



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



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

The screenshot shows the NCBI website interface. At the top, the browser address bar displays <https://www.ncbi.nlm.nih.gov/>. The NCBI logo and navigation links are visible. A search bar is present with a dropdown menu set to "All Databases" and the search term "NM_000531.5" entered. A "Search" button is located to the right of the search bar. The main content area features a "Welcome to NCBI" message and a grid of service tiles: "Submit" (Deposit data or manuscripts into NCBI databases), "Download" (Transfer NCBI data to your computer), "Learn" (Find help documents, attend a class or watch a tutorial), "Develop" (Use NCBI APIs and code libraries to build applications), "Analyze" (Identify an NCBI tool for your data analysis task), and "Research" (Explore NCBI research and collaborative projects). On the right side, there are sections for "Popular Resources" (PubMed, Bookshelf, PubMed Central, BLAST, Nucleotide, Genome, SNP, Gene, Protein, PubChem) and "NCBI News & Blog" (Read about NCBI resources in 2020 Nucleic Acids Research database issue, The 2020 Nucleic Acids Research database issue features papers from, NLM announces Curation at Scale Workshop, Data curation plays a critical role in today's biomedical research and ensures, Rapid access to 2019-nCoV (Wuhan coronavirus) data from the current public health emergency).

3) Search for CDS and click NP_number

exon 1..170
/gene="OTC"
/gene_synonym="OCTD"
/inference="alignment:Splign:2.1.0"

misc_feature 85..87
/gene="OTC"
/gene_synonym="OCTD"
/note="upstream in-frame stop codon"

CDS 94..1158
/gene="OTC"
/gene_synonym="OCTD"
/EC_number="2.1.3.3"
/note="ornithine transcarbamylase; ornithine carbamoyltransferase, mitochondrial; OTCase"
/codon_start=1
/product="ornithine carbamoyltransferase, mitochondrial precursor"
/protein_id="NP_000522.3"
/db_xref="CDS:CDS14247.1"
/db_xref="GeneID:5009"
/db_xref="HGNC:HGNC:8512"
/db_xref="MIM:300461"
/translation="MLFNLRILLNNAAFRNGHNFMRNFRCGQPLQNKVQLKGRDLLT LKNFTGEEIKYMLWLSADLKFRKQKGEYLPLLQGKSLGMIFEKRSTRTRLSTETGFA LLGGHPCFLTQTQDIHLGVNESLTDARVLSMADAVLARVYQSDLDTLAKEASIPII NGLSDLYHPIQILADYLTQEHYSSLKGLTLSWIGDGNILHSIMMSAAKFGMHLQAA TPKGYEPDASVTKLAEQYAKENGTKLLLNDPLEAAHGGNVLITDTWISMGQEEKKK RLQAFQGYQVTMKTAKVAASDWTFLHCLPRKPEEVDDVEVFYSPRSLVPEAENRKWTI MAVMVSLLTDYSPQLQKPKF"

transit_peptide 94..189
/gene="OTC"
/gene_synonym="OCTD"

4) Get FASTA from NP_number

The screenshot shows the NCBI protein search interface. The search bar contains the text 'Protein' and a search button. The page title is 'ornithine carbamoyltransferase, mitochondrial precursor [Homo sapiens]'. The NCBI Reference Sequence is NP_000522.3. There are links for 'Identical Proteins', 'FASTA', and 'Graphics'. The 'GenPept' tab is selected, and the 'Send to' dropdown is set to 'Change region shown'.

NCBI Reference Sequence: NP_000522.3
[Identical Proteins](#) [FASTA](#) [Graphics](#)

[Go to:](#)

LOCUS NP_000522 354 aa linear PRI 02-JAN-2020
DEFINITION ornithine carbamoyltransferase, mitochondrial precursor [Homo sapiens].
ACCESSION NP_000522
VERSION NP_000522.3
DBSOURCE REFSEQ: accession [NM_000531.6](#)
KEYWORDS RefSeq; MANE Select.
SOURCE Homo sapiens (human)
ORGANISM [Homo sapiens](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.
REFERENCE 1 (residues 1 to 354)
AUTHORS Chongsrisawat V, Damrongphol P, Ittiwut C, Ittiwut R,

The screenshot shows the NCBI FASTA view for the protein. The search bar is empty. The page title is 'ornithine carbamoyltransferase, mitochondrial precursor [Homo sapiens]'. The NCBI Reference Sequence is NP_000522.3. There are links for 'GenPept', 'Identical Proteins', and 'Graphics'. The 'FASTA' tab is selected, and the 'Send to' dropdown is set to 'Change region shown'. The FASTA sequence is displayed in a monospace font.

