


Protein homology modelling using MODELLER

Protein 1: PF3D7_0220800; cytoadherence linked asexual protein 2

Sequence

>PF3D7_0220800

<https://blast.ncbi.nlm.nih.gov/Blast.cgi#>



National Library of Medicine
National Center for Biotechnology Information

Log in

BLAST® » blastp suite » results for RID-S7Z1F9JM013

[Home](#) [Recent Results](#) [Saved Strategies](#) [Help](#)

[← Edit Search](#)

[Save Search](#)

[Search Summary ▾](#)

[How to read this report?](#) [BLAST Help Videos](#) [Back to Traditional Results Page](#)

Job Title

PF3D7_0220800

RID

S7Z1F9JM013 Search expires on 12-22 15:49 pm [Download All ▾](#)

Program

BLASTP [Citation ▾](#)

Database

nr [See details ▾](#)

Query ID

lcl|Query_426278

Description

PF3D7_0220800

Molecule type

amino acid

Query Length

1440

Other reports

[Distance tree of results](#) [Multiple alignment](#) [MSA viewer](#) [?](#)

Filter Results

Organism only top 20 will appear ☐ exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity

E value

Query Coverage

to

to

to

Filter

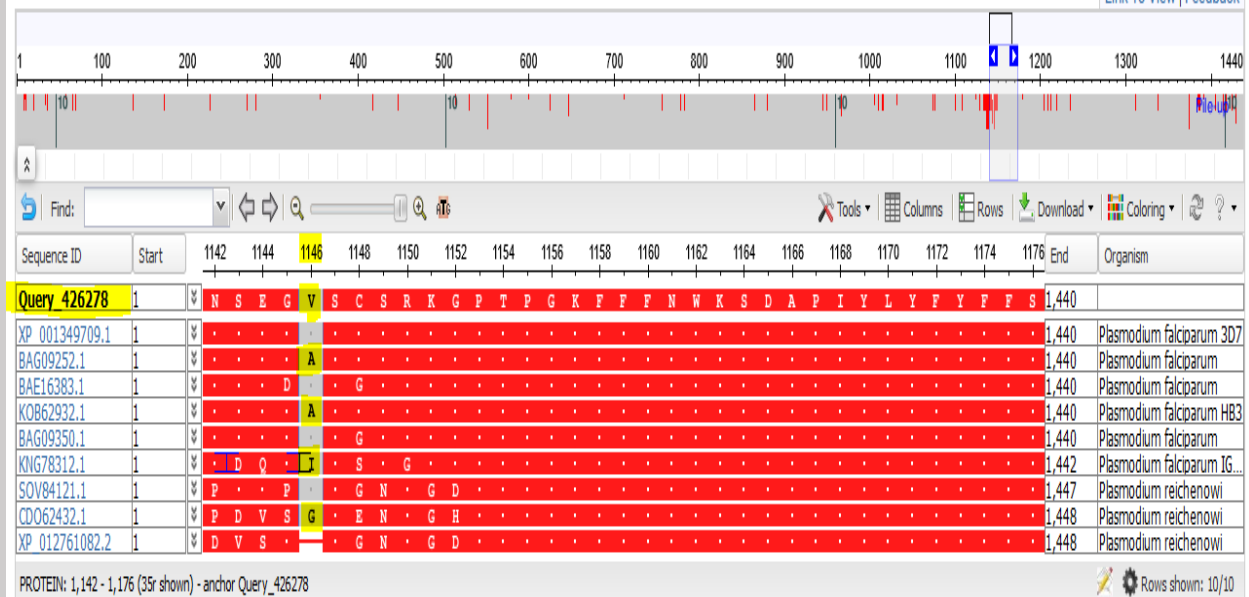
Reset

	Description	Scientific Name	Score	Score	Cover	value	Ident	Len	Accession
<input checked="" type="checkbox"/>	cytoadherence linked asexual protein 2 [Plasmodium falciparum 3D7]	Plasmodium falciparum 3D7	2939	2939	100%	0.0	100.00%	1440	XP_001349709.1
<input type="checkbox"/>	hypothetical protein PFMA1P_00366 [Plasmodium falciparum MaliPS096_E11]	Plasmodium falciparum MaliPS096_E11	2926	2926	100%	0.0	99.51%	1440	ETW51634.1
<input checked="" type="checkbox"/>	RhopH1/Clag2 [Plasmodium falciparum]	Plasmodium falciparum	2925	2925	100%	0.0	99.51%	1440	BAG09262.1
<input checked="" type="checkbox"/>	RhopH1/Clag2 [Plasmodium falciparum]	Plasmodium falciparum	2924	2924	100%	0.0	99.58%	1440	BAE16383.1
<input type="checkbox"/>	hypothetical protein PFAG_04640 [Plasmodium falciparum Santa Lucia]	Plasmodium falciparum Santa Lucia	2924	2924	100%	0.0	99.58%	1440	EUT80000.1
<input type="checkbox"/>	hypothetical protein C923_00384 [Plasmodium falciparum UGT5.1]	Plasmodium falciparum UGT5.1	2921	2921	100%	0.0	99.44%	1440	EWG78927.1
<input type="checkbox"/>	hypothetical protein PFDG_03837 [Plasmodium falciparum Dd2]	Plasmodium falciparum Dd2	2920	2920	100%	0.0	99.51%	1440	KOB88893.1
