Editors

Dr. Ghulam Md Ashraf
(Email: ashraf.gm@gmail.com)

Dr. Ishfaq Ahmed Sheikh
(Email: sheikhishfaq@gmail.com)

King Fahd Medical Research Center
King Abdulaziz University, P.O. Box 80216
Jeddah, Saudi Arabia
1. Introduction

Large scale genome sequencing projects are generating tremendous quantity of protein sequences; however, complete understanding of the biological function of these proteins necessitates detailed knowledge about their structure and function [1]. Majority of the functions of an organism are mediated by proteins, and all these functions are generally, determined by three-dimensional (3D) structure of the proteins [2]. One of the common goals in biological sciences is to functionally characterize a protein sequence by solving an accurate 3D structure of the protein under study [3]. X-ray crystallography and NMR spectroscopy are the most powerful experimental methods to study the structure of a protein [4,5] and the improvement in these methods has led to an increase in the number of protein structures in PDB. Moreover, together with the enhancement of computational technologies in recent times and the progress of latest and powerful computer programs, it has become easier to predict new structure models based on the huge growth in the number of protein structures deposited in the PDB [4]. These protein structures have been extremely helpful in the refinement of experimental structures [6,7], the design of site-directed mutants [8], the characterization of molecular function and structure-based drug design [9].

Therefore, in view of the growing number of 3D structures and models, and their indispensable role in accurate functioning of the proteins, there is a need to develop various computational resources that can aid in the storage and analysis of these structures. PDB serves as a primary source in areas of structural biology along with other web-based protein structure databases which come in a large variety of types and levels of knowledge content. Some of them exhibiting general purpose interest cover all experimentally determined structures and provide valuable links, analyses, and graphical representations involving their 3D structure and biological function. Many other databases have attempted to organize 3D structures based on their folds as these can provide insights into their evolutionary relationships which might not be easy to identify from sequence comparison only. There are many servers that compare folds which are predominantly helpful for newly determined structures, and especially those having unknown function. The other more specialized databases deal with specific families, diseases and various structural features [10]. In addition to these databases for experimentally determined 3D structures, some databases aim at storing 3D models of proteins based on homology or comparative modeling.

Similarly, computational structure prediction methods offer important information for the large fraction of sequences whose structures are not experimentally determined so far. Among the various protein structure prediction methods, threading and comparative modeling depends on similarity across most of the modeled sequence and at least one known structure. However, in case of de novo or Ab initio methods, the structure is predicted from sequence alone and does not require any prior similarity at the fold level between the modeled sequence and any of the known structures. The aim of this chapter is to provide a comprehensive list of web based protein structure databases and state of the art 3D structure prediction tools. In the end, a brief overview of the CASP experiment is also provided.

2. 3D Structure Databases

Protein structure databases serve as a resource for a variety of experimentally determined protein structures. The main aim of the majority of these protein structure databases is to categorize and annotate the protein structures, thereby presenting
the biological society access to the experimental data in a constructive manner. Data incorporated in protein structure databases often consists of three dimensional coordinates as well as experimental details, such as unit cell dimensions and angles for x-ray crystallography resolved structures. Since PDB is the master protein 3D structure database, therefore, before highlighting some of the major protein structure databases, a brief historical background on PDB is provided in the next section.

2.1 Historical background of PDB

In 1968, a small but increasing number of protein structures determined by X-ray diffraction, and the newly existing molecular graphics display, known as the Brookhaven Raster Display (BRAD), to visualize these protein structures in 3D were the driving forces that led to the birth of the PDB. Later, with the support of Walter Hamilton a chemist at the Brookhaven National Laboratory, Edgar Meyer (Texas A&M University) began to write software to store atomic coordinate files in a universal format to make them accessible for geometric and graphical assessment. By 1971, one of Meyer’s programs, SEARCH, facilitated researchers to distantly access information from the database to study protein structures offline [11]. SEARCH was helpful in enabling networking, thus marking the practical establishment of the PDB.

Soon after the death of Hamilton’s in 1973, Tom Koetzle took over the direction of the PDB, and in 1974 the first PDB Newsletter was circulated to explain the details of data deposition and remote access. At this instant, only thirteen structures were prepared for distribution and four were pending. The PDB continued in Brookhaven until 1998, and in 1999 the PDB was transferred to the Research Collaboratory for Structural Bioinformatics (RCSB) [12], under the directorship of Helen M. Berman of Rutgers University [13].

In 2003, the Worldwide PDB (wwPDB) was formed which made PDB an international organization, with RCSB PDB, the Macromolecular Structure Database at the European Bioinformatics Institute (MSD-EBI) and PDB Japan (PDBj) at the Institute for Protein Research at Osaka University, Japan, acting as the founding members [14]. This laid the foundation for the PDB to remain a universal worldwide resource of structural biology data [15].

2.2 Protein structure databases

The structure databases are divided into two subsections based on whether it consists of experimentally determined structures or structure models. All these databases are listed in Table 1.

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Database</th>
<th>Link</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>PDB</td>
<td><a href="http://www.rcsb.org/pdb/home/home.do">http://www.rcsb.org/pdb/home/home.do</a></td>
<td>X-ray &amp; NMR data for biological macromolecules</td>
</tr>
<tr>
<td>2.</td>
<td>PDBsum</td>
<td><a href="http://www.ebi.ac.uk/pdbsum/">http://www.ebi.ac.uk/pdbsum/</a></td>
<td>Pictorial representation of 3D structures</td>
</tr>
<tr>
<td>3.</td>
<td>CATH</td>
<td><a href="http://www.cathdb.info/">http://www.cathdb.info/</a></td>
<td>Protein domain structures</td>
</tr>
<tr>
<td>4.</td>
<td>SCOP</td>
<td><a href="http://scop.mrc-imb.cam.ac.uk/scop/">http://scop.mrc-imb.cam.ac.uk/scop/</a></td>
<td>Familial and structural protein relationships</td>
</tr>
<tr>
<td>5.</td>
<td>MMDB</td>
<td><a href="http://www.ncbi.nlm.nih.gov/structure">http://www.ncbi.nlm.nih.gov/structure</a></td>
<td>3D structures that are linked to NCBI Entrez</td>
</tr>
<tr>
<td>6.</td>
<td>ModBase</td>
<td><a href="http://modbase.compbio.ucsf.edu/modbase-cgi/index.cgi">http://modbase.compbio.ucsf.edu/modbase-cgi/index.cgi</a></td>
<td>Comparative protein structure models</td>
</tr>
<tr>
<td>8.</td>
<td>Database of Macromolecular Movements</td>
<td><a href="http://bioinfo.mbb.yale.edu/MolMovDB">http://bioinfo.mbb.yale.edu/MolMovDB</a></td>
<td>Motions that occur in proteins and other macromolecules</td>
</tr>
<tr>
<td>9.</td>
<td>PROCARB</td>
<td><a href="http://procarb.org/">http://procarb.org/</a></td>
<td>3D structures of protein-carbohydrate complexes &amp; comparative models of glycoproteins</td>
</tr>
<tr>
<td>10.</td>
<td>PMDB</td>
<td><a href="http://mi.caspur.it/PMDB/">http://mi.caspur.it/PMDB/</a></td>
<td>3D protein models obtained by structure prediction methods.</td>
</tr>
<tr>
<td>11.</td>
<td>PDBTM</td>
<td><a href="http://pdbtm.enzim.hu/">http://pdbtm.enzim.hu/</a></td>
<td>Protein Data Bank of Transmembrane Proteins</td>
</tr>
<tr>
<td>12.</td>
<td>OPM</td>
<td><a href="http://opm.phar.umich.edu/">http://opm.phar.umich.edu/</a></td>
<td>OPM provides spatial arrangements of membrane proteins with respect to the hydrocarbon core of the lipid bilayer</td>
</tr>
</tbody>
</table>

Table 1: List of various important protein structure and protein model databases.

2.2.1 Experimentally solved structure databases:

2.2.1.1 Protein Data Bank (PDB)

1. It is the chief source of structural data for biological macromolecules. PDB was founded at Brookhaven National Laboratories (BNL) [16] in 1971 as an archive for biological macromolecular crystal structures [17]. As of May 2013, there are 90206 biological macromolecular structures deposited in PDB [18]. The PDB database contains information regarding experimentally-determined structures of proteins, nucleic acids, and complex assemblies. The data obtained by X-ray crystallography or NMR spectroscopy is deposited globally by biologists and biochemists which is freely accessible to the world wide community. Each PDB file contains xyz coordinates of atoms and an entry in the PDB also includes information regarding the chemistry of the macromolecule, the small-molecule ligands, various particulars of the data collection and structure refinement, and some structural descriptors. Altogether, a characteristic PDB entry has about 400 unique items of data. The PDB file format was formulated in 1976 and is very simple, human readable as well as used by countless computer applications [19].

2.2.1.2 SCOP

The proteins in PDB have structural similarities with other proteins and, may share a common evolutionary source. Therefore, the Structural Classification of Proteins (SCOP) database [20] was created so that access to this information could be facilitated. Besides all the proteins in the current version of PDB, it also includes many proteins for which there are published descriptions but whose co-ordinates do not exist yet. The classification in SCOP is based on hierarchical levels
where the initial two levels, family and superfamily, illustrate close and distant evolutionary relationships where as the third fold describes geometrical relationships. The organization of proteins in SCOP has been created by visual examination and comparison of structures, which offers a possibility of most accurate and useful results keeping in mind the current limitations of purely automatic procedures. A protein domain represents the unit of classification. Small and medium sized proteins have a single domain and are, therefore, treated as a whole where as the domains in large proteins are generally classified separately [21].

