Chapter 24
Protein Homology Modeling in Phytochemical Research

Aditya Narayan, Ajeet Singh, and Shailesh Kumar

24.1 Introduction

24.1.1 Phytochemicals

In the broadest sense, we may define phytochemicals as chemical compounds produced by plants with very sharp, and distinctive properties such as taste, odor and so on. More significantly, they may also play key roles in metabolism, defense mechanisms and other aspects of plant physiology. Although they are non-essential nutrients, intake of some phytochemicals has been shown to provide protective and disease preventive properties. There are burgeons of phytochemicals belonging to different classes with multipartite functions. Some of the functions include antioxidant action (well-known examples include carotenoids and flavonoids), influencing hormone release (isoflavones in soy imitate estrogen), enzyme stimulation (indoles in cabbage), impairment of cell replication (saponins), DNA protective mechanisms (capsaicin) and protection from pathogens by physically inhibiting access to cell walls (Hamuel 2012).

Absolutely critical to the study of phytochemistry is the basic understanding of the 3D structures of the vast array of existing chemicals. Experimental validation of these structures is time-consuming and inefficient, thereby limiting the knowledge of the vast majority of protein structures. Till now, for exploring the biochemical role of a number of phytochemicals at any level, it becomes necessary to employ in-silico methods of structure modeling.

A. Narayan
University of Virginia, Charlottesville, VA, USA

A. Singh · S. Kumar (✉)
National Institute of Plant Genome Research (NIPGR), New Delhi, India
e-mail: shailesh@nipgr.ac.in

S. Kumar, C. Egbuna (eds.), *Phytochemistry: An in-silico and in-vitro Update*,
https://doi.org/10.1007/978-981-13-6920-9_24

The utility of such tools may be assessed by how the homology modeling and molecular docking analysis are used on phytochemicals for better understanding and designing solutions for disease-causing mechanisms along with their prevention. Some supportive examples include screening phytochemicals against the ligand binding sites of the E1 protein of chikungunya virus responsible for host-binding, docking phytoconstituents in psoriasis-causing protein corneodesmosin for better understanding of its active site and testing flavonoid inhibitors against the hyperactive 26S proteasome subunit in cancer cells (Vasavi et al. 2010; Panda et al. 2014; Salomi et al. 2016). Homology modeling may even be applied to whole plant proteomes in order to understand the set of phytoconstituents contributing to antioxidant activity in spinach (Sahay and Shakya 2010).

24.1.1.1 Molecular Modeling

Molecular modeling comprises of the wide range of theoretical, computational, and biochemical techniques employed to generate and study the structure of different molecules. As stated previously, the most commonly applied method for molecular modeling is known as homology modeling (also referred to as comparative modeling or template-based modeling), which employs template structures which are evolutionarily related to the structure of interest to serve as a model from which the structure can be predicted. Although not as empirically accurate as certain techniques such as x-ray crystallography, homology modeling may yield large volumes of structural information that are critical in the study of molecular structure, dynamics, and protein-ligand interactions. Thus, the development and use of these tools bear wide-reaching industrial and medical implications. In particular, given the experimental difficulties related to structural determination, faster methods of protein structure prediction yielding accurate models have significantly emerged nowadays.

24.1.2 Homology Modeling Steps

The process of homology modeling can be further categorized into a number of steps which are largely identical across the variety of in-silico homology modeling tools (Fig. 24.1).

24.1.2.1 Template Recognition and Initial Alignment

Homology modeling hinges on the fact that 3D structures among evolutionarily related protein structures are highly analogous. A variety of databases, particularly

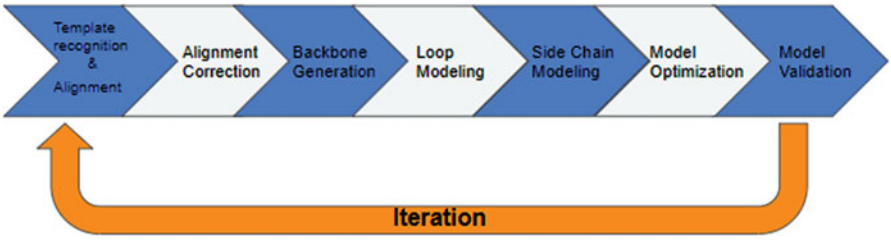


Fig. 24.1 Workflow/methodology of protein homology modeling tools

the Protein Data Bank (http://www.wwpdb.org/), and database scanning tools exist which allow users to enter query sequences/structural information that will be used to enquire appropriate template homologs. The best-known tool for this purpose is BLAST. Once a template is found, it is necessary to create a multiple sequence alignment of the template, the structure of interest and other related proteins to gain an insight into related domains, motifs, features within the protein family and other relevant structural information.

