NEAR EAST UNIVERSITY
DESAM INSTITUTE

II. BIOINFORMATICS WINTER SCHOOL:
COMPUTER METHODS IN MOLECULAR SCIENCES

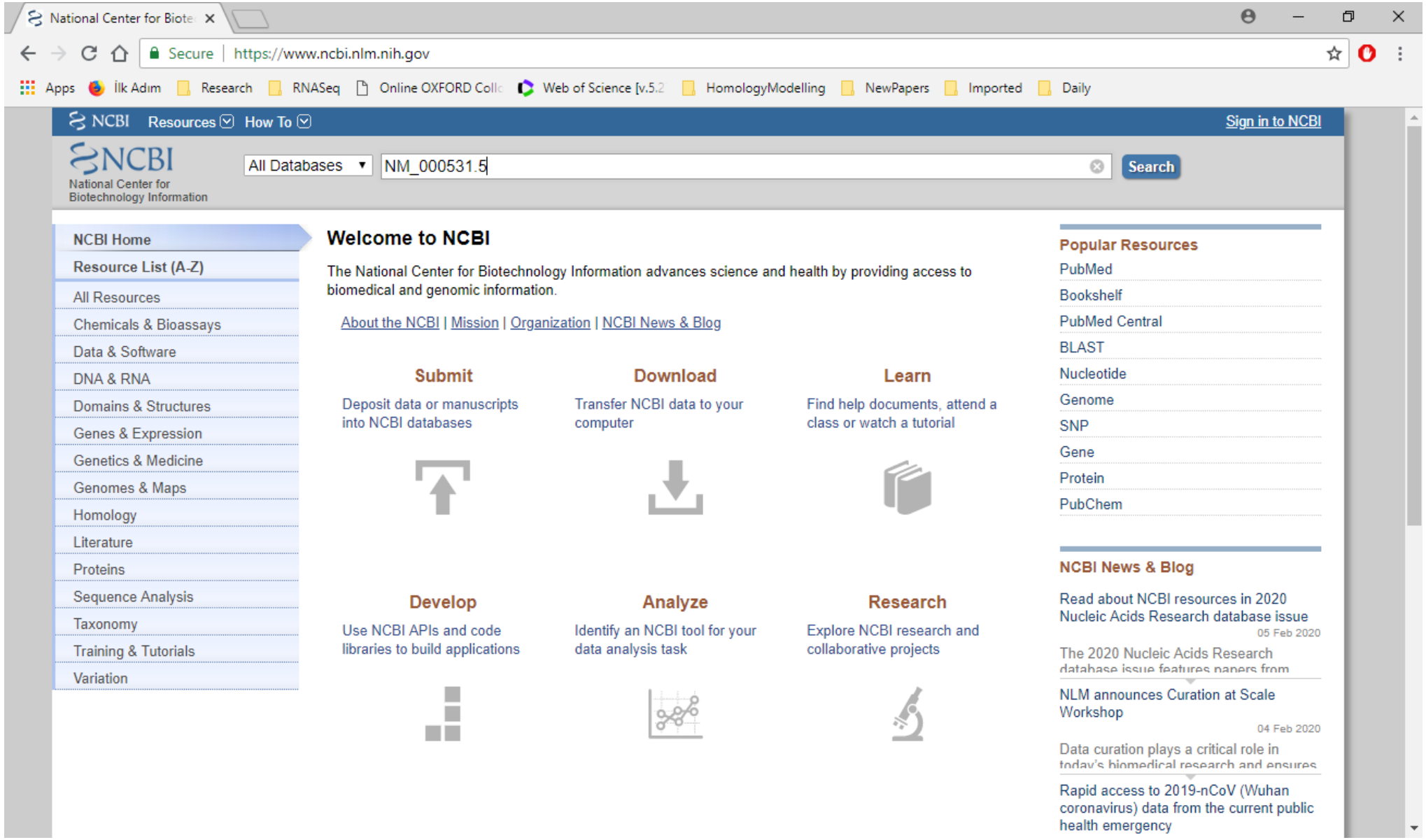
In Silico Homology Modelling of Proteins - Workshop

Şeref Gül, PhD.
Chemical and Biological Engineering
Koç University

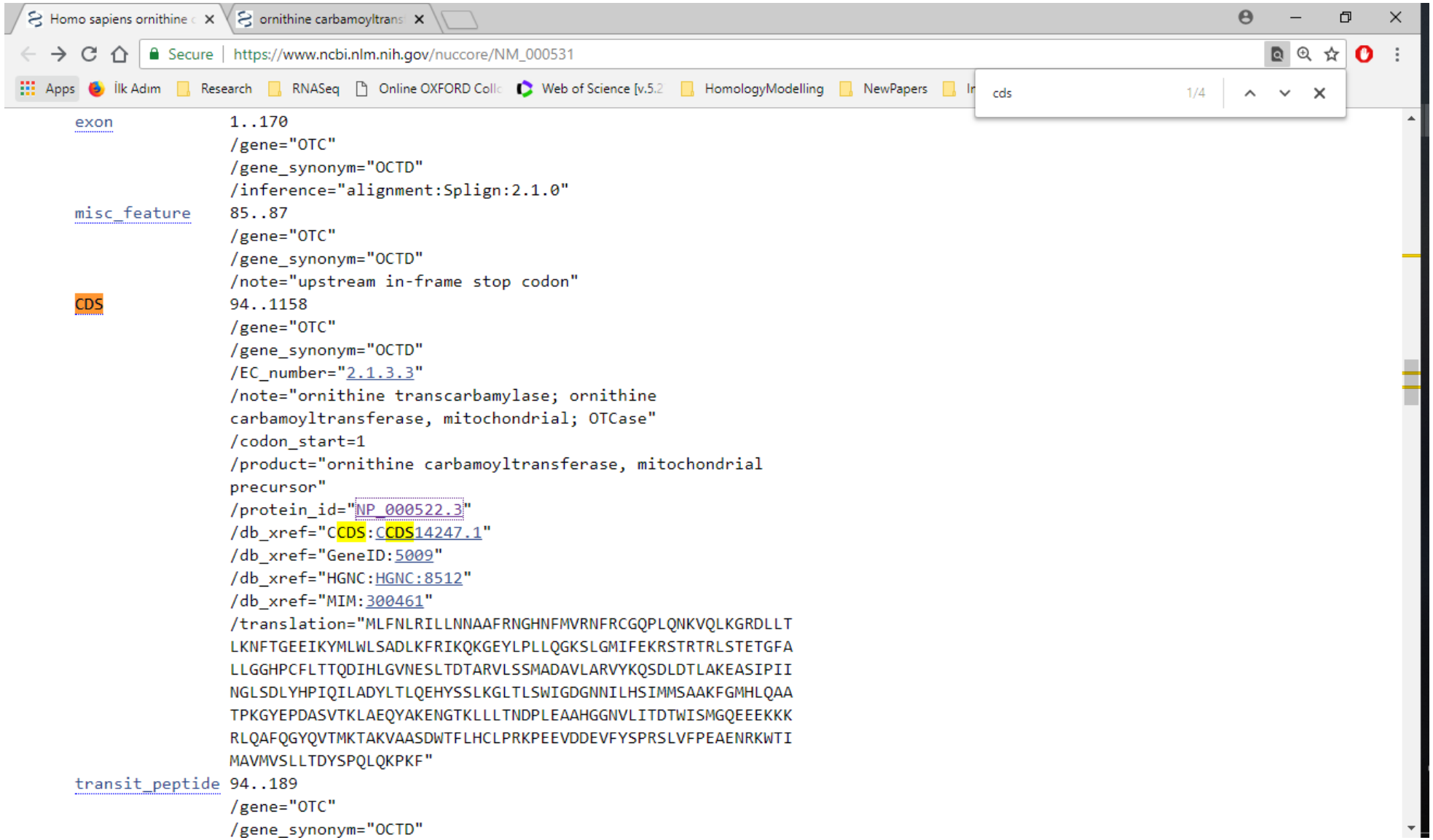
KOÇ UNIVERSITY

1) Search for ncbi and NM_number (e.g. NM_000531.1)
2) Go to NCBİ (https://www.ncbi.nlm.nih.gov/)-> Nucleotide search NM_number there