For each entry, SCOP provides links to co-ordinates, images of the structure, interactive viewers, sequence data and literature references. The user can search the SCOP database by using two methods. The homology based search allows users to enter a sequence and get a list of structures to which it has significant levels of sequence similarity. The key word search method retrieves, matches from both the text of the SCOP database and the headers of PDB structure files [20]. The most update release (23 Feb 2009) contains 8221 PDB Entries, 110800 domains and 1 literature reference [http:// scop.mrc-lmb.cam.ac.uk/scop/count.html#scop-1.75].

2.2.1.3 CATH

The CATH (Class, Architecture, Topology, Homology) database [22] is a hierarchical classification of protein domain structures, using labor-intensive curation supported by a range of classification and prediction algorithms; for instance, structural comparison [23] and hidden-Markov model (HMM)-based methods [24]. Before splitting of constituent chains each protein structure is verified to make certain it meets the selection criteria. Consecutively, these chains are divided into one or more individual domains and then classified into homologous super families depending on their structure and function [25]. Top of the hierarchy is represented by the Class, or C-level, in which the domains are classified by their secondary structure content—i.e. Mostly alpha-helical (Class 1), mostly beta-sheet (Class 2), both alpha-helical and beta-sheet secondary structure elements (Class 3) or have very little secondary structure (Class 4).Inside each class, the domains are then classified based on their Architecture (A-level)—i.e. similarities in the arrangement of secondary structures in 3D space, which is further sub-divided into one or more topology, or fold groups (T-level), where the connectivity between these secondary structures is taken into account. Lastly, the domains are classified into their particular Homologous super families (H-level), based on the similarities in structure, sequence and/or function. Sequence clustering at the H-level creates sequence families at <35% sequence identity (S-level), <60% (O-level), <95% (I-level) and 100% (I level) [25]. CATH now includes 173536 domains, 1313 folds and 2626 super families.

2.2.1.4 PDBsum

PDBsum is a web-based database that aims to complement the data already available on protein and nucleic acid structures from various sources like CATH [22] SCOP [20] MMDB [26] and NDB [27] for nucleic acids, etc. The database provides a summary of the proteins, nucleic acids, ligands, water molecules and metals in each PDB file in addition to various analyses of their structural features [28]. These include the images of the structure, annotated plots of each protein chain’s secondary structure, thorough structural analyses created by the PROMOTIF program, PROCHECK summary results and schematic representation of protein–ligand and protein–DNA interactions [29]. For each new structure deposited in the PDB, there is a corresponding summary in PDBsum which can be accessed by its four letter PDB code. The latest version of PDBsum contains 93,419 entries (Last update: 4 May, 2013).

2.2.1.5 PDBTM

PDBTM is a protein databank of transmembrane proteins with known structures, and aims to bring together all transmembrane proteins that are submitted in the PDB and to determine their membrane-spanning regions. The PDBTM database was created by scanning all PDB entries with the TMDET algorithm, which employs only structural information to find the most probable position of the lipid bilayer and to differentiate between transmembrane and globular proteins. The database is updated weekly as soon as the new PDB entries are available, by running the TMDET algorithm on every new PDB file. The PDBTM database can be considered as an expansion of the PDB database, since it contains added information for each PDB entry. The database is structured in the same way as the PDB; the entries are identified by their PDB code and are grouped in subdirectories according to the middle two characters of their codes. This database provides a useful resource for people interested in studying transmembrane proteins. For example, it could be used to assign whether a binding site is situated in the lipid or in the aqueous phase, which in turn is of significant value design a drug that binds to a certain part of a receptor [30]. The current version consists of 1891 transmembrane proteins, which include 1626 alpha and 258 beta proteins [http://pdbtm.enzim.hu/].

2.2.1.6 Database of Macromolecular Movements

Motions of proteins are involved in several biological functions such as catalysis, regulation of activity, transport of metabolites, formation of large assemblies and cellular locomotion. Many of the instances of protein structures solved in multiple conformations can now be studied in a database framework and are freely accessible [http://bioinfo.mbb.yale.edu/ MolMovDB]. Protein motions are classified hierarchically initially on the basis of size (distinguishing between fragment, domain and subunit motions) and then based on the packing. Packing classification further divides motions into different categories (e.g. shear, hinge, other) which depends on whether or not they involve sliding over a constantly sustained and firmly packed interface. Additionally, the database gives some hint about the evidence behind each motion (i.e. the type of experimental information or whether the motion is inferred based on structural similarity) and efforts to illustrate numerous aspects of a motion in terms of a standardized nomenclature (e.g. the maximum rotation, the residue selection of a fixed core, etc.). As compared to an individual protein identifier, each entry in the database is indexed by a unique motion
identifier, since a single macromolecule can have a numerous motions and the same vital motion can be shared between diverse macromolecules. Also, each entry has links to graphics and movies describing the motion, frequently portraying a possible interpolated pathway [31].

2.2.1.7 Orientations of Proteins in Membranes (OPM)

OPM database is a compilation of transmembrane, monotopic and peripheral proteins derived from PDB whose spatial arrangements in the lipid bilayer have been theoretically calculated and compared with the experimental data. The OPM database allows various different types of analysis, for instance, sorting and searching of membrane proteins on the basis of their structural classification, species, destination membrane, numbers of transmembrane segments and subunits, numbers of secondary structures and the calculated hydrophobic thickness or tilt angle concerning the bilayer normal. OPM can be browsed either by searching proteins by their name or PDB ID or by sorting of proteins in tables for various specific categories like type, class, superfamily, family, destination membrane or biological source. An individual web page is created for each membrane protein complex, and coordinate files of all proteins with calculated membrane boundary planes are accessible for downloading individually for each protein or as a whole dataset. Currently, OPM has about 2172 proteins grouped into 393 superfamilies and 667 families representing 538 species [32].

2.2.1.8 PROCARB

PROCARB is an open-access database which consists of three separately functioning components, i.e., (i) Core PROCARB module, includes 3D structures of protein-carbohydrate complexes taken from PDB, (ii) Homology Models module, consisting of manually constructed 3D models of N-linked and O-linked glycoproteins by comparative modeling, and (iii) CBS-Pred prediction module, consists of web servers to predict carbohydrate-binding sites using single sequence or server-generated PSSM. In PROCARB, numerous pre-computed structural and functional properties of complexes are also included for rapid analysis. Particularly, information about function, secondary structure, solvent accessibility, hydrogen bonds and literature reference is included. Additionally, each protein in the database is mapped to Uniprot, Pfam, PDB, and so forth. Currently the PROCARB consists of 604 experimentally verified protein carbohydrate complexes and 26 N-linked and 20 O-linked models in which at least one experimentally verified glycosylation site was modeled [33]. Figure 1 shows a representative N-linked homology model of Human Lysosomal alpha-glucosidase with three experimentally verified glycosylation sites (Uniprot ID = P10253).

![Figure 1: 3D model of Human Lysosomal alpha-glucosidase (Uniprot ID = P10253) with three experimentally verified N-linked glycosylation sites ASN390 (green), ASN470 (red) and ASN652 (orange).](image)

2.2.2 Protein structure model databases

2.2.2.1 Protein Model Database (PMDB)

PMDB is a relational database of manually generated protein models deposited by users and attained with different structure prediction methods. The database provides easy access to models that have been published in the scientific literature, simultaneously with validating experimental data. Most of the models in PMDB are the predictions which have been submitted to the Critical Assessment of Techniques for Structure Prediction (CASP) experiment, as well as models generated by PMDB group, and the models uploaded based on published alignments [2]. For each protein target, one or more models could be available or several models for different regions of the same target protein. The database provides some information for each target and includes the protein name, sequence and length, organism and, whenever possible, links to the SwissProt sequence database. As soon as the structure of a target is determined, the PMDB database entry is also linked to the experimental structure in the PDB.

2.2.2.2 Molecular Modeling Database (MMDB)

MMDB offers simple access to the richness of 3D structure data and its huge potential for functional annotation [34]. MMDB reflects the contents of the PDB and is strongly integrated with NCBI’s Entrez search and retrieval system. In MMDB, protein 3D structure data is connected with sequence data, sequence classification resources and PubChem, providing easy access to 3D structure data for structural biologists, as well as for molecular biologists and chemists [35]. An
entire set of comprehensive and pre-computed structural alignments are obtained with the VAST algorithm [36] where as the visualization tools for 3D structure and structure/sequence alignment are provided by the molecular graphics viewer Cn3D [37].

As on April 29, 2013, there are 89,571 structure records total which includes 22,084 proteins, 715 DNA and 508 RNA molecules only. Additionally, MMDB also consist of 2571 protein-DNA complexes, 1113 protein-RNA complexes and 116 protein-DNA-RNA complexes, in addition to more than 60,000 proteins bound to chemicals.

### 2.2.2.3 ModBase

ModBase is a database of annotated homology based protein structure models. Models in ModBase are generated as an automated software pipeline for comparative protein structure modeling, known as ModPipe [3] which mostly rely on modules of Modeller [38]. For fold assignment and target–template alignment, ModPipe uses sequence–sequence [39] sequence–profile [40,41] and profile–profile [40,42] methods by using an E-value cut-off of 1.0 to augment the possibility of identifying the finest available template structure. 10 models are generated [38] for each target–template alignment, and the model with the top Discrete Optimized Protein Energy (DOPE) statistical potential [43] score is selected and further assessed by numerous additional quality criteria: (i) target–template sequence identity, (ii) GA341 score [44] (iii) Z-DOPE score [43] (iv) ModPipe Quality Score (MPQS) and (v) TSVMod score [45].

