

# HOMOLOGY MODELING AND STRUCTURAL ANALYSIS OF FIBROCYSTIN PROTEIN



# CONTENTS:

- 1. Literature Survey**
- 2. Protein Information**
- 3. Sequence**
- 4. BLAST**
- 5. MULTIPLE SEQUENCE ALIGNMENT**
- 6. HOMOLOGY MODELLING**
- 7. REFERENCES**
- 8. SUMMARY**

# LITERATURE SURVEY

## **Polycystic kidney disease 4, with polycystic liver disease:**

This disease is caused by variants affecting the gene. It is an autosomal recessive disease which is characterized by enlarged, echogenic kidneys with fusiform dilatation of the collecting ducts. Most patients progress to end-stage renal disease (ESRD), but at varying ages.

It has long been recognized that the age distribution of cases of polycystic kidneys has 2 peaks, one at birth and one between ages 30 to 60 years. The clinical spectrum is widely variable, with most cases presenting during infancy. The fetal phenotypic features classically include enlarged and echogenic kidneys, as well as oligohydramnios secondary to a poor urine output. Up to 50% of the affected neonates die shortly after birth, as a result of severe pulmonary hypoplasia and secondary respiratory insufficiency.

# LITERATURE SURVEY(cont...)

The signs of ARPKD frequently begin before birth, so it is often called "infantile PKD" but some people do not develop symptoms until later in childhood or even adulthood. Children born with ARPKD often, but not always, develop kidney failure before reaching adulthood; babies with the worst cases die hours or days after birth due to respiratory difficulties or respiratory failure. Liver scarring occurs in all patients. The condition is caused by a mutation in the PKHD1 gene and is inherited in an autosomal recessive manner. Some symptoms of the condition may be controlled by medicines, antibiotics, healthy diet, and growth hormones.

## **Symptoms:**

ARPKD is asymptomatic, but the symptoms become more frequent with age like jaundice, weight loss, pain in the lower back region etc.

# PROTEIN

## Function of Fibrocystin protein in disease:

- Promotes ciliogenesis in renal epithelial cells and therefore participates in the tubules formation and ensures the maintenance of the architecture of the lumen of the kidney.
- Has an impact on cellular symmetry by ensuring correct bipolar cell division through the regulation of centrosome duplication and mitotic spindle assembly.
- Regulates cell-cell and cell-matrix adhesion and participates in cell motility.
- Promotes cell-cell contact through the positive regulation.



# SEQUENCE RETRIEVAL

**Uniprot Id:** P08F94

**Protein:** Fibrocystin

**Gene:** PKHD1

**Organism:** Homosapiens

**Length:** 4,074

**Mass (Da):** 446,702

# SEQUENCE (FASTA FORMAT)

```
>sp|P08F94|PKHD1_HUMAN Fibrocystin OS=Homo sapiens OX=9606 GN=PKHD1 PE=1 SV=1
MTAWLISLMSIEVLLLAVRHLSLHIEPEEGLAGGTWITVIFDGLELGLVLYPNNGSQLEI
HLVNVNMVVPALRSVPCDVFPVFLDLPVVTCTRSVLSEAHEGLYFLEAYFGGQLVSSPN
PGPRDSCCTFKFSKAQTPIVHQVYPPSGVPGKLIHVYGWIITGRLETDFDFAEYIDSPVIL
EAQGDKWVTPCSLINRQMGSCYP IQEDHGLGLTLQCHVEGDYIGSQNVSFVFNKGKSMVH
KKAWLISAKQDLFLYQTHSEILSVFPETGSLGGRTNITITGDFFDNSAQVTIAGIPCDIR
HVSPRKIECTTRAPGKDVRLTTPQPGNRGLLFEVGDAVEGLELTEATPGYRWQIVPNASS
PFGFWSQEGQPFRARLSGFFVAPETNNYTFWIQADSQASLHFSWSEEPRTKVKVASISVG
TADWFDSEQNRDEGTWQQKTPKLELLGGAMYYLEAEHHGIAPSRGMRIQVQIHNTWLNPN
DVVTTYLREKHQIRVRAQRLPEVQVLNVSGRGNFFLTWDNVSSQPIPANATAHLIQTIE
ELLAVKCKLEPLWSNILLRLGFERGPEVSNSDGDLTSGTEPFPCGRFSLRQPRHLVLTTPPA
AQKGYRLDQYTHLCLAYKGHMNKILKMIVSFTIGFQNMVKNTTCDWSLTRTSPESWQFDC
TDLWETCVRCFGDLQPPPANSPVLVHQINLLPLAQETGLFYVDEIIIADTNVTVSQADSG
TARPGGNLVEVS SVVGSPPVYSVTSWLAGCGTELPLITARSVPTEGTEEGSGLVLVTTQR
RQRTSPPLGGHFRIQLPNTVISDVPVQISAHHLHLQNNADDFTSRYLNASDFTVKEDL
YTCYEHVWTL SWSTQIGDLPNFIRVSDENLTGVNPAAAATRVVYDGGVFLGP IFGDMLATA
NQHTQVVVRVNDVPAHCPGSCSFQYLQGGSTPCVHSHVWYSIDGDINLMIYITGTGFSGDSQ
FLQVTVNKTSCKVIFSNQTNVVCQTDLLPVGMMHRI LMLVLRPSGLAISATGEDLFLNVKPR
LDMVEPSRAADIGGLWATIRGSSLEGVSLILFGSSYSCAINVATSNSSRIQCKVPPRGKDG
RIVNVTVIRGDYSAVLPRAFTYVSSSLNPVIVTLSRNISNIAGGETLVIGVARLMNYTDLD
VEVHVQDALAPVHTQSAWGLEVALPPLPAGLHRISVSINGVSIHSQGVDLHIQYLTEVFS
IEPCCGSL LGGTILSISGIGFSRDPALVWVLVGNRSCDIVNLTEASIWCETLPAPQIPDA
GAPTVPAAVEVWAGNRFFARGPSPSLVGKGF TFMYEAAAATPVVTAMQGEITNSSLSLHVG
GSNLSNSVILLGNLNC DVETQS FQGNVSLSGCS IPLHSLEAGIYPLQVRQKQMGFANMSV
VLQQFAVMPRIMAIFPSQGSACGGTILTVRGLLLNSRRRSVRVDLSGPFTCVILSLGDHT
ILCQVSLGDP LPGA SFS LNVTVLVNGLTSECQGNCTLFIREEASPVMDALSTNTSGSLT
TVLIRGRLATTADEPMVFDVDDQLPCNV TFFNASHVVCQTRDLAPGPHYLSVFYTRNGYA
CSGNVSRHFYIMPQVFHYFPKNFSLHGGSLLTIEGTGLRGQNTTSVYIDQQTCLTVNIGA
ELIRCIVPTGNGSVALEIEVDGLWYHIGVIGYNKAF TPELISISQSDDILTFAVAQISGA
ANIDIFIGMSPCVGVSGNHTVLQCVVPSL PAGEYHVRGYDCIRGWASSALVFTSRV IITA
VTENFGCLGGRLVHVFGAGFSPGNVSAAVCGAPCRVLANATVSAFSCLVLP LVDVSLAFLC
GLKREEDSCEAARHTYVQCDLTVMATEQLLESWPYLI CESSQCLFVPDHWAE SMFPS
FSGLFISPKLIERDEVL IYNSCNITMETEAE MECETPNQPI TVKITEIRKRWGQNTQGNF
SLQFCRRWSRTHSWFPERLPQDGDNVTVENGQLLLLD TNTSILNLLHIKGGKLI FMAPGP
IELRAHAILVSDGGELRIGSEDKPFDQAGRAQITLYGSSYSTPFFPYGVKFLAVRNGTLSLH
GSLPEVIVTCLRATAHALDTVLAL EDAVDWNP GDEVVIISGTGVKGAKPMEEIVTVETVQ
DTDLYLKSP LRYSHNFTENWVAGEHHILKATVALLSR SITIQGNLTNEREKLLVSCQEAN
APEGNLQHCLYSMSEKMLGSRDMGARVIVQSFP EEPQSQVLKGVQFQVLGQAFHKHLSSL
TLVGAMRESFIQGCTVRNSFSRGLSMCGT LGLKVDSNVFYNI LGHALLVGTCTEMRYISW
EAIHGRKDDWSGHGNIIRNNV IIQVSGAEGLSNPEMLTPSGIYICSP TNVIEGNRVCAGAG
```