The discussion regarding the "twilight zone" of sequence similarity for proteins is demonstrated graphically by Sivakumar (http://biosiva.50webs.org/alignment.htm). In comparing sequences, the similarity with respect to the residue/size/charge/hydrophobicity is measured as sequence identity. When sequence identity is high, generally defined as greater than ~25%, it is possible to argue the protein's evolutionary relatedness (homology) with some degree of confidence. Below this limit, the sequence identity is said to fall within the "twilight zone", in which homologous sequences mix with randomly aligned sequences. Below 20%, homologous relationships may not be reliably determined – the "midnight zone." Sequence length must also be accounted for as shorter sequences will have a greater chance that alignments are a result of random chance. Irrespective of the sequence, however, there are many situations in which proteins falling in the "twilight zone" of sequence identity maintain similar folds. However, such situations are rare and thus the goal of modeling tools is to identify templates with sufficient sequence homology that they may be used to generate accurate structures.

24.1.2.2 Alignment Correction

Residue properties must be taken into consideration in aligning multiple sequences such that analogous core residues are sufficiently hydrophobic while the outermost residues may be more variable. It may become necessary to introduce small insertions and deletions in the non-conserved regions of the structure by hand to correct for sequence disparities. If sequence fragments are not present in the target but are

found in the template, they may be deleted from the template. If there is an insertion in the target, then the template will contain a gap. This ultimately improves the overall quality of the target-template alignment.

24.1.2.3 Backbone Generation

The backbone of the target structure is generated from analogous coordinates in the template given the conservation of 3D structures. This applies also to the conserved side chains.

24.1.2.4 Loop Modeling

Due to deletions and insertions in the alignment as a result of alignment correction, the template may contain gaps in the sequence. This prevents modeling of the target and such sequences must be constructed as flexible surface loops. These may be generated *ab initio* through identification of similar sequences in the PDB which bear similar residues/environments. Various tools may be used to determine the most accurate loop model which is then added to the target structure.

24.1.2.5 Side-Chain Modeling

Side chains of conserved residues are easily modeled in the target structure. However, non-conserved residues must be added by accounting for a number of constraints. Torsional angle is often conserved across homologs for the majority of residues and certain rotamers are optimal for certain backbone structures/steric constraints etc. Libraries of side chain structures under such constraints are analyzed to choose the optimal rotamers.

24.1.2.6 Model Optimization

In order to account for backbone and side chain changes over the course of modeling, each must be adjusted in accordance with the other as constraints are altered. Molecular dynamics simulations are performed within a force-field modeling tool such that protein folding may be modeled. Ultimately the side chain rotamers and backbone will be adjusted until the potential energy of the structure is at a minimum.

24.1.2.7 Model Validation

A number of model validation tools may be employed to determine if the final structure produced from the previous steps abides by acceptable ranges with respect to bond angles, torsion, bond lengths, distribution of side chains/side chain properties and how the model folds/active site compare to homologs.

24.2 Homology Modeling Tools

In order to determine the most effective homology modeling tools, it is common to refer to international blind trials of protein structure prediction methods known as the Critical Assessment of protein Structure Prediction or CASP. For majority of the commonly used protein modeling tools, CASP trials indicate relatively little difference in accuracy – particularly in situations when homology is not readily inferred with known structures (Kelley et al. 2015). However, here we will discuss differences in methodology, algorithms applied, input, output, and speed among commonly applied homology modeling tools. The overview of advantages and disadvantages of each tool discussed in the chapter is highlighted in Table 24.1.

24.2.1 Modeller

24.2.1.1 Outline/Methodology

MODELLER is the most commonly used tool used in generating 3D structures via homology modeling. This tool is applied to create a model through analysis of spatial restraints as well as statistical assessments of homologous structures in the PDB and it was conceived in order to minimize alignment errors in the comparative modeling

Table 24.1 Overview of each tool's advantages and disadvantages

Tools name	Pros	Cons
MODELLER	Fast, high quality, free, reliable	Command-line only interface, poor core and side-chain modelling
I-Tasser	Fast, high quality, free	Unreliable model selection, poor domain splitting
Swiss-Model	Fast, high quality, good core modeling, good stereochemistry	Unreliable crashes frequently
Phyre2	Fast, reliable, ease of use	Low accuracy
HHPred	Very fast, ease of use	Low accuracy
Robetta	High quality, ease of use	Computationally demanding

process to produce the most accurate possible target structure. It does so through the application of a genetic algorithm to optimize alignment, after which a comparative modeling procedure is used to create the 3D structure. Application of a genetic algorithm allows for individuals within the population to undergo recombination and random mutation (alignment changes) under the guidance of a fitness function to converge on an optimal model as evaluated by an assessment score. In summary, the MODELLER outputs a list of comparative models after undergoing an iterative process of alignment using a genetic algorithm, model building, and assessment until the top model is output (as calculated by composite modeling score) (John and Sali 2003).