NCBI Reference Sequence: NP_000522.3
[GenPept](#) [Identical Proteins](#) [Graphics](#)

```
>NP_000522.3 ornithine carbamoyltransferase, mitochondrial precursor [Homo sapiens]
MLFNLRIILLNNAAFRNGHNFMRNFRCGQPLQNKVQLKGRDLLTLKNFTGEEIKYMLWLSADLKFRIKQK
GEYLPLLQGKSLGMIFEKRSTRRLSTETGFALLGGHPCFLTTQDIHLGVNESLTDARVLSMADAVLA
RVYKQSDLDLAKAEASIPINGLSDLYHPYIQLADYLTQEHYSSLKGLTSLWIGDGNLILHSIMMSAAK
FGMHLQAATPKGYEPDASVTKLAEQYAKENGTKLLL TNDPLEAAHGGNVLITDTWISMGQEEKKRLQA
FQGYQVTMKTAKVAASDWFHLHCLPRKPEEVDDVFYSPRSLVFEAENRKTIMAVMVSLLTDYSPQLQ
KPKF
```

Copy FASTA to Notepad

The 'Protein 3D Structure' section shows a 3D ribbon diagram of the protein structure. The structure is colored in shades of purple and yellow. The resolution is 3.5 Å. The source is Ovis aries (goat), and the method is X-Ray Diffraction.

Analyze this sequence

- Run BLAST
- Identify Conserved Domains
- Highlight Sequence Features
- Find in this Sequence
- Show in Genome Data Viewer

Protein 3D Structure

Low Resolution Structure Of Ovine Ornithine
PDB: 1FB5
Source: Ovis aries
Method: X-Ray Diffraction
Resolution: 3.5 Å

To search for protein structures to our protein of interest 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

The image shows a screenshot of the NCBI BLASTP search interface. The browser address bar shows the URL <https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>. The page title is "BLAST >> blastp suite". The main heading is "Standard Protein BLAST".

The "Enter Query Sequence" section contains a text input field with the text "Paste FASTA". To the right of this field are "Clear" and "Query subrange" options. Below the text field are "Or, upload file" (with a "Choose File" button and "No file chosen" text) and "Job Title" (with a text input field and the instruction "Enter a descriptive title for your BLAST search"). There is also a checkbox for "Align two or more sequences".

The "Choose Search Set" section is highlighted with a red box. It contains a "Database" dropdown menu set to "Protein Data Bank proteins(pdb)". Below this are "Organism" (Optional) and "Exclude" (Optional) sections. The "Organism" section has a text input field and an "exclude" checkbox. The "Exclude" section has three checkboxes: "Models (XM/XP)", "Non-redundant RefSeq proteins (WP)", and "Uncultured/environmental sample sequences".

The "Program Selection" section contains an "Algorithm" dropdown menu with three options: "blastp (protein-protein BLAST)" (selected), "PSI-BLAST (Position-Specific Iterated BLAST)", and "PHI-BLAST (Pattern Hit Initiated BLAST)".

On the right side of the page, there is a notification box that says: "BLAST results will be displayed in a new format by default. You can always switch back to the Traditional Results page." There is also a "New" badge on a small icon next to the notification.

