Interaction Diagram…** in the **Tools** menu.

For queries and comments, mail to **pradeeph1157@gmail.com**

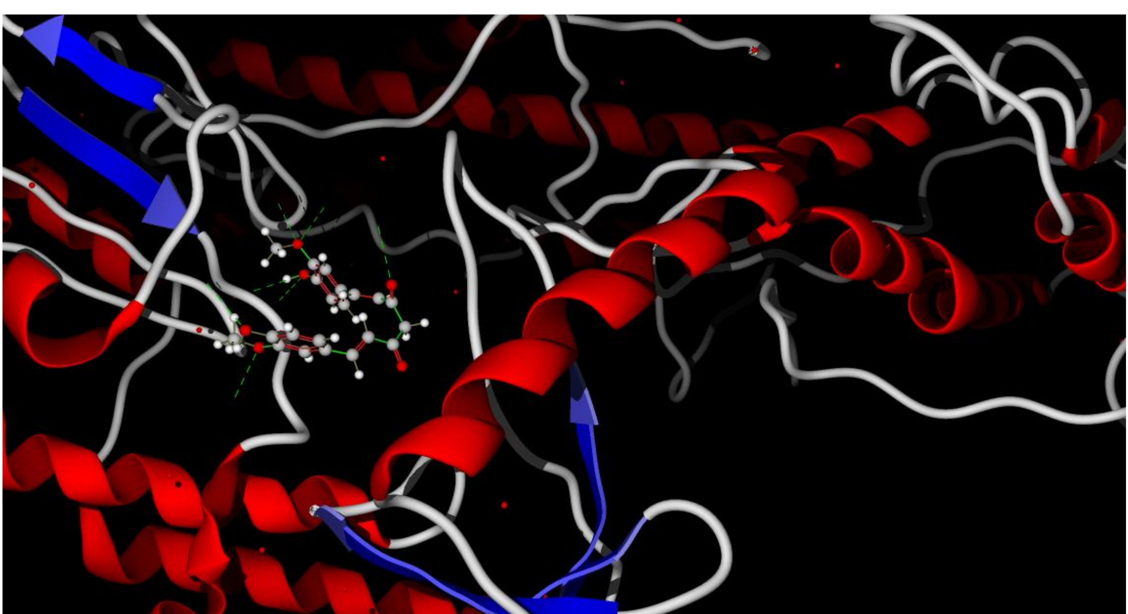
3.5. Other Software's for Molecular Docking

Molegro Virtual Docker- Practical Procedure

1. Download PDB structure of target ID - 1SJO from Protein Data Bank
2. Take ligand of interest (Curcumin). Copy its smiles notation and paste in CORINA 3D structure developer (http://www.molecular-networks.com/node/84). Download the generated 3D structure.
3. Open the molecule in Argus lab and optimize 3D.
4. Select the conformation with lowest energy and save it as PDB file.
5. Open the Molegro Virtual Docker software.
6. Import the receptor and ligand.
7. Prepare the molecule (both receptor and ligand) using the preparation menu.
8. Detect cavities in the protein molecule using the preparation menu.



9. Select Docking Wizard in the Docking menu.
10. Select the Reference ligand as current ligand.
11. Select the cavity of interest as the target constraint and for grid generation.
12. Start Docking.
13. After completion of the wizard drag the Pose Organizer to the Workspace.
14. Select the pose with lowest energy.
15. Check for H-bond interactions.
16. Analyse interactions



References

1. Prime, version 2.3, Schrödinger, LLC, New York, NY, 2011.
2. LigPrep, version 2.5, Schrödinger, LLC, New York, NY, 2011.
3. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem. 2004; 47:1739-1749.
4. Teli MK, Rajanikant GK. Identification of novel potential HIF-prolyl hydroxylase inhibitors by in silico screening. Mol Divers. 2012; 16:193-202.
5. Pradeep H, Rajanikant GK. A rational approach to selective pharmacophore designing: an innovative strategy for specific recognition of Gsk3β. Mol Divers. 2012 Aug;16(3):553-62.

Chapter 4

Quantitative Structure Activity Relationships (QSAR)

Mr. Mahesh Kumar Teli

Introduction

Quantitative structure–activity relationships (QSARs), are mathematical models that attempt to relate the structure-derived features of a compound to its biological or physicochemical activity. QSAR works on the assumption that structurally similar compounds have similar activities. Therefore, these methods have predictive and diagnostic abilities. They can be used to predict the biological activity (e.g., IC50) or class (e.g., inhibitor versus non-inhibitors) of compounds before the actual biological testing. They can also be used in the analysis of structural characteristics that can give rise to the properties of interest.

It can be used for predicting a specific interaction, such as the binding of a ligand to the active site of a specific protein and molecular properties such as bioavailability and toxicity. 3D-QSARs are being applied in many disciplines for example risk assessment, toxicity prediction, and regulatory decisions in addition to drug discovery and lead optimization.

4.1. Major types of QSAR analysis

a) Fragment based (group contribution)

Group or Fragment based QSAR is also known as GQSAR. GQSAR allows flexibility to study various molecular fragments of interest in relation to the variation in biological response. GQSAR also considers cross-terms fragment descriptors, which could be helpful in identification of key fragment interactions in determining variation of activity. Lead discovery using Fragnomics is an emerging paradigm. In this context FB-QSAR proves to be a promising strategy for fragment library design and in fragment-to-lead identification endeavours.

b) 3D-QSAR

3D-QSAR refers to the application of force field calculations requiring three-dimensional structures, e.g. based on protein crystallography or molecule superimposition. It uses computed potentials, e.g. the Lennard-Jones potential, rather than experimental constants and is concerned with the overall molecule rather than a single substituent. It examines the steric fields (shape of the molecule), the hydrophobic regions (water-soluble surfaces), and the electrostatic fields.

c) Chemical descriptor based

In this approach, descriptors quantifying various electronic, geometric, or steric properties of a catalyst are computed and used to develop a QSAR. This approach is different from the fragment (or group contribution) approach in that the descriptors are computed for the system as whole rather than from the properties of individual fragments. This approach is different from the 3D-QSAR approach in that the descriptors are computed from scalar quantities (e.g., energies, geometric parameters) rather than from 3D fields.

4.2 Practical Tutorial

Choose the following in the main window: Applications > Phase > Develop Common Pharmacophore Hypotheses

a) Adding Ligands

1. Click from File, which is located near the top of the panel.
2. Select the file ligand.maegz, and click Open. This file contains 44 ligands. These ligands have been prepared and conformers generated for each. If you are starting with 2D single structures, you would have to clean the structures and then generate conformers. When you click add, the Choose Activity Property dialog box opens. If you intend to build QSAR models (which you will do in this tutorial), this is where you should select the property to use as the experimental activity variable.

3. Choose all properties from the Subset option menu.
4. Select the pIC50-Exp property.
5. Ensure that a Convert property value is *not* selected.
6. Click OK. The Choose Activity Property dialog box closes and the ligands are imported.

b) Choosing the Active and Inactive Sets

1. Click Activity Thresholds. The Activity Thresholds dialog box appears, which allows you to specify the activity thresholds.
2. In the Active if activity above text box, type 7.0.
3. In the Inactive if activity below text box, type 4.9.
4. Click OK. The Pharm Set column now has active for each ligand whose pIC50 value is greater than 7 (11 ligands), and inactive for each ligand whose pIC50 value is less than 4.9 (8 ligands). The column is blank for ligands whose activity falls between these values.

c) Creating Pharmacophore Sites

1. Click the Create Sites button near the top of the panel.
2. Click Start in the Create Sites - Start dialog box.

Incorporation of results does not add any new information to the Ligands table, but the Find Common

Pharmacophores button at the bottom of the panel is now active, and you can proceed to the next step in the workflow.

