

Find a Gene Project

[Q1]

Protein Name: Sonic hedgehog protein isoform 1 preproprotein (SHH)

Accession: NP_000184.1

Species: Homo Sapiens

Known Function: chemical signal to control embryonic development

[Q2]

Method: NIH TBLASTN (2.12.0) search against nematode ESTs

Database: Expressed Sequence Tags (est)

Organism: All species

BLAST search output:

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

BLAST® » tblastn Home Recent Results

blastn blastp blastx **tblastn** tblastx Translated BLAST: tblastn

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) Query subrange [?](#)

NP_000184.1 From
To

Or, upload file no file selected [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

Choose Search Set

Database [+](#) Expressed sequence tags (est) [?](#)

Organism [Optional](#) ☐ exclude [Add organism](#)

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

☐ Sequences from type material

Limit to [Optional](#) [YouTube](#) [Create custom database](#)

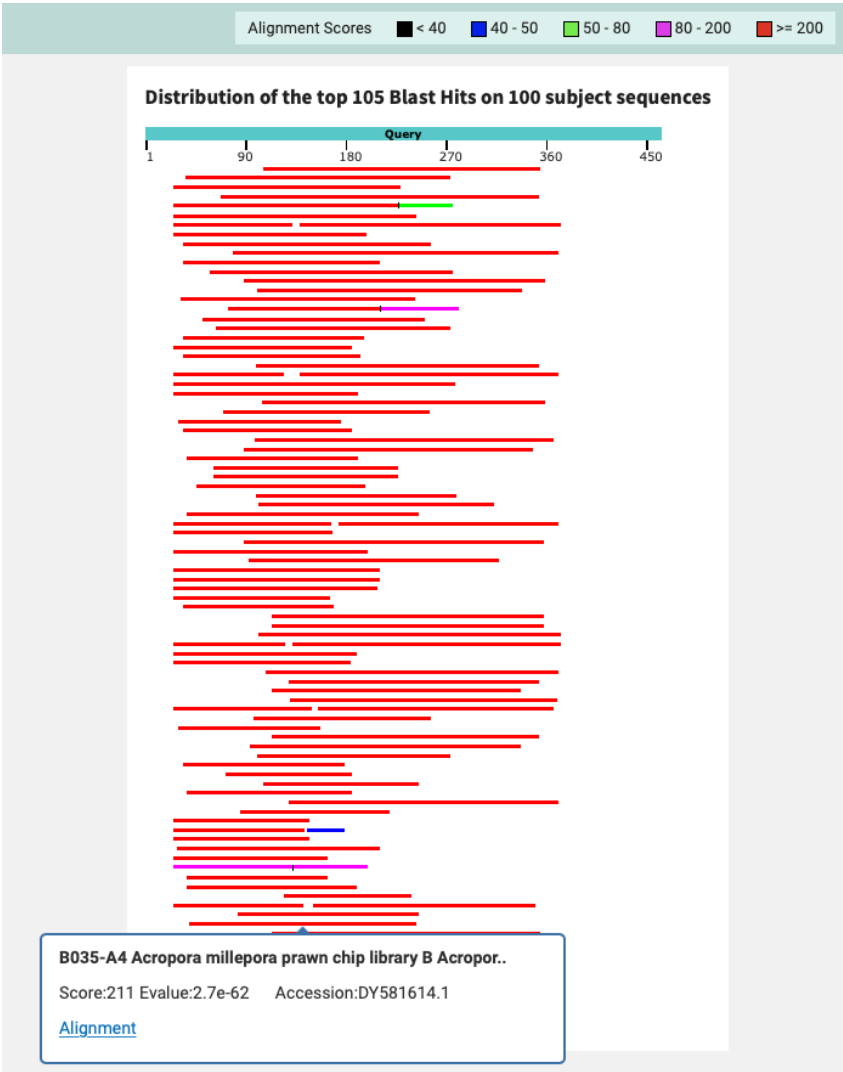
Entrez Query [Optional](#)

[BLAST](#) Search database est using Tblastn (search translated nucleotide databases using a protein query)

☐ Show results in a new window

Note: Parameter values that differ from the default are highlighted in yellow and marked with [+](#) sign

Chosen match: Accession DY581614.1 - a 859 base pair clone from *Acropora millepora* (a species of branching stony coral). See below for alignment details.