<input checked="" type="checkbox"/>	cytoadherence linked asexual protein 2 [Plasmodium falciparum HB3]	Plasmodium falciparum HB3	2920	2920	100%	0.0	99.38%	1440	KOB62932.1
<input type="checkbox"/>	hypothetical protein PFMC_00361 [Plasmodium falciparum CAMP/Malaysia]	Plasmodium falciparum CAMP/Malaysia	2919	2919	100%	0.0	99.38%	1440	ETW63802.1
<input checked="" type="checkbox"/>	RhopH1/Clag2 [Plasmodium falciparum]	Plasmodium falciparum	2912	2912	100%	0.0	99.17%	1440	BAG09350.1
<input checked="" type="checkbox"/>	RhopH1/Clag2 [Plasmodium falciparum IGH-CR14]	Plasmodium falciparum IGH-CR14	2904	2904	100%	0.0	99.03%	1442	KNG78312.1
<input type="checkbox"/>	hypothetical protein PFNF135_00384 [Plasmodium falciparum NF135/5_C10]	Plasmodium falciparum NF135/5_C10	2901	2901	100%	0.0	98.89%	1442	ETW45463.1
<input type="checkbox"/>	hypothetical protein PFBG_00220 [Plasmodium falciparum 7G8]	Plasmodium falciparum 7G8	2897	2897	100%	0.0	98.82%	1440	EUR81550.1
<input type="checkbox"/>	hypothetical protein PFFVO_00356 [Plasmodium falciparum Vietnam Oak-Knoll (FVO)]	Plasmodium falciparum Vietnam Oak-Knoll (FVO)	2895	2895	100%	0.0	98.68%	1442	ETW20801.1
<input type="checkbox"/>	cytoadherence linked asexual protein 2 [Plasmodium sp. gorilla clade G1]	Plasmodium sp. gorilla clade G1	2890	2890	100%	0.0	98.27%	1447	SOS76382.1
<input type="checkbox"/>	hypothetical protein PFUGPA_00364 [Plasmodium falciparum Palo Alto/Uganda]	Plasmodium falciparum Palo Alto/Uganda	2813	2813	100%	0.0	96.74%	1426	ETW57671.1
<input type="checkbox"/>	hypothetical protein PFFCH_00086 [Plasmodium falciparum FCH/4]	Plasmodium falciparum FCH/4	2810	2810	100%	0.0	96.60%	1396	ETW32493.1
<input checked="" type="checkbox"/>	cytoadherence linked asexual protein 2 [Plasmodium reichenowi]	Plasmodium reichenowi	2801	2801	100%	0.0	95.09%	1447	SOV84121.1
<input checked="" type="checkbox"/>	cytoadherence linked asexual protein 2 [Plasmodium reichenowi]	Plasmodium reichenowi	2793	2793	100%	0.0	95.03%	1448	CDQ62432.1
<input checked="" type="checkbox"/>	cytoadherence linked asexual protein 2 [Plasmodium reichenowi]	Plasmodium reichenowi	2790	2790	100%	0.0	95.10%	1448	XP_012761082.2

Alignment

[Link To View](#) | [Feedback](#)

Alignment

[Link To View](#) | [Feedback](#)