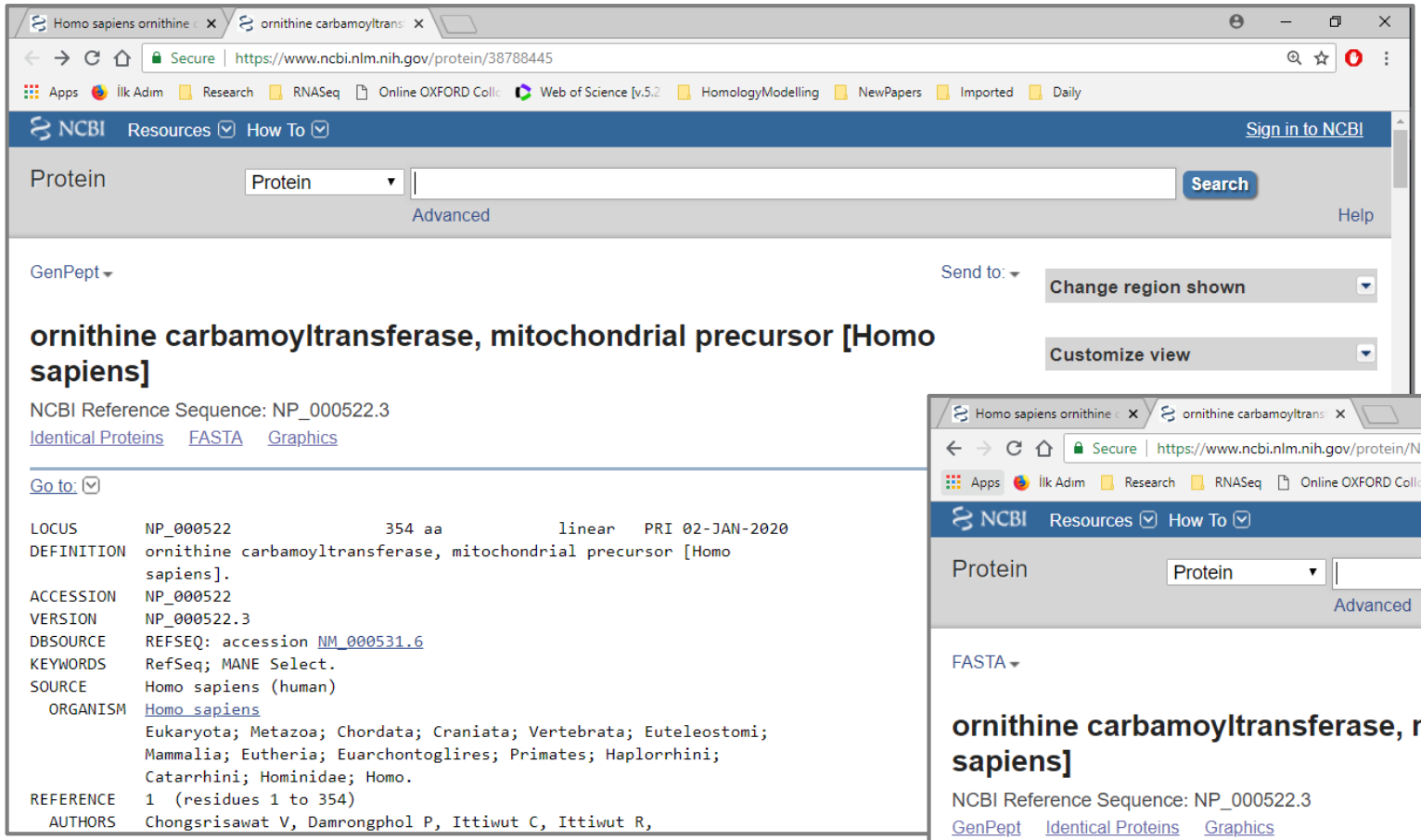
1) Search for ncbi and NM_number (e.g. NM_000531.1)
2) Go to NCBİ (https://www.ncbi.nlm.nih.gov/)-> Nucleotide search NM_number there

