

WEBLEM 3

Introduction to tertiary structure prediction

Proteins are involved in many cell activities (e.g., molecular transport, mechanical functions, message exchange) thus **knowing their 3D structure is crucial** in order to understand their function. **Protein tertiary structure prediction** is a research field which aims to **create models and software tools** able to predict the **three-dimensional shape of protein molecules** by describing the spatial disposition of each of its atoms starting from the sequence of its amino acids. There exist exact methods to **resolve the molecular structure with high precision**, but they are both time and resource consuming. **Computational based software techniques** can predict the tertiary structure of a protein with **acceptable precision** for many applications with high efficiency allowing for **genome-wide investigations**, otherwise not feasible.

Having a **computer-generated three-dimensional model** of a protein of interest has many ramifications, assuming it is reasonably correct. It may be of use for the **rational design of biochemical experiments**, such as **site-directed mutagenesis, protein stability, or functional analysis**. In addition to serving as a **theoretical guide to design experiments** for protein characterization, the model can help to **rationalize the experimental results** obtained with the protein of interest. In short, the modelling study helps to advance our **understanding of protein functions**.

METHODS:

There are **three computational approaches** to protein three-dimensional structural modelling and prediction. They are **homology modelling, threading, and ab initio prediction**.

HOMOLOGY MODELLING:

As the name suggests, **homology modelling** predicts protein structures based on **sequence homology** with known structures. It is also known as comparative modelling. The principle behind it is that if two proteins share a **high enough sequence similarity**, they are likely to have very **similar three-dimensional structures**. If one of the protein sequences has a **known structure**, then the structure **can be copied to the unknown protein** with a high degree of confidence. Homology modelling produces an **all-atom model** based on **alignment with template proteins**.

The overall homology modelling procedure consists of six steps.

1. **Template Selection** which involves identification of homologous sequences in the protein structure database to be used as templates for modelling
2. **Alignment** of target and template sequences.
3. **Building a framework structure** for the target protein consisting of main chain atoms.
4. **Refine and optimize** the entire model according to energy criteria.
5. **Evaluation** of the overall quality of the model obtained.

A number of **comprehensive modelling programs** are able to perform the complete procedure of homology modelling in an automated fashion. The **automation requires assembling a pipeline** that includes target selection, alignment, model generation, and model evaluation.

MODELLER:

MODELLER is a computer program for **comparative protein structure modelling**. In the simplest case, the input is an **alignment of a sequence** to be modelled with the **template structures**, the atomic coordinates of the templates, and a simple script file. **MODELLER** then automatically **calculates a model** containing all **non-hydrogen atoms**, within minutes on a modern PC and with no user intervention. Apart from model building, **MODELLER** can perform additional auxiliary tasks, including **fold assignment, alignment of two**

protein sequences or their profiles, **multiple alignment** of protein sequences and/or structures, **calculation of phylogenetic trees**, and **de novo modelling of loops** in protein structures.

THREADING AND FOLD RECOGNITION:

There are only **small number of protein folds available** (<1,000), compared to millions of protein sequences. This means that protein structures tend to be **more conserved** than protein sequences. Consequently, many proteins can share a **similar fold** even in the absence of **sequence similarities**. This allowed the development of computational methods to predict protein structures **beyond sequence similarities**. To determine whether a **protein sequence adopts** a known **three-dimensional structure** fold relies on **threading and fold recognition** methods. By definition, threading or structural fold recognition predicts the **structural fold** of an **unknown protein sequence** by fitting the sequence into a **structural database** and selecting the **best-fitting fold**. The comparison emphasizes matching of **secondary structures**, which are most evolutionarily conserved. Therefore, this approach can **identify structurally similar proteins** even without detectable sequence similarity.

The algorithms can be classified into two categories, **pairwise energy based** and **profile based**. The pairwise energy-based method was originally referred to as **threading** and the profile-based method was originally defined as **fold recognition**. However, the two terms are now often used **interchangeably without distinction** in the literature. A number of threading and fold recognition programs are available using **either or both prediction strategies**.

I-TASSER:

I-TASSER server is an on-line platform that implements the **I-TASSER based algorithms** for protein structure and function predictions. It allows academic users to **automatically generate high-quality model predictions** of 3D structure and **biological function** of protein molecules from their amino acid sequences. When user submits an amino acid sequence, the server **first tries to retrieve template proteins of similar folds** (or super-secondary structures) from the PDB library by LOMETS, a locally installed meta-threading approach.

In the **second step**, the continuous fragments excised from the PDB templates are reassembled into full-length models by **replica-exchange Monte Carlo simulations** with the threading unaligned regions (mainly loops) built by **ab initio modelling**. In cases where no appropriate template is identified by **LOMETs**, **I-TASSER** will build the whole structures by ab initio modelling. The low free-energy states are identified by **SPICKER** through clustering the simulation decoys.

In the **third step**, the fragment assembly simulation is performed again starting from the **SPICKER cluster centroids**, where the spatial restraints collected from **both the LOMETS templates and the PDB structures** by TM-align are used to **guide the simulations**. The purpose of the **second iteration** is to **remove the steric clash** as well as to **refine the global topology** of the cluster centroids. The decoys generated in the second simulations are then clustered and the lowest energy structures are selected. The **final full-atomic models** are obtained by **REMO** which builds the atomic details from the selected I-TASSER decoys through the optimization of the **hydrogen-bonding network**.

For **predicting the biological function** of the protein, the I-TASSER server matches the **predicted 3D models** to the **proteins in 3 independent libraries** which consist of proteins of known **enzyme classification (EC) number**, **gene ontology (GO) vocabulary**, and **ligand-binding sites**. The final results of function predictions are deduced from the consensus of **top structural matches** with the function scores calculated based on the confidence score of the I-TASSER structural models, the **structural similarity** between model and templates as **evaluated by TM-score**, and the sequence identity in the structurally aligned regions.

1. What is C-score?

C-score is a **confidence score** for estimating the quality of predicted models by I-TASSER. It is calculated based on the **significance of threading template alignments** and the **convergence parameters** of the

structure assembly simulations. C-score is typically in the **range of [-5,2]**, where a C-score of **higher value** signifies a model with a **high confidence** and vice-versa.

2. What is TM-score?

TM-score is a recently proposed scale for measuring the **structural similarity between two structures**. The purpose of proposing TM-score is to **solve** the problem of **RMSD** which is sensitive to the local error. Because RMSD is an **average distance** of all residue pairs in **two structures**, a **local** error (e.g., a misorientation of the tail) will arise a **big RMSD value** although the global topology is correct. In TM-score, however, the **small distance** is **weighted stronger** than the **big distance** which makes the score **insensitive to the local modelling error**. A **TM-score >0.5** indicates a model of correct topology and a **TM-score <0.17** means a random similarity. These cut-off does not depend on the protein length.

3. What is difference and relationship between C-score and TM-score?

TM-score (or RMSD) is a known **standard** for measuring **structural similarity** between two structures which are usually used to measure the **accuracy of structure modelling** when the native structure is known, while **C-score** is a metric that I-TASSER developed to **estimate the confidence of the modelling**. In case where the native structure is not known, it becomes necessary to predict the quality of the modelling prediction, i.e., what is the distance between the predicted model and the native structures? To answer this question, we tried **predicting the TM-score and RMSD** of the predicted models relative the native structures based on the **C-score**.

In a benchmark test set of 500 non-homologous proteins, we found that C-score is highly correlated with TM-score and RMSD. Correlation coefficient of C-score of the first model with TM-score to the native structure is 0.91, while the coefficient of C-score with RMSD to the native structure is 0.75. These data lay the base for the reliable prediction of the TM-score and RMSD using C-score. In the output section, I-TASSER only reports the quality prediction (TM-score and RMSD) for the first model, because it was found that the correlation between C-score and TM-score is weak for lower rank models. However, the C-score is listed for all models just for a reference.

AB INITIO PROTEIN STRUCTURAL PREDICTION

The limited knowledge of protein folding forms the basis of ab initio prediction. As the name suggests, the ab initio prediction method attempts to produce all-atom protein models based on sequence information alone without the aid of known protein structures. The perceived advantage of this method is that predictions are not restricted by known folds and that novel protein folds can be identified. However, because the physicochemical laws governing protein folding are not yet well understood, the energy functions used in the ab initio prediction are at present rather inaccurate. The folding problem remains one of the greatest challenges in bioinformatics today.

Current ab initio algorithms are not yet able to accurately simulate the protein folding process. They work by using some type of heuristics. Because the native state of a protein structure is near energy minimum, the prediction programs are thus designed using the energy minimization principle. These algorithms search for every possible conformation to find the one with the lowest global energy. However, searching for a fold with the absolute minimum energy may not be valid in reality. This contributes to one of the fundamental flaws of this approach. In addition, searching for all possible structural conformations is not yet computationally feasible. It has been estimated that, by using one of the world's fastest supercomputers (one trillion operations per second), it takes 10 20 years to sample all possible conformations of a 40-residue protein. Therefore, some type of heuristics must be used to reduce the conformational space to be searched. Some recent ab initio methods combine fragment search and threading to yield a model of an unknown protein. The following web program is such an example using the hybrid approach.

ROBETTA:

The ROBETTA server provides automated tools for protein structure prediction and analysis. For structure prediction, sequences submitted to the server are parsed into putative domains and structural models are generated using either comparative modelling or de novo structure prediction methods. If a confident match to a protein of known structure is found using BLAST, PSI-BLAST, FFAS03 or 3D-Jury, it is used as a template for comparative modelling. If no match is found, structure predictions are made using the de novo Rosetta fragment insertion method. Experimental nuclear magnetic resonance (NMR) constraints data can also be submitted with a query sequence for RosettaNMR de novo structure determination. Other current capabilities include the prediction of the effects of mutations on protein–protein interactions using computational interface alanine scanning. The Rosetta protein design and protein–protein docking methodologies will soon be available through the server as well.

INPUT AND OUTPUT:

Registration:

Users must register (<http://robetta.bakerlab.org/register.jsp>) before submitting jobs to Robetta.

Structure prediction server:

Sequences submitted to the structure prediction server must be in one-letter amino acid format. They can either be pasted into the submission form, or uploaded from a file. Users have the option to submit a sequence for either domain identification or full structure prediction. A user also has the option to specify the PDB id and chain for comparative modeling. For RosettaNMR submissions, a user must upload experimental NMR constraints data (chemical shifts, NOE data and/ or residual dipolar couplings). The required input format for each type of data is described at http://robetta.bakerlab.org/documents/data_formats.jsp.

Results for a specific job are provided through the web interface by clicking on the job id listed in the queue table (<http://robetta.bakerlab.org/queue.jsp>). For full structure predictions, coordinates are also emailed to the user. For added insight, the following results are displayed along with the predicted models:

1. The prediction of transmembrane helices using TMHMM.
2. Low-complexity regions assigned by the program SEG
3. Coiled-coils prediction using COILS
4. The prediction of disordered regions using DISOPRED
5. Secondary structure predictions using PSIPRED, SAM-T99, Jufo and Jufo3D
6. The results listed above, domain predictions and the NR PSI-BLAST multiple sequence alignment used
7. For the last step in the domain prediction protocol condensed into an image to help corroborate the domain prediction results
8. Domain repeats prediction using REPRO predicted boundaries are given if repeats are detected
9. The top NR PSI-BLAST results and annotations for the top 20 species determined by lowest E-values.

The models for the full query are displayed as images at the bottom of the page. The coordinates for these models can be downloaded from the web site by clicking on the icons represented below each model image. Specific results are also provided for each domain by clicking on the domain number listed in the Ginzu domain prediction results table. For comparative models, the KSync alignment used for modelling is displayed. For de novo models, the Mammoth structure-model comparison results are displayed for the top 10 matches with Z-scores >4.5. The actual Mammoth structure-model alignment can be downloaded by clicking on the Z-score and viewed for further inspection using a molecular viewer such as RasMol. Users can download domain models by clicking on the icons below each domain model image.

Thus, modeller, I-TASSER and Robetta can be used to predict tertiary structures of proteins. These tools give faster results than x-ray or NMR techniques and can be used by researchers to understand protein functions.

REFERENCES:

1. Xiong, J. (2008). Tertiary structure prediction. Essential bioinformatics. Cambridge: Cambridge University Press. 214-228.
2. Tradigo, Giuseppe (2018). Reference Module in Life Sciences || Algorithms for Structure Comparison and Analysis: Prediction of Tertiary Structures of Proteins. , (), -. doi:10.1016/B978-0-12-809633-8.20483-4
3. Bateman, Alex; Pearson, William R.; Stein, Lincoln D.; Stormo, Gary D.; Yates, John R. (2002). Current Protocols in Bioinformatics || Comparative Protein Structure Modeling Using MODELLER. , (), 5.6.1–5.6.37. doi:10.1002/cpbi.3
4. Tutorial. (n.d.). Salilab.org. Retrieved March 8, 2022, from <https://salilab.org/modeller/tutorial/basic.html>
5. I-TASSER server for protein structure and function prediction. (n.d.). Zhanggroup.org. Retrieved March 8, 2022, from <https://zhanggroup.org/I-TASSER/about.html>
6. Kim, D. E.; Chivian, D.; Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. , 32(0), 0–0. doi:10.1093/nar/gkh468

WEBLEM 3a

MODELLER

([URL:https://salilab.org/modeller/](https://salilab.org/modeller/))

AIM:

To perform tertiary structure prediction by comparative Modelling/Homology Modelling method using Modeller for query Rhodopsin

INTRODUCTION:

Rhodopsin, also called visual purple, pigment-containing sensory protein that converts light into an electrical signal. Rhodopsin is found in a wide range of organisms, from vertebrates to bacteria. In many seeing animals, including humans, it is required for vision in dim light and is located in the retina of the eye—specifically, within the tightly packed disks that make up the outer segment of the retina's photoreceptive rod cells, which are specially adapted for vision under low-light conditions.

MODELLER is a computer program for **comparative protein structure modelling**. In the simplest case, the input is an **alignment of a sequence** to be modelled with the **template structures**, the atomic coordinates of the templates, and a simple script file. **MODELLER** then automatically **calculates a model** containing all **non-hydrogen atoms**, within minutes on a modern PC and with no user intervention. Apart from model building, **MODELLER** can perform additional auxiliary tasks, including **fold assignment**, **alignment of two protein sequences** or their profiles, **multiple alignment** of protein sequences and/or structures, **calculation of phylogenetic trees**, and **de novo modelling of loops** in protein structures.

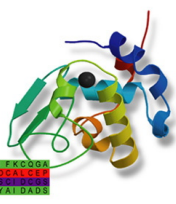
METHODOLOGY:

1. Install modeller. (URL: <https://salilab.prg/modeller>)
2. Retrieve FASTA sequence for enzyme rhodopsin
3. Follow the steps given in the tutorial section.
4. Run scripts for searching for structures related to query, selecting template target-template alignment and model building/
5. Observe and interpret the results.

OBSERVATION:

Modeller

Program for Comparative Protein Structure Modelling by Satisfaction of Spatial Restraints



```

M L V G S M P R R D G M E R < D L L K A N V K I F K Q G G A
K E V C P Y Q D F Y E G P N F L V I H P D E C I D C A L C E F
K A G K E S P R N F I E D S - R D M L A R S S I S S D S R
E - L A C G A C R P E G F V N I L Q G S - E Y A I D A D S

```

About MODELLER

MODELLER is used for homology or comparative modeling of protein three-dimensional structures (1,2). The user provides an alignment of a sequence to be modeled with known related structures and MODELLER automatically calculates a model containing all non-hydrogen atoms. MODELLER implements comparative protein structure modeling by satisfaction of spatial restraints (3,4), and can perform many additional tasks, including de novo modeling of loops in protein structures, optimization of various models of protein structure with respect to a flexibly defined objective function, multiple alignment of protein sequences and/or structures, clustering, searching of sequence databases, comparison of protein structures, etc. MODELLER is [available for download](#) for most Unix/Linux systems, Windows, and Mac.

Several graphical interfaces to MODELLER are [commercially available](#). There are also many other [resources and people using Modeller](#) in graphical or web interfaces or other frameworks.

1. B. Webb, A. Sali. Comparative Protein Structure Modeling Using Modeller. Current Protocols In Bioinformatics 54, John Wiley & Sons, Inc., 5.6.1-5.6.37, 2016.
2. M.A. Mari-Renom, A. Stuart, A. Fiser, R. Sánchez, F. Melo, A. Sali. Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 29, 291-325, 2000.
3. A. Sali & T.L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779-815, 1993.
4. A. Fiser, R.K. Do, & A. Sali. Modeling of loops in protein structures, Protein Science 9, 1753-1773, 2000.

The current release of Modeller is **10.2**, which was released on Nov 30th, 2021. Modeller is currently maintained by [Ben Webb](#).

[About MODELLER](#)

[MODELLER News](#)

[Download & Installation](#)

[Release Notes](#)

[Data file downloads](#)

[Registration](#)

[Non-academic use](#)

[Discussion Forum](#)

[Subscribe](#)

[Browse archives](#)

[Search archives](#)

[Documentation](#)

[FAQ](#)

[Tutorial](#)

[Online manual](#)

[Wiki](#)

[Developers' Pages](#)


[Contact Us](#)

MODELLER (copyright © 1989-2022 Andrej Sali) is maintained by [Ben Webb](#) at the Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, and California Institute for Quantitative Biomedical Research, Mission Bay Byers Hall, University of California San Francisco, San Francisco, CA 94143, USA. Any selling or distribution of the program or its parts, original or modified, is prohibited without a written permission from Andrej Sali. This file last modified: Mon Nov 29 23:46:55 PST 2021.

Fig1. Homepage for Modeller

Modeller

Program for Comparative Protein Structure Modelling by Satisfaction of Spatial Restraints



```

M L V G S M P R R D G M E R < D L L K A N V K I F K Q G G A
K E V C P Y Q D F Y E G P N F L V I H P D E C I D C A L C E F
K A G K E S P R N F I E D S - R D M L A R S S I S S D S R
E - L A C G A C R P E G F V N I L Q G S - E Y A I D A D S

```

Download & Installation

MODELLER is available free of charge to academic non-profit institutions; you will, however, need to [register for a license](#) in order to use the software. It is also [available through BIOVIA](#) for government research labs and commercial entities.

