

# Tutorial

## Protein modelling with online tools

Karel Berka

18.10.2023

# Task

- Prepare model of human UGT1A1
  - SwissModel
    - <https://swissmodel.expasy.org/>
  - AlphaFold
    - <https://colab.research.google.com/github/sokrypton/>
  - I-Tasser
    - <http://zhanglab.ccmb.med.umich.edu/I-TASSER/>
  - ESMfold
    - <https://esmatlas.com/resources?action=fold>

# Preparation of sequence

- Find Uniprot sequence of human  
<https://esmatlas.com/resources?action=fold>
  - Store it as FASTA
  - Remember UNIPROT ID
- Analyze overall topology
  - Feature viewer
- Find similar sequences with known PDB structure with BLAST

UniProtKB

BLAST Align Retrieve/ID mapping Peptide search

## UniProtKB - P22309 (UD11\_HUMAN)

Display

Entry Publications Feature viewer Feature table

Protein | UDP-glucuronosyltransferase 1-1

Gene | UGT1A1

Organism | *Homo sapiens* (Human)

Status | Reviewed - Annotation score: 100%

### Function<sup>i</sup>

UDPGT is of major importance in the conjugation and metabolism of bilirubin IX-alpha to form both the glucuronidation of 17beta-estradiol, 17alpha-ethinyl-17beta-estradiol and umbelliferone. Isoform 2 lacks transferase activity

### Catalytic activity<sup>i</sup>

UDP-glucuronate + acceptor = UDP + acceptor beta-D-glucuronide

# SwissModel

- Use FASTA as input
- First search for templates

## SWISS-MODEL

### Start a New Modelling Project

Target Sequence:

(Format must be Fasta, Clustal, Promod, plain string, or a valid UniProtKB AC)

Target GIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRYTGTRPSNLANTILVKWLPQNDL 360

Target LGHFPMTRAFITHAGSHGVYESICNGVPMVMPLFGDQMDNAKRMETKGAGVTLNVLEMTS 420

Target EDLENALKAVINDKSYKENIMRLSSLHKDRPEPLDLAVFWVEFVMRHKGAPHLRPAHD 480

Target LTWYQYHSLDVIGFLLAVVLTVAFITFKCCAYGYRKCLGKKGRVKKAHKSKTH 533

Reset Form

+ Upload Target Sequence File...

Project Title:

Untitled Project

Email:

Optional

Search For Templates

Build Model

### Untitled Project

Created: today at 21:39

Summary Templates Models

### Template Results

The search for templates matching your target sequence is currently **running**. Please wait.

...reusing cached query alignment  
...predicting residue burial status with ACCpro  
...searching PDB profile database with previously built query profile

If you want to come back later, bookmark this link:

<http://swissmodel.expasy.org/interactive/hVaTnM/>

MAVESQGG RPLVLG LLLCVLGPV VSHAGKILLIPVDG SHW

LSMLGAIQQ LQQRGHEIVVLAPDASLY I RDGAFYTLK TYP

VPFQREDVKESFVSLGHNV FENDSFLQ RVIKTYKKIKKDS

AMLLSGC SHLLHNKELMASLAESSFDV MLTDPFLPC SPIV

AQYLSLPTVFFLHALPCSLEFEATQCPN PFSYVPRPLSSH

# SwissModel

- Not all templates are made equal 😊
- Try to have
  - large coverage,
  - high identity,
  - best resolution
  - and preferably similar ligands (UDP)
- To compare select just 5 with best coverage

SWISS-MODEL

All Projects

Untitled Project Created: today at 21:39

Summary Templates **50** Models

Template Results ?