# SEQUENCE (FASTA FORMAT) cont...


```
YGYFFHLMTNQTSQAPLLSFTQNIASHSCTRYGLFVYPKFQPPWDNVTGTTLFFQSFTVWES
AGGAQIFRSSNLRLKNFKVYSCRDFGIDVLESDANTSVTDSL L LGHFAHKGSLCMSSGIK
TPKRWELMVSNTTTFVNFDLINCVAIRTCSDCSQGGQGGFTVKTSQ LKFTNSSNLVAFPPPH
AAILEDLDGSLSGKNRSHILASMETLSASCLVNSSFGRVVHGSACGGGVLFHRMSIGLAN
TPEVSYDLTMTDSRNKTTTTVN YVRDTLSNPRGWMALLLDQETYS LQSEN LWINRSLQYSA
TFDNFAPGN YLLL VHTDLPPYPDILLRCGSRVGLSFPFLPSPGQNQGC DWFFNSQLRQLT
YLVSGEGQVQVILRVKEGMPPTISASTSAPESALKWSLPETWQGV EEWGGYNNTIPGPG
DDVLILPNRTVLVDTDLPFFFKGLYVMGTLD FFPVDRSNVLSVACMVIAGGELKVG TLENPL
EKEQKLLILLRASEGVFCDRMNGIHIIDPGTIGVYGKVHLYSAYPKNSWTHLGADIASGNE
RIIVEDAVDWRPHDKIVLSSSSSYEPHEAEVLT VKEVKGHHVRIYERLKH RHIGSVHVTED
GRHIRLAAEVGLLTRNIQIQPDVSCRGR L FVG SFRKSSREEFSGVLQ L LNVEIQNF GSPL
YSSVEFSNV SAGSWIISSTLHQSCGGGIHAAASHGVL LNDNIVFGTAGHGIDLEGQAYTV
TNNLVVLMTQPAWSTI WVAGIKVNQVKDINLHGNVVAGSERLGFHIRGHKCSSCELLWSD
NVAHSSLHGLHLYKESGLDNCTRISGFLAFKNFDY GAMLHVENSVEIENITLVDNTIGLL
AVVYVFSAPQNSVKKVQIVLRNSVIVATSSSFDCIQDKVKPHSANLTSTDRA PSNPRGGR
IGILWPVFTSEP NQWPQEPWHKVRNDHSISGIMKLQDVT FSSFVKSCYSDDLDVCILPNA
ENSGIMHPITAERTRMLKIKDKNKFYFPSLQPRKDLGKVVCPELDCASPRKYLFKDLDGR
ALGLPPPVS VFPKTEAEWTASFFNAGTFREEQKCTYQFLMQGFI CKQTDQVVLILDSADA
IWAIQKLYPVVSVTSGFVDVFSSVNANIPCSTSGSVSTFY S ILPIRQITKVCFMDQTPQV
L RFFLLGNKSTSKLLLAVFYHELQSPHVFLGESFIPPTLVQSASLL LNESIGANYFNIMD
NLLYVVLQGE EPIEIRSGVSIHLALTVMVSVLEK GWEIVILERLTN FLQIGQNQIRFIHE
MPGHEETLKAIADSR AKRKRNCPTVTCTSHYRRVGQRRPLMMEMNSHRASPPMTVETISK
VIVIEIGDSPTVRSTGMIS SLSNKLQNL AHRVITAQQTGVL ENVLNMTIGALLVTQSKG
VIGYGNTSSFKTGNLIYIRPYALSILVQPSDGEVGNELPVQPQLVFLDEQNR RRVESLGPP
SEPWTISASLEGASDSVLKGCTQAETQDGYVSFYNLAVLISGSNWHFIFTVTSPPGVNFT
ARSKPFAVL PVTRKEKSTIILAASLSSVASWLALSCLVCCWLKRSKSRKTKPEEIPESQT
NNQNIHIIHISSKRRESQGPKKEDTVVGEDMRMKVMLGKV NQC PHQLMNGVSRRKVSRHIV
REEEA AVPAPGTTGITSHGHICAPGAPAQQVYLQETGNWKEGQEQLLRYQLAGQNQLLLL
CPDFRQERQQLP GQSR LSKQSGSLGLS QEKKASCGATEAFCLHSVHPETIQEQL
```




A large, horizontal, pink brushstroke graphic that serves as a background for the word 'BLAST'. The brushstroke has a textured, hand-painted appearance with visible bristles and varying shades of pink.

# BLAST

# BLAST SCREENSHOTS

 **U.S. National Library of Medicine**  
National Center for Biotechnology Information

Log in

 **COVID-19 Information**  
[Public health information \(CDC\)](#) | [Research information \(NIH\)](#) | [SARS-CoV-2 data \(NCBI\)](#) | [Prevention and treatment information \(HHS\)](#) | [Español](#)

**Notice of Scheduled eRA Maintenance:** Please note that eRA systems will be undergoing scheduled maintenance from 10am until 10pm Eastern US time on April 24, 2021. During this window, eRA-dependent services such as MyBibliography and Grant Reporting may be unavailable. More information is available on the [eRA website](#).

**BLAST® » blastp suite » results for RID-8C7SMPV8016**Home Recent Results Saved Strategies Help

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

Job Title	sp P08F94 PKHD1_HUMAN Fibrocystin OS=Homo...
RID	<a href="#">8C7SMPV8016</a> Search expires on 04-27 10:01 am <a href="#">Download All ▾</a>
Program	BLASTP <a href="#">?</a> <a href="#">Citation ▾</a>
Database	nr <a href="#">See details ▾</a>
Query ID	lcl Query_69080
Description	sp P08F94 PKHD1_HUMAN Fibrocystin OS=Homo sapien ...
Molecule type	amino acid
Query Length	4074
Other reports	<a href="#">Distance tree of results</a> <a href="#">Multiple alignment</a> <a href="#">MSA viewer</a> <a href="#">?</a>

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

**Filter Results**

Organism only top 20 will appear ☐ exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity

E value


Query Coverage

to

to

to

[Filter](#) [Reset](#)

 **Feedback**

# BLAST SCREENSHOTS

Descriptions									
Sequences producing significant alignments									
Download <span>▼</span> <span>New</span> Select columns <span>▼</span> Show 50 <span>▼</span> <span>?</span>									
<input checked="" type="checkbox"/> select all 50 sequences selected									
<a href="#">GenPept</a> <a href="#">Graphics</a> <a href="#">Distance tree of results</a> <a href="#">Multiple alignment</a> <span>New</span> <a href="#">MSA Viewer</a>									
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	<a href="#">fibrocystin isoform 1 precursor [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	8415	8415	100%	0.0	100.00%	4074	<a href="#">NP_619639.3</a>
<input checked="" type="checkbox"/>	<a href="#">polycystic kidney and hepatic disease 1 [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	8410	8410	100%	0.0	99.98%	4074	<a href="#">AAM44232.1</a>
<input checked="" type="checkbox"/>	<a href="#">fibrocystin isoform X2 [Homo sapiens]</a>	<a href="#">Homo sapiens</a>	8350	8350	100%	0.0	99.39%	4049	<a href="#">XP_016866434.1</a>
<input checked="" type="checkbox"/>	<a href="#">fibrocystin isoform X1 [Pan troglodytes]</a>	<a href="#">Pan troglodytes</a>	8328	8328	100%	0.0	98.94%	4074	<a href="#">XP_518534.4</a>
<input checked="" type="checkbox"/>	<a href="#">fibrocystin [Pan paniscus]</a>	<a href="#">Pan paniscus</a>	8323	8323	100%	0.0	98.87%	4074	<a href="#">XP_008956905.2</a>
<input checked="" type="checkbox"/>	<a href="#">fibrocystin [Gorilla gorilla gorilla]</a>	<a href="#">Gorilla gorilla gorilla</a>	8273	8273	100%	0.0	98.31%	4074	<a href="#">XP_030868334.1</a>
<input checked="" type="checkbox"/>	<a href="#">fibrocystin isoform X2 [Pan troglodytes]</a>	<a href="#">Pan troglodytes</a>	8271	8271	100%	0.0	98.40%	4052	<a href="#">XP_016811152.2</a>
<input checked="" type="checkbox"/>	<a href="#">fibrocystin isoform X3 [Pan troglodytes]</a>	<a href="#">Pan troglodytes</a>	8259	8259	100%	0.0	98.33%	4049	<a href="#">XP_016811153.2</a>
<input checked="" type="checkbox"/>	<a href="#">fibrocystin isoform X1 [Pongo abelii]</a>	<a href="#">Pongo abelii</a>	8179	8179	100%	0.0	97.32%	4074	<a href="#">XP_024104185.1</a>
<input checked="" type="checkbox"/>	<a href="#">fibrocystin [Hylobates moloch]</a>	<a href="#">Hylobates moloch</a>	8140	8140	100%	0.0	96.54%	4073	<a href="#">XP_031997014.1</a>
<input checked="" type="checkbox"/>	<a href="#">fibrocystin [Nomascus leucogenys]</a>	<a href="#">Nomascus leucogenys</a>	8121	8121	100%	0.0	96.42%	4073	<a href="#">XP_003254212.2</a>
<input checked="" type="checkbox"/>	<a href="#">PREDICTED: LOW QUALITY PROTEIN: fibrocystin [Rhinopithecus bieti]</a>	<a href="#">Rhinopithecus bieti</a>	8041	8041	100%	0.0	95.12%	4100	<a href="#">XP_017705097.1</a>
<input checked="" type="checkbox"/>	<a href="#">LOW QUALITY PROTEIN: fibrocystin [Theropithecus gelada]</a>	<a href="#">Theropithecus gelada</a>	8038	8038	100%	0.0	95.21%	4073	<a href="#">XP_025238760.1</a>
<input checked="" type="checkbox"/>	<a href="#">PREDICTED: LOW QUALITY PROTEIN: fibrocystin [Macaca fascicularis]</a>	<a href="#">Macaca fascicularis</a>	8034	8034	100%	0.0	95.16%	4073	<a href="#">XP_005552817.1</a>
<input checked="" type="checkbox"/>	<a href="#">LOW QUALITY PROTEIN: fibrocystin [Trachypithecus francoisi]</a>	<a href="#">Trachypithecus francoisi</a>	8033	8033	100%	0.0	95.09%	4074	<a href="#">XP_033073315.1</a>
<input checked="" type="checkbox"/>	<a href="#">LOW QUALITY PROTEIN: fibrocystin [Macaca nemestrina]</a>	<a href="#">Macaca nemestrina</a>	8033	8033	100%	0.0	95.16%	4073	<a href="#">XP_011755149.1</a>
<input checked="" type="checkbox"/>	<a href="#">LOW QUALITY PROTEIN: fibrocystin [Papio anubis]</a>	<a href="#">Papio anubis</a>	8031	8031	100%	0.0	95.19%	4073	<a href="#">XP_021793057.2</a>
<input checked="" type="checkbox"/>	<a href="#">PREDICTED: LOW QUALITY PROTEIN: fibrocystin [Colobus angolensis palliatus]</a>	<a href="#">Colobus angolensis palliatus</a>	8026	8026	100%	0.0	94.94%	4074	<a href="#">XP_011815634.1</a>

