

# HOMOLOGY MODELING AND STRUCTURAL ANALYSIS OF FIBROCYSTIN PROTEIN



AYUSHMA JOSHI(19BCB0143)

# **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.

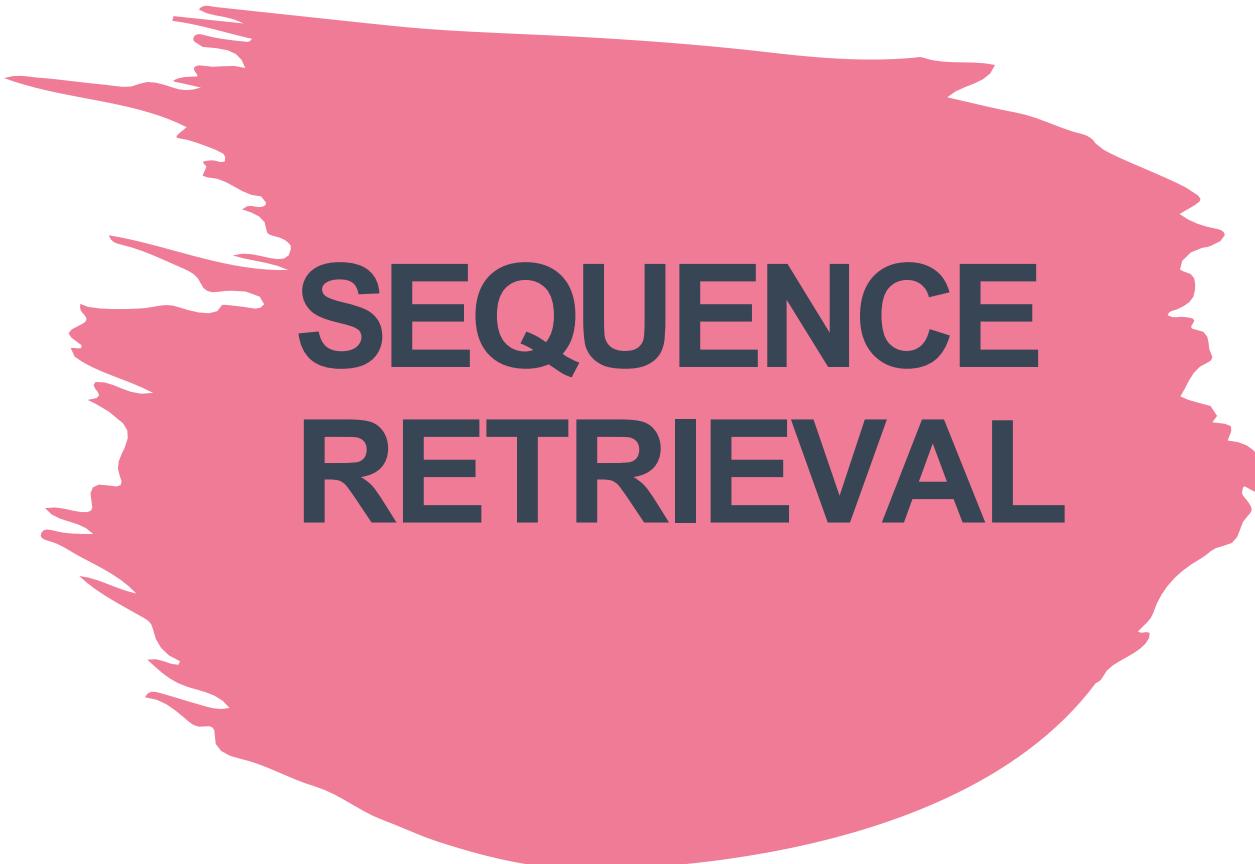
## Symptoms:

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:** Homo sapiens  
**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
MTAWLISLMSIEVLLAVRHLSLHIEPEEGSLAGGTWITVIIFDGLELGVLYPNNGSQLEI
HLVNVNMVVPAALRSVPVCDVFPVFLDLPVVTCRTRSVLSEAHEGLYFLEAYFGGQLVSSPN
PGPRDSCTFKFSKAQTPIVHQVYPPSGVPGKLIHVYGIITGRLETDFDAEYIDSPVIL
EAQGDKWVTPCSLINRQMGSYCPIQEDHGLGTLQCHVEGDIQGSQNVFSVFNKGKSMVH
KKAWLISAKQDLFLYQTHSEILSVPETGSLLGGRTNTITGDFFDNSAQVTIAGIIPCDIR
HVSPRKIECTTRAPGKDVRLLTTPQPGNRGLLFEVGDAVEGLELTEATPGYRWQIVPNASS
PFGFWSQEGQPFRARLSGFVAPETNNYTFWIQAQSASLHFWSWSEEPRTKVKVASISVG
TADWFDSWEQRNQDEGTWQQKTPKLELLGGAMYYLEAEHHGIAPSRGMRIGVQIHTWLNP
DVVTTYLREKHQIRVRAQRQLPEVQVLNVNSGRGNFFLTWDNVSSQPI PANATAHLIQTTIE
ELLAVKCKLEPLWSNILLRLGFERGPEVSNSDGDLSGTEPFCGRFSLRQPRHLVLTPPA
AQKGYRLDQYTHLCLAYKGHMNKILKMIVSFTIGFQNMVKNTTCDWSLRTSPESWQFDC
TDLWETCVRCFGDLQPPPANSPLVHQINLLPLAQETGLFYVDEIIIADTNVTVSQADSG
TARPGGNLVESVSVGSPPVSVTSWLAGCTELPLITARSVPTEGTEEGSGLVLVTTQR
RQRTSPLLGGHFRIQLPNTVISDPVQISAHHLHQLLQNNADDFTSRYLNASDFTVKEDL
YTCYEHVWTLSWSTQIGDLPNFIRVSDENLTGVNPAAATRVVYDGGVFLGPIFI GDMMLATA
NQHTQVVVRVNDVPAHCPSCSFQYLQGSTPCVHSWYSIDGDINLMIIYITGTGFSGDSQ
FLQVTVNKTSCKVIFSNQTNVVCQTDLLPVGMIHRLMVLVRPSGLAISATGEDLFNVKPR
LDMVEPSRAADIGGLWATIRGSSLEGVSLILFGSYSCAINVATSNSSRIQCKVPPRGKDG
RIVNVTVIRGDYSAVLRAFTYVSSLNPVIVTLSRNISNIAGGETLVIGVARLMNYTLD
VEVHVQDALAPVHTQSAGLEVALPPLPAGLHRISVISINGVSIHSQGVDLHIIQYLTEVFS
IEPCCGSLLGGTILSISGIGFSRDPALVWVLVGNRSCDIVNLTEASIWCE TLPA PQIPDA
GAPTVPAAVEVWAGNRRFARGPSPSLVGKGFTFMYEEAAATPVVTAMQGEITNSSL LHVG
GSNLSNSVILLGNLNCDVETQSFFQGNVSLSGCSIPLHSLEAGIYPLQVRQKQMGFANMSV
VLQQFAVMPRIMAIFPSQGSACGGTILT V RGLL LNSRRSRV DLSGPFTC V ILSL GDHT
ILCQVSLEGDPLPGASFSLNVTVLVNGLTSECQGNCLTFIREEASP VMDALSTNTSGSLT
TVLIRGQRLATTADEPMVFVDDQLPCNVTF FNASHVWCQTRDLAPGPHYLSVFYTRNGYA
CSGNVSRHFYIMPQVFHYFPKNFSLHGGSLLTIEGTGLRGQNTTSVYIDQQTCLTVNIGA
ELIRCIVPTGNGSVALEIEVDGLWYHIGVIGYNKAFTPELISISQSDDILTFAVAQISGA
ANIDIFIGMSPCVGSGNHTVLCQVVPSPAGEYHVRGYDCIRGWA SALLVFTSRVIITA
VTENFGCLGGRLVHVGAGFSPGNVSAAVCGAPCRVLANATVSAFSCLVLP LDVSLAFLC
GLKREEDSCEAARHTVQCQLTVAMATEQLLESWPYLYICEESSQCLFV P D HWA ESMF P S
FSGLFISP KLERDEV LIYNSCNITMETEAEEMECETPNQPI TVKITEIRKRWGQNTQGNF
SLQFCRRWSRTHSWFPERLPQDGDNVTVENQQLL LDTNTSILNLLHIKGKLI FMAPGP
IELRAHAILVSDGGELRIGSEDKPFQGRAQITLYGSSYSTPFFPYGVKF LA VRNGTL SLH
GSLPEVIVTC L RATAHALDTVLA EDAV DWNP GDEV VIIS GTGVKGAKPMEEIVTVETVQ
DTDLYLKSP LRYSHNFTENWVAGEHHILKATV ALLSRSITI IQGNLT NEREKLLVSCQEA N
APEGNLQHCLYSMSEKMLGS RD MGAR VIV QSFPEEPSQVQLKG VQFQV LQGA FH KHLSSL
TLVGAMRESFIQGCTVRNSFSRGLSMCGTLGLKVDSN VFY NILGHALLVGT CTEMRY ISW
EAIHGRKDDW SGHGN IIIRNNVIIQVSGAEGLSNPEMLTPSGIYICSP TNVIEGNRVC GAG
```