Because of the rapid growth of the public sequence databases, models in ModBase are structured in data sets that are useful for specific projects. Currently, ModBase includes about 27,288,148 models and 4,332,658 unique sequences modeled for more than 50 complete genomes [46]. ModBase can be queried through its web interface by querying with UniprotKB [47] and GI [48] identifiers, gene names, annotation keywords, PDB codes [49] data set names, organism names, sequence similarity to the modeled sequences (BLAST [41]) and model-specific criteria such as model reliability, model size and target–template sequence identity. Additionally, the coordinate and alignment files can also be retrieved as text files [50].

### 2.2.2.4 SWISS-MODEL Repository

SWISS-MODEL Repository is a database of 3D protein structure models constructed by using the SWISS-MODEL homology-modelling pipeline based on protein sequences from the UniProt database [47]. The SWISS-MODEL pipeline integrates various steps like: template selection, target sequence and template structure alignment, model building, energy minimization and/or refinement and model quality assessment [51]. Model target sequences are individually identified by their md5 cryptic hash of the full sequence. This permits the redundancy in protein sequence databases entries to be reduced, and in turn assists cross-referencing with databases by means of different access codes. The current SWISS-MODEL Repository release contains 3143784 model entries for 2286870 unique sequences in the UniProt database (2013_02).

The database could be queried for particular proteins by using diverse database accession codes (e.g. UniProt AC and ID, GenBank, IPI, Refseq) or directly by means of the protein amino acid sequence. For a particular query protein, a graphical outline demonstrating the segments for which models or experimental structures are available is shown. SWISS-MODEL Repository users can review the quality of the models in the database; search for alternative template structures, and construct models interactively by the use of SWISS-MODEL Workspace [52]. Repository is updated on a regular basis to reflect the growth of the sequence and structure databases.

### 3. 3D Structure Prediction

#### 3.1 A brief history of molecular modelling

The first homology based model dates back to 1969 when a wire and plastic models of bonds and atoms of α-lactalbumin was constructed by using the coordinates of a hen’s egg-white lysozyme and adjusting, physically, those amino acids that did not match the structure [53]. The two proteins exhibited 39% of sequence identity. Afterwards, the crystal structure of lysozyme was used to generate a model for α-lactalbumin [54]. These models were created by taking the existing coordinates of the well-known structure, and mutating side chains that were not identical in the protein to be modeled. This approach to protein modeling is still used at present with substantial success, particularly when the proteins share a considerable degree of sequence similarity [55].

McLachlan and Shotton [56] used the structures of mammalian chymotrypsin and elastase, and modeled the structure of α-lytic proteinase of the fungus Myxobacter 495. The modeling was not easy as the sequence similarity between the target and the template was only about 18%. Subsequently, the crystal structure of α-lytic proteinase was determined and compared with the homology model [57]. Although the domains of the model were constructed accurately, it was found that misalignment of the sequences led to local errors.

The modeling of variable regions was introduced in proteins on the basis of equivalent regions from homologous proteins of known structures [58,59]. Therefore, in order to construct the homology models of various serine proteases, structures of trypsin, chymotrypsin and elastase were superimposed, and it was found many equivalent Ca atoms were within 1.0Å of one another. The regions comprising of the amino acids of these Ca atoms were described as structurally conserved regions (SCRs). All the other remaining sites correspond to structurally variable or loop regions (VR) where the insertions/deletions were located. The backbone of SCRs and VRs was generated from the fragments of known serine proteases, where as the side chains were modeled based on the conformation found at the equivalent locations for those identical side chains in the well-known structures.
Furthermore, the initial models for the aspartic proteinases renin and renin-inhibitor complexes were built by using the 3D structure of the remotely related fungal proteinases [60-62]. Later on, the models for renin were constructed by employing the structures of mammalian aspartic proteases, pepsin and chymosin [63,64]. Comparative analysis of fungal and mammalian renin models revealed that the inaccuracies in the models occurred due to the dissimilarity in the arrangement of helices and strands between the fungal and mammalian proteinases, as well as the slightly different variable regions. On the other hand, the modeling of renin active site was reasonably accurate [65].

Early in the eighties, manual homology modeling was assisted by manoeuvring of protein molecules on the graphics terminal that was made achievable by computer programs like FRODO [66]. Since then, many different homology modeling packages have been developed [42], which can be grouped into three different groups: rigid-body assembly, segment matching, or modeling by satisfaction of spatial restraints [67].

### 3.2 Protein 3D structure prediction tools

The prediction of 3D structures of proteins remains an exceedingly complicated and uncertain undertaking. However, these difficulties can be addressed up to a certain extent by using some of the key state of the art tools which have been developed over the years. These tools (Table 2) either employ homology based methods or *Ab initio* methods in case of no significant similarities are found.

<table>
<thead>
<tr>
<th>SNo.</th>
<th>Software</th>
<th>Link</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>MODELLER</td>
<td><a href="http://salilab.org/modeller/">http://salilab.org/modeller/</a></td>
<td>Satisfaction of spatial restraints</td>
</tr>
<tr>
<td>2.</td>
<td>SWISS-MODEL</td>
<td><a href="http://swissmodel.expasy.org/">http://swissmodel.expasy.org/</a></td>
<td>Local similarity/fragment assembly</td>
</tr>
<tr>
<td>3.</td>
<td>3D-JIGSAW</td>
<td><a href="http://bmm.cancerresearchuk.org/~3djigsa/">http://bmm.cancerresearchuk.org/~3djigsa/</a></td>
<td>Fragment assembly</td>
</tr>
<tr>
<td>4.</td>
<td>ROBETTA</td>
<td><a href="http://robetta.bakerlab.org/">http://robetta.bakerlab.org/</a></td>
<td>Rosetta homology modeling and ab initio fragment assembly with Ginuzu domain prediction</td>
</tr>
<tr>
<td>5.</td>
<td>RaptorX</td>
<td><a href="http://raptorx.uchicago.edu/">http://raptorx.uchicago.edu/</a></td>
<td>Remote homology detection, protein 3D modeling, binding site prediction</td>
</tr>
<tr>
<td>6.</td>
<td>ESYPred3D</td>
<td><a href="http://www.unamur.be/sciences/biologie/urbm/bioinfo/espred/">http://www.unamur.be/sciences/biologie/urbm/bioinfo/espred/</a></td>
<td>Template detection, alignment, 3D modeling</td>
</tr>
<tr>
<td>7.</td>
<td>HHpred</td>
<td><a href="http://toolkit.tuebingen.mpg.de/hhpred">http://toolkit.tuebingen.mpg.de/hhpred</a></td>
<td>Template detection, alignment, 3D modeling</td>
</tr>
<tr>
<td>8.</td>
<td>EasyModeller</td>
<td>NA</td>
<td>GUI to MODELLER</td>
</tr>
<tr>
<td>9.</td>
<td>CPHModel</td>
<td><a href="http://www.cbs.dtu.dk/services/CPHmodels/">http://www.cbs.dtu.dk/services/CPHmodels/</a></td>
<td>Fragment assembly</td>
</tr>
<tr>
<td>10.</td>
<td>BHAGEERATH-H</td>
<td><a href="http://www.scbio-itd.res.in/bhageerath/hbageerath_h.jsp">http://www.scbio-itd.res.in/bhageerath/hbageerath_h.jsp</a></td>
<td>Combination of ab initio folding and homology methods</td>
</tr>
<tr>
<td>13.</td>
<td>PSIPRED Protein Structure Analysis Workbench</td>
<td><a href="http://bioinf.cs.ucl.ac.uk/psipred/">http://bioinf.cs.ucl.ac.uk/psipred/</a></td>
<td>Several high quality protein structure prediction and function annotation algorithms</td>
</tr>
<tr>
<td>14.</td>
<td>I-TASSER</td>
<td><a href="http://zhanglab.ccb.med.umich.edu/I-TASSER/">http://zhanglab.ccb.med.umich.edu/I-TASSER/</a></td>
<td>Combination of ab initio folding and threading methods</td>
</tr>
<tr>
<td>15.</td>
<td>QUARK</td>
<td><a href="http://zhanglab.ccb.med.umich.edu/QUARK/">http://zhanglab.ccb.med.umich.edu/QUARK/</a></td>
<td>Monte Carlo fragment assembly</td>
</tr>
<tr>
<td>16.</td>
<td>MUSTER</td>
<td><a href="http://zhanglab.ccb.med.umich.edu/MUSTER/">http://zhanglab.ccb.med.umich.edu/MUSTER/</a></td>
<td>Profile-profile alignment</td>
</tr>
<tr>
<td>17.</td>
<td>SPARKS-X</td>
<td><a href="http://sparksinformatics.iupui.edu/yueyang/sparks-x/">http://sparksinformatics.iupui.edu/yueyang/sparks-x/</a></td>
<td>3D structure modeling by Fold recognition according to Sequence profiles and structural profiles</td>
</tr>
<tr>
<td>18.</td>
<td>Phyre and Phyre2</td>
<td><a href="http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index">http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index</a></td>
<td>Remote template detection, alignment, 3D modeling, multi-templates, ab initio</td>
</tr>
</tbody>
</table>

Table 2: List of protein 3D structure prediction tools.