Modelling w Swiss Model

Sequence1

>NP_233195.1 hypothetical protein VCA0809 [Vibrio cholerae O1 biovar El Tor str. N16961]
MRYSVRLILGDQLNHAHSWFSEHRDDVLYLIAELHQEQEYVRHHIQKQCAFFAAMQAFADYLSAEGHHV
WHLDLASAQYNDLPDLIAQICQQVQADAFQYQRPDEYRLLEQMANLRLSGITIGCVDTTEHFLLPFAEIP
EQFPASKAVLMEHFYRRMRKRFGYLMTADGKPEGGQWNFDADNRNKLKSPDLLQLPTPLCFDNPVASIKA
RIERHRIPSIGQVGESLLWPINRAQALSLLAHFCQICLPNFGRFQDAMTAQHPRWSLYHSRSLFALNSK
LLSPREVIEATISAYRAAQGQISLAQVEGFVRQILGWREYVRGMYWSNMPHYQTRNHLGAQRPLPSYFWN
GQTKMRCLQQAITQSLDFGYAHHIQRLMVTGNFALLTECDPDQVDAWYLGIIYIDAIEWVELPNTRGMALF
ADGGLIATKPYSASGSYINKMSDYCASCA YQVKLKS GEKACPLNSLYWRFMLKHRDRLANNPRIGMLYKT
WDKMTSDSQAILSTADAYLSQIESL

- 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

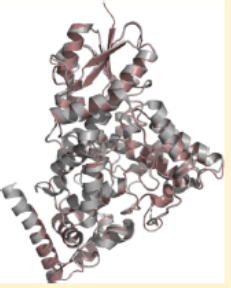
Biochemistry Article
Cite This: *Biochemistry* 2019, 58, 4352–4360 pubs.acs.org/biochemistry

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

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

Job Title NP_233195.1 hypothetical protein VCA0809 [Vibrio...]
RID [45CWDC05014](#) *Search expires on 02-12 21:00 pm* [Download All](#)
Program BLASTP [Citation](#)
Database pdb [See details](#)
Query ID Ict|Query_80633
Description NP_233195.1 hypothetical protein VCA0809 [Vibrio cholerae (...]
Molecule type amino acid
Query Length 516
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 to **E value** to **Query Coverage** to
[Filter](#) [Reset](#)

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

Sequences producing significant alignments [Download](#) [Manage Columns](#) [Show](#) 100 [?](#)

select all 4 sequences selected [GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#)

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	Chain A_Cryptochrome B [Rhodobacter sphaeroides 2.4.1]	387	387	98%	2e-129	42.47%	3ZXS_A
<input checked="" type="checkbox"/>	Chain A_Photosynthase [Agrobacterium fabrum str. C58]	370	370	98%	7e-123	39.06%	4DJA_A
<input checked="" type="checkbox"/>	Chain A_(6-4) photolyase [Agrobacterium fabrum str. C58]	370	370	98%	7e-123	39.06%	5KCM_A
<input checked="" type="checkbox"/>	Chain A_(6-4) photolyase [Agrobacterium fabrum str. C58]	369	369	98%	3e-122	38.87%	5LFA_A

Results

[Quaternary Structure](#) [Sequence Similarity](#) [Alignment of Selected Templates](#) [More](#)

[Build Models 1](#)
[Clear Selection](#)

	Sort	Name	Title	Coverage	GMQE	QSQE	Identity	Method	Oligo State	Ligands
<input checked="" type="checkbox"/>		5kcm.1.A	(6-4) photolyase		0.74	-	40.16	X-ray, 2.1Å	monomer	1 x SF4 ^{CS} , 1 x FAD ^{CS}
<input type="checkbox"/>		3zxs.1.A	CRYPTOCHROME B		0.74	-	43.75	X-ray, 2.7Å	monomer	1 x DLZ ^{CS} , 1 x SF4 ^{CS} , 1 x GD ^{CS} , 1 x MG ^{CS} , 1 x FAD ^{CS}
<input type="checkbox"/>		3zxs.1.A	CRYPTOCHROME B		0.74	-	41.08	X-ray, 2.7Å	monomer	1 x DLZ ^{CS} , 1 x SF4 ^{CS} , 1 x GD ^{CS} , 1 x MG ^{CS} , 1 x FAD ^{CS}
<input type="checkbox"/>		5kcm.1.A	(6-4) photolyase		0.74	-	38.65	X-ray, 2.1Å	monomer	1 x SF4 ^{CS} , 1 x FAD ^{CS}
<input checked="" type="checkbox"/>		4dja.1.A	Photolyase		0.74	-	40.16	X-ray, 1.4Å	monomer	1 x DLZ ^{CS} , 1 x SF4 ^{CS} , 1 x FAD ^{CS}
<input type="checkbox"/>		4dja.1.A	Photolyase		0.74	-	38.65	X-ray, 1.4Å	monomer	1 x DLZ ^{CS} , 1 x SF4 ^{CS} , 1 x FAD ^{CS}
<input checked="" type="checkbox"/>		5lfa.1.A	(6-4) photolyase		0.73	-	38.45	X-ray, 2.5Å	monomer	1 x DLZ ^{CS} , 1 x SF4 ^{CS} , 1 x FAD ^{CS}
<input type="checkbox"/>		5lfa.1.A	(6-4) photolyase		0.72	-	39.96	X-ray, 2.5Å	monomer	1 x DLZ ^{CS} , 1 x SF4 ^{CS} , 1 x FAD ^{CS}