# SEQUENCE (FASTA FORMAT) cont...

```
YGYFFHLMTNQTSQAPLLSFTQNIAHSCTRYGLFVYPKFQPPWDNVVTGTLFQSFTVWES
AGGAQIFRSSNLRLKNFKVYSCRDFGIDVLESANTSVTDSLLLGHFAHKGSLCMSSGIK
TPKRWELEMVSNTTFVNFDLINCAIRTCSDCSQGGGGFTVKTSQLKFTNSSLVAFPFPH
AAILEDLDGSLSGKNRSHILASMETLSASCLVNSSFGRVVHGSACGGGVLFHRMSIGLAN
TPEVSYDLTM TD SRN KTTV NY VRDT LSN PRGW MALL LDQET YSL QSE NLW INR SLQ YSA
TFDNFAPGN Y LLLVHTDLPPYDILLRCGSRVGLSFPLPSPGQNQGCDWF FNSQLRQLT
YLVS GEGQVQV I LRVKEGMPPTISASTSAPESALKWSLPETWQGV E EGWGGYNNTIPGPG
DDVLILPNRTV LVD TDL PFF KGL YVM GT LDF PVDRS NVL SVACM VIAGGELKVGTLENPL
EKEQKLLILL RASEGVFC DRMNGI HID PGTIGVYGVHLYSAYPKNSWTHLGADIASGNE
RIIVED A DV DWR PHDKIVLSSSSYEPHEAEVLT VKEVKGHHV RIYERLKHRHIGSVHVTED
GRHIRLAAEVGLLTRNIQIQPDVSCRGRLFVGFSFRKSSREEFSGVLQLLNVEIQNFGSPL
YSSVEFSNVSAGSWIISSTLHQSCGGGIHAAAASHGVLLNDNIVFGTAGHGIDLEGQAYTV
TNNLVVLMTQPAWSTI WAGIKVNQVKDINLHGNVVAGSERLGFHIRGHKCSSCELLWSD
NVAHSSLHGLHLYKESGLDNCTRISGF LAFKNFDYGA MLHVENSVEIENITLVDNTIGLL
AVVYVFSAPQNSVKKVQIVLRNSVIVATSSSFDCI QDKV KPH SANLTSTDRA PSNPRGGR
IGILWPVFTSEPNQWPQE PWHKVRNDHSISGIMKLQDVTFSSFVKSCYSDDLVCILPNA
ENSGIMHPITAERTRMLKIKDKNKFYFP SLQPRKDLGKVVCPELDCASPRKYLFKDLDGR
ALGLPPPVS VFPKTEAEWTASFFNAGTFREEQKCTYQFLMQQFICKQTDQVVLILDSADA
IWAIQKLYPVVSVTSGFVDVFSSVNANIPCSTSGSVSTFYSILPIRQITKVCMDQTPQV
LRFFLLGNKSTS KL LAVFYHELQSPHVFLGESFIPPTLVQSASLLNESIGANYFNIMD
NLLYVVLQGE EPIEIRSGVSIHLALTVMVSVLEKGWEIVILERLTNFLQIGQNZIRFIHE
MPGHEETLKAIA DSRAKRKRNCPTVTCTSHYRRVGQRRPLMMEMNSHRASPPMTVETISK
VIVIEIGDSPTVRSTGMISLSSNKLQNL AH RVITAQQTGVLENVLNMTIGALLVTQSKG
VIGYGN TSSFKTGNL IYIRPYALSILVQPSDGEVGNELPVQPQLVF LDEQNR RVESLGPP
SEPWTISASLEGASDSVLKGCTQAETQDG YV SFYNLA VLISGSNWHFIFTV TSPPGVNFT
ARSKPFAVLPVTRKEKSTII LA ASLSSV ASWL ALSCLVCCWLKR SKSRKTKPEEIPESQT
NNQNIHIHISSKRRESQGP KKEDTVVG EDMRMK VMLGKV NQC PHQLMNGV SRRK VSR HIV
REEEAAVPAPGTTGITSHGHICAPGAPAQ QVYLQETGNWKEGQEQLLRYQLAGQNQ LLL
CPDFRQERQQLPGQSRLSKQSGSLGLSQEKKA SCGATEAFCLHSVH PETI QEQL
```



# BLAST

---

# BLAST SCREENSHOTS

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

Log in

**COVID-19 Information** X

[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 8C7SMPV8016 Search expires on 04-27 10:01 am Download All ▾  
Program BLASTP ? Citation ▾  
Database nr See details ▾  
Query ID lcl|Query\_69080  
Description sp|P08F94|PKHD1\_HUMAN Fibrocystin OS=Homo sapien ...  
Molecule type amino acid  
Query Length 4074  
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      E value      Query Coverage  
 to        to        to   
Filter Reset

Descriptions Graphic Summary Alignments Taxonomy

Feedback

# BLAST SCREENSHOTS

**Descriptions**   Graphic Summary   Alignments   Taxonomy

**Sequences producing significant alignments**

Download Select columns Show 50 ?

select all 50 sequences selected

	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

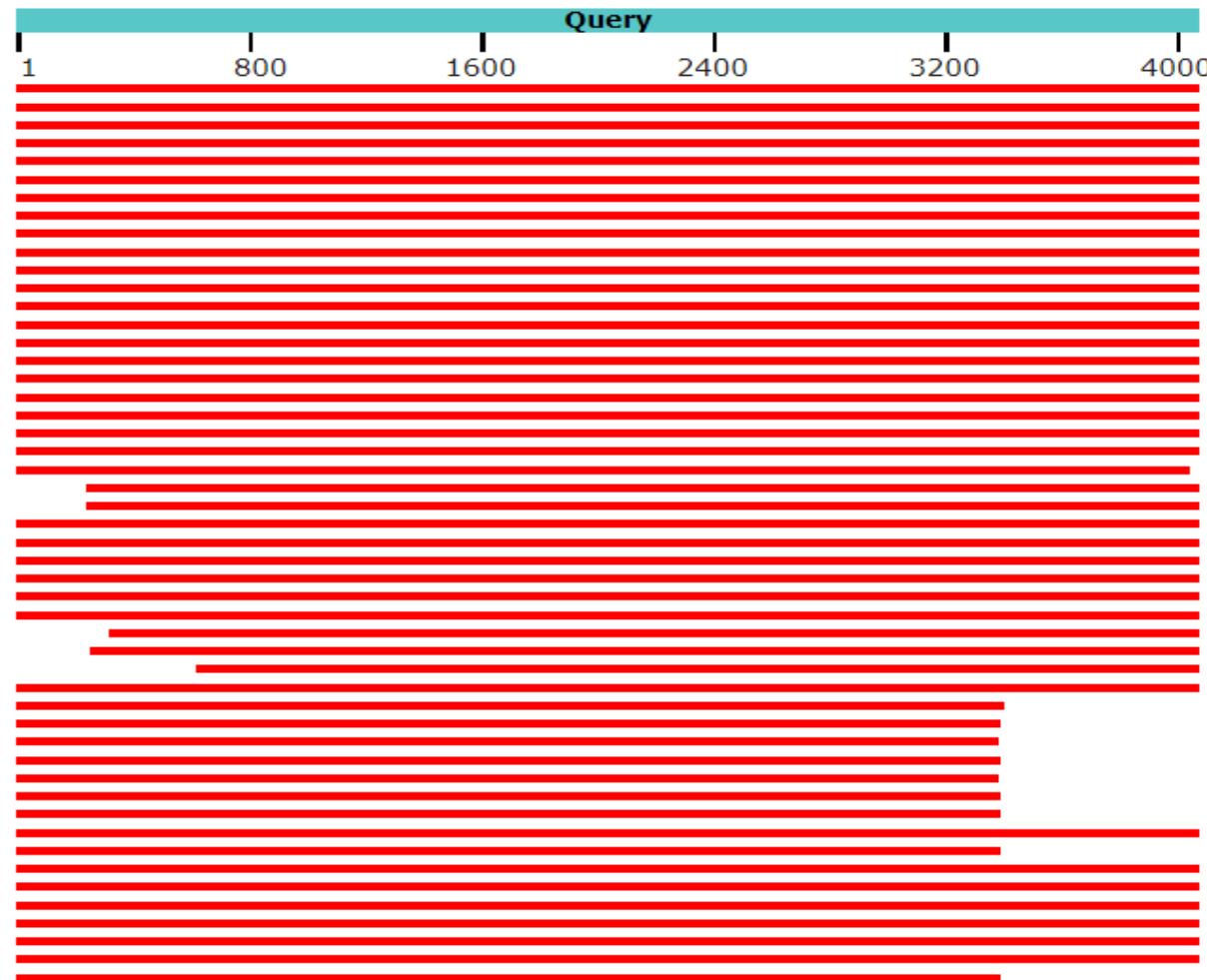
# BLAST

Among all 50 sequences, the top 10 sequences were selected, which includes the target protein (Fibrocystin isoform 11 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	4074	NP_619639.3