### 3.2.1 Homology modeling

3D structure of a protein is capable of providing invaluable information about the function of a protein and allowing an efficient design of experiments, for instance site-directed mutagenesis, studies of disease-related mutations or the structure based drug designing efforts [68]. Traditional approaches to determine the 3D structure of a protein includes X-ray crystallography or NMR spectroscopy. Other theoretical methods have not shown much promise in providing high-resolution information for the bulk of proteins. The number of structurally characterized proteins is very less in comparison to the number of known protein sequences. As of May 07, 2013, there are 90,424 structures in PDB [http://www.rcsb.org/pdb/home/home.do] which is extremely low as compared to UniProtKB/Swiss-Prot (http://www.uniprot.org/) which contains 5,40,052 sequence entries as of May 01, 2013. Nevertheless, it seems quite unreasonable to believe that it is possible to experimentally determine the structures of all these hundreds and thousands of proteins regardless of immense growth in the efforts of structural genomics. Therefore, in view of the above, homology modeling (also known as comparative modeling) methods offer the only possible way to get structural information for such a huge number of proteins [69].

One of the prerequisites of successful model building requires the availability of at least one experimentally determined 3D structure known as template that shares a significant amino acid sequence similarity to the target sequence [68]. The main steps that are required to create a homology based model are summarized in Figure 2 and include: (1) identification of homologs that can be used as templates for modeling; (2) target-template sequence alignment; (3) building a model for the target based on the information from the alignments; and (4) evaluation of the model [70,71]. These modeling steps usually involve extensive expertise in structural biology and the use of extremely specialized computational tools [72]. Some of these highly specialized and frequently used homology based modeling tools are summarized below.
Target EPPVPDTCGHVAEERTQFAELTTKLSELQENVVTNFHGNC

Template search

Structure Database

Target-Template alignment

Model Building & Evaluation

Figure 2: The target sequence is first searched against a structure database (preferably PDB) for an appropriate template. Target-template sequence alignment is created after the template with significant sequence similarity is found. The model is generated based on the target-template alignment and evaluated finally.

3.21.1 Modeller

Modeller is prediction software regularly used for creating homology or comparative protein structure models for proteins lacking experimentally determined 3D structures. It employs a method known as satisfaction of spatial restraints motivated by NMR spectroscopy data processing, by means of which a set of geometrical criteria are used to produce probability density functions (pdfs) for the location of each atom in the protein.

The Ca-Ca distances, main-chain N-O distances, main-chain and side-chain dihedral angles are restrained by these pdfs. In order to reduce the problem of a sparse database a smoothing method is used in the derivation of these relationships. The 3D model of a protein is acquired by optimization of the molecular pdf so that the model defies the input restraints as slightly as possible. The optimization method is a variable target function process that applies the conjugate gradients algorithm to positions of all non-hydrogen atoms. All these steps in Modeller are fully automated [38].

Modeller can perform additional tasks also, like de novo modeling of loops in protein structures, optimization of various models of protein structure with respect to a flexibly defined objective function, multiple alignments of protein sequences and/or structures, clustering, sequence database search and comparison of protein structures [73]. MODELLER was originally written in FORTRAN 90 and maintained at the University of California, San Francisco. It runs on UNIX, windows and MAC computers via scripts written in the Python language and is freely available for academic use. It does not provide any graphical interface however; a freely available GUI to MODELLER called Easy Modeller is developed and available for free download for Linux and Windows OS. A graphical user interfaces and commercial versions are also distributed as part of Accelrys' InsightII, and Discovery Studio interactive molecular modeling programs, which also contain many other tools for protein modeling and structural analysis. These programs facilitate preparation of input files for MODELLER as well as an analysis of the results.

3.21.2 SWISS-MODEL

SWISS-MODEL is an automated web server to build protein structure homology models, accessible via the ExPASy web server and program Deep View. The SWISS-MODEL provides a personalized web-based area for each user in which protein homology models can be built, and the results are stored and analyzed [74]. A personal account for the SWISS-MODEL workspace can be made and accessed at http://swissmodel.expasy.org/workspace. The user data is stored in a password-protected personal user space. For a given target protein in SWISS-MODEL, a library of experimental protein structures is searched to identify suitable templates. On the basis of a sequence alignment between the target protein and the template structure, a three-dimensional model for the target protein is generated. For template identification is carried out using various options like BLAST, PSI-BLAST and HMM-HMM-based searching. Automated mode can be applied if the alignment between the target and the template sequences exhibits a sufficiently high similarity where as alignment mode can be useful for distantly related target and template sequences. In case the sequence alignments fall within so-called ‘twilight zone’, i.e. when the sequence identity between target and template is lower than 30%, it is worthwhile to visually inspect and manually edit the target–template alignments. This leads to a considerable improvement in the quality of the resulting model and project mode can be applied [75]. To estimate the quality of model(s), programs provided in the ‘Structure Assessment’ section under tools using options sequence identity, stereochemistry check, global model quality and local model quality estimation are available. The quality assessment methods that are embedded in SWISS-MODEL workspace consists of Anolea mean force potential plots [76], GROMOS empirical force field energy [77], Verify3D profile
The SWISS-MODEL Repository is also available for the annotation of three-dimensional comparative protein structure models generated by homology-modeling pipeline SWISS-MODEL [51].

3D-JIGSAW is a web-based server which generates 3D models for proteins based on homologues of known structure. The program can either be run in completely automatic mode by means of a web server or the individual modules of the program may be executed separately and the intermediate files saved which can be modified, if necessary, before the next module in the series is run [75]. In automatic mode, the program looks for homologous templates in its sequence databases and splits the query sequence into domains and the best covered domain is modeled using a maximum of two good templates, if found. The process may take up to an hour on a system. An e-mail with the alignment between query and template and a PDB formatted set of coordinates is sent to user, while in an interactive mode an e-mail is sent back to user with a link to a graphical display of the domain arrangement and useful information extracted from the PFAM database. From this link, user can select the domains needed to be modeled and can select the templates and correct the alignments before submitting a modeling job. Templates are ranked according to the coverage of the query, their sequence identity and their crystallographic resolution. Information from each template is easily accessed, including its alignment to the query sequence. In order to remove the steric clashes, the models are subjected to 100 steps of steepest descents energy minimization (unrestrained) by using the program CHARMM [81]. Additionally, error approximations were made by measuring the range of equivalent atomic displacements, as calculated from the superposition of all significant homologues [82]. A new version of 3D-JIGSAW (version 3.0) is also available where templates are identified using HMM [83] and the returned alignments are used to generate the models.

3D-JIGSAW

3D-JIGSAW is a web-based server which generates 3D models for proteins based on homologues of known structure. The server utilizes the completely automated structure prediction process that constructs a model for a whole protein sequence whether the sequence homology to proteins of known structure is available or not. Robetta breaks down the input sequences into domains and generates models for domains with sequence homology to proteins of known structure using comparative modeling where as models for domains lacking such homology are constructed using the Rosetta de novo structure prediction method. Domain predictions and molecular coordinates of models across the full-length query are given as results [84].

The server can also exploit nuclear magnetic resonance (NMR) constraints data supplied by the user to resolve protein structures using the RosettaNMR [85,86] protocol. These tools can be used in combination with existing structural genomics initiatives to aid accelerate structure determination and expand structural insight for targeted open reading frames (ORFs). Robetta uses a somewhat modified edition of the de novo structure prediction method [87] so that the queries can be run within reasonable timescales. Similar to the original protocol, Robetta generates three- and nine-residue fragment libraries that correspond to local conformations seen in the PDB, and then assembles models by fragment insertion using a scoring function that favors protein-like characteristics. About 10000 decoys are generated for the original query where as 5000 decoys for up to two sequence homologs are generated by Robetta. Subsequently, 2000 query decoys and 1000 decoys of each homolog are selected on the basis of score, and on whether they surpass filters that remove decoys having too many local contacts or unlikely strand topologies. The selected decoys are then clustered based on Ca root-mean-square deviation (RMSD) above the entire ungapped positions. The top 9 cluster centers are selected as the top ranked models, and the finest scoring model that passed the filters is chosen as the 10th model. Searches with these final models are then carried out against a representative set of PDB chains to discover comparable structures using Mammoth [88] to recognize potential similarities to proteins of known structure.

In view of the fact that it is difficult to do de novo structure predictions with high accuracy, two moderately liberal definitions of correct predictions were defined. A domain is considered to be accurately modeled if at least 1 of the 10 models has a Mammoth alignment of 50 or more residues with an RMSD of 4Å or less to the native structure, or a Mammoth Z-score of 6 or better to the native structure. Robetta produces correct predictions for more than half of the domains on the basis of these definitions, especially with domains that consist of either all-alpha or alpha-beta secondary structure.

3.21.4 Robetta

Robetta web server offers automated structure prediction and analysis tools to deduce protein structural information from genomic data. The server utilizes the completely automated structure prediction process that constructs a model for a whole protein sequence whether the sequence homology to proteins of known structure is available or not. Robetta breaks down the input sequences into domains and generates models for domains with sequence homology to proteins of known structure using comparative modeling where as models for domains lacking such homology are constructed using the Rosetta de novo structure prediction method. Domain predictions and molecular coordinates of models across the full-length query are given as results [84].

The server can also exploit nuclear magnetic resonance (NMR) constraints data supplied by the user to resolve protein structures using the RosettaNMR [85,86] protocol. These tools can be used in combination with existing structural genomics initiatives to aid accelerate structure determination and expand structural insight for targeted open reading frames (ORFs). Robetta uses a somewhat modified edition of the de novo structure prediction method [87] so that the queries can be run within reasonable timescales. Similar to the original protocol, Robetta generates three- and nine-residue fragment libraries that correspond to local conformations seen in the PDB, and then assembles models by fragment insertion using a scoring function that favors protein-like characteristics. About 10000 decoys are generated for the original query where as 5000 decoys for up to two sequence homologs are generated by Robetta. Subsequently, 2000 query decoys and 1000 decoys of each homolog are selected on the basis of score, and on whether they surpass filters that remove decoys having too many local contacts or unlikely strand topologies. The selected decoys are then clustered based on Ca root-mean-square deviation (RMSD) above the entire ungapped positions. The top 9 cluster centers are selected as the top ranked models, and the finest scoring model that passed the filters is chosen as the 10th model. Searches with these final models are then carried out against a representative set of PDB chains to discover comparable structures using Mammoth [88] to recognize potential similarities to proteins of known structure.