 Feedback

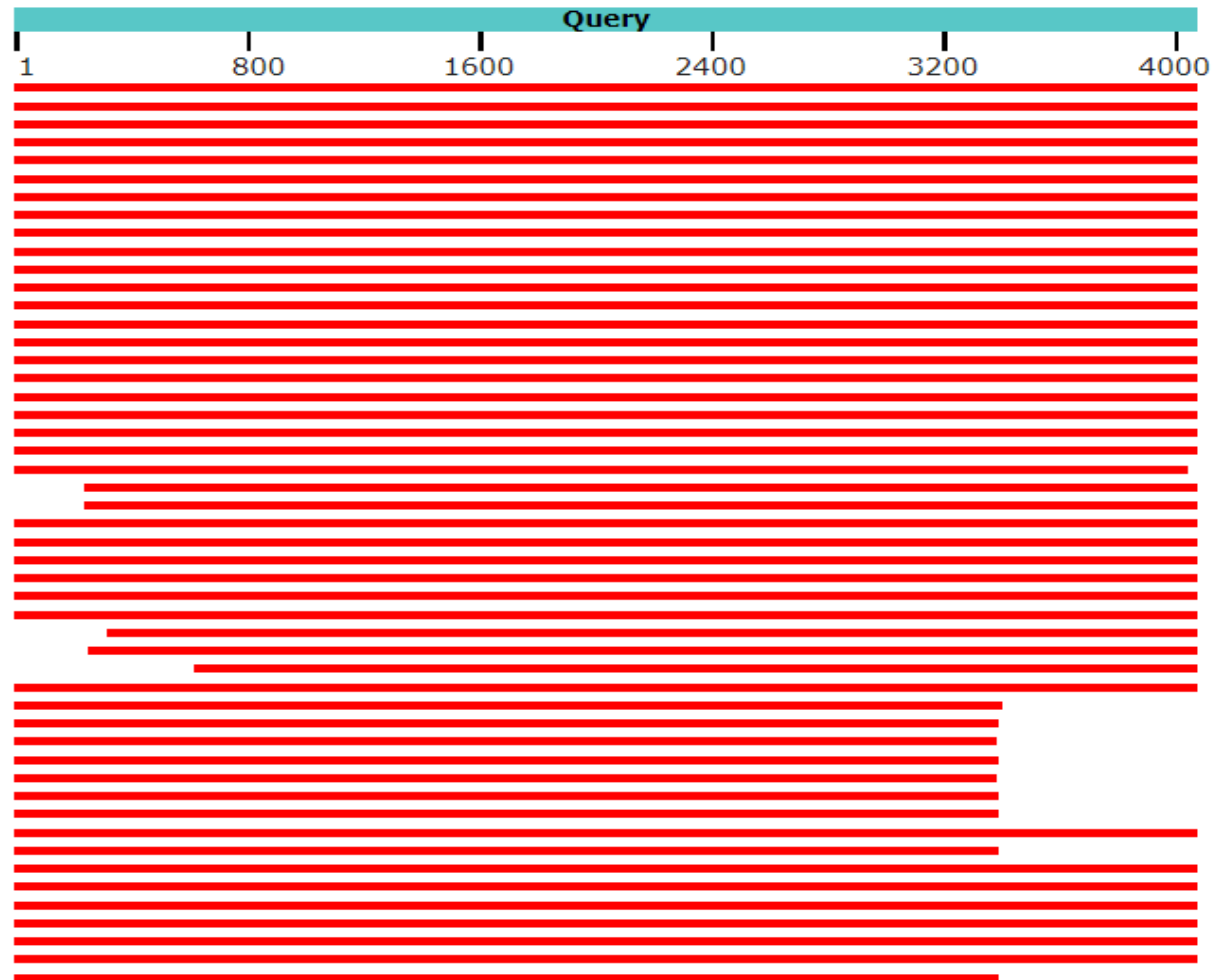
# BLAST

Among all 50 sequences, the top 10 sequences were selected, which includes the target protein (Fibrocystin isoform 1 precursor [Homo Sapiens]) and, the next 9 protein sequences other than Homo Sapiens..

Description	Scientific Name	Common Name	Max Score	Total Score	Query Cov	E value	Per. ident	Acc. Len	Accession
fibrocystin isoform 1 precursor [Homo sapiens]	Homo sapiens	human	8415	8415	100%	0	100		
fibrocystin isoform X1 [Pan troglodytes]	Pan troglodytes	chimpanzee	8328	8328	100%				
fibrocystin [Pan paniscus]	Pan paniscus	pygmy chimpanzee	8323	832					
fibrocystin [Gorilla gorilla gorilla]	Gorilla gorilla gorilla	western lowland gorilla							
fibrocystin isoform X2 [Pan troglodytes]	Pan troglodytes	chimp							
fibrocystin isoform X3 [Pan troglodytes]	Pan troglod								
fibrocystin isoform X1 [Pongo abelii]									
fibrocystin [Hylobates moloch]									
fibrocystin [Nomascus leu									
PREDICT									

# BLAST SCREENSHOTS

Distribution of the top 50 Blast Hits on 50 subject sequences





# BLAST SCREENSHOTS

<div> <div>Descriptions</div> <div>Graphic Summary</div> <div>Alignments</div> <div><b>Taxonomy</b></div> </div>				
<div> <div>Reports</div> <div><b>Lineage</b></div> <div>Organism</div> <div>Taxonomy</div> </div>				
50 sequences selected ?				
Organism	Blast Name	Score	Number of Hits	Description
<a href="#">Boreoeutheria</a>	<a href="#">placentals</a>		61	
• <a href="#">Primates</a>	<a href="#">primates</a>		55	
• <a href="#">Simiiformes</a>	<a href="#">primates</a>		53	
• <a href="#">Catarrhini</a>	<a href="#">primates</a>		45	
• <a href="#">Hominoidea</a>	<a href="#">primates</a>		33	
• <a href="#">Hominidae</a>	<a href="#">primates</a>		31	
• <a href="#">Homininae</a>	<a href="#">primates</a>		27	
• <a href="#">Homo sapiens</a>	<a href="#">primates</a>	8415	19	<a href="#">Homo sapiens hits</a>
• <a href="#">Pan troglodytes</a>	<a href="#">primates</a>	8328	6	<a href="#">Pan troglodytes hits</a>
• <a href="#">Pan paniscus</a>	<a href="#">primates</a>	8323	1	<a href="#">Pan paniscus hits</a>
• <a href="#">Gorilla gorilla gorilla</a>	<a href="#">primates</a>	8273	1	<a href="#">Gorilla gorilla gorilla hits</a>
• <a href="#">Pongo abelii</a>	<a href="#">primates</a>	8179	4	<a href="#">Pongo abelii hits</a>
• <a href="#">Hylobates moloch</a>	<a href="#">primates</a>	8140	1	<a href="#">Hylobates moloch hits</a>
• <a href="#">Nomascus leucogenys</a>	<a href="#">primates</a>	8121	1	<a href="#">Nomascus leucogenys hits</a>
• <a href="#">Rhinopithecus bieti</a>	<a href="#">primates</a>	8041	1	<a href="#">Rhinopithecus bieti hits</a>
• <a href="#">Theropithecus gelada</a>	<a href="#">primates</a>	8038	1	<a href="#">Theropithecus gelada hits</a>
• <a href="#">Macaca fascicularis</a>	<a href="#">primates</a>	8034	1	<a href="#">Macaca fascicularis hits</a>
• <a href="#">Trachypithecus francoisi</a>	<a href="#">primates</a>	8033	1	<a href="#">Trachypithecus francoisi hits</a>
• <a href="#">Macaca nemestrina</a>	<a href="#">primates</a>	8033	1	<a href="#">Macaca nemestrina hits</a>
• <a href="#">Papio anubis</a>	<a href="#">primates</a>	8031	1	<a href="#">Papio anubis hits</a>
• <a href="#">Colobus angolensis palliatus</a>	<a href="#">primates</a>	8026	1	<a href="#">Colobus angolensis palliatus hits</a>

For BLAST, we pasted the sequence of our target protein, i.e, P08F94 Fibrocytin, which was retrieved in FASTA format from uniprot to the blastp program. The top 50 results were analysed, from which the top 10 sequences (including the target protein) were chosen for multiple sequence alignment. Graphic summary of distribution of the top 50 blast hits on 50 subject sequences was also performed. And the lineage descendency in the taxonomy section was visualized to have a glance at the lineage from ancestors to descendants of species listed in the top 50 BLAST results.