Modeller 10.2, released Nov. 30th, 2021

To install MODELLER on this machine, we recommend the **Windows** package.

Anaconda Python ("conda")

Windows (32-bit) [\[GPG signature\]](#) [Installation guide](#)

Windows (64-bit) [\[GPG signature\]](#) [Installation guide](#)

Mac (Intel or Apple Silicon) [\[GPG signature\]](#) [Installation guide](#)

A Homebrew package is also available for both Intel and Apple Silicon (M1)

Linux (32-bit RPM) [Installation guide](#)

Linux (64-bit x86_64 RPM) [Installation guide](#)

Linux (64-bit ARM RPM) [Installation guide](#)

Linux (32-bit Debian/Ubuntu package) [\[GPG signature\]](#) [Installation guide](#)

Linux (64-bit x86_64 Debian/Ubuntu package) [\[GPG signature\]](#) [Installation guide](#)

Linux (32-bit ARM Raspberry Pi OS/Ubuntu package) [\[GPG signature\]](#) [Installation guide](#)

Linux (64-bit ARM Debian/Ubuntu package) [\[GPG signature\]](#) [Installation guide](#)

Generic Unix tarball [\[GPG signature\]](#) [Installation guide](#)

GPG signatures are provided for the security conscious, so that you can verify that the files have not been tampered with. (No signatures are given for the RPM packages, because the signatures are already included in those packages.) In order to verify them, you will need [this GPG key](#) and a copy of GNU Privacy Guard (GPG).

[Supplemental data file downloads](#)

[Old \(and unsupported\) Modeller releases](#)

[About MODELLER](#)

[MODELLER News](#)

[Download & Installation](#)

[Release Notes](#)

[Data file downloads](#)

[Registration](#)

[Non-academic use](#)

[Discussion Forum](#)

[Subscribe](#)

[Browse archives](#)

[Search archives](#)

[Documentation](#)

[FAQ](#)

[Tutorial](#)

[Online manual](#)

[Wiki](#)

[Developers' Pages](#)

[Contact Us](#)

Fig2. Page to install Modeller

UniProtKB - P02699 (OPSD_BOVIN)

Display | Help video | BLAST | Align | Format | Add to basket | History | Add a publication | Feedback

Entry
 Publications
 Feature viewer
 Feature table

Protein | **Rhodopsin**
Gene | **RHO**
Organism | *Bos taurus* (Bovine)
Status | Reviewed - Annotation score: ●●●●● - Experimental evidence at protein level¹

Function¹
 Photoreceptor required for image-forming vision at low light intensity. Required for photoreceptor cell viability after birth (By similarity).
 Light-induced isomerization of 11-cis to all-trans retinal triggers a conformational change that activates signaling via G-proteins (PubMed:10926528, PubMed:12044163, PubMed:11972040, PubMed:16908857, PubMed:16586416, PubMed:17060607, PubMed:17449675, PubMed:18818650, PubMed:21389983, PubMed:22198838, PubMed:23579341, PubMed:25205354, PubMed:27458239).
 Subsequent receptor phosphorylation mediates displacement of the bound G-protein alpha subunit by the arrestin SAG and terminates signaling (PubMed:1396673, PubMed:15111114).

Sites

Feature key	Position(s)	Description	Actions	Graphical view	Length
Site ¹	113	Plays an important role in the conformation switch to the active conformation	3 Publications		1
Metal binding ¹	201	Zinc	2 Publications		1
Metal binding ¹	279	Zinc	2 Publications		1

GO - Molecular function¹

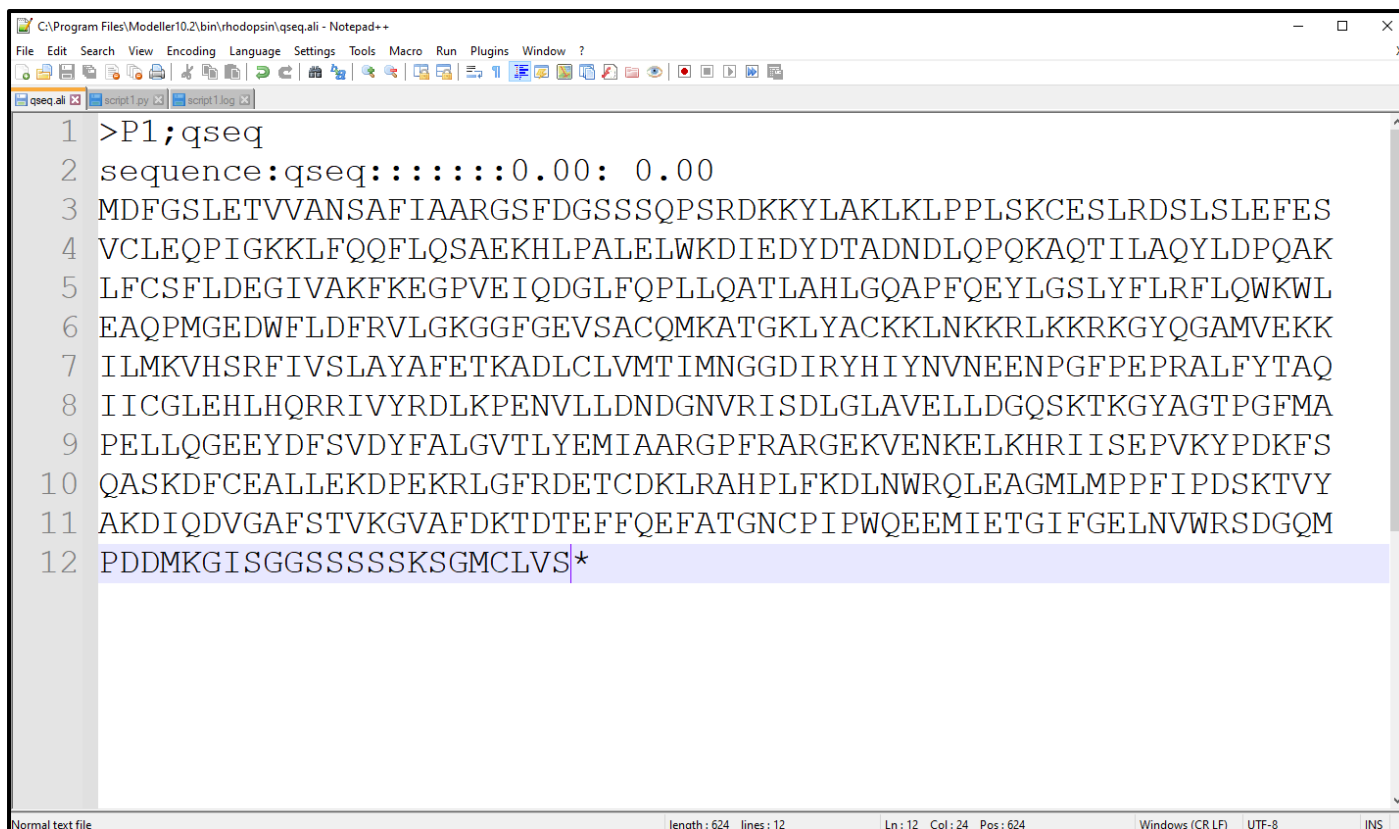
- 11-cis retinal binding Source: UniProtKB
- arrestin family protein binding Source: CAFA
- G-protein alpha-subunit binding Source: CAFA
- G-protein-coupled photoreceptor activity Source: UniProtKB
- guanyl-nucleotide exchange factor activity Source: UniProtKB
- identical protein binding Source: IntAct
- opsin binding Source: CAFA
- zinc ion binding Source: CAFA

Complete GO annotation on QuickGO ...

Fig3. Result page for Rhodopsin in UniProt database

```
>sp|Q15835|GRK1_HUMAN Rhodopsin kinase GRK1 OS=Homo sapiens OX=9606 GN=GRK1 PE=1
SV=1
MDFGSLETVVANSAFIAARGSF DGSSSQPSRDKKYLAKLKLPPLSKCESLRDSLSEFES
VCLEQPIGKKLFQQFLQSAEKHLPAL ELWKDIEDYDTADNDLQPQKAQTILAQYLDPQAK
LFCNFLDEGIVAKFKEGPVEIQDGLFQPLLQATLAHLGQAPFQEYLGSLYFLRFLQWKWL
EAQPMGEDWFLDFRVLGKGGFGEVSACQMKATGKLYACKKLNKKRLKKRKG YQGAMVEKK
ILMKVHSRFRIVSLAYAFETKADLCLVMTIMNGGDIRYHIYNVNEENPGFPEPRALFYTAQ
IICGLEHLHQRRIVYRDLKPENVLLDNDGNVRISDLGLAVELLDGQSKTKGYAGTPGFMA
PELLQGEEYDFSVDYFALGVTLYEMIAARGPFRARGEKVENKELKHRIISEPVKYPDKFS
QASKDFCEALLEKDPEKRLGFRDETCDKLR AHPLFKDLNWRQLEAGMLMPFFIPDSKTVY
AKDIQDVGA FSTVKGVAFDKTDTEFFQEFATGNCPIPWQEEMIETGIFGELNVWRSDGQM
PDDMKGISGGSSSSSKSGMCLVS
```

Fig4. FASTA sequence for Rhodopsin



```
C:\Program Files\Modeller10.2\bin\rhodopsin\qseq.ali - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
qseq.ali script1.py script1.log
1 >P1;qseq
2 sequence:qseq:::::::::0.00: 0.00
3 MDFGSLETVVANSAFIAARGSFDFGSSSQPSRDKKYLAKLKLPPLSKCESLRDSLSLEFES
4 VCLEQPIGKKLFQQFLQSAEKHLPALFLWKDIEDYDTADNDLQPQKAQTILAQYLDPOAK
5 LFCFSLDEGIVAKFKEGPVEIQDGLFQPLLQATLAHLGQAPFQEYLGSLYFLRFLQWKWL
6 EAQPMGEDWFLDFRVLGKGGFGEVSACQMKATGKLYACKKLNKKRLKKRKGYPGAMVEKK
7 ILMKVHSRFRIVSLAYAFETKADLCLVMTIMNGGDIRYHIYNVNEENPGFPEPRALFYTAQ
8 IICGLEHLHQRRIRIVYRDLKPENVLLDNDGNVRISDLGLAVELLDGQSKTKGYAGTPGFMA
9 PELLQGEEDYDFSDYFALGVTLYEMIAARGPFRARGEKVENKELKHRIISEPVKYPDKFS
10 QASKDFCEALLEKDPEKRLGFRDETCDKLRHPLFKDLNWRQLEAGMLMPPFIPDSKTVY
11 AKDIQDVGAFFSTVKGVAFDKTDTEFFQEAFATGNCPWPQEEIETGIFGELNVWRSDGQM
12 PDDMKGISGGSSSSSSKSGMCLVS*
```

