Irene Hsieh (A16197563)
chsieh@ucsd.edu

Questions:

[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 it's function is known.

If you do not have a favorite protein, select human RBP4 or KIF11. Do not use beta globin as this is in the worked example report that I provide you with online.

Name: retinol binding protein 4 (RBP4)

Accession: KAI4076848

Species: Homo sapiens

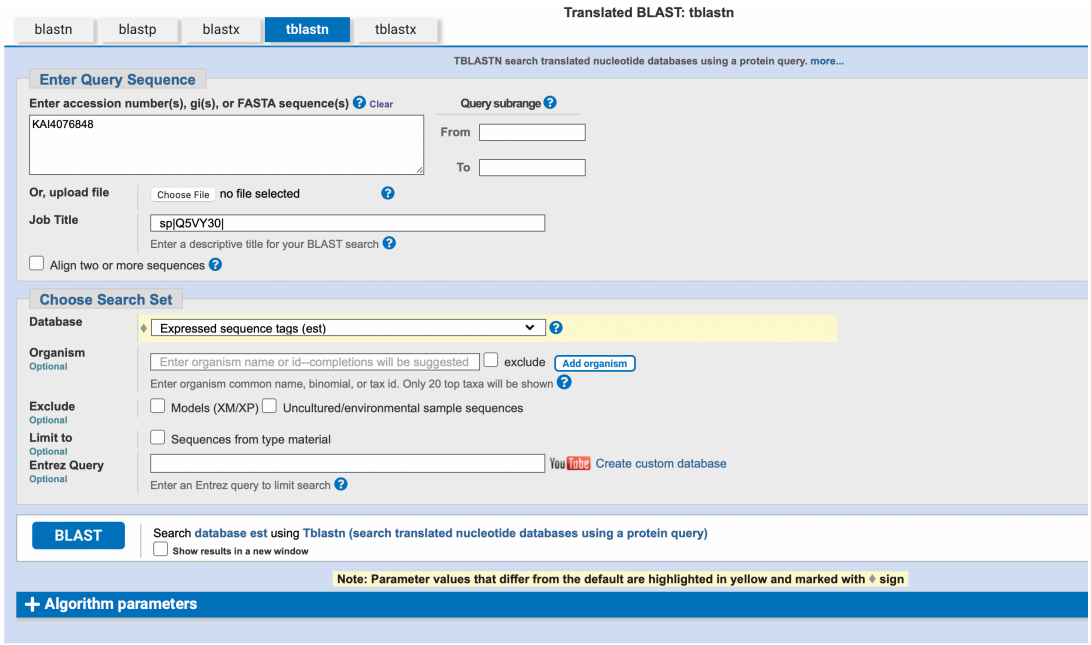
Function: Bind and transport Vitamin A, retinol, from liver to peripheral tissues

[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).