5kcm.1.A

[NGL](#) [Tube](#) [Camera](#) [Play](#) [Up](#) [Down](#) [Close](#)

- 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 potentials 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

Model Results Order by: GMQE

Oligo-State: Monomer
Ligands: None

Global Quality Estimate

QMEAN	-2.74
C β	-2.31
All Atom	-1.08
solvation	-0.26
torsion	-2.24

Local Quality Estimate: Residue Number vs. Quality Estimate plot

Comparison: Protein Size (Residues) vs. Quality Estimate plot

Model 01 Structure Assessment

Template	Seq Identity	Coverage	Description
5kcm.1.A	40.16%		(6-4) photolyase

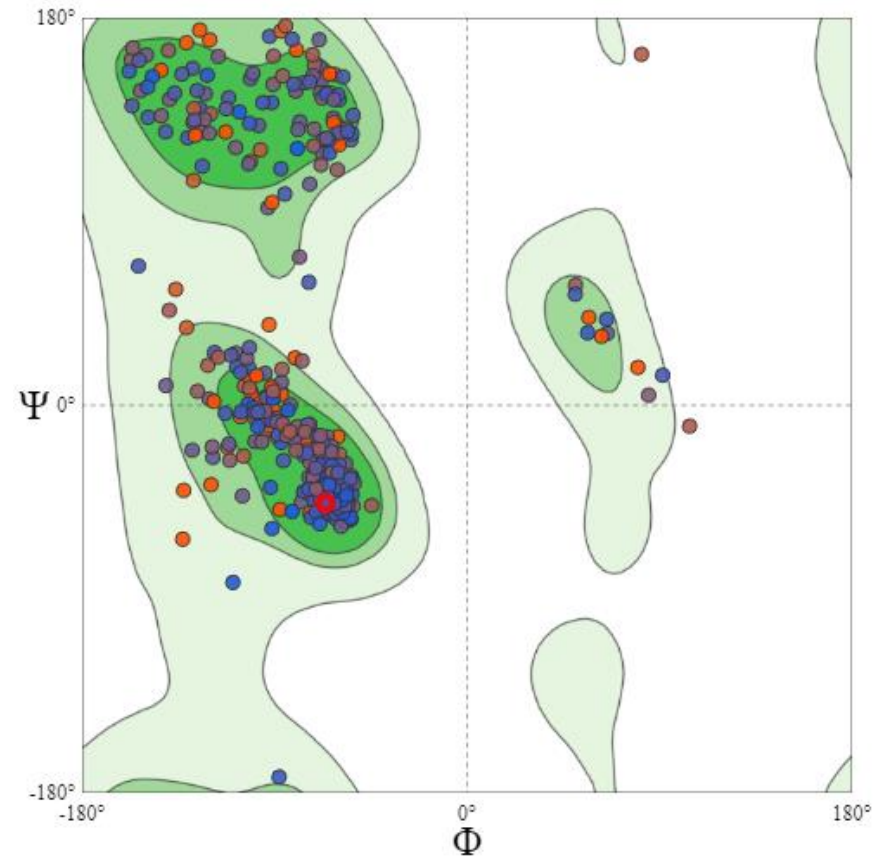
Model-Template Alignment

```
Model_01 MRYSVVR LILGDQLNHAHSWFS--EHRDDVLYLIAELHQEQEYVRRHHIQKQCAFFAAMQAFADYLSAEGHHVWHLDLDAQAQYN 83
5kcm.1.A -----DILGDQLSPS(AA)DGVDKKQD(TI-VI)EVMMAEASVYVGH(KK)KIAFNFSSAMRHFAEELRGEYR(VRY)AIDDDADNAGS 82
Model_01 LPDLIAQICQQVQADAFQYQRPDEYRLLEQMANLRLS-GITIGCVDTTEHFLLPFAEIPQFPASKAVLMEHFYRRMRKREGYLMT 167
5kcm.1.A GE(D)SVGGRWNE(DA)NRQPAP-PDLLRPKHPVFAPD(KITKEVIDEYER)FPDNEGKL-ENFGFAVT(RTD)AERALS(FID)S(LCNF) 167
Model_01 ADGKPEGGQWNFDADNRNKLKSPDLLQLPPTPL-CFDNPVASIKARIERHRIPSIQGVGESLLWPINRAQALSLLAHFCQIQLPNE 251
5kcm.1.A GE(D)SVGGRWNE(DA)NRQPAP-PDLLRPKHPVFAPD(KITKEVIDEYER)FPDNEGKL-ENFGFAVT(RTD)AERALS(FID)S(LCNF) 249
Model_01 CRFODAMTAQHPRWSLYHSRSLSFALNSKLLSPREVIEATISAYRAAQQGISLAQVEGFVRQILGWREYVRCGMYSNMPHYQTRN 326
5kcm.1.A GATQDA(M)X(D)PN---LNHSLI(SFY)NCGLLI(ALLVCKAAREPAY)---EGGAT(LNAVEGFIRQI)IGWREYVRCGMYSNMPHYQTRN 329
Model_01 HLGARPLPSYFWNGQTKMRCLOQAITQSLDEGYAHHIQRLMVTGNFALDTECDPDQVDAWYLGIIYIDAIEWVELPNTRCGMALFA 421