Rows selection ✕											
Anchor row Query_426278 cannot be hidden											
<input checked="" type="checkbox"/>	Name	Sequence ID	Organism	Date	Country	Host	Source	Gene	Identity	Coverage	Mismatches
<input checked="" type="checkbox"/>	cytoadherence linked asexual protein 2 [...]	XP_001349709.1	Plasmodium falciparum 3D7					PF3D...	100	100	0
<input checked="" type="checkbox"/>	RhopH1/Clag2 [Plasmodium falciparum]	BAG09252.1	Plasmodium falciparum (ma...		Honduras				99.51	100	7
<input checked="" type="checkbox"/>	RhopH1/Clag2 [Plasmodium falciparum]	BAE16383.1	Plasmodium falciparum (ma...						99.58	100	6
<input checked="" type="checkbox"/>	cytoadherence linked asexual protein 2 [...]	KOB62932.1	Plasmodium falciparum HB3						99.38	100	9
<input checked="" type="checkbox"/>	RhopH1/Clag2 [Plasmodium falciparum]	BAG09350.1	Plasmodium falciparum (ma...		Brazil				99.17	100	12
<input checked="" type="checkbox"/>	RhopH1/Clag2 [Plasmodium falciparum IG...	KNG78312.1	Plasmodium falciparum IGH...						99.03	100	12
<input checked="" type="checkbox"/>	cytoadherence linked asexual protein 2 [...]	SOV84121.1	Plasmodium reichenowi						95.09	100	64
<input checked="" type="checkbox"/>	cytoadherence linked asexual protein 2 [...]	CDO62432.1	Plasmodium reichenowi						95.03	100	64
<input checked="" type="checkbox"/>	cytoadherence linked asexual protein 2 [...]	XP_012761082.2	Plasmodium reichenowi	13-Feb-2...	Cameroo...	Pan trogl...	peripher...	PRSY...	95.1	99.93	61

Prepared by Najneen Rejwanda

Protein 1: PF3D7_0220800; cytoadherence linked asexual protein 2

Sequence Protein 1: PF3D7_0220800; cytoadherence linked asexual protein 2

Sequence

>PF3D7_0220800

MVSSVKSSLFLLIFFLYLKKNVICSINDNVNENITEGLDEYEFGNENINESITENVNVNVTENEKDNLIYDDNNNIEEL
KSMIGNDELHKNLSILEKLILDSLKKDKLKLPLIKEGTEEYLDISKFKKKILTDSDDKTYILPTLESSFYDITKYEHILKEQLIE
EYNSKISDAVKKKLLIVRTLKTIKMLIPLNAYKEKNDLKIALEELNNVITHRTYETLKKSPIENPGEFFRKLTHVKEVKE
SKEIENKGEYLILGNDKIEIMDAHDFFFTTNSNIKFMETLDSISNQYGLGLINDLGPHLIALGHFMVLKALKNYKNYF
EAKNTKFFSWQKILEFSLTDRFKILDMMCDHDVVYYSQDKRRKTYLNVDTSGSSMECNILEFLIHYFNKYQLEIIKAT
QDTEFLHGMMEHKNIKDYFFSFCNDPKECIYHTNQFKKEAKEENTFPEEPNREISAYNLYNYYYFMKRYSSYG
IKKTLVHLLNLTGLLNYDTRSIVTSLYLPGYNNVEMSFTEDVEFTTLFNNLLKCIKKCHKEETNTNTSLMDSNSSHN
YFLHEITKCDLCKGAFLYSNMKFEEVPSMLQKFYIYLTEGLRIQKVSLFMKTLDIYQDYSNFLSHDINWYTFLLFRLLTS
FEEISKSVGEAMYLNIQDEDSFHKITITTNYWFPSPIKKYTYLVVRKHLPNLLDELEKLMKSSTLEKMKKSINFLVHV
NSFLQLDFFHQLNEPPVGLPRSYPLSLILEHKFKWMMNSSPAGFYFSNYHNPYIRKELHRKVLTEKFEPPKMKNKWNE
VLKSIECAYDMYFEQRHVKNLYKNHNIYNINNKIMLMRDSVDLYKKNFKDVIFFADIFNLRKYLTAATPLIKKTWDR
MYFYIRNTGNSVNFYKYGIYGFKINKVYLKEVVDELYSIYNFNTDIFSDTSFLQTVYLLFRKIEDSYRTHRRNDHIGV
NNIFFMNVANNYSKLNNEEREMEIHNSMASRYYSKTMFAAFQMLFSTMLSNDANNLDKVYKSSNIQVATSTTA
FLTFAVYNGSIMDSLNRLLPPYAKKPITQLKYGKTFVFSNYFMLASQIYEMLNKYKNSLLCEYQAVASANYYSAKK
LGQFVGRKYFPLTTYLSLRIRASYGWVHGTETKICNSEGVSCSRKGPTPGKFFFNWKSDAPIYLYFYFFSNLYLDSAK
YFPGGFSTSLKEQTEHVSQKGFKKKPMVHELTKNLILDVTNGFMYAFCFYSIMPLYAYFENVNFYIISNFRFLDRYYN
AFNKYFINFFKTKLKKYTTDVFIKYEYDAYTSMKKYGYLNEVIGSRLSSKNRIVKYIYDSNDDIMNNLRRYDMENRFR
NKMSTYVDEYAFFDDCGKNEVFLNDRCDYCPIVEDLCEPDTKEYQPHTSNIQKVTDKNTTYINYEKLHEESYSQETQ
SDNTDDEKDNDLPDTELMITRLQ