☒ B035-A4 *Acropora millepora* prawn chip library B *Acropora millepora* cDNA clone B035-A4, mRNA sequence [Acropora millepora](#) 211 211 44% 3e-62 53.14% 859 [DY581614.1](#)

B035-A4 *Acropora millepora* prawn chip library B *Acropora millepora* cDNA clone B035-A4, mRNA sequence
Sequence ID: [DY581614.1](#) Length: 859 Number of Matches: 1

Range 1: 223 to 843		GenBank	Graphics				
				▼ Next Match	▲ Previous Match		
Score	Expect	Method	Identities	Positives	Gaps	Frame	
211 bits(538)	3e-62	Compositional matrix adjust.	110/207(53%)	136/207(65%)	3/207(1%)	+1	
Query 40	TPLAYKQFIPNVAEKT	LGASGRYEGKISRNSERFKELTPNYPDIIFKDEENTGADRLMT	99				
	+PL Q +P+++E + GASG +GKI+RNS F++L P YN IIFKDEE TGADRLM+						
Sbjct 223	SPLMLYQCVPDLS	SENSQGASGPAKGKITRNSPEFEKLEPCYNTAIIFKDEEGTGADRLMS	402				
Query 100	QRCKDKLNALAI	SVMNQWPGVKLRVTEGWDEEDGHSEESLHYEGRAVDITTSRDRS--K	157				
	+RCK+KL LA V NQWP +KL VTE WDE G HS+ SLHYEGRAVD+ SD +S K						
Sbjct 403	KRCKEKLIELAS	LVKNQWPSLKLVTTEAWDEQGHKNSLHYEGRAVDLRLSDTYKSNPK	582				
Query 158	YGMLARLAVEAG	FDWVYYESKAHIHCSVKAENSV-AAKSGGC	FPGSATVHLEQGGTKLVK	216			
	+L RLAV AGFD+V YESK HIH SV+ ++ V K GCF +TV LE G V						
Sbjct 583	LALLGRLAVNAG	FYVLYESKTHIASVREDSYVDKTRTGCFSSSESTVRLENGAVLRVD	762				
Query 217	DLSPGDRVLAAD	DQGRLLYSDFLTFLD	243				
	L DRV G +YS+ + F D						
Sbjct 763	HLKISDRVQVM	MQDGTIGYSEVIMFAD	843				

Alignment details:

B035-A4 Acropora millepora prawn chip library B Acropora millepora cDNA clone B035-A4, mRNA
sequence

Sequence ID: [DY581614.1](#) Length: 859 Number of Matches: 1

Score	Expect	Method	Identities	Positives	Gaps
211 bits(538)	3e-62	Compositional matrix adjust.	110/207(53%)	136/207(65%)	3/207(1%)
Query 40	TPLAYKQFIPNVAEKT LGASGRYEGKISRNSERFKELTPNYPDIIIFKDEENTGADRLMT 99				
	+PL Q +P+++E + GASG +GKI+RNS F++L P YN IIFKDEE TGADRLM+				
Sbjct 223	SPLMLYQCVPDLSSENSQGASGPAKGKITRNSPEFEKLEPCYNTAIIFKDEEGTGADRLMS 402				
Query 100	QRCKDKLNALAISVMNQWPGVKLRVTEGWDEDGHHSEESLHYEGRAVDITTSRDRS--K 157				
	+RCK+KL LA V NQWP +KL VTE WDE G HS+ SLHYEGRAVD+ SD +S K				
Sbjct 403	KRCKEKLIELASLVKNQWPSLKLVVTEAWDEQGQHSKNSLHYEGRAVDLRLSDTYKSNPK 582				
Query 158	YGMLARLAVEAGFDWVYYESKAHIHCSVKAENSV-AAKSGGCFPGSATVHLEQGGTKLVK 216				
	+L RLAV AGFD+V YESK HIH SV+ ++ V K GCF +TV LE G V				
Sbjct 583	LALLGRLAVNAGFDYVLYESKTHIHASVREDSYVDKTKRTGCFSSSESTVRLENGAVLRVD 762				
Query 217	DLSPGDRVLAADDQGRLLYSDFLTFLD 243				
	L DRV G + YS+ + F D				
Sbjct 763	HLKISDRVQVMMQDGTIGYSEVIMFAD 843				

[Q3]

Chosen sequence:

>A. millepora protein (sequence taken from BLAST result)
SPLMLYQCVPDLSSENSQGASGPAKGKITRNSPEFEKLEPCYNTAIIFKDEEGTGADRLMSKRCK
EKLIELASLVKNQWPSLKLVVTEAWDEQGQHSKNSLHYEGRAVDLRLSDTYKSNPKLALLGRLA
VNAGFDYVLYESKTHIHASVREDSYVDKTKRTGCFSSSESTVRLENGAVLRVDHLKISDRVQVMM
QDGTIGYSEVIMFAD

Name: A. millepora SHH


Species: Acropora millepora

Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Cnidaria; Anthozoa; Hexacorallia;
Scleractinia; Astrocoeniina; Acroporidae; Acropora

[Q4]

A BLASTP search against NR database (see setup in first screen-shot below) yielded a top hit result to a protein from *Acropora millepora* (a species of branching stony coral).

See additional screen shots below for top hits and selected alignment details:

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

BLAST® >> blastp suite

blastn **blastp** blastx tblastn tblastx **Standard Protein BLAST**

BLASTP programs search protein databases using a protein query. [more...](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#) [Clear](#)

>A. millepora protein (sequence taken from BLAST result)
SPLMLYQCVPDLSSENSQGASGPAKGKITRNSPEFEKLEPCYNTAIFKDEEGT
GADRLMSKRCKEKLIEL
ASLVKNQWPSLKLVTAEWDEQQHNSKNSLHYEGRAVDLRLSDTYKSNPKL
ALLGRLAVNAGFDYVLYES
KTHIHASVREDSYVDKTKRTGCFSSSESTVRLENGAVLRVDHLKISDRVQVMM
QDGTIGYSEVIMFAD

Query subrange [?](#)

From

To

Or, upload file no file selected [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

Choose Search Set

Database ?

Organism [Optional](#) ☐ exclude [Add organism](#)

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown. [?](#)

Exclude [Optional](#) ☐ Models (XM/XP) ☐ Non-redundant RefSeq proteins (WP) ☐ Uncultured/environmental sample sequences

Program Selection

Algorithm ☐ Quick BLASTP (Accelerated protein-protein BLAST)

☒ blastp (protein-protein BLAST)

☐ PSI-BLAST (Position-Specific Iterated BLAST)

☐ PHI-BLAST (Pattern Hit Initiated BLAST)

☐ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm [?](#)

BLAST Search **database nr** using **Blastp (protein-protein BLAST)**

☐ Show results in a new window

Score	Expect	Method	Identities	Positives	Gaps
430 bits(1106)	3e-149	Compositional matrix adjust.	205/207(99%)	206/207(99%)	0/207(0%)
Query 1		SPLMYLQCVPDLSSENSQGASGPAKGKITRNSPEFEKLEPCYNTAIIKFDEEGTGADRLMS			60
Sbjct 40		SPMLY QYCVPDLSSENSQGASGPAKGKITRNSPEFEKLEPCYNTAIIKFDEEGTGADRLMS			99
Query 61		KRCKEKLIELASLVKNQWPSLKLVTWEADDEQGHSKNSLHYEGRVLDRLSDTYKSNPK			12
Sbjct 100		KRCKEKLIELASLVKNQWPSLKLVTWEADDEQGHSKNSLHYEGRVLDRLSDTYKSNPK			15
Query 121		LALLGLRAVNAAGFDYLVYESKTHIASVREDSVYDKTKRTGCFSSSETVRLENGAVLRVD			18
Sbjct 160		LALLGLRAVNAAGFDYLVYESKTHIASVREDSVYDKTKRTGCFSSSETVRLENGAVLRVD			21
Query 181		HLKISDRVQVMMDGTGTGYSEVIMFAD	287		
Sbjct 220		HLKISDRVQVMMDGTGTGYSEVIMFAD	246		

[Q5]

Re-labeled sequences for alignment:

>Human_SHH | NP_000184.1 | sonic hedgehog protein isoform 1 preproprotein
[Homo sapiens]