3. Click Next. The Find Common Pharmacophores step is displayed.

d) Finding Common Pharmacophores

1. Ensure that maximum five and minimum four variants are selected, and then click Find.
2. Click Start in the dialog box.

e) Scoring Hypotheses

In the scoring step, a scoring procedure is applied to identify the pharmacophore from each box that yields the best alignment of the active ligands. The scoring procedure also provides a ranking of the different hypotheses.

1. Click Next to enter the Score Hypotheses step. The Score Hypotheses step is displayed. The Hypotheses table is empty because scoring has not yet been done.
2. Click Score Actives.
3. The Score Actives dialog box is displayed.
4. Click Start, and click Start again in the Start dialog box.

Once the job is incorporated, the Hypotheses table contains information about the highest scoring hypotheses from each variant.

5. Click Score Inactives. The Score Inactives dialog box opens.
6. Click Start, and click Start again in the Start dialog box.

The job takes less than a minute. When it finishes, the Survival-inactive column of the Hypotheses table is populated with values. The survival score is reduced fairly evenly for all the hypotheses.

f) Viewing Hypotheses and Ligand Alignments

In this exercise you will examine the nature of the top-scoring hypotheses and the quality of the associated ligand alignments.

1. Sort the Hypotheses table by survival score, by clicking twice on the Survival column heading. The top scoring hypotheses come from an ADRRR variant and ADDRR.
2. Click the In column for ADRRR in the Hypotheses table. The features of the hypothesis are displayed in the Workspace. If the hypothesis is *not* visible in the Workspace, click the Display Hypothesis toolbar button at the top of the Develop Pharmacophore Model panel.

The Alignments table is filled with a record for each ligand. The records for the ligands not used for the model (the *non-model* ligands) are dark gray, indicating that no alignment was done for these ligands. The records provide information on the conformation whose pharmacophores yielded the best multi-ligand alignment to the hypothesis when it was selected as the reference

from its box. Fitness measures the quality of each alignment using a weighted sum of alignment and volume scores, just as in the total scoring function.

Note that the eighth row of the Alignments table is blue, indicating that it is the reference ligand, i.e., the ligand from which the hypothesis came. Its fitness score is a perfect 3.0—the maximum possible score with the scoring options that were selected— because it corresponds to the alignment of train08 onto itself.

3. Place the reference ligand in the workspace by clicking its In box in the Alignments table. train08 is overlaid onto the hypothesis. The features of the hypothesis are perfectly positioned on the matching sites of this ligand, because the hypothesis comes from the reference ligand, so its features must coincide with those of the reference ligand.

4. Include the remaining actives in the Workspace. Most ligands superimpose on each other with only minimal offset. If you place this ligand in the Workspace with the reference ligand, you can see the deviation in the alignment.

5. Examine the hypothesis (ADRRR) by selecting its row in the Hypotheses table.

6. Click Alignment Options below the Alignments table. The Alignment Options dialog box opens. In addition to selecting the option to align the “non-model” ligands (those that were not in the pharm set or did not match all sites in the hypothesis), you can require certain features to match when the alignment is performed. If the features do not match for a particular ligand, the ligand is not aligned.

g) Proceeding to Build QSAR Model

Phase QSAR models are based on partial least-squares (PLS) regression, applied to a large set of binary-valued variables that encode whether or not ligand atoms or ligand features occupy various cube-shaped elements of space. When you build QSAR models, you should try several hypotheses to ensure that you have a good model. You can build models for multiple hypotheses simultaneously.

h) Assigning Training and Test Set Memberships

By default, all ligands are placed in the training set, so you must separate them into the proper training and test sets.

1. The hypothesis is displayed in the Workspace, and the Alignments table is populated with data for this hypothesis.
2. Select all the test set ligands in the Alignments table. Make sure you do not click in the QSAR Set column. If the ligands are not sorted by name, click the Ligand Name column heading to sort them.

i) Setting QSAR Model Options

Phase can generate QSAR models that are atom-based or pharmacophore-based. The independent variables in the QSAR model are derived from a regular grid of cubic volume elements that span the space occupied by the training set ligands. Each ligand is represented by a set of bit values (0 or 1) that indicate which volume elements are occupied by either a van der Waals model of the atoms in that ligand or by pharmacophore features of that ligand. In this exercise, you will be developing an atom-based model, and will set parameters to control the sizes of the cubes and the maximum number of PLS factors to include in the model.

1. Click Options. The Build QSAR Model options dialog box is displayed. Since we have selected the training set and the test set, there is no need to change settings in the Training set section.
2. Set Grid spacing to 1.0.
3. Set Maximum PLS Factors 6 (total no. of training molecules/5).
4. Under Model type, ensure that Atom-based is selected.
5. Click OK. The cubes that define the independent variables will be 1 Å on each side, and atom-based linear regression models will be built containing PLS factors.

j) Building the QSAR Models

In this exercise, PLS regression models are created using the training set ligands and then applied to the test set ligands.

1. Click Build Models. The Start dialog box opens.
2. Select a host, and click Start. The job status icon at the top right turns green and spins. When it stops, data for PLS regression models fills the QSAR results table. Each regression is a QSAR model.

The R-squared value always increases as the number of PLS factors increases, but the same is not necessarily true of Q-squared. In order to have a good QSAR, the R-squared should be near to one and Q-squared should be more than 0.5.

3. In the QSAR Results table, click the column (ADRRR). The hypothesis is displayed in the Workspace, and the Alignments table is populated with data for this hypothesis, including activity predictions for each ligand using PLS factors.

k) Visualizing the QSAR Model

Three-dimensional aspects of the QSAR model are examined to help gain an understanding of how the structures of the ligands contribute to the computed activity.

1. Ensure that the View QSAR model toolbar button is selected.
2. Place the most active ligand, in the Workspace.
3. Click QSAR Visualization. The QSAR Visualization Settings panel is displayed. This panel has various options for displaying characteristics of the QSAR model.

You can view the cubic volume elements occupied by the ligands in the Workspace, or view all the cubes in the QSAR model.

4. Ensure that all items in the View effects from list are selected, and that Combine effects is selected
5. Ensure that Workspace ligand is selected under View volume occupied by.
6. Select PLS factors.
7. Move the positive and negative coefficient threshold sliders to an intermediate value, such as 0.015 and –0.015. In the Workspace, you will see many blue cubes and a small number of red cubes. The blue cubes indicate regions that are favorable for activity and the red cubes indicate regions that are unfavorable.

8. Change the positive and negative coefficient thresholds to more extreme values, such as 0.03 and –0.03. The number of cubes in the Workspace drops significantly because now you are viewing only very significant terms in the model.

9. Change the thresholds back to 0.015 and –0.015.

10. Replace most active ligand in the Workspace with the least active ligand. The Workspace now contains many more red cubes than blue cubes, indicating a preponderance of unfavorable interactions.

11. Under View effects from, deselect Combine effects, and select (D) H-bond donor. This category includes polar hydrogens bonded to nitrogen, oxygen and sulfur.