Method: tBLASTn search against Expressed sequence tags

Database: Expressed sequence tags

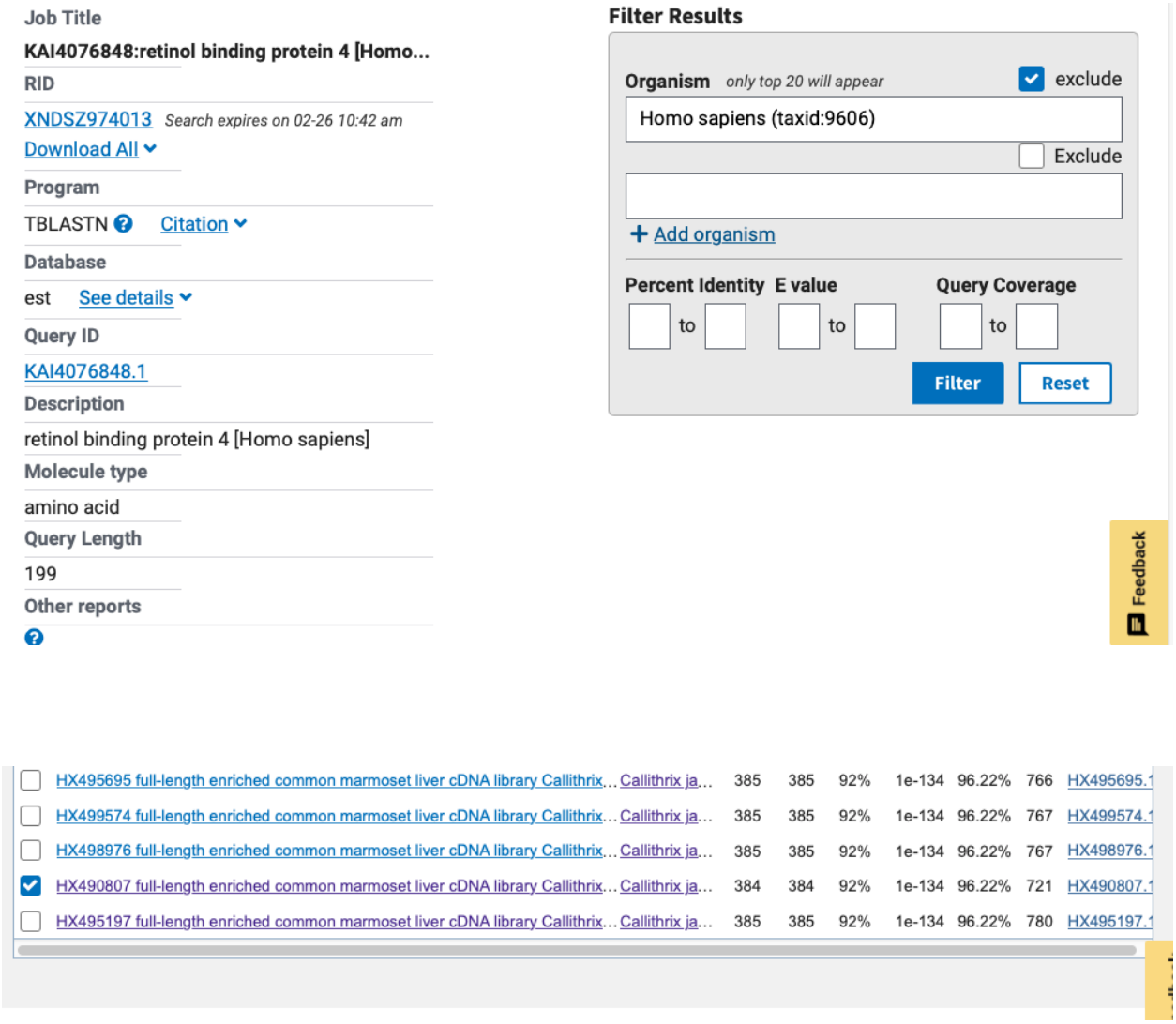
Species: Excluded Homo sapiens

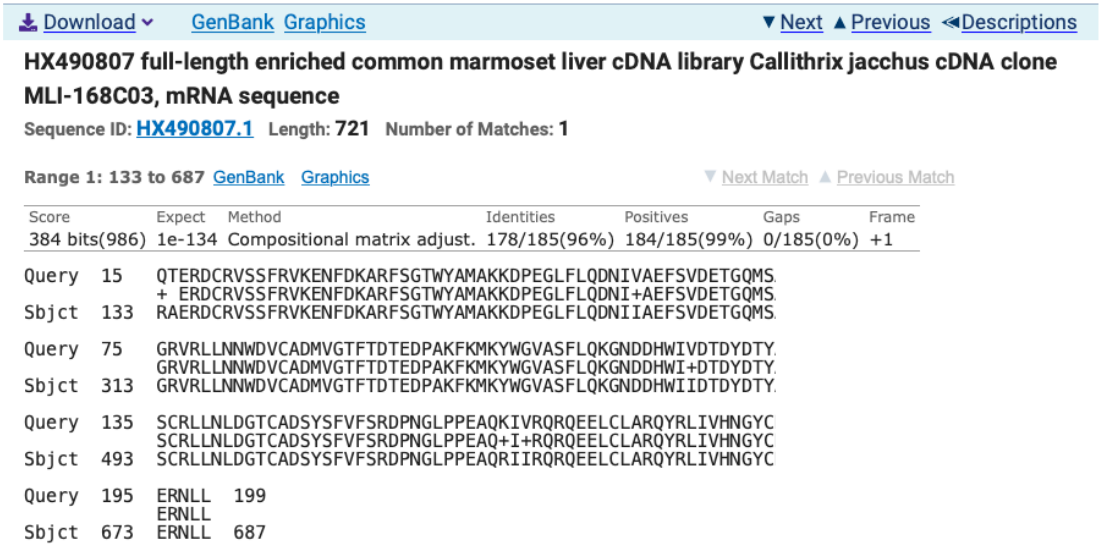


Also include the output of that BLAST search in your document. If appropriate, change the font to Courier size 10 so that the results are displayed neatly. You can also screen capture a BLAST output (e.g. alt print screen on a PC or on a MAC press ⌘-shift-4. The pointer becomes a bulls eye. Select the area you wish to capture and release. The image is saved as a file called Screen Shot [].png in your Desktop directory). It is **not** necessary to print out all of the blast results if there are many pages.

On the BLAST results, clearly indicate a match that represents a protein sequence, encoded from some DNA sequence, that is homologous to your query protein. I need to be able to inspect the pairwise alignment you have selected, including the E value and score. It should be labeled a "genomic clone" or "mRNA sequence", etc. - but include no functional annotation.