24.2.1.2 Pros/Cons

MODELLER has the advantage of being able to create high-quality figures relatively quickly and for free though it may be improved with respect to core and side chain modeling (Wallner and Elofsson 2005). However, its availability as command-line only interface may be difficult to use for the inexperienced users.

24.2.1.3 Input

(a) Multiple sequence alignment of the target (PIR format) with homologs.
(b) One or more template structures.
(c) Template profile of multiple structural alignments of the templates with close homologs.

24.2.1.4 Output

Target–template alignments alongside comparative models for the sequence ranked in order of the model score.

24.2.1.5 Working Steps (Basic Modeling)

I. Searching for structures related to the target.

   (a) Put the target sequence into PIR format (. ali) in which the first line contains ">P1; code". The second line contains information of the structure file with the term "sequence" and the model file name separated by a colon. The sequence follows with "*" at the end.

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

Fig. 24.2 MODELLER output log displaying a summary of models built

(b) The profile. Build (sequence.ali) MODELLER command is used to identify homologs.

II. Selecting template.

(a) The output of the profile. build command, "build_profile.log" contains PDB sequences for identified templates, the corresponding sequence identity, and e-value. The greatest similarity sequences (identifiable by sequence identity and e-value) are used as an argument in the alignment. compare_structures command. Within this command, structural and sequence similarity with templates are iteratively compared with the input. The results are evaluated and presented in a tabular format (compare.log file).

(b) The output table presents relevant pairwise sequence distances as well as a template clustering tree to illustrate the distinction between templates. From this result, the final template may be selected. Note that it is possible to consider crystallographic R-factor in making the selection.

III. Template alignment.

(a) The align2D command accepts the selected template and is used to generate the sequence alignment in PIR and PAP formats (model-selected_template. ali and model-selected_template.pap).

IV. Model building.

(a) The auto model class is automatically applied by MODELLER. This process loads the auto model class, generates an auto model object, names the file containing the target-template alignment with alnfile, defines the template structure with the known function, obtains alignment scores with assess methods, defines the target sequence name with sequence, defines the number of calculated models with starting model and ending model, and finally applies the make command.

(b) The most significant output file is "model-single.log" which contains a summary of models built with associated confidence scores (Fig. 24.2).

V. Model Evaluation.

(a) It is possible to evaluate the validity of the model with external programs or within MODELLER by using the MODELLER objective function, DOPE/SOAP assessment scores or GA341 assessment score.

Notes

1. PSI-BLAST may also be used to identify target and template profiles.
2. Additional features of MODELLER include de-novo modeling of loops, optimization of the structure through user-defined functions, multiple sequence/structure alignment, and clustering.

24.2.2 I-TASSER

24.2.2.1 Outline/Methodology

I-Tasser is a Unix based tool utilised for fully automated protein structure prediction and functional annotation. It does so by assembling the target model from fragments of threading templates. The process begins by threading the target sequence through a PDB library with LOMETS while generating alignments through a variety of tools (HMM models, Psi-BLAST, Needleman-Wunsch, Smith-Waterman). Consistent/aligned fragments are applied to the model while non-aligned regions are constructed through *ab initio* modeling by replica-exchange Monte Carlo simulations. Structural decoys are clustered iteratively through SPICKER and the lowest energy structures of the clusters matching the desired properties are selected. REMO and FG-MD are used for the final model and refinement with the top structures rated by I-TASSER confidence score (C-Score). Structural annotation is performed with COFACTOR and COACH (Zhang 2008).

24.2.2.2 Pros/Cons

I-TASSER, similar to MODELLER, provides a relatively fast structural prediction for free and based on CASP trials has outperformed other tools with respect to accuracy. Minor issues include poor domain splitting and a propensity for unreliable model selection that greatly skews results.

