

Q5.

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

>HUMAN\_RASGRP1

MGTLGKAREAPRKPSHGCRAASKARLEAKPANSPPFSPHPSLAHITQFRMMVSLGHLAKGASLDDLI DSCIQSFDADGNLCR  
SNQLLQVMLTMHRIVISSAELLQKVITLYKDALAKNSPGLCLKICYFVRYWITEFWVMFKMDASLTDTMEEFQELVKAKGE  
ELHCRLIDTTQINARDWSRKLTQRIKSNTSKKRKVSLLFDHLEPEELSEHLTYLEFKSFRRISFSDYQNYLVNSCVKENPT  
MERSIALCNGISQWVQLMVLRSRPTPQLRAEVFIKFIQVAQKLHQLQNFNTLMAVIGGLCHSSISRLKETSSHPHEINKVL  
GEMTELLSSSRNYDNYRRAYGECTDFKIPILGVHLKDLISLYEAMPDYLEDGKVN VHKLLALYNHISELVQLQEVAPPLEA  
NKDLVHLLLTSLDLYYTEDEIYELSYAREPRNHRAPPLTPSKPPVVVDWASGVSPKDPKTI SKHVQRMVDSVFKNYDHDQ  
DGYISQEEFEKIAASFPPSFCVMDKDREGLISRDEITAYFMRASSIYSKLGLGFPHNFQETTYLKPTFCDCNAGFLWGVIK  
QGYRCKDCGMNCHKQCKDLVVFECKKRAKNPVAPTENNTSVGPVSNLCSLGAKDLLHAPEEGPFTFPNGEAVEHGEESKDR  
TIMLMGVSSQKISLRLKRAVAHKATQTESQPWIGSEGPSGFVLSSPRKTAQDTLYVLPSPSPVLRKRAVFKWEN  
KDSLKSKEELRHLRLPTYQELEQEINTLKADNDALKIQLKYAQKKIESLQLEKSNHVLAQMEQGDSCS

>Novel\_protein\_Serinus\_canaria

VAQKLHQLQNFNTLMAVIGGLCHSSISRLKETSSCVPHDVIKVFNEMTELLSSYRNYDSYRRAYNECSNFKIPILGVHLKD  
LISLYEGMPDYLEDKKINVKLYSLYNHIDELIQLQEMPLPLEANMDLVHLLLTSLDLYYTEDEIYELSYAREPRSHRAAP  
MTPSKPPVVADWASGVAPKDPKTI SKHVQRMVDSVFKNYDHDQDGYISQEEFEKIAASFPPSFCVMAKDWEGL

>Amazona\_guildingii

VFNEMTELLSSYRNYDSYRRAYNECSNFKIPILGVHLKDLISLYEGMPDYLEDKKINIYKLYSLYNHINELIQLQEMPLPL  
EANMDLVHLLLTSLDLYYTEDEIYELSYAREPRSHRAAPLTPSRPPVVADWASGVAPKDPKTI SKHVQRMVDSVFKNYDH  
DQDGYISQEEFEKIAASFPPSFCVMAKDWEGLISRDEITAYFMRASSIYSKLGLGFAHNFQETTYLRPTFCDCNAGFLWGV  
IKQGYRCKDCGMNCHKQCKDLVVIECKRRPKTSIPDSSPTSALASSLCPVGVKEQFHGQEEGLFTFPNGEVVEHSEDSKDR  
TIMLMGSSAQKISVRLKPSVVHEGTQTDPVLLAGDVSRRQIEKKEHKMPENPYLQLAPSPFPSPILGRKKAYVKWENKDS  
SQKKKEEHYSCKPSYQELEQERNILKAHNEGLKIQLEQAHTIESLTIHRRNHVVDNLQHRDCS