Chosen match: HX490807 , a 721 base pair from Callithrix jacchus. See below for alignment details





Download GenBank Graphics ▼Next ▲Previous ◀◀Descriptions

HX490807 full-length enriched common marmoset liver cDNA library Callithrix jacchus cDNA clone MLI-168C03, mRNA sequence

Sequence ID: HX490807.1 Length: 721 Number of Matches: 1

Range 1: 133 to 687 GenBank Graphics ▼Next Match ▲Previous Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
384 bits(986)	1e-134	Compositional matrix adjust.	178/185(96%)	184/185(99%)	0/185(0%)	+1

```
Query  15   QTERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIVAEFSVDETGQMS
            + ERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNI+AEFSVDETGQMS
Sbjct  133  RAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIIAEFSVDETGQMS

Query  75   GRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGNDDHWIVDTDYDTY
            GRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGNDDHWI+DTDYDTY
Sbjct  313  GRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGNDDHWIIDTDYDTY

Query  135  SCRLLNLDGTCADSYSFVFSRDPNGLPPEAQKIVRQRQEELCLARQYRLIVHNGYC
            SCRLLNLDGTCADSYSFVFSRDPNGLPPEAQ+I+RQRQEELCLARQYRLIVHNGYC
Sbjct  493  SCRLLNLDGTCADSYSFVFSRDPNGLPPEAQRIIRQRQEELCLARQYRLIVHNGYC

Query  195  ERNLL  199
            ERNLL
Sbjct  673  ERNLL  687
```

In general, [Q2] is the most difficult for students because it requires you to have a "feel" for how to interpret BLAST results. You need to distinguish between a perfect match to your query (i.e. a sequence that is not "novel"), a near match (something that might be "novel", depending on the results of [Q4]), and a non-homologous result.

[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. Don't forget to translate all six reading frames; the ORF (open reading frame) is likely to be the longest sequence without a stop codon. It may not start with a methionine if you don't have the complete coding region. Make sure the sequence you provide includes a header/subject line and is in traditional FASTA format.

>HX490807 full-length enriched common marmoset liver cDNA library Callithrix jacchus cDNA clone MLI-168C03, mRNA sequence
RAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIIAEFSVDETGQMSATA
KGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGNDDHWIIDTDYDTY
AVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPEAQRIIRQRQEELCLARQYRLIVHNGYC
DGQSERNLL

Here, tell me the name of the novel protein, and the species from which it derives. It is very unlikely (but still definitely possible) that you will find a novel gene from an organism such as *S. cerevisiae*, human or mouse, because those genomes have already been thoroughly annotated. It

is more likely that you will discover a new gene in a genome that is currently being sequenced, such as bacteria or plants or protozoa.

Name: HX490807 full-length enriched common marmoset liver cDNA library Callithrix jacchus cDNA clone MLI-168C03, mRNA sequence

Species: Callithrix jacchus

> Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Platyrrhini; Cebidae; Callitrichinae; Callithrix; Callithrix.

HX490807 full-length enriched common marmoset liver cDNA library Callithrix jacchus cDNA clone MLI-168C03, mRNA sequence

GenBank: HX490807.1

FASTA Graphics

Go to:

```
LOCUS       HX490807                721 bp    mRNA    linear   EST 11-FEB-2014
DEFINITION  HX490807 full-length enriched common marmoset liver cDNA library
            Callithrix jacchus cDNA clone MLI-168C03, mRNA sequence.
ACCESSION   HX490807
VERSION     HX490807.1
DBLINK      BioSample: SAMN01998206
KEYWORDS    EST.
SOURCE      Callithrix jacchus (white-tufted-ear marmoset)
  ORGANISM  Callithrix jacchus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Platyrrhini; Cebidae; Callitrichinae; Callithrix; Callithrix.
REFERENCE   1  (bases 1 to 721)
  AUTHORS   Tatsumoto,S., Adati,N., Tohtoki,Y., Sakaki,Y., Boroviak,T.,
            Habu,S., Okano,H., Suemizu,H., Sasaki,E. and Satake,M.
  TITLE     Development and characterization of cDNA resources for the common
            marmoset: one of the experimental primate models
  JOURNAL   DNA Res. 20 (3), 255-262 (2013)
   PUBMED   23543116
COMMENT     Contact: Shoji Tatsumoto
            Comparative Genomics Laboratory
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411-8540, Japan
            URL:http://dna.brc.riken.jp/en/marmoseten.html
            Description: This library was made using Vector-capping method.
            See, Kato, S. et al. (2005) Vector-capping: a simple method for
            preparing a high-quality Full-Length cDNA Library. DNA Res.
            12:53-62.
```

[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.