fibrocystin isoform X1 [Pan troglodytes]	Pan troglodytes	chimpanzee	8328	8328	100%	0	98.94	4074	XP_518534.4
fibrocystin [Pan paniscus]	Pan paniscus	pygmy chimpanzee	8323	8323	100%	0	98.87	4074	XP_008956905.2
fibrocystin [Gorilla gorilla gorilla]	Gorilla gorilla gorilla	western lowland gorilla	8273	8273	100%	0	98.31	4074	XP_030868334.1
fibrocystin isoform X2 [Pan troglodytes]	Pan troglodytes	chimpanzee	8271	8271	100%	0	98.4	4052	XP_016811152.2
fibrocystin isoform X3 [Pan troglodytes]	Pan troglodytes	chimpanzee	8259	8259	100%	0	98.33	4049	XP_016811153.2
fibrocystin isoform X1 [Pongo abelii]	Pongo abelii	Sumatran orangutan	8179	8179	100%	0	97.32	4074	XP_024104185.1
fibrocystin [Hylobates moloch]	Hylobates moloch	silvery gibbon	8140	8140	100%	0	96.54	4073	XP_031997014.1
fibrocystin [Nomascus leucogenys]	Nomascus leucogenys	northern white-cheeked	8121	8121	100%	0	96.42	4073	XP_003254212.2
PREDICTED: LOW QUALITY PROTEIN: fibrocystin [Rhinopithecus bieti]	Rhinopithecus bieti	black snub-nosed monkey	8041	8041	100%	0	95.12	4100	XP_017705097.1

# BLAST SCREENSHOTS

**Distribution of the top 50 Blast Hits on 50 subject sequences**



# BLAST SCREENSHOTS

Descriptions	Graphic Summary	Alignments	Taxonomy
Reports	Lineage	Organism	Taxonomy
50 sequences selected <span>?</span>			
Organism	Blast Name	Score	Number of Hits
<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="#">Pan troglodytes</a>	<a href="#">primates</a>	8328	6
<a href="#">Pan paniscus</a>	<a href="#">primates</a>	8323	1
<a href="#">Gorilla gorilla gorilla</a>	<a href="#">primates</a>	8273	1
<a href="#">Pongo abelii</a>	<a href="#">primates</a>	8179	4
<a href="#">Hylobates moloch</a>	<a href="#">primates</a>	8140	1
<a href="#">Nomascus leucogenys</a>	<a href="#">primates</a>	8121	1
<a href="#">Rhinopithecus bieti</a>	<a href="#">primates</a>	8041	1
<a href="#">Theropithecus gelada</a>	<a href="#">primates</a>	8038	1
<a href="#">Macaca fascicularis</a>	<a href="#">primates</a>	8034	1
<a href="#">Trachypithecus francoisi</a>	<a href="#">primates</a>	8033	1
<a href="#">Macaca nemestrina</a>	<a href="#">primates</a>	8033	1
<a href="#">Papio anubis</a>	<a href="#">primates</a>	8031	1
<a href="#">Colobus angolensis palliatus</a>	<a href="#">primates</a>	8026	1
Description			
<a href="#">Homo sapiens hits</a>			
<a href="#">Pan troglodytes hits</a>			
<a href="#">Pan paniscus hits</a>			
<a href="#">Gorilla gorilla gorilla hits</a>			
<a href="#">Pongo abelii hits</a>			
<a href="#">Hylobates moloch hits</a>			
<a href="#">Nomascus leucogenys hits</a>			
<a href="#">Rhinopithecus bieti hits</a>			
<a href="#">Theropithecus gelada hits</a>			
<a href="#">Macaca fascicularis hits</a>			
<a href="#">Trachypithecus francoisi hits</a>			
<a href="#">Macaca nemestrina hits</a>			
<a href="#">Papio anubis hits</a>			
<a href="#">Colobus angolensis palliatus hits</a>			

For BLAST, I pasted the sequence of our target protein, i.e, P08F94 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 descendancy 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	MTAWLISLMSIEVLLLAVPHLSLHIEPEEGSLAGGTWITVIFDGLESGVLYPNNGSQLEI	60