Experiment with different atom classes and different thresholds to view the various ways in which the QSAR model distinguishes ligands with high and low activities.

For queries and comments, mail to **maheshkumar.teli@gmail.com**

References

1. Yee, L. C. and Wei, Y. C. (2012) Current Modeling Methods Used in QSAR/QSPR, in Statistical Modelling of Molecular Descriptors in QSAR/QSPR, Volume 2 (eds M. Dehmer, K. Varmuza and D. Bonchev), Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany. doi: 10.1002/9783527645121.ch1
2. User manual Phase, version 3.0, Schrödinger, LLC, New York, NY, 2008.
3. Dixon SL, Smondyrev AM, Knoll EH, Rao SN, Shaw DE, Friesner RA (2006) PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. J Comput Aided Mol Des 20:647-671

APPENDIX - I

No	Tool Name	Cheminformatics Tool Details	Web Link
1	PASS	Predicted activity spectrum for small molecules	http://www.pharmaexpert.ru/PASSOnline/
2	KEGG	Genomic information Pathway knowledgebase	http://www.genome.jp/kegg/
3	*GUSAR (General Unrestricted Structure-Activity Relationships)*	tool to create models on quantitative structure-activity (structure-property) relationships	http://www.pharmaexpert.ru/PASSOnline/products.php
4	ChemSketch	structure developing	http://www.acdlabs.com/download/
5	VEGA ZZ	VEGA ZZ is the evolution of the well known VEGA OpenGL package and includes several new features and enhancements that make very easy your cheminfo research jobs.	http://www.msg.ucsf.edu/local/programs/Vega/pages/gl_index.htm
6	PreAdmet	ADMET prediction	http://www.bmdrc.org/04_product/01_preadme.asp
7	CORINA	3D chemical Structure developer	http://www.molecular-networks.com/online_demos/corina_demo
8	Structural Bioinformatics-Expasy	Tools for Structural Bioinformatics	http://expasy.org/structural_bioinformatics
9	Systems biology Expasy	Systems biology & pathway simulation	http://expasy.org/systems_biology
10	CARBOHYDRATES	3D models of saccharides from their sequences	http://molbiol-tools.ca/Carbohydrates.htm
		Molecular Modelling	
1	MODELLER	MODELLER is used for homology or comparative modeling of protein three-dimensional structures	http://salilab.org/modeller/
2	I-TASSER	I-TASSER Suite is a package of standalone computer programs which were developed for protein structure prediction, refinement, and structure-based function annotations	http://zhanglab.ccmb.med.umich.edu/I-TASSER/download/
3	SwissPDB	Molecular modeling, graphics, and drug design program for Windows operating systems	http://spdbv.vital-it.ch/

4	ArgusLab	ArgusLab is a molecular modeling, graphics, and drug design program for Windows operating systems	http://www.arguslab.com/arguslab.com/ArgusLab.html
5	GenTHREADER	Secondary structure prediction, rapid fold recognition, matching your sequence against a library of whole PDB chains.	bioinf.cs.ucl.ac.uk/threader/
6	Protparam	Predicting protein parameters and properties	http://web.expasy.org/protparam/
7	MEMSAT3	Transmembrane prediction	http://bioinf.cs.ucl.ac.uk/?id=756
8	ProCheck	Ramachandran Plot	http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/
9	SABLE	Predicted Solvent accessibility	http://sable.cchmc.org/
10	Errat	ERRAT is a protein structure verification algorithm that is especially well-suited for evaluating the progress of crystallographic model building and refinement	http://nihserver.mbi.ucla.edu/ERRATv2/
11	SAVS server	Prove (Z-score), Procheck	http://nihserver.mbi.ucla.edu/SAVES/
12	3DLigandSite server	Potential Binding sites	http://bip.weizmann.ac.il/toolbox/structure/binding.htm
13	Phyre2	Secondary structure (graphical)	http://bioinformatictools.wordpress.com/tag/phyre2/
14	What_Check	finding binding sites of protein	http://swift.cmbi.ru.nl/gv/whatcheck/
Molecular Docking			
1	Hex 6.3	Protein- protein Docking	http://hex.loria.fr/
2	HADDOCK	Next to protein-protein docking, HADDOCK has been applied to the modelling of protein-DNA, protein-RNA, protein-oligosaccharides and protein-ligand complexes.	http://www.nmr.chem.uu.nl/haddock/
3	MVD	Protein- ligan flexible docking	http://www.molegro.com/products.php
4	AutoDock	Protein- ligan docking	http://autodock.scripps.edu/
5	GOLD	Protein- ligan docking	http://www.ccdc.cam.ac.uk/products/life_sciences/gold/
6	ParDOCK	Protein- ligan rigid docking	http://www.scfbioiitd.res.in/dock/pardock.jsp
7	RosettaDock	protein-protein docking server	http://rosettadock.graylab.jhu.edu/
8	DrugDesign-Expasy	Drugdesign tools and databases	http://expasy.org/drug_design

Protein Structure Annotation Tools			
1	SAS	SAS is a tool for applying structural information to a given protein sequence. It uses FASTA to scan a given protein sequence against all the proteins of known 3D structure in the Protein Data Bank (PDB).	http://www.ebi.ac.uk/thornton-srv/databases/sas/
2	ESPript	ESPript, Easy Sequencing in Postscript, is a utility to generate a pretty PostScript output from aligned sequences.	http://espript.ibcp.fr/ESPript/ESPript/
3	SPICE	SPICE is a browser for protein sequences, structures and their annotations. It can display annotations for PDB, UniProt and Ensembl Peptides.	http://www.efamily.org.uk/software/dasclients/spice/
Protein Structural Analysis			
1	DALI	The Dali server is a network service for comparing protein structures in 3D. Access also to related tools: SRS search for FSSP (families of structurally similar proteins); search for DSSP; search for HSSP, (homology-derived structures of proteins), a derived database merging structural (2-D and 3-D) and sequence information (1-D).	http://ekhidna.biocenter.helsinki.fi/dali_server/
2	PDBeMotif	PDBeMotif is an extremely fast and powerful search tool that facilitates exploration of the Protein Data Bank (PDB) by combining protein sequence, chemical structure and 3D data in a single search.	http://www.ebi.ac.uk/pdbe-site/pdbemotif/
3	WHAT IF	The WHAT IF Web Interface proposes a varied set of analyses on protein structure.	http://swift.cmbi.ru.nl/servers/html/index.html
4	PDBeFold	PDBeFold (SSM) is an interactive service for comparing protein structures in 3D.	http://www.ebi.ac.uk/msd-srv/ssm/
5	ProFunc	The ProFunc server had been developed to help identify the likely biochemical function of a protein from its three-dimensional structure.	http://www.ebi.ac.uk/thornton-srv/databases/ProFunc/
6	PIC	Protein Interactions Calculator (PIC) is a server which recognizes various kinds of interactions;	http://crick.mbu.iisc.ernet.in/~PIC/
7	Pride2Server	PRIDE2 server calculates the PRobability of IDEntity between three-dimensional domains (or whole structures) and offer varied various compararisons with known structures (from fold identification, CATH and PDB search, to NMR-NOE set comparison).	http://hydra.icgeb.trieste.it/pride/index.html
8	PBE server 2.0	PBE server 2.0 aims to provide a platform for protein structure	http://bioinformatics.univ-reunion.fr/?page_id=153
9	ConSurf	The ConSurf server is a useful and user-friendly tool that enables the identification of functionally important regions on the surface of a protein or domain, of known three-dimensional	http://consurf.tau.ac.il/