- If there is a match with 100% amino acid identity to a protein in the database, from the same species, then your protein is NOT novel (even if the match is to a protein with a name such as "unknown"). Someone has already found and annotated this sequence, and assigned it an accession number.

- If the top match reported has less than 100% identity, then it is likely that your protein is novel, and you have succeeded.

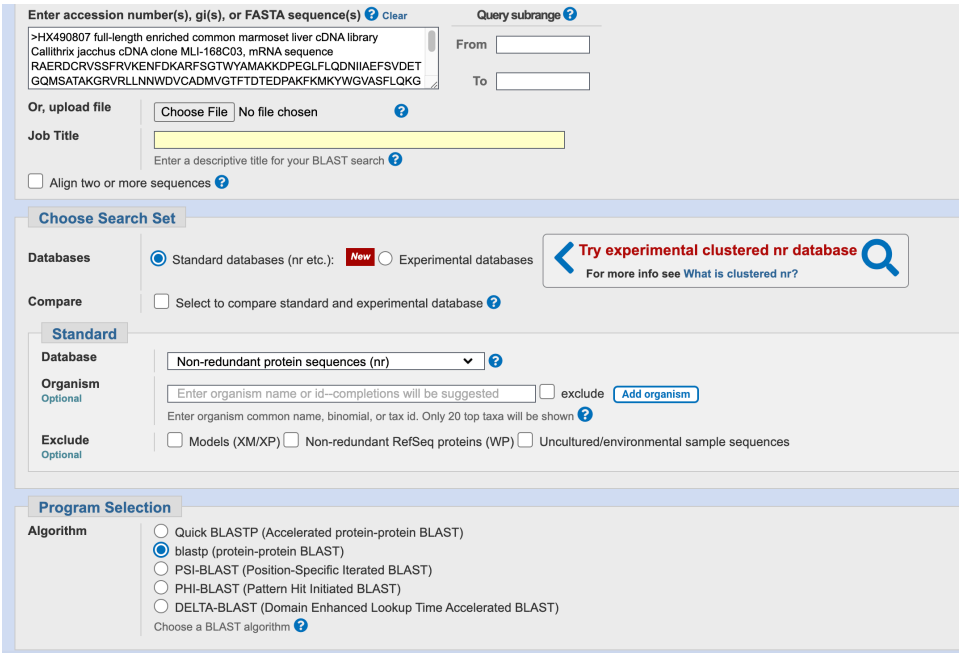
- If there is a match with 100% identity, but to a different species than the one you started with, then you have likely succeeded in finding a novel gene.

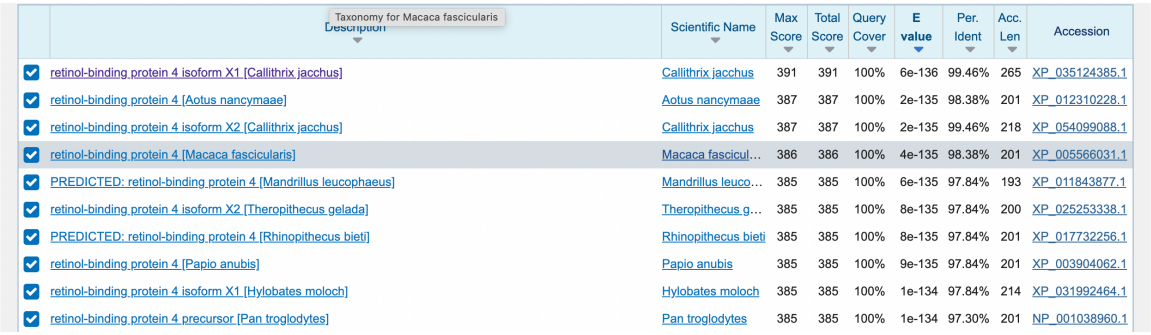
- If there are no database matches to the original query from [Q1], this indicates that you have partially succeeded: yes, you may have found a new gene, but no, it is not actually homologous to the original query. You should probably start over.