XP_003254212.2	MTAWLISLMSIEVLLLAVPHLSLHIEPEEGSLGGGTWITVIFDGLELGVLYPNNGSQLEI	60
XP_031997014.1	MTAWLISLMSIEVLLLAVPYLSLHIEPEEGSLGGGTWITVIFDGLELGVLYPNNGSQLEI	60
XP_024104185.1	MTAWLISLMSIEVLLLAVPHLSLHIEPEEGSLAGGTWITVIFDGLELGVLYPNNGSQLEI	60
XP_030868334.1	MTAWLISLMSIEVLLLAVPHLSLHIEPEEGSLAGGTWITVIFDGLELGVLYPNNGSQLEI	60
NP_619639.3	MTAWLISLMSIEVLLLAVRHLSLHIEPEEGSLAGGTWITVIFDGLELGVLYPNNGSQLEI	60
XP_008956905.2	MTAWLISLMSIEVLLLAVPHLSLHIEPEEGSLAGGTWITVIFDGLELGVFYPNNGSQLEI	60
XP_016811153.2	MTAWLISLMSIEVLLLAVPHLSLHIEPEEGSLAGGTWITVIFDGLELGVLYPNNGSQLEI	60
XP_016811152.2	MTAWLISLMSIEVLLLAVPHLSLHIEPEEGSLAGGTWITVIFDGLELGVLYPNNGSQLEI	60
XP_518534.4	MTAWLISLMSIEVLLLAVPHLSLHIEPEEGSLAGGTWITVIFDGLELGVLYPNNGSQLEI	60
	***** : ***** . ***** * : *****	
XP_017705097.1	HLVNMMVAPALPSIPCDVFPVFLDPVVMCQTRSLSEAHEGLYSLEAYFGQLVSSPS	120
XP_003254212.2	HLVNMMVMPALRSIPCDVFPVFLDPVVTCCRTRSVLSEAHEGLYSLEAYFGQLVSSPN	120
XP_031997014.1	HLVNMMVMPALRSIPCDVFPVFLDPLPVTCCRTRSVLSEAHEGLYSLEAYFGQLVSSPN	120
XP_024104185.1	HLVNMMVMPALRSIPCDVFPVFLDLPVVTCCRTRSVLSEAHEGLYFLEAYFRGQLVSSPN	120
XP_030868334.1	HLVNMMVMPALRSIPCDVFPVFLDLPVVTCCRTRSVLSEAHEGLYFLEAYFGQLVSSPN	120
NP_619639.3	HLVNMMVMPALRSVPDCVFPVFLDLPVVTCCRTRSVLSEAHEGLYFLEAYFGQLVSSPN	120
XP_008956905.2	HLVNMMVMPALRSVPDCVFPVFLDLPVVTCCRTRSVLSEAHEGLYFLEAYFGQLVSSPN	120
XP_016811153.2	HLVNMMVMPALRSVPDCVFPVFLDLPVVTCCRTRSVLSEAHEGLYFLEAYFGQLVSSPN	120
XP_016811152.2	HLVNMMVMPALRSVPDCVFPVFLDLPVVTCCRTRSVLSEAHEGLYFLEAYFGQLVSSPN	120
XP_518534.4	HLVNMMVMPALRSVPDCVFPVFLDLPVVTCCRTRSVLSEAHEGLYFLEAYFGQLVSSPN	120
	*** : *** . *** : ***** : * : *** : ***** * : *** : ***** . *** : *** .	
XP_017705097.1	PGPRDSCTFKFSKAQTPIVHKVYPPSGVPGKLIHVYGIWTGRLETDFDAEYIDSPVIL	180
XP_003254212.2	PGPRDRCTFKFSKAQTPIVHQVYPPSGVPGKLIHIYGIWTGRLESFDFDAEYIDSPVIL	180
XP_031997014.1	PGPRDRCTFKFSKAQTPIVHQVYPPSGVPGKLIHIYGIWTGRLESFDFDAEYIDSPVIL	180
XP_024104185.1	PGPRDSCTFKFSKAQTPIVHHVYPPSGVPGKLIHVYGIWTGRLETDFDAEYIDSPVIL	180
XP_030868334.1	PGPRDSCTFKFSKAQTPIVHQVYPPSGVPGKLIHVYGIWTGRLETDFDAEYIDSPVIL	180
NP_619639.3	PGPRDSCTFKFSKAQTPIVHQVYPPSGVPGKLIHVYGIWTGRLETDFDAEYIDSPVIL	180
XP_008956905.2	PGPRDSCTFKFSKAQTPIVHQVYPPSGVPGKLIHVYGIWTGRLETDFDAEYIDSPVIL	180
XP_016811153.2	PGPRDSCTFKFSKAQTPIVHQVYPPSGVPGKLIHVYGIWTGRLETDFDAEYIDSPVIL	176
XP_016811152.2	PGPRDSCTFKFSKAQTPIVHQVYPPSGVPGKLIHVYGIWTGRLETDFDAEYIDSPVIL	180
XP_518534.4	PGPRDSCTFKFSKAQTPIVHQVYPPSGVPGKLIHVYGIWTGRLETDFDAEYIDSPVIL	180
	***** . ***** . ***** . ***** . ***** . ***** . ***** . ***** .	

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

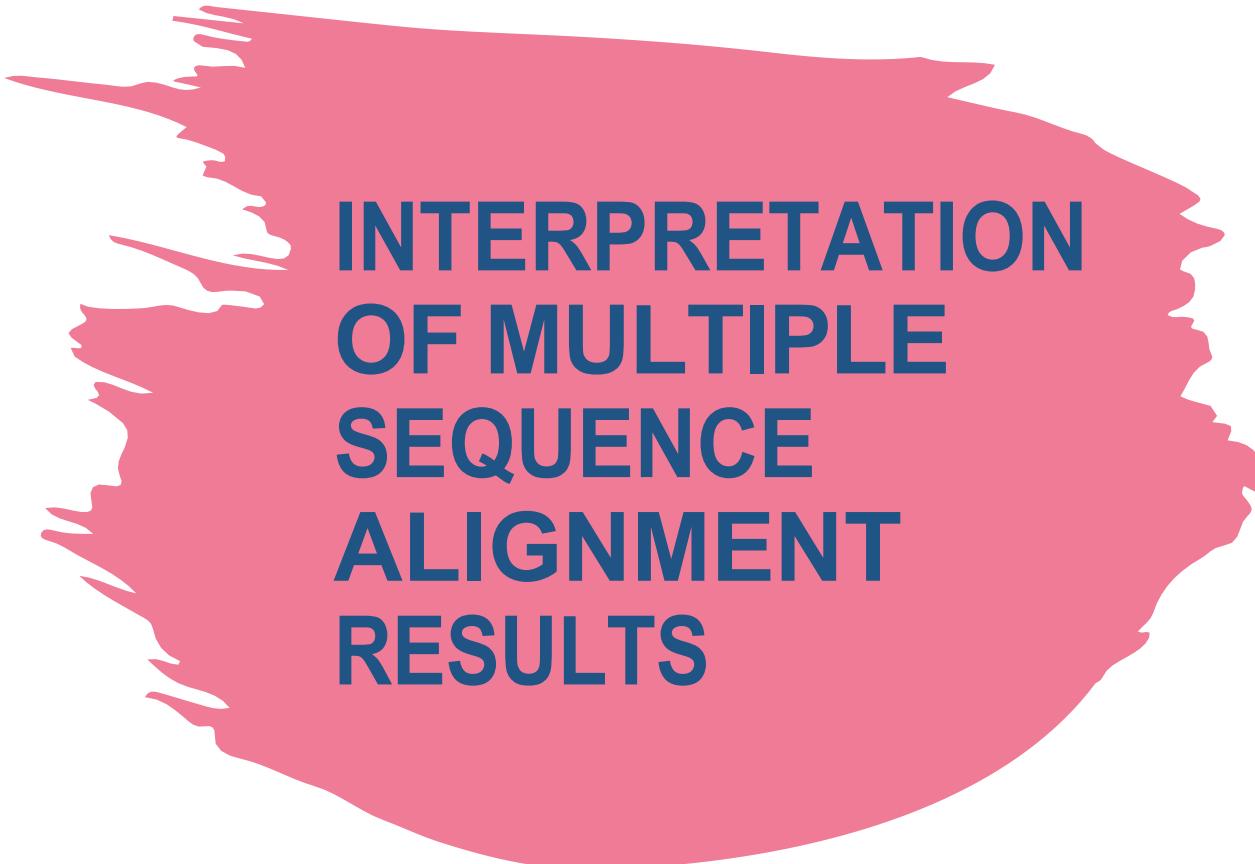
XP_017705097.1	EAQGDKWVTPCSLVNRQTGSCYPIQEDHGLGTLCCHVEGYIGSQNVSFSVFNKGSMIH	240
XP_003254212.2	EAQGDKWVTPCSLVNKQTGSCYPIQEDHGLGTLCCHVEGYIGSQNVSFSVFNKGSMVH	240
XP_031997014.1	EAQGDKWVTPCSLVNRQTGSCYPIQADHGLGTLCCHVEGYIGSQNVSFSVFNKGSMVH	240
XP_024104185.1	EAQGDKWVTPCSLINRQTGRCPYPIQEDHGLGTLCRVEGYIGSQNVSFSVFNKGSMVH	240
XP_030868334.1	EAQGDKWVTPCSLINRQMGSRYPIQEDHGLGTLCRVEGYIGSQNVSFSVFNKGSMVH	240
NP_619639.3	EAQGDKWVTPCSLINRQMGSYPIQEDHGLGTLCRVEGYIGSQNVSFSVFNKGSMVH	240
XP_008956905.2	EAQGDKWVTPCSLINRQMGSYPIQEDHGLGTLCRVEGYIGSQNVSFSVFNKGSMVH	240
XP_016811153.2	-----YPIQEDHGLGTLCRVEGYIGSQNVSFSVFNKGSMVH	215
XP_016811152.2	EAQGDKWVTPCSLINRQMGSYPIQEDHGLGTLCRVEGYIGSQNVSFSVFNKGSMVH	240
XP_518534.4	EAQGDKWVTPCSLINRQMGSYPIQEDHGLGTLCRVEGYIGSQNVSFSVFNKGSMVH	240
	***** *****: *****: *****: *	
XP_017705097.1	KKAWLISAKQDLFLYQTHSEILSVFPETGSLGGRTDITITGDFFDNAQVTIAGIPCDIR	300
XP_003254212.2	KKAWLISAKQDLFLYQTHSEILSVFPETGSLGGRTDITITGDFFHSSAQVTIAGIPCDIR	300
XP_031997014.1	KKAWLISAKQDLFLYQTHSEILSVFPETGSLGGRTDITITGDFFDNAQVTIAGIPCDIR	300
XP_024104185.1	KKAWLISAKQDLFLYQTHSEILSVFPETGSLGGRTDITITGDFFDSSAQVTIAGIPCDIR	300
XP_030868334.1	KKAWLISAKQDLFLYQTHSEILSVFPETGSLGGRTDITITGDFFDNAQVTIAGIPCDIR	300
NP_619639.3	KKAWLISAKQDLFLYQTHSEILSVFPETGSLGGRTNTITITGDFFDNAQVTIAGIPCDIR	300
XP_008956905.2	KKAWLISAKQDLFLYQTHSEILSVFPETGSLGGRTDITITGDFFDNAQVTIAGIPCDIR	300