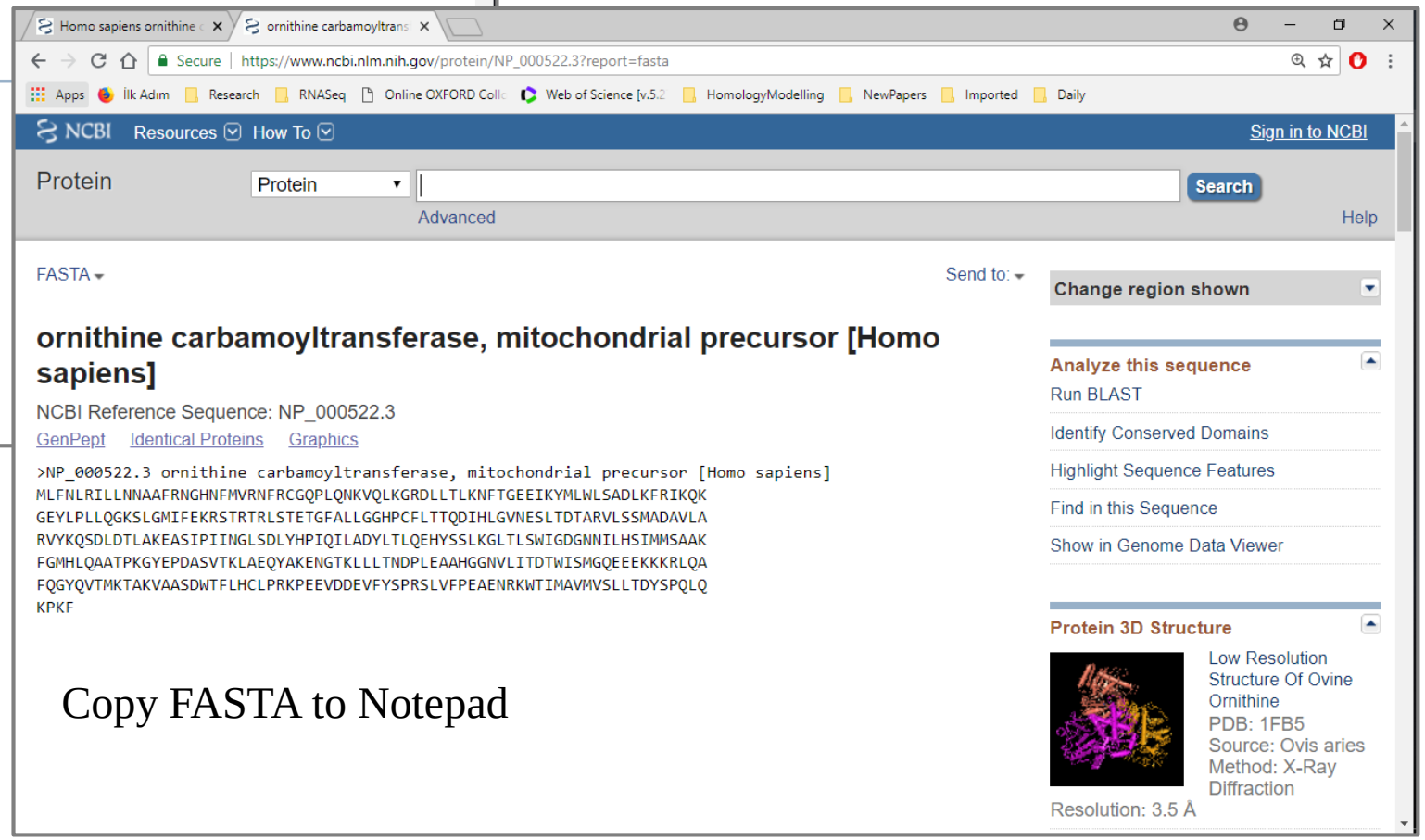


3) Search for CDS and click NP_number

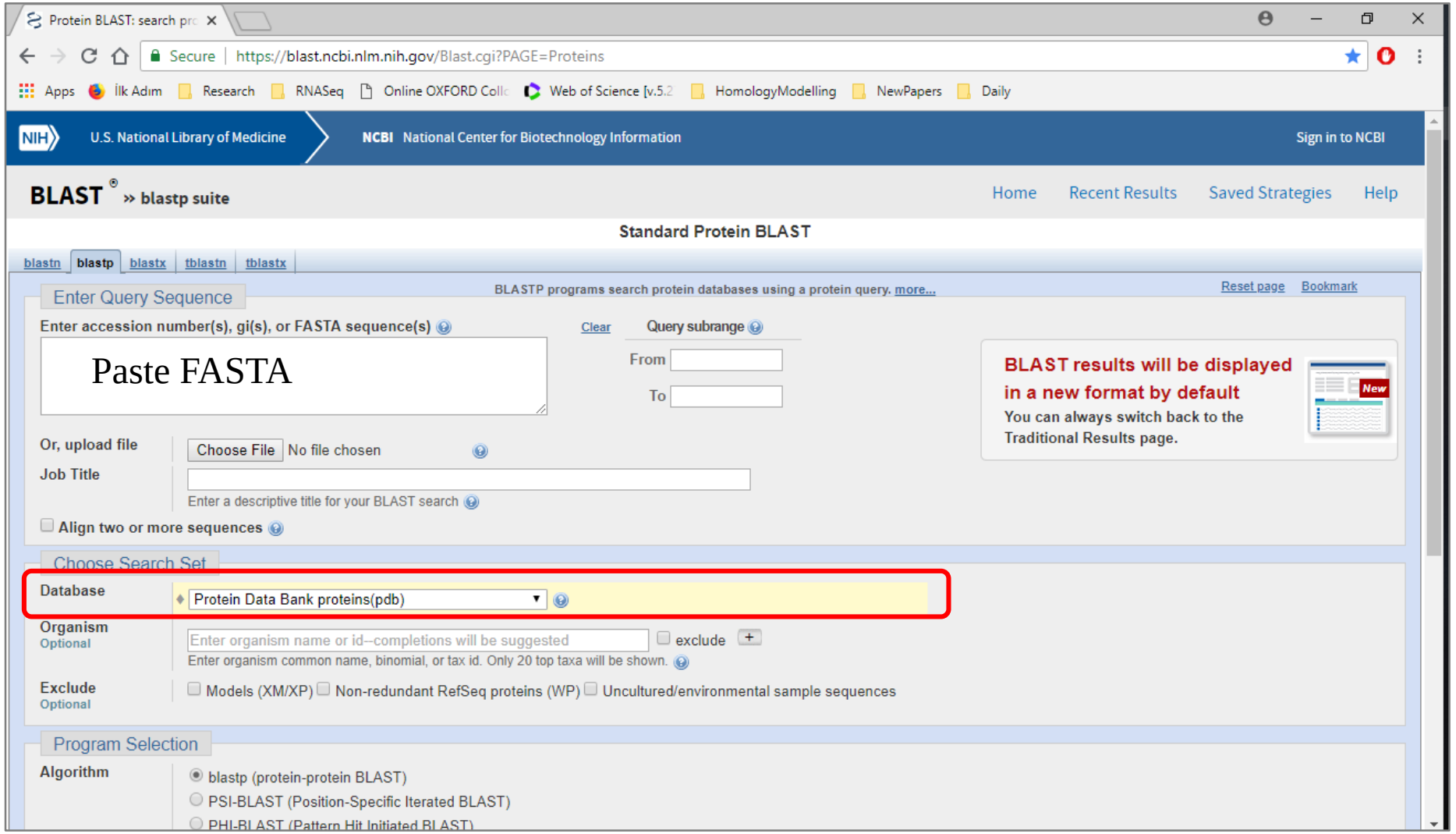


4) Get FASTA from NP_number





To search for protein structures to our protein of interst use BLASTP module of NCBI:
https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins
Select «Protein Data Bank» as Database to Get FASTA from NP_number



