**Q5.**

**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. 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. 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.**

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 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.**

**A structure of a protein

Description automatically generated with medium confidence**

**Q10.**

# 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 is RasGRP1 in humans, which is an isoform of RasGRP3. My novel protein may have the 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)