MR. MUHAMMED TILAHUN MUHAMMED (Orcid ID : 0000-0003-0050-5271)

Article type : Review

Homology Modeling in Drug Discovery: Overview, Current Applications and Future Perspectives

Muhammed Tilahun Muhammed*[1,2], Esin Aki-Yalcin[3]

[1] Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Suleyman Demirel University, 32260 Isparta, Turkey.

[2] Department of Basic Biotechnology, Institute of Biotechnology, Ankara University, Tandogan, 06100 Ankara, Turkey.

[3] Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Ankara University, Tandogan, 06100 Ankara, Turkey; Esin.Aki@ankara.edu.tr

* Correspondence: muh.tila@gmail.com; Tel.: +90-246-211-0342

Abstract

Homology modeling is one of the computational structure prediction methods that are used to determine protein 3D structure from its amino acid sequence. It is considered to be the most accurate of the computational structure prediction methods. It consists of multiple steps that are straightforward and easy to apply.

There are many tools and servers that are used for homology modeling. There is no single modeling program or server which is superior in every aspect to others. Since the functionality of the model depends on the quality of the generated protein 3D structure, maximizing the quality of homology modeling is crucial.

Homology modeling has many applications in the drug discovery process. Since drugs interact with receptors, which consists mainly of proteins in their structure, protein 3D structure determination, and thus homology modeling is important in drug discovery. Accordingly, there has been the clarification of protein interactions using 3D structures of proteins that are built with homology modeling. This contributes to the identification of novel drug candidates.

Homology modeling plays an important role in making drug discovery faster, easier, cheaper and more practical. As new modeling methods and combinations are introduced, the scope of its applications widens.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/cbdd.13388

Keywords: Current application; Drug discovery; Homology modeling; Structure prediction; 3D structure

Introduction

The world wide Protein Data Bank (wwPDB) (https://www.wwpdb.org/) contains approximately 144, 000 experimentally determined protein three dimensional (3D) structures currently [1]. In contrast the last reference sequence, which is a non redundant sequence, release of National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/) consists of annotated 155 million sequences including approximately 106 million protein sequences [2]. This represents a protein sequence number that is 736 times larger than the protein 3D strucuture deposited in the wwPDB. In 2006 the annotated sequence in NCBI was nearly 120 times larger than experimentally solved 3D structures deposited in wwPDB [3]. This means the number of protein sequences has increased 6 times faster than the number of experimentally determined protein 3D structures. Since the protein data banks available contain redundancy but the sequences in NCBI are non redundant, the difference is higher than the numbers given here. This growing gap between the sequences available and the protein 3D structures determined is in an alarming condition. Thus, computational structural determination methods are needed in filling this widening gap between the number of sequences available and protein 3D structures solved experimentally.

Since crystal structure of the first protein myoglobin was solved in 1960, there has been an improvement in the quality of the 3D structures determined. This has been achieved with the introduction of experimental methods like X-ray crystallography and NMR spectroscopy [4]. However, these experimental methods can not be used for each protein. For NMR analysis protein molecules should be small and for X-ray crystallography the molecules should be crystallized. Additionally, these methods are time consuming. Thus, there is deficiency in high resolution 3D structure of proteins, especially membrane proteins due to the difficulties in purification and crystallization of such proteins in relative to other small water soluble proteins [5]. Since membrane proteins constitute important proportion of therapeutic drug targets, advances in the determination of membrane proteins will speed up the drug discovery process. Here computational protein 3D structure prediction can play a crucial role.

Homology modeling (comparative modeling) is one of the computational structure prediction methods that are used to determine 3D structure of a protein from its amino acid sequence based on its template. The basis for homology modeling are two major observations. First protein 3D structure is particularly determined by its amino acid sequence. Second the structure of proteins is more conserved and the change happens at a much slower rate in relative to the sequence during evolution. As a result, similar sequences fold into identical structures and even sequences with low relation take similar structures [6].

Homology modeling is considered to be the most accurate of the computational structure prediction methods [7]. 3D structure predictions made by computational methods like *de novo* prediction and threading were compared to homology modeling using Root Mean Square Deviation (RMSD) as a criteria. Homology modeling was found to give 3D structures with the highest accuracy [8]. Furthermore, it is a protein 3D structure prediction method that needs less time and lower cost with

clear steps. Thus, homology modeling is widely used for the generation of 3D structures of proteins with high quality. This has changed the ways of docking and virtual screening methods that are based on structure in the drug discovery process [9].

In this review, the main features of steps of homology modeling are presented. The popular tools and servers that have been used for homology modeling in recent years are also summarized. Overview of the striking homology modeling applications in the prediction of protein 3D structures and recent applications in the drug discovery are also discussed. This review also provides insight into the opportunities and possible challenges in homology modeling.