Modelling w Swiss Model

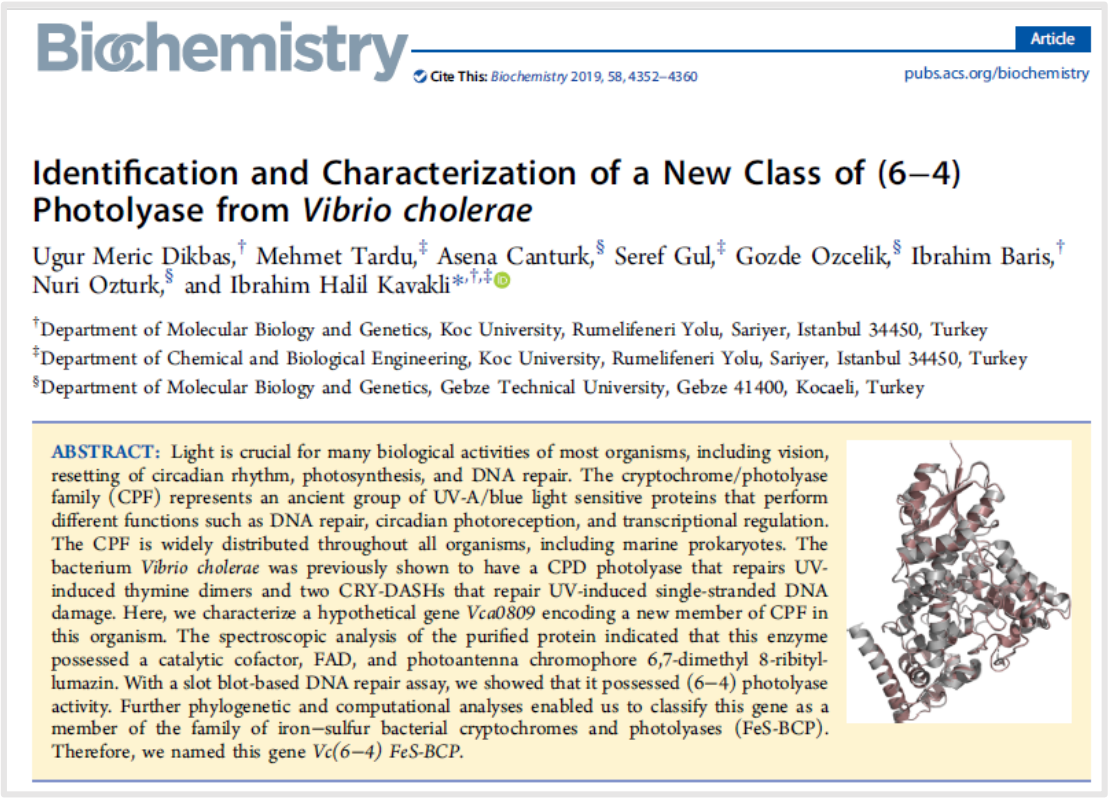
Sequence1

>NP_233195.1 hypothetical protein VCA0809 [Vibrio cholerae O1 biovar El Tor str. N16961]
MRYSVVRLILGDQLNHAHSWFSEHRDDVLYLIAELHQEQEYVRHHIQKQCAFFAAMQAFADYLSAEGHHV
WHLDLDASAQYNDLPDLIAQICQQVQADAFQYQRPDEYRLLEQMANLRLSGITIGCVDTEHFLLPFAEIP
EQFPASKAVLMEHFYRRMRKRFGYLMTADGKPEGGQWNFDADNRNKLKSPDLLQLPTPLCFDNPVASIKA
RIERHRIPSIGQVGESLLWPINRAQALSLLAHFCQICLPNFGRFQDAMTAQHPHRWSLYHSRLSFALNSK
LLSPREVIEATISAYRAAQGQISLAQVEGFVRQILGWREYVRGMYWSNMPHYQTRNHLGAQRPLPSYFWN
GQTKMRCLQQAITQSLDFGYAHHIQRLMVTGNFALLTECDPDQVDAWYLGIYIDAIEWVELPNTRGMALF
ADGGLIATKPYSASGSYINKMSDYCASCAYQVKLKSGEKACPLNSLYWRFMLKHRDRLANNPRIGMLYKT
WDKMTSDSQQAILSTADAYLSQIESL

1) Go to https://swissmodel.expasy.org/
   Paste sequence search for templates

2) BlastP the sequence; choose PDB to search for similar structures

3) Choose the best fit template recognized in BlastP for modelling



Biochemistry
Article
Cite This: Biochemistry 2019, 58, 4352−4360
pubs.acs.org/biochemistry

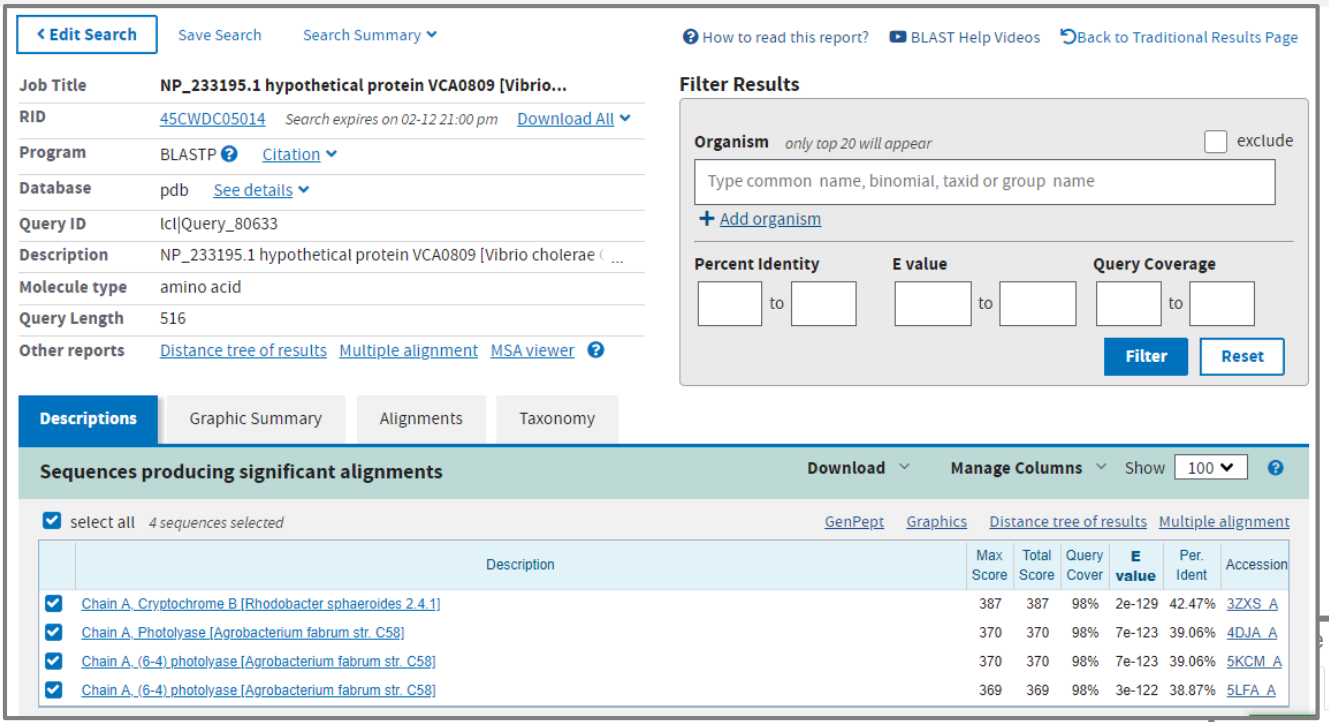
Identification and Characterization of a New Class of (6−4) Photolyase from *Vibrio cholerae*