Templates Sequence Similarity Alignment of Selected Templates More ▾

◆ Name ◆	Title	▼ Coverage	◆ Identity ◆	◆ Method ◆	◆ Oligo State ◆	◆ Ligands ◆
<input checked="" type="checkbox"/> 2pq6.1.A	UDP-glucuronosyl/UDP-glucosyltransferase		17.96	X-ray, 2.1Å	monomer	None
<input type="checkbox"/> 2iya.1.A	OLEANDOMYCIN GLYCOSYLTRANSFERASE		18.30	X-ray, 1.7Å	monomer	1 x ZIO <sup>Ⓞ</sup> , 1 x UDP <sup>Ⓞ</sup>
<input type="checkbox"/> 2yjn.1.A	GLYCOSYLTRANSFERASE		16.08	X-ray, 3.1Å	hetero-oligomer	None
<input type="checkbox"/> 4lei.1.B	NDP-forosamyltransferase		16.88	X-ray, 3.1Å	homo-dimer	2 x TYD <sup>Ⓞ</sup>
<input type="checkbox"/> 4lei.1.A	NDP-forosamyltransferase		16.88	X-ray, 3.1Å	homo-dimer	2 x TYD <sup>Ⓞ</sup>
<input type="checkbox"/> 1iir.1.A	glycosyltransferase GtIB		13.89	X-ray, 1.8Å	monomer	2 x MG <sup>Ⓞ</sup>
<input type="checkbox"/> 2c1x.1.A	UDP-GLUCOSE FLAVONOID 3-O GLYCOSYLTRANSFERASE		17.22	X-ray, 1.9Å	monomer	1 x UDP <sup>Ⓞ</sup> , 1 x B3P <sup>Ⓞ</sup>

# SwissModel



- Closeness in the sequence space is important, but structure comparison of templates is more important –
- template is the sole predictor of model


Untitled Project Created: today at 21:39

Summary Templates **50** Models   

### Template Results

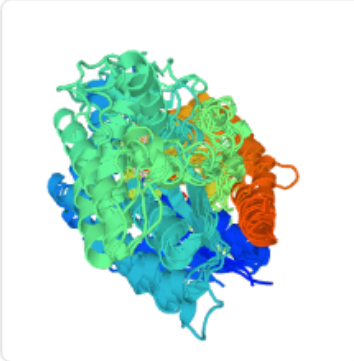
Templates Sequence Similarity Alignment of Selected Templates More ▾










SMTL ID	1iir.1.A
Title	glycosyltransferase GtfB
Coverage	
Identity	13.89
Similarity	0.26
Ligands	2 x MG
Method	X-ray 1.80Å
Oligo State	monomer
Found_by	HHblits


**Build Models 5**


Clear Selection





**View ▾**      

2pq6.1.A  

2iya.1.A  

2yjn.1.A  

4lei.1.A  

1iir.1.A  

# SwissModel – Model building

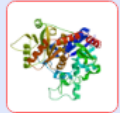
- Template selection enables model building
- Quality of model is shown with General Model Quality Estimate - Mine are not good...