## SUMMARY OF BLAST

# RESULTS OF MULTIPLE SEQUENCE ALIGNMENT USING CLUSTAL OMEGA

CLUSTAL O(1.2.4) multiple sequence alignment

```

XP_017705097.1    MTAWLISLMSIEVLLAVPHLSLHIEPEEGSLAGGTWITVIFDGLLEGLVYPNNGSQLEI    60
XP_003254212.2    MTAWLISLMSIEVLLAVPHLSLHIEPEEGSLGGGTWITVIFDGLLEGLVYPNNGSQLEI    60
XP_031997014.1    MTAWLISLMSIEVLLAVPYLSLHIEPEEGSLGGGTWITVIFDGLLEGLVYPNNGSQLEI    60
XP_024104185.1    MTAWLISLMSIEVLLAVPHLSLHIEPEEGSLAGGTWITVIFDGLLEGLVYPNNGSQLEI    60
XP_030868334.1    MTAWLISLMSIEVLLAVPHLSLHIEPEEGSLAGGTWITVIFDGLLEGLVYPNNGSQLEI    60
NP_619639.3       MTAWLISLMSIEVLLAVRHLSLHIEPEEGSLAGGTWITVIFDGLLEGLVYPNNGSQLEI    60
XP_008956905.2    MTAWLISLMSIEVLLAVPHLSLHIEPEEGSLAGGTWITVIFDGLLEGLVYPNNGSQLEI    60
XP_016811153.2    MTAWLISLMSIEVLLAVPHLSLHIEPEEGSLAGGTWITVIFDGLLEGLVYPNNGSQLEI    60
XP_016811152.2    MTAWLISLMSIEVLLAVPHLSLHIEPEEGSLAGGTWITVIFDGLLEGLVYPNNGSQLEI    60
XP_518534.4       MTAWLISLMSIEVLLAVPHLSLHIEPEEGSLAGGTWITVIFDGLLEGLVYPNNGSQLEI    60
*****:*****:***** **;*****

XP_017705097.1    HLVNMNMVAPALPSIPCDVFPVFLDLPVVMCQTRSLLEAHEGLYSLEAYFGGQLVSSPS    120
XP_003254212.2    HLVNMNMVAPALPSIPCDVFPVFLDLPVVTCTRSVLSEAHEGLYSLEAYFGGQLVSSPN    120
XP_031997014.1    HLVNMNMVAPALPSIPCDVFPVFLDLPVTCRTRSVLSEAHEGLYSLEAYFGGQLVSSPN    120
XP_024104185.1    HLVNMNMVAPALPSIPCDVFPVFLDLPVTCRTRSVLSEAHEGLYFLEAYFGGQLVSSPN    120
XP_030868334.1    HLVNMNMVAPALPSIPCDVFPVFLDLPVTCRTRSVLSEAHEGLYFLEAYFGGQLVSSPN    120
NP_619639.3       HLVNMNMVAPALPSVPCDVFPVFLDLPVVTCTRSVLSEAHEGLYFLEAYFGGQLVSSPN    120
XP_008956905.2    HLVNMNMVAPALPSVPCDVFPVFLDLPVTCRTRSVLSEAHEGLYFLEAYFGGQLVSSPN    120
XP_016811153.2    HLVNMNMVAPALPSVPCDVFPVFLDLPVTCRTRSVLSEAHEGLYFLEAYFGGQLVSSPN    120
XP_016811152.2    HLVNMNMVAPALPSVPCDVFPVFLDLPVTCRTRSVLSEAHEGLYFLEAYFGGQLVSSPN    120
XP_518534.4       HLVNMNMVAPALPSVPCDVFPVFLDLPVTCRTRSVLSEAHEGLYFLEAYFGGQLVSSPN    120
****:***:*** *:*****:* *:***:***** ***** *:*****

XP_017705097.1    PGPRDSCFTKFSKAQTPIVHKVYPPSGVPGKLIHVYGWIITGRLETDFDFAEYIDSPVIL    180
XP_003254212.2    PGPRDRCTFKFSKAQTPIVHQVYPPSGVPGKLIHIYGWIITGRLESFDFDAEYIDSPVIL    180
XP_031997014.1    PGPRDRCTFKFSKAQTPIVHQVYPPSGVPGKLIHIYGWIITGRLESFDFDAEYIDSPVIL    180
XP_024104185.1    PGPRDSCFTKFSKAQTPIVHHVYPPSGVPGKLIHVYGWIITGRLETDFDFAEYIDSPVIL    180
XP_030868334.1    PGPRDSCFTKFSKAQTPIVHQVYPPSGVPGKLIHVYGWIITGRLETDFDFAEYIDSPVIL    180
NP_619639.3       PGPRDSCFTKFSKAQTPIVHQVYPPSGVPGKLIHVYGWIITGRLETDFDFAEYIDSPVIL    180
XP_008956905.2    PGPRDSCFTKFSKAQTPIVHQVYPPSGVPGKLIHVYGWIITGRLETDFDFAEYIDSPVIL    180
XP_016811153.2    PGPRDSCFTKFSKAQTPIVHQVYPPSGVPGKLIHVYGWIITGRLETDFDFAEYIDS----    176
XP_016811152.2    PGPRDSCFTKFSKAQTPIVHQVYPPSGVPGKLIHVYGWIITGRLETDFDFAEYIDSPVIL    180
XP_518534.4       PGPRDSCFTKFSKAQTPIVHQVYPPSGVPGKLIHVYGWIITGRLETDFDFAEYIDSPVIL    180
*****:*****:*****:*****:*****:*****

```

# RESULTS OF MULTIPLE SEQUENCE ALIGNMENT USING CLUSTAL OMEGA (CONT ..)

XP_017705097.1	EAQGDKWVTPCSLVNRQTGSCYPIQEDHGLGTLQCHVEGDYIGSQNVSFVFNKGKSMIH	240
XP_003254212.2	EAQGDKWVTPCSLVNRQTGSCYPIQEDHGLGTLQCHVEGDYIGSQNVSFVFNKGKSMVH	240
XP_031997014.1	EAQGDKWVTPCSLVNRQTGSCYPIQADHGLGTLQCHVEGDYIGSQNVSFVFNKGKSMVH	240
XP_024104185.1	EAQGDKWVTPCSLVNRQTGSCYPIQEDHGLGTLQCRVEGDYIGSQNVSFVFNKGKSMVH	240
XP_030868334.1	EAQGDKWVTPCSLVNRQMGSRYPYPIQEDHGLGTLQCRVEGDYIGSQNVSFVFNKGKSMVH	240
NP_619639.3	EAQGDKWVTPCSLVNRQMGSCYPIQEDHGLGTLQCHVEGDYIGSQNVSFVFNKGKSMVH	240
XP_008956905.2	EAQGDKWVTPCSLVNRQMGSCYPIQEDHGLGTLQCRVEGDYIGSQNVSFVFNKGKSMVH	240
XP_016811153.2	-----YPIQEDHGLGTLQCRVEGDYIGSQNVSFVFNKGKSMVH	215
XP_016811152.2	EAQGDKWVTPCSLVNRQMGSCYPIQEDHGLGTLQCRVEGDYIGSQNVSFVFNKGKSMVH	240
XP_518534.4	EAQGDKWVTPCSLVNRQMGSCYPIQEDHGLGTLQCRVEGDYIGSQNVSFVFNKGKSMVH	240
	**** * : * : * : * : * : * : * : * : * : *	
XP_017705097.1	KKAWLISAKQDLFLYQTHSEILSVFPETGSLGGRTDITITGDDFNSAQVTIAGIPCDIR	300
XP_003254212.2	KKAWLISAKQDLFLYQTHSEILSVFPETGSLGGRTDITITGDDFHSSAQVTIAGIPCDIR	300
XP_031997014.1	KKAWLISAKQDLFLYQTHSEILSVFPETGSLGGRTDITITGDDFNSAQVTIAGIPCDIR	300
XP_024104185.1	KKAWLISAKQDLFLYQTHSEILSVFPETGSLGGRTDITITGDDFNSAQVTIAGIPCDIR	300
XP_030868334.1	KKAWLISAKQDLFLYQTHSEILSVFPETGSLGGRTDITITGDDFNSAQVTIAGIPCDIR	300
NP_619639.3	KKAWLISAKQDLFLYQTHSEILSVFPETGSLGGRTDITITGDDFNSAQVTIAGIPCDIR	300
XP_008956905.2	KKAWLISAKQDLFLYQTHSEILSVFPETGSLGGRTDITITGDDFNSAQVTIAGIPCDIR	300
XP_016811153.2	KKAWLISAKQDLFLYQTHSEILSVFPETGSLGGRTDITITGDDFNSAQVTIAGIPCDIR	275
XP_016811152.2	KKAWLISAKQDLFLYQTHSEILSVFPETGSLGGRTDITITGDDFNSAQVTIAGIPCDIR	300
XP_518534.4	KKAWLISAKQDLFLYQTHSEILSVFPETGSLGGRTDITITGDDFNSAQVTIAGIPCDIR	300
	***** : ***** : ***** : ***** : *****	
XP_017705097.1	HVSPKKIECTTRAPGKDARLTPQPGNRGLLFVGDAVEGLELTEATPGYRWQIVPNASS	360
XP_003254212.2	HVSPRKIECTTRAPGKDARLTPQPGNRGLLFVGDAVEGLELTEATPGYRWQIVPNASS	360
XP_031997014.1	HVSPRKIECTTRAPGKDARLTPQPGNRGLLFVGDAVEGLELTEATPGYRWQIVPNASS	360
XP_024104185.1	HVSPRKIECTTRAPGKDARLTPQPGNRGLLFVGDAVEGLELTEATPGYRWQIVPNASS	360
XP_030868334.1	HVSPRKIECTTRAPGKDARLTPQPGNRGLLFVGDAVEGLELTEATPGYRWQIVPNASS	360
NP_619639.3	HVSPRKIECTTRAPGKDARLTPQPGNRGLLFVGDAVEGLELTEATPGYRWQIVPNASS	360
XP_008956905.2	HVSPRKIECTTRAPGKDARLTPQPGNRGLLFVGDAVEGLELTEATPGYRWQIVPNASS	360
XP_016811153.2	HVSPRKIECTTRAPGKDARLTPQPGNRGLLFVGDAVEGLELTEATPGYRWQIVPNASS	335
XP_016811152.2	HVSPRKIECTTRAPGKDARLTPQPGNRGLLFVGDAVEGLELTEATPGYRWQIVPNASS	360
XP_518534.4	HVSPRKIECTTRAPGKDARLTPQPGNRGLLFVGDAVEGLELTEATPGYRWQIVPNASS	360
	**** : ***** : ***** : ***** : *****	