>Neopelma\_chrysocephalum

MGTLGKRRENQQSAQACSTAPESALELKQTSCHPSLSNHTQVMMVPLGHLAKGATLEDLLETICIQSFDLEGNAYQNNQLLK  
IILAMHQFI ISSADMLQKLIDLYLNALENNSSMLCVKICYFVRYWITEFWIMFKMDSKLSTTMEEFQELVRANGEELHCRL  
IDTSQINSRDWSRKLTQRVKANTSKKRKVSLLFDHLEPEELSDHLTYLEFKSFRRITFSDYQNYIVNSCVKENPTMERSIS  
LCNGISQWVQLMVLRSRPTPQLRAEVFIKFIHVAQKLHQLQNFNTLMAVIGGLCHSSISRLKETSSCVPHDVIKVFNEMTEL  
LSSYRNYDSYRRAYNECSNFKIPILGVHLKDLISLYEGMPDYLEDKKINIYKLYSLYNHINELIQLQEMPLPLEANMDLVH  
LLTSLDLYYTEDEIYELSYAREPRSHRAAPLTPCKPPVVADWASGVAPKDPKTI SKHVQRMVDSVFKNYDHDQDGYISQ  
EEFEKIAASFPPSFCVMAKDWEGLISRDEITAYFMRASSIYSKLGLGFAHNFQETTYLRPTFCDCNAGFLWGVIKQGYRCK  
DCGMNCHKQCKDLVVIECKRRPKTSVADSSPTSALASSLCPVGVKEQFHGKKRSH

>Theropithecus\_gelada

GAAAGTCTCAAACCAGTTATTACCTGAGCCACCTGGATGAACCTTGATGAAGACTTCTGCTCGGAGCTGCGGGGTGGGGCGGCTGAGAA  
CCATCAGTTGTACCCACTGGGAGATACCGTTGCACAGAGCAATAGATCTCTCCATGGTGGGGTTTTCTTCACACAGCTATTTACAAGG  
TAATTCTGATAATCAGA

>Sagittarius\_serpentarius

DLEGNAYQNNQLLKIILAMHQFI ISSADMLQKLFSTYLNALENKSSALCVKICYFVRYWITEFWVMFKMDSKLSTTMEEFQ  
ELVKANGEELHCHLIDTTQINSRDWSRKLTQRVKANTSKKRKVSLLFDHLEPEELSDHLTYLEFKSFRRISFSDYQNYIVN  
SCVKENPTMERSIALCNGISQWVQLMVLRSRPTPQLRAEVFIKFIHVAQKLHQLQNFNTLMAVIGGLCHSSISRLKETSSCV  
PHDVIKVFNEMTELLSSYRNYDSYRRAYNECSNFKIPILGVHLKDLISLYEGMPNYLEDKKINIYKLYSLYNHINELIQLQ  
EMPLPLEANMDLVHLLLTSLDLYYTEDEIYELSYAREPRSHRAAPLTPSRPPVVADWASGVAPKDPKTI SKHVQRMVDSV  
FKNYDHDQDGYISQEEFEKIAASFPPSFCVMAKDWEGL

>Macaca\_mulatta

TACAAATCCAGCACAGTTGTCACAAAAAGTGGGCTTCAGGTAGGTGGTCTCTTGGAAAGTTGTGAGGAAAGCCCAGGCCAGCTTGGAAAT  
AGATTGAGCTGGCTCGCATGAAGTAGGCTGTGATCTCATCCCTGCTGATGAGGCCTTCCCTGCCAGCAAATGACCAAGGCAAGGATGTG  
AGTATACG

