***BGGN-213: FOUNDATIONS OF BIOINFORMATICS***

***Find-a-gene project assignment!***

Name: Blair Chang

UCSD email: [yac046@ucsd.edu](mailto:yac046@ucsd.edu)

PID: A59000602

**Q1.** Tell me the name of a protein you are interested in. Include the species and the accession number. This can be a human protein or a protein from any other species as long as its function is known.

1. Protein name:   
   RAS guanyl-releasing protein 1 (RASGRP1) isoform b
2. Species:   
   Home sapiens (taxid:9606)
3. Accession number:   
   NP\_001122074.1
4. Function known:   
   RASGRP1 is guanine nucleotide exchange factor (GEF) that activates RAS small GTPase by converting GDP to GTP. It mediates MAPK signaling in T cells and is important for T cell activation and proliferation.

**Q2**. Perform a BLAST search against a DNA database, such as a database consisting of genomic DNA or ESTs. The BLAST server can be at NCBI or elsewhere. Include details of the BLAST method used, database searched, and any limits applied (e.g. Organism). Also include the output of that BLAST search in your document. If appropriate, change the font to Courier size 10 so that the results are displayed neatly. You can also screen capture a BLAST output.

[Search Results]

1. Blast method:   
   TBLASTN

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Description | Scientific Name | Max Score | Query Cover | E-value | Per. Ident | Accession |
| 1  Picked | Egretta garzetta isolate BGI\_Z169 unplaced genomic scaffold, ASM68718v1 scaffold2 | Egretta garzetta | 145 | 89% | 1e-33 | 57.26% | NW\_009258894.1 |
| 2 | Egretta garzetta isolate BGI\_Z169 unplaced genomic scaffold, ASM68718v1 scaffold54 | Egretta garzetta | 94.4 | 41% | 7e-18 | 53.57% | NW\_009259313.1 |
| 3 | Egretta garzetta isolate BGI\_Z169 unplaced genomic scaffold, ASM68718v1  scaffold621 | Egretta garzetta | 70.9 | 6% | 9e-11 | 49.02% | NW\_009258687.1 |
| 4 | Egretta garzetta isolate BGI\_Z169 unplaced genomic scaffold, ASM68718v1  scaffold287 | Egretta garzetta | 68.6 | 6% | 5e-10 | 49.02% | NW\_009260435.1 |
| 5 | Egretta garzetta isolate BGI\_Z169 unplaced genomic scaffold, ASM68718v1  scaffold59 | Egretta garzetta | 67.0 | 7% | 2e-09 | 45.76% | NW\_009267230.1 |

1. Search output list (top 5 hits):
2. Database searched:   
   refseq\_genomes
3. Limits applied:   
   Organism limited to Egretta (taxid:56073)
4. Alignment of choice:   
   Egretta garzetta isolate BGI\_Z169 unplaced genomic scaffold, ASM68718v1 scaffold2

A screenshot of a computer

Description automatically generated

1. E-value and other alignment stats:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| E-value | Score | Method | Identities | Positives | Gaps |
| 1e-33 | 145 bits (365) | Compositional matrix adjust | 79/124(64%) | 75% | 0% |
|  |  |  |  |  |  |

**Q3.** Gather information about this “novel” protein. At a minimum, show me the protein sequence of the “novel” protein as displayed in your BLAST results from [Q2] as FASTA format (you can copy and paste the aligned sequence subject lines from your BLAST result page if necessary) or translate your novel DNA sequence using a tool called EMBOSS Transeq at the EBI.