In view of the fact that it is difficult to do de novo structure predictions with high accuracy, two moderately liberal definitions of correct predictions were defined. A domain is considered to be accurately modeled if at least 1 of the 10 models has a Mammoth alignment of 50 or more residues with an RMSD of 4Å or less to the native structure, or a Mammoth Z-score of 6 or better to the native structure. Robetta produces correct predictions for more than half of the domains on the basis of these definitions, especially with domains that consist of either all-alpha or alpha-beta secondary structure.

3.21.5 ESyPred3D

ESyPred3D is an automated homology modeling web server. It implements four steps of homology modeling approach which includes database searching to identify structure homolog, target template alignment, model building & optimization and model evaluation. ESyPred3D is available for use at http://www.fundp.ac.be/urbm/bioinfo/esypred/. The data and parameters required to submit the simplest job are the sequence, email id and description of job. In the advance parameter, PDB ID with chain name can be given to use this as template or PDB ID can be submitted to discard it as template. User has
the option to get results of submitted job in different format and alignment method provided at the web server. The method in EasyPred3D program gets benefit of the increased alignment performances of a new alignment strategy using neural networks, where alignments are obtained by combining, weighting and screening the results of several multiple alignment programs. The final three dimensional structures are built using the modeling package Modeller by the satisfaction of spatial and geometric restraints and an extremely rapid molecular dynamic annealing. Comparison has been made to test the accuracy of EasyPred3D models in CASP4 experiment and it was found that alignment strategy of EasyPred3D is better as compared to PSI-BLAST alignment and also models generated by EasyPred3D were among the best of CASP3 experiments [89].

### 3.21.6 CPHModel

In CPHModel, template recognition is done on the basis of profile-profile alignment guided by secondary structure and exposure predictions and in the new version, improvement in the alignment algorithm for the remote homology modeling has been introduced by structure dependant gap penalty [90]. The response time of the CPHModel web server is less than 20 minutes. Only one sequence per submission is allowed with not less than 15 and not more than 4,000 amino acids and the sequences are kept confidential and will be deleted after processing. In Query sequence section, the query sequence submitted and template hits obtained with the e-value are shown. In retrieving template section, if any significant hits were found in the PDB, the PDB entry name and the chain identifier are listed for the template that is used to construct the model. In making profile-profile alignment section, the resulting sequence alignments of using the PDB-Blas hit are shown. The results from the remote homology modeling is shown in another section and if a MSA is submitted by clicking on the link ‘query.pdb’ can be downloaded in the in PDB format and finally if using a java-enabled browser, the trace of the model will be shown [90].

### 3.21.7 BHAGEERATH-H

BHAGEERATH-H is a homology-Ab initio hybrid web server for protein structure prediction, and is available at http://www.scbio-iitd.res.in/bhageerath/bhageerath_h.jsp. For each given input sequence, the web server predicts 5 structures. For sequences greater than 100 amino acids the results may take longer time to complete. In the template information section, user can also select auto template search by using Bhageerath-H template search protocol. User can also give as many as references with option PDB ID and its chain ID by clicking ADD button. For sequences with homologs, it has the potential to predict a structure with higher resolution and accuracy in less time [91]. The models are ranked using all atom energy based empirical scoring function [92] and selects 100 lowest energy structures. The number of models are further reduced to five using solvent accessible surface areas (SASA) [http://www.bioinf.manchester.ac.uk/naccess/] which is further refined to five models by an optional short molecular dynamics simulations with explicit solvent.

### 3.21.8 EasyModeller

EasyModeller is a GUI standalone tool for homology modeling using MODELLER in the backend and was available initially for Windows at http://www.uohyd.ernet.in/modellergui/ and now Linux platform also. The GUI eliminates the requirement of prior knowledge of backend applications, consequently increasing the number of users of MODELLER and assists them to exploit the unique features of package more efficiently. EasyModeller uses default parameters for most commands to make the process as simple as possible. User can change the parameters manually by editing the associated script file in the working directory. It is freely available and its usage requires pre installed Modeller, perl, python, Microsoft Excel plot function and also Rasmol to view the modeled structure. User is required to follow the numbered steps one by one and is guided by the provided help panel. A very simple color coding consisting of green and red buttons is followed by EasyModeller where green represents the optional steps where as red are the essential steps to obtain a model. A maximum of six templates can be used in case of multi-template based modeling [93].

### 3.21.9 GeneSilico

GeneSilico is a WWW ‘meta-server’ which serves as a gateway to numerous protein structure prediction methods, and addresses the key issues of multiple sequence alignment (MSA) submission and data confidentiality [94]. The input to this meta-server could be either a single sequence or the alignment of the sequences. In case of single sequences each method generates its own MSA, whereas, if a MSA is submitted, the user can select between submissions of a full-length query or limit the analysis to a region with less than 30% gaps in the alignment. This allows elimination of highly divergent loops that may cause difficulty when the core structures of the template and the target are matched. The user-defined MSA is submitted to those fold recognition servers which allow the submission of MSA. For servers which allow the submissions of single sequences only and generate their own MSA, the user defined MSA is converted into a ‘consensus sequence’. The consensus sequence can be generated by any of the alternate methods available. The quality of the target-template alignments obtained by GeneSilico meta-server ensured its win in the CASP5 homology modeling contest [95].

The various components that coordinately work in GeneSilico include: a) HMMPFAM: A tool for the identification of PFAM domains [96]; b) Secondary structure is predicted by using methods like PSIPRED [97], SAM-T02 [98] and PROF [99]; c) Identification of potential transmembrane helices is carried out by methods like MEMSAT2 [100], TMHMM [101], TMPRED [102]; d) A local PDB-BLAST filter is employed for the detection of closely related sequences of known structures in PDB; e) The FR methods comprising of RAPTOR [103], 3DPPSSM [104], FUGUE [105], mGENTHREADER [106], FFAS [107], SAM-T02 [108] and BIOINBGU [109] represent the 3D structure prediction core; f) The results retrieved from these FR servers are analyzed using the consensus server PCONS [110], which ranks the best models generated by the FR methods; g) Based on the target-template alignments from FR methods, initial 3D models of the query structure are created using SCWRL [111], on the basis of the template backbone. These preliminary models do not contain the features
corresponding to gaps in the FR alignment (for example insertions in the target, absent from the template), however, the structure of the hydrophobic core is generally inferred sufficient to carry out 3D structure assessment using VERIFY3D [112]. Thus, all FR alignments achieved from diverse servers go through unified evaluation by energetic criteria employed in VERIFY3D in addition to the ranking criterion obtained by the PCONS server.

3.21.10 Geno3D

It is an automatic web server that uses distance geometry, simulated annealing and energy minimization algorithms to build the protein 3D models. The steps taken by the server to generate the homology models of the query sequence include: (a) identification of homologous proteins with known 3D structures by using PSI-BLAST; (b) in the second step all potential templates are provided to the user via a very handy user interface for target selection; (c) the alignment of both target and template sequences is performed; (d) the geometrical restraints (dihedral angles and distances) for corresponding atoms between the query and the template are extracted; (e) finally, the 3D models of the protein are constructed by using a distance geometry approach which are sent to the user by e-mail. The results which are sent to the user includes files containing atomic coordinates of each model (3 is the default value) that best satisfies the spatial restraints and the alignment between target and template sequences, and the superimposition (which highlights the poorly defined regions that correspond to gaps in the alignment) of the models with the template. In addition of these, a matrix of binaries root mean square deviations between coordinates of equivalent α-carbons is also sent to the user, in order to check the homogeneity of the models. The advantages in employing a distance geometry based approach in homology modeling are no a priori preference in loops construction and effortless recognition of well defined regions [113].

3.21.11 The PSIPRED Protein Sequence Analysis Workbench

The PSIPRED server maintained by the UCL Bioinformatics Group presents several high quality protein structure prediction and function annotation algorithms including PSIPRED- for the prediction of secondary structure (Jones, 1999), pGenTHREADER (fold recognition) and pDomTHREADER (domain recognition) [114], MEMSAT- transmembrane topology prediction [115,116], MetSite- metal binding sites [117], DISOPRED2- regions of protein disorder [118], DomPred- prediction of protein domain boundaries [119] and FFPred- protein function prediction [120] respectively. The web portal also offers a fully automated 3D modeling pipeline based on fragment-assembly approach, “BioSerf”, which was placed in the top five servers in the de novo modeling category in CASP8 experiment [121].

3.2.2 Protein Threading

With the rapid growth in PDB and the improvement in structure prediction methods, Template-based modeling which includes homology modeling and protein threading is proving to be more powerful and key method for protein structure prediction. This suggests that the structures of several novel proteins can be predicted by means of template-based methods. The inaccuracy of a template-based model results from the selection of template and the alignment between sequence and template [122]. In case of higher sequence identity (>50%), template-based models can be accurately applied in virtual screening studies [123,124], designing site-directed mutagenesis experiments [125,126], small ligand docking prediction, and function prediction [1,127]. However, when sequence identity falls in the twilight zone (<30%), then it becomes difficult to identify the top most template and generate precise sequence-template alignments [128,129]. The alignments between the sequence and template can be significantly improved by integrating additional information in the form of sequence profile into the scoring function [122].