Steps of Homology Modeling

Homology modeling is a structure prediction method that consists of multiple steps. Homology modeling has common standard procedures with minor differences. The standard steps of homology modeling are summarized in Figure 1 and the detail explanation is given below the figure.

Identification and Selection of Templates

In this step of the process target (query) sequence is used for the identification of template structures in the PDB (https://www.rcsb.org/) [10] or similar databases. There are popular tools in searching for eligible templates for target sequence with different approaches. Among of them, Basic Local Alignment Search Tool (BLAST) [11] is the one which provides pairwise sequence-sequence alignment. This service is available inside databases like NCBI [2] and UniProt (http://www.uniprot.org/) [12]. The other approaches used in template identification are profile-profile alignments [13] and Hidden Markov Models (HMMs) [14]. Some other advanced approaches use profile-profiles and HMMs in combination with structural properties.

After template candidates are identified, the best structures must be selected. Sequence similarity level of the template sequence in relative to the target sequence is important in generating 3D structures with high accuracy. However, sequence similarity is not the only factor that determines the accuracy of the structures generated in homology modeling. Regarding the minimum sequence similarity limit in homology modeling, there are ambiguities about the exact value but >25% suggests that the template and target will take similar 3D structures [15].

Apart from high sequence similarity, various factors are considered in choosing an eligible template. These factors include phylogenetic similarity between template and target sequences. Templates from identical or analogous phylogenetic tree to the target sequence may result in a 3D structure with high accuracy [16]. The other factors are environmental factors such as pH, solvent type and existence of bound ligand. These are also important in choosing the most eligible template as it has a role in ensuring the most optimal conditions in building an accurate target structure. The resolution of the experimental structure under consideration is also a factor in choosing the eligible template [17].

Sequence Alignments and Alignment Correction

After the most appropriate alignments are selected, alignments and correction of them in case it is necessary is undertaken. The alignments are target-template and template-template when more than one template is used. The error in the alignment of a residue causes shifting of α carbon. A single residue gap in an α helix section triggers rotation of the rest of the residues in the helix. As a result, the alignment of sequences in the right way is crucial in homology modeling [18]. Careful checkups and correction while performing alignments may enhance building 3D protein structures with high quality. The most widely used alignment methods are Clustal W (http://www.genome.jp/tools-bin/clustalw) [19], T-Coffee (http://tcoffee.crg.cat/) [20], 3Dcoffee (http://phylogeny.lirmm.fr/phylo_cgi/) [21] and MUSCLE (https://www.ebi.ac.uk/Tools/msa/muscle/) [22].

Model Building

Various methods are used to generate 3D models for the target sequence based on its templates. Model building approaches can be classified as rigid body-assembly methods, segment matching methods, spatial restraint methods and artificial evolution methods.

In rigid-body assembly the protein structure is broken down into basic conserved core regions, loops and side chains. This approach depends on the natural dissection that enables the building of a protein 3D structure by bringing this rigid bodies together which are picked up from the aligned template protein structures [23]. This can be done by tools like 3D-JIGSAW [24], BUILDER [25] and SWISS-MODEL [26].

In segment matching method a cluster of atomic positions obtained from the template structures are used as leading positions. Selection of segments from known structures in a database for matching the segments is done based on the sequence identity, geometry and energy. Then the entire atom model is generated by using the leading structure as a pillar to lay the segments. This can be done by using SEGMOD/ENCAD [27].

Spatial restraint method builds the model by meeting restraints came from the template structure. The restraints are framed onto the target structure depending on the alignment. These restraints are determined by stereochemical restraints on bond length, bond angle, dihedral angles and van der waals contact distances. This can be performed with MODELLER [28].

Artificial evolution method uses rigid-body assembly method and stepwise template evolutionary mutations together until the template sequence is the same as the target sequence. This can be performed with NEST [29].

Table 1 displays summary of general features of the popular tools and servers that can be used for model building. Researchers reported that when the sequence identity is high, the homology models derived from different packages are comparable to each other. When the sequence identity is lower, the results tend to vary, with some packages performing noticeably better than others [30]. The quality of the models is related with the performance of packages in sequence alignment and model building. MODELLER is found to be one of the best tools in homology modeling [31]. In addition to

this critical assessment of methods of protein structure prediction (CASP) assesses modeling methods in a number of different categories. I-TASSER was ranked as the best server for protein structure prediction in recent CASP experiments [32]. These tools and servers have their own pros and cons. As a result, there is no single modeling tool or server which is superior in every aspects to others.

Loop Modeling

Gaps or insertions called loops are present in sequences of homologous proteins. The structures of loops are not conserved during evolution. Even without deletions or insertions different loop conformations in query and template are often found. The specificity of the function of a protein structure is often determined by the loops. Accuracy of loop modeling is an important factor which determines the value of the generated models for further applications. Since loops show higher structural variability than strands and helices, the prediction of their structure is more difficult than strands and helices [44].