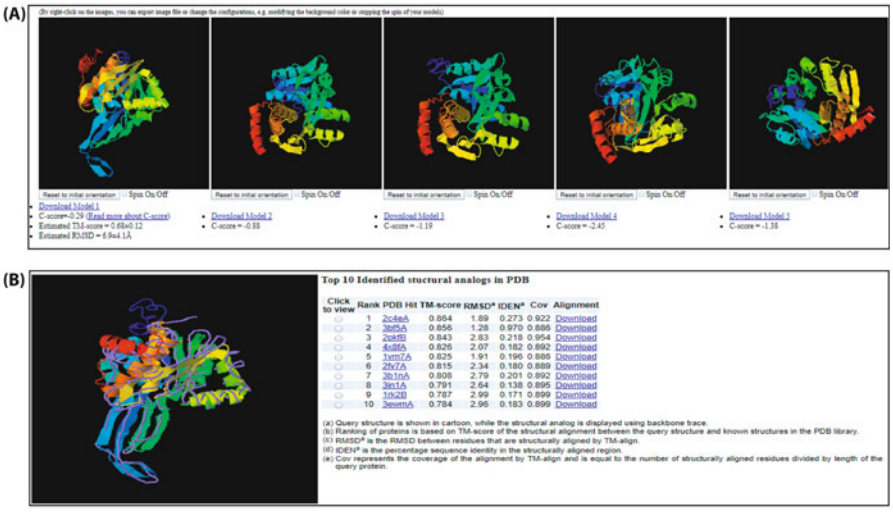


Fig. 24.3 Demonstration of I-Tasser overall output for a given query sequence. (**A**) Top 5 modeling results displayed in a browser window from decoy clustering. (**B**) Top 10 structurally similar homologs to the given target

24.2.2.3 Input

The primary amino acid sequence of protein query in the FASTA format.

24.2.2.4 Output

The output page includes information on the sequence, secondary structure/structural feature prediction, predicted normalized B factor (thermal mobility), top structural models with associated accuracy estimation from TM-Align (Fig. 24.3a), top threading alignments (Fig. 24.3b), solvent accessibility, functional annotation, EC number/GO term and editable models exportable as gif, jpg, or png.

24.2.2.5 Working Steps

I. Submit protein amino acid sequence using the form on I-TASSER site or upload directly and press the Run button.
II. Results will be provided via input email.

Notes
The I-Tasser suite contains a variety of tools including PSSpred, LOMETS, SPICKER (decoy clustering), ModRefiner, ResQ, COFACTOR, COACH, and NW-align.

24.2.3 Swiss-Model

24.2.3.1 Outline/Methodology

SWISS-MODEL, a web server was created initially with the goal of creating a flexible, easy-to-use protein modeling tool for structural analysis. This flexibility is evident in the multiple interfaces available including a "first approach mode" which minimizes complications for the user by only allowing an amino acid sequence input. It is also possible to use an "alignment mode" which accepts a sequence alignment as well as a "project mode" allowing manual control of a wide variety of parameters. SWISS-MODEL functions by first selecting a template through template structure library analysis (specifically the Swiss-Model Template Library derived from PDB), after which a structural alignment is constructed by iteratively superimposing structures and removing incompatible templates. If no candidate is found, a conformational space search is applied. The backbone structure is constructed as an average of templates with loops created with *ab initio* modeling. Finally, side chains are chosen from a rotamer library and the structures are energy minimized via the CHARMM27 force field and the optimal structures output (Biasini et al. 2014).

24.2.3.2 Pros/Cons

SWISS-MODEL has shown to provide highly accurate structures, particularly in the protein core and with respect to stereochemistry, at high speed. However, the program has been shown to crash, particularly when only low sequence identity is available (Wallner and Elofsson 2005).

24.2.3.3 Input

There are a number of potential inputs for SWISS-MODEL within the web interface. The simplest one is an amino acid sequence in FASTA format, with one letter sequence or through UniProt accession code. Additional inputs include a target-template MSA or a Deep View project file.