MLLLARCLLLVSVSLLVCSGLACGPGRGFGKRRHPKKLTPLAYKQFIPNVAEKT LGASGRYEGKISRNS
ERFKELTPNYPDIIFKDEENTGADRLMTQRCKDKLNALAI SVMNQWPGVKLRVTEGWDEDGHHSEESLH
YEGRAVDITTSDDRDRSKYGMLARLAVEAGFDWVYYESKAHIHCSVKAENSVAAKSGGCFPGSATVHLEQG
GTKLVKDLSPGDRVLAADDQGRLLYSDFLTFLDRDDGAKKVFIYVIETREPRERLLLTAHLLFVAPHNDS
ATGEPEASSSGSGPPSGGALGPRALFASRVRPQGQVYVVAERDGDRLLPAAVHSVTLSEEAAGAYAPLTA
QGTILINRVLASCYAVIEHSHWAHRAFPFRLAHALLAALAPARTDRGGDSGGGDRGGGGGRVALTAPGA
ADAPGAGATAGIHWYSQLLYQIGTWLLDSEALHPLGMAVKSS

> A_millepora_protein (sequence taken from BLAST result) | XP_015756004.1:40-
246 PREDICTED | tiggy-winkle hedgehog protein-like [Acropora digitifera]
SPLMLYQCVPLSENSQGASGPAKGKITRNSPEFEKLEPCYNTAIIFKDEEGTGADRLMSKRCKEKLIEL
ASLVKNQWPSLKLVTAEWDEQGQHSKNSLHYEGRVDLRLSDTYKSNPKLALLGRLAVNAGFDYVLYES
ETHIHASVREDSYVDKTKRTGCFSSSESTVRLENGAVSRVDHLKISDRVQVMMQDGTIGYSEVIMFAD

>Yellow_Fever_Mosquito | XP_001657979.2:58-260 | protein hedgehog [Aedes
aegypti]

PLVFKQHVPNVSENSLGASGMQEGPISRNSDKFRNLETNYNKDIIFKDEEGTGADRVMTQRCKEKLNILA
VSVMNQWPGRLRLMVTEGWDEDDHMHARESLHYEGRVDIMTSDKDRSKIGMLARLAVEAGFDWVYYESRS
IHCSVKSDSSQSNHASGCFTGDSTVQTSTGETRKLSELQIGEQLSVDSGNTVYSEVIMFMD

>Wolf_Eel | XP_031705848.1:40-243 | sonic hedgehog protein [Anarrhichthys
ocellatus]

TPLAYKQFIPNVAEKT LGASGRYEGKITRNSERFKELTPNYNTDIIFKDEENTGADRLMTQRCKDKLNLSL
AISVMNQWPGVKLRVTEGWDEDDGHHFEESLHYEGRVDITTSDDRDRSKYGTLSRLAVEAGFDWVYYESKA
HIHCSVKAENSVAAKSGGCFPGSSVTTLQDGTAKVHKLQTDGDRVLAADDDGKPIYTDIFIMFID

>Oak_Gall_Wasp | XP_033210718.1:66-269 | sonic hedgehog protein A isoform X1
[Belonocnema kinseyi]

TPLVFKQHVPNVSENTLPASGLSEGRVSRHDSRFRNLVPNYNTDIIFKDEEGTGADRLMTQRCKEKLNTL
AISVMNQWPGVKLRVTEGWDEEGKHAIDSLHYEGRVDVTTSDRDRSKYGMLARLAVEAGFDWVYYESRS
HIHCSVKSESSSAGKSGGCFPGKSIVRTEKGARRRLDELHIGERIAALDSRGDIVYSEVISFLD

>Florida_Carpenter_Ant | XP_011260776.1:65-268 |sonic hedgehog protein A
[Camponotus floridanus]

TPLVFKQHVPNVSENTLPASGLGEGRISRDDSRFRDLVPNYNSDIVFKDEEGTGADRLMTQRCKEKLNAL
AISVMNQWPGIKLRVTEGWDEEGKHAVIDSLHYEGRVDITTSDDRDRSKYGMLARLAVEAGFDWVYYESRS
HIHCSVKSESSSTGKSGGCFPAKSLIRTEHGGTKRLDEVHLGERIAALNSHGDIVYSEVIAFLD

Multiple Sequence Alignment:
Obtained using MUSCLE (version 3.8) at EBI:

CLUSTAL multiple sequence alignment by MUSCLE (3.8)

A_millepora_protein
Human_SHH
Wolf_Eel
Yellow_Fever_Mosquito
Oak_Gall_Wasp
Florida_Carpenter_Ant