# RESULTS OF MULTIPLE SEQUENCE ALIGNMENT USING CLUSTAL OMEGA (CONT ..)

```

XP_017705097.1  PFGFWSEKGGQPFARLSGFFVAPETNNTFWIQADSQASLHFSWSEEPRTKVKVASISVG  420
XP_003254212.2  PFGFWSEKGGQPFARLSGFFVAPETNNTFWIQADSQASLHFSWSEEPRTKVKVASISVG  420
XP_031997014.1  PFGFWSEKGGQPFARLSGFFVAPETNNTFWIQADSQASLHFSWSEEPRTKVKVASISVG  420
XP_024104185.1  PFGFWSEKGGQPFARLSGFFVAPETNNTFWIQADSQASLHFSWSEEPRTKVKVASISVG  420
XP_030868334.1  PFGFWSEKGGQPFARLSGFFVAPETNNTFWIQADSQASLHFSWSEEPRTKVKVASISVG  420
NP_619639.3      PFGFWSEKGGQPFARLSGFFVAPETNNTFWIQADSQASLHFSWSEEPRTKVKVASISVG  420
XP_008956905.2  PFGFWSEKGGQPFARLSGFFVAPETNNTFWIQADSQASLHFSWSEEPRTKVKVASISVG  420
XP_016811153.2  PFGFWSEKGGQPFARLSGFFVAPETNNTFWIQADSQASLHFSWSEEPRTKVKVASISVG  395
XP_016811152.2  PFGFWSEKGGQPFARLSGFFVAPETNNTFWIQADSQASLHFSWSEEPRTKVKVASISVG  420
XP_518534.4      PFGFWSEKGGQPFARLSGFFVAPETNNTFWIQADSQASLHFSWSEEPRTKVKVASISVG  420
*****

XP_017705097.1  TADWFDSEQNTDERTWQKTPKLELLGGAKYYLEAEHHGIAPSRGMRIGVQIHNTWLN  480
XP_003254212.2  TADWFDSEQNGDEGTWQKSPKLELLGGAKYYLEAEHHGIAPSRGMRIGVQIHNTWLN  480
XP_031997014.1  TADWFDSEQNRDEGTWQKSPKLELLGGAKYYLEAEHHGIAPSRGMRIGVQIHNTWLN  480
XP_024104185.1  TADWFDSEQNRDEGTWQKTPKLELLGGAKYYLEAEHHGIAPSRGMRIGVQIHNTWLN  480
XP_030868334.1  TADWFDSEQNRDEGTWQKTPKLELLGGAMYYLEAEHHGIAPSRGMRIGVQIHNTWLN  480
NP_619639.3      TADWFDSEQNRDEGTWQKTPKLELLGGAMYYLEAEHHGIAPSRGMRIGVQIHNTWLN  480
XP_008956905.2  TADWFDSEQNRDEGTWQKTPKLELLGGAMYYLEAEHHGIAPSRGMRIGVQIHNTWLN  480
XP_016811153.2  TADWFDSEQNRDEGTWQKTPKLELLGGAMYYLEAEHHGIAPSRGMRIGVQIHNTWLN  455
XP_016811152.2  TADWFDSEQNRDEGTWQKTPKLELLGGAMYYLEAEHHGIAPSRGMRIGVQIHNTWLN  480
XP_518534.4      TADWFDSEQNRDEGTWQKTPKLELLGGAMYYLEAEHHGIAPSRGMRIGVQIHNTWLN  480
*****

XP_017705097.1  DWVTTYLREKHQIRARAQRLPEVQVLNVSGRGNFFLTWDNVSSQIPANATAHLIQT  540
XP_003254212.2  DWVTTYLREKHQIRVRAQRLPEVQVLNVSGRGNFFLTWDNVSSQIPANATAHLIQT  540
XP_031997014.1  DWVTTYLREKHQIRVRAQRLPEVQVLNVSGRGNFFLTWDNVSSQIPANATAHLIQT  540
XP_024104185.1  DWVTTYLREKHQIRVQAQRLPEVQVLNVSGRGNFFLTWDNVSSQIPANATAHLIQT  540
XP_030868334.1  DWVTTYLREKHQIRVRAQRLPEVQVLNVSGRGNFFLTWDNVSSQIPANATAHLIQT  540
NP_619639.3      DWVTTYLREKHQIRVRAQRLPEVQVLNVSGRGNFFLTWDNVSSQIPANATAHLIQT  540
XP_008956905.2  DWVTTYLREKHQIRVRAQRLPEVQVLNVSGRGNFFLTWDNVSSQIPANATAHLIQT  540
XP_016811153.2  DWVTTYLREKHQIRVRAQRLPEVQVLNVSGRGNFFLTWDNVSSQIPANATAHLIQT  515
XP_016811152.2  DWVTTYLREKHQIRVRAQRLPEVQVLNVSGRGNFFLTWDNVSSQIPANATAHLIQT  540
XP_518534.4      DWVTTYLREKHQIRVRAQRLPEVQVLNVSGRGNFFLTWDNVSSQIPANATAHLIQT  540
*****

```



# PERCENTAGE IDENTITY MATRIX

```
# Percent Identity Matrix - created by Clustal2.1
```

```
#
```

```
#
```

1:	XP_017705097.1	100.00	94.55	94.75	94.97	94.94	95.14	95.29	95.38	95.36	95.36
2:	XP_003254212.2	94.55	100.00	98.50	96.44	96.27	96.44	96.56	96.67	96.59	96.61
3:	XP_031997014.1	94.75	98.50	100.00	96.51	96.49	96.56	96.69	96.76	96.72	96.73
4:	XP_024104185.1	94.97	96.44	96.51	100.00	97.03	97.32	97.45	97.46	97.38	97.42
5:	XP_030868334.1	94.94	96.27	96.49	97.03	100.00	98.31	98.38	98.52	98.47	98.50
6:	NP_619639.3	95.14	96.44	96.56	97.32	98.31	100.00	98.87	98.94	98.91	98.94
7:	XP_008956905.2	95.29	96.56	96.69	97.45	98.38	98.87	100.00	99.63	99.61	99.63
8:	XP_016811153.2	95.38	96.67	96.76	97.46	98.52	98.94	99.63	100.00	99.98	100.00
9:	XP_016811152.2	95.36	96.59	96.72	97.38	98.47	98.91	99.61	99.98	100.00	99.98
10:	XP_518534.4	95.36	96.61	96.73	97.42	98.50	98.94	99.63	100.00	99.98	100.00

A large, irregular pink brushstroke shape serves as a background for the title text.

## INTERPRETATION OF MULTIPLE SEQUENCE ALIGNMENT RESULTS

Color indicates physio-chemical properties of the amino-acid, and the symbols below each position in the sequence indicate the amount of conservation ('\*': Exact i.e. all sequences have a common aa in that, ':': Conserved Substitution i.e. all except 1 sequences have a common aa residue on that location '!' : Semi-conserved substitution).

All parts of sequences are conserved exactly or approximately (since \* or : is observed for all aa in regions), showing the close relationship among the sequences and hence the organisms.

# Phylogenetic Tree

In a phylogenetic tree, the relatedness of two species has a very specific meaning.