5kcm.1.A PFENIDSLPV(YW)GKTHM(CMAKAVITETIE)AYAH(HI)QRLMVTGNFALDTECDPDQVDAWYLGIIYIDAIEWVELPNTRCGMALFA 414
Model_01 DGGLIATKPYSASGSYINKMSDYCASCAYQVKKLSGEKACPLNSLYWRFMLKRRDRLDANNPRIGMLYKTWDKMTSDSQAILSTA 506
5kcm.1.A DGGFLGKTPYAS(GNYINR)SDYCD(CRY)EKERIGD(S)CFENALYWFLEAS(REK)KSN(HRLAQPYATW)M(CDVRHDLRAKA 499
Model_01 DAYLSQTE SL 516
5kcm.1.A AAFLRKLD-- 507
```

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	
<input type="checkbox"/> Clash Score	5.8	(A469 ARG-A507 ASP), (A437 TYR-A441 MET), (A21 PHE-A68 HIS)
Ramachandran Favoured	95.45%	
<input type="checkbox"/> Ramachandran Outliers	0.79%	A280 LYS, A221 GLY, A262 HIS, A427 ALA
<input type="checkbox"/> Rotamer Outliers	0.93%	A205 VAL, A275 PHE, A497 ASP, A395 ASP
<input type="checkbox"/> C-Beta Deviations	7	A441 MET, A361 ALA, A96 GLN, A137 ALA, A186 LYS, A291 THR, A106 ASP
Bad Bonds	0 / 4225	
<input type="checkbox"/> 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]
MAMRELVEAECGGANPLMKLAGHFTQDKALRQEGLRPGWPWPPGAPAS
EAVSVLEVESPGAASEAASKPLGVASEDELVAEFLQDQNAPLVSRAPQTF
KMDDLAEQQIEQSNFRQAPQRAPGVADLALSENWAQEFLAAGDAV
DVTQDYNEDTWSQEFISEVTDPLSVSPARWAEYLEQSEEKWLWGEPEGT
ATDRWYDEYHPEEDLQHTASDFVAKVDDPKLANSEFLKFVRQIGEGQVS
LESGAGSGRAQAEQWAAEFIQQQGTSDAWVDQFTRPVNTSALDMEFER
AKSAIESDVDFWDKLAELAELEMAKRDAEAHPWLSDYDDLTSATYDKGY
QFEEENPLRDHPQPFEGLRRLQEGDLPNAVLLFEAAVQQDPKHMEAW
QYLGTTQAENEQELLAISALRRCLELKPDNQTALMALAVSFTNESLQRQA
CETLRDWLRYTPAYAHLVTPAEEGAGGAGLGPSKRILGSLSDSLFLEVKE
LFLAAVRLDPTSIDPDVQCGLGVLFNLSGEYDKAVDCFTAALSVRPNDYL
LWNKLGATLANGNQSEEAVAAAYRRALELQPGYIRSRYNLGISINLGAH
REAVEHFLEALNMQRKSRGPRGEGGAMSENIWSTLRLALSMLGQSDAYG
AADARDLSTLLTMFGLPQ