Fig5. FASTA sequence in PIR format

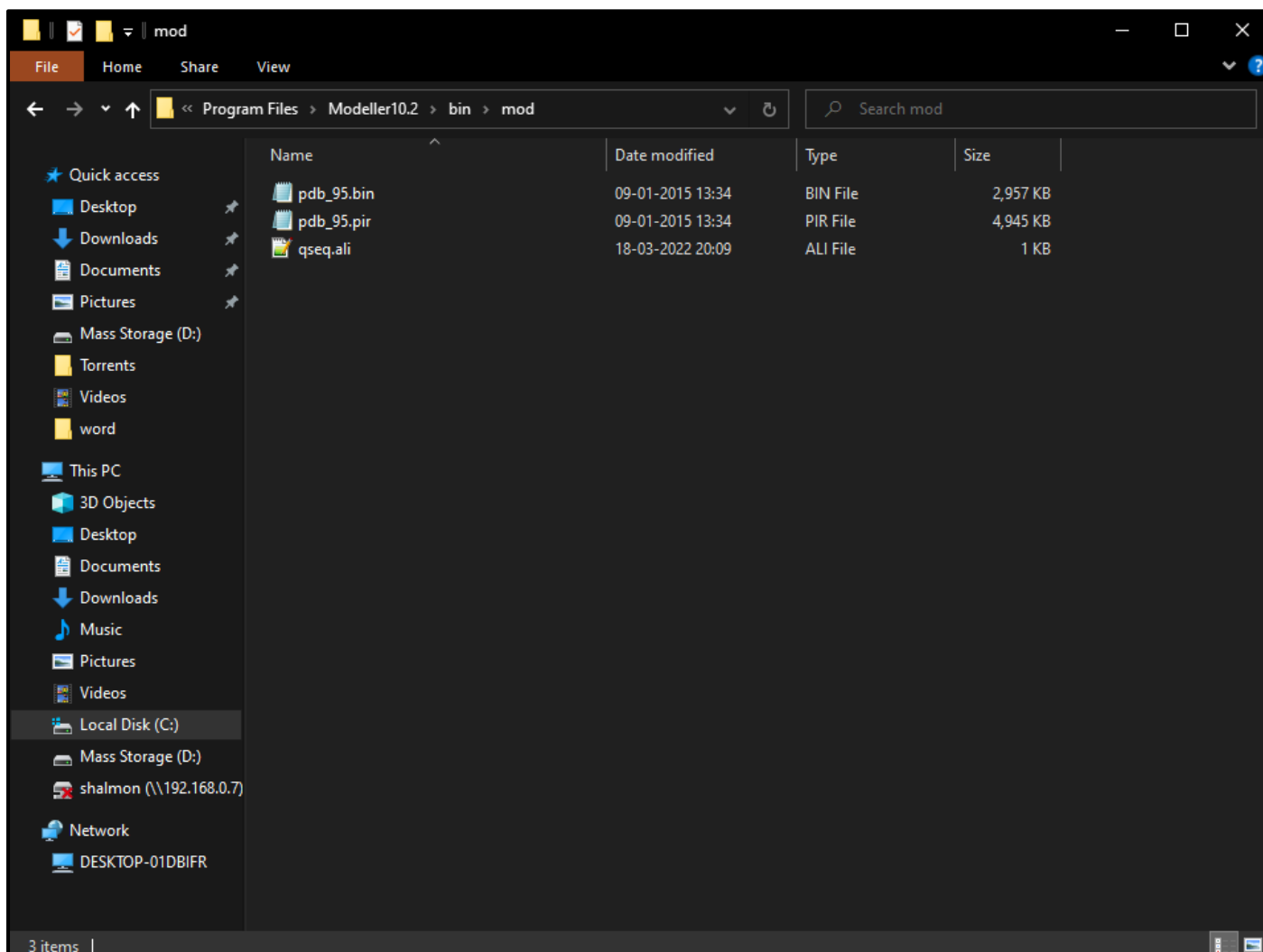


Fig6. Target sequence saved in .ali format

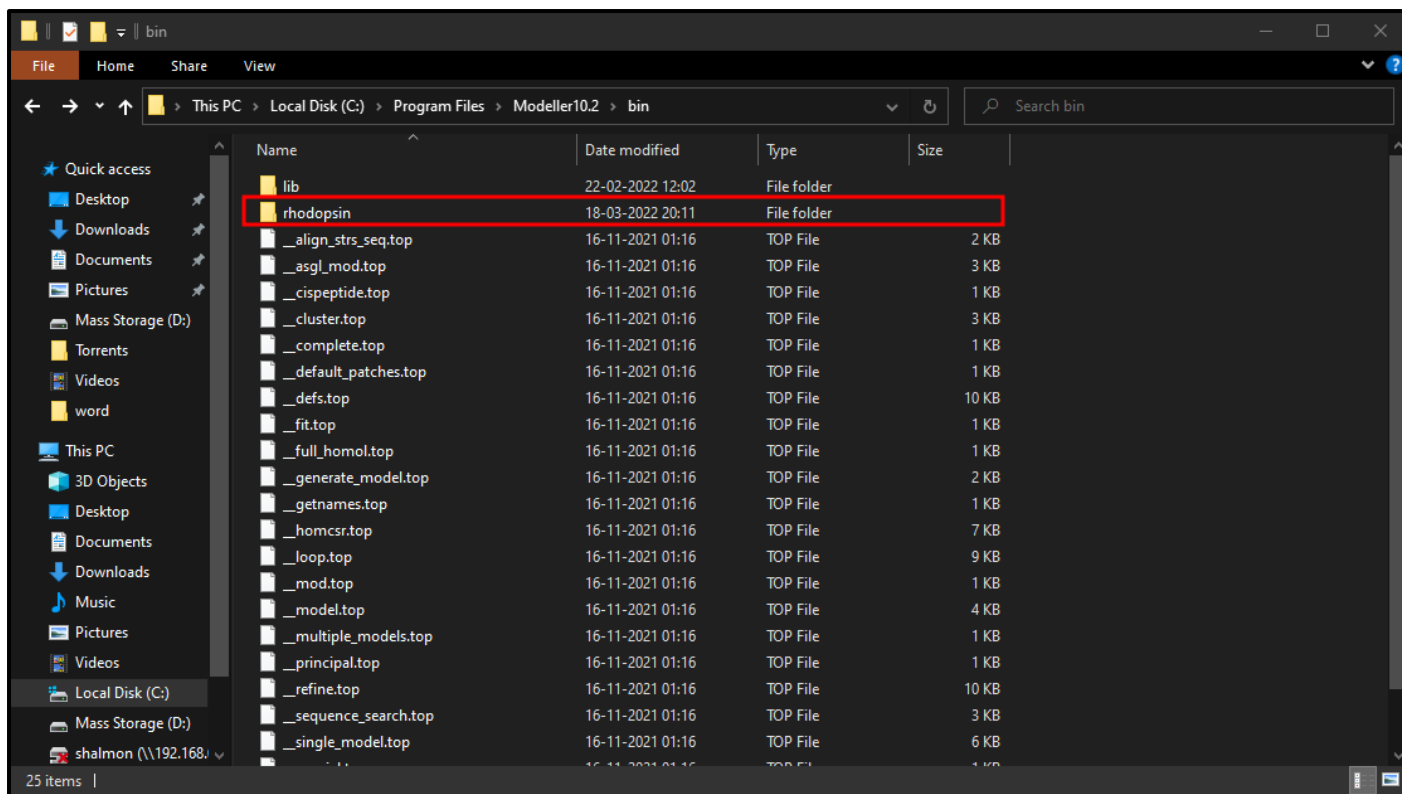


Fig7. Rhodopsin folder saved in the bin folder of modeller

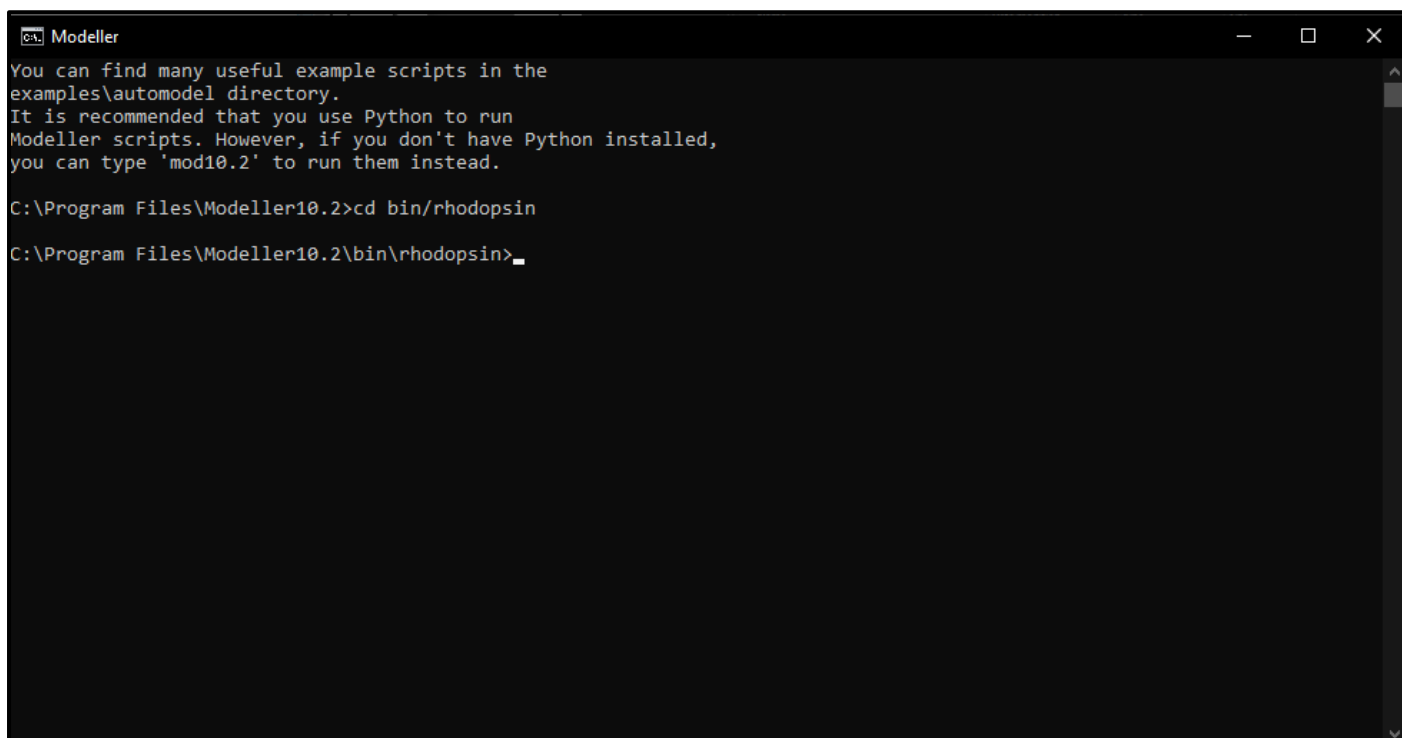
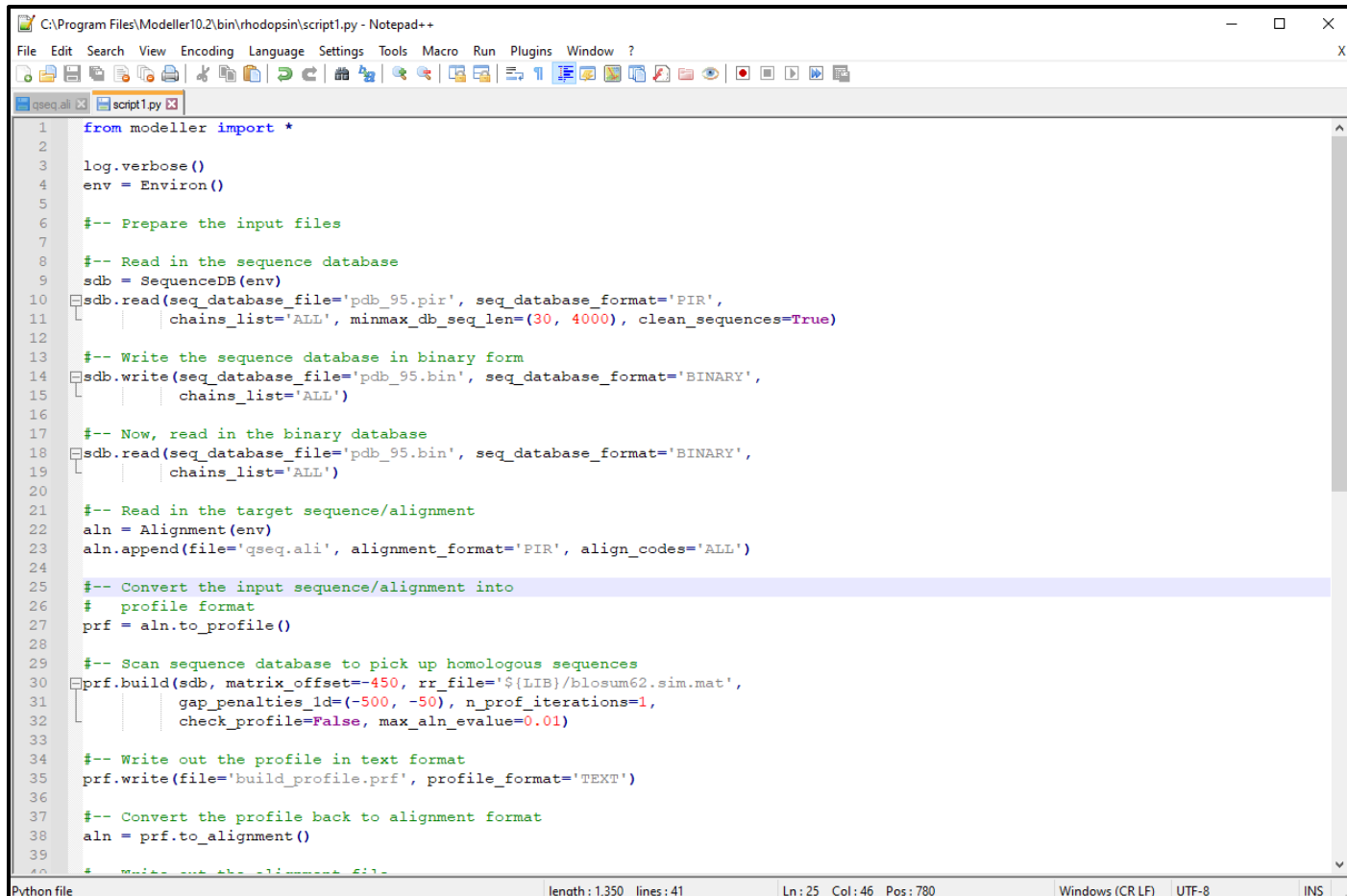
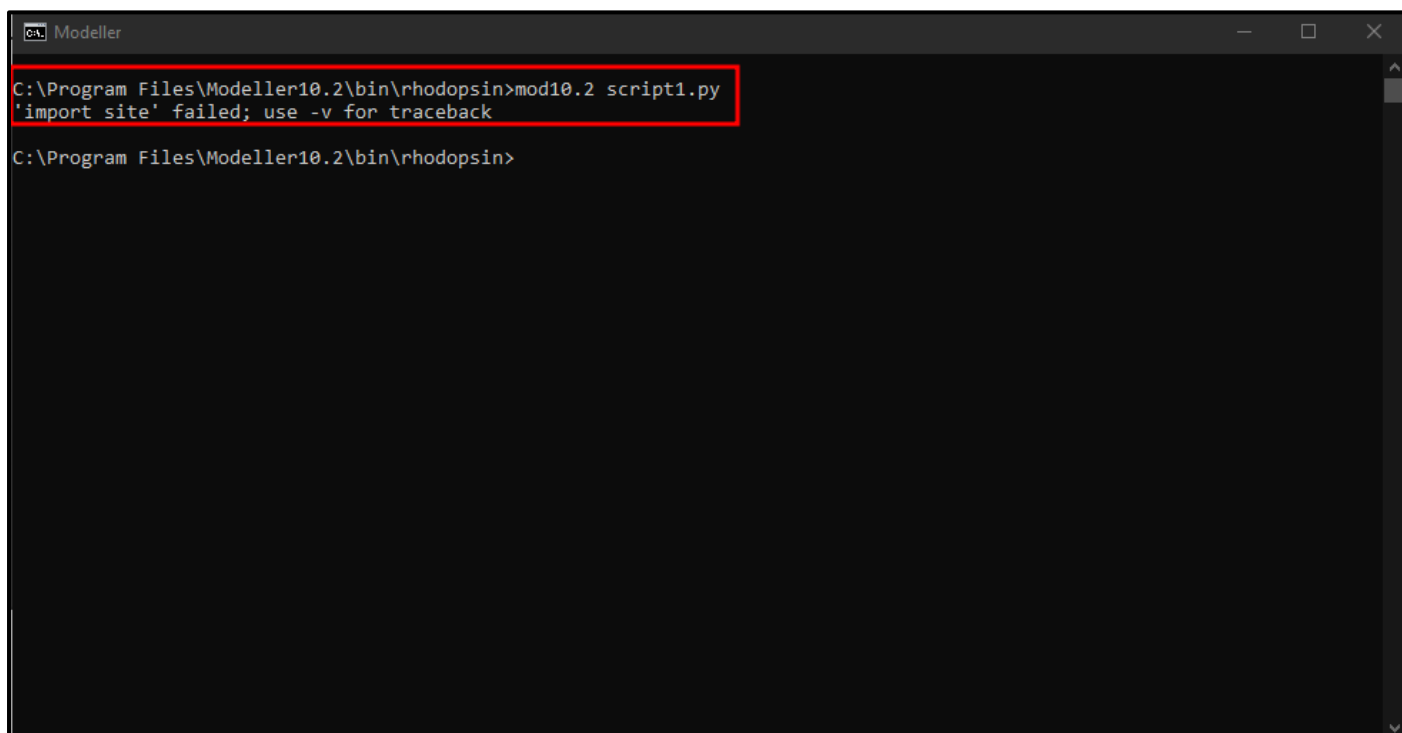


Fig8. Setting Working directory in Modeller command line



```
1 from modeller import *
2
3 log.verbose()
4 env = Environ()
5
6 #-- Prepare the input files
7
8 #-- Read in the sequence database
9 sdb = SequenceDB(env)
10 sdb.read(seq_database_file='pdb_95.pir', seq_database_format='PIR',
11         chains_list='ALL', minmax_db_seq_len=(30, 4000), clean_sequences=True)
12
13 #-- Write the sequence database in binary form
14 sdb.write(seq_database_file='pdb_95.bin', seq_database_format='BINARY',
15          chains_list='ALL')
16
17 #-- Now, read in the binary database
18 sdb.read(seq_database_file='pdb_95.bin', seq_database_format='BINARY',
19          chains_list='ALL')
20
21 #-- Read in the target sequence/alignment
22 aln = Alignment(env)
23 aln.append(file='qseq.ali', alignment_format='PIR', align_codes='ALL')
24
25 #-- Convert the input sequence/alignment into
26 # profile format
27 prf = aln.to_profile()
28
29 #-- Scan sequence database to pick up homologous sequences
30 prf.build(sdb, matrix_offset=-450, rr_file='${LIB}/blosum62.sim.mat',
31          gap_penalties_ld=(-500, -50), n_prof_iterations=1,
32          check_profile=False, max_aln_evalue=0.01)
33
34 #-- Write out the profile in text format
35 prf.write(file='build_profile.prf', profile_format='TEXT')
36
37 #-- Convert the profile back to alignment format
38 aln = prf.to_alignment()
39
40 # Write out the alignment file
```

Fig9. Python script for searching for structures relation to rhodopsin



```
C:\Program Files\Modeller10.2\bin\rhodopsin>mod10.2 script1.py
'import site' failed; use -v for traceback

C:\Program Files\Modeller10.2\bin\rhodopsin>
```

Fig10. Running script1.py

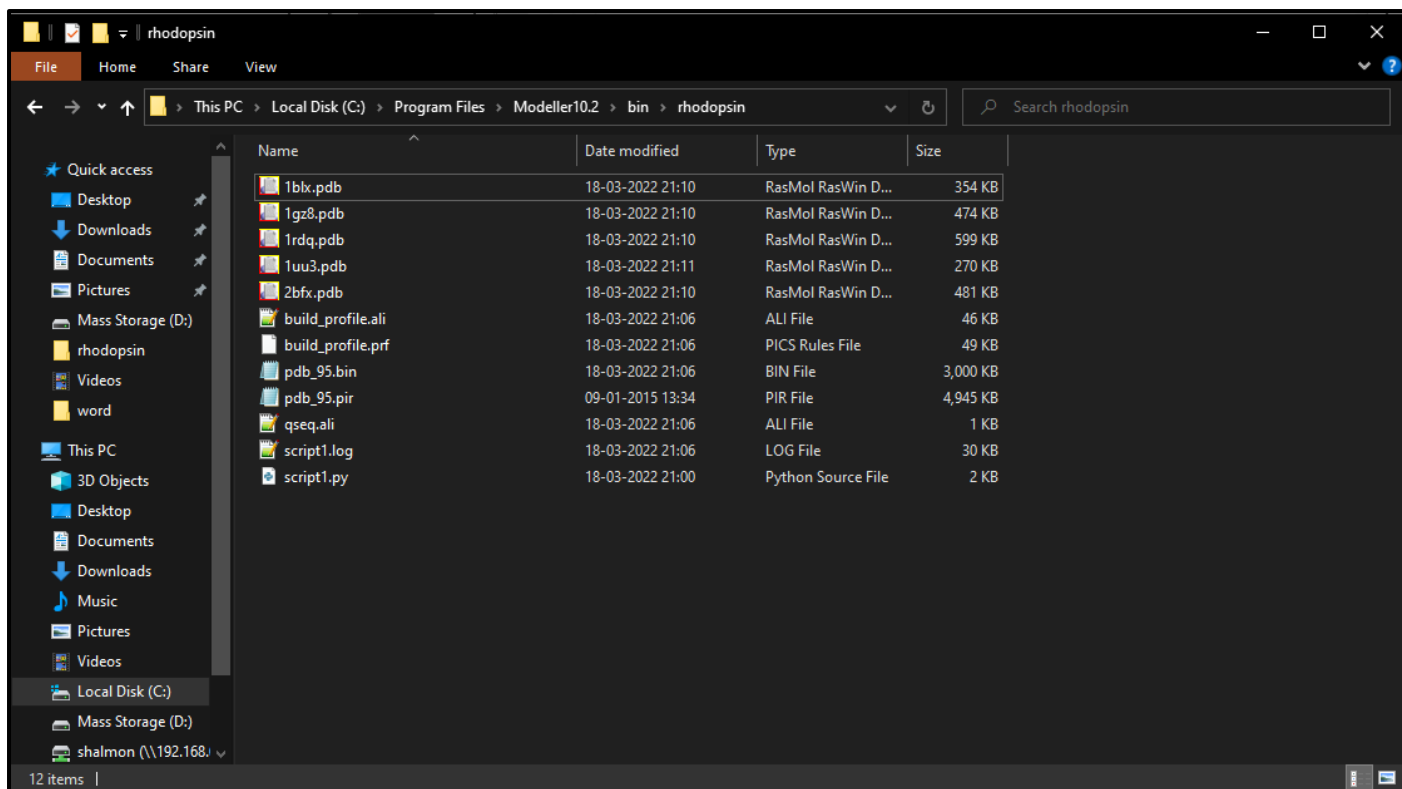


Fig12. Five structure download in PDB format

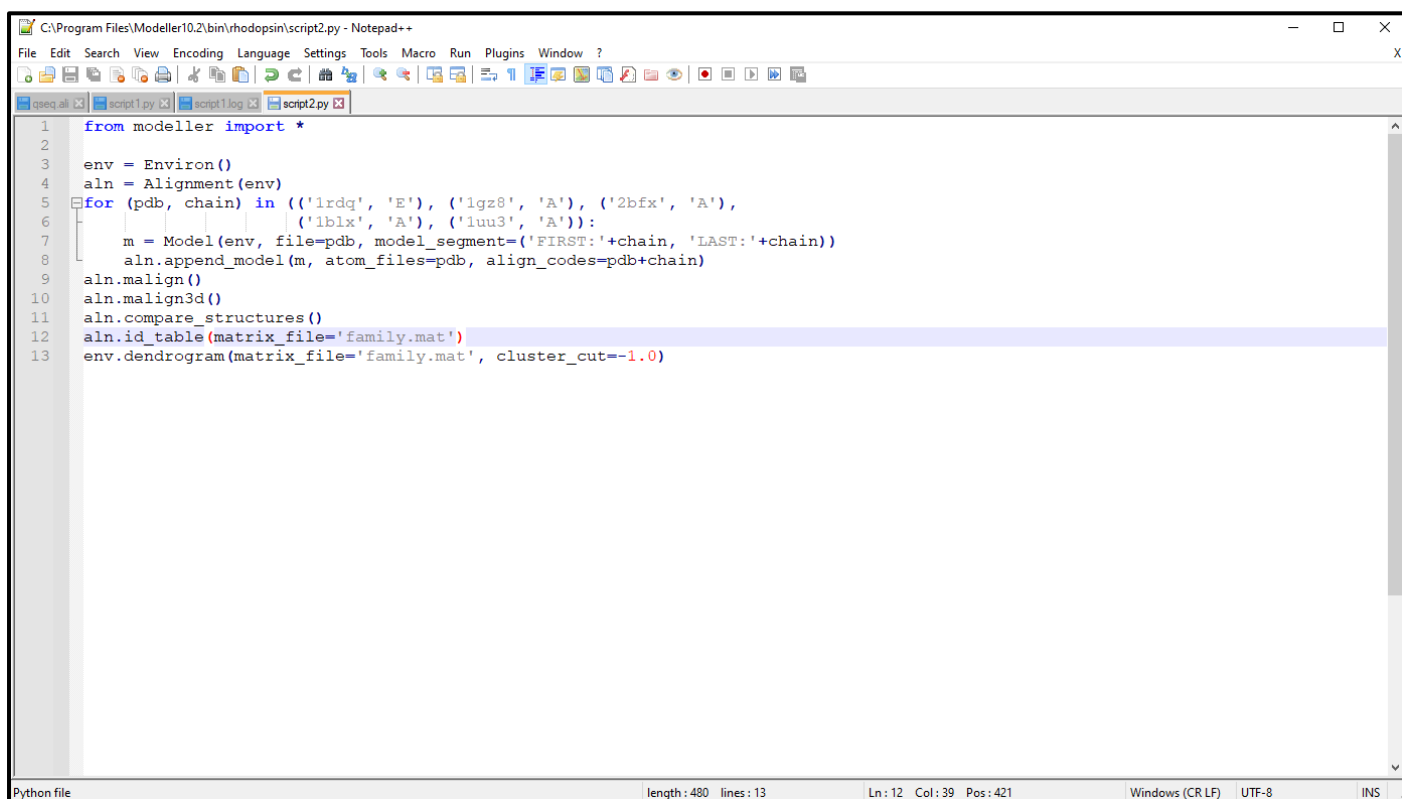


Fig13. Python script for selecting a template

```
Modeller

C:\Program Files\Modeller10.2\bin\rhodopsin>mod10.2 script2.py
'import site' failed; use -v for traceback

C:\Program Files\Modeller10.2\bin\rhodopsin>
```