XP_016811153.2	KKAWLISAKQDLFLYQTHSEILSVFPETGSLGGRTDITITGDFFDNAQVTIAGIPCDIR	275
XP_016811152.2	KKAWLISAKQDLFLYQTHSEILSVFPETGSLGGRTDITITGDFFDNAQVTIAGIPCDIR	300
XP_518534.4	KKAWLISAKQDLFLYQTHSEILSVFPETGSLGGRTDITITGDFFDNAQVTIAGIPCDIR	300
	***** , ***** . ***** . *****	
XP_017705097.1	HVSPKKIECTTRAPGKDARLTTPQPGNRGLLFEVGDAVEGLELTEATPGYRWQIVPNASS	360
XP_003254212.2	HVSPRKIECTTRAPGKDARLTTPQPGNRGLLFEVGDAVEGLELTEATPGYRWQIVPNASS	360
XP_031997014.1	HVSPRKIECTTRAPGKDARLTTPQPGNRGLLFEVGDAVEGLELTEATPGYRWQIVPNASS	360
XP_024104185.1	HVSPRKIECTTRAPGKDVRLLTTPQPGNRGLLFEVGDAVEGLELTEATPGYRWQIVPNASS	360
XP_030868334.1	HVSPRKIECTTRAPGKDVRLLTTPQPGNRGLLFEVGDAVEGLELTEATPGYRWQIVPNASS	360
NP_619639.3	HVSPRKIECTTRAPGKDVRLLTTPQPGNRGLLFEVGDAVEGLELTEATPGYRWQIVPNASS	360
XP_008956905.2	HVSPRKIECTTRAPGKDVRLLTTPQPGNRGLLFEVGDAVEGLELTEATPGYRWQIVPNASS	360
XP_016811153.2	HVSPRKIECTTRAPGKDVRLLTTPQPGNRGLLFEVGDAVEGLELTEATPGYRWQIVPNASS	335
XP_016811152.2	HVSPRKIECTTRAPGKDVRLLTTPQPGNRGLLFEVGDAVEGLELTEATPGYRWQIVPNASS	360
XP_518534.4	HVSPRKIECTTRAPGKDVRLLTTPQPGNRGLLFEVGDAVEGLELTEATPGYRWQIVPNASS	360
	***: ***. *****. *****. *****	

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

XP_017705097.1	PFGFWSKEGQPFRARLSGFFVAPETNNYTFWIQADSQASLHFWSSEEPRTKVKVASISVG	420
XP_003254212.2	PFGFWSKEGQPFRARLSGFFVAPETNNYTFWIQADSQASLHFWSSEEPRTKVKVASISVG	420
XP_031997014.1	PFGFWSKEGQPFRARLSGFFVAPETNNYTFWIQADSQASLHFWSSEEPRTKVKVASISVG	420
XP_024104185.1	PFGFWSKEGQPFRARLSGFFVAPETNNYTFWIQADSQASLHFWSSEEPRTKVKVASISVG	420
XP_030868334.1	PFGFWSKEGQPFRARLSGFFVAPETNNYTFWIQADSQASLHFWSSEEPRTKVKVASISVG	420
NP_619639.3	PFGFWSKEGQPFRARLSGFFVAPETNNYTFWIQADSQASLHFWSSEEPRTKVKVASISVG	420
XP_008956905.2	PFGFWSKEGQPFRARLSGFFVAPETNNYTFWIQADSQASLHFWSSEEPRTKVKVASISVG	420
XP_016811153.2	PFGFWSKEGQPFRARLSGFFVAPETNNYTFWIQADSQASLHFWSSEEPRTKVKVASISVG	395
XP_016811152.2	PFGFWSKEGQPFRARLSGFFVAPETNNYTFWIQADSQASLHFWSSEEPRTKVKVASISVG	420
XP_518534.4	PFGFWSKEGQPFRARLSGFFVAPETNNYTFWIQADSQASLHFWSSEEPRTKVKVASISVG	420
	***** : *****	
XP_017705097.1	TADWFDSWEQNTDERTWQQKTPKLELLGGAKYYLEAEHHGIAPSRGMRIGVQIHNTWLNP	480
XP_003254212.2	TADWFDSWEQNQDEGTWQQKSPKLELLGGAKYYLEAEHHGIAPSRGMRIGVQIHNTWLNP	480
XP_031997014.1	TADWFDSWEQRNDEGTWQQKSPKLELLGGAKYYLEAEHHGIAPSRGMRIGVQIHNTWLNP	480
XP_024104185.1	TADWFDSWEQRNDEGTWQQKTPKLELLGGAKYYLEAEHHGIAPSRGMRIGVQIHNTWLNP	480
XP_030868334.1	TADWFDSWEQRNDEGTWQQKTPKLELLGGAMYYLEAEHHGIAPSRGMRIGVQIHNTWLNP	480
NP_619639.3	TADWFDSWEQRNDEGTWQQKTPKLELLGGAMYYLEAEHHGIAPSRGMRIGVQIHNTWLNP	480
XP_008956905.2	TADWFDSWEQRNDEGTWQQKTPKLELLGGAMYYLEAEHHGIAPSRGMRIGVQIHNTWLNP	480
XP_016811153.2	TADWFDSWEQRNDEGTWQQKTPKLELLGGAMYYLEAEHHGIAPSRGMRIGVQIHNTWLNP	455
XP_016811152.2	TADWFDSWEQRNDEGTWQQKTPKLELLGGAMYYLEAEHHGIAPSRGMRIGVQIHNTWLNP	480
XP_518534.4	TADWFDSWEQRNDEGTWQQKTPKLELLGGAMYYLEAEHHGIAPSRGMRIGVQIHNTWLNP	480
	***** : *** : *** : *****	
XP_017705097.1	DWTTYLREKHQIRARAQRLPEVQLNVSGRGNFFLTWDNVSSQPIPANATAHLIQTIE	540
XP_003254212.2	DWTTYLREKHQIRVRAQRLPEVQLNVSGRGNFFLTWDNVSSQPIPANATAHLIQTIE	540
XP_031997014.1	DWTTYLREKHQIRVRAQRLPEVQLNVSGRGNFFLTWDNVSSQPIPANATAHLIQTIE	540
XP_024104185.1	DWTTYLREKHQIRVQAQRLPEVQLNVSGRGNFFLTWDNVSSQPIPANATAHLIQTIE	540
XP_030868334.1	DWTTYLREKHQIRVRAQRLPEVQLNVSGRGNFFLTWDNVSSQPIPANATAHLIQTIE	540
NP_619639.3	DWTTYLREKHQIRVRAQRLPEVQLNVSGRGNFFLTWDNVSSQPIPANATAHLIQTIE	540
XP_008956905.2	DWTTYLREKHQIRVRAQRLPEVQLNVSGRGNFFLTWDNVSSQPIPANATAHLIQTIE	540
XP_016811153.2	DWTTYLREKHQIRVRAQRLPEVQLNVSGRGNFFLTWDNVSSQPIPANATAHLIQTIE	515
XP_016811152.2	DWTTYLREKHQIRVRAQRLPEVQLNVSGRGNFFLTWDNVSSQPIPANATAHLIQTIE	540
XP_518534.4	DWTTYLREKHQIRVRAQRLPEVQLNVSGRGNFFLTWDNVSSQPIPANATAHLIQTIE	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
```



# 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.

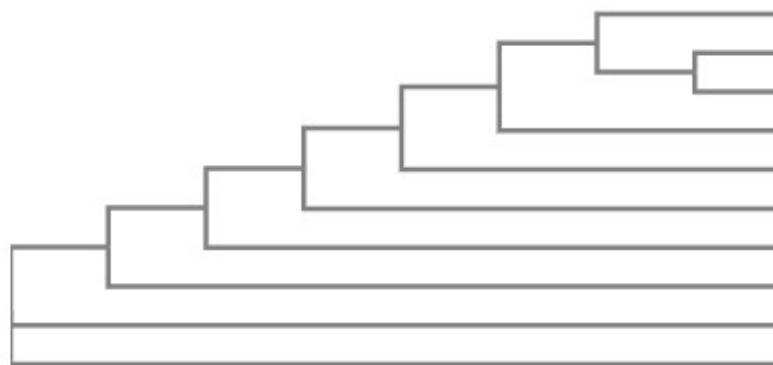
Two species are more related if they have a more recent common ancestor, and less related if they have a less recent common ancestor.

