



OPEN

# EasyModel: a user-friendly web-based interface based on MODELLER

Seyed Shahriar Arab  & Alireza Dantism

Three-dimensional protein structures are invaluable sources of information for the functional annotation of protein molecules. Describing the function of a protein sequence is one of the most common problems in biology. Generally, this problem can be facilitated by studying the tertiary structure of proteins. In the lack of protein structures, comparative modeling often provides a useful three-dimensional model of the protein associated with at least one known protein structure. Comparative modeling predicts the tertiary structure of a certain protein sequence (target) mainly based on its homological sequence to the sequence of one or more proteins with known structures (templates). MODELLER is one of the most widely used tools for homology or comparative modeling of three-dimensional protein structures. However, most users find it challenging to start with MODELLER as it is a command line based and requires knowledge of basic Python scripting to use it efficiently. In this study, a web-based interface has been designed to predict the tertiary structure of proteins based on Modeller, which does the comparative modeling automatically, and uses PHP and Python programming languages. This tool is called “EasyModel” and is available at <http://bioinf.modares.ac.ir/software/easymodel/>. EasyModel provides a straightforward graphical interface for Modeller that can be used in only one browser.

The basis of prediction by homology modeling is that the protein sequence is similar to one or more proteins with known structures. Based on the fact that proteins with similar sequences have identical structures. Homology modeling predicts a given protein sequence (target) establish on its alignment with one or more proteins of known structure (templates) and the alignment of target and template(s) sequences.

The steps of this process are: (1) identification of homologs that can serve as template(s) for modeling; (2) alignment of the target sequence to the template(s); (3) backbone generation; (4) loop modeling; (5) side-chain modeling; (6) model optimization; and (7) validation of the model<sup>1</sup>.


Several computer programs and web servers serve as a graphical user interface for Modeller<sup>2</sup>. These interfaces are categorized into two types, web-based and desktop applications.

Among the web-based applications, Esyspred3D<sup>3</sup>, HHpred<sup>4</sup>, and Modbase<sup>5</sup> can be mentioned. For desktop applications, ChimeraX<sup>6</sup>, and EasyModeller<sup>7</sup> are notable choices.

Modeller is a widely used program for the homology modeling of proteins, though some users find it challenging to navigate due to its command line interface and reliance on Python expertise. As a command line application without a graphical interface, users must possess basic knowledge of both the command line and Python to achieve optimal performance. To address this issue, various companies and individuals have developed visual interface software to make Modeller more accessible to users. However, these interfaces are depending on the operating system (Windows or Mac) or they consider it necessary to install Python on the operating system, also, some of them are no longer available or not under active development. Most of these programs are limited to basic modeling, and many older programs are no longer available due to their incompatibility with newer versions of Modeller.

Therefore, the approach adopted in this study is developing and implementing an online bioinformatics interface for predicting the tertiary structure of proteins. For this prediction, the homology modeling method and Modeller are used. All the parts that researchers required, including Python, and the command line, have been removed so that researchers can complete all their modeling steps with only a straightforward graphical interface.

The main focus of this study is to create an online bioinformatics tool that can predict the tertiary structure of proteins. To achieve this goal, the homology modeling method and Modeller are utilized. The tool has been designed to offer a simple graphical interface, eliminating the need for researchers to have experience with

Department of Biophysics, Faculty of Biological Sciences, Tarbiat Modares University, 1411713116 Terhan, Iran.  
 email: sh.arab@modares.ac.ir

Python scripts or the command line. This simplifies the modeling process, allowing researchers to complete all necessary steps easily.

## Results

We utilized the lactate dehydrogenase enzyme sequence as a sample for modeling purposes. Upon completing the initial modeling process, we obtained a diagram named the DOPE (Discrete Optimized Protein Energy)<sup>8</sup> score diagram (Fig. 4), which displays the quality of the constructed model per residue in comparison with the target protein. Access to all data generated by the Modeller program is also provided through the provided link (Fig. 1).