Untitled Project Created: today at 21:39

Summary Templates 50 Models 5

Model Results

Order by: QMEAN4



Model 01

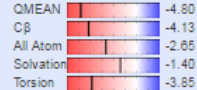
Oligo-State  
MONOMER

Ligands  
None

GMQE 0.46

QMEAN -4.80

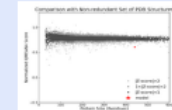
Global Quality



Local Quality



Comparison



Template

2pq6.1.A

Seq Identity

17.96%

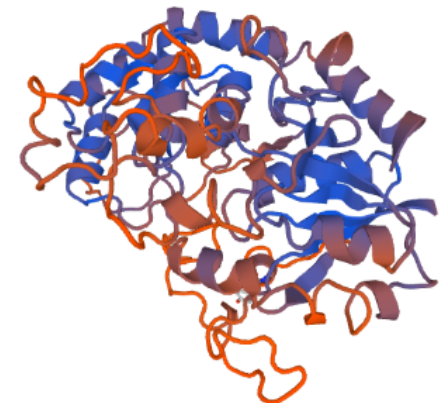
Coverage

Description

UDP-glucuronosyl/UDP-glucosyltransferase

Model-Template Alignment

```
Model_01 MAVESQGRPLVLGLLLCVLGPVVSFAGKILLIPVDGSHWLSMLGATQQLQQRGHEIWLAPDASLYIRDCAFYTLKTYFVPFQREDVKESFVSLGRHV
2pq6.1.A -----KPTQNTFFPFQCHINFLFLKALLLRGSHITFQSYNNGHQLNRGPHAFDGFTEFNEFQDGLTME-GD-GD---V--
Model_01 FENDSFQQRVIKTKKIKKQDSAMLLSGCHLLHNKELMASLAESSFDVMLTDPELPQSPIVAQYLSLPTVFLLHALPCSLEFEATQCPNPFYVPRPLSSHSDDH
2pq6.1.A --SQ---D-----QFLLQSRQNNLQPY--ELATFLHSTWTFPTCLVQDCMCTTQAAGTFLSLVYSSSCSLNLCGFSTFAGIIFPKIESY--TH
Model_01 YFLQRVKHMLIAFSQNFLCDWYSYPATLASEFLQREVTVDQLSSASVWLFSDFVMDYFRPIMPHNVTVGGINCLHQHPL-----SQEF
2pq6.1.A LLDVQWNPGLKNGSGDIDFRTTNENLMEFFTE--VLDWNRQTTLLNTNLEESVINALSTTPDVGGLPELLSTPILHLDNLNKGEDCE
Model_01 EAYINASGERGIWFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRYTGTSPSNLANNTILVXWLPQNDLLGHPNTRAFITHAGSHGVYESICNGVPM
2pq6.1.A LLDGSHKSPGCVVQGGSTTVMDPEQLLEFANLLQGGFLNFGDQVIGGVYFSEPTNEAGGLQWNRQKQDHPGGEFLTHGNTSTIESICGVPM
Model_01 VMLPLGDDQNDNAKRMEIKGAGVTLNVLENTSEDLNALKAVINDKSYKENIMHLSLHKDRPV--EPDLDAVERVEFVRHNGAPHLPAAHDLTWIYHSLDV
2pq6.1.A LPPFAQPTQCRFFQSGDIDTHVLELLMLNHEVLAQKKKQKQKMLGGQALTRPGGCEHMLKQKTHV
Model_01 IGFLNAVLTVAPIIFKCCAYGKAKLGGKGRVKKAKHKKHTH
2pq6.1.A -----
```



# AlphaFold

- Google Colab
  - <https://colab.research.google.com/github/sokrypton/>
- AlphaFold
  - <https://www.alphafold.ebi.ac.uk/entry/P22309>



# I-Tasser

- Insert FASTA
- Use as much knowledge as possible (e.g. PDB, secondary structure...)
- Submit
- And compare to SwissModel

[\[Home\]](#) [\[Server\]](#) [\[Queue\]](#) [\[Forum\]](#) [\[About\]](#) [\[Remove\]](#) [\[Statistics\]](#)

## I-TASSER submission for job id S298758

### Submitted Sequence

```
>UGT1A1
MAVESQGGRLVLGLLLCVLGPVSHAGKILLIPVDGSHWLSMLGAIQQLQQRGHEIVVL
APDASLYIRDGAFTYTKTYPVPFQREDVKESFVSLGHNVFENDSFQRVIKTYKKIKKDS
AMLLSGCSHLLHNKELMASLAESSFDVMLTDPFLPCSPIVAQYLSLPTVFFLHALPCSLE
FEATQCPNPF SYVPRPLSSHSDHMTFLQRVKNMLIAFSQNFLCDVVYSPYATLASEFLQR
EVTVQDLLSSASVWLFPSDFVKDYPRPIMPNMVFGGINCLHQNPLSQEFAYINASGEH
GIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRYTGTRPSNLANNTILVKWLPQNDL
LGHPMTRAFITHAGSHGVYESICNGVPMVMMP LFGDQMDNAKRMETKGAGVT LNVLEMTS
EDLENALKAVINDKSYKENIMRLSSLHKDRPVEPLD LAVFWVEFVMRHKGAPHLRPAHD
LTWYQYHSLDVIGFLLAVVLTVAFITFKCCAYGYRKCLGKKGRVKKAHKSKTH
```

Your sequence has been successfully submitted and in processing.....

You will receive a notification email once the job is completed. You may want to bookmark the page to check the results later.

# I-Tasser results

- Predict secondary structure
- Predict solvent accessibility
- Identification of templates
- 5 models sorted by C-score (quality estimate)
- Structural analogs
- Predicted function
- Ligand binding site
- EC number and GO

# ESMfold – language model

- <https://esmatlas.com/resources?action=fold> – max 400 aa
- [https://esmatlas.com/resources?action=search sequence](https://esmatlas.com/resources?action=search_sequence) – gives list of known sequences
  - Or look them up in Mgnify database
- UGT has 533 AA.
  - > 3 models – beginning, center, end
  - Division ideally beyond some helical segment that enable concatenation
  - Later / concatenation of models in Pymol

# Comparison of models

- in Pymol
  - compare to known best template
  - compare between each other
  - identify parts which seems to be ok in most models
  - look at positions of disorder
  - C-terminal TM helix – is it in the model?
- Quality check
  - QMEAN
    - <https://swissmodel.expasy.org/qmean/>
  - AlphaFold score