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



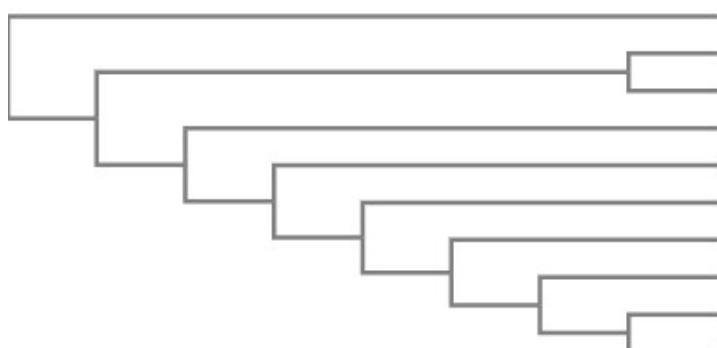
XP_017705097.1	0.03361
XP_003254212.2	0.00816
XP_031997014.1	0.00681
XP_024104185.1	0.01426
XP_030868334.1	0.00954
NP_619639.3	0.00618
XP_008956905.2	0.00222
XP_016811153.2	0
XP_016811152.2	0.00026
XP_518534.4	0

# 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



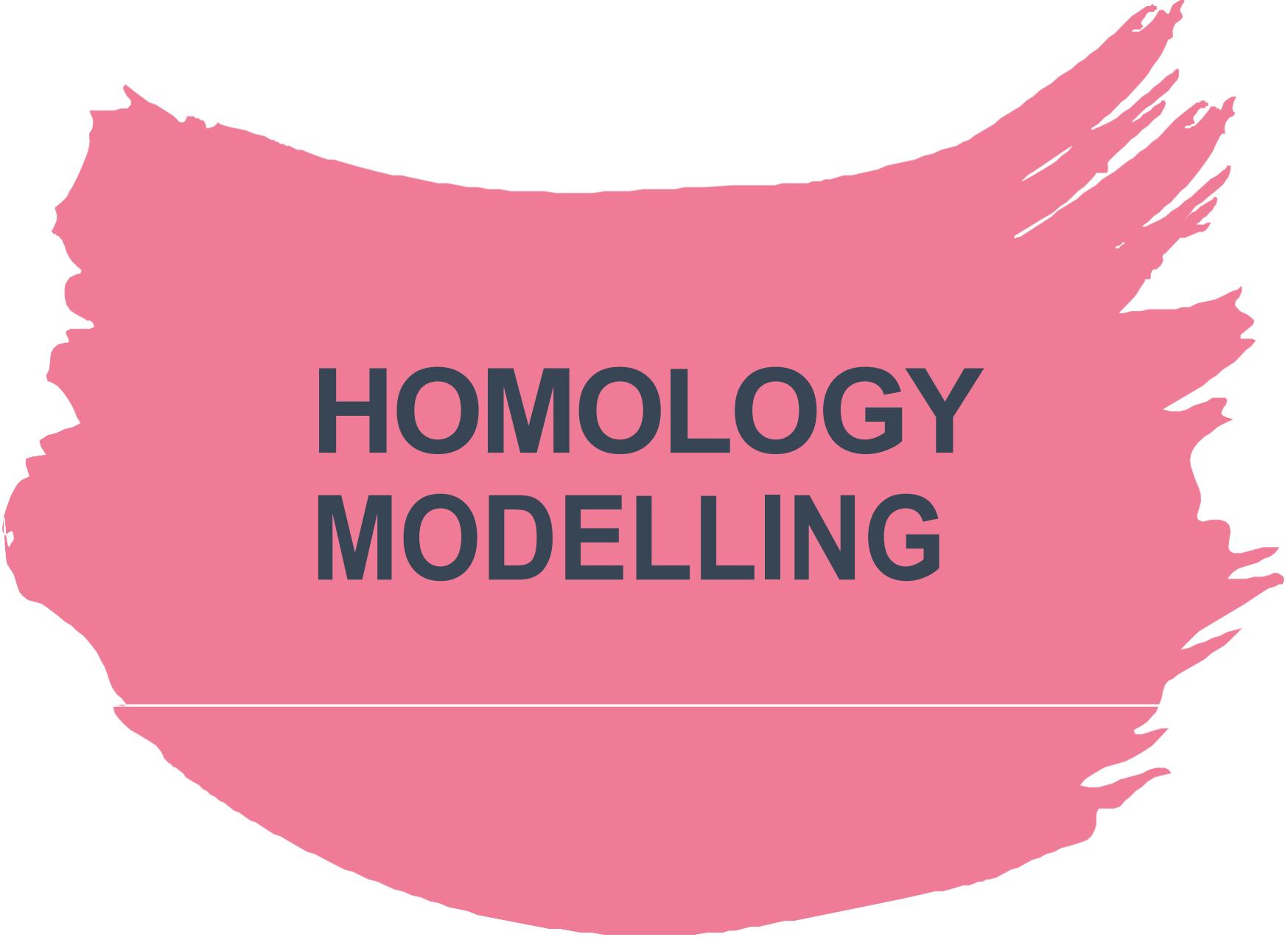
XP\_017705097.1 0.0263695  
XP\_003254212.2 0.00748834  
XP\_031997014.1 0.00748834  
XP\_024104185.1 0.0140023  
XP\_030868334.1 0.00819316  
NP\_619639.3 0.00561208  
XP\_008956905.2 0.0021269  
XP\_016811153.2 0.00191405  
XP\_016811152.2 0.000370188  
XP\_518534.4 0.000370188

## Guide Tree

```
(  
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



# **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**?

Search SWISS-MODEL Repository



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: MTAWLISILMSIEVLLLA  
Target: CRTRSVLSEAHEGLYF  
Target: EAQGDKWVTPCSL  
Target: LGGRTNITITGDFFDNSA  
Add Hetero Target Reset

Project Title: Untitled Project

Email: Optional

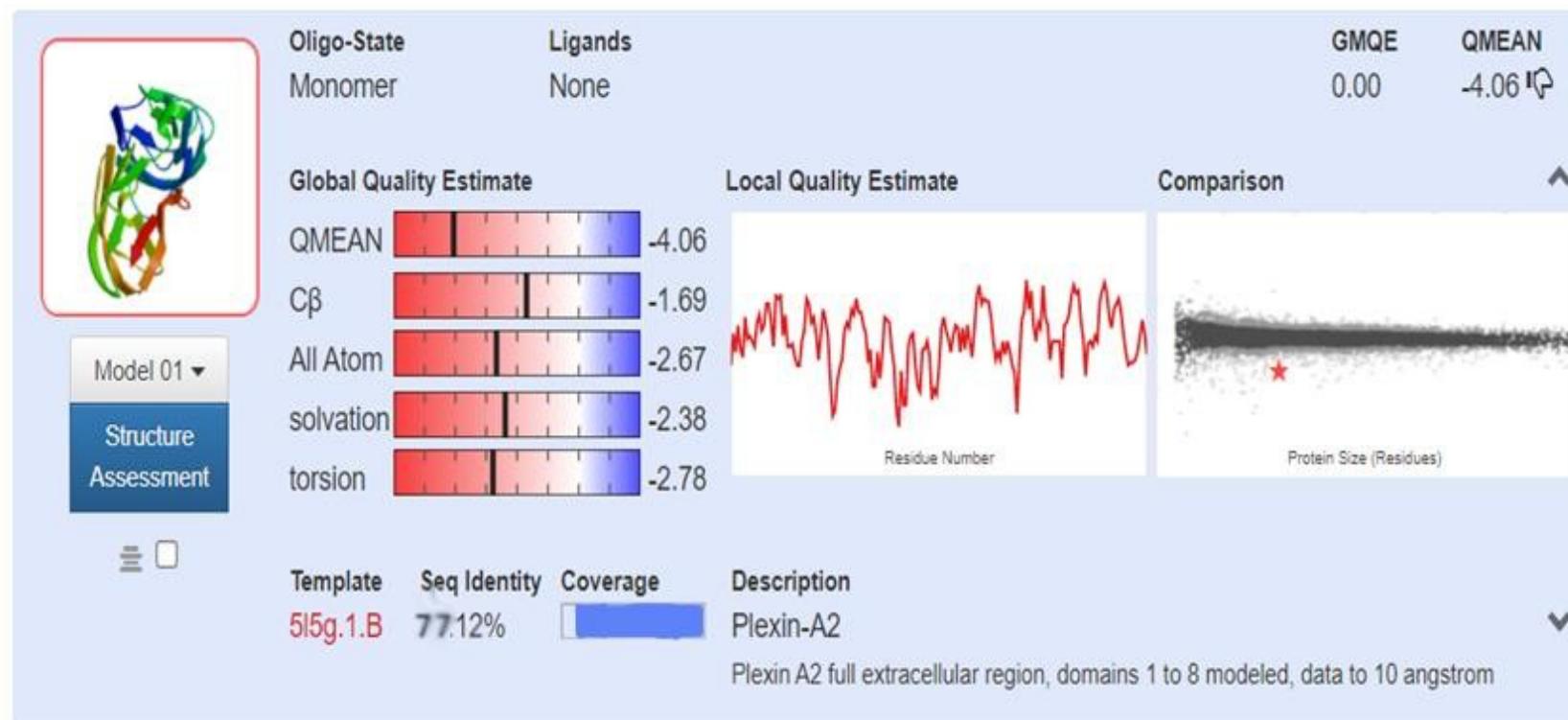
Search For Templates Build Model

Supported Inputs

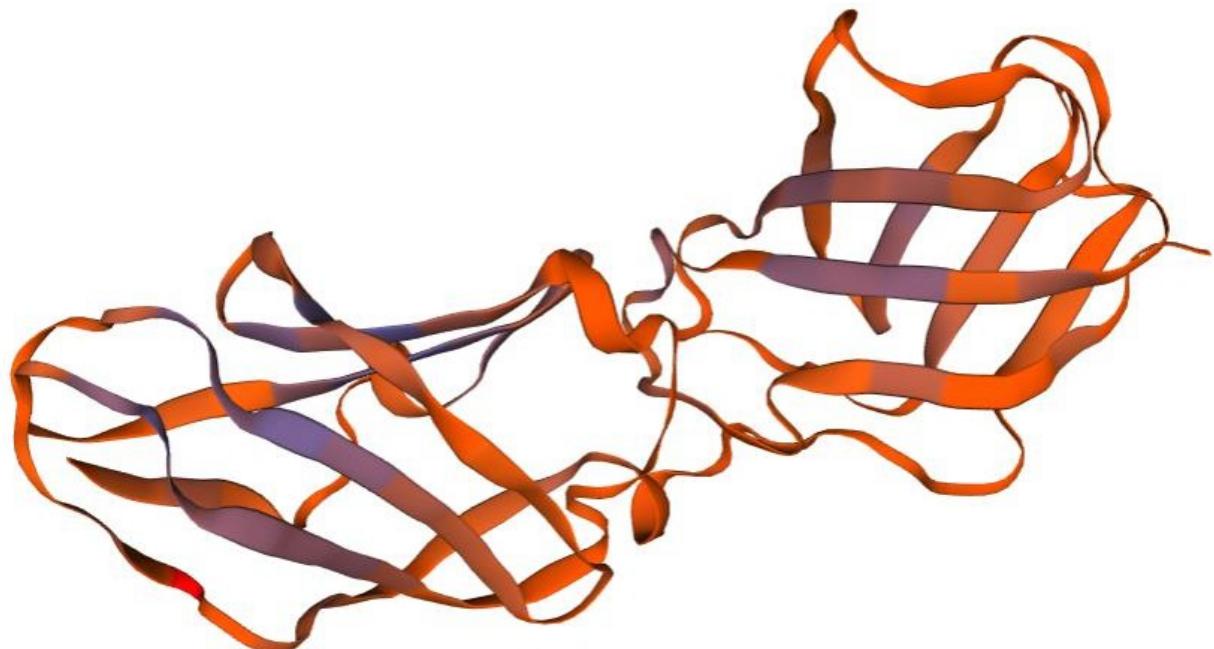
- Sequence(s)
- Target-Template Alignment
- User Template
- DeepView Project

By using the SWISS-MODEL server, you agree to comply with the following [terms of use](#) and to cite the corresponding [articles](#).