The NGL viewer displays the best model on the right-hand side, while a list of available and downloadable models is provided on the left-hand side. The model with the lowest DOPE Score is also identified as the BEST MODEL. In addition to the built models, alignment files in 'pir' and 'ali' formats, as well as log files, can be downloaded individually or as a compressed file. To ensure the accuracy of the protein model, it is strongly advised to thoroughly review the sequence alignment. As such, the report includes an illustration of the sequence alignment at the end. All of these modeling tasks can be performed in just a few minutes, without requiring any programming or Python knowledge, or the installation of necessary software. The successful integration of the Modeller academic license into EasyModel's server has enabled the user-friendly graphical interface for Modeller to be up and running. To make use of EasyModel, ensure you possess a valid Modeller academic license key, which can be obtained by visiting the website of Andrej Sali's lab at <https://salilab.org>.

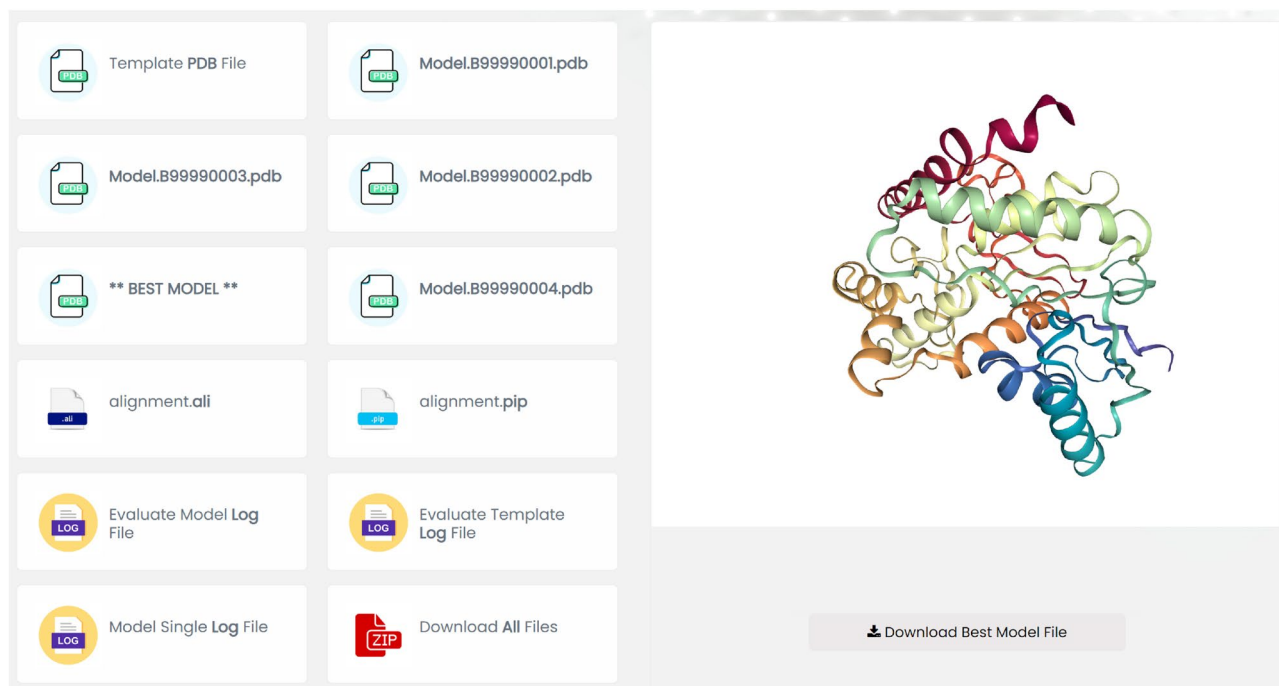
## Discussion

In addition to the results obtained from this study, exploring other capabilities of modeller, such as analyzing protein-ligand interactions and generating independent diagrams of protein profiles, can further aid scientists in predicting the tertiary structure of sequences.