A screenshot of a computer

Description automatically generated

1. Protein sequence of choice matches Subject above:  
   VAQKLHQLQNFNTLMAVIGGLCHSSISRLKETSSCVPHDVIKVFNEMTELLSSYRNYDSYRRAYNECSNFKIPILGVHLKDLISLYEGMPDYLEDKKINVYKLYSLYNHIDELIQLQEMPLPLEANMDLVHLLTLSLDLYYTEDEIYELSYAREPRSHRAAPMTPSKPPVVADWASGVAPKPDPKTISKHVQRMVDSVFKNYDHDQDGYISQEEFEKIAASFPFSFCVMAKDWEG
2. Name in header:

>AGX29484.1 RAS guanyl releasing protein 1, partial [Serinus canaria]

1. Species:   
   Egretta garzetta

**Q4.** Prove that this gene, and its corresponding protein, are novel. For the purposes of this project, “novel” is defined as follows. Take the protein sequence (your answer to [Q3]) and use it as a query in a blastp search of the nr database at NCBI.

1. Blastp output list with identities and Evalue:  
   A screenshot of a computer

   Description automatically generated

A screenshot of a computer

Description automatically generated

1. Top alignment shown with alignment statistics:  
   A screenshot of a computer

   Description automatically generated
2. Results indicates a “novel” gene found:

My top match has 98.33% identity to a different species - Serinus canaria. Thus, this gene from NW\_009258894.1 and its corresponding protein are novel.

 “Hi! It’s me, again!”

**Q5.** Generate a multiple sequence alignment with your novel protein, your original query protein, and a group of other members of this family from different species. A typical number of proteins to use in a multiple sequence alignment for this assignment purpose is a minimum of 5 and a maximum of 20 - although the exact number is up to you. Include the multiple sequence alignment in your report. Use Courier font with a size appropriate to fit page width.

**Re-labeled sequences for my alignment:**

>HUMAN\_RASGRP1

MGTLGKAREAPRKPSHGCRAASKARLEAKPANSPFPSHPSLAHITQFRMMVSLGHLAKGASLDDLIDSCIQSFDADGNLCRSNQLLQVMLTMHRIVISSAELLQKVITLYKDALAKNSPGLCLKICYFVRYWITEFWVMFKMDASLTDTMEEFQELVKAKGEELHCRLIDTTQINARDWSRKLTQRIKSNTSKKRKVSLLFDHLEPEELSEHLTYLEFKSFRRISFSDYQNYLVNSCVKENPTMERSIALCNGISQWVQLMVLSRPTPQLRAEVFIKFIQVAQKLHQLQNFNTLMAVIGGLCHSSISRLKETSSHVPHEINKVLGEMTELLSSSRNYDNYRRAYGECTDFKIPILGVHLKDLISLYEAMPDYLEDGKVNVHKLLALYNHISELVQLQEVAPPLEANKDLVHLLTLSLDLYYTEDEIYELSYAREPRNHRAPPLTPSKPPVVVDWASGVSPKPDPKTISKHVQRMVDSVFKNYDHDQDGYISQEEFEKIAASFPFSFCVMDKDREGLISRDEITAYFMRASSIYSKLGLGFPHNFQETTYLKPTFCDNCAGFLWGVIKQGYRCKDCGMNCHKQCKDLVVFECKKRAKNPVAPTENNTSVGPVSNLCSLGAKDLLHAPEEGPFTFPNGEAVEHGEESKDRTIMLMGVSSQKISLRLKRAVAHKATQTESQPWIGSEGPSGPFVLSSPRKTAQDTLYVLPSPTSPCPSPVLVRKRAFVKWENKDSLIKSKEELRHLRLPTYQELEQEINTLKADNDALKIQLKYAQKKIESLQLEKSNHVLAQMEQGDCS

>Novel\_protein\_Serinus\_canaria

VAQKLHQLQNFNTLMAVIGGLCHSSISRLKETSSCVPHDVIKVFNEMTELLSSYRNYDSYRRAYNECSNFKIPILGVHLKDLISLYEGMPDYLEDKKINVYKLYSLYNHIDELIQLQEMPLPLEANMDLVHLLTLSLDLYYTEDEIYELSYAREPRSHRAAPMTPSKPPVVADWASGVAPKPDPKTISKHVQRMVDSVFKNYDHDQDGYISQEEFEKIAASFPFSFCVMAKDWEG