[Edit Search](#)
[Save Search](#)
[Search Summary](#)
[How to read this report?](#)
[BLAST Help Videos](#)
[Back to Traditional Results Page](#)

Job Title **PF3D7_0220800**
RID [S84409MJ016](#) Search expires on 12-22 17:16 pm [Download All](#)
Program BLASTP [Citation](#)
Database pdb [See details](#)
Query ID lcl|Query_16583
Description PF3D7_0220800
Molecule type amino acid
Query Length 1440
Other reports [Distance tree of results](#) [Multiple alignment](#) [MSA viewer](#)

Filter Results

Organism only top 20 will appear ☐ exclude

[Add organism](#)

Percent Identity to
E value to
Query Coverage to

[Filter](#)
[Reset](#)

[Descriptions](#)
[Graphic Summary](#)
[Alignments](#)
[Taxonomy](#)

Sequences producing significant alignments

[Download](#) [Select columns](#) Show

☒ select all 2 sequences selected
[GenPept](#)
[Graphics](#)
[Distance tree of results](#)
[Multiple alignment](#)
[MSA Viewer](#)

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	Chain A_Cytoadherence linked asexual protein 3.1 [Plasmodium falciparum NF54]	Plasmodium falciparum NF54	2141	2141	100%	0.0	71.69%	1417	7MRW_A
<input checked="" type="checkbox"/>	Chain A_Cytoadherence linked asexual protein 3 [Plasmodium falciparum]	Plasmodium falciparum	2127	2127	100%	0.0	71.21%	1505	7KIY_A

7MRW_A = 71.69% = **reference_proteinA** = <https://www.rcsb.org/structure/7MRW>

7KIY_A = 71.22% = **reference_proteinB** = <https://www.rcsb.org/structure/7KIY>

= ##### = **reference_proteinC**

We need reference proteins, which is why we used BLASTP

Scripts

For using Modeller, we need 6 scripts

<https://salilab.org/modeller/tutorial/basic.html>

- 1) main_protein.ali
- 2) build_profile.py
- 3) compare.py
- 4) align2d.py
- 5) model-single.py
- 6) evaluate_model.py (optional)

main_protein.ali

>P1; main_protein

sequence: main_protein:::0.00: 0.00

MVSSVKSSLFLLIFFLYLKKNVICSINDNVNENITEGLDEYEFGNENINESITENVNVNVTENEKDNLIYNDNNNIEEL
KSMIGNDELHKNLSILEKLILDSLKKDKLKLPLIKEGTEEYLDISKFKKKILTSDDKTYILPTLESSFYDITKYEHILKEQLIE
EYNSKISDAVKKLLIVRTLKTIKMLIPLNAYKEKNDLKIALEELNNVITHRTYETLKKSPIENPGEFFRKLTHVKEVKE
SKEIENKGEYLILGNDKIEIMDAHDDFFTTNSNIKFMETLDSISNQYGLGLINDLGPHLIALGHFMVLKALKNYKNYF
EAKNTKFFSWQKILEFSLTDRFKILDMMCDHDVVYYSQDKRRKTYLNVDTSGSSMECNILEFLIHYFNKYQLEIIKAT
QDTEFLHGMMEHKNIKDYFFSFCNDPKECIYHTNQFKKEAKEENTFPEEPNREISAYNLNLYNYFFMKRYSSYG
IKKTLYVHLLNLTGLLNYDTRSYVTSYLPGYNNVEMSFTEDVEFTTLFNNLLKCIKKCHKEETNTNTSLMDSNSSHN
YFLHEITKCDLCKGAFLYSNMKFEEVPSMLQKFYIYLTEGLRIQKVSFLMKTLDIYQDYSNFLSHDINWYTFRLFRLTS
FEEISKSVGEAMYLNIQDEDSFHKITTNYWFPSPIKKYTYLVRKHLPNLLDELEKLMKSSTLEKMKKSINFLVHV
NSFLQLDFFHQLNEPPVGLPRSYPLSLILEHKFKWMMNSSPAGFYFSNYHNPYIRKELHRKVLTEKFEPPKMKNWNE
VLKSLIECAYDMYFEQRHVKNLYKNHNIYNINNKIMLMRDSVDLYKKNFKDVIFFADIFNLKRYLTATPLIKKTWDR
MYFYIRNTGNSVNFYKYGIYGFKINKVYLKEVDELYSIYNFNTDIFSDTSFLQTVYLLFRKIEDSYRTHRRNDHIGV
NNIFFMMNVANNYSKLNNEEREMEIHNMSASRYYSKTMFAAFQMLFSTMLSNDANNLDKVYKGSSNIQVATSTTA
FLTFAVYVNGSIMDSLNRLLPPYAKKPITQLKYGKTFVFSNYFMLASQIYEMLNKYKNSLLCEYQAVASANYYSAKK
LGQFVGRKYFPLTTYLSLRIRASYGWVHGTETKICNSEGVSCSRKGPTPGKFFFNWKSDAPIYLYFYFFSNLYLDSAK
YFPGGFSTSLKEQTEHVSQKGFKKKPMVHELTKNLILDVTNGFMYAFCFYSIMPLYAYFENVNFYIISNFRFLDRYYN
AFNKYFINFFKTKLKKYTTDVFIKYEYDAYTSMKKYGYLNEVIGSRLSSKNRIVKYIYDSNDDIMNNLRRYDMENRFR
NKMSTYVDEYAFFDDCGKNEVFLNDRCDYCPIVEDLCEPDTKEYQPHTSNIQKVTDKNTTYINYEKLHEESYSQETQ