Fig14. Running script2.py

```
C:\Program Files\Modeller10.2\bin\rhodopsin\script2.log - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
script1.py script1.log script2.py script2.log

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2549
2550
2551
2552
2553
2554
2555
2556
2557
2558
2559
2560
2561
2562
2563
2564
2565
2566
2567
2568
2569
2570
2571
2572
2573
2574
2575
2576
2577
2578
2579
2580
2581
2582
2583
2584
2585
2586
2587
2588
2589
2590
2591
2592
2593
2594
2595
2596
2597
2598
2599
2600
2601
2602
2603
2604
2605
2606
2607
2608
2609
2610
2611
2612
2613
2614
2615
2616
2617
2618
2619
2620
2621
2622
2623
2624
2625
2626
2627
2628
2629
2630
2631
2632
2633
2634
2635
2636
2637
2638
2639
2640
2641
2642
2643
2644
2645
2646
2647
2648
2649
2650
2651
2652
2653
2654
2655

```

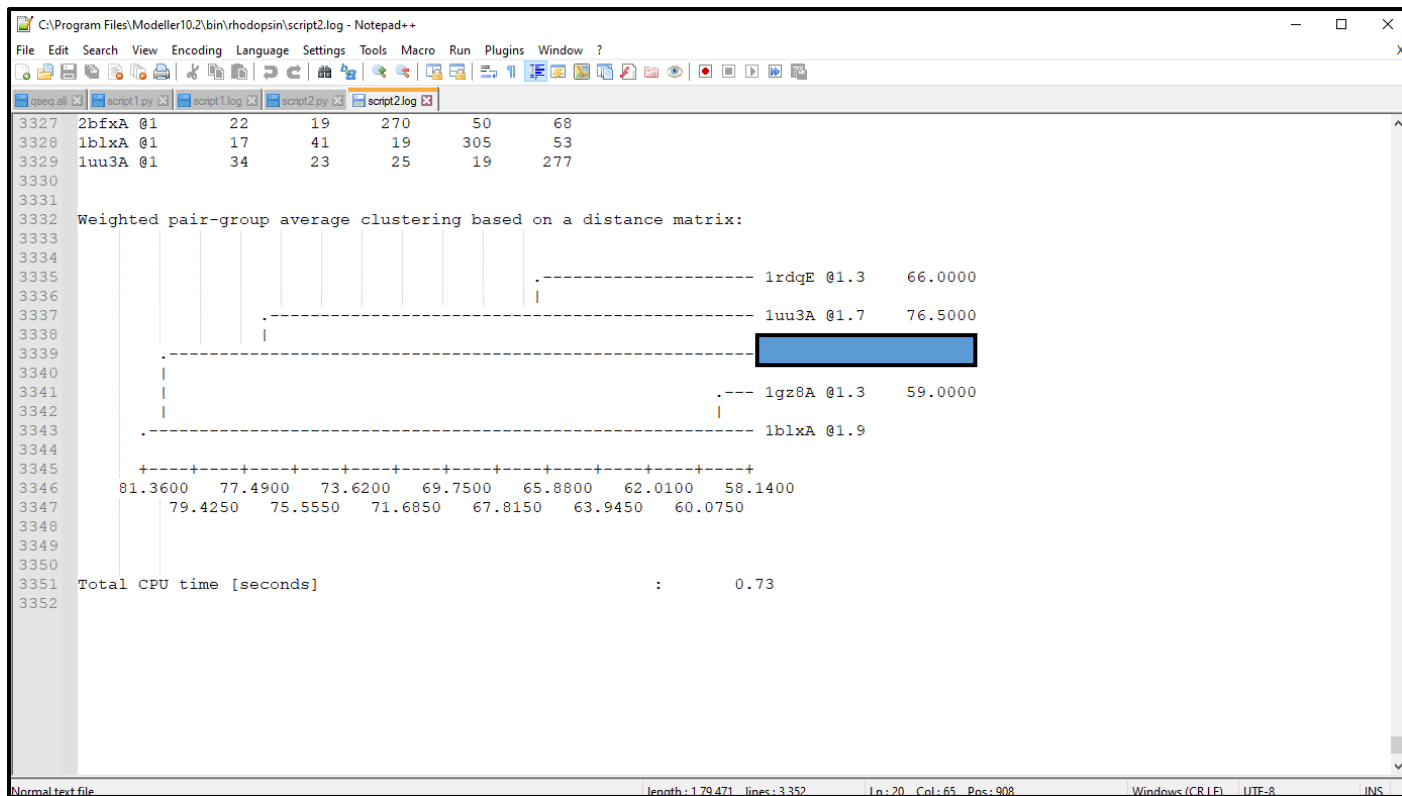


Fig15.1 Structure selected with low x-ray crystallography value and high NMR value

```

1 from modeller import *
2
3 env = Environ()
4 aln = Alignment(env)
5 mdl = Model(env, file='2bfx', model_segment=('FIRST:A','LAST:A'))
6 aln.append_model(mdl, align_codes='2bfxA', atom_files='2bfx.pdb')
7 aln.append(file='qseq.ali', align_codes='qseq')
8 aln.align2d(max_gap_length=50)
9 aln.write(file='qseq-2bfxA.ali', alignment_format='PIR')
10 aln.write(file='qseq-2bfxA.pap', alignment_format='PAP')

```

Fig16. Python script for aligning query with the template

```
Modeller

C:\Program Files\Modeller10.2\bin\rhodopsin>mod10.2 script2.py
'import site' failed; use -v for traceback

C:\Program Files\Modeller10.2\bin\rhodopsin>mod10.2 script3.py
'import site' failed; use -v for traceback

C:\Program Files\Modeller10.2\bin\rhodopsin>
```

Fig17. Running script3.py

```
C:\Program Files\Modeller10.2\bin\rhodopsin\script3.log - Notepad++

File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?

C:\Program Files\Modeller10.2\bin\rhodopsin\script3.log
script1.py script1.log script2.py script2.log script3.py script3.log

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35

MODELLER 10.2, 2021/11/15, r12267

PROTEIN STRUCTURE MODELLING BY SATISFACTION OF SPATIAL RESTRAINTS

Copyright(c) 1989-2021 Andrej Sali
All Rights Reserved

Written by A. Sali
with help from
B. Webb, M.S. Madhusudhan, M-Y. Shen, G.Q. Dong,
M.A. Marti-Renom, N. Eswar, F. Alber, M. Topf, B. Oliva,
A. Fiser, R. Sanchez, B. Yerkovich, A. Badretdinov,
F. Melo, J.P. Overington, E. Feyfant
University of California, San Francisco, USA
Rockefeller University, New York, USA
Harvard University, Cambridge, USA
Imperial Cancer Research Fund, London, UK
Birkbeck College, University of London, London, UK

Kind, OS, HostName, Kernel, Processor: 4, Windows Vista build 9200, DESKTOP-01DBIFR, SMP, unknown
Date and time of compilation : 2021/11/15 19:44:35
MODELLER executable type : x86_64-w64
Job starting time (YY/MM/DD HH:MM:SS): 2022/03/18 21:20:49

fndatmi_285W> Only 269 residues out of 270 contain atoms of type CA
(This is usually caused by non-standard residues, such
as ligands, or by PDB files with missing atoms.)
mkapsa_637W> No residue topology library is in memory.
Better radii would be used if topology.read() is called first.
iup2crm_280W> No topology library in memory or assigning a BLK residue.
Default CHARMM atom type assigned: N --> N
This message is written only for the first such atom

Normal text file length:3,105 lines:62 Ln:1 Col:1 Pos:1 Windows (CR LF) UTF-8 INS
```

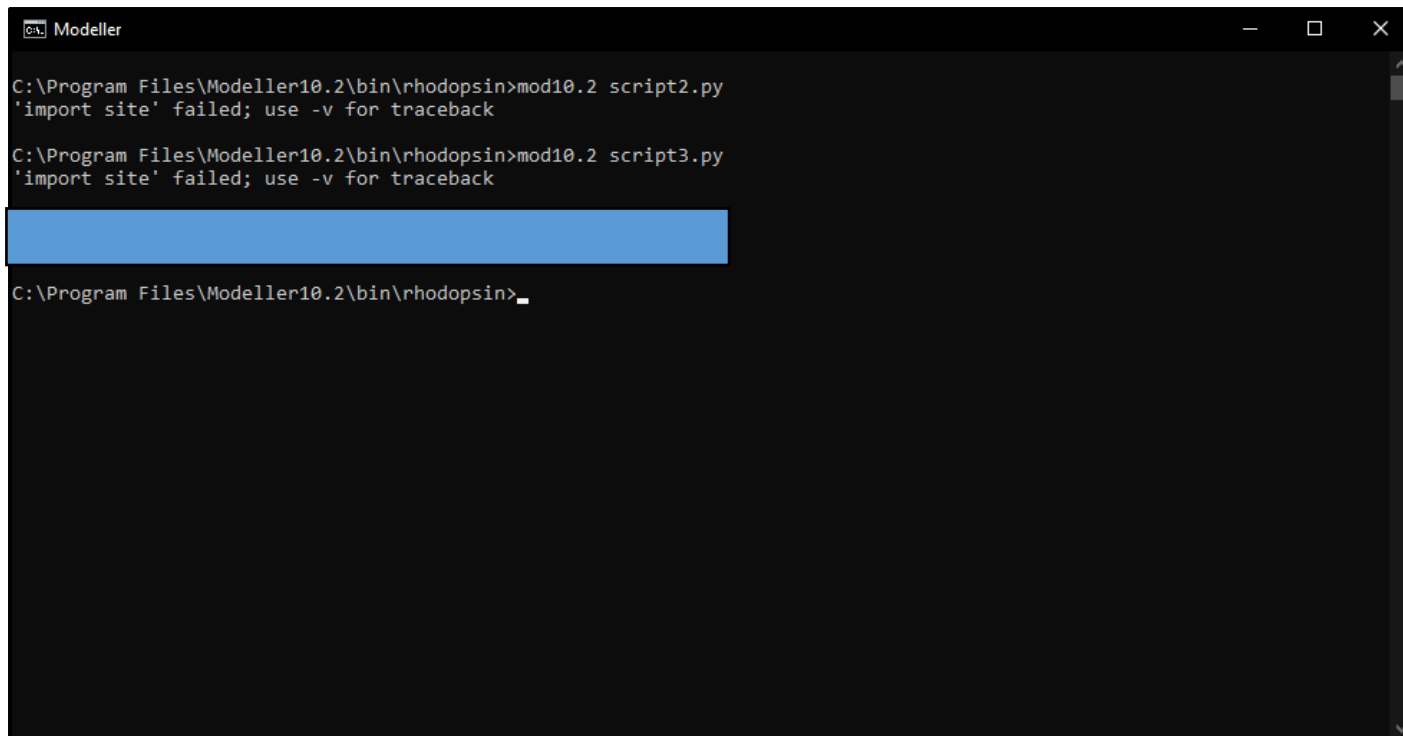
Fig18. Log file for script3


```
C:\Program Files\Modeller10.2\bin\rhodopsin\qseq-2bfxA.pap - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
qseq.all script1.py script1.log script2.py script2.log script3.py script3.log qseq-2bfxA.pap
1 _aln.pos 10 20 30 40 50 60
2 2bfxA -----R-----KFTI-----D---DFD---IGRPLG
3 qseq MDFGSLETVVANSAFIAARGSGFDGSSSQPSRDKKYLAKLKLPLSKCESLRDLSLEFESVCLEQPIG
4 _consrvd * * * * *
5
6 _aln.p 70 80 90 100 110 120 130
7 2bfxA KGKFGNVYLAREK-----QN-----KFIMALKVLFSQLEKEGVVEHQLR
8 qseq KKLQQQLQSLQSAEKHLPALELWKDIEDYDTADNDLQPKAQITLAQYLDPAKLFCFLDEGIVAKFKE
9 _consrvd * * * * *
10
11 _aln.pos 140 150 160 170 180 190 200
12 2bfxA REIEIQSHLRHPNILRMNYN-----FHD-RKRIYLM-----LEFAPR-----
13 qseq GPVEIQDGLFQPLLOATLAHLGQAPFQOEYLGSLYFLRFLQWKWLEAQPMPGEDWFLDFRVLGKGGFGEV
14 _consrvd * * * * *
15
16 _aln.pos 210 220 230 240 250 260 270
17 2bfxA -----GELY--KELQKH-----G-----RF-----
18 qseq SACQMKATGKLYACKKLNLKRLKRRKGYQGAMVEKKILMKVHSRFFVSLAYAFETKADLCLVMTIMNG
19 _consrvd * * * * *
20
21 _aln.pos 280 290 300 310 320 330 340
22 2bfxA -----DEQRSATFMEELADALHYCHERKVIHRDIKPENLLMGYKGEKLIADFGWSV
23 qseq GDIRYHIYNVNEENPGFPEPALFYTAQIICGLEHLHORRIVYRDLPENVLNDGNNVRISDLGLAV
24 _consrvd * * * * *
25
26 _aln.pos 350 360 370 380 390 400
27 2bfxA HAP---SLRRMCGTLDYLPPEMIEGKTHDEKVDLWCAGVLCYEFVLGMPPF---DSPSHTETHHRI
28 qseq ELLDQSKTKGYAGTGGFMAPELLQGEEYDFSVDFALGVTLYEMIAARGPFRARGEKVENKELKHRI
29 _consrvd * * * * *
30
31 _aln.p 410 420 430 440 450 460 470
32 2bfxA VNVDLKFPPFLSDGSKDLISLLRYHPPQRLPLK-----GVMEHPWVKANSRRVI-----PP-----
33 qseq ISEPVKYPDKFSQASKDFCEALLEKDPKRLGFRDETCDKLRHPLFKDLNWRQLEAGMLMPPFIPDS
34 _consrvd * * * * *
35
Normal text file length: 2,535 lines: 45 Ln: 1 Col: 71 Pos: 71 Unix (LF) UTF-8 INS
```

Fig19. Sequence alignment

```
C:\Program Files\Modeller10.2\bin\rhodopsin\script4.py - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
qseq.all script1.py script1.log script2.py script2.log script3.py script3.log qseq-2bfxA.pap script4.py
1 from modeller import *
2 from modeller.automodel import *
3 #from modeller import soap_protein_od
4
5 env = Environ()
6 a = AutoModel(env, alnfile='qseq-2bfxA.ali',
7               knowns='2bfxA', sequence='qseq',
8               assess_methods=(assess.DOPE,
9                               #soap_protein_od.Scorer(),
10                              assess.GA341))
11 a.starting_model = 1
12 a.ending_model = 5
13 a.make()
14
Python file length: 408 lines: 13 Ln: 13 Col: 9 Pos: 409 Windows (CR LF) UTF-8 INS
```