		(3D) structure, based on the phylogenetic relations between its close sequence homologues.	
10	StrucToolsn	StrucToolsn is a set of tools intended to provide a convenient web interface to simple, commonly used structural biology calculations with PDB files.	http://helixweb.nih.gov/structbio/basic.html
11	STINGMillennium	STING Millennium is a web based suite of programs that starts with the visualization of a molecular structure and leads the user through a series of operations resulting in an extensive structural analysis of the molecule	http://luna.bioc.columbia.edu/SMS/sms.html
12	firestar	firestar server: predicting functional residues from structural templates and alignment reliability	http://firedb.bioinfo.cnio.es/Php/FireStar.php
13	ESBRI	Evaluating the salt brIdges in proteins.	http://bioinformatica.isa.cnr.it/ESBRI/introduction.html
14	CMA	This program allows to analyse contacts between two chains or within one chain in a given PDB file.	http://ligin.weizmann.ac.il/cma/
15	HORI	HORI - Higher Order Residue Interactions, is a web server to compute higher order interactions (pairwise interaction, triplet interactions and quadruple interactions) in a protein structure.	http://caps.ncbs.res.in/hori/
16	SRIDE	Identification of Stabilizing Residues in proteins.	http://sride.enzim.hu/
17	MolTalk	MolTalk is a computational environment for doing Structural Bioinformatics. At the base of i.Moltalk	http://www.moltalk.org/
18	FoldX	FoldX provides a fast and quantitative estimation of the importance of the interactions contributing to the stability of proteins and protein complexes.	http://foldx.crg.es/
19	Seq2Struct	Web resource for the identification of sequence-structure links. The resource consists of an exhaustive collection of annotated links between the Swiss-Prot + TrEMBL sequence database entries and the PDB and SCOP structure database entries.	http://surface.bio.uniroma2.it/seq2struct/
20	iMolTalk	iMolTalk is an interactive, Internet-based service for computational analyses in Structural Biology that Compute Ramachandran plot (angles), Compute Distance Matrix of C atoms, Find Contacts for a residue or Distances between a pair of residues, Find Interface between two chains of a structure, Compute Secondary structure assignment, Structural Alignment Structural Alignment computation, derivation and verification. Structurally compare all models in a structure, and Analyse Domain Motion Analyse domain motion in homologous structures.	http://i.moltalk.org/
21	MaxSprout	MaxSprout is a fast database algorithm for generating protein backbone and side chain co-ordinates from a C(alpha) trace.	http://www.ebi.ac.uk/maxsprout/

Domain Prediction, Globularity and Assembly Analysis			
1	pDomain	pDomain resource is centered around defining structural domains from 3D coordinates. This resource brings together current state-of-the-art algorithmic methods for partitioning proteins into domains.	
2	CDD server	Search for Conserved Domains within a protein sequence.	http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi
3	DomPred	Dompred is a server designed to predict putative protein domains and their boundaries for a given protein sequence.	http://bioinf.cs.ucl.ac.uk/dompred/
4	Protein Peeling 3D	The approach 'Protein Peeling' aims at cutting the 3D protein structure into a limited set of 'Protein Units'.	http://www.dsimb.inserm.fr/dsimb_tools/peeling3/
5	DOMpro	DOMpro is a web server to predict protein domain boundaries using 1D-Recursive Neural Networks and statistical methods.	http://www.ics.uci.edu/~baldig/dompro.html
6	DIAL	DIAL is a web server for the automatic identification of structural domains given the three-dimensional coordinates of a protein.	http://caps.ncbs.res.in/DIAL/home.html
7	GLobPlot	Intrinsic Protein Disorder, Domain & Globularity Prediction	http://globplot.embl.de/
8	PURE	PURE is a bioinformatics protocol to identify putative domains in the Unassigned Regions.	http://caps.ncbs.res.in/PURE/PURE.html
9	Shandy	Prediction of Protein Domains at University College Dublin.	http://distill.ucd.ie/shandy/
10	Domain 3D	Domain 3D is a method developed by Willie Taylor to identify globular domains in protein structure.	http://www.ibi.vu.nl/programs/domain3Dwww/
11	DHcL	DHcL (Domain Hierarchy and closed Loops)is a server for the analysis of basic structural units of a protein. The server calculates domain structures at different levels of energy hierarchy and elements of the loop-n-lock structure, closed loops and van derWaals locks.	http://sitron.bccs.uib.no/dhcl/
Hinge & Flexibility			
1	DynDom	DynDom is a program to determine domains, hinge axes and hinge bending residues in proteins where two conformations are available.	http://fizz.cmp.uea.ac.uk/dyndom/dyndomMain.do
2	HINGEprot	An Algorithm For Protein Hinge Prediction Using Elastic Network Models	http://www.prc.boun.edu.tr/appserv/prc/hingeprot/index.html
3	FlexWeb	Analysis of Flexibility in Biomolecules and Networks.	http://flexweb.asu.edu/software/first/

4	DSSPcont	Analysis of Flexibility in Biomolecules and Networks.	https://rostlab.org/owiki/index.php/DSSPcont
5	TLSMD	TLS Motion Determination (TLSMD) analyzes a protein crystal structure for evidence of flexibility, e.g. local or inter-domain motions.	http://skuld.bmsc.washington.edu/~tlsmd/index.html
Surface & Cavity Analysis			
1	DynDom	DynDom is a program to determine domains, hinge axes and hinge bending residues in proteins where two conformations are available.	http://fizz.cmp.uea.ac.uk/dyndom/dyndomMain.do
2	HINGEprot	An Algorithm For Protein Hinge Prediction Using Elastic Network Models	http://www.prc.boun.edu.tr/appserv/prc/hingeprot/index.html
3	FlexWeb	Analysis of Flexibility in Biomolecules and Networks.	http://flexweb.asu.edu/software/first/
4	DSSPcont	Analysis of Flexibility in Biomolecules and Networks.	https://rostlab.org/owiki/index.php/DSSPcont
5	TLSMD	TLS Motion Determination (TLSMD) analyzes a protein crystal structure for evidence of flexibility, e.g. local or inter-domain motions.	http://skuld.bmsc.washington.edu/~tlsmd/index.html
Surface & Cavity Analysis			
1	CAVER	CAVER is a software tool for analysis and visualisation of channels (tunnels) in protein structures. Channels are void pathways leading from a cavity buried in a protein core to the surrounding solvent. Studying of these pathways is highly important for drug design and molecular enzymology.	http://www.caver.cz/
2	3D-SURFER	3D-SURFER is web-based software for protein surface comparison and analysis. The server integrates various repertoire of methods to assist in high throughput screening and visualization of protein surface comparisons.	http://dragon.bio.purdue.edu/3d-surfer/index.php?home
3	MolAxis	Accurate identification of channels in macromolecules.	http://bioinfo3d.cs.tau.ac.il/MolAxis/
4	SURF's_UP	SURF's_UP! is a web tool for analysis of functional relationships in protein families as inferred from protein surface maps comparison.	http://asia.genesilico.pl/surfs_up/
5	eF-surf	Molecular surface of proteins with the electrostatic potential is a representation of protein three dimensional structures, which often gives some clues to infer the function of proteins. eF-surf is a web server to calculate the molecular surface of the up-loaded file with PDB format.	http://ef-site.protein.osaka-u.ac.jp/eF-surf/top.do
6	CASTp	Binding sites and active sites of proteins and DNAs are often	http://sts.bioengr.uic.edu/castp/