There are two important methods that are used in developing the loops. One is database search approach and the other is conformation search approach. The database search method browses all the known protein structures to detect segments providing the critical core regions. The conformational search approach depends on a scoring function optimization [8]. Loop searches are done for loops of length 4-7 residues these days. This is because of the conformation variation increase as the length of the loop increases.

To deal with these drawbacks, *de novo* methods that are used for loop conformation predictions by looking for conformational space have been developed. Monte Carlo simulations, simulated annealing, genetic algorithms and molecular dynamics simulations often in combination with knowledge-based potentials are examples for this. In such methods the length of loop that can be modelled is not limited but as the length increases possible conformation number increases rapidly which makes the modeling very time consuming [45]. There are servers such as ArchPRED (http://www.bioinsilico.org/ARCHPRED/) [46] and Congen (http://www.congenomics.com/congen/doc/) [6,47] that are used in loop modeling.

Side Chain Modeling

Side chain modeling is usually done by putting side chains onto the backbone coordinates that are derived from a parent structure and/or from *ab initio* modeling simulations. In practice side chain prediction works at high levels of sequence identity. Protein side chains are present in a limited number of structures with low energy known as rotamers. Depending on defined energy functions and search strategies, rotamers are selected in accordance with the preferred protein sequence and the given backbone coordinates. The accuracy of prediction is usually high for the hydrophobic core residues but low for water exposed residues on the surface [48]. Tools like RAMP (http://www.ram.org/computing/ramp/) [41] and SCWRL [49] can be used in side chain modeling.

Model Optimization

Optimization of the model usually begins with an energy minimization utilizing molecular mechanics force fields [50]. At each energy minimization a few big errors are eliminated but many other small errors are introduced at the same time and start accumulating. Therefore, restraining the atom positions, implementing energy minimization with a few hundred steps and using more precise force fields like quantum force fields [51] and self-parameterizing force fields [52] can be utilized to decrease the errors in model optimization. For further model optimization methods such as molecular dynamics and Monte Carlo can be used [53].

Model Validation

Accuracy of the constructed model can determine its further application in various areas. Thus, verification and validation of models are necessary. Depending on sequence similarity, environmental parameters and the quality of the templates, the generated models have different accuracy.

Analysis of the stereochemistry of the model is one basic requirement. This analysis is done with parameters such as bond length, torsion angle and rotational angle. WHATCHECK (https://swift.cmbi.umcn.nl/gv/whatcheck/) [54], PROCHECK (https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/) [55] and Molprobity (http://molprobity.biochem.duke.edu/) [56] are popular tools used for the determination of the stereochemistry of the model in homology modeling. The Ramachandran plot (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) is also powerful determinant of the quality of protein structure. Residues with a problem of stereochemistry will fall out of the acceptable regions of the Ramachandran plot [57].

There are also tools that focus on the determination of the spatial features of the model based on 3D conformations and mean force statistical potentials. VERIFY3D (http://servicesn.mbi.ucla.edu/Verify3d/) [58] and PROSAII (https://www.came.sbg.ac.at/prosa.php) [7] are examples for this. These tools consider model construction environmental parameters in relative to the expected environmental conditions.

Applications of Homology Modeling

Homology modeling has a vast range of applications and its importance is increasing as the number of structures determined increases. It has applications in structure based drug design, analysis of mutations, insight into binding mechanisms, identification of active sites, looking for ligands and designing of novel ligands, modeling of substrate specificity, protein–protein docking simulations, molecular replacement in experimental structural refinements, rationalizing of known experimental results and planning of future computational experiments by using the generated models [59].

Homology modeling has many applications in drug discovery process. This makes the drug discovery process faster, easier, cheaper and more practical. In general homology modeling applications in the drug discovery need high quality models. As a result, high sequence similarity, good side chain modeling and loop modeling are crucial in determining further applications of the model build in the drug discovery.

As an illustration of the case application, for example, homology modeling was used to discover novel acetohydroxy acid synthase (AHAS, EC 2.2.1.6) inhibitors against *M. Tuberculosis*. Several studies demonstrated that the plant AHAS inhibitors of sulfonylurea chemicals such as sulfometuron methyl (SMM) exhibit antituberculosis activity. However, the 3D structure of *M. tuberculosis* AHAS remains to be elucidated. Thus, homology modeling was performed based on the *S. cerevisiae* AHAS to build a 3D structure of *M. Tuberculosis* AHAS. Through docking simulation and similarity searches, 23 novel AHAS inhibitors of *E. coli* AHAS II enzymatic activity were identified. Five of the identified chemicals showed strong inhibitory effects against multidrug-resistant and extensively drug-resistant strains. Three of the compounds exhibited more activity than the positive control SMM [60].