SDNTDDEKDNDLPDTELMITRLQ*

2) build_profile.py

```
from modeller import *

log.verbose()
env = Environ()

#-- Prepare the input files

#-- Read in the sequence database
sdb = SequenceDB(env)
sdb.read(seq_database_file='pdb_95.pir', seq_database_format='PIR',
        chains_list='ALL', minmax_db_seq_len=(30, 4000),
        clean_sequences=True)

#-- Write the sequence database in binary form
sdb.write(seq_database_file='pdb_95.bin', seq_database_format='BINARY',
        chains_list='ALL')

#-- Now, read in the binary database
sdb.read(seq_database_file='pdb_95.bin', seq_database_format='BINARY',
        chains_list='ALL')

#-- Read in the target sequence/alignment
aln = Alignment(env)
aln.append(file='main_protein.ali', alignment_format='PIR',
        align_codes='ALL')

#-- Convert the input sequence/alignment into
#   profile format
prf = aln.to_profile()

#-- Scan sequence database to pick up homologous sequences
prf.build(sdb, matrix_offset=-450, rr_file='${LIB}/blosum62.sim.mat',
        gap_penalties_1d=(-500, -50), n_prof_iterations=1,
        check_profile=False, max_aln_evalue=0.01)

#-- Write out the profile in text format
prf.write(file='build_profile.prf', profile_format='TEXT')

#-- Convert the profile back to alignment format
aln = prf.to_alignment()

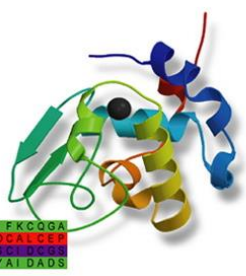
#-- Write out the alignment file
aln.write(file='build_profile.ali', alignment_format='PIR')
```

For pdb_95.bin >>

← → ↻ https://salilab.org/modeller/tutorial/basic.html

Modeller

Program for Comparative Protein Structure Modelling by Satisfaction of Spatial Restraints



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

- About MODELLER
- MODELLER News
- Download & Installation
 - Release Notes
 - Data file downloads
- Registration
 - Non-academic use
- Discussion Forum
 - Subscribe
 - Archives
- Documentation

Tutorial

Basic example: Modeling lactate dehydrogenase from *Trichomonas vaginalis* based on a single template.

All input and output files for this example are available to download, in either [zip format \(for Windows\)](#) or [tar.gz format \(for Unix/Linux\)](#).

A novel gene for lactate dehydrogenase was identified from the genomic sequence of *Trichomonas vaginalis* (TvLDH). The corresponding protein has a higher similarity to the malate dehydrogenase of the same species (TvMDH) than to any other LDH. We hypothesized that TvLDH arose from TvMDH by convergent evolution relatively recently. Comparative models were constructed for TvLDH and TvMDH to study the sequences in the structural context and to suggest site-directed mutagenesis experiments for elucidating specificity changes in this apparent case of convergent evolution of enzymatic specificity. The native and mutated enzymes were expressed and their activities were compared.