>Amazona\_guildingii

VFNEMTELLSSYRNYDSYRRAYNECSNFKIPILGVHLKDLISLYEGMPDYLEDKKINIYKLYSLYNHINELIQLQEMPLPLEANMDLVHLLTLSLDLYYTEDEIYELSYAREPRSHRAAPLTPSRPPVVADWASGVAPKPDPKTISKHVQRMVDSVFKNYDHDQDGYISQEEFEKIAASFPFSFCVMAKDWEGLISRDEITAYFMRASSIYSKLGLGFAHNFQETTYLRPTFCDNCAGFLWGVIKQGYRCKDCGMNCHKQCKDLVVIECKRRPKTSIPDSSPTSALASSLCPVGVKEQFHGQEEGLFTFPNGEVVEHSEDSKDRTIMLMGSSAQKISVRLKPSVVHEGTQTDPVLLAGDVSRRQIEKKEHKMPENPYLQLAPPSPFPSPILGRKKAYVKWENKDSSQKKKEEHYSCKPSYQELEQERNILKAHNEGLKIQLEQAHKTIESLTIHRRNHVVDNLQHRDCS

>Neopelma\_chrysocephalum

MGTLGKRRENQQSAQACSTAPESALELKQTSHCPSLSNHTQVMMVPLGHLAKGATLEDLLETCIQSFDLEGNAYQNNQLLKIILAMHQFIISSADMLQKLIDLYLNALENNSSMLCVKICYFVRYWITEFWIMFKMDSKLSTTMEEFQELVRANGEELHCRLIDTSQINSRDWSRKLTQRVKANTSKKRKVSLLFDHLEPEELSDHLTYLEFKSFRRITFSDYQNYIVNSCVKENPTMERSISLCNGISQWVQLMVLSRPTPQLRAEVFIKFIHVAQKLHQLQNFNTLMAVIGGLCHSSISRLKETSSCVPHDVIKVFNEMTELLSSYRNYDSYRRAYNECSNFKIPILGVHLKDLISLYEGMPDYLEDKKINIYKLYSLYNHINELIQLQEMPLPLEANMDLVHLLTLSLDLYYTEDEIYELSYAREPRSHRAAPLTPCKPPVVADWASGVAPKPDPKTISKHVQRMVDSVFKNYDHDQDGYISQEEFEKIAASFPFSFCVMAKDWEGLISRDEITAYFMRASSIYSKLGLGFAHNFQETTYLRPTFCDNCAGFLWGVIKQGYRCKDCGMNCHKQCKDLVVIECKRRPKTSVADSSPTSALASSLCPVGVKEQFHGKKRSH

>Theropithecus\_gelada

GAAAGTCTCAAACCAGTTATTCACCTGAGCCACCTGGATGAACTTGATGAAGACTTCTGCTCGGAGCTGCGGGGTGGGGCGGCTGAGAACCATCAGTTGTACCCACTGGGAGATACCGTTGCACAGAGCAATAGATCTCTCCATGGTGGGGTTTTCCTTCACACAGCTATTTACAAGGTAATTCTGATAATCAGA

>Sagittarius\_serpentarius

DLEGNAYQNNQLLKIILAMHQFIISSADMLQKLFSTYLNALENKSSALCVKICYFVRYWITEFWVMFKMDSKLSTTMEEFQELVKANGEELHCHLIDTTQINSRDWSRKLTQRVKANTSKKRKVSLLFDHLEPEELSDHLTYLEFKSFRRISFSDYQNYIVNSCVKENPTMERSIALCNGISQWVQLMVLSRPTPQLRAEVFIKFIHVAQKLHQLQNFNTLMAVIGGLCHSSISRLKETSSCVPHDVIKVFNEMTELLSSYRNYDSYRRAYNECSNFKIPILGVHLKDLISLYEGMPNYLEDKKINIYKLYSLYNHINELIQLQEMPLPLEANMDLVHLLTLSLDLYYTEDEIYELSYAREPRSHRAAPLTPSRPPVVADWASGVAPKPDPKTISKHVQRMVDSVFKNYDHDQDGYISQEEFEKIAASFPFSFCVMAKDW