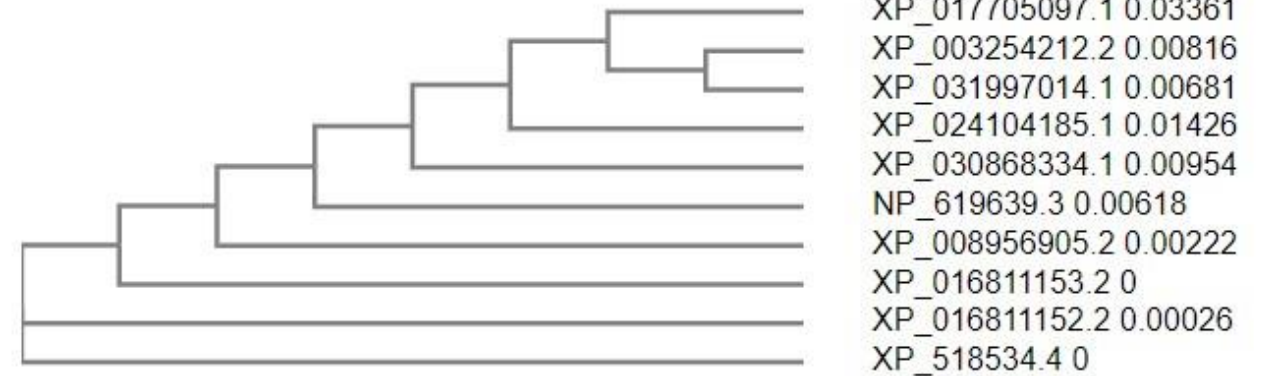
Two species are more related if they have a more recent common ancestor, and less related if they have a less recent common ancestor.

At each branch point lies the most recent common ancestor of all the groups descended from that branch point.

## Phylogenetic Tree

*This is a Neighbour-joining tree without distance corrections.*

Branch length: ☒ Cladogram ☐ Real

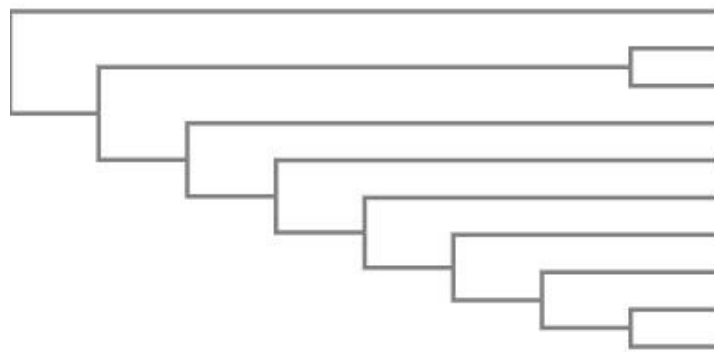


# Tree Data

```
(
(
(
(
(
(
XP_017705097.1:0.03361,
(
XP_003254212.2:0.00816,
XP_031997014.1:0.00681)
:0.01243)
:0.00175,
XP_024104185.1:0.01426)
:0.00562,
XP_030868334.1:0.00954)
:0.00115,
NP_619639.3:0.00618)
:0.00292,
XP_008956905.2:0.00222)
:0.00147,
XP_016811153.2:0.00000)
:0.00000,
XP_016811152.2:0.00026,
XP_518534.4:0.00000);
```

## Phylogram

Branch length: ☒ Cladogram ☐ Real



```
XP_017705097.1 0.0263695
XP_003254212.2 0.00748834
XP_031997014.1 0.00748834
XP_024104185.1 0.0140023
XP_030868334.1 0.00819316
NP_619639.3 0.00561208
XP_008956905.2 0.0021269
XP_016811153.2 0.00191405
XP_016811152.2 0.000370188
XP_518534.4 0.000370188
```

## GuideTree

```
(
XP_017705097.1:0.0263695
,
(
(
XP_003254212.2:0.00748834
,
XP_031997014.1:0.00748834
):0.010512
,
(
XP_024104185.1:0.0140023
,
(
XP_030868334.1:0.00819316
,
(
NP_619639.3:0.00561208
,
(
XP_008956905.2:0.0021269
,
(
XP_016811153.2:0.00191405
,
(
XP_016811152.2:0.000370188
,
XP_518534.4:0.000370188
):0.00154387
):0.00021285
):0.00348518
):0.00258108
):0.00580911
):0.00399812
):0.00836915
)
);
```

For the multiple sequence alignment, the top 50 blast results were analyzed, from which the top 10 sequences were selected, which includes the target protein (Fibrocystin isoform 1 precursor [Homo Sapiens]) and, the next 9 protein sequences other than homo sapiens. Results obtained were analysed based on the relationship between the sequences by looking at the symbols, '\*', ':' and '.'. Percentage identity matrix was generated. And the phylogenetic tree, phylogram, tree data and guide tree were built to understand the relationship of homology among the sequences.


## SUMMARY OF MSA



A large, irregular pink brushstroke shape serves as a background for the title text. The stroke is thick and has a textured, hand-painted appearance with visible bristles and varying shades of pink.

# **HOMOLOGY MODELLING**

# Steps:

**BIOZENTRUM**  
University of Basel  
The Center for Molecular Life Sciences

SWISS-MODEL

Modelling Repository Tools Documentation Log in Create Account

## SWISS-MODEL


is a fully automated protein structure homology-modelling server, accessible via the **Expasy web server**, or from the program DeepView (Swiss Pdb-Viewer).

The purpose of this server is to make protein modelling accessible to all life science researchers worldwide.

[Start Modelling](#)

## Repository

Every week we model all the sequences for thirteen core species based on the latest UniProtKB proteome. Is your protein already modelled and up to date in **SWISS-MODEL Repository**?





SWISS-MODEL

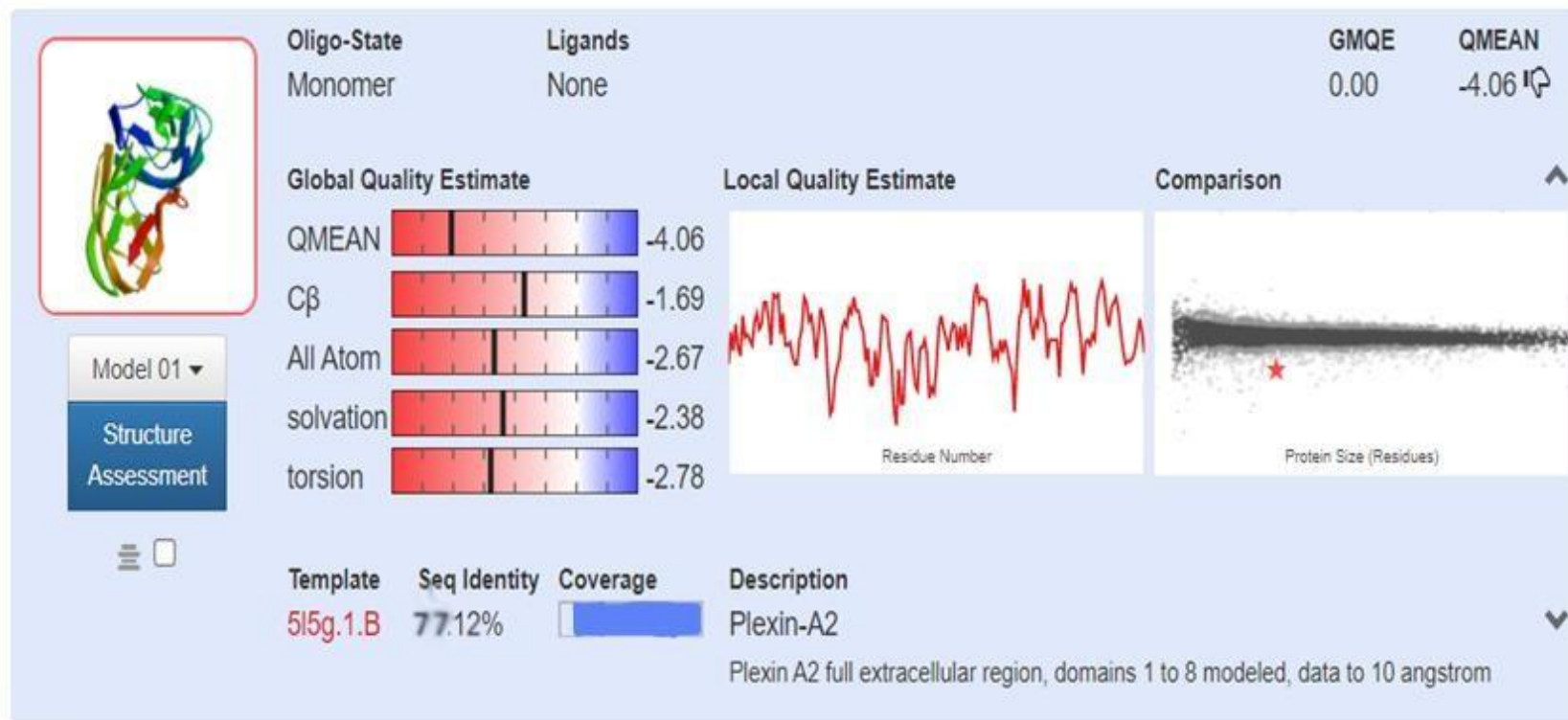
Modelling Repository Tools Documentation Log in Create Account

### Start a New Modelling Project

Target  
Sequence(s):  
(Format must be  
FASTA, Clustal,  
plain string, or a valid  
UniProtKB AC)