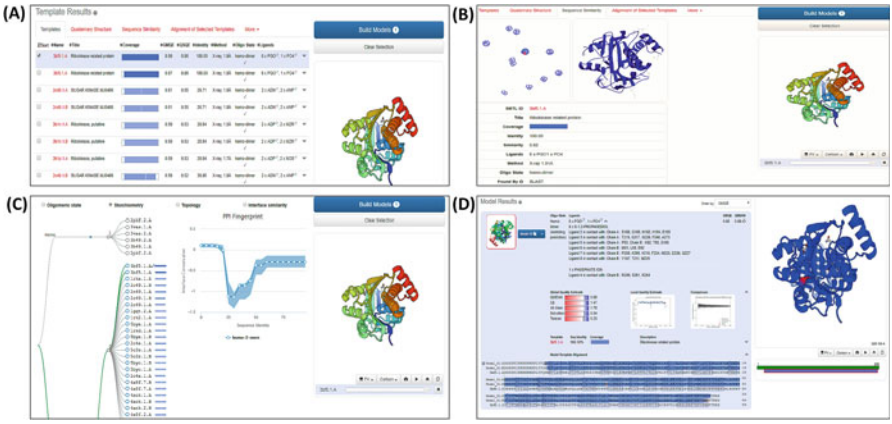


Fig. 24.4 Representation of the working modules of the tool SwissModel. (**A**) List of top alignments found by SwissModel and model constructed from the respective alignments on the right. (**B**) Sequence similarity of alignments visualized. The target protein is shown in red and each template as a blue circle with the N terminus of the target protein shown at the top of the circle and the C terminal at the clockwise end. The distance between circles represents similarity. Clicking on a circle takes the user to template-specific information. (**C**) Phylogenetic tree displaying the relationship between template homologs. (**D**) Modeling results for a specific template with relevant information including ligand binding sites, model-template alignments, and quality scores

24.2.3.4 Template Search Output

The output of the template search includes template results (Fig. 24.4a), relationships between detected templates and the query. It is possible to manually select the template within the browser interface through inspection of visual representations of the alignments (Fig. 24.4b). Alternatively, the templates may be selected automatically. Results are presented both graphically (interactive 3D and 2D formats), as a relationship tree (Fig. 24.4c), and also in tabular format alongside relevant information about each template.

From each of the selected templates, the program provides model coordinates and structure, predicted accuracy and similar relevant information (Fig. 24.4d). This information can be downloaded from the website for future use as zip files containing models in PNG, PDB, and JSON file formats as well as information on templates.

24.2.3.5 Working Steps

I. Submit protein the amino acid sequence through the form on the Swiss-Model site into the "Target Sequence" box.
II. Press "Build Model" and results will be presented within the web interface.

24.2.4 Phyre2 (ProteinHomology/Analogy Recognition Engine)

24.2.4.1 Outline/Methodology

Phyre2, a web tool, which distinguished by its ease of use and speed while providing advanced tools such as batch submission, Backphyre (detects homology across genomes) and Phyre Investigator, which allows for detailed model analysis. Phyre2 first scans libraries to construct an evolutionary profile, or statistical distribution of residues across homologs, with HHBlits that is then used in conjunction with a secondary structure prediction (Psi-Pred) to convert the query to an HMM. This HMM is scanned and aligned against HMMs in a fold library using HHsearch, to generate the protein fold while loops are modeled from a fragment (2–15 amino acid length) library or *ab initio* modeling to correct for insertions/deletions. Finally, side chains are modeled from a rotamer library with the R3 protocol (Kelley et al. 2015).

24.2.4.2 Pros/Cons

Phyre2 provides relatively high ease of use for those lacking expertise in protein modeling while sacrificing little in the terms of accuracy. The limitations of Phyre2 are not unique to the program and include difficulty in producing accurate structures when no homologs of known structure are detected, difficulty predicting the structural effect of mutations and inability to model multimeric proteins.

24.2.4.3 Input

The initial input for the standard protocol used to model a single sequence is the primary amino acid sequence in one letter code (FASTA format).

24.2.4.4 Output

Once complete, the user will be sent via an email, several pieces of information about the models produced including confidence in the models and a list of the highest scoring models. A web page of results will be attached containing the interactive views of all models produced with confidence scores, secondary structure/function predictions and summary tables including information on homologs and ligand binding predictions (Fig. 24.5a, b). Models may be downloaded in the PDB format or analyzed through the built-in Phyre Investigator tool to explore mutations, model quality, and functional assignments.