Details:

A BLASTP search against non redundant protein sequence yielded a top hit research is to a protein from Callithrix jacchus

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





Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
retinol-binding protein 4 isoform X1 [Callithrix jacchus]	Callithrix jacchus	391	391	100%	6e-136	99.46%	265	XP_035124385.1
retinol-binding protein 4 [Aotus nancymaae]	Aotus nancymaae	387	387	100%	2e-135	98.38%	201	XP_012310228.1
retinol-binding protein 4 isoform X2 [Callithrix jacchus]	Callithrix jacchus	387	387	100%	2e-135	99.46%	218	XP_054099088.1
retinol-binding protein 4 [Macaca fascicularis]	Macaca fascicul…	386	386	100%	4e-135	98.38%	201	XP_005566031.1
PREDICTED: retinol-binding protein 4 [Mandrillus leucophaeus]	Mandrillus leuco…	385	385	100%	6e-135	97.84%	193	XP_011843877.1
retinol-binding protein 4 isoform X2 [Theropithecus gelada]	Theropithecus g…	385	385	100%	8e-135	97.84%	200	XP_025253338.1
PREDICTED: retinol-binding protein 4 [Rhinopithecus bieti]	Rhinopithecus bieti	385	385	100%	8e-135	97.84%	201	XP_017732256.1
retinol-binding protein 4 [Papio anubis]	Papio anubis	385	385	100%	9e-135	97.84%	201	XP_003904062.1
retinol-binding protein 4 isoform X1 [Hylobates moloch]	Hylobates moloch	385	385	100%	1e-134	97.84%	214	XP_031992464.1
retinol-binding protein 4 precursor [Pan troglodytes]	Pan troglodytes	385	385	100%	1e-134	97.30%	201	NP_001038960.1

retinol-binding protein 4 isoform X1 [Callithrix jacchus]

Sequence ID: XP_035124385.1 Length: 265 Number of Matches: 1

Range 1: 81 to 265 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
391 bits(1004)	6e-136	Compositional matrix adjust.	184/185(99%)	185/185(100%)	0/185(0%)

```
Query  1    RAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIIAEFSVDETGQMSATAK  60
            RAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIIAEFSVDETGQMSATAK
Sbjct  81   RAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIIAEFSVDETGQMSATAK  140

Query  61   GRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGNDDHWIIDTDYDTYAVQY  120
            GRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGNDDHWIIDTDYDTYAVQY
Sbjct  141  GRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGNDDHWIIDTDYDTYAVQY  200

Query  121  SCRLLNLDGTCADSYSFVFSRDPNGLPPEAQRIIRQRQEELCLARQYRLIVHNGYCDGQS  180
            SCRLLNLDGTCADSYSFVFSRDPNGLPPEAQRIIRQRQEELCLARQYRLIVHNGYCDG+S
Sbjct  201  SCRLLNLDGTCADSYSFVFSRDPNGLPPEAQRIIRQRQEELCLARQYRLIVHNGYCDGKS  260

Query  181  ERNLL  185
            ERNLL
Sbjct  261  ERNLL  265
```

[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.
Side-note: Indicate your sequence in the alignment by choosing an appropriate name for each sequence in the input unaligned sequence file (i.e. edit the sequence file so that the species, or short common, names (rather than accession numbers) display in the output alignment and in the subsequent answers below). The goal in this step is to create an interesting an alignment for building a phylogenetic tree that illustrates species divergence.

Re-labeled sequences for alignment:

>Homo sapiens retinol binding protein 4 (RBP4)
MNYSKIPAQVDLRRQTERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIVA
EFSVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKG

NDDHWIVDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPEAQKIVRQRQEEL
CLARQYRLIVHNGYCDGRSERNLL

>HX490807 full-length enriched common marmoset liver cDNA library Callithrix jacchus cDNA clone MLI-168C03
RAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIIAEFSVDETGQMSATA
KGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGNDDHWIIDTDYDTY
AVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPEAQRIIRQRQEELCLARQYRLIVHNGYC
DGQSERNLL

>Theropithecus gelada retinol binding protein 4 (RBP4)
MNYSKIPAQVDLRGGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNI
VAEFSVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQ
KGNDDHWIIDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPEAQKIVRQRQE
ELCLARQYRLIVHNGYCDGRSERNLL

>Pan troglodytes retinol binding protein 4 (RBP4)
RAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIVAEFSVDETGQMSATA
KGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGNDDHWIVDTDYDT
YAVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPEAQKIVRQRQEELCLARQYRLIVHNG
YCDGRSERNLL

> Symphalangus syndactylus retinol binding protein 4 (RBP4)
MNYSKIPAQVDLRRQTERDCRVSSFRVKENFDKARFSGIWYAMAKKDPEGLFLQDNIVA
EFSVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKG
NDDHWIIDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPEAQKIVRQRQEEL
CLARQYRLIVHNGYCDGRSERNLL

>Mandrillus leucophaeus retinol binding protein 4
RAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIVAEFSVDETGQMSATA
KGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGNDDHWIIDTDYDTY
AVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPEAQKIVRQRQEELCLARQYRLIVHNGY
CDGRSERNLL