		associated with structural pockets and cavities. castP server uses the weighted Delaunay triangulation and the alpha complex for shape measurements. It provides identification and measurements of surface accessible pockets as well as interior inaccessible cavities, for proteins and other molecules.	
7	PIC	Protein Interactions Calculator (PIC) is a server which recognizes various kinds of interactions;	http://crick.mbu.iisc.ernet.in/~PIC/
8	HotPatch	HotPatch finds unusual patches on the surface of proteins, and computes just how unusual they are (patch rareness), and how likely each patch is to be of functional importance (functional confidence (FC).) The statistical analysis is done by comparing your protein's surface against the surfaces of a large set of proteins whose functional sites are known.	http://hotpatch.mbi.ucla.edu/
9	MolLoc	MolLoc takes in input two molecules and superimposes them on their two most extended similar regions. This server can be used to test if two molecules with different sequences and folds share any local surface similarity.	http://bcb.dei.unipd.it/MolLoc/
10	3V	Web tool for volume calculation.	http://geometry.molmovdb.org/3v/
Binding Pocket and Binding Site Prediction			
1	IBIS	IBIS is the NCBI Inferred Biomolecular Interactions Server. For a given protein sequence or structure query, IBIS reports physical interactions observed in experimentally-determined structures for this protein.	http://www.ncbi.nlm.nih.gov/Structure/ibis/ibis.cgi
2	Pocket-finder	Pocket-Finder is a pocket detection algorithm based on Ligsite written by Hendlich et al (1997). Pocket-Finder works by scanning a probe radius 1.6 angstoms along all gridlines of a grid resolution 0.9 angstroms surrounding the protein.	http://www.modelling.leeds.ac.uk/pocketfinder/
3	FindSite	FINDSITE is a threading-based binding site prediction/protein functional inference/ligand screening algorithm that detects common ligand binding sites in a set of evolutionarily related proteins. Crystal structures as well as protein models can be used as the target structures.	http://cssb.biology.gatech.edu/findsite
4	LIGSITE	LIGSITE is a web server for the automatic identification of pockets on protein surface using the Connolly surface and the degree of conservation	http://projects.biotec.tu-dresden.de/pocket/
5	PocketPIcker	PocketPicker - Binding Site Prediction	http://gecco.org.chemie.uni-frankfurt.de/pocketpicker/index.html
6	MetaPocket	metaPocket is a meta server to identify pockets on protein surface to predict ligand-binding sites!	http://metapocket.eml.org/

7	CastP	castP server uses the weighted Delaunay triangulation and the alpha complex for shape measurements. It provides identification and measurements of surface accessible pockets as well as interior inaccessible cavities, for proteins and other molecules.	http://sts.bioengr.uic.edu/castp/
8	3DLigandSite	3DLigandSite is an automated method for the prediction of ligand binding sites.	http://www.sbg.bio.ic.ac.uk/~3dligandsite/
9	Qsite-Finder	Ligand Binding Site Prediction	http://www.modelling.leeds.ac.uk/qsitefinder/
10	PASS	Fast Prediction and Visualization of Protein Binding Pockets.	http://mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py?form=PASS
11	MEDock	The MEDock (Maximum-Entropy based Docking) web server is aimed at providing an efficient utility for prediction of ligand binding site.	http://medock.ee.ncku.edu.tw/
12	Q-SiteFinder	Q-SiteFinder is a new method of ligand binding site prediction. It works by binding hydrophobic (CH_3) probes to the protein, and finding clusters of probes with the most favourable binding energy. These clusters are placed in rank order of the likelihood of being a binding site according to the sum total binding energies for each cluster.	http://www.modelling.leeds.ac.uk/qsitefinder/
13	PINUP	Protein binding site prediction with an empirical scoring function.	http://sparks.informatics.iupui.edu/PINUP/
14	SOAK	Metal ion binding sites, affinities, and specificities from structure	http://protinfo.compbio.washington.edu/soak/
15	ProPred	The aim of this server is to predict MHC Class-II binding regions in an antigen sequence, using quantitative matrices.	http://www.imtech.res.in/raghava/propred/
Ligand Interaction			
1	STITCH	STITCH is a resource to explore known and predicted interactions of chemicals and proteins.	http://stitch.embl.de/
2	LPCSU	Ligand-Protein Contacts & Contacts of Structural Units.	http://bip.weizmann.ac.il/oca-bin/lpccsu
3	PLATINUM	PLATINUM is designed for calculation of hydrophobic properties of molecules and their match or mismatch in receptor–ligand complexes. These properties may help to analyze results of molecular docking.	http://model.nmr.ru/platinum/
4	LigPlot	Program for automatically plotting protein-ligand interactions.	http://www.ebi.ac.uk/thornton-srv/software/LIGPLOT/
5	PDBeMotif	PDBeMotif is an extremely fast and powerful search tool that facilitates exploration of the Protein Data Bank (PDB) by	http://www.ebi.ac.uk/pdbe-

		combining protein sequence, chemical structure and 3D data in a single search.	site/pdbemotif/
6	Catalytic Site Atlas (CSA)	The Catalytic Site Atlas (CSA) is a database documenting enzyme active sites and catalytic residues in enzymes of 3D structure.	http://www.ebi.ac.uk/thornton-srv/databases/CSA/
7	SuMo	SuMo allows you to screen the Protein Data Bank (PDB) for finding ligand binding sites matching your protein structure or inversely, for finding protein structures matching a given site in your protein. This method is neither based on aminoacid sequence nor on fold comparisons.	http://sumo-pbil.ibcp.fr/cgi-bin/sumo-welcome
8	ProSMoS	All structures are reduced to matrices that contain just enough information to define a fold, so the definition is general and large deviations in coordinates are tolerated. A user supplies a matrix for a motif, and ProSMoS lists all structures that exactly match this motif.	http://prodata.swmed.edu/ProSMoS/
9	MultiBind	Multiple alignment of protein binding sites recognizes spatial chemical binding patterns common to a Set of protein Structures.	http://bioinfo3d.cs.tau.ac.il/MultiBind/
10	SiteEngine	SiteEngine recognizes regions on the surface of one protein that resemble a specific binding site of another.	http://bioinfo3d.cs.tau.ac.il/SiteEngine/
11	PINTS	Patterns In Non-homologous Tertiary Structures. PINTS finds similarities between protein structures consisting of amino acids that are close in space, but not necessarily close or co-linear in sequence (local structural patterns, for example the catalytic triad). Unlike other tools, PINTS does not aim to find proteins adopting a similar fold.	http://www.russell.embl.de/pints/
12	FunClust	FunClust is a web server for the identification of local functional motifs in a set of non homologous protein structures	http://pdbfun.uniroma2.it/funclust/help.py
13	LabelHash	The LabelHash suite of programs and scripts can be used to match point-based structural motifs to a set of target proteins. A motif is defined by the C-alpha positions of a number of residues. Associated with each motif point is a number of allowed residue labels.	http://kavrakilab.org/labelhash/labelhash.html
14	RASMOT-3D PRO	RASMOT-3D PRO searches in protein structure files for proteins possessing a group of residues in a topology similar to that adopted by a 3D motif given in input.	http://biodev.extra.cea.fr/rasmot3d/
15	PESDserv	Compare binding sites of proteins.	http://reccr.chem.rpi.edu/Software/pesdserv/
16	SeeMotif	Given a set of motifs and a reference sequence, seeMotif helps to visualize and exploring these motifs in appropriate structures selected based on a reference sequence.	http://seemotif.csie.ntu.edu.tw/index.html
17	ProtMot	ProtMot, a PROTeins MOTif analysis tool. This web site analyzes structure of proteins and output the list of hydrogen bonding	http://bioinfo.weizmann.ac.il/protm