The individual modeling steps of this example are explained below. Note that we go through every step in this tutorial to build a model leaving only

3_compare.py

```
from modeller import *

env = Environ()
aln = Alignment(env)
for (pdb, chain) in (('reference_protein1_id', 'A'),
                    ('reference_protein2_id', 'A'), ('reference_protein3_id', 'A')):
    m = Model(env, file=pdb, model_segment=('FIRST:'+chain, 'LAST:'+chain))
    aln.append_model(m, atom_files=pdb, align_codes=pdb+chain)
aln.malign()
aln.malign3d()
aln.compare_structures()
aln.id_table(matrix_file='family.mat')
env.dendrogram(matrix_file='family.mat', cluster_cut=-1.0)
```

In compare.log file

Sequence identity comparison (ID_TABLE):

Diagonal ... number of residues;
Upper triangle ... number of identical residues;
Lower triangle ... % sequence identity, id/min(length).

	1b8pA	@11bdmA	@11civA	@25mdhA	@27mdhA	@21smkA	@2
1b8pA @1	327	194	147	151	153	49	
1bdmA @1	61	318	152	167	155	56	

1civA @2	45	48	374	139	304	53
5mdhA @2	46	53	42	333	139	57
7mdhA @2	47	49	87	42	351	48
1smkA @2	16	18	17	18	15	313

Weighted pair-group average clustering based on a distance matrix:



4_align2d.py

```
from modeller import *

env = Environ()
aln = Alignment(env)
mdl = Model(env, file='best_model_id', model_segment=('FIRST:A','LAST:A'))
aln.append_model(mdl, align_codes='best_model_id',
atom_files='best_model_id.pdb')
aln.append(file='main_protein.ali', align_codes='main_protein')
aln.align2d(max_gap_length=50)
aln.write(file='main_protein-best_model_id.ali', alignment_format='PIR')
aln.write(file='main_protein-best protein id.pap', alignment_format='PAP')
```

5_model_single.py

```
from modeller import *
from modeller.automodel import *
#from modeller import soap_protein_od

env = Environ()
a = AutoModel(env, alnfile='main_protein-best_model_id.ali',
```

```
knowns='best_model_idA', sequence='main_protein',  
assess_methods=(assess.DOPE,  
                 #soap_protein_od.Scorer(),  
                 assess.GA341))  
  
a.starting_model = 1  
a.ending_model = 5  
a.make()
```

Prepared by Najneen Rejwana

commands:

2) mod10.4 2_build_profile.py

```
/home/dinesh/Documents/Rejwana/MODELLER/iqbal_malaria/build_profile.ali - dinesh@192.168.15.91 - Editor - WinSCP

>P1;main_protein
sequence:main_protein: 0: : 0: ::-1.00:-1.00
MVSSVKSSLFLLIFFLYLKKNVICSINDNVNENITEGLDEYEFGNENINESITENVNVNVTENEKDNLIYNDNN
NIEELKSMIGDELHKNSILEKLILDSLKKDKLKLPLIKEGTEEYLDISKFKKKILTSDDKTYILPTLESSFY
DITKYEHLKEQLIEEYNSKISDAVKKLLIVRTLKTIKMLIPLNAYKEKNDLKIALEELNNVITHRTYETLKK
SPIENPGEFFRKLLTHVKEVKESKEIENKGEYLILGNDKIEIMDAHDFFTTNSNIKFMETLDSISNQYGLGLIN
DLGPHLIALGHFMVLKALKKNYFEAKNTKFFSWQKILEFSLDRFKILDMMCDHDVVVYSQDKRRKTYLNV
TSGSSMECNILEFLIHFNKYQLIEIKATQDTDFELHGMMEHKNIKDYFFSFCNDPKECIIYHTNQFKKEAKEE
NTFPEEPNREISAYNLYLNYFFMKRYSSYGIKKTLVHLLNLTGLLNDTRSYVTSLYLPGYNVVEMSFTEDV
EFTTLFNLLKCIKKCHKEETNTNTSLMDSNSSHYFLHEITKCDLCKGAFLYSNMKFEEVPSMLQKFYIYLTG
LRIQKVSFLMKTLDIYQDYSNFLSHDINWYTFLLFRLTSFEEISKKSVEGAMYLNIQDEDSFHKTITTYWFP
PIKKYTYLYVRKHLNNLLDELEKLMKSSLEKMKKSINFLVHVSFLQLDFFHQLNEPPVGLPRSYPLSLILEH
KFKEWMNSSPAGFYFSNYHNPYIRKELHRKVLTEKFEPKMNKWEVLKSLIECAYDMYEQRHVKNLKYNHNIY
NINNKMIMRSDVLYKKNFKDVIFFADIFNLRKYLTAPLIKKTWDRMYFYIYRNTGNSVNFYKYGIYGFKN
KVYLKEVWDELYSIYNFNTDIFSOTSFLQTVYLLFRKIEDSYRTHRRNDHIGVNNIFFMNVANNYSKLNEEREM
EIHNSMASRYYSKTMFAAFQMLFSTMLSDANNLDKVYKSSNIQVATSTTAFLTFAYVYNGSIMDSLNRLLPP
YAKKPITQLKYGKTFVFSNYFMLASQIYEMLNKNSLLCEYQAVASANYYSAKKLGQFVGRKYFPLTTYLSLR
IRASYGWVHGTETKICNSEGVSCSRKGPTPGKFFFNWKSADAPYLYFYFFSNLYLDSAKYFPGGFSTSLKEQTEH
VSQKGFKKPMVHELTKNLILDVTNGFMAYFCFYSIMPLYAYFENVNFYIISNFRFLDRYYNAFNKYFINFFKTK
LKKYTTDVFICYEYDAYTSMKKYGYLNEVIGSRLSSKNRIVKYIYDSNDDIMNNLRYYDMENRFRNKMSTYVDEY
AFFDDCGKNEVFLNDRCDYCPIVEDLCEPDTKEYQPHTSNIQKVTDKNTTYINYEKLHEESYSQETQSDNTDDEK
DNDLPDTELMITRLQ*
```

```
a$ mod10.4 2_build_profile.py
Could not find platform independent libraries <prefix>
Could not find platform dependent libraries <exec_prefix>
Consider setting $PYTHONHOME to <prefix>[:<exec_prefix>]
'import site' failed; use -v for traceback
```

