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**:

Graphical user interface, application

Description automatically generated

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

Chart, bar chart

Description automatically generated



**Text

Description automatically generated**

**Alignment details**:

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

Sequence ID: [DY581614.1](https://www.ncbi.nlm.nih.gov/nucleotide/DY581614.1?report=genbank&log$=nuclalign&blast_rank=85&RID=ZSEK6YMY01R) Length: 859 Number of Matches: 1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Alignment statistics for match #1 | | | | | | |
| **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 TPLAYKQFIPNVAEKTLGASGRYEGKISRNSERFKELTPNYNPDIIFKDEENTGADRLMT 99

+PL Q +P+++E + GASG +GKI+RNS F++L P YN IIFKDEE TGADRLM+

Sbjct 223 SPLMLYQCVPDLSENSQGASGPAKGKITRNSPEFEKLEPCYNTAIIFKDEEGTGADRLMS 402

Query 100 QRCKDKLNALAISVMNQWPGVKLRVTEGWDEDGHHSEESLHYEGRAVDITTSDRDRS--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 LALLGRLAVNAGFDYVLYESKTHIHASVREDSYVDKTKRTGCFSSESTVRLENGAVLRVD 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)** **SPLMLYQCVPDLSENSQGASGPAKGKITRNSPEFEKLEPCYNTAIIFKDEEGTGADRLMSKRCKEKLIELASLVKNQWPSLKLVVTEAWDEQGQHSKNSLHYEGRAVDLRLSDTYKSNPKLALLGRLAVNAGFDYVLYESKTHIHASVREDSYVDKTKRTGCFSSESTVRLENGAVLRVDHLKISDRVQVMMQDGTIGYSEVIMFAD**

**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:

Graphical user interface, text, application, Word, email

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

[**Q5**]

**Re-labeled sequences for alignment:**

>Human\_SHH | NP\_000184.1 | sonic hedgehog protein isoform 1 preproprotein [Homo sapiens]

MLLLARCLLLVLVSSLLVCSGLACGPGRGFGKRRHPKKLTPLAYKQFIPNVAEKTLGASGRYEGKISRNS

ERFKELTPNYNPDIIFKDEENTGADRLMTQRCKDKLNALAISVMNQWPGVKLRVTEGWDEDGHHSEESLH

YEGRAVDITTSDRDRSKYGMLARLAVEAGFDWVYYESKAHIHCSVKAENSVAAKSGGCFPGSATVHLEQG

GTKLVKDLSPGDRVLAADDQGRLLYSDFLTFLDRDDGAKKVFYVIETREPRERLLLTAAHLLFVAPHNDS

ATGEPEASSGSGPPSGGALGPRALFASRVRPGQRVYVVAERDGDRRLLPAAVHSVTLSEEAAGAYAPLTA

QGTILINRVLASCYAVIEEHSWAHRAFAPFRLAHALLAALAPARTDRGGDSGGGDRGGGGGRVALTAPGA

ADAPGAGATAGIHWYSQLLYQIGTWLLDSEALHPLGMAVKSS

**> A\_millepora\_protein (sequence taken from BLAST result)  | XP\_015756004.1:40-246 PREDICTED | tiggy-winkle hedgehog protein-like [Acropora digitifera]**

**SPLMLYQCVPDLSENSQGASGPAKGKITRNSPEFEKLEPCYNTAIIFKDEEGTGADRLMSKRCKEKLIEL**

**ASLVKNQWPSLKLVVTEAWDEQGQHSKNSLHYEGRAVDLRLSDTYKSNPKLALLGRLAVNAGFDYVLYES**

**ETHIHASVREDSYVDKTKRTGCFSSESTVRLENGAVSRVDHLKISDRVQVMMQDGTIGYSEVIMFAD**

>Yellow\_Fever\_Mosquito | XP\_001657979.2:58-260 | protein hedgehog [Aedes aegypti]

PLVFKQHVPNVSENSLGASGMQEGPISRNDSKFRNLETNYNKDIIFKDEEGTGADRVMTQRCKEKLNILA

VSVMNQWPGLRLMVTEGWDEDHMHARESLHYEGRAVDIMTSDKDRSKIGMLARLAVEAGFDWVYYESRSH

IHCSVKSDSSQSNHASGCFTGDSTVQTSTGETRKLSELQIGEQVLSVDSSGNTVYSEVIMFMD

>Wolf\_Eel | XP\_031705848.1:40-243 | sonic hedgehog protein [Anarrhichthys ocellatus]

TPLAYKQFIPNVAEKTLGASGRYEGKITRNSERFKELTPNYNTDIIFKDEENTGADRLMTQRCKDKLNSL

AISVMNQWPGVKLRVTEGWDEDGHHFEESLHYEGRAVDITTSDRDKSKYGTLSRLAVEAGFDWVYYESKA

HIHCSVKAENSVAAKSGGCFPGSSTVTLQDGTEKAVKHLQTGDRVLAADDDGKPIYTDFIMFID

>Oak\_Gall\_Wasp | XP\_033210718.1:66-269 | sonic hedgehog protein A isoform X1 [Belonocnema kinseyi]

TPLVFKQHVPNVSENTLPASGLSEGRVSRHDSRFRNLVPNYNTDIIFKDEEGTGADRLMTQRCKEKLNTL

AISVMNQWPGVKLRVTEGWDEEGKHAIDSLHYEGRAVDVTTSDRDRSKYGMLARLAVEAGFDWVYYESRS

HIHCSVKSESSSAGKSGGCFPGKSIVRTEKGARRRLDELHIGERIAALDSRGDIVYSEVISFLD

>Florida\_Carpenter\_Ant | XP\_011260776.1:65-268|sonic hedgehog protein A [Camponotus floridanus]

TPLVFKQHVPNVSENTLPASGLGEGRISRDDSRFRDLVPNYNSDIVFKDEEGTGADRLMTQRCKEKLNAL

AISVMNQWPGIKLRVTEGWDEEGKHAVDSLHYEGRAVDITTSDRDRSKYGMLARLAVEAGFDWVYYESRS

HIHCSVKSESSSTGKSGGCFPAKSLIRTEHGGTKRLDEVHLGERIAALNSHGDIVYSEVIAFLD

**Multiple Sequence Alignment:**

**Obtained using MUSCLE (version 3.8) at EBI:**



Table

Description automatically generated

[**Q6**]

**Neighbor-joining Phylogenetic Tree:**

**Obtained using MUSCLE (version 3.8) at EBI:**

Table

Description automatically generated with low confidence

[**Q7**]

**Heatmap generated using R package ‘pheatmap’:**

Chart, box and whisker chart

Description automatically generated

[**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

Map

Description automatically generated with medium confidence

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 ChEMBEL 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>