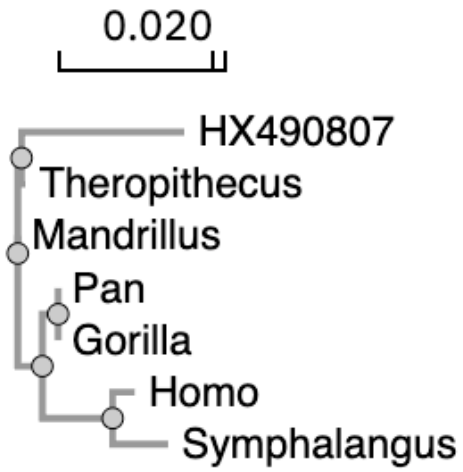
>Gorilla gorilla gorilla retinol binding protein 4 (RBP4)
RAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIVAEFSVDETGQMSATA
KGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGNDDHWIVDTDYDT
YAVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPEAQKIVRQRQEELCLARQYRLIVHNG
YCDGRSERNLL

Alignment -> Obtained using CLUSTAL OMEGA at EBI:

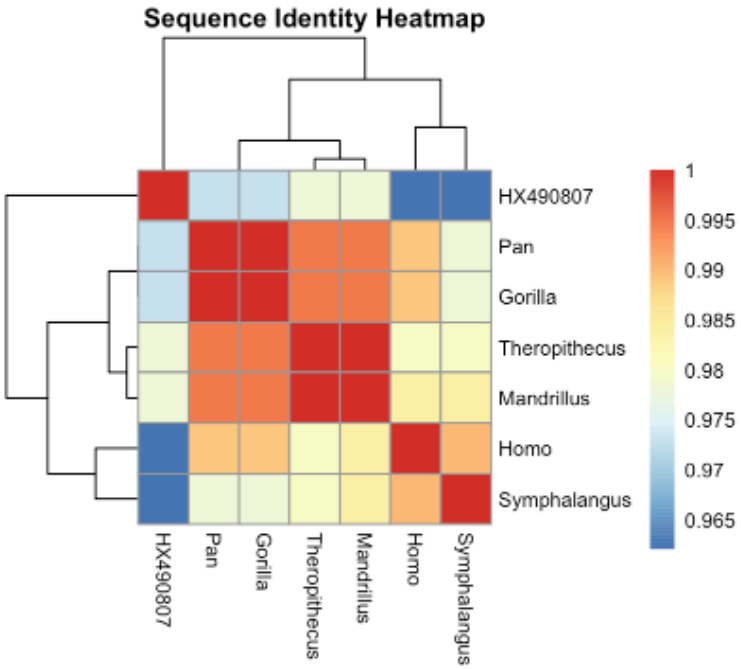
```
CLUSTAL O(1.2.4) multiple sequence alignment

HX490807          ----------------RAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIIA 45
Theropithecus     MNYSKIPAQVDLRGGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIVA 60
Mandrillus        ----------------RAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIVA 45
Pan               ----------------RAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIVA 45
Gorilla           ----------------RAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIVA 45
Homo              MNYSKIPAQVDLRR-QTERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIVA 59
Symphalangus      MNYSKIPAQVDLRR-QTERDCRVSSFRVKENFDKARFSGIWYAMAKKDPEGLFLQDNIVA 59
                               ::**********************  ******************:*

HX490807          EFSVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGNDD 105
Theropithecus     EFSVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGNDD 120
Mandrillus        EFSVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGNDD 105
Pan               EFSVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGNDD 105
Gorilla           EFSVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGNDD 105
Homo              EFSVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGNDD 119
Symphalangus      EFSVDETGQMSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFLQKGNDD 119
                  ************************************************************

HX490807          HWIIDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPEAQRIIRQRQEELCLAR 165
Theropithecus     HWIIDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPEAQKIVRQRQEELCLAR 180
Mandrillus        HWIIDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPEAQKIVRQRQEELCLAR 165
Pan               HWIVDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPEAQKIVRQRQEELCLAR 165
Gorilla           HWIVDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPEAQKIVRQRQEELCLAR 165
Homo              HWIVDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPEAQKIVRQRQEELCLAR 179
Symphalangus      HWIIDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLPPEAQKIVRQRQEELCLAR 179
                  ***:*******************************************:*:***********

HX490807          QYRLIVHNGYCDGQSERNLL 185
Theropithecus     QYRLIVHNGYCDGRSERNLL 200
Mandrillus        QYRLIVHNGYCDGRSERNLL 185
Pan               QYRLIVHNGYCDGRSERNLL 185
Gorilla           QYRLIVHNGYCDGRSERNLL 185
Homo              QYRLIVHNGYCDGRSERNLL 199
Symphalangus      QYRLIVHNGYCDGRSERNLL 199
                  *************:******
```

[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” online from the EBI



[Q7] Generate a sequence identity based heatmap of your aligned sequences using R.
If necessary convert your sequence alignment to the ubiquitous FASTA format (Seaview can read in clustal format and "Save as" FASTA format for example). Read this FASTA format alignment into R with the help of functions in the Bio3D package. Calculate a sequence identity matrix (again using a function within the Bio3D package). Then generate a heatmap plot and add to your report. Do make sure your labels are visible and not cut at the figure margins.