Note:

#Here, mod10.4 is the version of modeler, it can be different based on the different version of modeler.

#reference protein name and scripts must be same

main_protein.ali

main_protein

```
main_protein 12/21/2023 6:32 PM ALI File 2 KB
main_protein - Notepad
File Edit Format View Help
>P1; main_protein
sequence: main_protein:::0.00: 0.00
MVSSVKSSLFLLIFFLYLKKNVICSINDNVNENITEGLDEYEFGNENINESITENVNVNVTENEKONLIYNDNNNIEEL
KSMIGNDELHKNLSILEKLILDSLKKDKLKLPLIKEGTEEYLDISKFKKKILTSDDKTYILPTLESSFYDITKYEHILKEQLIE
EYNSKISDAVKKKLLIVRTLKTIKMLIPLNAYKEKNDLKIAEELNNVITHRTYETLKKSPIENPGEFFRKLTHVKEVKE
SKEIENKGEYLILGNDKIEIMDAHDFFTTNSNIKFMETLDSISNQYGLGLINDLGPHLIALGHFMVLKLALKNYKNYF
EAKNTKFFSWQKILEFSLTDRFKILDMMCDHVDVYYSQDKRRKTYLNVDTSGSSMECNILEFLIHYFNKYQLEIKAT
QDQDFELHGMMEHKNIKDYFFSFCNDPKECIIYHTNQFKKEAKEENTFPEEPNREISAYNLYLNYYYFMKRYSSYG
IKKTLTYHLLNLTGLLNYDTRSYVTSYLYLPGYYNVVEMSFTEDVEFTTLFNNLLKCIKKCHKEETNTNTSLMDSNSSHN
YFLHEITKCDLCKGAFLYSNMKFEEVPSMLQKFYIYLTGLRIQKVSFLMKTLDIYQDYSNFLSHDINWYTFLLFLRLTS
FEEISKKSVEAMYLNIQDEDSFHKITTNYWFPSPIKKYTYLYVRKHLPNLLDELEKLMKSSTLEKMKKSINFLVHV
NSFLQLDFFHQLEPPVGLPRSYPLSLILEHKFKWMMNSSPAGFYFSNYHNPYIRKELHRKVLTEKFEPKMNKWNE
```

3) mod10.4 3_compare.py

```
/home/dinesh/Documents/Rejwana/MODELLER/3_compare.py -
from modeller import *

env = Environ()
aln = Alignment(env)
for (pdb, chain) in (('7mrw', 'A'), ('7kiy', 'A'), ('AF-096279-F1-model_v4', 'A')):
    m = Model(env, file=pdb, model_segment=('FIRST:'+chain, 'LAST:'+chain))
    aln.append_model(m, atom_files=pdb, align_codes=pdb+chain)
aln.malign()
aln.malign3d()
aln.compare_structures()
aln.id_table(matrix_file='family.mat')
env.dendrogram(matrix_file='family.mat', cluster_cut=-1.0)
```

```
(base) $ mod10.4 3_compare.py
Could not find platform independent libraries <prefix>
Could not find platform dependent libraries <exec_prefix>
Consider setting $PYTHONHOME to <prefix>[:<exec_prefix>]
'import site' failed; use -v for traceback
```

Note: minimum 3 PDBs are required. Otherwise phylogenetic tree may not generate.