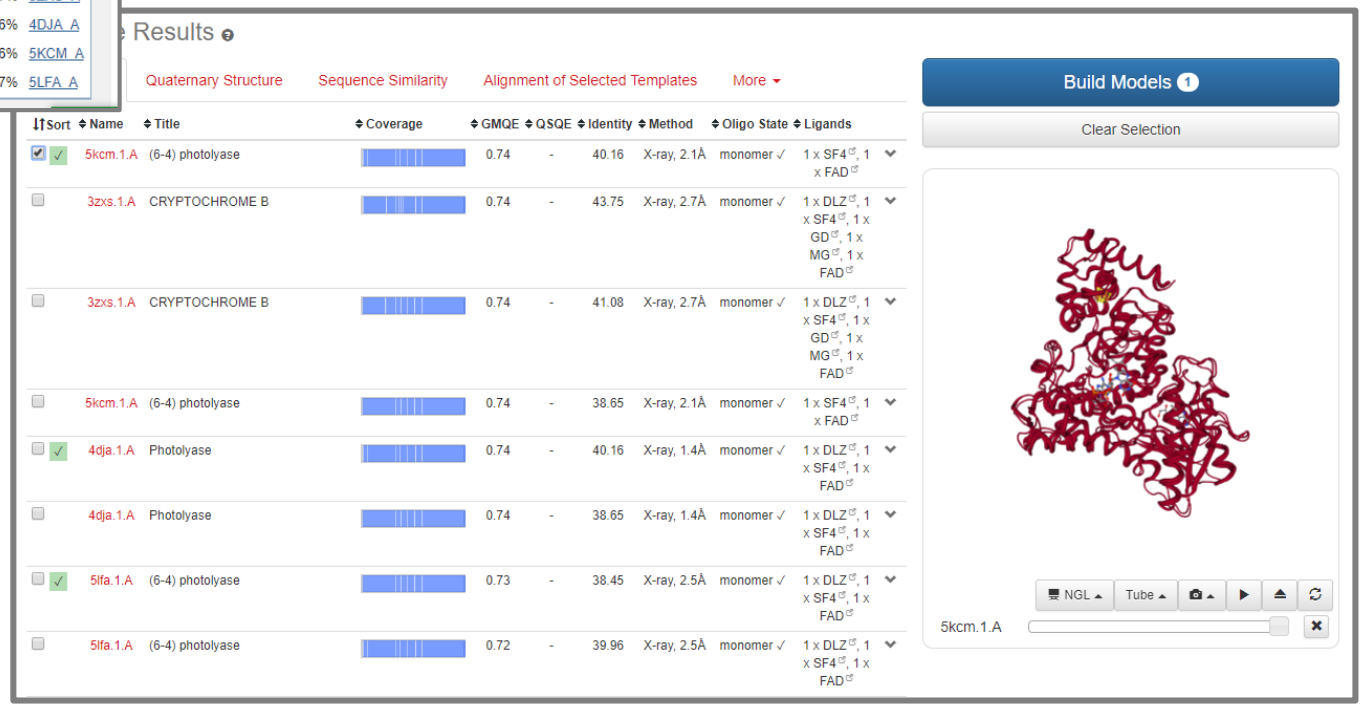
Ugur Meric Dikbas,[†] Mehmet Tardu,[‡] Asena Canturk,[§] Seref Gul,[‡] Gozde Ozcelik,[§] Ibrahim Baris,[†] Nuri Ozturk,[§] and Ibrahim Halil Kavakli*[,†,‡]

[†]Department of Molecular Biology and Genetics, Koc University, Rumelifeneri Yolu, Sariyer, Istanbul 34450, Turkey
[‡]Department of Chemical and Biological Engineering, Koc University, Rumelifeneri Yolu, Sariyer, Istanbul 34450, Turkey
[§]Department of Molecular Biology and Genetics, Gebze Technical University, Gebze 41400, Kocaeli, Turkey

ABSTRACT: Light is crucial for many biological activities of most organisms, including vision, resetting of circadian rhythm, photosynthesis, and DNA repair. The cryptochrome/photolyase family (CPF) represents an ancient group of UV-A/blue light sensitive proteins that perform different functions such as DNA repair, circadian photoreception, and transcriptional regulation. The CPF is widely distributed throughout all organisms, including marine prokaryotes. The bacterium *Vibrio cholerae* was previously shown to have a CPD photolyase that repairs UV-induced thymine dimers and two CRY-DASHs that repair UV-induced single-stranded DNA damage. Here, we characterize a hypothetical gene *Vca0809* encoding a new member of CPF in this organism. The spectroscopic analysis of the purified protein indicated that this enzyme possessed a catalytic cofactor, FAD, and photoantenna chromophore 6,7-dimethyl 8-ribityl-lumazin. With a slot blot-based DNA repair assay, we showed that it possessed (6−4) photolyase activity. Further phylogenetic and computational analyses enabled us to classify this gene as a member of the family of iron−sulfur bacterial cryptochromes and photolyases (FeS-BCP). Therefore, we named this gene *Vc(6−4) FeS-BCP*.

Modelling w Swiss Model





1) «Build models» => will build depending on the selected PDBs

2) Evaluate models

Evaluating Models

Expected values for experimental structure with similar size QMEAN is calculated based on 4-terms; e.g. interaction potentais of these terms

Interatomic distance in model vs homologous proteins; <0.6 poor model

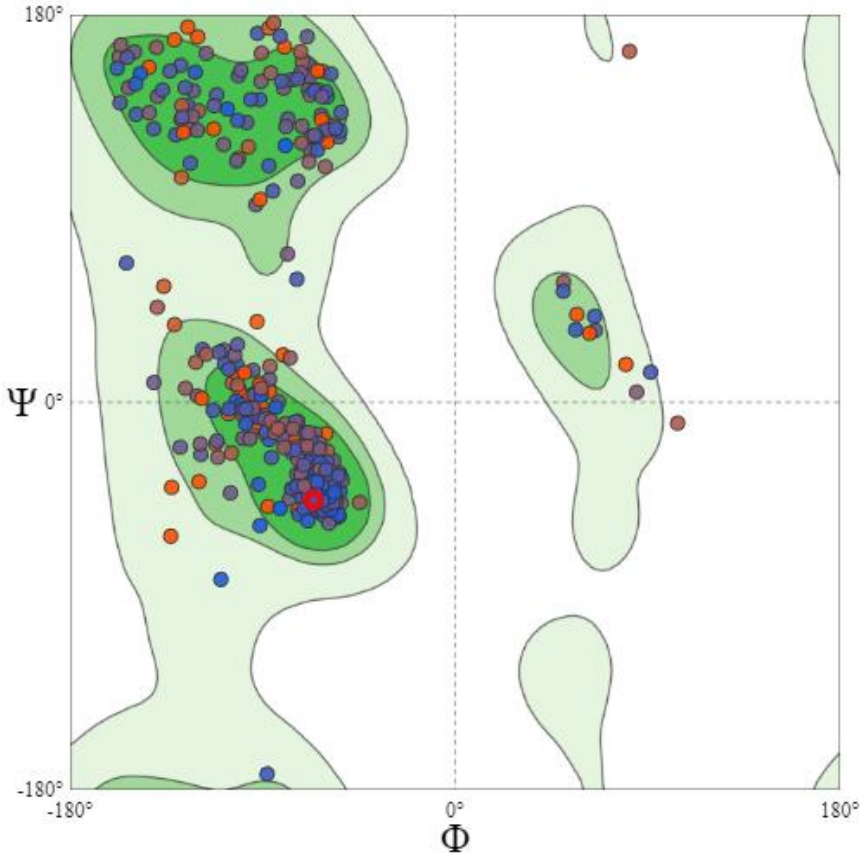
Reflecting the expected accuracy of a model built GMQE score is 0-1 Alignment; template; coverage

The QMEAN Z-score provides an estimate of the «degree of nativeness» QMEAN Z-scores ~0 indicate good agreement; If QMEAN < -4.0 low quality