## Methods

The primary objective of this study has been to create a frontend graphical user interface called "EasyModel" for Modeller, which has been built using PHP and Python. It is worth mentioning that the frontend of the tool has been developed using HTML5, Bootstrap, CSS, and JavaScript, furthermore, NGL viewer<sup>9</sup> is used to display molecules. This web-based interface has been designed to assist users who lack programming skills, knowledge of the command line, and proficiency in the Python programming language. With EasyModel, users can easily model, evaluate, analyze, and optimize protein models. The interface is capable of identifying the target sequence and template(s) structure, predicting the tertiary structure of the target sequence, and presenting the results in the form of graphs, log files, alignments, and data in constructed protein files. The screenshot of the EasyModel interface is shown in Fig. 2.

EasyModel offers three different sections to predict the tertiary structures of proteins. Based on their specific requirements, users can select one of these sections. The three sections available are: (1) basic modeling, (2) Advanced modeling-loop refining, (3) Advanced modeling-multiple templates.



**Figure 1.** Show results after basic modeling and advanced modelling (loop refinement and multiple templates) in EasyModel.

**EasyModel**  
advanced protein modeling

Home Basic Modelling Loop Refining Modelling Multiple Templates Modelling Contact Us

EasyModel: Web-based bioinformatics interface for protein modeling based on MODELLER

In this interface, we use the homology modeling method. The basis of prediction by homologous modeling is that the protein sequence is similar to one or more proteins with a known structure.

**Basic Modeling**

*Model a sequence with high identity to a template.*

This section use a simple case where the template selection and target-template alignments are not a problem. This means that there is no gap between the amino acids.

**Advanced Modeling - Loop Refining**

*The loop optimization relies on a scoring function and optimization schedule adapted for loop modeling.*

Select any set of atoms of your choosing model and generating a number of loop models.

**Advanced Modeling - Multiple Templates**

*Model a sequence with high identity to multiple templates.*

Build a model using information from multiple templates, provide an alignment between all of the templates and your target sequence and list all of the templates in the knowns argument.

**Figure 2.** Screenshot of the EasyModel that consists of basic modelling, advanced modelling (loop refinement), and advanced modelling (multiple templates).

### Basic modelling

In cases where template selection and target-template alignments pose no challenge, basic modeling can be readily employed as a straightforward approach. Detailed step-by-step instructions on how to perform these steps are fully available at <https://salilab.org/modeller/tutorial/basic.html>.

In the basic modeling section of the designed tool, users are able to import their desired protein PDB file. Here, the various steps of protein modeling using this tool for the lactate dehydrogenase protein sequence are demonstrated (Fig. 3, 4).

Upon basic modeling operations, EasyModel automatically generates a log file. Following the Basic Modeling step, a graph illustrating the DOPE scores of both the input template and the generated model is presented. The DOPE score is a statistical potential-based scoring function utilized to assess the quality of a protein structure model by estimating its relative energy. This score enables the comparison and ranking of different models, with lower scores indicating higher-quality protein structure models. The resulting chart can be downloaded in various file formats, including jpg, png, pdf, and others (Fig. 4).

In Fig. 4, the yellow area highlights structural gaps in the template, resulting in high DOPE scores due to the lack of a modeling template in this region. To address these regions, advanced modeling—loop modeling—will be employed to improve the model's accuracy. By clicking on the “See All Generated Files” button, users are directed to the modeling results page. More details about the results will be provided in the results section. Additionally, you can find a sample result at <http://bioinf.modares.ac.ir/software/easymodel/basic-modelling/sample>.

### Advanced modelling—multiple templates

In protein modeling, multiple templates are employed when there is no single template that closely matches the target protein sequence. In such scenarios, various templates can be utilized to model different segments of the target protein sequence, and subsequently, these models can be combined to generate a final model. The utilization of multiple templates can enhance the accuracy of the resulting model, particularly when the target protein possesses distinct structural features that are absent in any of the individual templates.

As it is shown in Fig. 5, EasyModel then automatically starts to perform modeling operations with several models. After the completion of the modeling process, the results page will be presented, providing detailed information in the results section. You can also view a sample result at <http://bioinf.modares.ac.ir/software/easymodel/advanced-modelling/multiple-templates/sample>.