Check 3_compare.log

	7mrwA	@37kiyA	@2AF-096@1
7mrwA @3	1209	1013	856
7kiyA @2	87	1169	892
AF-096@1	71	76	1440

Weighted pair-group average clustering based on a distance matrix:

```

      .--- 7mrwA @3.7    13.0000
      |
      |----- 7kiyA @2.9    26.5000
      |
      |----- AF-096@1.0

+---+---+---+---+---+---+---+---+---+---+
27.0400  24.6100  22.1800  19.7500  17.3200  14.8900  12.4600
      25.8250  23.3950  20.9650  18.5350  16.1050  13.6750

```

Here, 7kiyA is the best template model because of low(est) crystallographic resolution (2.9A) & highest similarity (26.5%)

4) mod10.4 4_align2d.py

```

/home/dinesh/Documents/Rejwana/MODELLER/.../4_align2d.py - ...
[Icons] [ab] [ac] [Icons] Encoding [v] Color [v] [?]

from modeller import *

env = Environ()
aln = Alignment(env)
mdl = Model(env, file='7kiy', model_segment=('FIRST:A','LAST:A'))
aln.append_model(mdl, align_codes='7kiyA', atom_files='7kiy.pdb')
aln.append(file='main_protein.ali', align_codes='main_protein')
aln.align2d(max_gap_length=50)
aln.write(file='main_protein-7kiy.ali', alignment_format='PIR')
aln.write(file='main_protein-7kiy.pap', alignment_format='PAP')

```

\$ mod10.4 4_align2d.py

```
Could not find platform independent libraries <prefix>  
Could not find platform dependent libraries <exec_prefix>  
Consider setting $PYTHONHOME to <prefix>[:<exec_prefix>]  
'import site' failed; use -v for traceback
```

5) mod10.4 5_model_single.py

```
from modeller import *  
from modeller.automodel import *  
#from modeller import soap_protein_od  
  
env = Environ()  
a = AutoModel(env, alnfile='main_protein-7kiy.ali',  
               knowns='7kiyA', sequence='main_protein',  
               assess_methods=(assess.DOPE,  
                               #soap_protein_od.Scorer(),  
                               assess.GA341))  
  
a.starting_model = 1  
a.ending_model = 5  
a.make()
```

\$ mod10.4 5_model_single.py

```
Could not find platform independent libraries <prefix>  
Could not find platform dependent libraries <exec_prefix>  
Consider setting $PYTHONHOME to <prefix>[:<exec_prefix>]  
'import site' failed; use -v for traceback
```

Check >> 5_model_single.log

```
>> Summary of successfully produced models:  
Filename                                molpdf      DOPE score    GA341 score  
-----  
main_protein.B99990001.pdb             14399.21387  -161037.59375  1.00000  
✓ main_protein.B99990002.pdb             13218.66992  -163234.59375  1.00000  
main_protein.B99990003.pdb             14407.94727  -160762.10938  1.00000  
main_protein.B99990004.pdb             15190.11719  -160926.57812  1.00000  
main_protein.B99990005.pdb             14338.67578  -161750.53125  1.00000
```

Note: Least DOPE score indicates the best model

6) mod10.4 6_evaluate_model.py

```
evaluate_model - Notepad
File Edit Format View Help
from modeller import *
from modeller.scripts import complete_pdb

log.verbose()    # request verbose output
env = Environ()
env.libs.topology.read(file='$(LIB)/top_heav.lib') # read topology
env.libs.parameters.read(file='$(LIB)/par.lib') # read parameters

# read model file
mdl = complete_pdb(env, 'main_protein.B999990002.pdb')

# Assess with DOPE:
s = Selection(mdl) # all atom selection
s.assess_dope(output='ENERGY_PROFILE NO_REPORT', file='TvLDH.profile',
              normalize_profile=True, smoothing_window=15)
```

```
[C:\Users\user\AppData\Local\Temp\mod10.4_6_evaluate_model.py]
$ mod10.4 6_evaluate_model.py

Could not find platform independent libraries <prefix>
Could not find platform dependent libraries <exec_prefix>
Consider setting $PYTHONHOME to <prefix>[:<exec_prefix>]
'import site' failed; use -v for traceback
```