Fig20. Python script for model building

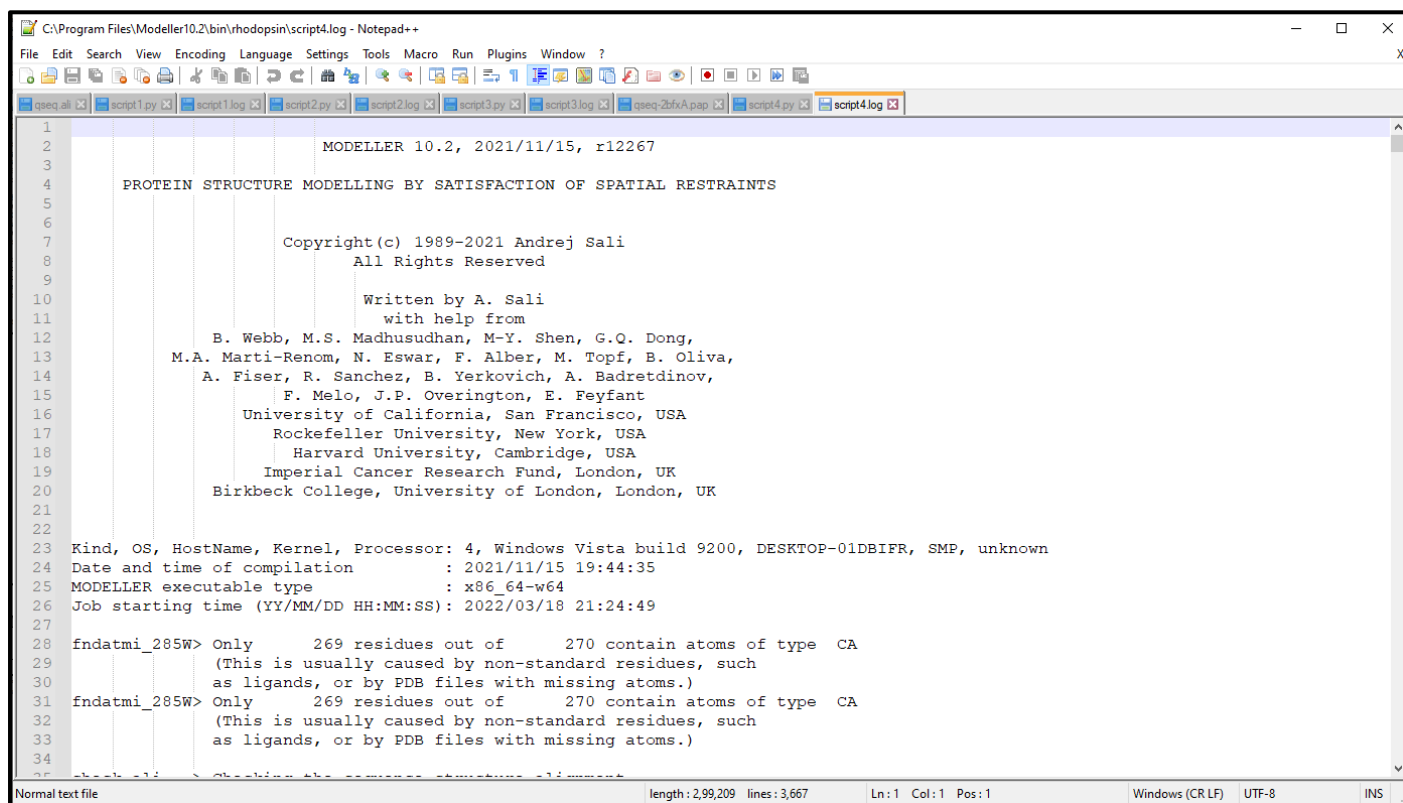


```
C:\Program Files\Modeller10.2\bin\rhodopsin>mod10.2 script2.py
'import site' failed; use -v for traceback

C:\Program Files\Modeller10.2\bin\rhodopsin>mod10.2 script3.py
'import site' failed; use -v for traceback

C:\Program Files\Modeller10.2\bin\rhodopsin>
```

Fig21. Running script4.py



```
C:\Program Files\Modeller10.2\bin\rhodopsin\script4.log - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?

1
2      MODELLER 10.2, 2021/11/15, r12267
3
4      PROTEIN STRUCTURE MODELLING BY SATISFACTION OF SPATIAL RESTRAINTS
5
6
7      Copyright (c) 1989-2021 Andrej Sali
8      All Rights Reserved
9
10     Written by A. Sali
11     with help from
12     B. Webb, M.S. Madhusudhan, M-Y. Shen, G.Q. Dong,
13     M.A. Marti-Renom, N. Eswar, F. Alber, M. Topf, B. Oliva,
14     A. Fiser, R. Sanchez, B. Yerkovich, A. Badretdinov,
15     F. Melo, J.P. Overington, E. Feyfant
16     University of California, San Francisco, USA
17     Rockefeller University, New York, USA
18     Harvard University, Cambridge, USA
19     Imperial Cancer Research Fund, London, UK
20     Birkbeck College, University of London, London, UK
21
22
23 Kind, OS, HostName, Kernel, Processor: 4, Windows Vista build 9200, DESKTOP-01DBIFR, SMP, unknown
24 Date and time of compilation      : 2021/11/15 19:44:35
25 MODELLER executable type        : x86_64-w64
26 Job starting time (YY/MM/DD HH:MM:SS): 2022/03/18 21:24:49
27
28 fndatmi_285W> Only      269 residues out of      270 contain atoms of type  CA
29 (This is usually caused by non-standard residues, such
30 as ligands, or by PDB files with missing atoms.)
31 fndatmi_285W> Only      269 residues out of      270 contain atoms of type  CA
32 (This is usually caused by non-standard residues, such
33 as ligands, or by PDB files with missing atoms.)
34
35 check -f1 -f2 -f3 -f4 -f5 -f6 -f7 -f8 -f9 -f10 -f11 -f12 -f13 -f14 -f15 -f16 -f17 -f18 -f19 -f20 -f21 -f22 -f23 -f24 -f25 -f26 -f27 -f28 -f29 -f30 -f31 -f32 -f33 -f34 -f35 -f36 -f37 -f38 -f39 -f40 -f41 -f42 -f43 -f44 -f45 -f46 -f47 -f48 -f49 -f50 -f51 -f52 -f53 -f54 -f55 -f56 -f57 -f58 -f59 -f60 -f61 -f62 -f63 -f64 -f65 -f66 -f67 -f68 -f69 -f70 -f71 -f72 -f73 -f74 -f75 -f76 -f77 -f78 -f79 -f80 -f81 -f82 -f83 -f84 -f85 -f86 -f87 -f88 -f89 -f90 -f91 -f92 -f93 -f94 -f95 -f96 -f97 -f98 -f99 -f100 -f101 -f102 -f103 -f104 -f105 -f106 -f107 -f108 -f109 -f110 -f111 -f112 -f113 -f114 -f115 -f116 -f117 -f118 -f119 -f120 -f121 -f122 -f123 -f124 -f125 -f126 -f127 -f128 -f129 -f130 -f131 -f132 -f133 -f134 -f135 -f136 -f137 -f138 -f139 -f140 -f141 -f142 -f143 -f144 -f145 -f146 -f147 -f148 -f149 -f150 -f151 -f152 -f153 -f154 -f155 -f156 -f157 -f158 -f159 -f160 -f161 -f162 -f163 -f164 -f165 -f166 -f167 -f168 -f169 -f170 -f171 -f172 -f173 -f174 -f175 -f176 -f177 -f178 -f179 -f180 -f181 -f182 -f183 -f184 -f185 -f186 -f187 -f188 -f189 -f190 -f191 -f192 -f193 -f194 -f195 -f196 -f197 -f198 -f199 -f200 -f201 -f202 -f203 -f204 -f205 -f206 -f207 -f208 -f209 -f210 -f211 -f212 -f213 -f214 -f215 -f216 -f217 -f218 -f219 -f220 -f221 -f222 -f223 -f224 -f225 -f226 -f227 -f228 -f229 -f230 -f231 -f232 -f233 -f234 -f235 -f236 -f237 -f238 -f239 -f240 -f241 -f242 -f243 -f244 -f245 -f246 -f247 -f248 -f249 -f250 -f251 -f252 -f253 -f254 -f255 -f256 -f257 -f258 -f259 -f260 -f261 -f262 -f263 -f264 -f265 -f266 -f267 -f268 -f269 -f270 -f271 -f272 -f273 -f274 -f275 -f276 -f277 -f278 -f279 -f280 -f281 -f282 -f283 -f284 -f285 -f286 -f287 -f288 -f289 -f290 -f291 -f292 -f293 -f294 -f295 -f296 -f297 -f298 -f299 -f300 -f301 -f302 -f303 -f304 -f305 -f306 -f307 -f308 -f309 -f310 -f311 -f312 -f313 -f314 -f315 -f316 -f317 -f318 -f319 -f320 -f321 -f322 -f323 -f324 -f325 -f326 -f327 -f328 -f329 -f330 -f331 -f332 -f333 -f334 -f335 -f336 -f337 -f338 -f339 -f340 -f341 -f342 -f343 -f344 -f345 -f346 -f347 -f348 -f349 -f350 -f351 -f352 -f353 -f354 -f355 -f356 -f357 -f358 -f359 -f360 -f361 -f362 -f363 -f364 -f365 -f366 -f367 -f368 -f369 -f370 -f371 -f372 -f373 -f374 -f375 -f376 -f377 -f378 -f379 -f380 -f381 -f382 -f383 -f384 -f385 -f386 -f387 -f388 -f389 -f390 -f391 -f392 -f393 -f394 -f395 -f396 -f397 -f398 -f399 -f400 -f401 -f402 -f403 -f404 -f405 -f406 -f407 -f408 -f409 -f410 -f411 -f412 -f413 -f414 -f415 -f416 -f417 -f418 -f419 -f420 -f421 -f422 -f423 -f424 -f425 -f426 -f427 -f428 -f429 -f430 -f431 -f432 -f433 -f434 -f435 -f436 -f437 -f438 -f439 -f440 -f441 -f442 -f443 -f444 -f445 -f446 -f447 -f448 -f449 -f450 -f451 -f452 -f453 -f454 -f455 -f456 -f457 -f458 -f459 -f460 -f461 -f462 -f463 -f464 -f465 -f466 -f467 -f468 -f469 -f470 -f471 -f472 -f473 -f474 -f475 -f476 -f477 -f478 -f479 -f480 -f481 -f482 -f483 -f484 -f485 -f486 -f487 -f488 -f489 -f490 -f491 -f492 -f493 -f494 -f495 -f496 -f497 -f498 -f499 -f500 -f501 -f502 -f503 -f504 -f505 -f506 -f507 -f508 -f509 -f510 -f511 -f512 -f513 -f514 -f515 -f516 -f517 -f518 -f519 -f520 -f521 -f522 -f523 -f524 -f525 -f526 -f527 -f528 -f529 -f530 -f531 -f532 -f533 -f534 -f535 -f536 -f537 -f538 -f539 -f540 -f541 -f542 -f543 -f544 -f545 -f546 -f547 -f548 -f549 -f550 -f551 -f552 -f553 -f554 -f555 -f556 -f557 -f558 -f559 -f560 -f561 -f562 -f563 -f564 -f565 -f566 -f567 -f568 -f569 -f570 -f571 -f572 -f573 -f574 -f575 -f576 -f577 -f578 -f579 -f580 -f581 -f582 -f583 -f584 -f585 -f586 -f587 -f588 -f589 -f590 -f591 -f592 -f593 -f594 -f595 -f596 -f597 -f598 -f599 -f600 -f601 -f602 -f603 -f604 -f605 -f606 -f607 -f608 -f609 -f610 -f611 -f612 -f613 -f614 -f615 -f616 -f617 -f618 -f619 -f620 -f621 -f622 -f623 -f624 -f625 -f626 -f627 -f628 -f629 -f630 -f631 -f632 -f633 -f634 -f635 -f636 -f637 -f638 -f639 -f640 -f641 -f642 -f643 -f644 -f645 -f646 -f647 -f648 -f649 -f650 -f651 -f652 -f653 -f654 -f655 -f656 -f657 -f658 -f659 -f660 -f661 -f662 -f663 -f664 -f665 -f666 -f667 -f668 -f669 -f670 -f671 -f672 -f673 -f674 -f675 -f676 -f677 -f678 -f679 -f680 -f681 -f682 -f683 -f684 -f685 -f686 -f687 -f688 -f689 -f690 -f691 -f692 -f693 -f694 -f695 -f696 -f697 -f698 -f699 -f700 -f701 -f702 -f703 -f704 -f705 -f706 -f707 -f708 -f709 -f710 -f711 -f712 -f713 -f714 -f715 -f716 -f717 -f718 -f719 -f720 -f721 -f722 -f723 -f724 -f725 -f726 -f727 -f728 -f729 -f730 -f731 -f732 -f733 -f734 -f735 -f736 -f737 -f738 -f739 -f740 -f741 -f742 -f743 -f744 -f745 -f746 -f747 -f748 -f749 -f750 -f751 -f752 -f753 -f754 -f755 -f756 -f757 -f758 -f759 -f760 -f761 -f762 -f763 -f764 -f765 -f766 -f767 -f768 -f769 -f770 -f771 -f772 -f773 -f774 -f775 -f776 -f777 -f778 -f779 -f780 -f781 -f782 -f783 -f784 -f785 -f786 -f787 -f788 -f789 -f790 -f791 -f792 -f793 -f794 -f795 -f796 -f797 -f798 -f799 -f800 -f801 -f802 -f803 -f804 -f805 -f806 -f807 -f808 -f809 -f810 -f811 -f812 -f813 -f814 -f815 -f816 -f817 -f818 -f819 -f820 -f821 -f822 -f823 -f824 -f825 -f826 -f827 -f828 -f829 -f830 -f831 -f832 -f833 -f834 -f835 -f836 -f837 -f838 -f839 -f840 -f841 -f842 -f843 -f844 -f845 -f846 -f847 -f848 -f849 -f850 -f851 -f852 -f853 -f854 -f855 -f856 -f857 -f858 -f859 -f860 -f861 -f862 -f863 -f864 -f865 -f866 -f867 -f868 -f869 -f870 -f871 -f872 -f873 -f874 -f875 -f876 -f877 -f878 -f879 -f880 -f881 -f882 -f883 -f884 -f885 -f886 -f887 -f888 -f889 -f890 -f891 -f892 -f893 -f894 -f895 -f896 -f897 -f898 -f899 -f900 -f901 -f902 -f903 -f904 -f905 -f906 -f907 -f908 -f909 -f910 -f911 -f912 -f913 -f914 -f915 -f916 -f917 -f918 -f919 -f920 -f921 -f922 -f923 -f924 -f925 -f926 -f927 -f928 -f929 -f930 -f931 -f932 -f933 -f934 -f935 -f936 -f937 -f938 -f939 -f940 -f941 -f942 -f943 -f944 -f945 -f946 -f947 -f948 -f949 -f950 -f951 -f952 -f953 -f954 -f955 -f956 -f957 -f958 -f959 -f960 -f961 -f962 -f963 -f964 -f965 -f966 -f967 -f968 -f969 -f970 -f971 -f972 -f973 -f974 -f975 -f976 -f977 -f978 -f979 -f980 -f981 -f982 -f983 -f984 -f985 -f986 -f987 -f988 -f989 -f990 -f991 -f992 -f993 -f994 -f995 -f996 -f997 -f998 -f999 -f1000 -f1001 -f1002 -f1003 -f1004 -f1005 -f1006 -f1007 -f1008 -f1009 -f1010 -f1011 -f1012 -f1013 -f1014 -f1015 -f1016 -f1017 -f1018 -f1019 -f1020 -f1021 -f1022 -f1023 -f1024 -f1025 -f1026 -f1027 -f1028 -f1029 -f1030 -f1031 -f1032 -f1033 -f1034 -f1035 -f1036 -f1037 -f1038 -f1039 -f1040 -f1041 -f1042 -f1043 -f1044 -f1045 -f1046 -f1047 -f1048 -f1049 -f1050 -f1051 -f1052 -f1053 -f1054 -f1055 -f1056 -f1057 -f1058 -f1059 -f1060 -f1061 -f1062 -f1063 -f1064 -f1065 -f1066 -f1067 -f1068 -f1069 -f1070 -f1071 -f1072 -f1073 -f1074 -f1075 -f1076 -f1077 -f1078 -f1079 -f1080 -f1081 -f1082 -f1083 -f1084 -f1085 -f1086 -f1087 -f1088 -f1089 -f1090 -f1091 -f1092 -f1093 -f1094 -f1095 -f1096 -f1097 -f1098 -f1099 -f1100 -f1101 -f1102 -f1103 -f1104 -f1105 -f1106 -f1107 -f1108 -f1109 -f1110 -f1111 -f1112 -f1113 -f1114 -f1115 -f1116 -f1117 -f1118 -f1119 -f1120 -f1121 -f1122 -f1123 -f1124 -f1125 -f1126 -f1127 -f1128 -f1129 -f1130 -f1131 -f1132 -f1133 -f1134 -f1135 -f1136 -f1137 -f1138 -f1139 -f1140 -f1141 -f1142 -f1143 -f1144 -f1145 -f1146 -f1147 -f1148 -f1149 -f1150 -f1151 -f1152 -f1153 -f1154 -f1155 -f1156 -f1157 -f1158 -f1159 -f1160 -f1161 -f1162 -f1163 -f1164 -f1165 -f1166 -f1167 -f1168 -f1169 -f1170 -f1171 -f1172 -f1173 -f1174 -f1175 -f1176 -f1177 -f1178 -f1179 -f1180 -f1181 -f1182 -f1183 -f1184 -f1185 -f1186 -f1187 -f1188 -f1189 -f1190 -f1191 -f1192 -f1193 -f1194 -f1195 -f1196 -f1197 -f1198 -f1199 -f1200 -f1201 -f1202 -f1203 -f1204 -f1205 -f1206 -f1207 -f1208 -f1209 -f1210 -f1211 -f1212 -f1213 -f1214 -f1215 -f1216 -f1217 -f1218 -f1219 -f1220 -f1221 -f1222 -f1223 -f1224 -f1225 -f1226 -f1227 -f1228 -f1229 -f1230 -f1231 -f1232 -f1233 -f1234 -f1235 -f1236 -f1237 -f1238 -f1239 -f1240 -f1241 -f1242 -f1243 -f1244 -f1245 -f1246 -f1247 -f1248 -f1249 -f1250 -f1251 -f1252 -f1253 -f1254 -f1255 -f1256 -f1257 -f1258 -f1259 -f1260 -f1261 -f1262 -f1263 -f1264 -f1265 -f1266 -f1267 -f1268 -f1269 -f1270 -f1271 -f1272 -f1273 -f1274 -f1275 -f1276 -f1277 -f1278 -f1279 -f1280 -f1281 -f1282 -f1283 -f1284 -f1285 -f1286 -f1287 -f1288 -f1289 -f1290 -f1291 -f1292 -f1293 -f1294 -f1295 -f1296 -f1297 -f1298 -f1299 -f1300 -f1301 -f1302 -f1303 -f1304 -f1305 -f1306 -f1307 -f1308 -f1309 -f1310 -f1311 -f1312 -f1313 -f1314 -f1315 -f1316 -f1317 -f1318 -f1319 -f1320 -f1321 -f1322 -f1323 -f1324 -f1325 -f1326 -f1327 -f1328 -f1329 -f1330 -f1331 -f1332 -f1333 -f1334 -f1335 -f1336 -f1337 -f1338 -f1339 -f1340 -f1341 -f1342 -f1343 -f1344 -f1345 -f1346 -f1347 -f1348 -f1349 -f1350 -f1351 -f1352 -f1353 -f1354 -f1355 -f1356 -f1357 -f1358 -f1359 -f1360 -f1361 -f1362 -f1363 -f1364 -f1365 -f1366 -f1367 -f1368 -f1369 -f1370 -f1371 -f1372 -f1373 -f1374 -f1375 -f1376 -f1377 -f1378 -f1379 -f1380 -f1381 -f1382 -f1383 -f1384 -f1385 -f1386 -f1387 -f1388 -f1389 -f1390 -f1391 -f1392 -f1393 -f1394 -f1395 -f1396 -f1397 -f1398 -f1399 -f1400 -f1401 -f1402 -f1403 -f1404 -f1405 -f1406 -f1407 -f1408 -f1409 -f1410 -f1411 -f1412 -f1413 -f1414 -f1415 -f1416 -f1417 -f1418 -f1419 -f1420 -f1421 -f1422 -f1423 -f1424 -f1425 -f1426 -f1427 -f1428 -f1429 -f1430 -f1431 -f1432 -f1433 -f1434 -f1435 -f1436 -f1437 -f1438 -f1439 -f1440 -f1441 -f1442 -f1443 -f1444 -f1445 -f1446 -f1447 -f1448 -f1449 -f1450 -f1451 -f1452 -f1453 -f1454 -f1455 -f1456 -f1457 -f1458 -f1459 -f1460 -f1461 -f1462 -f1463 -f1464 -f1465 -f1466 -f1467 -f1468 -f1469 -f1470 -f1471 -f1472 -f1473 -f1474 -f1475 -f1476 -f1477 -f1478 -f1479 -f1480 -f1481 -f1482 -f1483 -f1484 -f1485 -f1486 -f1487 -f1488 -f1489 -f1490 -f1491 -f1492 -f1493 -f1494 -f1495 -f1496 -f1497 -f1498 -f1499 -f1500 -f1501 -f1502 -f1503 -f1504 -f1505 -f1506 -f1507 -f1508 -f1509 -f1510 -f1511 -f1512 -f1513 -f1514 -f1515 -f1516 -f1517 -f1518 -f1519 -f1520 -f1521 -f1522 -f1523 -f1524 -f1525 -f1526 -f1527 -f1528 -f1529 -f1530 -f1531 -f1532 -f1533 -f1534 -f1535 -f1536 -f1537 -f1538 -f1539 -f1540 -f1541 -f1542 -f1543 -f1544 -f1545 -f1546 -f1547 -f1548 -f1549 -f1550 -f1551 -f1552 -f1553 -f1554 -f1555 -f1556 -f1557 -f1558 -f1559 -f1560 -f1561 -f1562 -f1563 -f1564 -f1565 -f1566 -f1567 -f1568 -f1569 -f1570 -f1571 -f1572 -f1573 -f1574 -f1575 -f1576 -f1577 -f1578 -f1579 -f1580 -f1581 -f1582 -f1583 -f1584 -f1585 -f1586 -f1587 -f1588 -f1589 -f1590 -f1591 -f1592 -f1593 -f1594 -f1595 -f1596 -f1597 -f1598 -f1599 -f1600 -f1601 -f1602 -f1603 -f1604 -f1605 -f1606 -f1607 -f1608 -f1609 -f1610 -f1611 -f1612 -f1613 -f1614 -f1615 -f1616 -f1617 -f1618 -f1619 -f1620 -f1621 -f1622 -f1623 -f1624 -f1625 -f1626 -f1627 -f1628 -f1629 -f1630 -f1631 -f1632 -f1633 -f1634 -f1635 -f1636 -f1637 -f1638 -f1639 -f1640 -f1641 -f1642 -f1643 -f1644 -f1645 -f1646 -f1647 -f1648 -f1649 -f1650 -f1651 -f1652 -f1653 -f1654 -f1655 -f1656 -f1657 -f1658 -f1659 -f1660 -f1661 -f1662 -f1663 -f1664 -f1665 -f1666 -f1667 -f1668 -f1669 -f1670 -f1671 -f1672 -f1673 -f1674 -f1675 -f1676 -f1677 -f1678 -f1679 -f1680 -f1681 -f1682 -f1683 -f1684 -f1685 -f1686 -f1687 -f1688 -f1689 -f1690 -f1691 -f1692 -f1693 -f1694 -f1695 -f1696 -f1697 -f1698 -f1699 -f1700 -f1701 -f1702 -f1703 -f1704 -f1705 -f1706 -f1707 -f1708 -f1709 -f1710 -f1711 -f1712 -f1713 -f1714 -f1715 -f1716 -f1717 -f1718 -f1719 -f1720 -f1721 -f1722 -f1723 -f1724 -f1725 -f1726 -f1727 -f1728 -f1729 -f1730 -f1731 -f1732 -f1733 -f1734 -f1735 -f1736 -f1737 -f1738 -f1739 -f1740 -f1741 -f1742 -f1743 -f1744 -f1745 -f1746 -f1747 -f1748 -f1749 -f1750 -f1751 -f1752 -f1753 -f1754 -f1755 -f1756 -f1757 -f1758 -f1759 -f1760 -f1761 -f1762 -f1763 -f1764 -f1765 -f1766 -f1767 -f1768 -f1769 -f1770 -f1771 -f1772 -f1773 -f1774 -f1775 -f1776 -f1777 -f1778 -f1779 -f1780 -f1781 -f1782 -f1783 -f1784 -f1785 -f1786 -f1787 -f1788 -f1789 -f1790 -f1791 -f1792 -f1793 -f1794 -f1795 -f1796 -f1797 -f1798 -f1799 -f1800 -f1801 -f1802 -f1803 -f1804 -f1805 -f1806 -f1807 -f1808 -f1809 -f1810 -f1811 -f1812 -f1813 -f1814 -f1815 -f1816 -f1817 -f1818 -f1819 -f1820 -f1821 -f1822 -f1823 -f1824 -f1825 -f1826 -f1827 -f1828 -f1829 -f1830 -f1831 -f1832 -f1833 -f1834 -f1835 -f1836 -f1837 -f1838 -f1839 -f1840 -f1841 -f1842 -f1843 -f1844 -f1845 -f1846 -f1847 -f1848 -f1849 -f1850 -f1851 -f1852 -f1853 -f1854 -f1855 -f1856 -f1857 -f1858 -f1859 -f1860 -f1861 -f1862 -f1863 -f1864 -f1865 -f1866 -f1867 -f1868 -f1869 -f1870 -f1871 -f1872 -f1873 -f1874 -f1875 -f1876 -f1877 -f1878 -f1879 -f1880 -f1881 -f1882 -f1883 -f1884 -f1885 -f1886 -f1887 -f1888 -f1889 -f1890 -f1891 -f1892 -f1893 -f1894 -f1895 -f1896 -f1897 -f1898 -f1899 -f1900 -f1901 -f1902 -f1903 -f1904 -f1905 -f1906 -f1907 -f1908 -f1909 -f1910 -f1911 -f1912 -f1913 -f1914 -f1915 -f1916 -f1917 -f1918 -f1919 -f1920 -f1921 -f1922 -f1923 -f1924 -f1925 -f1926 -f1927 -f1928 -f1929 -f1930 -f1931 -f1932 -f1933 -f1934 -f1935 -f1936 -f1937 -f1938 -f1939 -f1940 -f1941 -f1942 -f1943 -f1944 -f1945 -f1946 -f1947 -f1948 -f1949 -f1950 -f1951 -f1952 -f1953 -f1954 -f1955 -f1956 -f1957 -f1958 -f1959 -f1960 -f1961 -f1962 -f1963 -f1964 -f1965 -f1966 -f1967 -f1968 -f1969 -f1970 -f1971 -f1972 -f1973 -f1974 -f1975 -f1976 -f1977 -f1978 -f1979 -f1980 -f1981 -f1982 -f1983 -f1984 -f1985 -f1986 -f1987 -f1988 -f1989 -f1990 -f1991 -f1992 -f1993 -f1994 -f1995 -f1996 -f1997 -f1998 -f1999 -f2000 -f2001 -f2002 -f2003 -f2004 -f2005 -f2006 -f2007 -f2008 -f2009 -f2010 -f2011 -f2012 -f2013 -f2014 -f2015 -f2016 -f2017 -f2018 -f2019 -f2020 -f2021 -f2022 -f2023 -f2024 -f2025 -f2026 -f2027 -f2028 -f2029 -f2030 -f2031 -f2032 -f2033 -f2034 -f2035 -f2036 -f2037 -f2038 -f2039 -f2040 -f2041 -f2042 -f2043 -f2044 -f2045 -f2046 -f2047 -f2048 -f2049 -f2050 -f2051 -f2052 -f2053 -f2054 -f2055 -f2056 -f2057 -f2058 -f2059 -f2060 -f2061 -f2062 -f2063 -f2064 -f2065 -f2066 -f2067 -f2068 -f2069 -f2070 -f2071 -f2072 -f2073 -f2074 -f2075 -f2076 -f2077 -f2078 -f2079 -f208
```

```

C:\Program Files\Modeller10.2\bin\rhodopsin\script4.log - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
qseq.all x script1.py x script1.log x script2.py x script2.log x script3.py x script3.log x qseq-2fxA.pap x script4.py x script4.log x
3640 254 558G 558G N CA 4432 4433 161.88 174.60 8.50
3641 255 15981 558G 559M C N 4434 4436 -75.58 -73.00 3.75 0.24 -63.40 174.21 26.97
3642 255 559M 559M N CA 4436 4437 145.71 143.00 -40.50
3643 256 15983 560C 561L C N 4448 4450 -71.54 -70.70 6.86 0.49 -63.50 170.58 24.15
3644 256 561L 561L N CA 4450 4451 148.41 141.60 -41.20
3645
3646
3647 report_____> Distribution of short non-bonded contacts:
3648
3649
3650 DISTANCE1: 0.00 2.10 2.20 2.30 2.40 2.50 2.60 2.70 2.80 2.90 3.00 3.10 3.20 3.30 3.40
3651 DISTANCE2: 2.10 2.20 2.30 2.40 2.50 2.60 2.70 2.80 2.90 3.00 3.10 3.20 3.30 3.40 3.50
3652 FREQUENCY: 0 0 0 0 0 55 82 389 425 588 492 567 641 684 714
3653
3654
3655 << end of ENERGY.
3656
3657 >> Summary of successfully produced models:
3658 Filename molpdf DOPE score GA341 score
3659 -----
3660 qseq.B99990001.pdb 5123.33984 -39670.18359 0.69617
3661 qseq.B99990002.pdb 5371.25439 -39643.84766 0.70775
3662 qseq.B99990003.pdb 4576.40967 -39959.83984 0.92847
3663 qseq.B99990004.pdb 4718.91113 -38850.38672 0.95707
3664
3665
3666 Total CPU time [seconds] : 219.75
3667
Normal text file length: 2,99,209 lines: 3,667 Ln: 3,664 Col: 57 Pos: 2,99,121 Windows (CR LF) UTF-8 INS

