Tutorial
Homology modelling with online tools

Karel Berka

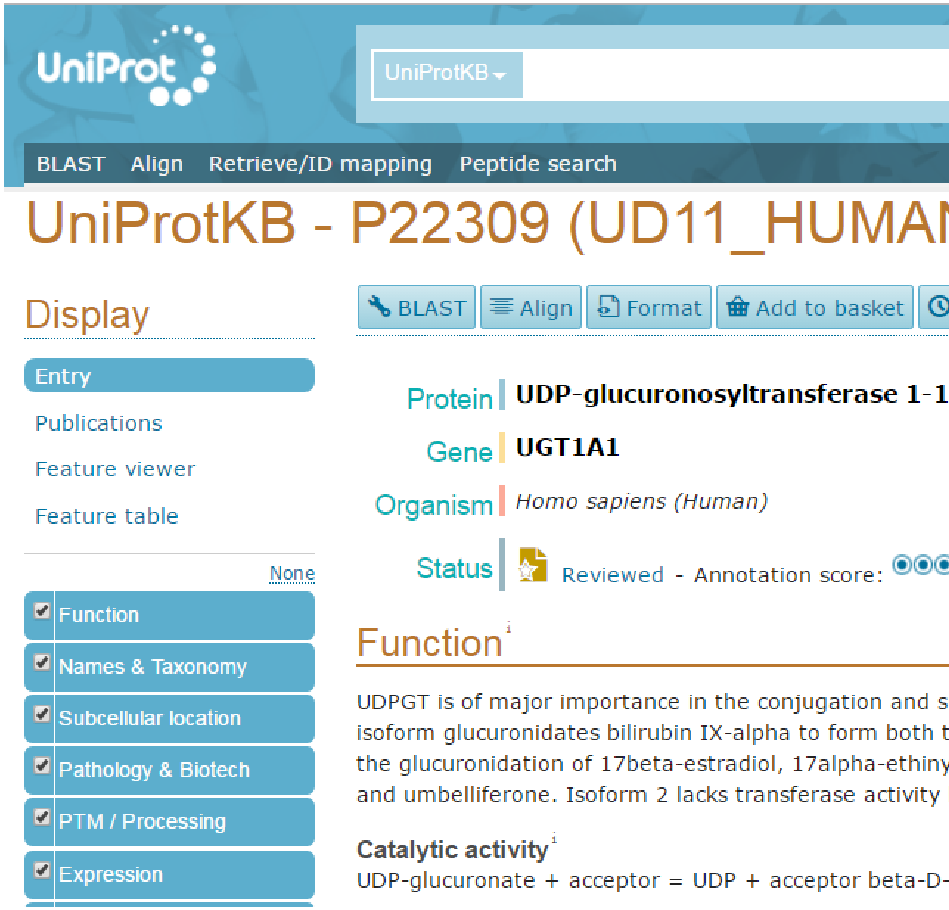
22.11.2021

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/
  - Evfold
    - http://evfold.org/evfold-web/evfold.do

Preparation of sequence

- Find Uniprot sequence of human UGT1A1
  - Store it as FASTA
  - Remember UNIPROT ID
- Analyze overall topology
  - Feature viewer
- Find similar sequences with known PDB structure with BLAST



UniProtKB - P22309 (UD11_HUMA

Protein UDP-glucuronosyltransferase 1-1

Gene UGT1A1

Organism *Homo sapiens (Human)*

Status Reviewed - Annotation score:

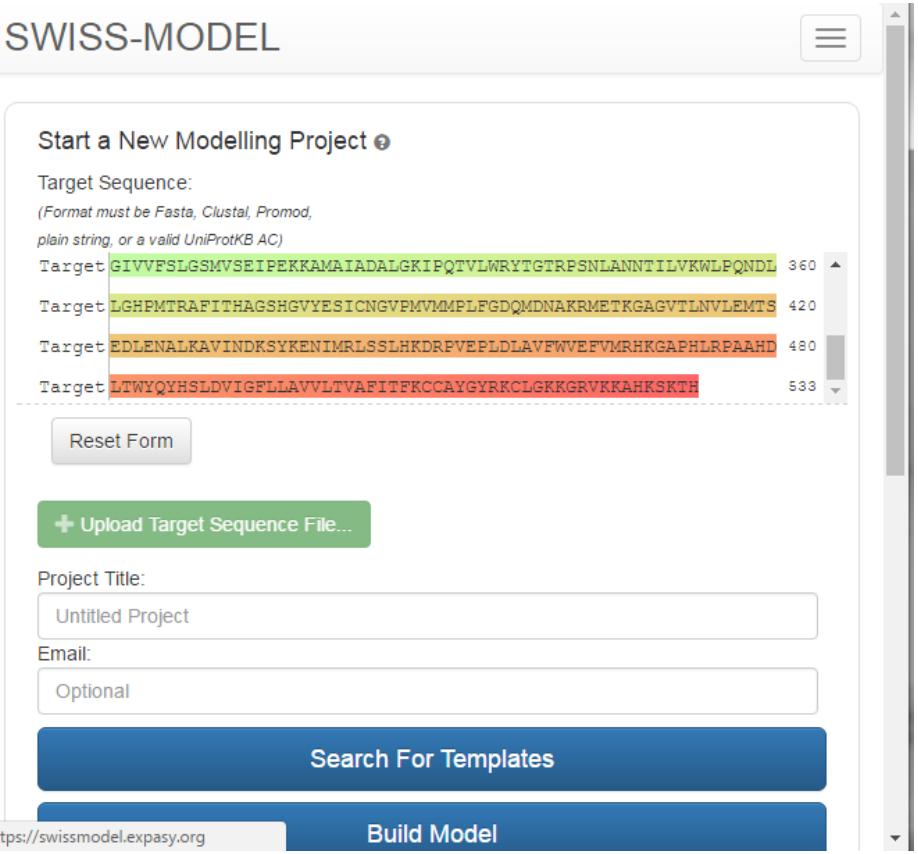
Function

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

Catalytic activity
UDP-glucuronate + acceptor = UDP + acceptor beta-D-

SwissModel

- Use FASTA as input - First search for templates



Start a New Modelling Project

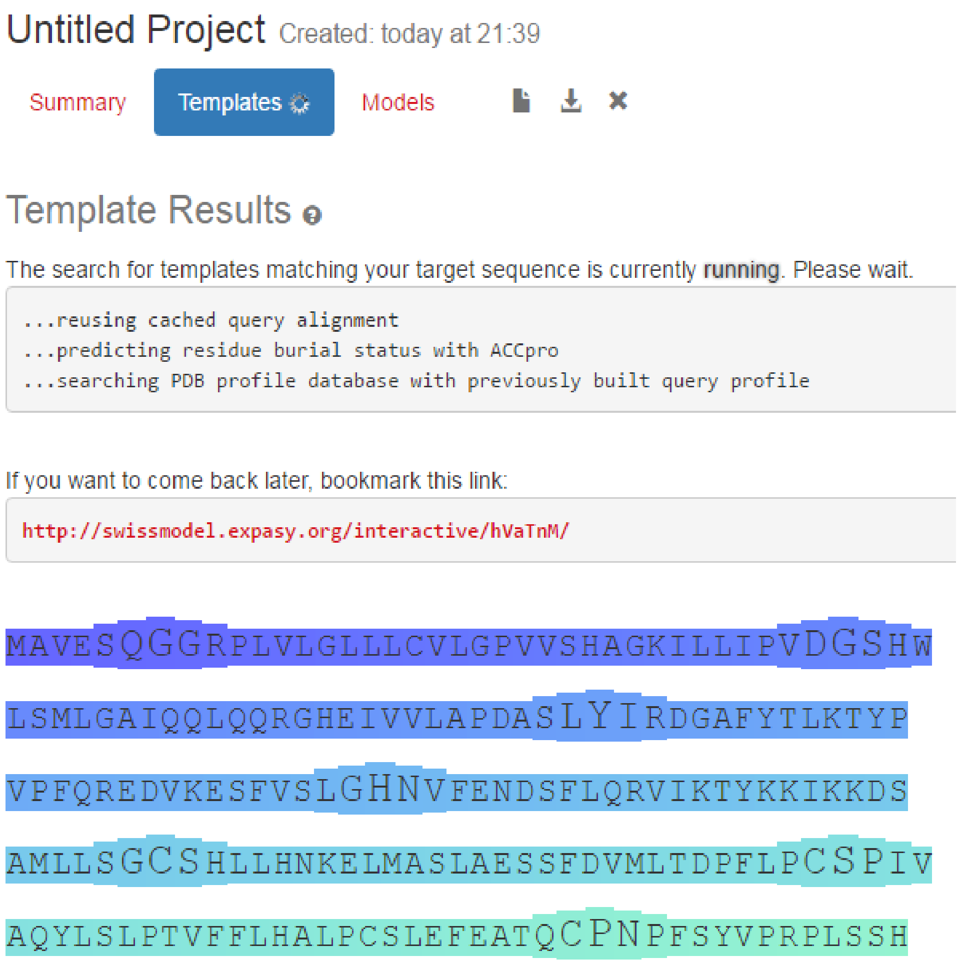
Target Sequence:

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

```
Target GIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRYTGTRPSNLANNTILVKWLPQNDL 360
Target LGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFGDQMDNAKRMETKGAGVTLNVLEMTS 420
Target EDLENALKAVINDKSYKENIMRLSSLHKDRPVEPLDLAVFWVEFVMRHKGAPHLRPAAHD 480
Target LTWYQYHSLDVIGFLLAVVLTVAFITFKCCAYGYRKCLGKKGRVKKAHKSKTH         533
```

Project Title: Untitled Project

Email: Optional



Untitled Project Created: today at 21:39

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
```

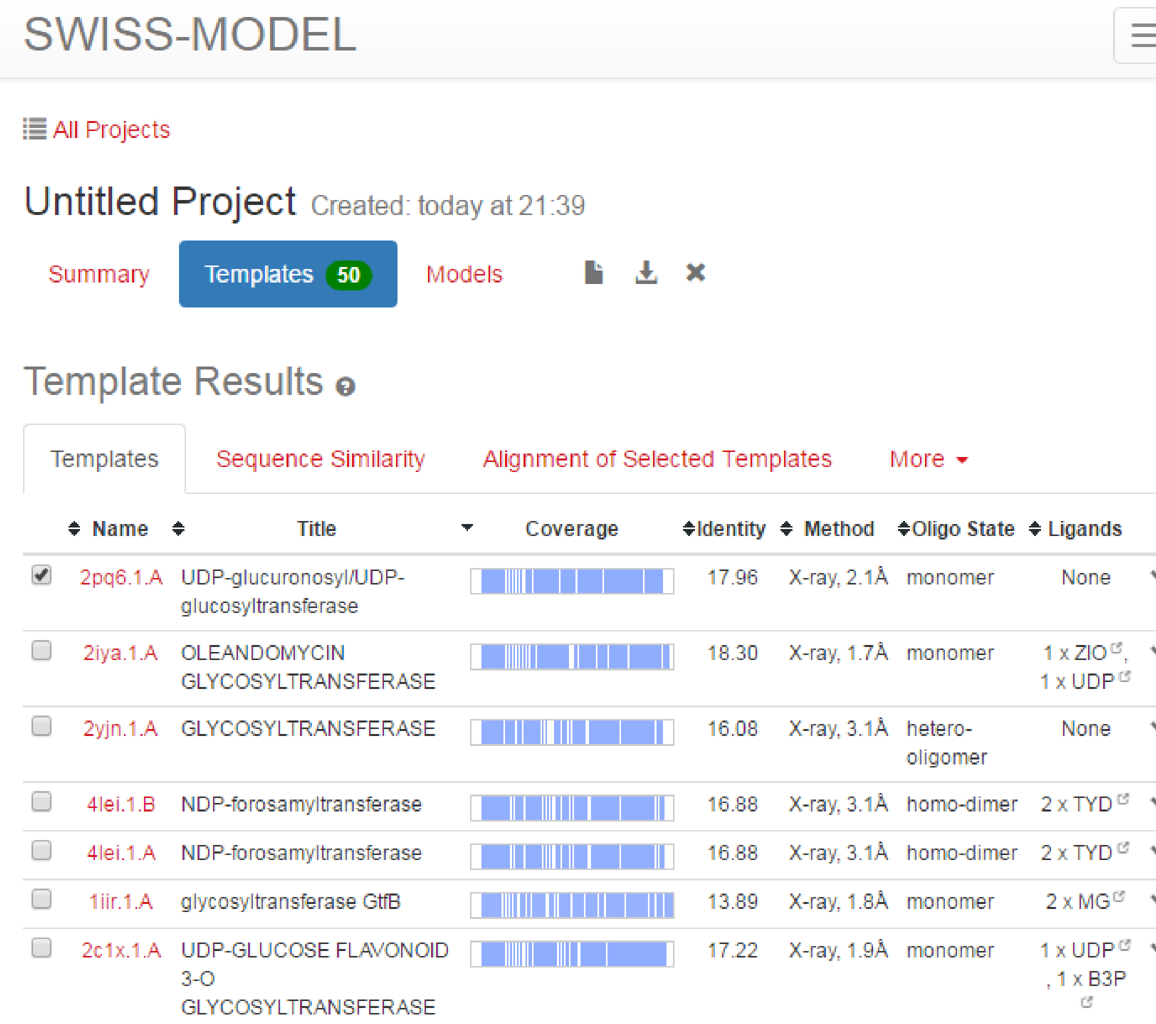
If you want to come back later, bookmark this link:

`http://swissmodel.expasy.org/interactive/hVaTnM/`