Normal vs Intensive Mode

Phyre links:

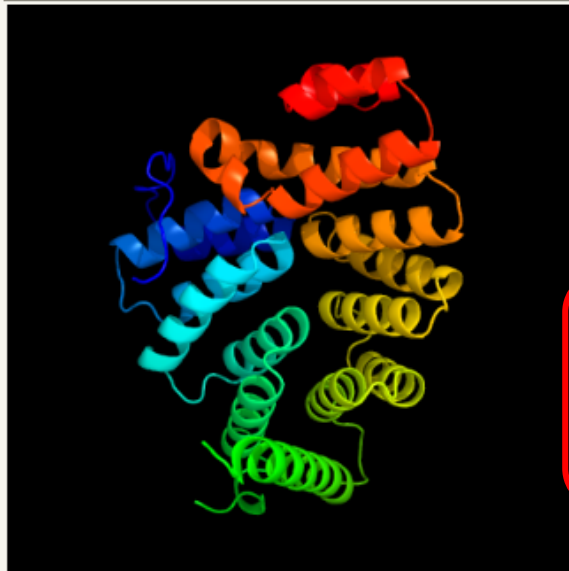
Normal Mode: http://www.sbg.bio.ic.ac.uk/phyre2/phyre2_output/425381f5c127e597/summary.html

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

The screenshot displays the Phyre2 website interface. At the top left is the Phyre2 logo, and below it is the text "Protein Homology/analogy Recognition Engine V 2.0". In the top right corner, there are social media links: "Subscribe to Phyre at Google Groups" with an email input field and a "Subscribe" button, and "Visit Phyre at Google Groups" with a "Follow @Phyre2server" button. Below these are navigation icons for home, search, help, and contact. A "New resources" section lists "Missense3D: Analyse structural impact of missense variants" and "PhyreRisk: A dynamic database to view human sequences and structures and map genetic variants". Below this is a horizontal line and a row of links: "Cambridge 2019 Workshop | Older Workshops | Phyre2 paper". The main search form is a dark grey box with the following fields: "E-mail Address" (input field), "Optional Job description" (input field), "Amino Acid Sequence" (input field with a blue 'x' icon), "Modelling Mode" (radio buttons for "Normal" and "Intensive", with "Normal" selected), and "Please tick as appropriate." (radio buttons for "NOT for Profit", "FOR Profit (Commercial)", and "Other", with "Other" selected). At the bottom of the form are "Phyre Search" and "Reset" buttons. A link "Or try the sequence finder" is located above the "Modelling Mode" section.