```

Fig23. Structure with lowest DOPE score selected as final model

RESULT:

Modeller was used to predict the tertiary structure of Rhodopsin

CONCLUSION:

Thus, modeller can be used to predict tertiary structures of proteins by comparative protein structure modelling. These tools give faster results than x-ray or NMR techniques and can be used by researchers to understand protein functions and drug designing.

REFERENCES:

1. Xiong, J. (2008).Tertiary structure prediction. Essential bioinformatics. Cambridge: Cambridge University Press. 214-228.
2. Encyclopædia Britannica, inc. (n.d.). *Rhodopsin*. Encyclopædia Britannica. Retrieved March 18, 2022, from <https://www.britannica.com/science/rhodopsin>
3. Uniprot. (n.d.). Retrieved March 18, 2022, from <https://www.uniprot.org/uniprot/Q15835.fasta>

WEBLEM 3b

I-TASSAR

(URL: <https://zhanggroup.org/I-TASSER/>)

AIM:

To perform tertiary structure prediction by threading approach using I-TASSER server for query rhodopsin.

INTRODUCTION:

Rhodopsin, also called visual purple, pigment-containing sensory protein that converts light into an electrical signal. Rhodopsin is found in a wide range of organisms, from vertebrates to bacteria. In many seeing animals, including humans, it is required for vision in dim light and is located in the retina of the eye—specifically, within the tightly packed disks that make up the outer segment of the retina's photoreceptive rod cells, which are specially adapted for vision under low-light conditions.

I-TASSER server is an on-line platform that implements the I-TASSER based algorithms for protein structure and function predictions. It allows academic users to automatically generate high-quality model predictions of 3D structure and biological function of protein molecules from their amino acid sequences.

METHODOLOGY:

1. Open homepage for I-TASSER. (URL: <https://zhanggroup.org/I-TASSER/>)
2. Complete registration.
3. Submit FASTA sequence for kinase.
4. Observe and interpret results.

OBSERVATION:

The screenshot displays the UniProtKB entry for Rhodopsin (P02699). The entry is for the protein Rhodopsin (RHO) from Bos taurus (Bovine). The status is 'Reviewed' with an annotation score of 5.0. The function is described as a photoreceptor required for image-forming vision at low light intensity. The entry includes a list of sites, with the most prominent being Site 1 at position 113, which plays an important role in the conformation switch to the active conformation. The entry also lists various molecular functions, including 11-cis retinal binding, arrestin family protein binding, G-protein alpha-subunit binding, G-protein-coupled photoreceptor activity, guanyl-nucleotide exchange factor activity, identical protein binding, and opsin binding.

Feature key	Position(s)	Description	Actions	Graphical view	Length
Site 1	113	Plays an important role in the conformation switch to the active conformation	3 Publications	1 Publication	1
Metal binding 1	201	Zinc	Combined sources	2 Publications	1
Metal binding 1	279	Zinc	Combined sources	2 Publications	1

GO - Molecular function 1

- 11-cis retinal binding (Source: UniProtKB)
- arrestin family protein binding (Source: CAFA)
- G-protein alpha-subunit binding (Source: CAFA)
- G-protein-coupled photoreceptor activity (Source: UniProtKB)
- guanyl-nucleotide exchange factor activity (Source: UniProtKB)
- identical protein binding (Source: IntAct)
- opsin binding (Source: CAFA)

Fig1. Result page for Rhodopsin in UniProt database

```
>sp|P02699|OPSD_BOVIN Rhodopsin OS=Bos taurus OX=9913 GN=RHO PE=1 SV=1
MNGTEGPNFYVPFSNKTGVVRSPFEAPQYYLAEPWQFSMLAAYMFLLIMLGFPINFLTLY
VTVQHKKLRTPLNYILLNLAVADLFMVFGGFTTTTLYTSLHGYFVFGPTGCNLEGFFATLG
GEIALWSLVVLAIERVYVVKPMSNFRFGENHAIMGVAF TWVMALACAAPPLVGWSRYIP
EGMQCSCGIDYYTPHEETNNE SFVIYMFVVHFI IPLIVIFFCYGQLVFTVKEAAAQQQES
ATTQKAEKEVTRMVIIMVIAFLICWLPYAGVAFYIFTHQGSDFGP IFMTIPAFFAKTSAV
YNPVIYIMMNKQFRNCMVTTLC CGKNPLGDDEASTTVSKTETSQVAPA
```

Fig2. FASTA sequence for Rhodopsin

Online Services

- I-TASSER
- C-I-TASSER
- QUARK
- C-QUARK
- LOMETS
- COACH
- COFACTOR
- MetaGO
- MUSTER
- CEThreader
- SEGMENT
- FG-MD
- ModRefiner
- REMO
- DEMO
- SPRING
- COTH
- Threpp
- BSpreD
- ANGLOD
- EDock
- BSP-SLIM
- SAXSTER
- FUpred
- ThreaDom
- ThreaDomEx
- EvoDesign
- CR-I-TASSER
- GPOR-I-TASSER
- MAGELLAN
- BindProf
- BindProX
- SSIPa

(The server completed predictions for 676386 proteins submitted by 163852 users from 159 countries)
(The template library was updated on 2022/03/14)

I-TASSER (Iterative Threading ASSEMBLY Refinement) is a hierarchical approach to protein structure prediction and structure-based function annotation. It first identifies structural templates from the PDB by multiple threading approach LOMETS, with full-length atomic models constructed by iterative template-based fragment assembly simulations. Function insights of the target are then derived by re-threading the 3D models through protein function database BiGLEP. I-TASSER (as 'Zhang-Server') was ranked as the No. 1 server for protein structure prediction in recent community-wide CASP7, CASP8, CASP9, CASP10, CASP11, CASP12, CASP13, and CASP14 experiments. It was also ranked the best for function prediction in CASP9. The server is in active development with the goal to provide the most accurate protein structure and function predictions using state-of-the-art algorithms. The server is only for non-commercial use. Please report problems and questions at [I-TASSER message board](#) and our developers will study and answer the questions accordingly. [>> More about the server...](#)

Structure models for the SARS-CoV2 Coronavirus genome by C-I-TASSER

[\[Queue\]](#) [\[Forum\]](#) [\[Download\]](#) [\[Search\]](#) [\[Registration\]](#) [\[Statistics\]](#) [\[Remove\]](#) [\[Potential\]](#) [\[Decoys\]](#) [\[News\]](#) [\[Annotation\]](#) [\[About\]](#) [\[FAQ\]](#)

I-TASSER On-line Server ([View an example of I-TASSER output](#)):

Copy and paste your sequence within [10, 1500] residues in FASTA format. [Click here for a sample input.](#)

Or upload the sequence from your local computer:
[Choose File](#) No file chosen

Email: (mandatory, where results will be sent to)

Password: (mandatory, please click [here](#) if you do not have a password)

ID: (optional, your given name of the protein)

► **Option I:** Assign additional restraints & templates to guide I-TASSER modeling.