Methods such as HHpred [83], MUSTER [130], Phyre2 [131] and SPARKS/SP3/SP5 [132-136], are some of the prominent threading based methods which combine homologous information with a variety of structure based information for remote homolog detection.

3.2.2.1 HHpred

Functional information of a protein or gene can be predicted from homologous proteins or genes identified by various sequence based search methods like BLAST, FASTA or PSI-BLAST [41,137,138]. However, there are certain groups of proteins for which no orthologs have been indentified; therefore, it becomes quite difficult to establish a relationship to a protein with known function [83]. HHpred is one of the web-based servers which could be utilized when database search results from BLAST or PSI-BLAST retrieves insignificant hits against proteins having well known structures and functions [83]. For each query sequence in HHpred, an alignment of homologs is generated using PSI-BLAST search against non-redundant (NR) database of NCBI. PSIPRED [97] is used to predicted secondary structure and annotate the alignment from PSI-BLAST. This annotated multiple alignment is then used to generate a profile hidden Markov models (HMMs). Input to HHpred can be a single sequence or a multiple alignment. Search preferences include both local or global alignment and scoring secondary structure similarity. As an output HHpred produces pair wise query-template alignments, multiple alignments of the query with a set of templates selected from the search results, as well as 3D structural models generated by the MODELLER software from these alignments. HHpred also provides a high level of flexibility and user-friendliness while maintaining exceptional sensitivity.

HHpred is a web based server for remote protein homology detection and structure prediction and is the first to employ pair wise comparison of profile hidden Markov models (HMMs) and allows searching a variety of databases, like PDB, SCOP, Pfam, SMART, COGs and CDD. HHpred is very fast and performs uniformly well for single-domain as well as multi-domain query sequences and can be used to predict domain boundaries.
3.2.2.2 MUSTER

MUSTER (MUlti-Source ThreadER) is an extension of sequence profile–profile alignment (PPA) threading algorithm [139] which includes a variety of sequence and structural resources generated from several other tools. PPA algorithm has been effectively employed in tools like I-TASSER and LOMETS. MUSTER uses sequence and structure information like (1) sequence profiles; (2) secondary structures; (3) structure fragment profiles; (4) solvent accessibility; (5) dihedral torsion angles; (6) hydrophobic scoring matrix and merges them into single-body terms which can be easily used in dynamic programming search [130].

The major aim in this method is to thoroughly examine the increase that can be obtained in fold recognition when the different resources of structural features are combined with the powerful sequence profile–profile alignment methods. In order to identify the top match between the query and the template sequences, the Needleman-Wunsch [140] dynamic programming algorithm is used. The output of the MUSTER server includes top five template proteins and the query-template alignments and the models generated by MODELLER. The performance of MUSTER is improved over PPA because of the contribution of more accurate alignments by incorporating new structure features.

3.2.2.3 PHYRE

Phyre employs a library of well-known protein structures [131] retrieved from the Structural Classification of Proteins (SCOP) database [20] and increased with updated additions in the PDB. A profile is created and deposited in the fold library by scanning the sequence of all these structures against a non-redundant (NR) database. Additionally, the known and predicted secondary structure of these proteins is stored in the fold library as well.

A query sequence is scanned against the NR sequence database to construct profile in a similar manner. To accommodate together close and remote sequence homologs, five iterations of PSI-Blast are used. The resulting large numbers of pair wise alignments generated by PSI-Blast are pooled into a single alignment with the query sequence as the master. After the profile has been constructed, the three-state secondary structure (alpha helix (H), beta strand (E—for extended) and coil (C)) of the query is predicted using Psi-Pred [141] SSPro [142] and jNet [143].

Afterwards, the profile and predicted secondary structure is scanned against the fold library using a profile–profile alignment algorithm [144] which in turn returns a score on which the alignments are ranked. An E-value is generated by fitting these scores to an extreme value distribution. Finally, 3D models of the query are generated based on the top ten highest scoring alignments. Additionally, loop library and reconstruction procedure are used to repair missing or inserted regions caused by insertions and deletions in the alignment wherever possible. Moreover, by using a fast graph-based algorithm and side chain rotamer library, side chains are placed on the models. As an input, Phyre web server accepts an amino-acid query sequence and after about 30 minutes, a link related to the results including full downloadable 3D models of the query protein are sent to the user via email. The present publicly available Phyre server showed performance typical of the majority.

3.2.2.4 SPARKS-X

A chain of various single fold recognition methods like SPARKS, SP2, SP3, SP4 and SP5 were developed based on weighted matching of multiple profiles [145] including sequence profiles generated from multiple sequence alignment [41] predicted versus actual secondary structures [134,136,146] knowledge-based profile (single-body) score function [134] depth-dependent sequence profiles derived from template structures [136] predicted versus actual solvent accessible surface area [147] and predicted versus actual dihedral angles [133]. Among these, SPARKS, SP3 and SP4 were ranked amid the top performers for automatic servers in CASP 6 [135,148] and CASP 7 [147,149] experiments. However, a known concern in the above mentioned methods is that matching predicted 1D profiles of query sequence with actual profiles of templates is based on simple difference matrices.

These methods do not account for the likelihood of errors in predicted 1D structural properties; for instance secondary structure, backbone torsion angles and solvent accessible surface area. Therefore, in order to overcome these issues, energy terms based on the estimated probability of a match between predicted and actual 1D structural properties were introduced, a common technique used in fold recognition based on hidden Markov models [150]. This new web-based method is called as SPARKS-X and takes an additional advantage of newly improved accuracy in predicted secondary structure, torsion angles, mean absolute error and solvent accessibility [145]. The models are generated by modeler [38] based on the alignment created by SPARKS-X. In case of gaps of > 30 residues in the termini, the program will be evoked again to construct a model for the missing parts in the region. Subsequently, these different models are linked and steric clashes are removed by using the DFIRE potential functions [151,152].

SPARKS-X turned out to be one of the best single-method fold recognition servers as indicated by CASP 9 when tested for its alignment accuracy, fold recognition and structure prediction by using several benchmarks and compared it to several state-of-the-art techniques. The authors believe that the method can be further improved significantly by integrating the techniques of multiple templates and refinement in model building that are already being employed by many other automatic servers.

3.2.2.5 RAPTORX

RAPTOR – RApid Protein Threading by Operation Research technique is a novel linear programming approach to do predict 3D structure of proteins. In this approach, the protein threading dilemma is devised as a large scale integer
programming (IP) problem based on the contact map graph of the protein 3D structure template. Furthermore, the IP formulation is then relaxed to a linear programming (LP) problem, and later resolved by the canonical branch-and-bound scheme. RAPTOR extensively outperforms other programs at the fold similarity level as shown by benchmark tests for fold recognition. RAPTOR was ranked as top 1 among individual prediction servers, in terms of the recognition capability and alignment accuracy for Fold Recognition (FR) family targets as evaluated by CAFASP3 [153].

However, RAPTOR was significantly outperformed by RaptorX which is good at aligning proteins with sparse sequence profile. RaptorX utilizes a statistical knowledge based method to design a new threading scoring function, which aims at enhanced measuring the compatibility between a target sequence and a template structure. Additionally, RaptorX also has a multiple-template threading component and contains a new module for alignment quality prediction. RaptorX includes three main components: single-template threading, alignment quality prediction, and multiple-template threading. Initially, RaptorX aligns a specific target to each of its templates using the single-template alignment algorithm. Then, the alignment quality is predicted and all the templates are ranked by this predicted quality in a descending order. If the target is not appropriate for multiple-template threading, RaptorX generates a 3D model for the target from the pairwise alignment by means of the highest predicted quality. Else, RaptorX runs multiple-template threading for the target and constructs a corresponding 3D model. RaptorX does extremely well at the alignment of hard targets, which have less than 30% sequence identity with experimentally determined structures in PDB. RaptorX outperformed all the CASP9 participating servers including those using consensus and refinement methods, blindly tested on the 50 hardest CASP9 template-based modeling targets [154,155]. As an input, RaptorX takes an amino acid sequence and predicts its secondary and tertiary structures as well as solvent accessibility and disordered regions. It also allocates confidence scores like P-value for the relative global quality, GDT (global distance test) and uGDT (un-normalized GDT) for the absolute global quality, and RMSD for the absolute local quality of each residue in the model to indicate the quality of a predicted 3D model [156].

3.2.3 Ab initio structure prediction

For a given protein amino acid sequence, the aim of Ab initio structure prediction is to predict the 3D structure of its native state. A protein sequence folds to a native conformation or a collection of conformations that is at or close to the global free-energy minimum. Therefore, the difficulty in finding native-like conformations for a given sequence has to address the development of a precise potential and an efficient method for searching the resultant energy landscape [157]. Traditionally, the most acclaimed structure prediction methods have been homology-based comparative modeling and fold recognition [158]. These two approaches construct protein models by aligning query sequences against solved template structures. If close templates are recognized, high-resolution models could be generated by the template-based modeling. However, if templates are lacking from the Protein Data Bank (PDB), the models require to be built from scratch, i.e. Ab initio folding, the most complicated category of protein-structure prediction [159,160]. As the size of the protein increases, the conformational phase space of sampling also increases sharply making the Ab initio modeling of bigger proteins exceedingly tricky [161]. Current Ab initio predictions are mainly focused on small proteins. Existing Ab initio predictions are primarily focused on small proteins and quite a few successful examples have been described in literature [139]. In this chapter, we highlight some of the well known current methods for predicting tertiary structure which employ Ab initio methods in the absence of homology to a known structure.