Modelling w Phyre2

Top model



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](#)

Image coloured by rainbow N → C terminus

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

Final Model

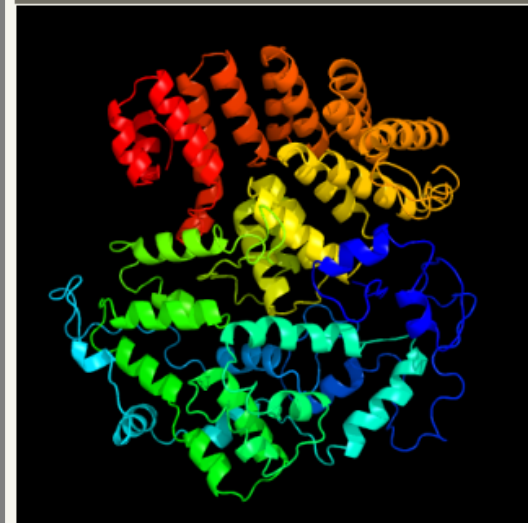


Image coloured by rainbow N → C terminus

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

Download Model

Download zip of all results

Confidence Summary



Confidence Key

High(9) Low (0)

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

Publication-ready images

[Hi-Res image \(black background\)](#)


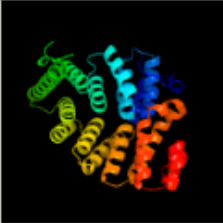

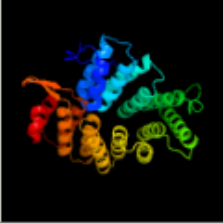

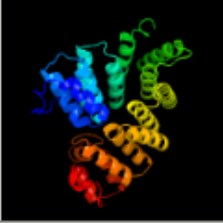

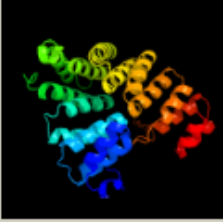
[Hi-Res image \(white background\)](#)

JSmol Viewer

[Interactive 3D view in JSmol](#)

Download both models and evaluate with «SAVES»

Analysing the Phyre2 Models

#	Template	Alignment Coverage	3D Model	Confidence	% i.d.	Template Information
1	c1fchB <input type="radio"/> <input type="checkbox"/>	 <input type="button" value="Alignment"/>		100.0	99	<p>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</p> <p>View investigator results</p>
2	d1fcha <input type="radio"/> <input type="checkbox"/>	 <input type="button" value="Alignment"/>		100.0	99	<p>Fold:alpha-alpha superhelix Superfamily:TPR-like Family:Tetratricopeptide repeat (TPR)</p> <p>Phyre2 Run Investigator</p>
3	c4eqfA <input type="radio"/> <input type="checkbox"/>	 <input type="button" value="Alignment"/>		100.0	62	<p>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</p> <p>Phyre2 Run Investigator</p>
4	c3cypA <input type="radio"/> <input type="checkbox"/>	 <input type="button" value="Alignment"/>		100.0	40	<p>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)</p> <p>Phyre2 Run Investigator</p>

Analysing the Phyre2 Models



Job Description	pex5	Date	Mon Feb 10 18:32:47 GMT 2020
Confidence	100.00%	Aligned Residues	297
Rank	1	Template	c1fchB_
% Identity	99%	PDB header	signaling protein
PDB info	Chain: B; PDB Molecule: peroxisomal targeting signal 1 receptor;	PDBTitle	crystal structure of the pts1 complexed to the tpr region2 of human pex5
Resolution	2.20 Å		

Info section

Beta testing - Please contact [Lawrence Kelley](#) with problems/suggestions

ProQ2 quality assessment

ProQ2 is a model quality assessment algorithm that uses support vector machines to predict local as well as global quality of protein models. If you use this information, please cite: Improved model quality assessment using ProQ2. Arjun Ray, Erik Lindahl and Bjorn Wallner. BMC Bioinformatics 2012, 13:224.
[Download raw data](#)

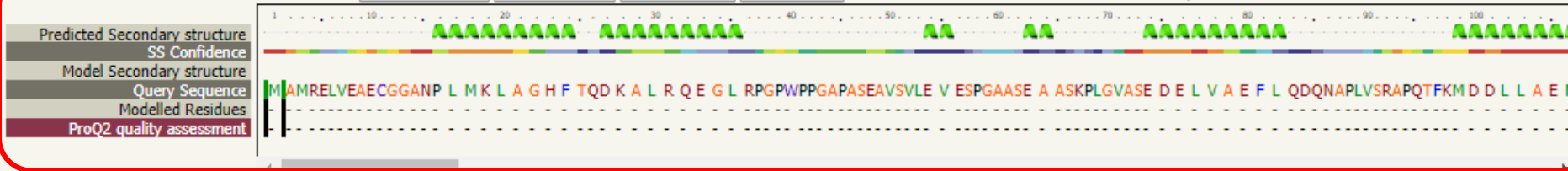
Analyses

Residue: MET 1



3D structure and analyses section

Take Jmol snapshot Show All analyses Hide All analyses Clear Selection Hover over a residue below to see info. Click to spacefill.