### Basic Modelling

Modeller License Key

Model Sequence

```

CEIAMLHAKNLKPEN
FSSLSMLDQNRAYEVASKLGVVDKVDVHDIIVWGNHGESMVADLTQATFTKEGKTQKV
VDVLDHDYVFDFFKKI
GHRAWDILEHRGFTSAASPTKAAIQHMKAWLFGTAPGEVLSMGIPVPEGNPYGIKPGV
VFSFPCNVKKEGKIHVV
EGFKVNDWLREKLDFTKDLFHEKEIALNHLAQGG

```

Template pdb file

1b8p

Chain

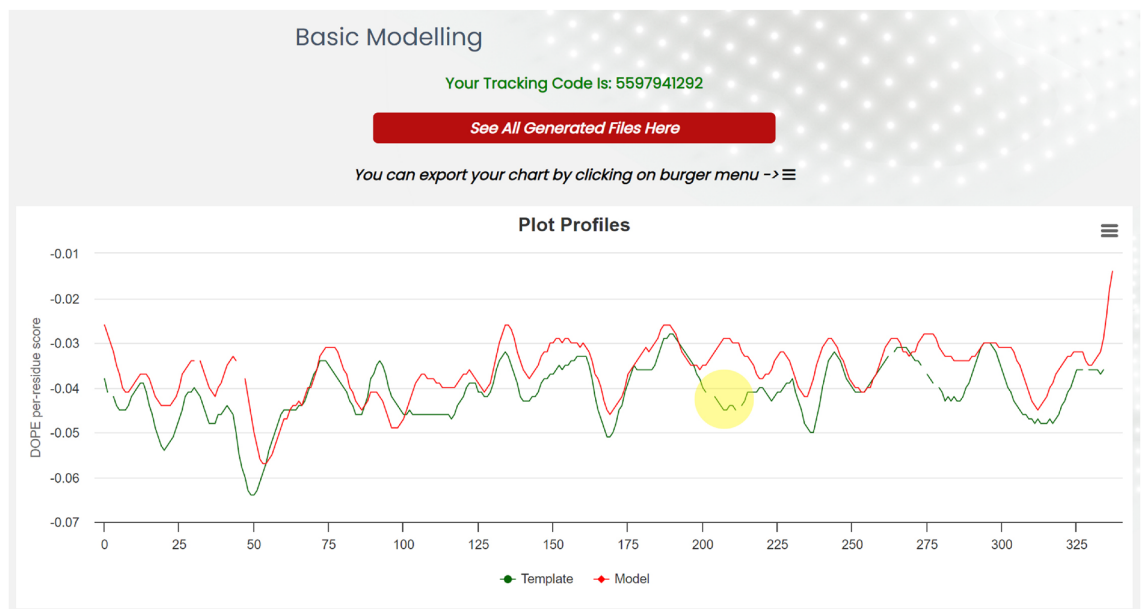
A

Model Numbers

5

Reset Submit

**Figure 3.** Screenshot of the basic modeling section in EasyModel.



**Figure 4.** The DOPE scores for individual residues are depicted with green and red lines representing the scores of the template and model structures, respectively. The yellow area indicates structural gaps in the template that have resulted in a poor DOPE score due to the absence of a modeling template in this region. These regions need to improve their modeling.

### Advanced Modelling - Multiple Templates

Modeller License Key

Your Model Sequence

```

CEIAMLHAKNLKPEN
FSSLSMLDQNRAYYEVASKLGVVDKDVHDIIVWGNHGESMVADLTQATFTKEGKTQKV
VDVLDHDYVFDTFKKI
GHRAWDILEHRGFTSAASPTKAAIQHMKAWLFGTAPGEVLSMGIPVPEGNPYGIKPGV
VFSFPCNVKKEGKIHVV
EGFKVNDWLREKLDFTKDLFHEKEIALNHLAQGG