3.2.3.1 I-TASSER

I-TASSER is a web-based hierarchical protein structure modelling method which employs the Profile-Profile threading Alignment (PPA) improved by secondary-structure [162] and the iterative implementation of the Threading ASSEmbly Refinement (TASSER) program [163]. In I-TASSER, the query sequences are initially threaded through a representative PDB structure library (using 70% sequence identity as cut-off) to explore for the potential folds by four straightforward variants of PPA methods, with diverse combinations of the hidden Markov model [164] and PSI-BLAST [41] profiles and the Needleman-Wunsch [140] and Smith-Waterman [39] alignment algorithms. The continuous fragments are then eliminated from the threading aligned regions which are used to reconstruct full-length models whereas the threading unaligned regions (e.g. loops) are built by Ab initio modeling [165]. The replica-exchange Monte Carlo simulation [166] is used to search conformational space. SPICKER [167,168] is used to cluster the structure trajectories and the cluster centroids are acquired by the averaging the coordinates of the entire clustered structures.

Fragment assembly simulation is implemented again, which begins from the cluster centroid of the initial round simulation, so that the steric clashes on the centroid structures are excluded and promote additional model refinement. Spatial restraints are removed from the centroids and the PDB structures are explored by the structure alignment program TM-align [169], which are used to direct the second round of simulation. In the end, the structure decoys are clustered and the structures with the lowest possible energy in each cluster are selected, which has the Ca atoms and the side-chain centres' of mass specified. Backbone atoms (N, C, O) are added by Pulchra [170] and Scwrl_3.0 [171] is used to build side-chain rotamers. The absolute quality of the final model in comparison with the native structure is given by TM-score.

3.2.3.2 QUARK

QUARK is a web-based Ab initio protein structure prediction method which focuses on the elaborate design of the force field and the search engine by taking a semi-reduced model to represent protein residues by the full backbone atoms and the side-chain center of mass [172]. Initially, QUARK predicts a wide range of carefully selected structural features by using neural network (NN) for a query sequence. Based on the idea borrowed from Rosetta and I-TASSER, the global fold is then created by replica-exchange Monte Carlo (REMC) simulations by assembling the small fragments as generated...
by gapless threading through template library. However, the fragments in QUARK have constantly multiple sizes which range from 1 to 20 residues where as in Rosetta and I-TASSER the fragments are in either 3/9-mer or from threading alignments. There are three main steps in which the QUARK Ab initio structure prediction method can be divided; A). Multiple feature predictions and fragment generation starting from one query sequence. B) Structural constructions using REMC simulation based on the semi-reduced protein model. C) Decoy structure clustering and full atomic refinement.

For each query amino acid sequence, multiple sequence alignment (MSA) is generated by PSI-BLAST [41] using a non-redundant database where as secondary structure as predicted by Protein Secondary Structure prediction program PSSpred [173] based on multiple neural network (NN) trainings of sequence profiles computed from the MSA. Features like Solvent accessibility (SA), real-value phi and psi angles, b-turn positions are predicted by using different NNs based on the checkpoint file by PSI-BLAST and SS types predicted by PSSpred. These features are then used to generate structural fragments for each segment of the query sequence. Moreover, the total energy of the QUARK force field is the sum of the 11 knowledge based energy terms that can be categorized into three levels of structural packing: atom-, residue-, and topology-level energy terms. The structural quality of the final models is measured by TM-score in comparison to the native structures. The QUARK method was also tested in the CASP9 experiment where it showed novel advancement in the field of Ab initio protein structure predictions.

3.2.3.3 PEP-FOLD

It is on-line resource for 3D structure prediction of peptides of length up to 36 amino acids with well-defined structures in aqueous solution and also accepts cyclic peptides using disulfide bonds defined by the user [174]. The PEP-FOLD predicts the structure in three steps. (i) Structural Alphabet (SA) letters from the amino acid sequence are predicted. The sequence is also used to generate a psi-blast profile which is further used as input for SVM that returns a likelihood profile of each SA letter at each position of the sequence. This SA profile is then analyzed to choose a few letters at each position. (ii) In the second step the 3D assembly of the prototype fragments linked with the letters selected is performed, which relies on the sOPEP coarse grained force field [175]. An improved greedy procedure [176] that builds the peptide residue by residue is utilized for 3D generation which is followed by a Monte-Carlo method for ultimate refinement. (iii) The final step creates all-atom conformations from the coarse grained models returned by the 100 simulations and executes a clustering procedure. At the end of the program run, the server gives the outcome of the cluster analysis. Additionally, if a reference structure was provided, the cRMSD, the GDT-TS and TM scores of each model are also reported. PEP-FOLD was tested on a benchmark of 34 cyclic peptides with one, two and three disulphide bonds and the best PEP-FOLD models deviated by an average RMS of 2.75Å from the full NMR structures. Similarly, a test on benchmark of 37 linear peptides showed that PEP-FOLD finds lowest-energy conformations deviating by 3Å RMS from the NMR rigid cores.

4. Critical Assessment of Protein Structure Prediction (CASP)

CASP is a community-wide, worldwide experiment for protein structure prediction which takes place after every two years ever since it started in 1994 [177]. CASP offers research groups with an opportunity to test their structure prediction algorithms and conveys an independent evaluation of the state of the art in protein structure modeling to the research community and software users. Currently, large number of servers and tools are extensively available for generating a structural model for the protein of interest which can then be used as a structural scaffold for designing additional experiments, interpreting functional data, assigning molecular function to the protein, or as a target for drug design or even as a means for solving the experimental structure of the protein. Nevertheless, the possible applications of a model depends on its quality, and therefore, the admittedly difficult problem of assessing in advance the quality of models created by various methods is of great significance [178].

CASP targets or the sequences of the protein for which experimental biologists are about to resolve a protein structure by X-ray crystallography or NMR spectroscopy are made available. Predictors generate and deposit models for these the CASP targets ahead of the structures are made publically available. If the given target sequence is found to exhibit a certain degree of similarity with known structure called template, then comparative protein modelling or homology based modeling may be used to predict the 3D structure or else, de novo protein structure prediction method must be applied, which is much less accurate but can occasionally yield models with the correct fold. A panel of three assessors compares the models with the structures the moment they are available and try to assess the quality of the models and to draw some conclusions about the state of the art of the diverse methods. To make certain that no predictor can have aforementioned information regarding a protein’s structure, experiment is carried out in a double-blind fashion where neither the predictors nor the organizers and assessors know the structures of the target proteins at the time when predictions are made.

The main way of evaluation [179] is a comparison of the predicted model a-carbon positions with those in the target structure. The comparison is shown by cumulative plots of distances between pairs of corresponding a-carbon in the alignment of the model and the structure, and is assigned a numerical score GDT-TS (Global Distance Test — Total Score) [180] describing percentage of accurately-modeled residues in the model with regard to the target. Template-free, or de novo modeling (Free modeling) is also assessed visually by the assessors, given that the statistical scores do not work as well for finding loose similarities in the most complex cases [181].

In CASP10 [182] evaluation of the results was carried out in the following two main prediction categories:

4.1 Tertiary structure predictions (TS)

- Template Based Modeling (TBM) - includes domains where an appropriate template can be recognized that covers all
or almost the entire target.

- **Template free modeling (FM)** - includes models of proteins for which no reliable template can be identified.

- **Refinement** - one of the best models received during the prediction season was reissued as a starting structure for refinement.

- **Contact-assisted structure modeling** - this category demonstrates how the information of a few (usually 3 to 5) long-range contacts influences the capability of predictors to model the complete structure.

- **Chemical shifts guided modeling** of NMR structures was performed if the chemical shifts table from the NMR-spectroscopists was available for the selected CASP10 targets.

- **Structure modeling based on molecular replacement with Ab initio models and crystallographic diffraction data** was carried out for selected targets provided the structure factors from the crystallographers was available.

### 4.2 Other prediction categories

- Residue-residue contact prediction in proteins (RR).

- Disordered regions in target proteins (DR).

- Function prediction (prediction of binding sites) (FN).

- Quality assessment of models and the reliability of predicting certain residues in particular (QA).

The results of CASP are available in special supplement issues of the scientific journal *Proteins: Structure, Function and Bioinformatics*, all of which are accessible through the CASP website. CASP proceedings comprises of papers describing the structure and ways in which the experiments are conducted, the statistical evaluation procedures, reports from the evaluation teams highlighting state of the art in diverse prediction categories, methods from some of the most successful prediction teams, and advancement in different aspects of the modeling [182,183].

### 5. Molecular Visualization Tools

#### 5.1 2D-GraLab (two-dimensional graphics lab for biosystem interactions)

2D-GraLab is a program that takes PDB file as an input and automatically generates schematic representations of nonbonding interactions across the protein binding interfaces. The program outputs two-dimensional PostScript diagrams providing intuitive and informative report of the protein–protein interactions (PPIs) and several energetics properties, such as hydrogen bond, salt bridge, van der Waals interaction, hydrophobic contact, π–π stacking, disulfide bond, desolvation effect, and loss of conformational entropy. The three main points on which 2D-GraLab emphasizes are a) reliability-which is ensured by the use of widely acclaimed programs embedded in 2D-GraLab; b) comprehensiveness- ability to handle nearly all the nonbonding interactions across binding interface of protein complexes, such as hydrogen bond, salt bridge, van der Waals (vdW) interaction, hydrophobic contact, π–π stacking, disulfide bond, desolvation effect, and loss of conformational entropy; c) artistry- with the aim of creating aesthetically pleasing 2D images of PPIs, the layout, color match, and page style for different diagrams are richly designed [183]. Additionally, 2D-GraLab also provides a graphical user interface (GUI), which allows users to interact with the program and displays the spatial structure and interfacial characteristics of protein complexes. 2D-GraLab is written in C++ and OpenGL, and the resultant output of 2D schematic diagrams of nonbinding interactions are depicted in PostScript. The current version of 2D-GraLab is freely available to academic users on request.