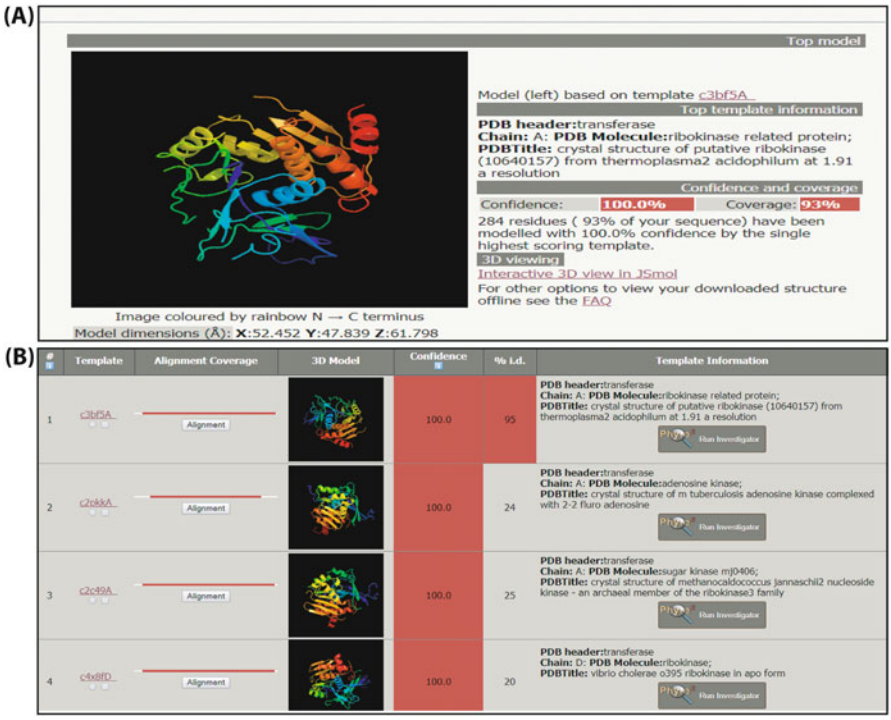


Fig. 24.5 The Phyre2 Output displaying the generation of 3D model. (**A**) Summary of information for top model generated as well as confidence score and download link for 3D model. (**B**) List of top models with alignments applied to generate them

24.2.4.5 Working Steps

I. In the Phyre2 homepage, enter relevant information such as email, job name etc.
II. Enter the amino acid sequence in the form provided.
III. Click the "Phyre search" button which directs one to the job monitoring page.
IV. On completion, an email update will be sent containing summary information and a job identifier used to explore models within the website.

24.2.5 HHPred

24.2.5.1 Outline/Methodology

The HHPred server was designed as an intermediary between fast, but lower accuracy homology search programs such as BLAST and high accuracy but slow protein prediction programs with the ultimate goal of predicting protein structure and

function. The program proceeds by first generating a multiple sequence alignment through iterative PSI-BLAST searches by applying the HHSearch software. An HMM is constructed from the query and used to search for homologous templates in alignment databases such as Pfam via HMM-HMM alignment with a neural network used to then re-rank the potential templates. Distinct from other modeling programs, HHPred then generates multiple alignments with progressively lower diversity for the target sequence. Each alignment receives a TM-score for the structural models, allowing for template selection based on the optimal alignment diversity for each domain. Finally, MODELLER is run with the optimal template to generate a 3D structure (Hildebrand et al. 2009).

24.2.5.2 Pros/Cons

HHPred has the benefit of taking extremely low computational time but does not explore alternative alignments, optimize the side chain, model loops or use model assessments - thus providing slightly lower accuracy.

24.2.5.3 Input

Protein sequence or MSA (FASTA/Clustal/A3M format) in the HHPred form online. The user may provide additional parameters such as databases to search.

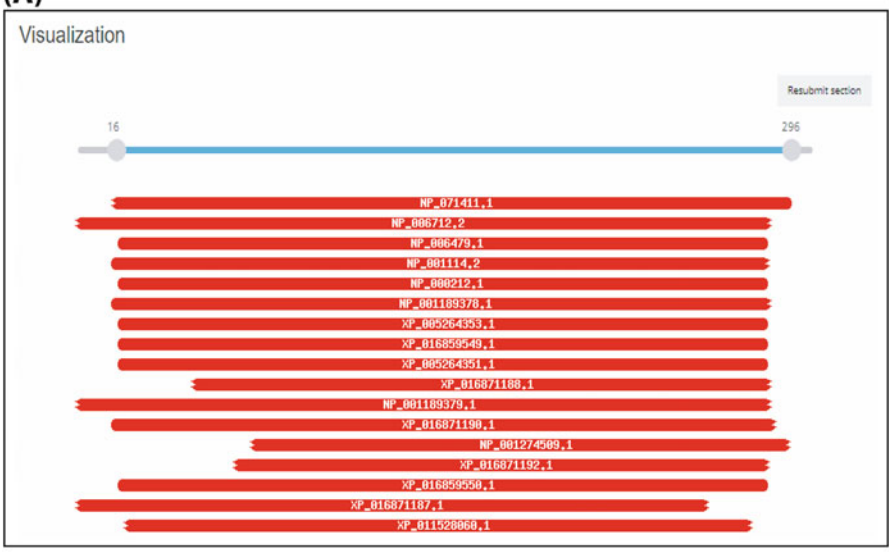
24.2.5.4 Output

Predicted alignments with visualization of overlap for top hits as well as NCBI reference sequences and accompanying confidence scores displayed within the browser window (Fig. 24.6a, b). Additional information on each protein may be identified on the NCBI website.