-----SPLMLYQCV~~PDL~~SENSQ~~GA~~SG
MLLLA~~RC~~LLLLV~~SS~~LLVCSGLACG~~PGR~~GFGKRRHP~~PK~~LTPLAYKQFIPNVA~~EK~~LTGASG
-----TPLAYKQFIPNVA~~EK~~LTGASG
-----PLVFKQ~~HVP~~NVNS~~SL~~LGASG
-----TPLVFKQ~~HVP~~NVNS~~SL~~LPASG
-----TPLVFKQ~~HVP~~NVNS~~SL~~LPASG

** * * : : : : * : : *

[illegible][illegible]

```
A_millepora_protein      THIHASVREDSDYVDKTKRTGCFSEESTVRLENGAVSRVDHLKISDRVQVMMDGTIGYSE
Human_SHH                AHIHCSVKAEANSVA-AKSGGCGFPGSATVHLEQGQTCLKVKDLSFGDRVLAAADGGCRLLYSYD
Wolf_Eel                 AHIHCSVKAEANSVA-AKSGGCGFPGSSTVLDQGETAKVKHLQTDGRVLAAADDGCKPIITYD
Yellow_Fever_Mosquito    SHIHCSVKSDSSQS-NHASGCTSLTDSVTQTSTGTETKGLSELQIQEQLVSDSSNGNTVSYSE
Oak_Gall_Wasp            SHIHCSVKSESSSA-GKSGGCGFPKGSIIVRTEKGARRRLDELHIGERIAALDSRGDDIVYSE
Florida_Carpenter_Ant     SHIHCSVKSESSST-GKSGGCGFPAKSLRIETHGGTKRLDEVHLGERIAALNSHGDIVYSE
:::*,*:,*:::*::**::***::**::**::**::*
```

```

A_millepora_protein      VIMFAD
Human_SHH                FLTFLDRDDGAKKVFYVIEETREPRERLLLTAHHLLFVAPHNDSATGPEASSGSGPPSGG
Wolf_Eel                 FIMFID
Yellow_Fever_Mosquito    VIMFMD
Oak_Gall_Wasp            VISFLD
Florida_Carpenter_Ant    VIAFLD
.: * *

```

A_millepora_protein
Human_SHH
Wolf_Eel
Yellow_Fever_Mosquito
Oak_Gall_Wasp
Florida_Carpenter_Ant

A_millepora_protein
Human_SHH
Wolf_Eel
Yellow_Fever_Mosquito
Oak_Gall_Wasp
Florida_Carpenter_Ant

A_millepora_protein
Human_SHH
Wolf_Eel
Yellow_Fever_Mosquito
Oak_Gall_Wasp
Florida_Carpenter_Ant

[Q6]

Neighbor-joining Phylogenetic Tree:
Obtained using MUSCLE (version 3.8) at EBI:

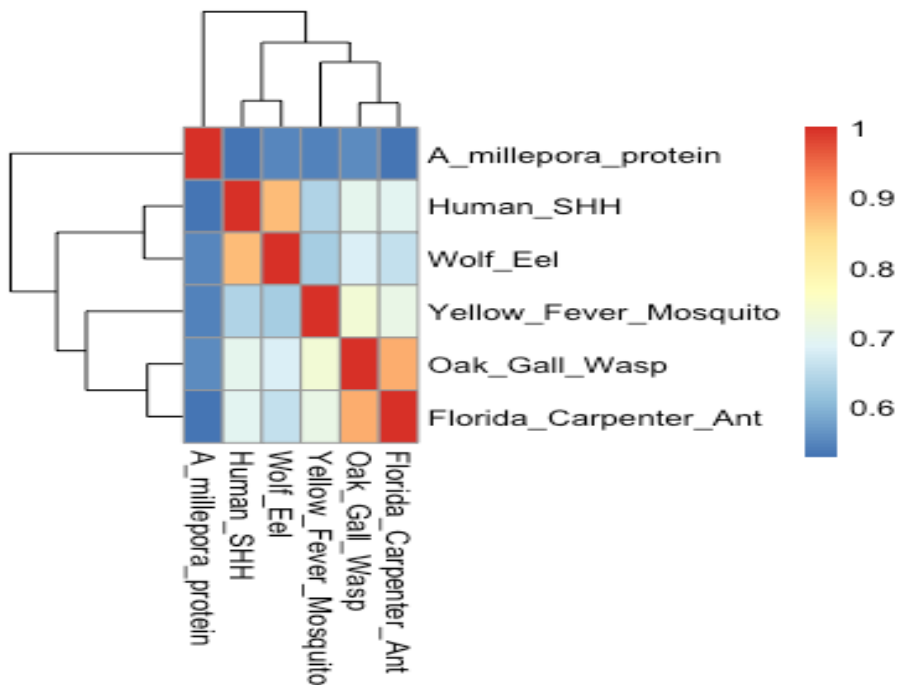
Phylogenetic Tree

This is a Neighbour-joining tree without distance corrections.