In recent years 3D structure of targets in cancer that can be used for discovering effective chemotherapeutic agents has been generated using homology modeling [61,62]. Reliable 3D structures of G-protein coupled receptors (GPCRs) which are targets of nearly a third of FDA approved drugs has been built similarly [63]. Another recent application of homology modeling is 3D structure determination of RNA polymerase of the Ebola virus that helps in the detection of potential therapeutic agents [64]. Furthermore 3D structure of NS5 protein of the Zika virus has been determined by homology modeling that leads to the discovery of its potential inhibitors [65]. Recent case applications of homology modeling in drug discovery are summarized in Table 2.

Opportunities and Possible Challenges in Homology Modeling

The number of high quality protein 3D structures has increased in the last decades. The introduction of new experimental methods like Cryo-electron microscopy (Cryo-EM) is anticipated to increase the number of 3D structures determined experimentally [80]. As the experimentally determined number of high quality 3D protein structures of protein families increases, the role of homology modeling in determining the 3D structures of the rest of the sequences in these families increases. However, 3D structures of all protein distinct folds in nature has not been completed yet. As a result, there are some difficulties in building 3D structures of proteins in which the structures of their protein families have not been determined [18].

There are dozens of methods used for model building in homology modeling. New methods with new algorithms have been developed. Various studies have demonstrated that there is no single modeling program or server which is superior in every properties to others [81]. So, selecting the method/s to be used according to the protein in hand and specific aim of future applications of the model is important.

In classical homology modeling the model is built mainly based on sequence similarity. In the experimental structure determination, ligands are absent as they are often lost during the purification process. Thus, the resulting models that are built without considering the ligand

information in the template represent an unliganded state. This shortcoming has been dealed with the introduction of ligand sensitive approaches. However, such approaches need expertise and manual interventions that takes time. Hence the introduction of fully automated homology modeling tools that can deal with such problems is an important issue [82]. Furthermore, there are efforts to integrate it with post modeling applications. For instance, there are works to integrate modeling tools with thermostabilizing mutations [83].

Homology modeling may leave some unresolved questions in the computational models. This can be reduced by using models that came from more experimentally determined structures which allow better conceivable templates for targets. As consistent, accurate and progressive methods for the improvement of models by shifting the coordinates parallel to the native state are developed, coverage increases [84].

Another limitation of homology modeling is presence of loops and inserts as it is difficult to model them without template data [85]. In order to have a model with high accuracy, optimization of the loop region and side chains is important. Optimization encompasses refinement of the generated models with molecular dynamics simulations. In case there is low sequence similarity level between target and template, using multiple templates is advantageous. But using multiple templates may lead to aberrations in the alignment unless templates which are from identical or analogous phylogenetic tree are used as the target sequence [86]. Using PSI BLAST algorithm instead of normal BLAST may provide optimal template selections in evolutionary distant cases.

At the end of the homology modeling process, many models of a target are built in general. Having many models is an opportunity but identification of the best model needs further investigation. In order to identify the best model, the constructed models are compared using various parameters. Discrete Optimized Protein Energy (DOPE) score [87], Template Modeling (TM) score [88] and Root Mean Square Deviation (RMSD) value [89] are used for comparison. The determinant parameter is decided depending on the purpose of modeling results.

Conclusion

The gap between protein sequences available and protein 3D structures determined experimentally is growing. Homology modeling aims at building 3D structure of proteins from their sequences by using templates with an accuracy which is similar to the experimental methods. Thus, it has a big role in filling the widening gap.

In recent years there are many advances in the tools and servers of homology modeling that improve the accuracy of modeling results. This has an impact on each step of homology modeling. Better alignment methods, loop modeling, side chain modeling and validation techniques have been introduced. As the accuracy of models generated increases, their applications in the drug discovery process increase. So, homology modeling contributes much in the drug discovery. Furthermore, in the near future integration of homology modeling with other computer aided drug design methods and post modeling applications are expected.

Homology modeling is used in determining 3D structures of proteins and it has many applications in the drug discovery process.

Figure 1. Steps in Homology Modeling

Table 1. Popular Homology Modeling Tools and Servers

Table 2. Recent Applications of Homology Modeling

Disclosure

The authors declare that there is no conflict of interest in this work.

References

[1] B. Helen, H. Kim, N. Haruki, *Nat. Struct. Mol. Biol.* **2003**, *10*, 980.

[2] R. Agarwala, T. Barrett, J. Beck, D. A. Benson, C. Bollin, E. Bolton, et al., *Nucleic Acids Res.* **2016**, *44*, D7.