[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.
List the top 3 unique hits (i.e. not hits representing different chains from the same structure) along with their Evalue and sequence identity to your query. Please also add annotation details of these structures. For example include the annotation terms PDB identifier (structureId), Method used to solve the structure (experimentalTechnique), resolution (resolution), and source organism (source).
HINT: You can use a single sequence from your alignment or generate a consensus sequence from your alignment using the Bio3D function consensus(). The Bio3D functions blast.pdb(), plot.blast() and pdb.annotate() are likely to be of most relevance for completing this task. Note that the results of blast.pdb() contain the hits PDB identifier (or pdb.id) as well as Evalue and identity. The results of pdb.annotate() contain the other annotation terms noted above.
Note that if your consensus sequence has lots of gap positions then it will be better to use an original sequence from the alignment for your search of the PDB. In this case you could chose the sequence with the highest identity to all others in your alignment by calculating the row-wise maximum from your sequence identity matrix.

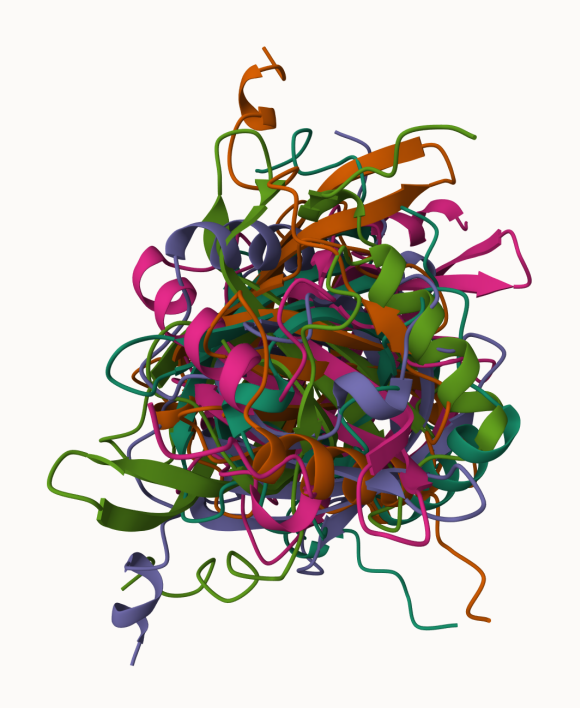
PDB ID	Technique	Resolution	Source	E-value	Identity
4O9S	X-ray Diffraction.	2.30 Å	Homo sapiens	2E-137	97.27%
1AQB	X-ray Diffraction.	1.65 Å	Sus scrofa domesticus	3E-130	91.80%
1HBQ	X-ray Diffraction.	1.70 Å	Bos taurus	2E-129	91.26%
1RLB	X-ray Diffraction.	3.10 Å	Gallus gallus	5E-123	91.38%

[Q9] Using AlphaFold notebook generate a structural model using the default parameters for your novel protein sequence.
Note that this can take some time depending upon your sequence length. If your model is taking many hours to generate or your input sequence yields a “too many amino acids” (i.e. length) error you can focus on a single domain from your sequence - identify region by searching for PFAM domain matches.
Once complete save the resulting PDB format file for your records. Finally, generate a molecular figure of your generated PDB structure using the Mol* viewer online (or VMD/PyMol/Chimera if you prefer). To complete your analysis you can optionally highlight conserved residues that are

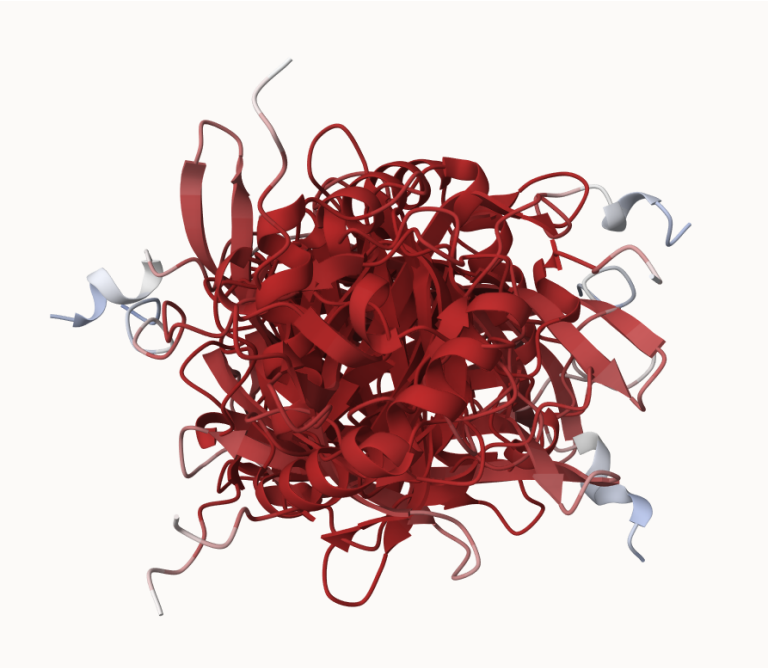
likely to be functional as spacefill and the protein as cartoon colored by local alpha fold pLDDT quality score. This score is contained in the B-factor column of your PDB downloaded file. Please use a white or transparent background for your figure (i.e. not the default black in PyMol/ VMD/Chimera etc.).