► **Option II:** Exclude some templates from I-TASSER template library.

► **Option III:** Specify secondary structure for specific residues.

☒ Keep my results public (uncheck this box if you want to keep your job private, and a key will be assigned for you to access the results. We received numerous requests from users who lost their key to access result. To save your time, please keep results public, or ensure you remember the key if you choose to keep job private)

[Run I-TASSER](#) [Clear form](#)

(Please submit a new job only after your old job is completed)

Waiting for www.goo...

Fig3. Homepage for I-TASSAR

- a. All the residues are coloured in black; however, those residues in template which are identical to the residue in the query sequence are highlighted in colour. Colouring scheme is based on the property of amino acids, where polar are brightly coloured while non-polar residues are coloured in dark shade. (More about the colours used)
- b. Rank of templates represents the top ten threading templates used by I-TASSER.
- c. Iden1 is the percentage sequence identity of the templates in the threading aligned region with the query sequence.
- d. Iden2 is the percentage sequence identity of the whole template chains with query sequence.
- e. Cov represents the coverage of the threading alignment and is equal to the number of aligned residues divided by the length of query protein.
- f. Norm. Z-score is the normalized Z-score of the threading alignments. Alignment with a Normalized Z-score >1 mean a good alignment and vice versa.
- g. Download Align. provides the 3D structure of the aligned regions of the threading templates.
- h. The top 10 alignments reported above (in order of their ranking) are from the following threading programs: 1: FFAS-3D 2: SPARKS-X 3: HHSEARCH2 4: HHSEARCH I 5: Neff-PPAS 6: HHSEARCH 7: pGenTHREADER 8: wdPPAS 9: PROSPECT2 10: SP3

Top 5 final models predicted by I-TASSER

(For each target, I-TASSER simulations generate a large ensemble of structural conformations, called decoys. To select the final models, I-TASSER uses the SPICKER program to cluster all the decoys based on the pair-wise structure similarity, and reports up to five models which corresponds to the five largest structure clusters. The confidence of each model is quantitatively measured by C-score that is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score is typically in the range of $[-5, 2]$, where a C-score of a higher value signifies a model with a higher confidence and vice-versa. TM-score and RMSD are estimated based on C-score and protein length following the correlation observed between these qualities. Since the top 5 models are ranked by the cluster size, it is possible that the lower-rank models have a higher C-score in rare cases. Although the first model has a better quality in most cases, it is also possible that the lower-rank models have a better quality than the higher-rank models as seen in our benchmark tests. If the I-TASSER simulations converge, it is possible to have less than 5 clusters generated; this is usually an indication that the models have a good quality because of the converged simulations.)

- [More about C-score](#)
- [Local structure accuracy profile of the top five models](#)

(By right-click on the images, you can export image file or change the configurations, e.g. modifying the background color or stopping the spin of your models)

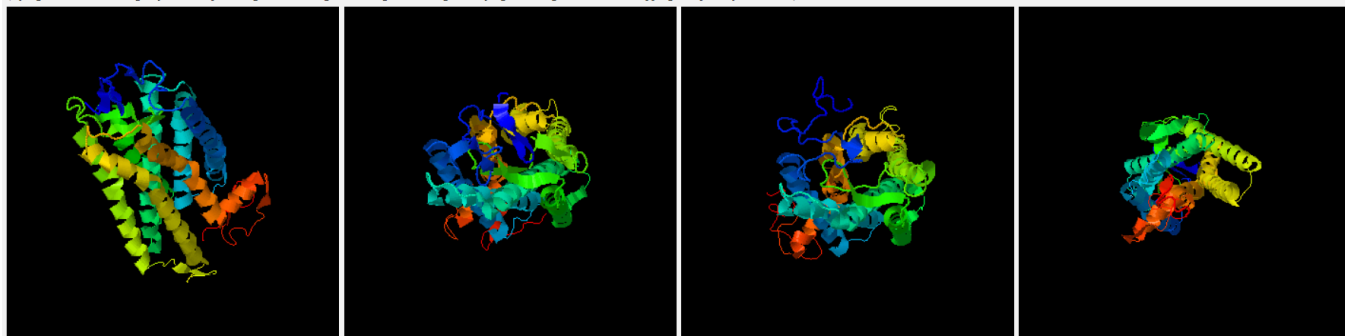
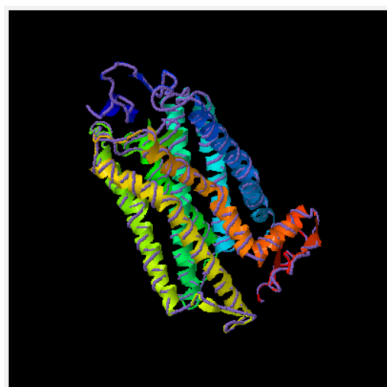


Fig8. Result for top 5 final models predicted

Proteins structurally close to the target in the PDB (as identified by TM-align)

(After the structure assembly simulation, I-TASSER uses the TM-align structural alignment program to match the first I-TASSER model to all structures in the PDB library. This section reports the top 10 proteins from the PDB that have the closest structural similarity, i.e. the highest **TM-score**, to the predicted I-TASSER model. Due to the structural similarity, these proteins often have similar function to the target. However, users are encouraged to use the data in the next section 'Predicted function using COACH' to infer the function of the target protein, since COACH has been extensively trained to derive biological functions from multi-source of sequence and structure features which has on average a higher accuracy than the function annotations derived only from the global structure comparison.)



Top 10 Identified structural analogs in PDB

Click to view	Rank	PDB Hit	TM-score	RMSD ^a	IDEN ^d	Cov	Alignment
<input type="radio"/>	1	1u19A	0.998	0.31	1.000	1.000	Download
<input type="radio"/>	2	2ks9A	0.941	1.40	0.212	0.977	Download
<input type="radio"/>	3	1f88B	0.872	0.52	0.990	0.876	Download
<input type="radio"/>	4	4zwjA	0.857	2.95	0.859	0.954	Download
<input type="radio"/>	5	2ziyA	0.822	2.93	0.235	0.911	Download
<input type="radio"/>	6	6j6kA	0.768	2.61	0.278	0.839	Download
<input type="radio"/>	7	6me6A	0.764	3.38	0.172	0.888	Download
<input type="radio"/>	8	4n6hA	0.730	3.79	0.176	0.859	Download
<input type="radio"/>	9	7m8wA	0.723	3.65	0.165	0.848	Download
<input type="radio"/>	10	6me2A	0.723	3.00	0.202	0.816	Download

(a) Query structure is shown in cartoon, while the structural analog is displayed using backbone trace.

(b) Ranking of proteins is based on TM-score of the structural alignment between the query structure and known structures in the PDB library.

(c) RMSD^a is the RMSD between residues that are structurally aligned by TM-align.

(d) IDEN^d is the percentage sequence identity in the structurally aligned region.

(e) Cov represents the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by length of the query protein.

☐ Spin On/Off

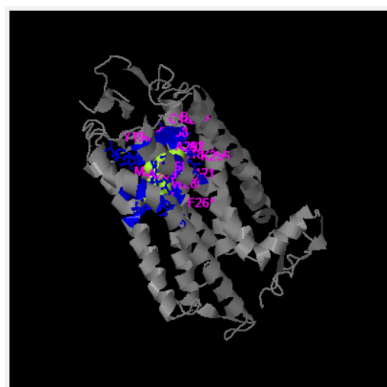
Fig9. Result for proteins that are structurally close to target

- Query structure is shown in cartoon, while the structural analog is displayed using backbone trace.
- Ranking of proteins is based on TM-score of the structural alignment between the query structure and known structures in the PDB library
- RMSD^a is the RMSD between residues that are structurally aligned by TM-align
- IDEN^d is the percentage sequence identity in the structurally aligned region.
- Cov represents the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by length of the query protein.

Predicted function using COFACTOR and COACH

(This section reports biological annotations of the target protein by COFACTOR and COACH based on the I-TASSER structure prediction. While COFACTOR deduces protein functions (ligand-binding sites, EC and GO) using structure comparison and protein-protein networks, COACH is a meta-server approach that combines multiple function annotation results (on ligand-binding sites) from the COFACTOR, TM-SITE and S-SITE programs.)

Ligand binding sites



Click to view	Rank	C-score	Cluster size	PDB Hit	Lig Name	Download Complex	Ligand Binding Site Residues
<input checked="" type="radio"/>	1	0.64	89	3oaxB	RET	Rep , Mult	113,117,118,121,122,186,187,188,189,191,207,212,261,265,268,269,292,296
<input type="radio"/>	2	0.06	7	1gzma	C8E	Rep , Mult	40,43,267,290,291,294
<input type="radio"/>	3	0.04	9	4a4mA	PEPTIDE	Rep , Mult	71,135,138,230,238,239,245,310,311,312
<input type="radio"/>	4	0.04	20	3zpgA	Y01	Rep , Mult	69,74,78,81,119,150,154,157,161
<input type="radio"/>	5	0.03	7	4ldeA	1WV	Rep , Mult	44,47,98,289,293

[Download](#) the residue-specific ligand binding probability, which is estimated by SVM.

[Download](#) the all possible binding ligands and detailed prediction summary.

[Download](#) the templates clustering results.

(a) **C-score** is the confidence score of the prediction. C-score ranges [0-1], where a higher score indicates a more reliable prediction.

(b) **Cluster size** is the total number of templates in a cluster.

(c) **Lig Name** is name of possible binding ligand. Click the name to view its information in the [BioLiP database](#).

(d) **Rep** is a single complex structure with the most representative ligand in the cluster, i.e., the one listed in the **Lig Name** column.

Mult is the complex structures with all potential binding ligands in the cluster.

[Reset to initial orientation](#) ☐ Spin On/Off

Fig10. Result for predicted fuctions

- C-score is the confidence score of the prediction. C-scores ranges [0-1], where a higher score indicated a more reliable prediction.
- Cluster size is the total number of templates in a cluster.
- Lig Name is name of possible binding ligand. Click the name to view its information in the BioLiP database.
- Rep is a single complex structure with the most representative ligand in the cluster, i.e., the one listed in the Lig Name column. Mult is the complex structures with all potential binding ligands in the cluster.

Enzyme Commission (EC) numbers and active sites



Click to view	Rank	Cscore ^{EC}	PDB Hit	TM-score	RMSD ^a	IDEN ^a	Cov	EC Number	Active Site Residues
<input type="radio"/>	1	0.348	3d4sA	0.690	3.34	0.212	0.787	3.2.1.17	NA
<input type="radio"/>	2	0.254	2occN	0.456	5.90	0.051	0.695	1.9.3.1	NA
<input type="radio"/>	3	0.254	1xmeA	0.472	5.83	0.051	0.727	1.9.3.1	NA
<input type="radio"/>	4	0.253	1occcA	0.456	5.89	0.048	0.695	1.9.3.1	NA
<input type="radio"/>	5	0.242	1m56A	0.473	5.13	0.038	0.649	1.9.3.1	NA

Click on the radio buttons to visualize predicted active site residues.

(a) Cscore^{EC} is the confidence score for the EC number prediction. Cscore^{EC} values range in between [0-1]; where a higher score indicates a more reliable EC number prediction.

(b) TM-score is a measure of global structural similarity between query and template protein.

(c) RMSD^a is the RMSD between residues that are structurally aligned by TM-align.

(d) IDEN^a is the percentage sequence identity in the structurally aligned region.

(e) Cov represents the coverage of global structural alignment and is equal to the number of structurally aligned residues divided by length of the query protein.

[Reset to initial orientation](#) ☐ Spin On/Off

Fig11. Enzyme commission (EC) numbers and active sites

- CscoreEC is the confidence score for the EC number prediction. CscoreEC values range in between [0-1]; where a higher score indicates a more reliable EC number prediction.
- TM-score is a measure of global structural similarity between query and template protein.
- RMSDa is the RMSD between residues that are structurally aligned by TM-align.
- IDENa is the percentage sequence identity in the structurally aligned region.
- Cov represents the coverage of global structural alignment and is equal to the number of structurally aligned residues divided by length of the query protein.

Gene Ontology (GO) terms

Top 10 homologous GO templates in PDB

Rank	Cscore ^{GO}	TM-score	RMSD ^a	IDEN ^a	Cov	PDB Hit	Associated GO Terms
1	0.74	0.9980	0.31	1.00	1.00	1u19A	GO:0018298 GO:0046872 GO:0004930 GO:0007601 GO:0009583 GO:0001750 GO:0009881 GO:0007602 GO:0050953 GO:0016021 GO:0050896 GO:0007165 GO:0004871 GO:0007186 GO:0004872 GO:0005515 GO:0009416 GO:0005886 GO:0016020 GO:0060342 GO:0042622 GO:0006468 GO:0001917
2	0.61	0.8715	0.52	0.99	0.88	1f88B	GO:0046872 GO:0018298 GO:0007186 GO:0050953 GO:0004872 GO:0009583 GO:0005515 GO:0016021 GO:0009416 GO:0009881 GO:0005886 GO:0050896 GO:0016020 GO:0007601 GO:0060342 GO:0007165 GO:0042622 GO:0007602 GO:0006468 GO:0004871 GO:0001917 GO:0001750 GO:0004930
3	0.47	0.6974	3.16	0.19	0.80	2x00B	GO:0004935 GO:0007186 GO:0016021 GO:0071875 GO:0004940
4	0.45	0.9412	1.40	0.21	0.98	2ks9A	GO:0004995 GO:0005886 GO:0007186 GO:0016021
5	0.41	0.8215	2.93	0.23	0.91	2ziyA	GO:0004871 GO:0016021 GO:0007601 GO:0018298 GO:0007186 GO:0007602 GO:0004872 GO:0009881 GO:0007165 GO:0004930 GO:0050896 GO:0016020
6	0.40	0.6716	3.20	0.21	0.77	3p0eA	GO:0016998 GO:0016787 GO:0003824 GO:0009253 GO:0003796 GO:0016798 GO:0008152 GO:0042742 GO:0019835 GO:0007186 GO:0016021
7	0.37	0.6573	3.46	0.18	0.78	2ydvA	GO:0001609 GO:0001973 GO:0007186 GO:0016021
8	0.36	0.6969	2.76	0.19	0.78	3rreA	GO:0005737 GO:0004872 GO:0007186 GO:0016021 GO:0010894 GO:0042742 GO:0008152 GO:0016798 GO:0003796 GO:0019835 GO:0003824 GO:0009253 GO:0016787 GO:0016998 GO:0004930 GO:0009629 GO:0045907 GO:0005887 GO:0045429 GO:0006954 GO:0005730 GO:0004969 GO:0051381 GO:0007165 GO:0007200 GO:0004871 GO:0005634 GO:0016020 GO:0005624 GO:0007268 GO:0005886
9	0.35	0.6937	3.31	0.20	0.80	2rh1A	GO:0003796 GO:0007186 GO:0009253 GO:0016021 GO:0016998 GO:0042742 GO:0008152 GO:0016798 GO:0003824 GO:0016787 GO:0019835
10	0.34	0.6847	3.01	0.27	0.77	3pblA	GO:0019835 GO:0016998 GO:0016798 GO:0016787 GO:0008152 GO:0003824 GO:0003796 GO:0009253 GO:0042742 GO:0007186 GO:0016021

Fig12. Gene Ontology (GO) terms

- CscoreGO is a combined measure for evaluating global and local similarity between query and template protein. It's range is [0-1] and higher values indicate more confident predictions.
- TM-score is a measure of global structural similarity between query and template protein.
- RMSDa is the RMSD between residues that are structurally aligned by TM-align.
- IDENa is the percentage sequence identity in the structurally aligned region.
- Cov represents the coverage of global structural alignment and is equal to the number of structurally aligned residues divided by length of the query protein.
- The second table shows a consensus GO terms amongst the top scoring templates. The GO-Score associated with each prediction is defined as the average weight of the GO term, where the weights are assigned based on CscoreGO of the template.

RESULT:

I-TASSER was used to predict the tertiary structure of Kinase based on threading approach. The information regarding solvent accessibility, normalized B-factor, top 10 threading templates, top 5 final models, proteins that are structurally close to target, functions and active sites were predicted.

CONCLUSION:

Thus, I-TASSER can be used to predict tertiary structures of proteins by threading method. These tools give faster results than x-ray or NMR techniques and can be used by researchers to understand protein functions and drug designing.

REFERENCES:

- Xiong, J. (2008).Tertiary structure prediction. Essential bioinformatics. Cambridge: Cambridge University Press. 214-228.
- kinase | Definition, Biology, & Function. (n.d.). Encyclopedia Britannica. Retrieved March 8, 2022, from <https://www.britannica.com/science/kinase>
- I-TASSER server for protein structure and function prediction. (n.d.-b). Zhanggroup.org. Retrieved March 8, 2022, from <https://zhanggroup.org/I-TASSER/>
- I-TASSER results. (n.d.). Zhanggroup.org. Retrieved March 8, 2022, from <https://zhanggroup.org/I-TASSER/output/S673761/>

WEBLEM 3c

ROBETTA

(URL: <https://robetta.bakerlab.org/>)

AIM:

To perform tertiary structure prediction by Ab-Initio approach using ROBETTA server for query Rhodopsin.

INTRODUCTION:

Rhodopsin, also called visual purple, pigment-containing sensory protein that converts light into an electrical signal. Rhodopsin is found in a wide range of organisms, from vertebrates to bacteria. In many seeing animals, including humans, it is required for vision in dim light and is located in the retina of the eye—specifically, within the tightly packed disks that make up the outer segment of the retina's photoreceptive rod cells, which are specially adapted for vision under low-light conditions.