[3] M. Levitt, *Pnas* **2007**, *104*, 3183.

[4] N. I. for G. M. S. U. NIGMS, NIGMS - 2011 - 50 Years of Protein Structure Determination Timeline. *https://publications.nigms.nih.gov/psi/timeline_text.html* **2011**.

[5] M. Baker, *Nat. Methods* **2010**, 429.

[6] E. Krieger, S.D. Nabuurs, G. Vriend, *Structural Bioinformatics*, P.E. Bourne, H. Weissig, Ed., Wiley-Liss, **2012**, pp. 507–520.

[7] C. N. Cavasotto, S. S. Phatak, *Drug Discov. Today* **2009**, *14*, 676.

[8] T. Werner, M. B. Morris, S. Dastmalchi, W. B. Church, *Adv. Drug Deliv. Rev.* **2012**, *64*, 323.

[9] T. Cheng, Q. Li, Z. Zhou, Y. Wang, S. H. Bryant, *AAPS J.* **2012**, *14*, 133.

[10] P. W. Rose, A. Prlić, A. Altunkaya, C. Bi, A. R. Bradley, C. H. Christie, et al., *Nucleic Acids Res.* **2017**, *45*, D271.

[11] S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, et al., *Nucleic Acids Res.* **1997**, *25*, 3389.

[12] R. Apweiler, M. J. Martin, C. O'Donovan, M. Magrane, Y. Alam-Faruque, R. Antunes, et al., *Nucleic Acids Res.* **2011**, *39*, 214.

[13] G. Wang, R. L. Dunbrack, Jr., *Protein Sci.* **2004**, *13*, 1612.

[14] J. Söding, *Bioinformatics* **2005**, *21*, 951.

[15] B. Rost, C. Sander, *Annu. Rev. Biophys. Biomol. Struct.* **1996**, *25*, 113.

[16] J. Peng, Statistical Inference for template-based protein structure prediction, Toyota Technological Institute, **2013**.

[17] M. S. Saxena A, Sangwan RS, *Sci. Imternational* **2013**, *7*.

[18] J. A. R. Dalton, R. M. Jackson, *Bioinformatics* **2007**, *23*, 1901.

[19] M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. Mcgettigan, H. McWilliam, et al., *Bioinformatics* **2007**, *23*, 2947.

[20] C. Notredame, D. G. Higgins, J. Heringa, *J. Mol. Biol.* **2000**, *302*, 205.

[21] O. O'Sullivan, K. Suhre, C. Abergel, D. G. Higgins, C. Notredame, *J. Mol. Biol.* **2004**, *340*, 385.

[22] R. C. Edgar, *Nucleic Acids Res.* **2004**, *32*, 1792.

[23] A. R. Katebi, A. Kloczkowski, R. L. Jernigan, *BMC Struct. Biol.* **2010**, *10*, 1.

[24] P. A. Bates, L. A. Kelley, R. M. MacCallum, M. J. E. Sternberg, *Proteins Struct. Funct. Genet.* **2001**, *45*, 39.

[25] T. R. Kim, S. Oh, J. S. Yang, S. Lee, S. Shin, J. Lee, *J. Comput. Chem.* **2012**, *33*, 1927.

[26] K. Arnold, L. Bordoli, J. Kopp, T. Schwede, *Bioinformatics* **2006**, *22*, 195.

[27] M. Levitt, *J. Mol. Biol.* **1992**, *226*, 507.

[28] B. T. Sali A, *J. Mol. Biol.* **1993**, *234*, 779.

[29] D. Petrey, Z. Xiang, C. L. Tang, L. Xie, M. Gimpelev, T. Mitros, et al., *Proteins Struct. Funct. Genet.* **2003**, *53*, 430.

[30] B. K. Kuntal, P. Aparoy, P. Reddanna, *BMC Res.Notes* **2010**, 1.

[31] B. Wallner, A. Elofsson, *Protein Sci.* **2005**, *14*, 1315.

[32] A. Kryshtafovych, B. Monastyrskyy, K. Fidelis, J. Moult, T. Schwede, A. Tramontano, *Proteins Struct. Funct. Bioinforma.* **2018**, *86*, 321.

[33] Y. Zhang, *BMC Bioinformatics* **2008**, *9*, 1.

[34] J. C. Almagro, M. P. Beavers, F. Hernandez-Guzman, J. Maier, J. Shaulsky, K. Butenhof, et al., *Proteins Struct. Funct. Bioinforma.* **2011**, *79*, 3050.

[35] L. A. Kelly, S. Mezulis, C. Yates, M. Wass, M. Sternberg, *Nat. Protoc.* **2015**, *10*, 845.

[36] J. Shi, T. L. Blundell, K. Mizuguchi, *J. Mol. Biol.* **2001**, *310*, 243.