>Macaca\_mulatta

TACAAATCCAGCACAGTTGTCACAAAAAGTGGGCTTCAGGTAGGTGGTCTCTTGGAAGTTGTGAGGAAAGCCCAGGCCCAGCTTGGAATAGATTGAGCTGGCTCGCATGAAGTAGGCTGTGATCTCATCCCTGCTGATGAGGCCTTCCCTGCCAGCAAATGACCAAGGCAAGGATGTGAGTATACG

**Multiple Sequence Alignment using MUSCLE at EMBL-EBI:**

CLUSTAL multiple sequence alignment by MUSCLE (3.8)

HUMAN\_RASGRP1 MGTLGKAREAPRKPSHGCRAASKARLEAKPANSPFPSHPSLAHITQFRMMVSLGHLAKGA

Amazona\_guildingii ------------------------------------------------------------

Sagittarius\_serpentarius ------------------------------------------------------------

Neopelma\_chrysocephalum MGTLGKRRENQQSAQACSTAPESALELKQTSHCPSLSNHTQV------MMVPLGHLAKGA

Novel\_protein\_Serinus\_canaria ------------------------------------------------------------

Theropithecus\_gelada ------------------------------------------------------------

Macaca\_mulatta ------------------------------------------------------------

HUMAN\_RASGRP1 SLDDLIDSCIQSFDADGNLCRSNQLLQVMLTMHRIVISSAELLQKVITLYKDALAKNSPG

Amazona\_guildingii ------------------------------------------------------------

Sagittarius\_serpentarius -------------DLEGNAYQNNQLLKIILAMHQFIISSADMLQKLFSTYLNALENKSSA

Neopelma\_chrysocephalum TLEDLLETCIQSFDLEGNAYQNNQLLKIILAMHQFIISSADMLQKLIDLYLNALENNSSM

Novel\_protein\_Serinus\_canaria ------------------------------------------------------------

Theropithecus\_gelada -------------------------------------------------------GAAAG

Macaca\_mulatta ------------------------------------------------------------

HUMAN\_RASGRP1 LCLKICYFVRYWITEFWVMFKMDASLTDTMEEFQELVKAKGEELHCRLIDTTQINARDWS

Amazona\_guildingii ------------------------------------------------------------

Sagittarius\_serpentarius LCVKICYFVRYWITEFWVMFKMDSKLSTTMEEFQELVKANGEELHCHLIDTTQINSRDWS

Neopelma\_chrysocephalum LCVKICYFVRYWITEFWIMFKMDSKLSTTMEEFQELVRANGEELHCRLIDTSQINSRDWS

Novel\_protein\_Serinus\_canaria ------------------------------------------------------------

Theropithecus\_gelada TC----------------------------------------------------------

Macaca\_mulatta ------------------------------------------------------------

HUMAN\_RASGRP1 RKLTQRIKSNTSKKRKVSLLFDHLEPEELSEHLTYLEFKSFRRISFSDYQNYLVNSCVKE

Amazona\_guildingii ------------------------------------------------------------

Sagittarius\_serpentarius RKLTQRVKANTSKKRKVSLLFDHLEPEELSDHLTYLEFKSFRRISFSDYQNYIVNSCVKE

Neopelma\_chrysocephalum RKLTQRVKANTSKKRKVSLLFDHLEPEELSDHLTYLEFKSFRRITFSDYQNYIVNSCVKE

Novel\_protein\_Serinus\_canaria ------------------------------------------------------------

Theropithecus\_gelada ------------------------------------------------------------

Macaca\_mulatta ------------------------------------------------------------

HUMAN\_RASGRP1 NPTMERSIALCNGISQWVQLMVLSRPTPQLRAEVFIKFIQVAQKLHQLQNFNTLMAVIGG

Amazona\_guildingii ------------------------------------------------------------

Sagittarius\_serpentarius NPTMERSIALCNGISQWVQLMVLSRPTPQLRAEVFIKFIHVAQKLHQLQNFNTLMAVIGG