Target                       

# STRUCTURE PREDICTION



# STRUCTURE OF THE PROTEIN



# VISUALIZATION OF PROTEIN

## MolProbity Results

MolProbity Score	2.51	
<input type="checkbox"/> Clash Score	14.93	(B1083 VAL-B1100 PHE), (B1130 VAL-B1131 ALA), (B1055 TYR-B1076 ARG), (B1052 PHE-B1100 PHE)
Ramachandran Favoured	88.00%	
<input type="checkbox"/> Ramachandran Outliers	4.00%	B1074 PRO, B1130 VAL, B1082 ILE, B1135 ASN, B1088 ILE, B1081 ARG, B1026 PRO
<input type="checkbox"/> Rotamer Outliers	2.03%	B1073 VAL, B1130 VAL, B1074 PRO
<input type="checkbox"/> C-Beta Deviations	5	B1066 SER, B1134 MET, B1133 LEU, B1112 THR, B1176 VAL
<input type="checkbox"/> Bad Bonds	1 / 1354	B1109 VAL
<input type="checkbox"/> Bad Angles	37 / 1850	B1184 HIS, (B1109 VAL-B1110 ILE), (B1077 GLY-B1078 LYS), B1176 VAL, B1136 TYR, B1179 ASN, B1181 VAL, B1084 ASN, B1074 PRO, (B1030 ALA-B1031 ASP), B1056 SER, (B1130 VAL-B1131 ALA), B1035 LEU, (B1034 GLY-B1035 LEU), (B1031 ASP-B1032 ILE), (B1074 PRO-B1075 PRO), B1134 MFT B1031 ASP B1112 THR B1087 VAL B1144 HIS (B1078
<input type="checkbox"/> Twisted Non-Proline	1 / 166	(B1129 GLY-B1130 VAL)
<input type="checkbox"/> Twisted Prolines	1 / 10	(B1025 GLU-B1026 PRO)

Results obtained using MolProbity version 4.4

## Quality Estimate





# TEMPLATE PROTEIN

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignments

Download ▼ New Select columns ▼ Show 100 ▼ ?

☐ select all 5 sequences selected
 

[GenPept](#)
[Graphics](#)
[Distance tree of results](#)
[Multiple alignment](#)
New [MSA Viewer](#)

	Description <span>▼</span>	Scientific Name <span>▼</span>	Max Score	Total Score	Query Cover	E value <span>▼</span>	Per. Ident	Acc. Len <span>▼</span>	Accession
<input checked="" type="checkbox"/>	<a href="#">Plexin A2 full extracellular region, domains 1 to 8 modeled, data to 10 angstrom [Mus musculus]</a>	<a href="#">Mus musculus</a>	8023	8023	100%	0.0	76.21%	4073	<a href="#">5L5G_A</a>

RCSB PDB Deposit Search Visualize Annotate

**PDB** PROTEIN DATA BANK

177009 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

Enter search term(s)

Advanced Search | Browse Annotations Help

PDB-101 EMBL Data Resource WorldWide Protein Data Bank Foundation


Celebrating 40 YEARS OF Protein Data Bank

f t y d

---

Structure Summary 3D View Annotations Experiment Sequence Genome Versions

< Biological Assembly 1 ? >



## 5L5G

Plexin A2 full extracellular region, domains 1 to 8 modeled, data to 10 angstrom

DOI: 10.2210/pdb5L5G/pdb

**Classification:** SIGNALING PROTEIN

**Organism(s):** Mus musculus

**Expression System:** Homo sapiens

**Mutation(s):** No

**Deposited:** 2016-05-28 **Released:** 2016-07-06


**Deposition Author(s):** Janssen, B.J.C., Kong, Y., Malinauskas, T., Vangoor, V.R., Coles, C.H., Kaufmann, R., Ni, T., Gilbert, R.J.C., Padilla-Parra, S., Pasterkamp, R.J., Jones, E.Y.

**Experimental Data Snapshot**

Metric	Percentile Ranks	Value
X-RAY DIFFRACTION		

wwPDB Validation

3D Report Full Report


















SMTL ID : 5l5g.1 (1 other biounit)

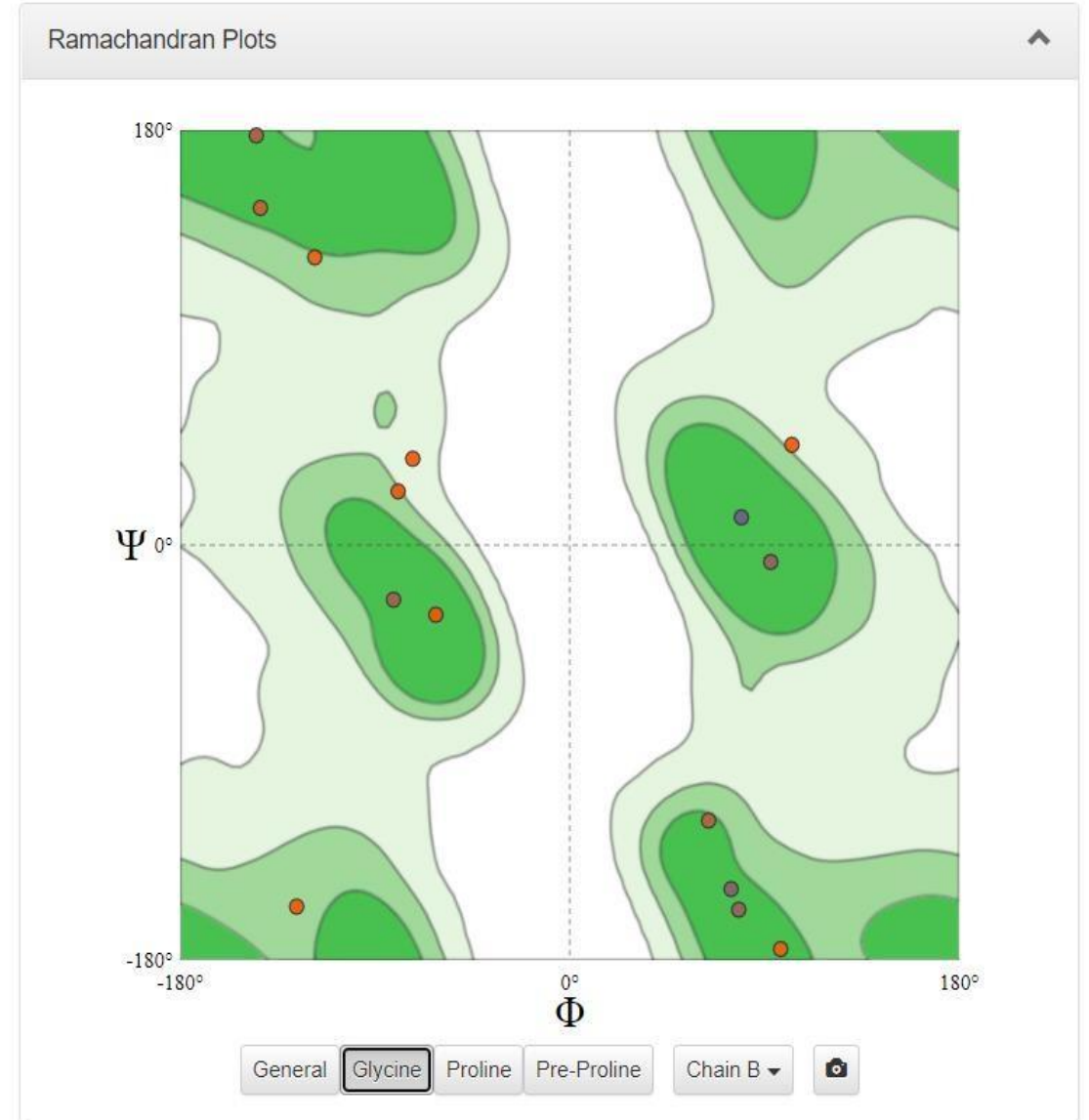
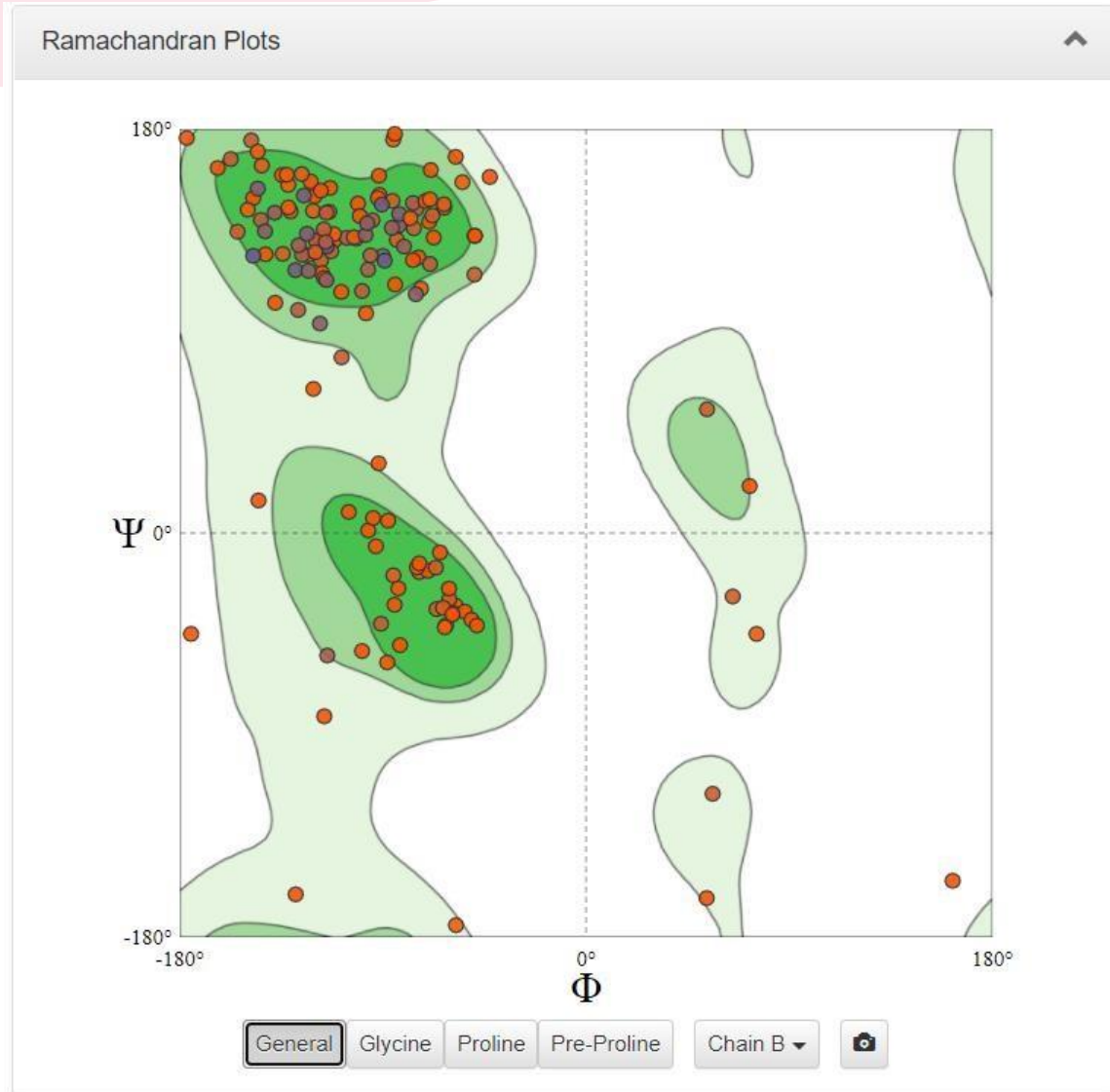
Plexin A2 full extracellular region, domains 1 to 8 modeled, data to 10 angstrom