```
MAVESQGGRPLVLGLLLCVLGPVVSHAGKILLIPVDGSHW
LSMLGAIQQLQQRGHEIVVLAPDASLYIRDGAFYTLKTYP
VPFQREDVKESFVSLGHNVFENDSFLQRVIKTYKKIKKDS
AMLLSGCSHLLHNKELMASLAESSFDVMLTDPFLPCSPIV
AQYLSLPTVFFLHALPCSLEFEATQCPNPFSYVPRPLSSH
```

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



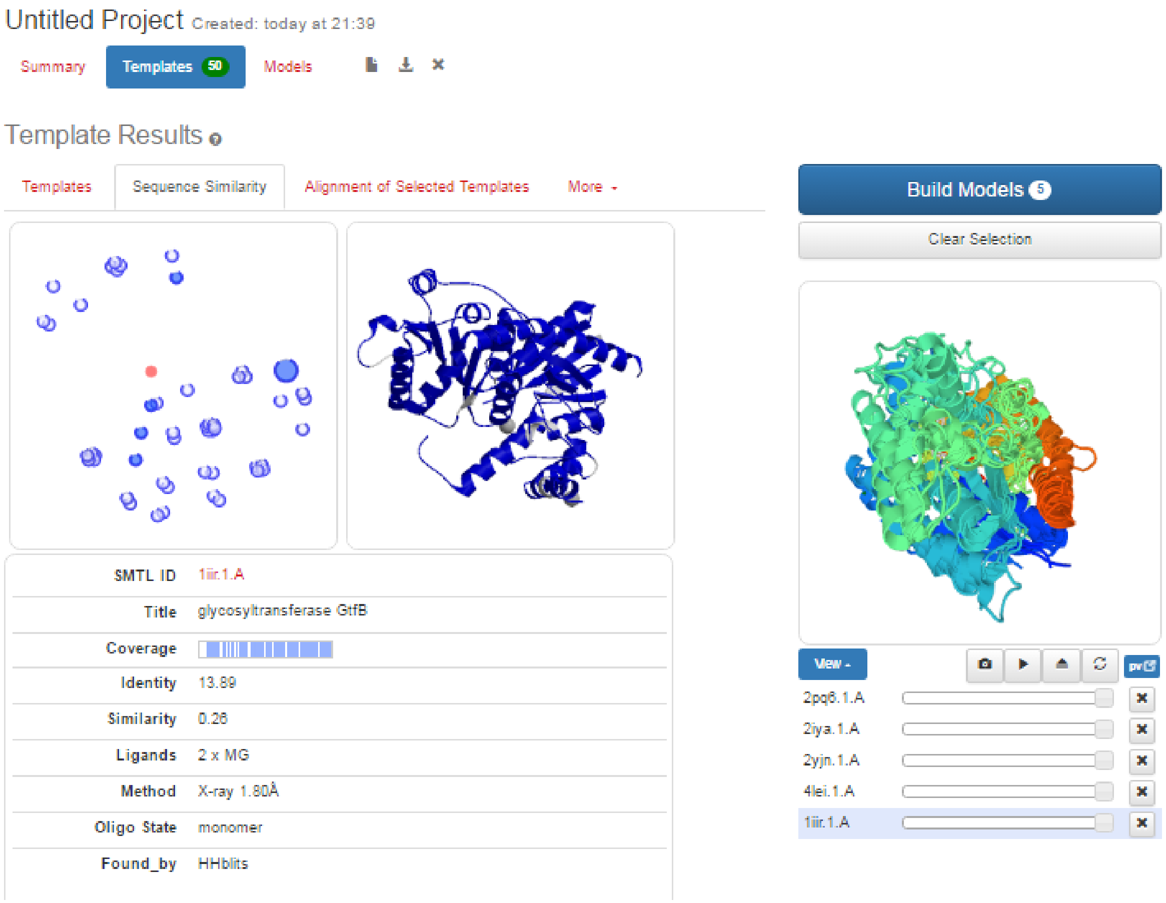
Untitled Project Created: today at 21:39

Template Results

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

SwissModel

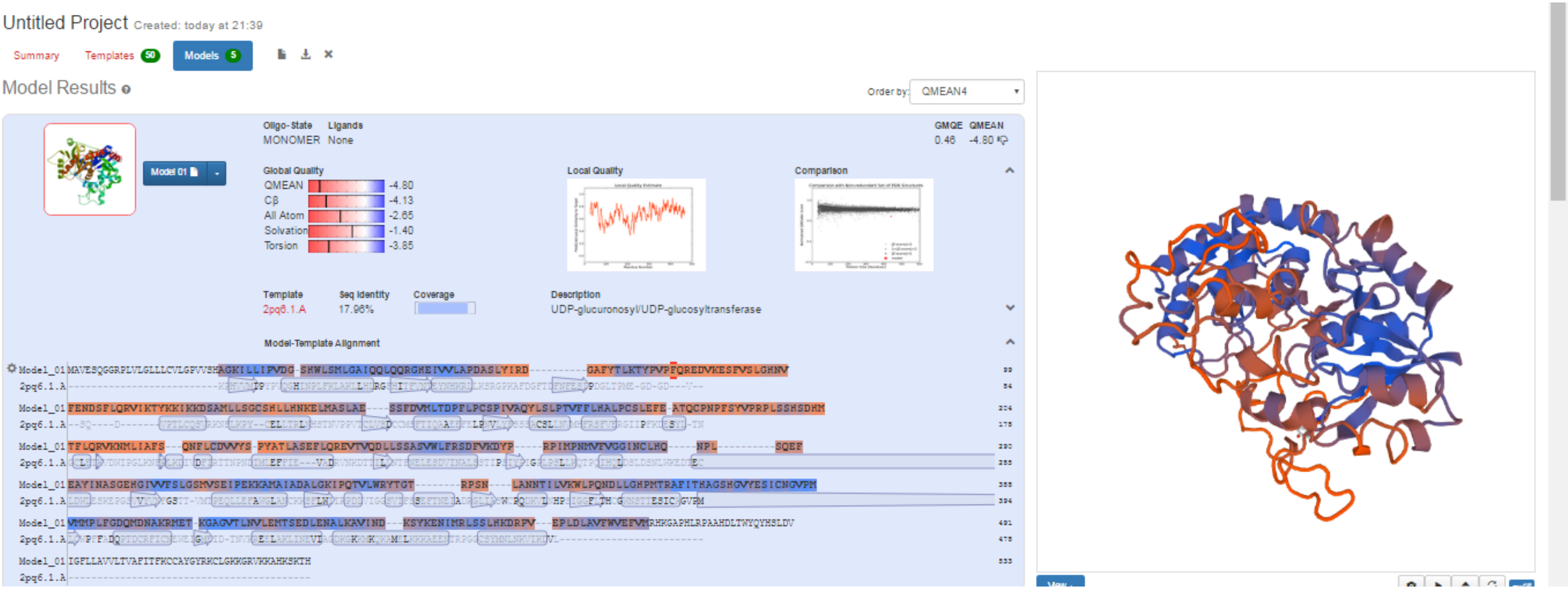
- Closeness in the sequence space is important, but structure comparison of templates is more important –
- template is the sole predictor of model



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

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

Oligo-State: MONOMER
Ligands: None
GMQE: 0.46
QMEAN: -4.80

Global Quality

QMEAN	-4.80
Cβ	-4.13
All Atom	-2.65
Solvation	-1.40
Torsion	-3.85

Template	Seq Identity	Coverage	Description