Neopelma\_chrysocephalum NPTMERSISLCNGISQWVQLMVLSRPTPQLRAEVFIKFIHVAQKLHQLQNFNTLMAVIGG

Novel\_protein\_Serinus\_canaria ----------------------------------------VAQKLHQLQNFNTLMAVIGG

Theropithecus\_gelada ---------------------------------------------------TCAAACCAG

Macaca\_mulatta ------------------------------------------------------TACAAA

HUMAN\_RASGRP1 LCHSSISRLKETSSHVPHEINKVLGEMTELLSSSRNYDNYRRAYGECTDFKIPILGV--H

Amazona\_guildingii ----------------------VFNEMTELLSSYRNYDSYRRAYNECSNFKIPILGV--H

Sagittarius\_serpentarius LCHSSISRLKETSSCVPHDVIKVFNEMTELLSSYRNYDSYRRAYNECSNFKIPILGV--H

Neopelma\_chrysocephalum LCHSSISRLKETSSCVPHDVIKVFNEMTELLSSYRNYDSYRRAYNECSNFKIPILGV--H

Novel\_protein\_Serinus\_canaria LCHSSISRLKETSSCVPHDVIKVFNEMTELLSSYRNYDSYRRAYNECSNFKIPILGV--H

Theropithecus\_gelada TTATTCACCTGAGCCAC-----CTGGATGAA---------------CTTGATGAAGACTT

Macaca\_mulatta TCCAGCACAGTTGTCACAAAAAGTGGG-------------------CTTCAGGTAGG--T

. \*: \*

HUMAN\_RASGRP1 LKDLISLYEAMPDYLEDGKVNVHKLLALYNHISELVQLQEVAPPLEANKDLVHLLTLSLD

Amazona\_guildingii LKDLISLYEGMPDYLEDKKINIYKLYSLYNHINELIQLQEMPLPLEANMDLVHLLTLSLD

Sagittarius\_serpentarius LKDLISLYEGMPNYLEDKKINIYKLYSLYNHINELIQLQEMPLPLEANMDLVHLLTLSLD

Neopelma\_chrysocephalum LKDLISLYEGMPDYLEDKKINIYKLYSLYNHINELIQLQEMPLPLEANMDLVHLLTLSLD

Novel\_protein\_Serinus\_canaria LKDLISLYEGMPDYLEDKKINVYKLYSLYNHIDELIQLQEMPLPLEANMDLVHLLTLSLD

Theropithecus\_gelada CTGCTCGGAGCTGCGGGGT-----------------------------------------

Macaca\_mulatta GGTCTCTTGGAAGTTGTGA-----------------------------------------

. . ..

HUMAN\_RASGRP1 LYYTEDEIYELSYAREPRNHRAPPLTPSKPPVVVDWASGVSPKPDPKTISKHVQRMVDSV

Amazona\_guildingii LYYTEDEIYELSYAREPRSHRAAPLTPSRPPVVADWASGVAPKPDPKTISKHVQRMVDSV

Sagittarius\_serpentarius LYYTEDEIYELSYAREPRSHRAAPLTPSRPPVVADWASGVAPKPDPKTISKHVQRMVDSV

Neopelma\_chrysocephalum LYYTEDEIYELSYAREPRSHRAAPLTPCKPPVVADWASGVAPKPDPKTISKHVQRMVDSV

Novel\_protein\_Serinus\_canaria LYYTEDEIYELSYAREPRSHRAAPMTPSKPPVVADWASGVAPKPDPKTISKHVQRMVDSV

Theropithecus\_gelada -----------------GGGGCGGCTGAGAACCATCAGTTGTACCCACTGGGAGATACCG

Macaca\_mulatta ------------------GGAAAGCCCAGGCCCAGCTTGGAATAGATTGAGCTGGCTCGC

. . . . : .. . . .