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

CLUSTAL multiple sequence alignment by MUSCLE (3.8)

```
HUMAN_RASGRP1          MGTLGKAREAPRKPSHGCRAASKARLEAKPANSFPSPSHPSLAHITQFRMMVSLGHLAKGA
Amazona_guildingii    -----
Sagittarius_serpentarius -----
Neopelma_chrysocephalum MGTLGKRRENQQSAQACSTAPESALELKQTSHCPSLSNHTQV-----MMVPLGHLAKGA
Novel_protein_Serinus_canaria -----
Theropithecus_gelada  -----
Macaca_mulatta        -----
```

```
HUMAN_RASGRP1          SLDDLIDSCIQSFADGNLCRSNQLLQVMLTMHRIVISSAELLQKVITLYKDALAKNSPG
Amazona_guildingii    -----
Sagittarius_serpentarius -----
Neopelma_chrysocephalum -----DLEGNAYQNNQLLKIIAMHQFIISSADMLQKLFSTYLNALENKSSA
Novel_protein_Serinus_canaria TLEDLLETCIQSFDEGNAYQNNQLLKIIAMHQFIISSADMLQKLIDLYLNALENNSSM
Theropithecus_gelada  -----
Macaca_mulatta        -----GAAAG
-----
```

```
HUMAN_RASGRP1          LCLKICYFVRYWITEFWVMFKMDASLDTMEEFQELVKAKGEELHCRLIDTTQINARDWS
Amazona_guildingii    -----
Sagittarius_serpentarius -----
Neopelma_chrysocephalum -----
Novel_protein_Serinus_canaria -----
Theropithecus_gelada  TC-----
Macaca_mulatta        -----
```

```
HUMAN_RASGRP1          RKLTQRIKSNSTSKRKVSLFLDHLEPEELSEHLTYLEFKSFRRISFSDYQNYLVNSCVKE
Amazona_guildingii    -----
Sagittarius_serpentarius -----
Neopelma_chrysocephalum -----
Novel_protein_Serinus_canaria -----
Theropithecus_gelada  -----
Macaca_mulatta        -----
```

```
HUMAN_RASGRP1          NPTMERSIALCNGISQWVQLMVLRSRPTQLRAEVFIKFIQVAQKLHQLQNFTLMAVIGG
Amazona_guildingii    -----
Sagittarius_serpentarius -----
Neopelma_chrysocephalum -----
Novel_protein_Serinus_canaria -----
Theropithecus_gelada  -----VAQKLHQLQNFTLMAVIGG
Macaca_mulatta        -----TCAAACCAG
-----TACAAA
```

```
HUMAN_RASGRP1          LCHSSISRLKETSSHPVPEINKVLGEMTELLSSSRNYDNYRRAYGECTDFKIPILGV--H
Amazona_guildingii    -----VFNEMTELLSSYRNYDSYRRAYNECSNFKIPILGV--H
Sagittarius_serpentarius -----
Neopelma_chrysocephalum -----
Novel_protein_Serinus_canaria -----
Theropithecus_gelada  TTATTCACCTGAGCCAC-----CTGGATGAA-----CTTGATGAAGACTT
Macaca_mulatta        TCCAGCACAGTTGTCACAAAAAGTGGG-----CTTCAGGTAGG--T
```

. \* :

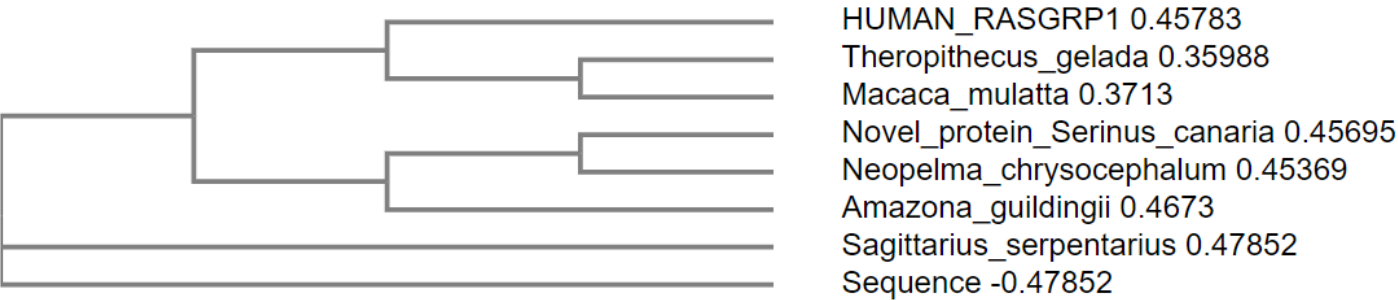
```
HUMAN_RASGRP1          LKDLISLYEAMPDYLEDGKVNvhKLLALYNHISELVQLQEVAPPLEANKDLVHLLTSLD
Amazona_guildingii    LKDLISLYEGMPDYLEDKKINIYKLYSLYNHINELIQLQEMPLPLEANMDLVHLLTSLD
Sagittarius_serpentarius -----
Neopelma_chrysocephalum -----
Novel_protein_Serinus_canaria -----
Theropithecus_gelada  LKDLISLYEAMPDYLEDKKINVYKLYSLYNHIDELIQLQEMPLPLEANMDLVHLLTSLD
CTGCTCGGAGCTGCGGGGT-----
```