2pq6.1.A	17.96%		UDP-glucuronosyl/UDP-glucosyltransferase

Model-Template Alignment

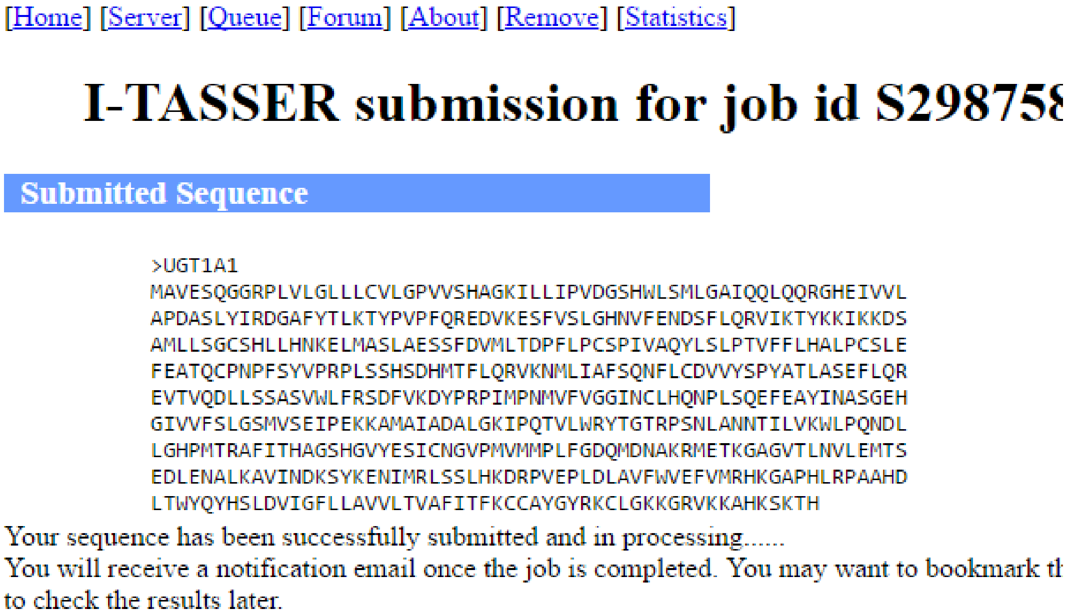
```
Model_01 MAVESQGGRPLVLGLLLCVLGPVVSHAGKILLIPVDG-SHWLSMLGAIQQLQQRGHEIVVLAPDASLYIRD----------GAFYTLKTYPVPFQREDVKESFVSLGHNV  99
Model_01 FENDSFLQRVIKTYKKIKKDSAMLLSGCSHLLHNKELMASLAE----SSFDVMLTDPFLPCSPIVAQYLSLPTVFFLHALPCSLEFE-ATQCPNPFSYVPRPLSSHSDHM  204
Model_01 TFLQRVKNMLIAFS---QNFLCDVVYS-PYATLASEFLQREVTVQDLLSSASVWLFRSDFVKDYP-----RPIMPNMVFVGGINCLHQ-----NPL----------SQEF  290
Model_01 EAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRYTGT--------RPSN----LANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPM  355
Model_01 VMMPLFGDQMDNAKRMET-KGAGVTLNVLEMTSEDLENALKAVIND---KSYKENIMRLSSLHKDRPV---EPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDV  491
Model_01 IGFLLAVVLTVAFITFKCCAYGYRKCLGKKGRVKKAHKSKTH  533
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 S29875

Submitted Sequence

```
>UGT1A1
MAVESQGGRPLVLGLLLCVLGPVVSHAGKILLIPVDGSHWLSMLGAIQQLQQRGHEIVVL
APDASLYIRDGAFYTLKTYPVPFQREDVKESFVSLGHNVFENDSFLQRVIKTYKKIKKDS
AMLLSGCSHLLHNKELMASLAESSFDVMLTDPFLPCSPIVAQYLSLPTVFFLHALPCSLE
FEATQCPNPFSYVPRPLSSHSDHMTFLQRVKNMLIAFSQNFLCDVVYSPYATLASEFLQR
EVTVQDLLSSASVWLFRSDFVKDYPRPIMPNMVFVGGINCLHQNPLSQEFEAYINASGEH
GIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRYTGTRPSNLANNTILVKWLPQNDL
LGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFGDQMDNAKRMETKGAGVTLNVLEMTS
EDLENALKAVINDKSYKENIMRLSSLHKDRPVEPLDLAVFWVEFVMRHKGAPHLRPAAHD
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 t
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

EV fold

- EV fold require only <500 amino acids
  - UGT has 533 AA.
  - >3 models – beginning, center, end
  - Division ideally in some helical segment
- 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