24.2.5.5 Working Steps

I. Submit protein amino acid sequence in an approved format using the form available on the HHPred site.
II. Press “Submit Job” and results will be displayed within the web interface.

(A)



(B)

Hitlist

Show 25 entries

Nr	Hit	Name	Probability	E-value	SS	Cols	Target Length
1	NP_071411.1	ribokinase isoform 1 [Homo sapiens]	100	2.3e-31	30.4	277	322
2	NP_006712.2	adenosine kinase isoform b [Homo sapiens]	99.96	2e-28	34.5	286	362
3	NP_006479.1	ketohexokinase isoform b [Homo sapiens]	99.95	2e-28	32.9	264	298
4	NP_001114.2	adenosine kinase isoform a [Homo sapiens]	99.95	2.5e-28	32.8	271	345
5	NP_000212.1	ketohexokinase isoform a [Homo sapiens]	99.95	2.4e-28	31.5	264	298
6	NP_001189378.1	adenosine kinase isoform c [Homo sapiens]	99.95	1.1e-28	28.1	272	327
7	XP_005264353.1	PREDICTED: ketohexokinase isoform X8 [Homo sapiens]	99.94	3.3e-27	31.8	264	299
8	XP_016859549.1	PREDICTED: ketohexokinase isoform X4 [Homo sapiens]	99.93	1.1e-25	33.7	264	343
9	XP_005264351.1	PREDICTED: ketohexokinase isoform X3 [Homo sapiens]	99.93	1.4e-26	27.7	269	344

Fig. 24.6 Visualization of working of the HHPred and model generation. (**A**) HHPred alignment overlap visualization. (**B**) HHPred hitlist displaying aligned proteins

24.2.6 ROBETTA

24.2.6.1 Outline/Methodology

Robetta is an automated web tool which functions through application of the integrated ROSETTA software which applies a fragment insertion method. Input sequences are separated into domains through the "Ginzu" method, a hierarchical screening procedure that assigns domains. If a PDB homolog is found with BLAST or HHSearch, the template is aligned through methods such as HHSearch or Compass, comparative models are generated, and loops assembled from fragments and finally are optimized to fit the aligned structure. If no homolog is available, the ROSETTA *de novo* fragment insertion method (Bonneau) is applied by assembling fragment libraries that are used to generate the model. To produce the final structure, an iterative domain assembly protocol is applied by introducing linker regions between successive domains. Notable advancements include the ability to apply experimental NMR data to more accurately construct the model as well as prediction of the effect of mutagenesis on protein-protein interactions (Kim et al. 2004).

24.2.6.2 Pros/Cons

Robetta is a significantly demanding tool, requiring 4–6 h to run a 150 residues query and the *de novo* algorithm is largely optimized for small single domain proteins (making the domain assignment step highly significant) although the program performs well compared to other servers.

24.2.6.3 Input

The user inputs the primary amino acid sequence (FASTA) into the form displayed within the Robetta website, specifying a domain identification or structure prediction job.

24.2.6.4 Output

Results are presented within the web interface and are accessible by clicking the job ID from the queue. The ten (10) full structure predictions are shown with accompanying relevant information including secondary structure prediction, disordered region prediction, domain prediction and top PSI-BLAST analogs with annotations (Fig. 24.7a–c).