Macaca_mulatta	GGTCTCTTGAAGTTGTGA----- .      .      .
HUMAN_RASGRP1	LYYTEDEIYELSYAREPRNHRAPPLTPSKPPVVVDWASGVSPKDPKTISKHVQRMVDSV
Amazona_guildingii	LYYTEDEIYELSYAREPRSHRAAPLTPSRPPVVADWASGVAPKDPKTISKHVQRMVDSV
Sagittarius_serpentarius	LYYTEDEIYELSYAREPRSHRAAPLTPSRPPVVADWASGVAPKDPKTISKHVQRMVDSV
Neopelma_chrysocephalum	LYYTEDEIYELSYAREPRSHRAAPLTPCKPPVVADWASGVAPKDPKTISKHVQRMVDSV
Novel_protein_Serinus_canaria	LYYTEDEIYELSYAREPRSHRAAPMTPSKPPVVADWASGVAPKDPKTISKHVQRMVDSV
Theropithecus_gelada	-----GGGGCGGCTGAGAACCATCAGTTGTACCCACTGGGAGATACCG
Macaca_mulatta	-----GGAAAGCCCAGGCCAGCTTGAATAGATTGAGCTGGCTCGC .      .      .      .      :      ..      .      .      .
HUMAN_RASGRP1	FKNYDHDQDGYISQEEFEKIAASFPPSFCVMDKDREGLISRDEITAYFMRASSIYSKLGL
Amazona_guildingii	FKNYDHDQDGYISQEEFEKIAASFPPSFCVMAKDWEGLISRDEITAYFMRASSIYSKLGL
Sagittarius_serpentarius	FKNYDHDQDGYISQEEFEKIAASFPPSFCVMAKDW-----
Neopelma_chrysocephalum	FKNYDHDQDGYISQEEFEKIAASFPPSFCVMAKDWEGLISRDEITAYFMRASSIYSKLGL
Novel_protein_Serinus_canaria	FKNYDHDQDGYISQEEFEKIAASFPPSFCVMAKDWE-----
Theropithecus_gelada	TTGCACAGAGCAATAG-----ATCTCTCCATGGTGGG-----
Macaca_mulatta	ATGAAGTAGGCTGT-GATCTCATCCCTGCTGATGAGG----- ..                  *      .                  *:      :      *.
HUMAN_RASGRP1	GFPHNFQETTYLKPTFCDNCAGFLWGVIKQGYRCKDCGMNCHKQCKDLVVFIECKKRAKNP
Amazona_guildingii	GFAHNFQETTYLRPTFCDNCAGFLWGVIKQGYRCKDCGMNCHKQCKDLVVFIECKRRPKTS
Sagittarius_serpentarius	-----
Neopelma_chrysocephalum	GFAHNFQETTYLRPTFCDNCAGFLWGVIKQGYRCKDCGMNCHKQCKDLVVFIECKRRPKTS
Novel_protein_Serinus_canaria	-----
Theropithecus_gelada	-----GTTTTCTTCACACAGCTATTTACA
Macaca_mulatta	-----CCTTC-----CCTGCCAGCAAATGACCAAGGCAA
HUMAN_RASGRP1	VAPTENNTSVGPVSNLCSLGAkdLLHAPEEGPFTFPNGEAVEHGEESKDRITIMLMGVSSQ
Amazona_guildingii	IPDSSPTSALA--SSLCPVGKQFHHGQEEGLFTFPNGEVVEHSEDSKDRITIMLMGSSAQ
Sagittarius_serpentarius	-----
Neopelma_chrysocephalum	VADSSPTSALA--SSLCPVGKQFHHGKKRSH-----
Novel_protein_Serinus_canaria	-----
Theropithecus_gelada	AGGTAATTCTGATAATCAGA-----
Macaca_mulatta	GGATGTGAGTA-----
HUMAN_RASGRP1	KISLRLKRAVAHKATQTESQPWIGSEGPSGPFVLSSPRKTAQDTLYVLPSPTSPCPSPVL
Amazona_guildingii	KISVRLKPSVVHEGTQTDPLLAGDVSRQ---IEKKEHKMPENPYLQLAPPSPFPSPIL
Sagittarius_serpentarius	-----
Neopelma_chrysocephalum	-----
Novel_protein_Serinus_canaria	-----
Theropithecus_gelada	-----
Macaca_mulatta	-----
HUMAN_RASGRP1	VRKRAVVKWENKDSLIIKSKEELRHLRLPTYQELEQEINTLKADNDALKIQLKYAQKKIES
Amazona_guildingii	GRKKAYVKWENKDSQKKKEEHYSCK-PSYQELEQERNILKAHNEGLKIQLEQAHTIES
Sagittarius_serpentarius	-----
Neopelma_chrysocephalum	-----
Novel_protein_Serinus_canaria	-----
Theropithecus_gelada	-----
Macaca_mulatta	-----
HUMAN_RASGRP1	LQLEKSNHVLQMEQGDCS
Amazona_guildingii	LTIHRRNHVVNDLQHRDCS
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).