Normalized QMEAN value wrt residue number

Evaluating Models

1) Ramachandran plot



2) Evaluation of plot; rotamers; bond angles-bonds

MolProbity Results

MolProbity Score	1.63	
Clash Score	5.8	(A469 ARG-A507 ASP), (A437 TYR-A441 MET), (A21 PHE-A68 HIS)
Ramachandran Favoured	95.45%	
Ramachandran Outliers	0.79%	A280 LYS, A221 GLY, A262 HIS, A427 ALA
Rotamer Outliers	0.93%	A205 VAL, A275 PHE, A497 ASP, A395 ASP
C-Beta Deviations	7	A441 MET, A361 ALA, A96 GLN, A137 ALA, A186 LYS, A291 THR, A106 ASP
Bad Bonds	0 / 4225	
Bad Angles	57 / 5731	A395 ASP, A20 TRP, (A203 ASN-A204 PRO), (A262 HIS-A263 PRO), (A248 LEU-A249 PRO), (A143 PHE-A144 PRO), (A411 LEU-A412 PRO), A114 MET, (A217 ILE-A218 PRO), A76 ASP, A266 TRP, A441 MET, A150 LEU, (A84 LEU-A85 PRO), A185 ASN, A106 ASP, (A481 ASN-A482 PRO), A368 PHE, A427 ALA, A292 ILE, A370 TYR, (A139 ILE-A140 PRO), A44 HIS, A153 HIS, (A202 ASP-A203 ASN), A132 PHE, A307 VAL, A373 HIS, A54 ALA, A18 HIS, (A279

Results obtained using MolProbity version 4.4

3) Evaluate models with «SAVES» server

Phyre links:

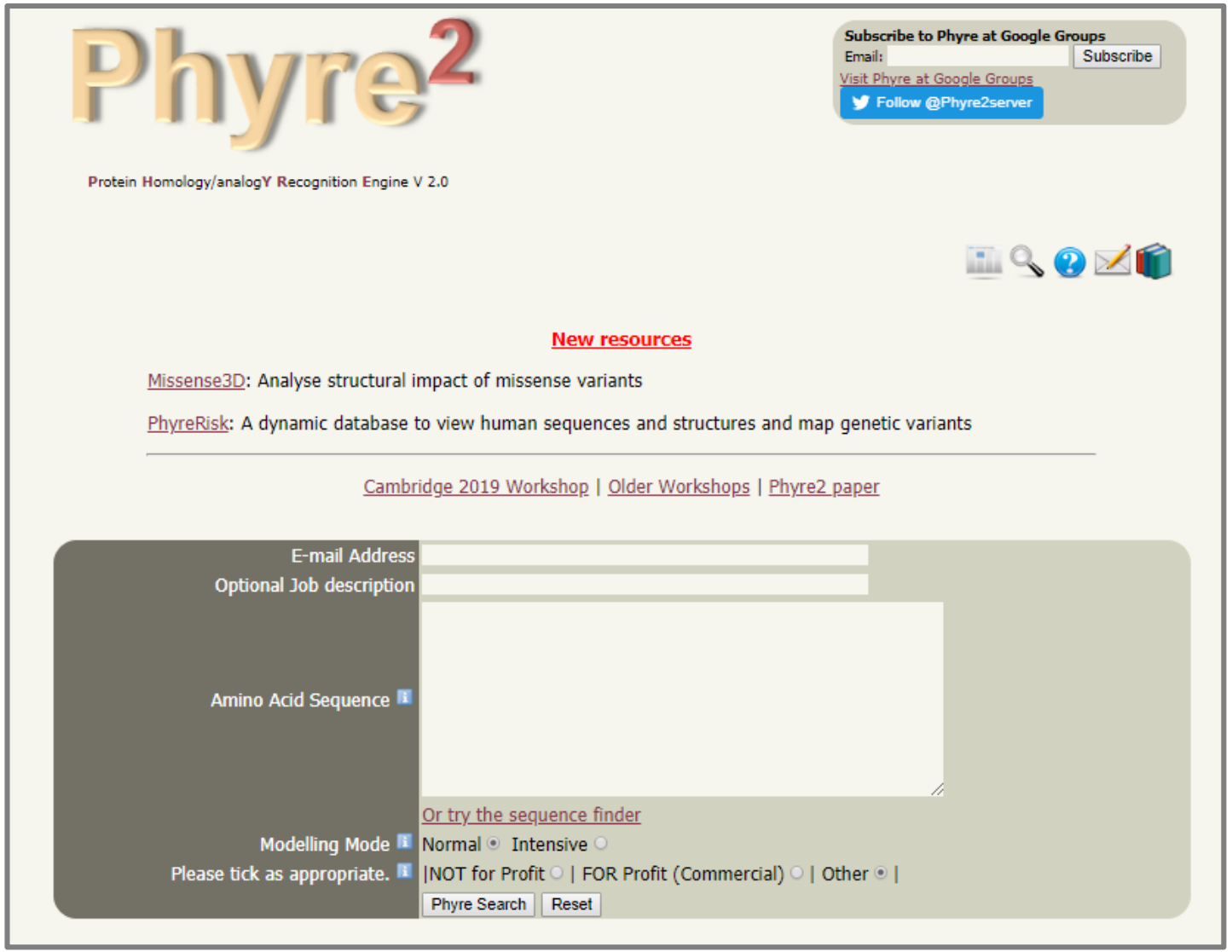
Normal Mode: http://www.sbg.bio.ic.ac.uk/phyre2/phyre2_output/425381f5c127e597/summary.html

Intensive Mode: http://www.sbg.bio.ic.ac.uk/phyre2/phyre2_output/8d49f93928e4c273/summary.html

Modelling w Phyre2

Sequence2:

>NP_001124495.1 peroxisomal biogenesis factor 5 isoform a [Homo sapiens]
MAMRELVEAECGGANPLMKLAGHFTQDKALRQEGLRPGPWPPGAPAS
EAVSVLEVESPGAASEAASKPLGVASEDELVAEFLQDQNAPLVSRAPQTF
KMDDLLAEMQQIEQSNFRQAPQRAPGVADLALSENWAQEFLAAGDAV
DVTQDYNETDWSQEFISEVTDPLSVSPARWAEEYLEQSEEKLWLGEPEGT
ATDRWYDEYHPEEDLQHTASDFVAKVDDPKLANSEFLKFVRQIGEGQVS
LESGAGSGRAQAEQWAAEFIQQQGTSDAWVDQFTRPVNTSALDMEFER
AKSAIESDVDFWDKLQAELEEMAKRDAEAHPWLSDYDDLTSATYDKGY
QFEEENPLRDHPQPFEEGLRRLQEGDLPNAVLLFEAAVQQDPKHMEAW
QYLGTTQAENEQELLAISALRRCLELKPDNQTALMALAVSFTNESLQRQA
CETLRDWLRYTPAYAHLVTPAEEGAGGAGLGPSKRILGSLLSDSLFLEVKE
LFLAAVRLDPTSIDPDVQCGLGVLFNLSGEYDKAVDCFTAALSVRPNDYL
LWNKLGATLANGNQSEEAVAAYRRALELQPGYIRSRYNLGISCINLGAH
REAVEHFLEALNMQRKSRGPRGEGGAMSENIWSTLRLALSMLGQSDAYG
AADARDLSTLLTMFGLPQ



Normal vs Intensive Mode

Phyre links:

Normal Mode: http://www.sbg.bio.ic.ac.uk/phyre2/phyre2_output/425381f5c127e597/summary.html