[37] D. E. Kim, D. Chivian, D. Baker, *Nucleic Acids Res.* **2004**, *32*, 526.

[38] J. Haas, S. Roth, K. Arnold, F. Kiefer, T. Schmidt, L. Bordoli, et al., *Database* **2013**, 1.

[39] T. Cardozo, M. Totrov, R. Abagyan, *Proteins Struct. Funct. Bioinforma.* **1995**, *23*, 403.

[40] D. Sindhikara, S. A. Spronk, T. Day, K. Borrelli, D. L. Cheney, S. L. Posy, *J. Chem. Inf. Model.* **2017**, *57*, 1881.

[41] G. G. Krivov, M. V. Shapovalov, R. L. Dunbrack, *Proteins Struct. Funct. Bioinforma.* **2009**, *77*, 778.

[42] D. B. Roche, M. T. Buenavista, S. J. Tetchner, L. J. McGuffin, *Nucleic Acids Res.* **2011**, *39*, 171.

[43] S. J. Lee J, Lee D, Park H, Coutsias EA, *Proteins* **2010**, *78*, 3428.

[44] S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid, A. Kolinski, *Chem. Rev.* **2016**, *116*, 7898.

[45] N. Fernandez-Fuentes, J. Zhai, A. Fiser, *Nucleic Acids Res.* **2006**, *34*, 173.

[46] R. E. Bruccoleri, *Mol. Simul.* **1993**, *10*, 151.

[47] V. K. Vyas, R. D. Ukawala, M. Ghate, C. Chintha, *Indian J Pharm Sci.* **2012**, *74*, 1.

[48] M. J. Samudrala R, *Protein Eng.* **1998**, *11*, 991.

[49] J. Zhu, H. Fan, X. Periole, B. Honig, *Proteins* **2008**, *72*, 1171.

[50] Y. W. Liu H, Elstner M, Kaxiras E, Frauenheim T, Hermans J, *Proteins Struct. Funct. Bioinforma.* **2001**, *44*, 484.

[51] E. Krieger, G. Koraimann, G. Vriend, *Proteins Struct. Funct. Genet.* **2002**, *47*, 393.

[52] R. Han, A. Leo-Macias, D. Zerbino, U. Bastolla, B. Contreras-Moreira, *Proteins Struct. Funct. Bioinforma.* **2007**, *71*, 175.

[53] R. W. W. Hooft, G. Vriend, C. Sander, *Nature* **1996**, *381*, 272.

[54] R. A. Laskowski, M. W. Macarthur, D. S. Moss, *J. Appl. Cryst.* **1993**, *26*, 283.

[55] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, et al., *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 12.

[56] O. Carugo, K. Djinovic Carugo, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2013**, *69*, 1333.

[57] B. Wallner, A. Elofsson, *Protein Sci.* **2006**, *15*, 900.

[58] M. Wiederstein, M. J. Sippl, *Nucleic Acids Res.* **2007**, *35*, 407.

[59] Y. Koseki, S. Aoki, *Curr. Top. Med. Chem.* **2014**, *14*, 176.

[60] J. Wang, W. Li, B. Wang, B. Hu, H. Jiang, B. Lai, et al., *Curr. Pharmacol. Reports* **2017**, *3*, 184.

[61] J. P. Overington, B. Al-Lazikani, A.L. Hopkins, *Nat. Rev. Drug Discov.* **2006**, *5*, 993.

[62] I. Kufareva, M. Rueda, V. Katritch, R. C. Stevens, *Structure* **2011**, *19*, 1108.

[63] M. Balmith, M. Faya, M. E. S. Soliman, *Chem. Biol. Drug Des.* **2017**, *89*, 297.

[64] P. Ramharack, M.E.S. Soliman, *J Biomol .Struct. Dyn.* **2018**, *36*, :1118.

[65] V. K. Vyas, M. Ghate, K. Patel, G. Qureshi, S. Shah, *Biomed. Pharmacother.* **2015**, *74*, 42.

[66] S. Mishra, V.S. Gomase, *J. Heal. Med. Informatics* **2016**, *7*.

[67] H. Tarazi, E. Saleh, R. El-Awady, *Biomed. Pharmacother.* **2016**, *83*, 693.

[68] C. Arimany-Nardi, A. Claudio-Montero, A. Viel-Oliva, P. Schmidtke, C. Estarellas, X. Barril, et al., *Mol. Pharm.* **2017**, *14*, 1980.

[69] S. Prabhu, S. Vijayakumar, P. Manogar, G. P. Maniam, N. Govindan, *Biomed. Pharmacother.* **2017**, *92*, 528.