[Q7]

Heatmap generated using R package ‘pheatmap’:



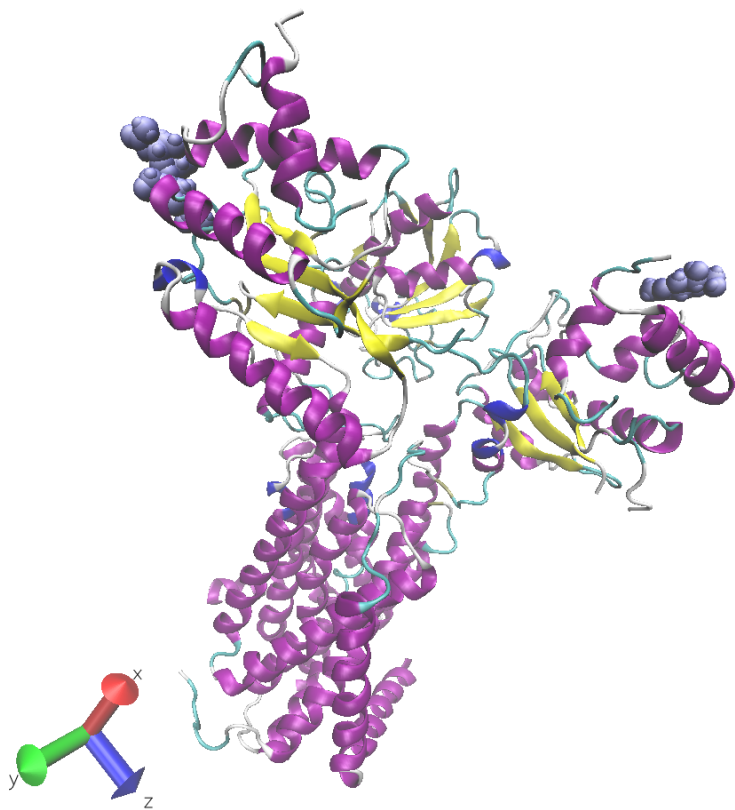
[Q8]

Top 3 unique hits of similar atomic resolution structures to the novel sequence

ID	Technique	Resolution	Source	Evalue	Identity
6TD6	Electron Microscopy	4.76 Å	Drosophila melanogaster	2.493e-62	53%
7E2I	Electron Microscopy	4.07 Å	Homo sapiens	2.258e-61	53%
4C4N	X-ray Diffraction	2.36 Å	Mus musculus	3.233e-52	59%

[Q9]

Visualization of 6TD6 from Drosophila melanogaster



This protein structure is similar in structure to the novel A_millepora_protein given the sequence similarity of 53% (>40% pairwise sequence identity for similarity). In the figure above, alpha

helix secondary structures are in purple and beta sheet secondary structures are in yellow.
There is no water molecule, and non-protein structures are in VDW shape.

[Q10]

There are 8 targets found from performing BLAST search of ChEMBL with the novel sequence.

One target from Homo sapiens (ID: ChEMBL5602) reported 21 binding assays, 4 functional assays and a ligand efficiency plot.

https://www.ebi.ac.uk/chembl/target_report_card/ChEMBL5602/

One of the binding assays linked to an article that discusses the Hedgehog-glioma-associated oncogene homologue zinc finger protein ("Hh-Gli") signaling pathway, "which plays an important role in the embryonic patterning and development of many tissues and somatic structures". The article discusses about phase I clinical trials of the Hh pathway inhibition for several cancers.

Mahindroo N, Punchihewa C, Fujii N. Hedgehog-Gli Signaling Pathway Inhibitors as Anticancer Agents. *Journal of Medicinal Chemistry* **2009** 52 (13), 3829-3845. doi: 10.1021/jm801420y

<https://pubs.acs.org/doi/10.1021/jm801420y>