The Robetta server provides automated tools for protein structure prediction and analysis. For structure prediction, sequences submitted to the server are parsed into putative domains and structural models are generated using either comparative modeling or de novo structure prediction methods. If a confident match to a protein of known structure is found using BLAST, PSI-BLAST, FFAS03 or 3D-Jury, it is used as a template for comparative modeling. If no match is found, structure predictions are made using the de novo Rosetta fragment insertion method. Experimental nuclear magnetic resonance (NMR) constraints data can also be submitted with a query sequence for RosettaNMR de novo structure determination. Other current capabilities include the prediction of the effects of mutations on protein–protein interactions using computational interface alanine scanning. The Rosetta protein design and protein–protein docking methodologies will soon be available through the server as well.

METHODOLOGY:

1. Open homepage for Robetta (URL: <https://robetta.bakerlab.org/>)
2. Complete registration
3. Submit FASTA sequence for kinase.
4. Observe and interpret results.

OBSERVATION:

UniProt

UniProtKB

Advanced Search

[BLAST](#) [Align](#) [Retrieve/ID mapping](#) [Peptide search](#) [SPARQL](#) [Help](#) [Contact](#)

UniProt

The new UniProt website is here! [Take me to UniProt BETA](#)

UniProtKB - P02699 (OPSD_BOVIN)

Basket

Display

Help video

BLAST

Align

Format

Add to basket

History

Add a publication

Feedback

Entry

Publications

Feature viewer

Feature table

Protein

Gene

Organism

Status

Rhodopsin

RHO

Bos taurus (Bovine)

Reviewed - Annotation score: - Experimental evidence at protein level¹

Functionⁱ

Photoreceptor required for image-forming vision at low light intensity. Required for photoreceptor cell viability after birth (By similarity).
Light-induced isomerization of 11-cis to all-trans retinal triggers a conformational change that activates signaling via G-proteins (PubMed:10926528, PubMed:12044163, PubMed:11972040, PubMed:16908857, PubMed:16586416, PubMed:17060607, PubMed:17449675, PubMed:18818650, PubMed:21389983, PubMed:22198838, PubMed:23579341, PubMed:25205354, PubMed:27458239).
Subsequent receptor phosphorylation mediates displacement of the bound G-protein alpha subunit by the arrestin SAG and terminates signaling (PubMed:1396673, PubMed:15111114).

By similarity

5 Publications

11 Publications

Sites

Feature key	Position(s)	Description	Actions	Graphical view	Length
Site ¹	113	Plays an important role in the conformation switch to the active conformation	<div>3 Publications</div> <div>1 Publication</div>	<div></div>	1
Metal binding ¹	201	Zinc <div>Combined sources</div>	<div>2 Publications</div>	<div></div>	1
Metal binding ¹	279	Zinc <div>Combined sources</div>	<div>2 Publications</div>	<div></div>	1

GO - Molecular function¹

- 11-cis retinal binding

Source: UniProtKB
- arrestin family protein binding

Source: CAFA
- G-protein alpha-subunit binding

Source: CAFA
- G protein-coupled photoreceptor activity

Source: UniProtKB
- guanyl-nucleotide exchange factor activity

Source: UniProtKB
- identical protein binding

Source: IntAct
- opsin binding

Source: CAFA

Fig1. Result page for Rhodopsin in UniProt database

```
>sp|P02699|OPSD_BOVIN Rhodopsin OS=Bos taurus OX=9913 GN=RHO PE=1 SV=1
MNGTEGPNFYVPFSNKTGVVRSPFEAPQYYLAEPWQFSMLAAYMFLLIMLGFPINFLTLY
VTVQHKKLRTPLNYILLNLAVADLFMVFGGFTTTLYTSLHGYFVFGPTGCNLEGFFATLG
GEIALWSLVVLAIERVYVVKPMSNFRFGENHAIMGVAFTWVMALACAAPPLVGWSRYIP
EGMQCSCGIDYYTPHEETNNESFVIYMFVVHFIIP LIVIFFCYGQLVFTVKEAAAQQQES
ATTQKAEKEVTRMVIIMVIAFLICWLPYAGVAFYIFTHQGSDFGPIFMTIPAFFAKTSAV
YNPVIYIMMNKQFRNCMVTTLCCKGNPLGDDEASTTVSKTETSQVAPA
```

Fig2. FASTA sequence for Rhodopsin

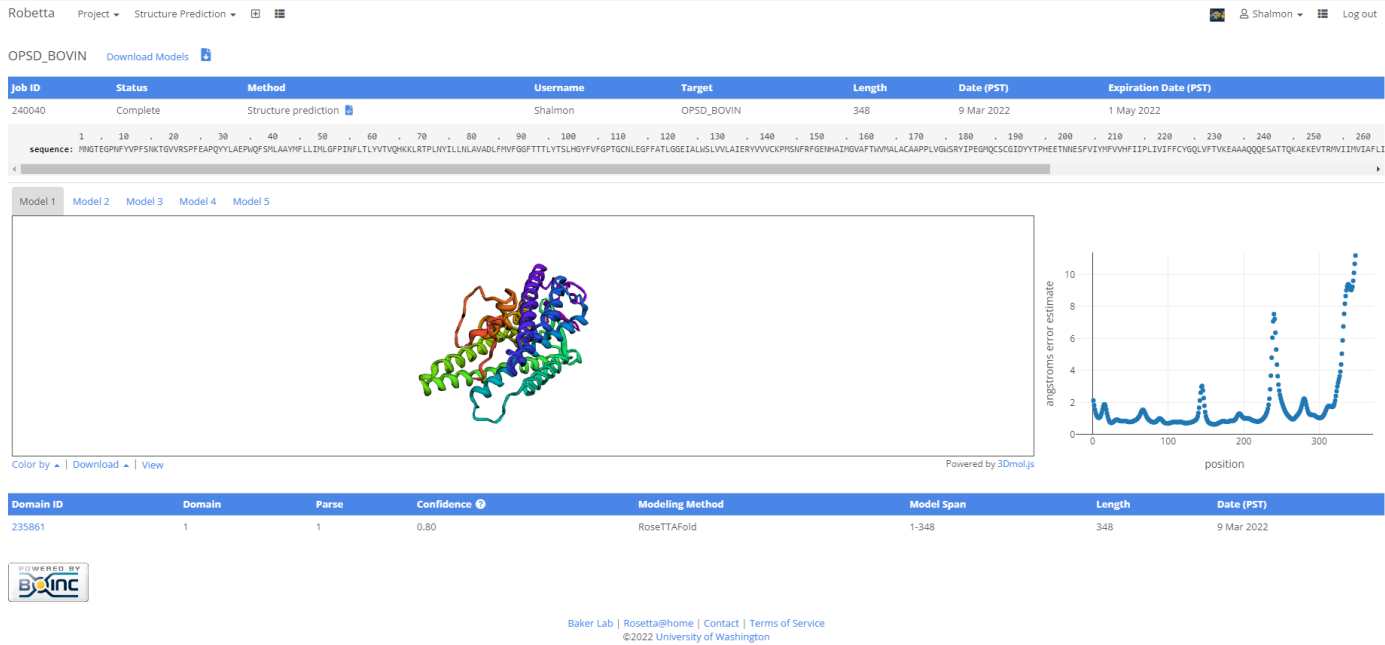


Fig3. Model 1 with atom co-ordinates

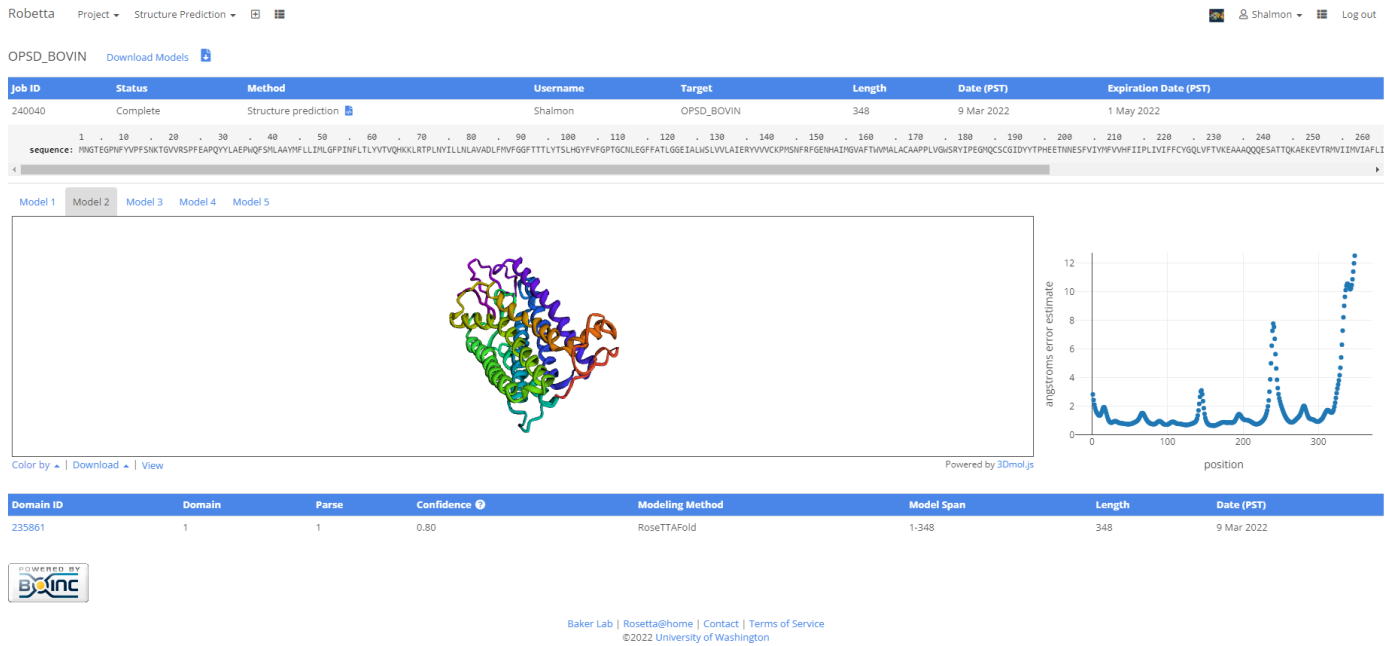


Fig4. Model 2 with atom co-ordinates



Fig5. Model 3 with atom co-ordinates

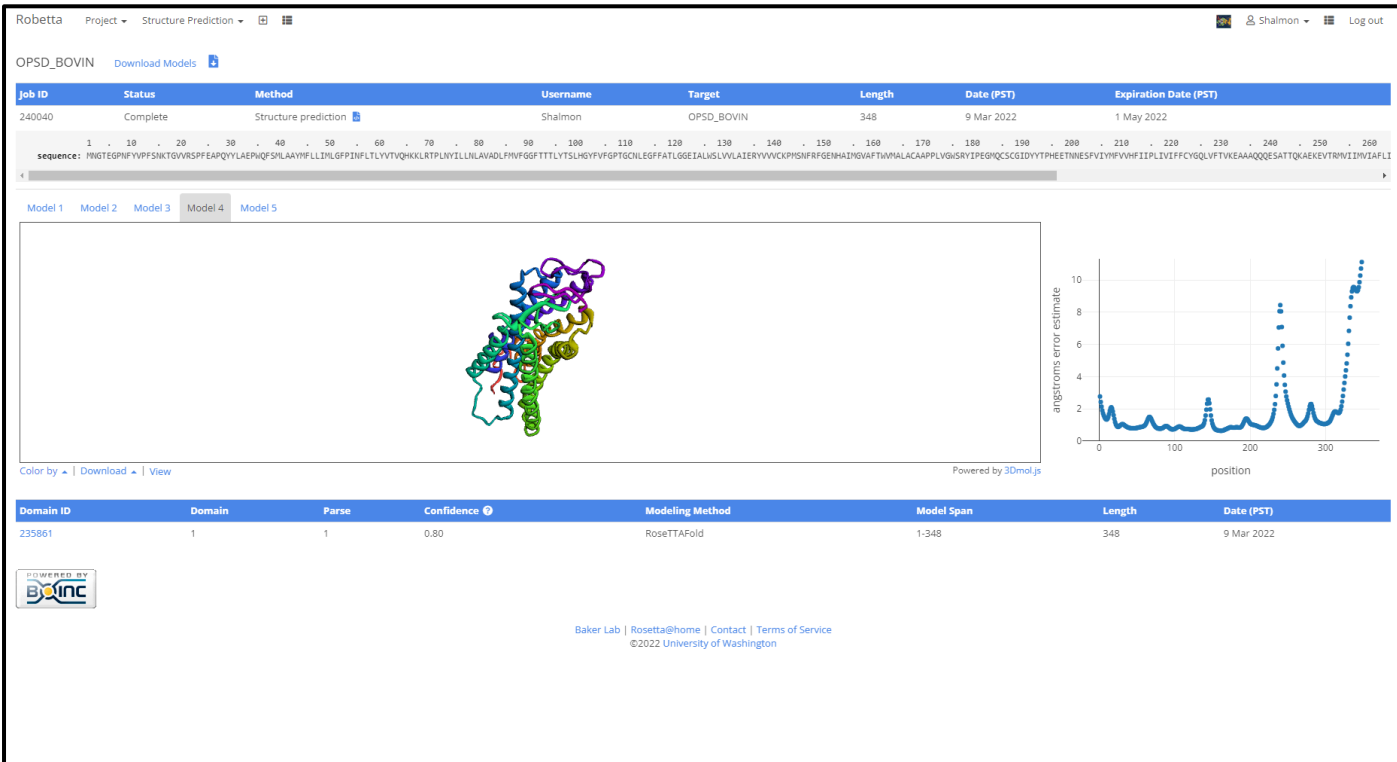


Fig6. Model 4 with atom co-ordinates

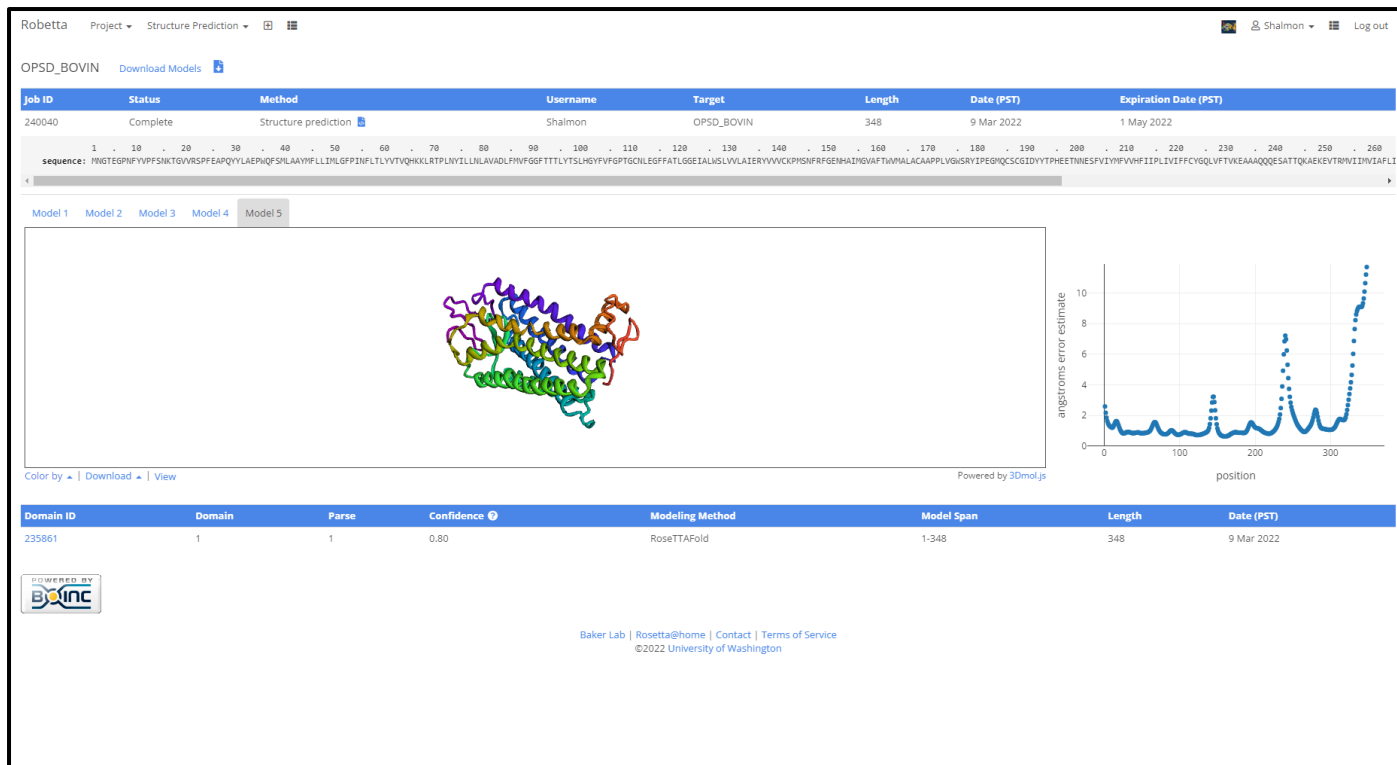


Fig7. Model 5 with atom co-ordinates

RESULT:

Robetta was used to predict the tertiary structure of Rhodopsin based on ab-initio approach.

CONCLUSION:

Thus, Robetta can be used to predict tertiary structures of proteins by ab-initio method. These tools give faster results than x-ray or NMR techniques and can be used by researchers to understand protein functions and drug designing.

REFERENCES:

1. Xiong, J. (2008). Tertiary structure prediction. Essential bioinformatics. Cambridge: Cambridge University Press. 214-228.
2. Encyclopædia Britannica, inc. (n.d.). *Rhodopsin*. Encyclopædia Britannica. Retrieved March 18, 2022, from <https://www.britannica.com/science/rhodopsin>
3. Robetta (2021b). Bakerlab.org. Retrieved March 18, 2022, from <https://robetta.bakerlab.org/results.php?id=240040>