		patterns network motifs. The list of motifs, a superset of the secondary structures, is shown both projected on the protein structure and as a network motif significance profile (SP), a chart which gives the occurrences of each motif in the protein.	ot/
Interface Analysis			
1	PIPSA	PIPSA service is provided for the comparison of the electrostatic interaction properties of proteins. It permits the classification of proteins according to their interaction properties.	http://pipsa.eml.org/pipsa/pipsa-index.jsp
2	PDBePISA	PDBePISA is an interactive tool for the exploration of macromolecular (protein, DNA/RNA and ligand) interfaces, prediction of probable quaternary structures (assemblies), database searches of structurally similar interfaces and assemblies, as well as searches on various assembly and PDB entry parameters.	http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html
3	MAPPIS	Multiple Alignment of Protein-Protein InterfaceS recognizes spatially conserved chemical interactions shared by a set of protein-protein interfaces.	http://bioinfo3d.cs.tau.ac.il/MAPPIS/
4	I2I	Interface-to-Interface (I2I)-SiteEngine compares pairs of interacting protein binding sites.	http://bioinfo3d.cs.tau.ac.il/I2I-SiteEngine/index.html
5	PROTORP	The PROTORP server is a bioinformatics tool designed to analyse the interfaces between protein chains in protein-protein associations.	http://www.bioinformatics.sussex.ac.uk/protorp/
6	MolSurfer	A Macromolecular Interface Navigator. MolSurfer is a graphical tool that links a 2D projection of a macromolecular interface to a 3D view of the macromolecular structures. MolSurfer can be used to study protein-protein and protein-DNA/RNA interfaces. The 2D projections of the computed interface aid visualization of complicated interfacial geometries in 3D. Molecular properties, including hydrophobicity and electrostatic potential, can be projected onto the interface. MolSurfer can thereby aid exploration of molecular complementarity, identification of binding 'hot spots' and prediction of the effects of mutations. MolSurfer can also facilitate the location of cavities at macromolecular interfaces.	http://projects.villa-bosch.de/dbase/molsurfer/
7	Proface	A server for the analysis of the physicochemical features of protein-protein interfaces.	http://resources.boseinst.ernet.in/resources/bioinfo/interface/
8	AquaProt	AquaProt analyses protein-protein binding interface, defines inter-residue interaction map within the interface and extracts related water molecules.	http://bioinfo.weizmann.ac.il/aquaprot/
9	VASCo	VASCo is a program pipeline including a visualization tool to calculate and visualize annotated surfaces with special emphasis on surface contact regions and protein-protein interactions.	http://genome.tugraz.at/VASCo/

10	LIGPLOT	LIGPLOT Automatically generates schematic diagrams of protein-ligand interactions for a given PDB file.	http://www.ebi.ac.uk/thornton-srv/software/LIGPLOT/
11	lpccsu	Ligand-Protein Contacts & Contacts of Structural Units	http://bip.weizmann.ac.il/oca-bin/lpccsu
Solvent Accessibility			
1	SPPIDER	Solvent accessibility based Protein-Protein Interface iDEntification and Recognition.	http://sppider.cchmc.org/
2	SABLE	Accurate sequence-based prediction of relative Solvent AccessiBiLitiEs, secondary structures and transmembrane domains for proteins of unknown structure.	http://sable.cchmc.org/
3	GETAREA	Calculation of Solvent Accessible Surface Areas, Atomic Solvation Energies and Their Gradients for Macromolecules.	http://curie.utmb.edu/getarea.html
4	POPS	POPS is a fast algorithm to calculate solvent accessible surface areas (SASAs) of proteins and nucleic acids at atomic (default) and residue (coarse-grained) level.	http://mathbio.nimr.mrc.ac.uk/wiki/POPS
5	ASAP	ASAP predicts solvent accessible surface area of proteins.	http://ccb.imb.uq.edu.au/ASAP/
6	ASA-View	This server provides graphical representation of solvent accessibility of amino acid residues in proteins, with known structures.	http://gibk26.bse.kyutech.ac.jp/jouhou/shandar/netasa/asaview/
Functional & Conserved Residues			
1	ConSurf	The ConSurf server is a useful and user-friendly tool that enables the identification of functionally important regions on the surface of a protein or domain, of known three-dimensional (3D) structure, based on the phylogenetic relations between its close sequence homologues.	http://consurf.tau.ac.il/
2	Crescendo	Crescendo can only be used to identify residues that are interacting with ligands or other proteins.	http://mordred.bioc.cam.ac.uk/~crescendo/crescendo.php
3	3dLOGO	3dLOGO is a web server that allows the identification of 3D locally conserved residues in a set of protein structures.	http://3dlogo.uniroma2.it/3dLOGO/index.html
4	Firestar	Firestar predicts functionally important residues from structural templates and alignment reliability.	http://firedb.bioinfo.cnio.es/Php/FireStar.php
Effects of Mutations in Protein (SNP/InDels)			
1	MUpro	Prediction of Protein Stability Changes for Single-Site Mutations from Sequences	http://www.ics.uci.edu/%7Ebaldig/mutation.html
2	FoldX	FoldX provides a fast and quantitative estimation of the importance of the interactions contributing to the stability of	http://foldx.crg.es/

		proteins and protein complexes.	
3	SIFT	SIFT predicts whether an amino acid substitution affects protein function based on sequence homology and the physical properties of amino acids. SIFT can be applied to naturally occurring nonsynonymous polymorphisms and laboratory-induced missense mutations.	http://sift.bii.a-star.edu.sg/
4	SNP Analysis	Estimates the likelihood of a particular nonsynonymous (amino-acid changing) coding SNP to cause a functional impact on the protein. It calculates the subPSEC (substitution position-specific evolutionary conservation) score based on an alignment of evolutionarily related proteins.	http://www.pantherdb.org/tools/csnpScoreForm.jsp
5	CUPSAT	CUPSAT is a tool to predict changes in protein stability upon point mutations. The prediction model uses amino acid-atom potentials and torsion angle distribution to assess the amino acid environment of the mutation site. Additionally, the prediction model can distinguish the amino acid environment using its solvent accessibility and secondary structure specificity.	http://cupsat.tu-bs.de/
6	SNAP	SNAP is a method for evaluating effects of single amino acid substitutions on protein function. It was developed by Yana Bromberg in Rost Lab, at Columbia University, New York.	http://cubic.bioc.columbia.edu/services/SNAP/
Disulphide Bonds			
1	DiSulFind	Cysteines Disulfide Bonding State and Connectivity Predictor	http://disulfind.dsi.unifi.it/
2	DiANNA	DiANNA: unified software for Cysteine state and Disulfide Bond partner prediction	http://bioinformatics.bc.edu/clotelab/DiANNA/
3	Disulfide by Design	Disulfide by Design is an application for the rational design of disulfide bonds in proteins. For a given protein structural model, all residue pairs are rapidly assessed for proximity and geometry consistent with disulfide formation	http://cptweb.cpt.wayne.edu/DbD/
4	CYSREDOX	Predicting the redox state of cysteins in proteins from multiple sequence alignments	http://manaslu.aecom.yu.edu/cysredox.html
5	DSDBASE	DSDBASE is a database on disulphide bonds in proteins that provides information on native disulphides and those which are stereochemically possible between pairs of residues in a protein.	http://caps.ncbs.res.in/dsdbase//dsdbase.html
Metal Bindind Sites			
1	MetalDetector	MetalDetector predicts metal binding sites in proteins using sequence information alone. Prediction is limited to transition metals (with the addition of heme and Fe/S clusters) and to CYS and HIS as ligands.	http://metaldetector.dsi.unifi.it/v2.0/