#### 5.2 Chimera

UCSF Chimera is an extremely extensible program for interactive visualization and analysis of molecular structures, including density maps, supra-molecular assemblies, sequence alignments, docking results, trajectories, and conformational ensembles [184]. High-quality images and animations can be created by using Chimera which provides complete documentation and several tutorials, and is freely available for academic, government, non-profit, and personal use. Chimera is divided into two sections, namely, a core and extensions. The core offers fundamental services and molecular graphics capabilities, where as all the advanced level functionality is provided through extensions. The extension mechanism ensures that this design is strong enough to handle the requirements of outside researchers who wish to broaden the scope of Chimera in novel ways. These extensions can be incorporated into the Chimera menu system, and can present a separate graphical user interface as desired by means of the Tkinter (http://www.python.org/topics/tkinter/), Tix (http://tixlibrary.sourceforge.net/) and/or Pmw (http://pmw.sourceforge.net/) toolkits. The Chimera core consists of a C++ layer that handles time-critical operations such as graphics rendering, and a Python layer that handles all additional functions. The entire major C++ data and functions are made available to the Python layer. The abilities of core consist of molecular file input/output, molecular surface generation via the MSMS algorithm [185] and characteristics of graphical display for example wire-frame, ball-and-stick, ribbon, and sphere representations, transparency control, near and far clipping planes, and lenses.

Some of the major extensions of Chimera include:

**Multiscale:** This extension appends capabilities for interactively investigating bulky molecular assemblies with special focus on viral structures, condensed chromosomes, and ribosomes; further examples consist of cytoskeletal fibers and motors, flagellar structures, and chaperonins. Multiscale makes use of Chimera’s core molecular display abilities, data
structures, file reading, and selection management, and the Volume Viewer extension for surface computation and rendering.

**Multalign Viewer:** This extension permits Chimera to show sequence alignments together with associated structures. Multalign Viewer can read and write sequence an alignment in a broad range of popular formats (Clustal, “aligned” FASTA, GCG MSF, GCG RSF, “aligned” NBRF/PIR, and Stockholm).

**ViewDock:** This extension assists in interactive screening of ligand orientations from DOCK [186], which calculates potential binding orientations given the structures of ligand and receptor molecules. ViewDock reads the DOCK output and gives a well-situated interface for screening results in the context of the target structure.

**Collaboratory:** Chimera’s Collaboratory extension makes possible for researchers at geographically remote sites to share a molecular modeling session in real time. By default, all users associated to the identical session have equivalent control over the models (structures) being viewed. An alteration made by any member is straight away reflected to all other participants, so that a coordinated view of the data is sustained during the entire session.

Chimera is developed and maintained by the Resource for Biocomputing, Visualization, and Informatics and funded by the NIH National Center for Research Resources, and one of their major challenges is to enhance Chimera’s performance in the context of the large-scale systems for which the Multiscale extension was created.

5.3 VMD: Visual molecular dynamics

VMD is a molecular graphics program intended for the display and analysis of molecular assemblies like proteins and nucleic acids. One of the most important features of VMD is that it can simultaneously display several numbers of structures by means of a broad array of rendering styles and coloring methods. Molecules are shown as one or more “representations,” in which every representation symbolizes a particular rendering technique and coloring scheme for a selected subset of atoms [187]. The atoms displayed in each representation are selected by using a wide atom selection syntax, which includes Boolean operators and regular expressions. Besides providing a complete graphical user interface for program control, VMD also has a text interface which uses the Tcl embeddable parser to permit complex scripting tasks with variable substitution, control loops, and function calls. VMD has also been explicitly designed with the facility to animate molecular dynamics (MD) simulation trajectories, which could either be imported from files or from a direct connection to a running MD simulation. VMD is the visualization module of MDScope [188], a set of tools for interactive problem solving in structural biology, which also comprises the parallel MD program NAMD [189], and the MDCOMM software used to unite the visualization and simulation programs. VMD is written in C++, using an object-oriented design; the program, including source code and all the documentation, is freely obtainable via anonymous ftp and through the World Wide Web.

5.4 Pymol

Pymol, a molecular visualization system developed by Warren Lyford DeLano and commercialized by DeLano Scientific LLC (a private software company committed to develop useful tools for scientific and educational communities), is a widely popular open source molecular visualization tool. Pymol can create superior quality 3D images of small molecules and biological macromolecules, such as proteins and nucleic acids. The Py segment of the software’s name refers to the fact that it extends, and is extensible by the Python programming language. More details about Pymol can be found at http://www.pymol.org/.

---

**Appendix**

<table>
<thead>
<tr>
<th>Section</th>
<th>Database/Software</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.1.1</td>
<td>PDB</td>
<td><a href="http://www.rcsb.org/pdb/home/home.do">http://www.rcsb.org/pdb/home/home.do</a></td>
</tr>
<tr>
<td>2.2.1.2</td>
<td>SCOP</td>
<td><a href="http://scop.mrc-lmb.cam.ac.uk/scop/">http://scop.mrc-lmb.cam.ac.uk/scop/</a></td>
</tr>
<tr>
<td>2.2.1.3</td>
<td>CATH</td>
<td><a href="http://www.cathdb.info/">http://www.cathdb.info/</a></td>
</tr>
<tr>
<td>2.2.1.4</td>
<td>PDBeum</td>
<td><a href="http://www.ebi.ac.uk/pdbeum/">http://www.ebi.ac.uk/pdbeum/</a></td>
</tr>
<tr>
<td>2.2.1.5</td>
<td>PDBTM</td>
<td><a href="http://pdbtm.enzim.hu/">http://pdbtm.enzim.hu/</a></td>
</tr>
<tr>
<td>2.2.1.6</td>
<td>Database of Macromolecular Movements</td>
<td><a href="http://www.molmovdb.org/">http://www.molmovdb.org/</a></td>
</tr>
<tr>
<td>2.2.1.7</td>
<td>Orientations of Proteins in Membranes (OPM)</td>
<td><a href="http://opm.phar.uminich.edu/">http://opm.phar.uminich.edu/</a></td>
</tr>
<tr>
<td>2.2.1.8</td>
<td>PROCARB</td>
<td><a href="http://www.procarb.org/">http://www.procarb.org/</a></td>
</tr>
<tr>
<td>2.2.2.1</td>
<td>Protein Model Database (PMDB)</td>
<td><a href="http://mi.caspur.it/PMDB/">http://mi.caspur.it/PMDB/</a></td>
</tr>
<tr>
<td>2.2.2.2</td>
<td>Molecular Modeling Database (MMDB)</td>
<td><a href="http://www.ncbi.nlm.nih.gov/Structure/MMDB/mmdb.shtml">http://www.ncbi.nlm.nih.gov/Structure/MMDB/mmdb.shtml</a></td>
</tr>
<tr>
<td>2.2.2.3</td>
<td>ModBase</td>
<td><a href="http://modbase.compbio.ucsf.edu/modbase/cgi/index.cgi">http://modbase.compbio.ucsf.edu/modbase/cgi/index.cgi</a></td>
</tr>
<tr>
<td>2.2.2.4</td>
<td>SWISS-MODEL Repository</td>
<td><a href="http://swissmodel.expasy.org/repository/">http://swissmodel.expasy.org/repository/</a></td>
</tr>
<tr>
<td>3.2.1.1</td>
<td>MODELLER</td>
<td><a href="http://salilab.org/modeller/">http://salilab.org/modeller/</a></td>
</tr>
<tr>
<td>3.2.1.2</td>
<td>SWISS-MODEL</td>
<td><a href="http://swissmodel.expasy.org/">http://swissmodel.expasy.org/</a></td>
</tr>
<tr>
<td>3.2.1.3</td>
<td>3D-JIGSAW</td>
<td><a href="http://bmm.cancerresearchuk.org/~3djigsaw/">http://bmm.cancerresearchuk.org/~3djigsaw/</a></td>
</tr>
<tr>
<td>3.2.1.4</td>
<td>Robetta</td>
<td><a href="http://robbetta.bakerlab.org/">http://robbetta.bakerlab.org/</a></td>
</tr>
<tr>
<td>3.2.1.5</td>
<td>E5yPred3D</td>
<td><a href="http://www.unamur.be/sciences/biologie/urbm/bioinfo/esypred">http://www.unamur.be/sciences/biologie/urbm/bioinfo/esypred</a></td>
</tr>
<tr>
<td>3.2.1.6</td>
<td>CPHModel</td>
<td><a href="http://www.cbs.dtu.dk/services/CPHmodels/">http://www.cbs.dtu.dk/services/CPHmodels/</a></td>
</tr>
<tr>
<td>3.2.1.7</td>
<td>BHAGEERATH-H</td>
<td><a href="http://www.scbio-itd.res.in/bhageerath/bhageerath_h.jsp">http://www.scbio-itd.res.in/bhageerath/bhageerath_h.jsp</a></td>
</tr>
<tr>
<td>3.2.1.8</td>
<td>EasyModeller</td>
<td><a href="http://www.uohyd.emet.in/modellergui/">http://www.uohyd.emet.in/modellergui/</a></td>
</tr>
<tr>
<td>3.2.1.9</td>
<td>GeneSilico</td>
<td><a href="https://genesilico.pl/meta2/">https://genesilico.pl/meta2/</a></td>
</tr>
</tbody>
</table>
References

12. RCSB: [http://www.rcsb.org/]


156. http://raptorx.uchicago.edu
173. http://zhanglab.ccb.med.umich.edu/PSSpred