# STRUCTURE PREDICTION



# STRUCTURE OF THE PROTEIN

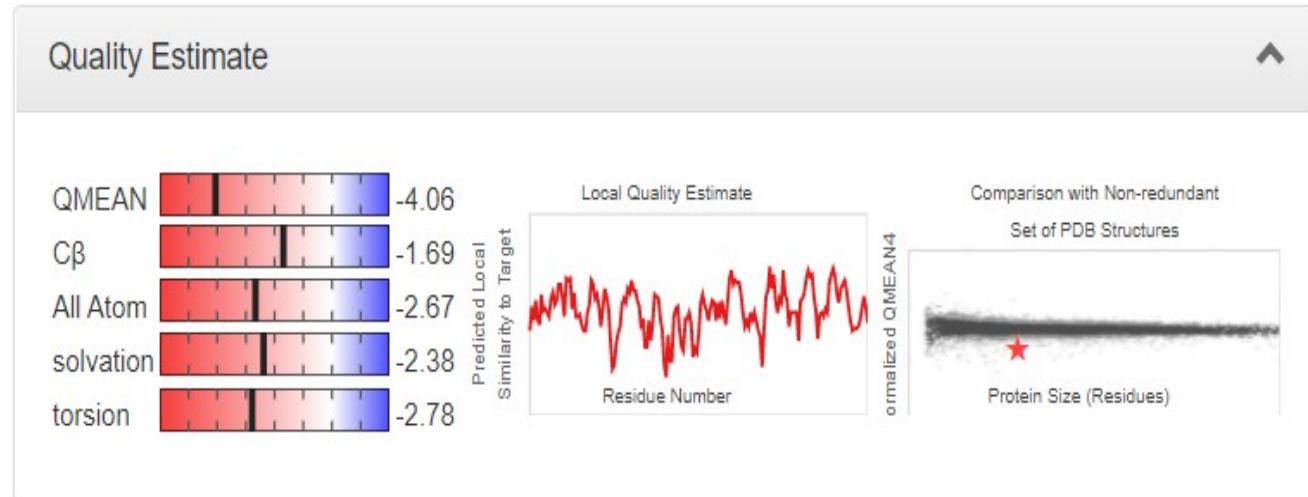


# VISUALIZATION OF PROTEIN

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 R1031 ASP R1112 THR R1087 VAI R1144 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



# TEMPLATE PROTEIN



SWISS-MODEL

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

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

Coordinates PDB Format

Method X-RAY DIFFRACTION 10.00 Å

Oligo State homo-dimer

Ligands None

Links RCSB PDBe PDBe-KB PDBsum CATH

Citation Kong, Y. et al., Structural Basis for Plexin Activation and Regulation. Neuron (2016)

Release Date 2016-07-06

Peptides Plexin-A2: A B

**SMTL:PDB** SMTL Chain Id: A B  
PDB Chain Id: A C

Plexin-A2

Segres	ETGMPQYSSTPHSENIRDWTFNHLTVHRRRTGAVVVGAINRVY	40
5l5g.1.A	POYSSTPHSENIRDWTFNHLTVHRRRTGAVVVGAINRVY	40
5l5g.1.B	POYSSTPHSENIRDWTFNHLTVHRRRTGAVVVGAINRVY	40
Segres	KLTGNLTIQVWAHKTGPEEDNKACYPPPLIVQFCSEVLTLTN	80
5l5g.1.A	KDGNLTIQVWAHKTGPEEDNKACYPPPLIVQFCSEVLTLTN	80
5l5g.1.B	KDGNLTIQVWAHKTGPEEDNKACYPPPLIVQFCSEVLTLTN	80

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

RCSB PDB Deposit Search Visualize An

**PDB** 177009 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

PDB-101 CPDB EMDDataResource Worldwide Protein Data Bank Foundation

Enter search term(s) Advanced Search | Browse Annotations Help Celebrating YEARS OF Protein Data Bank

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

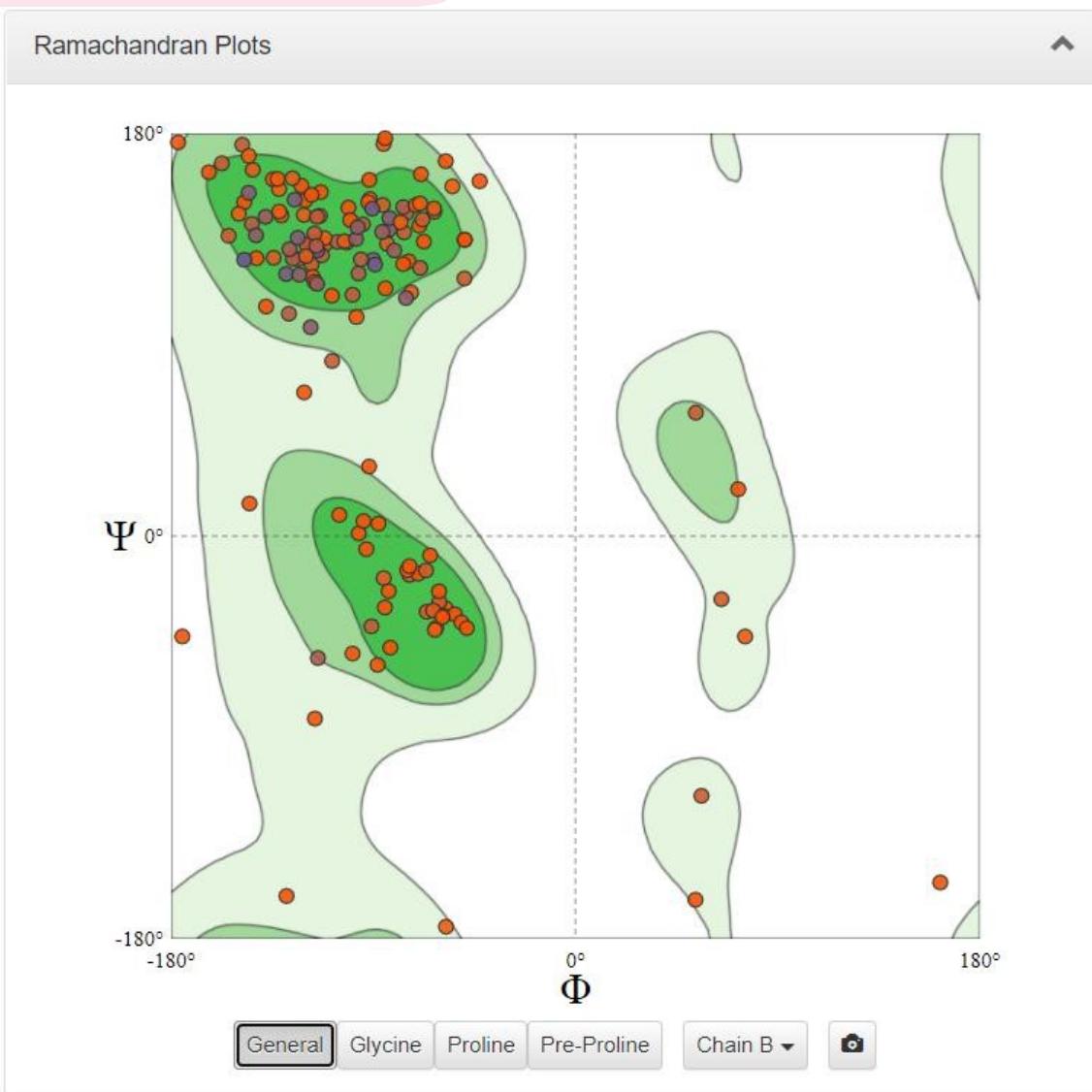
Deposited: 2016-05-28 Released: 2016-07-06  
Deposition Author(s): Janssen, B.J.C., Kong, Y., Malinauskas, T., Vangoor, V.R., Coles, C.H., Kaufmann, R., Ni, T., Gilbert, R.J.C., Padilla-Parra, S., Pasterkamp, R.J., Jones, E.Y.