Sequence view

Analysing the Phyre2 Models



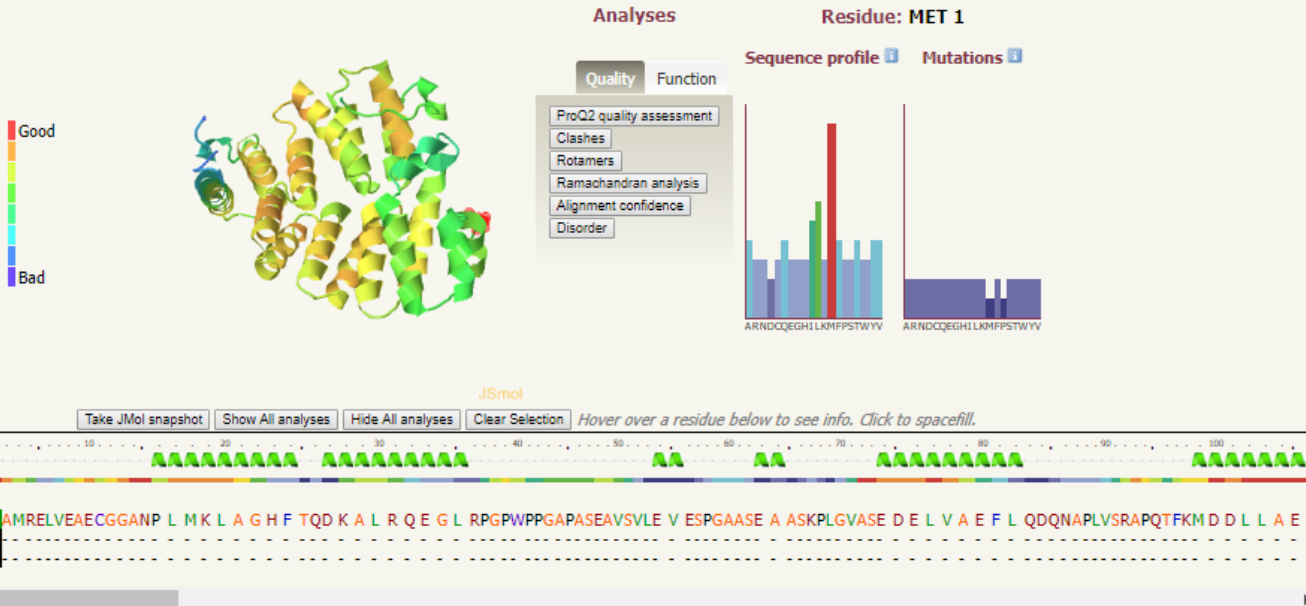
Job Description	pex5	Date	Mon Feb 10 18:32:47 GMT 2020
Confidence	100.00%	Aligned Residues	297
Rank	1	Template	c1fchB_
% Identity	99%	PDB header:signaling protein	Chain: B; PDB Molecule:peroxisomal targeting signal 1 receptor; PDBtitle: crystal structure of the pts1 complexed to the tpr region2 of human pex5
Resolution	2.20 Å		

Beta testing - Please contact [Lawrence Kelley](#) with problems/suggestions

ProQ2 quality assessment

ProQ2 is a model quality assessment algorithm that uses support vector machines to predict local as well as global quality of protein models. If you use this information, please cite: Improved model quality assessment using ProQ2. Arjun Ray, Erik Lindahl and Bjorn Wallner. BMC Bioinformatics 2012, 13:224.

[Download raw data](#)



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