HUMAN\_RASGRP1 FKNYDHDQDGYISQEEFEKIAASFPFSFCVMDKDREGLISRDEITAYFMRASSIYSKLGL

Amazona\_guildingii FKNYDHDQDGYISQEEFEKIAASFPFSFCVMAKDWEGLISRDEITAYFMRASSIYSKLGL

Sagittarius\_serpentarius FKNYDHDQDGYISQEEFEKIAASFPFSFCVMAKDW-------------------------

Neopelma\_chrysocephalum FKNYDHDQDGYISQEEFEKIAASFPFSFCVMAKDWEGLISRDEITAYFMRASSIYSKLGL

Novel\_protein\_Serinus\_canaria FKNYDHDQDGYISQEEFEKIAASFPFSFCVMAKDWEG-----------------------

Theropithecus\_gelada TTGCACAGAGCAATAG-----ATCTCTCCATGGTGGG-----------------------

Macaca\_mulatta ATGAAGTAGGCTGT-GATCTCATCCCTGCTGATGAGG-----------------------

.. \* . \*: : \*.

HUMAN\_RASGRP1 GFPHNFQETTYLKPTFCDNCAGFLWGVIKQGYRCKDCGMNCHKQCKDLVVFECKKRAKNP

Amazona\_guildingii GFAHNFQETTYLRPTFCDNCAGFLWGVIKQGYRCKDCGMNCHKQCKDLVVIECKRRPKTS

Sagittarius\_serpentarius ------------------------------------------------------------

Neopelma\_chrysocephalum GFAHNFQETTYLRPTFCDNCAGFLWGVIKQGYRCKDCGMNCHKQCKDLVVIECKRRPKTS

Novel\_protein\_Serinus\_canaria ------------------------------------------------------------

Theropithecus\_gelada -----------------------------------GTTTTCCTTCACACAGCTATTTACA

Macaca\_mulatta ---------------CCTTC----------------CCTGCCAGCAAATGACCAAGGCAA

HUMAN\_RASGRP1 VAPTENNTSVGPVSNLCSLGAKDLLHAPEEGPFTFPNGEAVEHGEESKDRTIMLMGVSSQ

Amazona\_guildingii IPDSSPTSALA--SSLCPVGVKEQFHGQEEGLFTFPNGEVVEHSEDSKDRTIMLMGSSAQ

Sagittarius\_serpentarius ------------------------------------------------------------

Neopelma\_chrysocephalum VADSSPTSALA--SSLCPVGVKEQFHGKKRSH----------------------------

Novel\_protein\_Serinus\_canaria ------------------------------------------------------------

Theropithecus\_gelada AGGTAATTCTGATAATCAGA----------------------------------------

Macaca\_mulatta GGATGTGAGTA-------------------------------------------------

HUMAN\_RASGRP1 KISLRLKRAVAHKATQTESQPWIGSEGPSGPFVLSSPRKTAQDTLYVLPSPTSPCPSPVL

Amazona\_guildingii KISVRLKPSVVHEGTQTDPVLLAGDVSRRQ---IEKKEHKMPENPYLQLAPPSPFPSPIL

Sagittarius\_serpentarius ------------------------------------------------------------

Neopelma\_chrysocephalum ------------------------------------------------------------

Novel\_protein\_Serinus\_canaria ------------------------------------------------------------

Theropithecus\_gelada ------------------------------------------------------------

Macaca\_mulatta ------------------------------------------------------------

HUMAN\_RASGRP1 VRKRAFVKWENKDSLIKSKEELRHLRLPTYQELEQEINTLKADNDALKIQLKYAQKKIES

Amazona\_guildingii GRKKAYVKWENKDSSQKKKEEHYSCK-PSYQELEQERNILKAHNEGLKIQLEQAHKTIES

Sagittarius\_serpentarius ------------------------------------------------------------

Neopelma\_chrysocephalum ------------------------------------------------------------

Novel\_protein\_Serinus\_canaria ------------------------------------------------------------

Theropithecus\_gelada ------------------------------------------------------------