Experimental Data Snapshot wwPDB Validation   
Method: X-RAY DIFFRACTION Metric Percentile Ranks Value

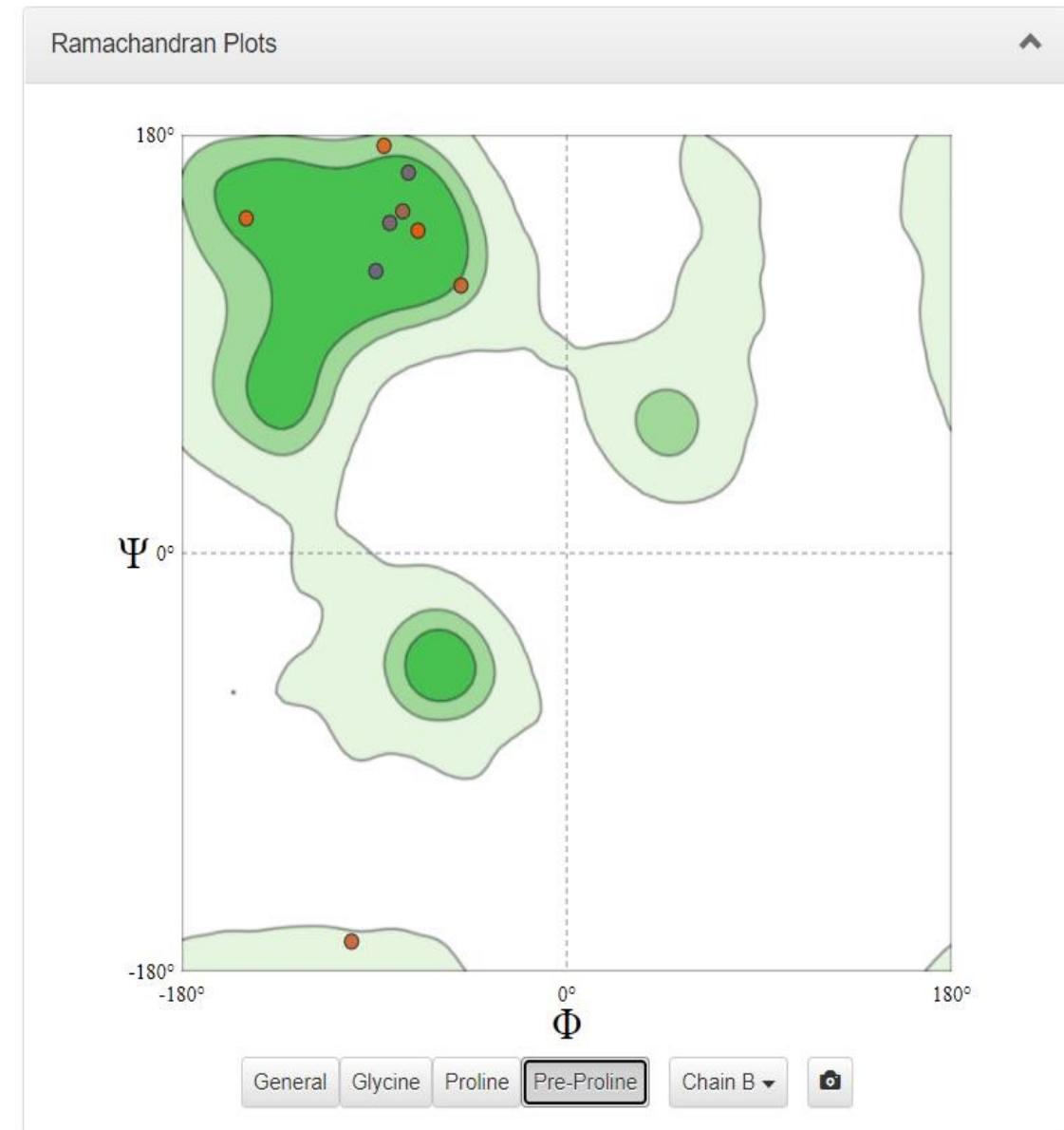
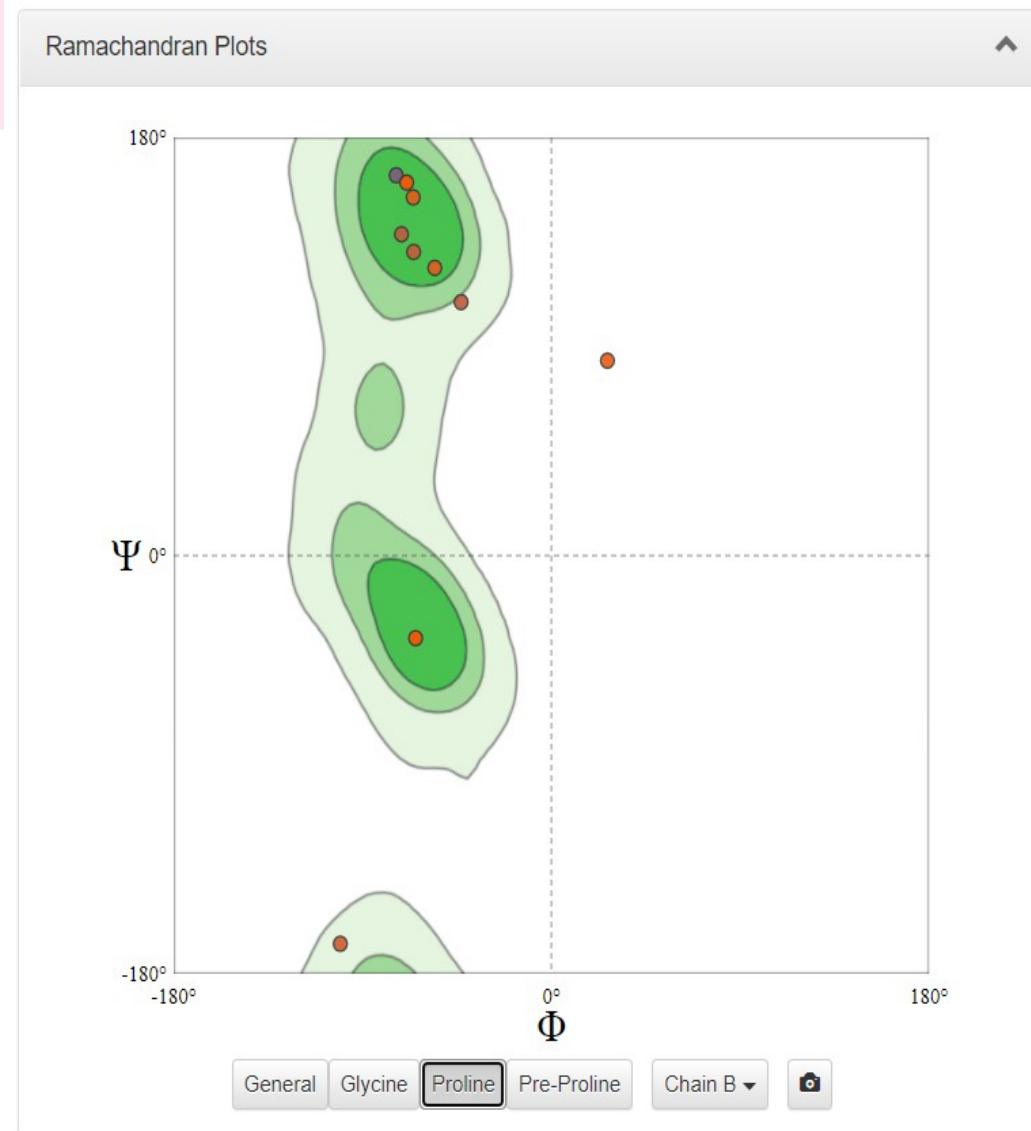
3D Report Full Report

Connect Us

# VALIDATION OF THE STRUCTURE



# VALIDATION OF THE STRUCTURE



# SUMMARY OF HOMOLGY MODELLING

For homology modelling, I 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. <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. <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, I pasted the sequence of our target protein, i.e, P08F94 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 descendancy 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.

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, I 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 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.

# SUMMARY



**THANK YOU !**