2	CHED	This site uses the 'CHED' algorithm to predict 3D intra-chain protein binding sites for transition metals (Zn, Fe, Mn, Cu, Ni, Co, and Ca and Mg sites that can be replaced by a transition metal).	http://ligin.weizmann.ac.il/~lpgerzon/mbs4/mbs.cgi
		Structure Activity Relationships	
1	QSAR tools	The Computational Toxicology Group supports the development of QSAR and statistical analysis software tools that are potentially useful for regulatory purposes.	http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/qsar_tools
2	Stat4tox (Software for the Statistical Evaluation of In Vitro Assays)	An open-source computer program to implement a variety of statistical analysis techniques for toxicological data, including dose-response data.	http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/qsar_tools/stat4tox
3	METIS (Metabolic Information Input System)	An open-source computer program for the storage and input of information on metabolism and degradation reactions.	http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/qsar_tools/METIS
4	Danish (Q)SAR Database	Danish Environmental Protection Agency (EPA) constructed a (Q)SAR database comprising predictions made by some 70 models for about 166,000 organic chemicals for a wide range of different endpoints. In 2004, a collaborative project was set up between the Danish EPA and the JRC to develop an internet-accessible version of this database. The internet version of the Danish (Q)SAR Database was constructed to enable different types of searching, including structure (substructure/exact match) searching, ID (CAS number, name) searching and parameter (endpoint) searching.	http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/qsar_tools/DDB
5	QSAR Reporting Formats and the JRC QSAR Model Database	a collection of robust summaries of (Q)SAR models compiled by using a standard (Q)SAR Model Reporting Format (QMRF). The QMRF template is available for download.	http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/qsar_tools/QRF
6	EasyQSAR	A beginners tool for QSAR in Drug Designing (Free software for drug designing and QSAR)	http://easyqsar.blogspot.in/2009/09/easyqsar-beginners-tool-for-qsar-in.html
7	TEST	Toxicity Estimation Software Tool (TEST)	http://www.epa.gov/nrmrl/std/qsar/qsar.html
8	OECD QSAR Toolbox	The QSAR Toolbox is a software intended to be used by governments, the chemical industry and other stakeholders to fill gaps in (eco-)toxicity data needed for assessing the hazards of chemicals. The Toolbox incorporates information and tools from various sources into a logical workflow. Grouping chemicals into chemical categories is crucial to this workflow.	http://www.qsartoolbox.org/
9	Abalone	Biomolecular simulations, protein folding.	http://www.biomolecularmodeling.com/Abalone/index.html

10	ACEMD	Molecular dynamics with CHARMM, Amber forcefields. Running on NVIDIA GPUs. Heavily optimized with CUDA.	http://www.acellera.com/acemd
11	ADUN	Charmm, AMBER, user specified (through force field markup language, FFML), QM/MM calculations with Empirical Valence Bond (EVB); Framework based (GNUStep/cocoa); SCAAS for spherical boundary conditions	http://adun.imim.es/
12	AMBER	Simulation of biomolecules	http://ambermd.org/
13	Ascalaph Designer	Molecular building (DNA, proteins, hydrocarbons, nanotubes).	http://www.biomolecularmodeling.com/Ascalaph/Packages.html
14	Automated Topology Builder (ATB)	Automated molecular topology building service for small molecules (< 99 atoms). GROMOS, GROMACS, CNS formats with validation	http://compbio.biosci.uq.edu.au/atb
15	Avogadro	Molecule building, editing (Peptides, small molecules, crystals), Conformational analysis, 2D/3D conversion. Extensible interfaces to other tools.	http://avogadro.openmolecules.net/wiki/
16	Balloon	2D/3D conversion and conformational analysis.	http://www.abo.fi/~mivainio/balloon/
17	BOSS	OPLS	http://zarbi.chem.yale.edu/software.html#boss
18	CHARMM	Commercial version with multiple graphical front ends is sold by Accelrys (as CHARMm)	http://www.charmm.org/
19	Chemitorium	Free 2D/3D graphical organic molecule builder, viewer and visualisation tool.	http://weltweitimnetz.de/software/Chemistry.en.page
20	ChemSketch	Fast 2-D graphical molecule builder and 3-D viewer. Contains simplified CHARMM for fast stable inaccurate optimization of single molecules up to 1000 atoms	http://www.acdlabs.com/resources/freeware/chemsketch/
21	COSMOS	Hybrid QM/MM COSMOS-NMR force field with fast semi-empirical calculation of electrostatic and/or NMR properties. 3-D graphical molecule builder and viewer.	http://www.cosmossoftware.de/ce_intro.html
22	Culgi	Atomistic simulations and mesoscale methods.	http://www.culgi.com/
23	Desmond	High Performance MD. Comes with a comprehensive GUI for building, visualizing, and reviewing results as well as calculation setup up and launching.	http://deshawresearch.com/resources.html
24	Discovery Studio	Discovery Studio is a comprehensive life science modeling and simulation suite of applications focused on optimizing the drug discovery process. Capabilities include, small molecule simulations, QM/MM, pharmacophore modeling, QSAR, protein-ligand docking, protein homology modeling, sequence analysis, protein-protein docking, antibody modeling, etc.	http://accelrys.com/products/discovery-studio/