Macaca\_mulatta ------------------------------------------------------------

HUMAN\_RASGRP1 LQLEKSNHVLAQMEQGDCS

Amazona\_guildingii LTIHRRNHVVDNLQHRDCS

Sagittarius\_serpentarius -------------------

Neopelma\_chrysocephalum -------------------

Novel\_protein\_Serinus\_canaria -------------------

Theropithecus\_gelada -------------------

Macaca\_mulatta ---------------TACG

**Q6.** Create a phylogenetic tree, using either a parsimony or distance-based approach. Bootstrapping and tree rooting are optional. Use “simple phylogeny” online from the EBI or any respected phylogeny program (such as MEGA, PAUP, or Phylip). Paste an image of your Cladogram or tree output in your report.

Using simple phylogeny from EBI to create a phylogenetic tree (distance-based approach).

**A diagram of a person's body

Description automatically generated with medium confidence**

**Q7.** Generate a sequence identity based heatmap of your aligned sequences using R. Making heatmap using R/Bio3D

1. Converting alignment file to fasta file by MEGA.
2. Drawing heatmap on R.

A screenshot of a computer

Description automatically generated

**Q8.** Using R/Bio3D (or an online blast server if you prefer), search the main protein structure database for the most similar atomic resolution structures to your aligned sequences.

Since my consensus sequence has lots of gap positions, I used my original sequence of Novel\_protein\_Serinus\_canaria for the search.

My top 3 unique hits are all from humans. They are RasGRP1, RasGRP4, RasGRP2. They are different proteins from RASGRP protein family.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ID | Method | Resolution | Source | E-value | Identity |
| 4L9M\_A | X-ray Diffraction | 3 Å | Homo sapiens | 1e-149 | 87.23% |
| [6AXG\_A](https://www.ncbi.nlm.nih.gov/protein/6AXG_A?report=genbank&log$=protalign&blast_rank=2&RID=R2U29S2V016) | X-ray Diffraction | 3.302 Å | Homo sapiens | 1e-58 | 59.02% |
| [6AXF\_A](https://www.ncbi.nlm.nih.gov/protein/6AXF_A?report=genbank&log$=protalign&blast_rank=3&RID=R2U29S2V016) | X-ray Diffraction | 3.1 Å | Homo sapiens | 3e-48 | 51.20% |

**Q9.** Using AlphaFold notebook generate a structural model using the default parameters for your novel protein sequence.

**A structure of a protein

Description automatically generated with medium confidence**

**Q10.** Perform a “Target” search of ChEMBEL ( https://www.ebi.ac.uk/chembl/ ) with your novel sequence. Are there any Target Associated Assays and ligand efficiency data reported that may be useful starting points for exploring potential inhibition of your novel protein? If there are no assays listed here simply list “non available as of [date]”.

# There are 8 Target Associated Assays for my novel protein.

# One is from an article published in [J Med Chem (2018) 61:6261-6276](http://dx.doi.org/10.1021%2Facs.jmedchem.8b00661): Activation of GFP-tagged RasGRP1 expressed in HEK293 cells assessed as ERK1/2 phosphorylation after 30 mins by immunoblot method. This paper proposed a ligand of Ras Guanine-Releasing Protein 3 (RasGRP3).

# Assay ID: CHEMBL4137150 (<https://www.ebi.ac.uk/chembl/assay_report_card/CHEMBL4137150/>)

# I think this ligand is promising since my interested protein (for Q1) is RasGRP1 in humans, which also belongs to RASGRP family. Thus, it’s likely that my novel protein has binding motif for this ligand.

# Yes, there’s ligand efficiency data in my report.

# [CHEMBL519741](https://www.ebi.ac.uk/chembl/compound_report_card/CHEMBL519741) shows the highest Binding Efficiency index (BEI).

# It is a small molecule that has a Molecular Weight of 413.51.

# A screenshot of a computer Description automatically generated

# A screenshot of a graph Description automatically generated

[CHEMBL519741](https://www.ebi.ac.uk/chembl/compound_report_card/CHEMBL519741)