```

Template PDB Files

**0.5 MB**

*1b8p.pdb*

**0.4 MB**

*2mdh.pdb*

**0.2 MB**

*1b8p.pdb*

PDB Name	Chain
2mdh.pdb	B
1b8p.pdb	B
1b8p.pdb	A

Start Multiple Modelling

**Figure 5.** Advanced modeling with multiple templates in EasyModel.

### Advanced modelling—loop refining

Loop Refining improves the accuracy of a protein structure model by optimizing poorly modeled loops. This method utilizes scoring functions and optimization protocols specially designed for loop modeling. It can also enhance the accuracy of a particular range of residues in the model<sup>10</sup>. Figure 6 demonstrates how to select the chain and the range of residues within a loop that requires refining after uploading the protein structure file.

Upon completing the Loop Refining process, 10 different models and a log file of the program execution are created and saved. You can explore a sample result page at <http://bioinf.modares.ac.ir/software/easymodel/advanced-modelling/loop-refining/sample>.

Comprehensive, detailed, and step-by-step instructions for performing Modeller advanced modeling steps are fully accessible at <https://salilab.org/modeller/tutorial/advanced.html>.

Advanced Modelling - Loop refining

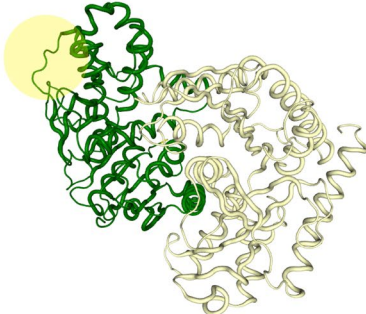
Modeller License Key

Template PDB File

0.5 MB  
lbdm.pdb

Chain      Residue Range From      Residue Range To

A      90      100



Start Loop Refining

**Figure 6.** Selection of side chain and amino acid range in loop refining in EasyModel.

### Code availability

EasyModel is available on Github, allowing users to access the source code, customize the interface to their requirements, and contribute to the development of EasyModel. Visit our GitHub repository at <https://github.com/alireza-dantism/EasyModel>.

Received: 6 June 2023; Accepted: 9 October 2023

Published online: 11 October 2023

### References

1. Krieger, E., Nabuurs, S. B. & Vriend, G. Homology modeling. *Methods Biochem. Anal.* **44**, 509–524 (2003).
2. Webb, B. & Sali, A. Comparative protein structure modeling using modeller. *Curr. Protoc. Bioinform.* <https://doi.org/10.1002/cpbi.3> (2016).
3. Lambert, C., Léonard, N., Bolle, X. D. & Depiereux, E. Espypred3d: Prediction of proteins 3d structures. *Bioinformatics* <https://doi.org/10.1093/bioinformatics/18.9.1250> (2002).
4. Söding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33**, W244–8 (2005).
5. Pieper, U. *et al.* Modbase, a database of annotated comparative protein structure models, and associated resources. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkq1091> (2011).
6. Pettersen, E. F. *et al.* Ucsf chimerax: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82. <https://doi.org/10.1002/pro.3943> (2021).
7. Kuntal, B. K., Aparoy, P. & Reddanna, P. EasyModeller: A graphical interface to MODELLER. *BMC. Res. Notes* **3**, 226 (2010).
8. Shen, M.-Y. & Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **15**, 2507–2524 (2006).
9. Rose, A. S. *et al.* Ngl viewer: Web-based molecular graphics for large complexes. *Bioinformatics* **34**, 3755–3758. <https://doi.org/10.1093/bioinformatics/bty419> (2018).

10. Fiser, A., Do, R. K. G. & Šali, A. Modeling of loops in protein structures. *Protein Sci.* <https://doi.org/10.1110/ps.9.9.1753> (2000).

### Acknowledgements

The authors would like to express their gratitude to all the members of the bioinformatics lab at Tarbiat Modares University (TMU) for their valuable contributions in the form of discussions and critical feedback on the manuscript.

### Author contributions

S.S.A. led the project and provided a comprehensive review of the manuscript. A.D. wrote the main manuscript and played a key role in designing, implementing, and developing the EasyModel web tool.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to S.S.A.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023