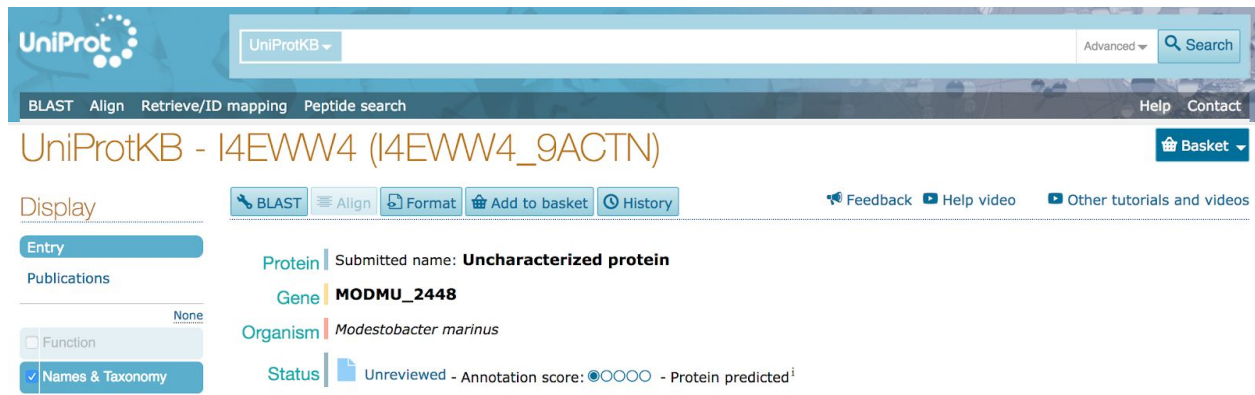


Modeling proteins based on homology

Exercise 1: Homology detection & structure prediction of an uncharacterized protein from *Modestobacter marinus* using HHpred

1. Retrieve the query sequence:



The screenshot shows the UniProtKB interface for the protein entry I4EWW4 (I4EWW4_9ACTN). The page includes a search bar at the top with the UniProtKB logo and a search button. Below the search bar, there are navigation links for BLAST, Align, Retrieve/ID mapping, and Peptide search. The main content area displays the protein name and accession number, along with a 'Basket' button. On the left, there are tabs for 'Display', 'Entry', 'Publications', 'Function', and 'Names & Taxonomy'. The 'Entry' tab is selected, showing the following information:

- Protein** Submitted name: **Uncharacterized protein**
- Gene** **MODMU_2448**
- Organism** *Modestobacter marinus*
- Status** Unreviewed - Annotation score: ●○○○○ - Protein predictedⁱ

You will work with a uncharacterized protein from *Modestobacter marinus*. Download the protein sequence of in fasta format and save into a directory.

Open <http://www.uniprot.org/> in your browser, and search for the gene name *MODMU_2448*. (Uniprot Id : *I4EWW4*)

2. Find template structures:

Standard Protein BLAST

blastn **blastp** blastx tblastn tblastx

BLASTP programs search protein databases using a protein query. [more...](#) [Reset page](#) [Bookmark](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

[Clear](#) **Query subrange**

From

To

Or, upload file No file selected.

Job Title

Enter a descriptive title for your BLAST search

Align two or more sequences

Choose Search Set

Database Protein Data Bank proteins(pdb)

Organism

Optional Enter organism name or id—completions will be suggested **Exclude**

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Exclude Models (XM/XP) Uncultured/environmental sample sequences

Optional

Entrez Query [YouTube](#) [Create custom database](#)

Optional Enter an Entrez query to limit search

We can template structures by performing a BLAST search against the Protein Data Bank 'PDB'.

Open the Ncbi Blast page in your browser.

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

Paste your FASTA sequence into the box

Enter a title and use default parameters

Choose 'pdb' as your search database

Run BLAST

Did you find any templates?.

[Hint: There are no templates available in PDB for this protein].

3. Find template structures by fold recognition:

Now we can try to use 'fold recognition' methods to find potential template structures.

We are going to use the '**HHpred**' server to search for protein folds. HHpred was used for sequence and structure database searching to find remote homologs.

HHpred - Homology detection & structure prediction by HMM-HMM comparison

[Help](#)

Input

Paste protein sequence or MSA

Search with pancreatitis induced protein **PIP49**.

or upload a local file

No file selected.

Select input format

FASTA

Search Options

Select HMM databases
(hold Ctr to select several)

Standard

pdb70_04Mar17
pdb_on_hold_7Mar17
mmcif70_Dec16
SCOPe95_2.06
SCOPe70_2.06

Proteomes

Arabidopsis_thaliana
Caenorhabditis_elegans
Drosophila_melanogaster
Homo_sapiens
Mus_musculus

MSA Generation Method

HHblits Psiblast

Max. MSA Generation iterations

3

Score secondary structure

yes no predicted vs predicted only

Alignment mode

local global

Realign with MAC

Open the HHpred web server.