<b>Coordinates</b>	PDB Format 
<b>Method</b>	X-RAY DIFFRACTION 10.00 Å
<b>Oligo State</b>	homo-dimer
<b>Ligands</b>	None
<b>Links</b>	RCSB  PDBe  PDBe-KB  PDBj  PDBsum  CATH 
<b>Citation</b>	Kong, Y. et al., Structural Basis for Plexin Activation and Regulation. Neuron (2016)  <a href="#">DOI</a> 
<b>Release Date</b>	2016-07-06
<b>Peptides</b>	Plexin-A2: A  B 
<b>SMTL:PDB</b>	SMTL Chain Id: A B PDB Chain Id: A C

Plexin-A2

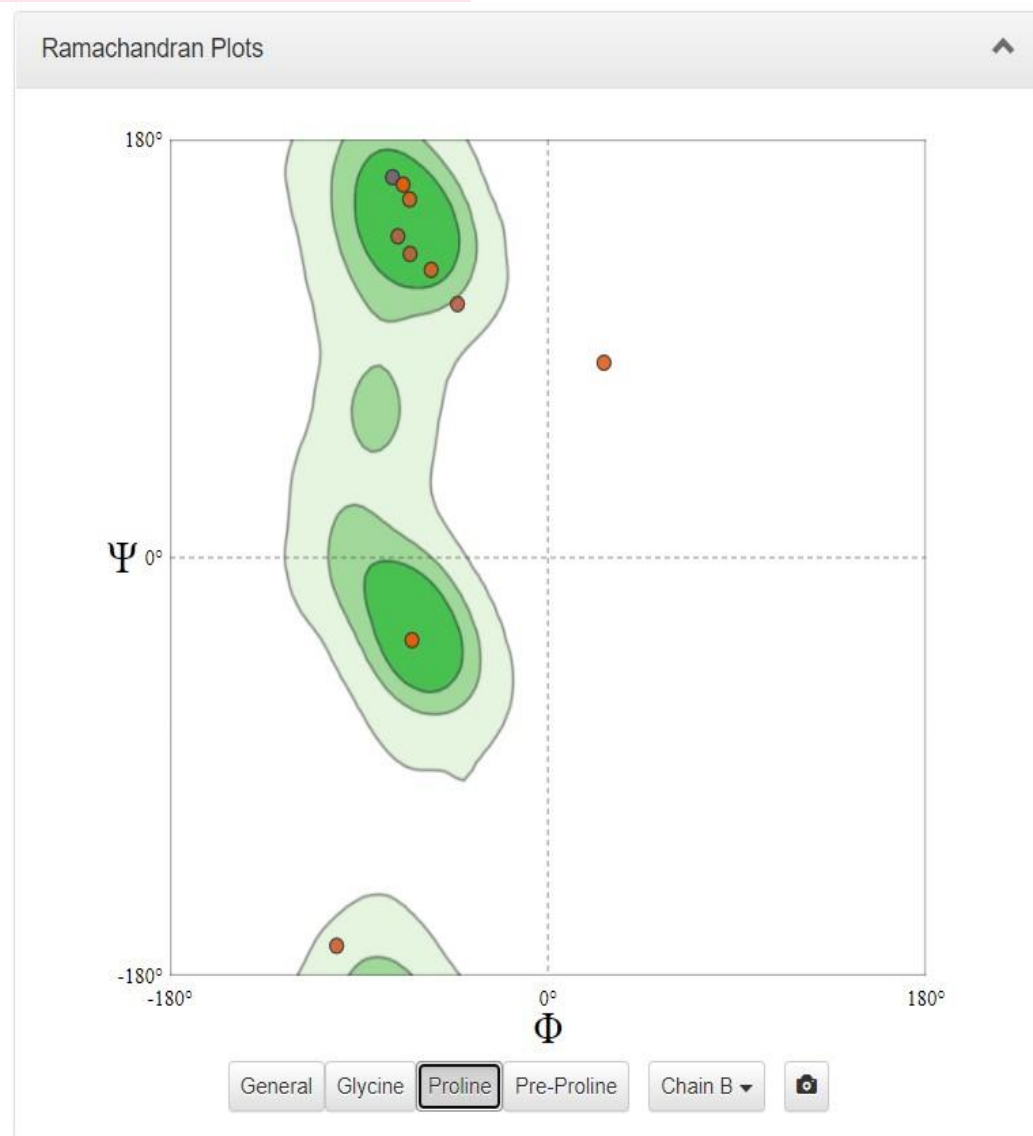
Seqres: **RTGMPQVISTFHSENRDWTFNHLTVHRRRTGAVIVGAINAVV** 40  
515g.1.A ----- **QVSTFHSENRDWTFNHLTVHRRRTGAVIVGAINAVV** 40  
515g.1.B ----- **QVSTFHSENRDWTFNHLTVHRRRTGAVIVGAINAVV** 40  
  
Seqres: **LTSTGNLTIQVAHKTSGFEENKACVYFLIVQPCSEVLTLTN** 80  
515g.1.A ----- **LTSTGNLTIQVAHKTSGFEENKACVYFLIVQPCSEVLTLTN** 80

# VALIDATION OF THE STRUCTURE



rotation

# VALIDATION OF THE STRUCTURE



# SUMMARY OF HOMOLGY MODELLING

For homology modelling, we searched for the templates which were most similar to our protein and then chose the one which had the max seq identity percentage. Then we selected an another template which was available in our blast result and checked whether the structure we got on swiss model is similar to the one available on PDB. We provided the predicted tertiary structure of our protein, visualization of protein and validation of structure as a result.

# REFERENCES

1.

1. <https://www.uniprot.org/>

2. [https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE\\_TYPE=BlastSearch&LINK\\_LOC=blasthome](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome)

3.

3. <https://swissmodel.expasy.org/interactive>

4. <https://www.ncbi.nlm.nih.gov/>

5. <https://www.ebi.ac.uk/Tools/msa/clustalo/>

6. <https://www.rcsb.org/>



For BLAST, we pasted the sequence of our target protein, i.e, Po8F94 Fibrocystin, which was retrieved in FASTA format from uniprot to the blastp program. The top 50 results were analysed, from which the top 10 sequences (including the target protein) were chosen for multiple sequence alignment. Graphic summary of distribution of the top 50 blast hits on 50 subject sequences was also performed. And the lineage descendency in the taxonomy section was visualized to have a glance at the lineage from ancestors to descendants of species listed in the top 50 BLAST results.

## SUMMARY

For the multiple sequence alignment, the top 50 blast results were analyzed, from which the top 10 sequences were selected, which includes the target protein (Fibrocystin isoform 1 precursor [Homo Sapiens]) and, the next 9 protein sequences other than homo sapiens. Results obtained were analysed based on the relationship between the sequences by looking at the symbols, '\*', ':' and, '.'. Percentage identity matrix was generated. And the phylogenetic tree, phylogram, tree data and guide tree were built to understand the relationship of homology among the sequences.

For homology modelling, we searched for the templates which were most similar to our protein and then chose the one which had the max seq identity percentage. Then we selected an another template which was available in our blast result and checked whether the structure we got on swiss model is similar to the one available on PDB. As a result we provided the predicted tertiary structure of our protein, visualization of protein and validation of structure.

A large, horizontal, pink brushstroke graphic with a textured, hand-painted appearance. It has irregular, feathered edges and a slightly wavy shape, resembling a thick stroke of paint. The color is a vibrant pink. Centered within this brushstroke is the text "THANK YOU !".

**THANK YOU !**