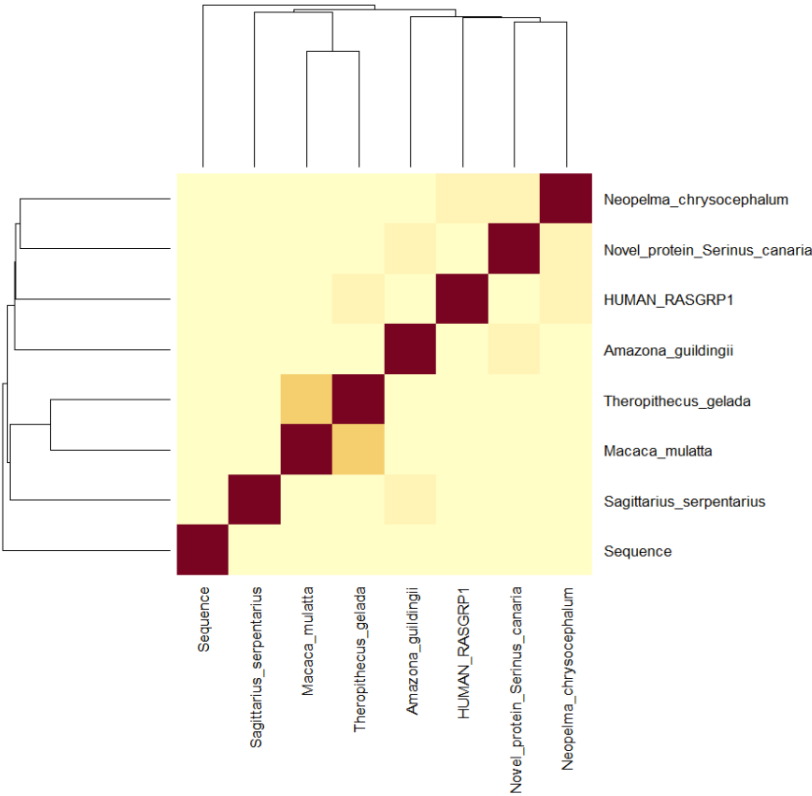
# Phylogram

Branch length: ☒ Cladogram ☐ Real



## Q7. Making heatmap using R/Bio3D

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



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	X-ray Diffraction	3.302 Å	Homo sapiens	1e-58	59.02%
6AXF_A	X-ray Diffraction	3.1 Å	Homo sapiens	3e-48	51.20%

Q9.



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: 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

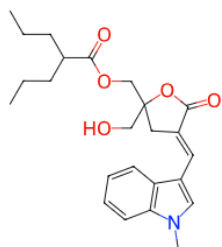
([https://www.ebi.ac.uk/chembl/assay\\_report\\_card/CHEMBL4137150/](https://www.ebi.ac.uk/chembl/assay_report_card/CHEMBL4137150/))