[70] R. B. Singh, G. K. Singh, K. Chaturvedi, D. Kumar, S. K. Singh, M. K. Zaman, *Med. Chem. Res.* **2017**, 1.

[71] N. C. Jadhav, A. R. Pahelkar, N. V. Desai, V. N. Telvekar, *Med. Chem. Res.* **2017**, *26*, 2675.

[72] Z. Payandeh, M. Rajabibazl, Y. Mortazavi, A. Rahimpour, *Int. J. Pept. Res. Ther.* **2017**, *0*, 0.

[73] V. Singh, N. Gohil, R. Ramírez-García, *J. Cell. Biochem.* **2018**, *119*, 2003.

[74] Nalini, N. Chadha, M. S. Bahia, M. Kaur, R. Bahadur, O. Silakari, *Mol. Divers.* **2017**, *22*, 1.

[75] P. S. Mohanty, A. K. Bansal, F. Naaz, U. D. Gupta, V. D. Dwivedi, U. Yadava, *Infect. Genet. Evol.* **2018**, *60*, 58.

[76] J. Iqbal, S. J. A. Shah, *Sci. Rep.* **2018**, *8*, 2581.

[77] N. Antoniou, D. Vlachakis, A. Memou, E. Leandrou, P.-E. Valkimadi, K. Melachroinou, et al., *Sci. Rep.* **2018**, *8*, 3455.

[78] L. Ferri, D. Malesci, A. Fioravanti, G. Bagordo, A. Filippini, A. Ficcadenti, et al., *Clin. Chim. Acta* **2018**, *481*, 25.

[79] M. Carroni, H. R. Saibil, *Methods* **2016**, *95*, 78.

[80] Z. Xiang, *Curr Protein Pept Sci.* **2006**, *7*, 217.

[81] T. Schmidt, A. Bergner, T. Schwede, *Drug Discov. Today* **2014**, *19*, 890.

[82] Y. Kajiwara, S. Yasuda, Y. Takamuku, T. Murata, M. Kinoshita, *J. Comput. Chem.* **2017**, *38*, 211.

[83] C. L. Gupta, S. Akhtar, P. Bajpai, *EXCLI J.* **2014**, *13*, 513.

[84] F. Beer, Tjaart A P De, Taştan Bishop, A. Özlem, Joubert, *S. Afr. J. Sci.* **2008**, *104*, 2.

[85] G. Munsamy, M. E. S. Soliman, *Lett. Drug Des. Discov.* **2017**, *14*.

[86] M. Shen, A. Sali, *Protein Sci.* **2006**, 2507.

[87] Y. Zhang, J. Skolnick, *Proteins Struct. Funct. Genet.* **2004**, *57*, 702.

[88] I. Kufareva, R. Abagyan, *Methods Mol Biol.* **2012**, *857*, 231.

[89] A. Nayeem, D. Sitkoff, S. Krystek, *Protein Sci.* **2006**, *15*, 808.

Homology Modeling Tools or Servers	URL address	Short Discription
MODELLER	http://www.salilab.org/modeller/	Is a homology modeling tool that generates protein 3D structures with spatial restraints method. It is available freely, has powerful features and gives reliable results [33].
I-TASSER	https://zhanglab.ccmb.med.umich.edu/I-TASSER/	Is a server that provides an internet based service for protein structure prediction. It was found to be one of the best methods in the servers section of CASP experiments [26].
SWISS-MODEL	http://swissmodel.expasy.org/	Is a server that gives protein 3D structure from its amino acid sequence. It provides user friendly web interface. This server uses model quality estimation to select the most appropriate templates and gives the expected accuracy of the models built approximately [34].
Molecular Operating Environment (MOE)	https://www.chemcomp.com/MOE-Molecular_Operating_Environment.htm	Is a combination of segment matching and modeling of insertion or deletion regions approaches. In addition to 3D structure prediction, it has advanced loop modeling, advanced alignment methods and powerful alignment visualizer and editor [35].
Phyre2	http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index	This modeling uses various detection tools to generate 3D structures. It has special features like ligand binding prediction and variant analysis among the protein amino acid sequence [36].
HHpred	http://toolkit.tuebingen.mpg.de/hhpred	This tool builds 3D structures using pairwise comparison of profile hidden Markov models (HMMs) from a single or multiple query sequence [37].
Robetta	http://www.robetta.org/	Based on the ROSETTA fragment insertion method, it gives both *ab initio* and homolog models of protein regions [38].
Protein Model Portal (PMP)	http://www.proteinmodelportal.org/	PMP provides interactive interface for model building and quality assessment [39].
ICM	https://www.molsoft.com/homology.html	Is one of the homology modeling tools that give 3D structure with good accuracy. Its features include fast model building, loop prediction, model validation and refinement [40].
Prime	https://www.schrodinger.com/prime	Is a powerful package for accurate protein structure prediction. In addition to building structures with high accuracy, it provides advanced simulation. It makes homology modeling and fold recognition merge into a

		package. It has an easy –to-use interface [41].
SCWRL4	http://dunbrack.fccc.edu/scwrl4/index.php	It is a tool that is rapid with good accuracy and easy-to-use [42].
IntFOLD	http://www.reading.ac.uk/bioinf/IntFOLD/	It is an independent server that predicts intrinsic disorders, domains and protein-ligand binding sites [43].