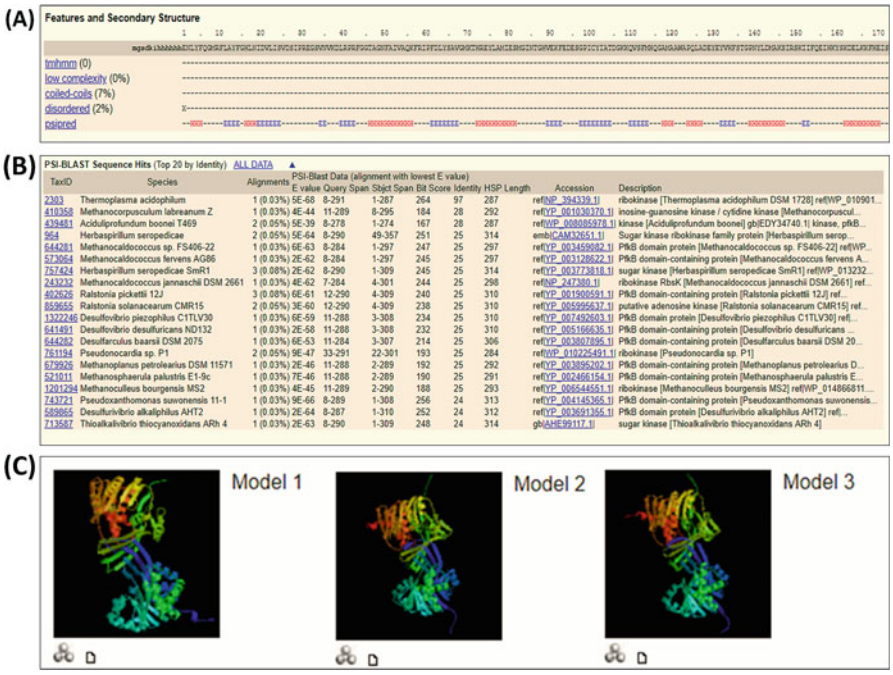


Fig. 24.7 Overall features and working of the modeling tool Robetta alongwith model generation. (**A**) Features and secondary structures of the query structure. TMHMM is applied to detect transmembrane regions, SEG to identify regions of low complexity, COILS to identify coiled-coil regions, DISOPRED to identify disordered regions, PSIPRED for secondary structures (*H* Helix, *E* Strand, *C* Coil). (**B**) PSI-BLAST hits (Top 20 by identity) displaying potential template homologs. (**C**) Top 3 generated models

24.2.6.5 Working Steps

I. From the Robetta site, select "Submit" under the section titled "Domain Parsing & 3-D Modeling."
II. Input the target name, FASTA amino acid sequence, and relevant contact information.
III. Results of modeling will be provided within the web interface.
IV. In order to generate models, from the results, one must click on "Predict domain structure with comparative modeling" under the respective Ginzu domain prediction.
V. Final models will be provided via email and within the web interface.

References

Biasini M, Bienert S, Waterhouse A, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 2014;42:W252–8. https://doi.org/10.1093/nar/gku340.

Hamuel J. Phytochemicals: extraction methods, basic structures and mode of action as potential chemotherapeutic agents. In: Phytochemicals – a global perspective of their role in nutrition and health. InTech. 2012.

Hildebrand A, Remmert M, Biegert A, Söding J. Fast and accurate automatic structure prediction with HHpred. Proteins: Struct Funct Bioinforma. 2009;77:128–32. https://doi.org/10.1002/prot.22499.

John B, Sali A. Comparative protein structure modeling by iterative alignment, model building and model assessment. Nucleic Acids Res. 2003;31:3982–92.

Kelley LA, Mezulis S, Yates CM, et al. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015;10:845–58. https://doi.org/10.1038/nprot.2015.053.

Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res. 2004;32:W526–31. https://doi.org/10.1093/nar/gkh468.

Panda PK, Ibrahim D, Kumar P, Gupta P. Computational modeling and analysis of theoretical structure of Corneodesmosin receptor protein with existing phytochemicals in psoriasis | request PDF. Indian J Fundam Appl Life Sci. 2014;4(4):346–55.

Sahay A, Shakya M. In silico analysis and homology modelling of antioxidant proteins of spinach. J Proteomics Bioinforma. 2010;03:148–54. https://doi.org/10.4172/jpb.1000134.

Salomi MV, Manimekalai R, Shoba G. Comparative in silico docking analysis of phytochemicals from Murraya koenigii and commonly used drugs in the treatment of cancer. Int J Curr Res Acta Rev. 2016;3:119–33.

Vasavi CS, Saptharshi, Radhika Devi R, et al. Homology modeling and protein ligand interaction to identify potential inhibitor for E1 protein of chikungunya. Berlin: Springer; 2010. p. 510–3.

Wallner B, Elofsson A. All are not equal: a benchmark of different homology modeling programs. Protein Sci: Publ Protein Soc. 2005;14:1315–27. https://doi.org/10.1110/ps.041253405.

Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinforma. 2008;9:40. https://doi.org/10.1186/1471-2105-9-40.