25	fold.it	University of Washington and The Baker Labs. Structure prediction. Protein folding.	http://fold.it/portal/
26	FoldX	Energy calculations and protein design	http://foldx.crg.es/
27	GoVASP	GoVASP is a sophisticated graphical user interface for the Vienna Ab-Initio Simulation Package (VASP). GoVASP comprises tools to prepare, perform and monitor VASP calculations and to evaluate and visualize the computed data.	http://www.govasp.com/
28	GPIUTMD	GPIUTMD stands for Graphic Processors at Isfahan University of Technology for Many-particle Dynamics. It performs general purpose particle dynamics simulations on a single workstation, taking the advantage of NVIDIA CUDA GPUs to attain a level of performance equivalent to hundreds of processors on a fast cluster.	http://gpiutmd.iut.ac.ir/index.php
29	GROMACS	High performance MD	http://www.gromacs.org/
30	GROMOS	Geared towards biomolecules	http://www.igc.ethz.ch/GROMOS/index
31	GULP	Molecular dynamics and Lattice optimization	https://projects.ivec.org/gulp/
32	HOOMD-blue	General-purpose Molecular Dynamics highly optimized for GPUs. Includes various pair potentials, Brownian dynamics, dissipative particle dynamics, rigid body constraints, energy minimization, etc...	http://codeblue.umich.edu/hoomdblue/index.html
33	ICM	Powerful global optimizer in an arbitrary subset of internal variables, NOEs, Protein docking, Ligand docking, Peptide docking, EM, Density placement	http://www.molsoft.com/
34	LAMMPS	Has potentials for soft and solid-state materials and coarse-grain systems	http://lammps.sandia.gov/
35	MacroModel	OPLS-AA, MMFF, GBSA solvent model, conformational sampling, minimization, MD. Includes the Maestro GUI which provides visualization, molecule building, calculation setup, job launching and monitoring, project-level organization of results and access to a suite of other modelling programs.	http://www.schrodinger.com/ProductDescription.php?mID=6&sID=8&cID=0
36	MAPS	Building, visualization and analysis tools in a single user interface together with access to multiple simulation engines.	http://www.scienomics.com/products/molecular-modeling-platform/
37	Materials Studio	Materials Studio is a software environment that brings the materials simulation technology to desktop computing, solving key problems throughout the R&D process.	http://accelrys.com/products/materialsstudio/
38	MedeA	MedeA combines leading experimental databases and major computational programs like the Vienna Ab-Initio Simulation Package (VASP), LAMMPS, GIBBS with sophisticated materials property prediction, analysis, and visualization.	http://www.materialsdesign.com/

39	MCCCS Towhee	Originally designed for the prediction of fluid phase equilibria	http://towhee.sourceforge.net/
40	MDynaMix	Parallel MD	http://www.fos.su.se/~sasha/mdynamix/
41	MOE	Molecular Operating Environment	http://www.chemcomp.com/
42	MOIL	Also includes action-based algorithms (Stochastic Difference Equation in Time and Stochastic Difference Equation in Length) and locally enhanced sampling.	http://cbsu.tc.cornell.edu/software/moil/moil.html
43	molecools	Simple Javascript molecular visualization tool	http://blahbleh.com/molecools.php
44	MOLDY	Parallel, only pair-potentials, Cell lists, modified Beeman's algorithm	http://www.ccp5.ac.uk/moldy/moldy.html
45	ORAC	A Molecular Dynamics Simulation Program to Explore Free Energy Surfaces in Biomolecular Systems at the Atomistic Level	http://www.chim.unifi.it/orac
46	NAB[4]	Generation of Models for "Unusual" DNA and RNA	http://www.scripps.edu/mb/case/
47	Packmol	Builds complex initial configurations for Molecular Dynamics	http://www.ime.unicamp.br/~martinez/packmol
48	Prime	Homology modeling, loop and side chain optimization, minimization, OPLS-AA, SGB solvent model, parallalized	http://www.schrodinger.com/ProductDescription.php?mID=6&sID=2&cID=0
49	Protein Local Optimization Program	Helix, loop, and side chain optimization. Fast energy minimization.	http://jacobson.compbio.ucsf.edu/plop_manual/plop_overview.htm
50	p4vasp	Python-based viewer, structure builder and VASP results browser. Shows band-structure, charge densities and simulates STM images.	http://p4vasp.at/
51	PyMol	Python-based viewer, many plugins to other software. Some mutagenisis.	http://www.pymol.org/
52	QMOL	Protein viewer, provided by DNASTAR	http://www.dnastar.com/products/qmol/index.html
53	RasMol	Fast viewer	www.bernsteinplussons.com/software/rasmol/
54	Raster3D	High quality raster images	http://skuld.bmsc.washington.edu/raster3d/raster3d.html
55	RedMD	Reduced MD. Package for coarse-grained simulations.	http://bionano.icm.edu.pl/software/redmd
56	StruMM3D (STR3DI32)	Sophisticated 3-D molecule builder and viewer, advanced structural analytical algorithms, full featured molecular modeling and quantitation of stereo-electronic effects, docking	http://www.exorga.com/

		and the handling of complexes.	
57	Selvita Protein Modeling Platform	Protein structure prediction, homology modeling, *ab initio* modeling, loop modeling, protein threading	http://www.selvita.com/selvita-protein-modeling-platform.html
58	SimBioSys' MoDeST	molecular docking, scoring functions for docking, "ligand-based", "fragment-based", "de-novo"	http://www.simbiosys.ca/products/index.html
59	Spartan	Small molecule (< 2000 a.m.u.) MM and QM tools for determining conformation, structure, property, spectra, reactivity, and selectivity.	http://www.wavefun.com/products/spartan.html
60	SwissParam	Web server to determine automatically parameters and topologies for small organic molecules, for use with the CHARMM all atoms force field. Files can be used with CHARMM and GROMACS.	http://www.swissparam.ch/
61	TeraChem	High performance GPU-accelerated *ab initio* Molecular Dynamics and TD/DFT software package for very large molecular or even *nanoscale* systems. The software runs on NVIDIA GPUs and 64-bit Linux, and is based on heavily optimized CUDA code.	http://petachem.com/
62	TINKER	Software tools for molecular design	http://dasher.wustl.edu/tinker/
63	Tremolo-X	Fast, parallel MD	http://www.tremolo-x.com/
64	UCSF Chimera	Visually appealing viewer, amino acid rotamers and other building, includes Antechamber and MMTK, Ambertools plugins in development.	http://www.cgl.ucsf.edu/chimera/index.html
65	VEGA ZZ	3D viewer, multiple file format support, 2D and 3D editor, surface calculation, conformational analysis, MOPAC and NAMD interfaces, MD trajectory analysis, molecular docking, virtual screening, database engine, parallel design, OpenCL acceleration, etc.	http://www.vegazz.net/
66	VLifeMDS	Complete Molecular Modelling Software, QSAR, Combinetorial Library generation, Pharmacophore, Cheminformatics, docking, etc.	http://www.vlifesciences.com/
67	VMD + NAMD	Fast, parallel MD, CUDA	http://www.ks.uiuc.edu/Research/vmd/
60	WHAT IF	Visualizer for MD. Interface to GROMACS.	http://swift.cmbi.ru.nl/whatif/
61	xeo	open project management for nanostructures	http://sourceforge.net/projects/xeo
62	YASARA	Molecular-graphics, -modeling and -simulation program	http://www.yasara.com/
63	Zodiac	Drug design suite	http://www.zeden.org/

64	ESPResSo	ESPResSo is a highly versatile software package for performing and analyzing scientific Molecular Dynamics many-particle simulations of coarse-grained atomistic or bead-spring models as they are used in soft-matter research in physics, chemistry and molecular biology.	http://espressomd.org/
65	LAMMPS Molecular Dynamics Simulator	LAMMPS is a classical molecular dynamics code, and an acronym for Large-scale Atomic/Molecular Massively Parallel Simulator.	http://lammps.sandia.gov/
66	MDTools suite	The MDTools suite is a collection of programs, scripts, and utilities we provide for researchers to make various modeling and simulation tasks easier, and to provide basic code and utilities which can be built up into larger toolsets.	http://www.ks.uiuc.edu/Development/MDTools/
67	Gromacs	GROMACS is a versatile package to perform molecular dynamics, i.e. simulate the Newtonian equations of motion for systems with hundreds to millions of particles.	http://www.gromacs.org/