Protein (Molecule)	**Application**	**Program/Server**
Human angiotensin II type I receptor [66]	Guide for designing novel therapeutic agents as angiotensin receptor antagonists	BLAST, CLUSTALW, SYBYL, MODELLER, I-TASSER, PROCHECK, SurflexDock
HSP70 from GWD [67]	Determination of 3D structure of hsp70 chaperone protein which is a target of new wide spectrum cancer therapeutics candidates	BLAST, SWISS-MODEL, QMEAN, PSVS
DNA Dependent Protein Kinase (DNA-PK) [68]	Screening of potential candidates as DNA-PK inhibitors	WHATIF, ProSA, AutoDock Vina, Discovery Studio, ClustalW, Phyre2, WHATCHECK
Human Concentrative Nucleoside Transporter 3 (hCNT3) [69]	Identification of sodium binding site and the determinant residues of nucleoside selectivity	MOE, Gold, Glide, CHARMM
Peroxisome Proliferator Activated Receptor gamma (PPARγ) [70]	Identification of new ligand molecules that reduce PPARγ receptor in Type II diabetes complications	Prime, GlideXP, Schrodinger
GABA Transporter 1 (GAT1) [71]	Discovering GAT1 inhibitor molecules that are potential anticonvulsant and antidepressant agents	ClustalW, Prime, GlideXP, Schrodinger
α-Glucosidase [72]	Designing of new classes of α-glucosidase inhibitors	BLAST, Prime, PROCHECK, Sitemap, GlideXP, Schrodinger, Maesterero
CD20 antigen [73]	Insight into the structure of CD20 antigen which is a target in developing new monoclonal antibodies	PSI-BLAST, T-Coffee, SWISS-MODEL, I-TASSER, Phyre2, MUSTER, Rampage
Histamine H2 receptor [74]	New perspectives into the development of a new potent drugs against peptic ulcer by targeting H2	BLAST, ClustalX, MODELLER, PROCHECK, AutoDock, STRING
Protein Kinase D 1 (PKD1) [75]	Designing new PKD1 inhibitors	Schrodinger suit, Maestero, MODELLER, PDBSum, Glide XP

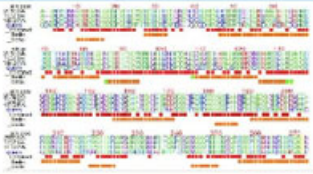
Ribonucleotide reductase [76]	Screening of novel drugs for drug resistant Leprosy therapy	SWISS-MODEL, HHPred, ProFunc, ERRAT, WHATIF, ProSA, Glide XP, Schrodinger
Ecto-Nucleoside Triphosphate Diphosphohydrolases (E-NTPDases) [77]	Structural insights into the binding of E-NTPDases to substrates and inhibitors	MOE, BLAST, Rampage, Marvin Sketch, FleX X, GROMACS
Parkinson's linked mutant Leucine-Rich Repeat Kinase 2 (LRRK2) [78]	Identification of multiple novel points within neuronal death signaling pathways that could be targeted by potential therapeutic candidates	MOE, Glide 1, Maestero, CHARMM
Alpha-Galactosidase A (α-Gal A) [79]	Detection of six GLA variants that cause α-Gal A activity deficiency and protein wild type structure loss	Alamut Visual, SWISS MODEL, PyMol

1

Identification and Selection of Templates
Sequence similarity, phylogenetic and environmental factors are used.

2

Sequence Alignments and Alignment Correction
Careful inspection and correction of alignments is crucial.

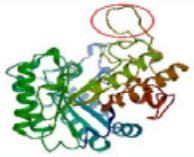


3

Model Building
There are various methods that can be used in model generation.

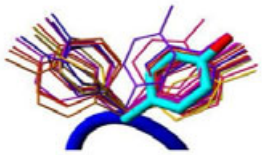
4

Loop Modeling
Accuracy of loop modeling determines the value of the models for further applications.



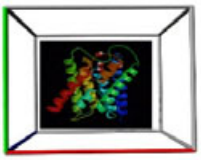
5

Side Chain Modeling
This is usually done by putting side chain onto the backbone.



6

Model Optimization
This step is done to improve the quality of the model.



7

Model Validation
The quality of the model determines the functionality of it.