Intensive Mode: http://www.sbg.bio.ic.ac.uk/phyre2/phyre2_output/8d49f93928e4c273/summary.html

Modelling w Phyre2

Top model

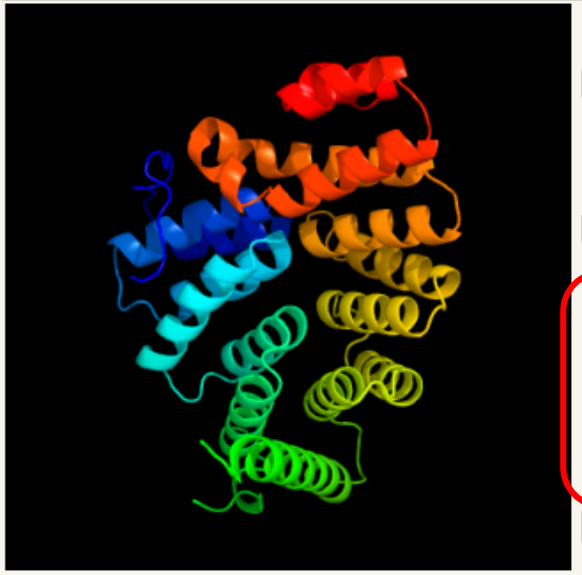


Image coloured by rainbow N → C terminus

Model dimensions (Å): **X**:46.374 **Y**:52.134 **Z**:54.791

Model (left) based on template c1fchB_

Top template information

PDB header:signaling protein
Chain: B: **PDB Molecule:**peroxisomal targeting signal 1 receptor;
PDBTitle: crystal structure of the pts1 complexed to the tpr region2 of human pex5

Confidence and coverage

Confidence: **100.0%** Coverage: **45%**

297 residues (45% of your sequence) have been modelled with 100.0% confidence by the single highest scoring template.
Additional confident templates have been detected (see Domain analysis) which cover other regions of your sequence.
448 residues (69%) could be modelled at >90% confidence using multiple-templates.
You may wish to try resubmitting your sequence in "intensive" mode to model more of your sequence.

3D viewing

Interactive 3D view in JSmol

For other options to view your downloaded structure offline see the FAQ

Download both models and evaluate with «SAVES»

Final Model

Image coloured by rainbow N → C terminus

Model dimensions (Å): **X**:82.767 **Y**:78.441 **Z**:53.849

Download Model

Download zip of all results

Confidence Summary

1 — 200
201 — 400
401 — 600
601

Confidence Key
High(9) Low (0)

69% of residues modelled at >90% confidence (Details)

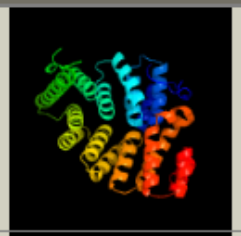
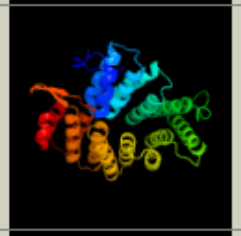
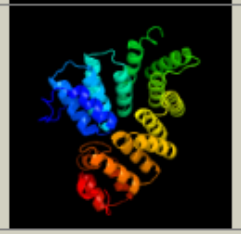
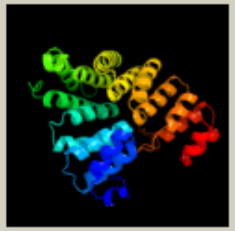
Publication-ready images

Hi-Res image (black background)
Hi-Res image (white background)