<https://toolkit.tuebingen.mpg.de/hhpred>

Past your target sequence in fasta format and choose default parameters for other options. Choose **pdb70** for the option 'Select HMM Databases'. Submit Job and wait for the results.

4. HHpred - Results

Analyze the results. The output is organized into three sections:

1. A bar graph summarizing the position
2. A tabular hit list with probabilities, E-values, scores, and match regions in query and templates.
3. The pairwise query-templates alignments

It is not possible to identify these templates by just sequence comparison. Note the poor percent identity of the alignment for the top hits.

5. Build model

From the HHpred results page click Results ---> Create model.

HHpred - Results Job-ID: 2441985 Date: 07:48 on Mar 08 2017 [Help](#)

Results Histogram Reduced alignment Representative alignment Full alignment

Create model Merge Q/T alignments Forward to PCoils Forward MSA Save Export

1. Manual template selection: Generate a PIR-alignment of your sequence with the selected template or templates in order to build a 3D model with MODELLER.

Create model from manually selected template(s)

2. Automatic template selection: Optimize diversities of query and template HMMs, rerank templates and automatically select best set. In further steps a multiple alignment is created from this set, and a 3D model is build with MODELLER using this alignment.

Automatically select best templates

Only hits found in PDB or SCOP can be used to create a model (other hits are disabled).

HHpred will connect to modeller and do the model building for you!. You can manually or ask hhpred to automatically select template from the results.

To select template manually, scroll down to tabular hit list and check the check box for the template of your preference. Then click 'create model from manually selected template(s)'. This will open the modeller web server page with the alignment already pasted in the box. You need to just submit the modeller job.

Enter your Modeller-key

Choose a Job-id

Input notification email

Submit Job.

6.Results

The result page displays the modelled structure pdb-file. You can choose pdb-file → save to save the model in pdb format.

Now use chimera to view your model and compare it with the template structure.

[Answer: The uncharacterized protein contains the ADENYLATE CYCLASE fold]

Exercise 2: Homology detection & structure prediction of an uncharacterized protein from *Candidatus Entothionella* using HHpred

You will work with a uncharacterized protein from *Candidatus Entotheonella*. Download the protein sequence of in fasta format and save into a directory.

Open <http://www.uniprot.org/> in your browser, and search for the gene name *ETSY2_32710*. (Uniprot Id : W4M0R9)

Repeat the same steps as Exercise 1.

Hint: Look at the template length and coverage for HHpred hits. You may need to choose more than one template for modelling.

Exercise 3: Homology detection & structure prediction of an uncharacterized protein of your interest

1. Retrieve the query sequence:

The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

UniProtKB
UniProt Knowledgebase
Swiss-Prot (553,655)
Manually annotated and reviewed.
TrEMBL (77,483,538)
Automatically annotated and not reviewed.

UniRef
Sequence clusters

UniParc
Sequence archive

Proteomes

Supporting data

- Literature citations
- Taxonomy
- Subcellular locations
- Cross-ref. databases
- Diseases
- Keywords

News

- Forthcoming changes
- Planned changes for UniProt
- UniProt release 2017_02
Freshwater fish see red | Cross-references to Araport, TAIR and IMGT/Gene-DB | Removal of sequence similarity annotations for domains
- UniProt release 2017_01
Sheep in wolves' clothing | Change of the UniRef FASTA header
- News archive

You will work with a uncharacterized protein of your interest.

Open <http://www.uniprot.org/> in your browser, and choose 'Advanced' in the search bar.

Searching in UniProtKB [Help](#)

Protein name [DE] Term

AND Term

Enter 'Uncharacterized protein' for protein name and search. Once you get the results you can choose any protein of your interest. Note that you can sort the results on the results page by sequence length, organism name etc.

Hint: You have more chance of finding a template for proteins with sequence length < 400, Also it is faster to build the model for smaller proteins.

Once you select your protein, repeat the same steps as Exercise 2.

Modeling proteins using automated modelling servers

Exercise 4: Automated protein structure prediction using Phyre2

Go to Phyre2 web server

<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>

Phyre²
Protein Homology/analogy Recognition Engine V 2.0

Subscribe to Phyre at Google Groups
Email:
Visit Phyre at Google Groups
[Follow @Phyre2server](#)

EBI 2016 Workshop | Older Workshops | New Phyre2 paper | Fast structural search with PhyreStorm (beta-testing)

E-mail Address

Optional Job description

Amino Acid Sequence

Or try the sequence finder

Modelling Mode Normal Intensive

Phyre2 is fully automated, you just need to give your protein sequence for modelling. The program will find the template, create alignment and build the model.

Use any of the previous example to build model and compare the results.

Exercise 5: Automated protein structure prediction using I-TASSER

Go to I-TASSER web server

<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>