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 shows the highest Binding Efficiency index (BEI).

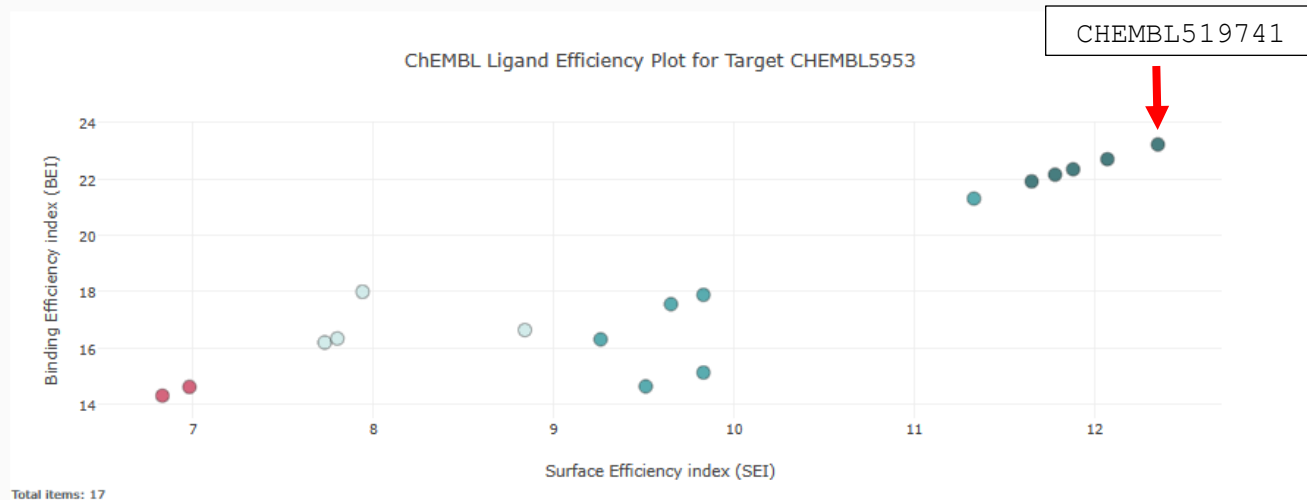
It is a small molecule that has a Molecular Weight of 413.51.



## Ligand Efficiencies



See all bioactivities for target CHEMBL5953 used in this visualisation



### Standard Value nM



The Ligand Efficiency chart plots Binding Efficiency Index (BEI) against Surface Efficiency Index (SEI), where:

**SEI** =  $(-\log_{10}(\text{Standard Value} \times 10^{-9})) \times 100 / \text{PSA}$

**BEI** =  $(-\log_{10}(\text{Standard Value} \times 10^{-9})) \times 1000 / \text{MWT}$