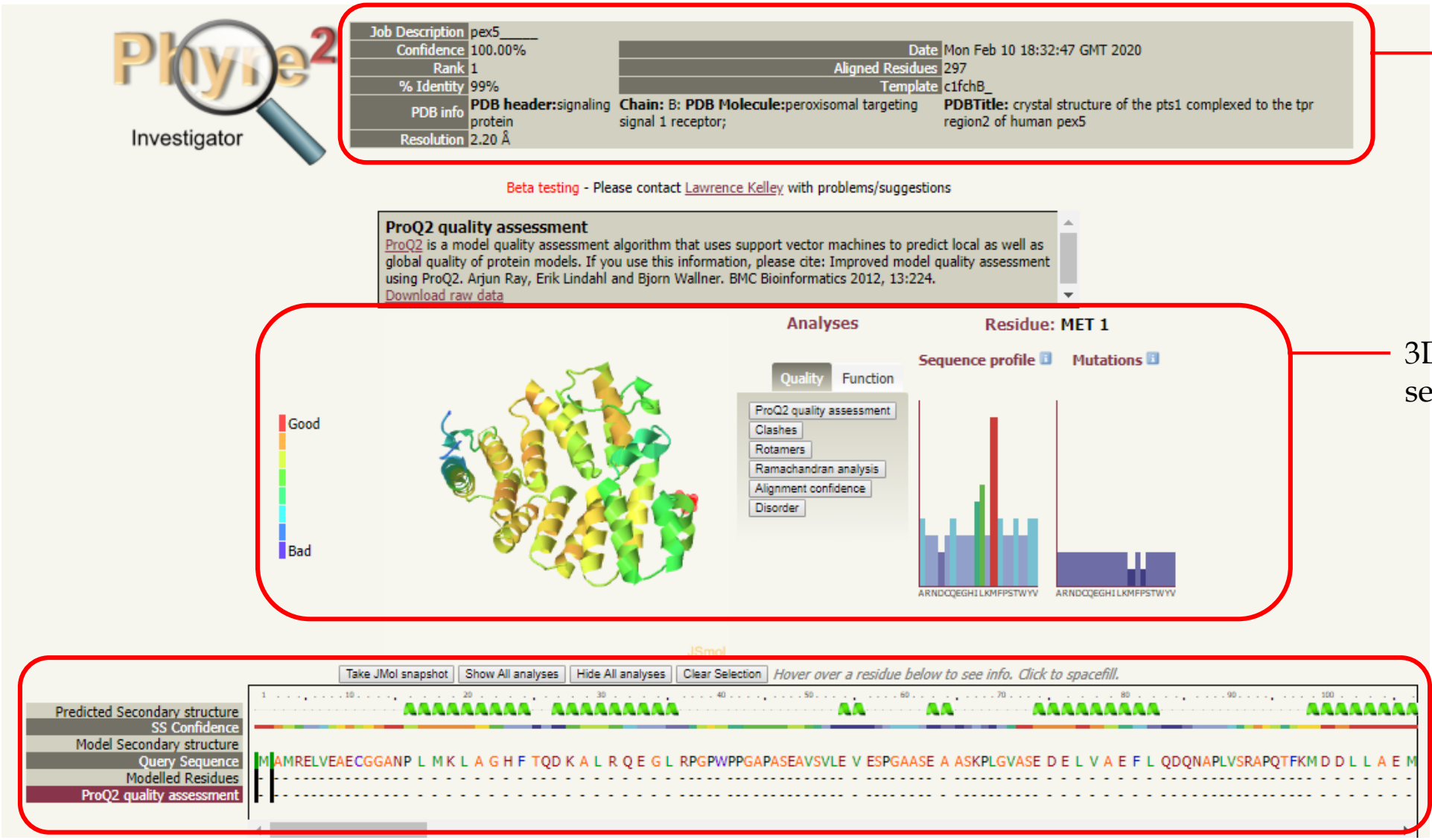
JSmol Viewer

Interactive 3D view in JSmol

Analysing the Phyre2 Models

#	Template	Alignment Coverage	3D Model	Confidence	% i.d.	Template Information
1	c1fchB_	Alignment		100.0	99	**PDB header:**signaling protein **Chain:** B: **PDB Molecule:**peroxisomal targeting signal 1 receptor; **PDBTitle:** crystal structure of the pts1 complexed to the tpr region2 of human pex5 ✓ View investigator results
2	d1fcha_	Alignment		100.0	99	**Fold:**alpha-alpha superhelix **Superfamily:**TPR-like **Family:**Tetratricopeptide repeat (TPR) Run Investigator
3	c4eqfA_	Alignment		100.0	62	**PDB header:**protein binding/transport protein **Chain:** A: **PDB Molecule:**pex5-related protein; **PDBTitle:** trip8b-1a#206-567 interacting with the carboxy-terminal seven residues2 of hcn2 Run Investigator
4	c3cvpA_	Alignment		100.0	40	**PDB header:**transport protein **Chain:** A: **PDB Molecule:**peroxisome targeting signal 1 receptor pex5; **PDBTitle:** structure of peroxisomal targeting signal 1 (pts1) binding domain of2 trypanosoma brucei peroxin 5 (tbpex5)complexed to pts1 peptide (10-3 skl) Run Investigator

Analysing the Phyre2 Models

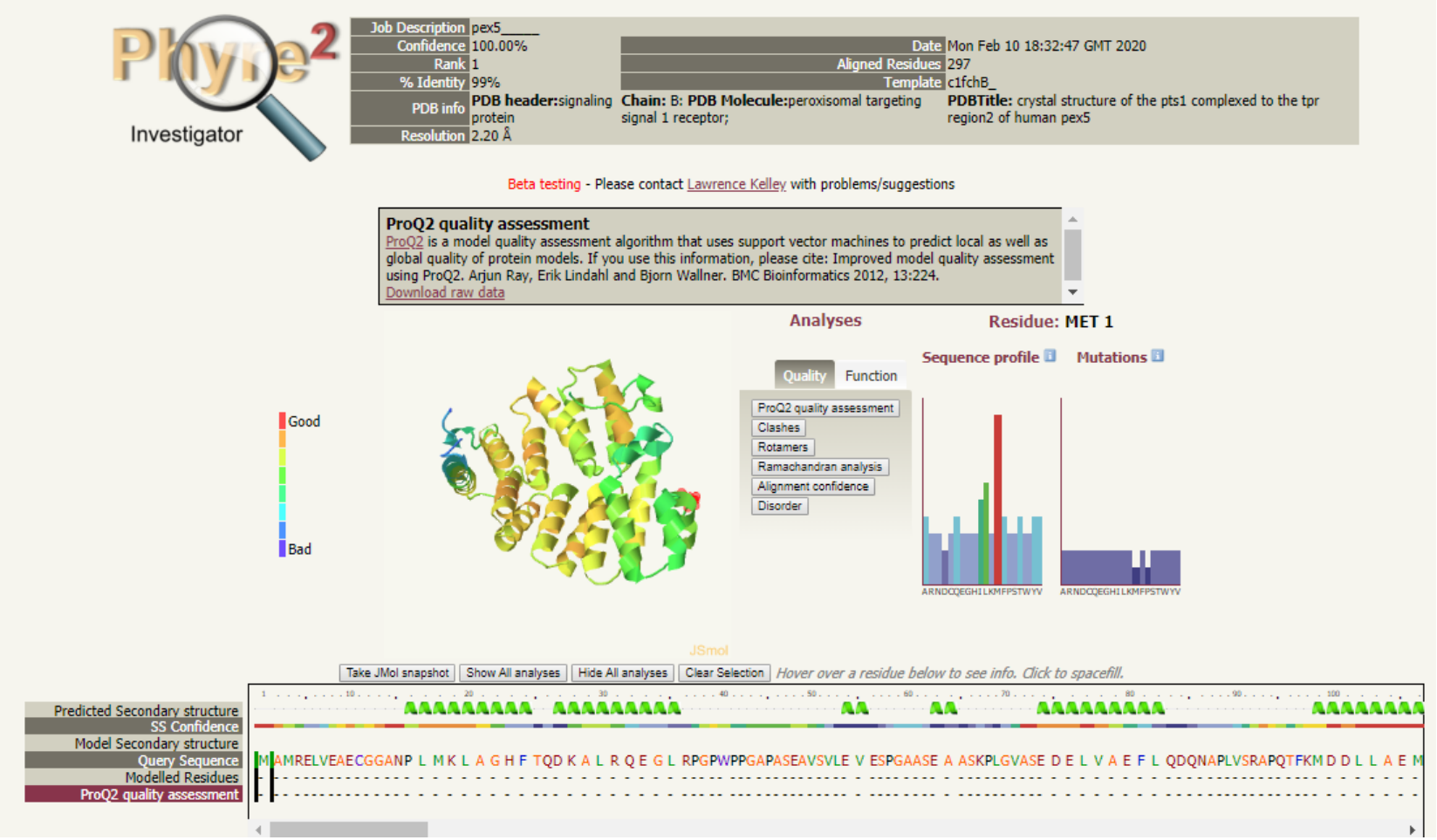


Info section

3D structure and analyses section

Sequence view

Analysing the Phyre2 Models



Mostly based on predictions from DBs

- Quality
  - ProQ2 quality
  - Clashes
  - Rotamers
  - Ramachandran
  - Alignment confidence
  - Disorder
- Function
  - Conservation
  - PI-Site Interface
  - Pocket detection
  - Mutational sensitivity
- Sequence profile
- Mutations

PyMOL Installation

1) Download and install Python2.7 for 64 (https://www.python.org/download/releases/2.7/)

2) Download / Install Pymol Open Source

http://tubiana.me/how-to-install-and-compile-pymol-windows-linux-mac/
Download Pymol_win_64 or Pymol_win_32

3) Copy unzipped folder to C:\

Start => command;
«dir» shows all files in current directory
«cd ../../»
«cd Pymol_win_64»

Open «install_pymol.bat» in text editor

Now, in command prompt
Follow installation steps in «install_pymol.bat»
C:\Python27\Scripts\pip.exe install wheel
C:\Python27\Scripts\pip.exe install --upgrade pip
C:\Python27\Scripts\pip.exe install Pmw-2.0.1-py2-none-any.whl
C:\Python27\Scripts\pip.exe install numpy-1.10.4+mkl-cp27-cp27m-win_amd64.whl
C:\Python27\Scripts\pip.exe install pymol-1.8.2.0-cp27-cp27m-win_amd64.whl
C:\Python27\Scripts\pip.exe install pymol_launcher-1.0-cp27-none-win_amd64.whl
Pymol.exe is in the C:\Python27

Case Study

Dr. Ergoren identified series of mutation on a gene in one of rare diseases. NM number of the gene is: NM_007055.4
Mutation is c.3568C>T (p.Gln1190Ter)

1) Find the sequence of protein

2) Model the protein with your favorite modeling tool

3) Visualize with Pymol

4) Discuss the effect of mutation