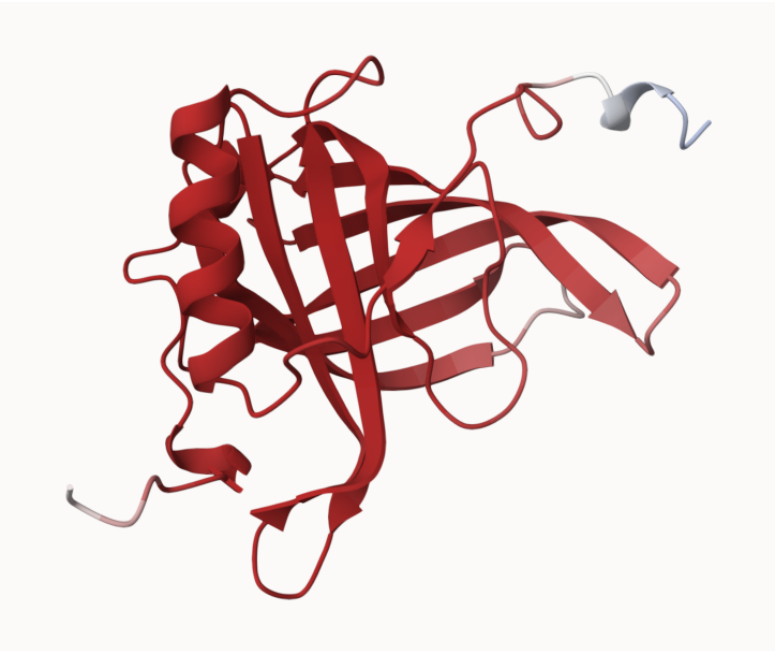
Between the novel RBP4 protein from Callithrix jacchus and the Homo Sapien RBP4 protein structure, there is about 97.27% identity between them.

Novel protein: (first is combination of 5 Different PDBs, the second is one single PDB)

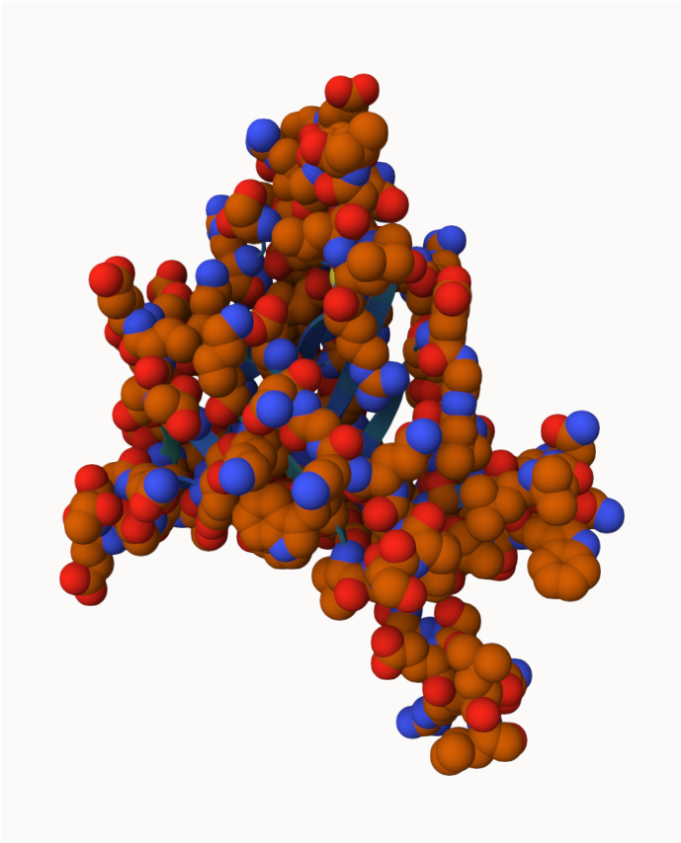


With cartoon colored pLDDT score





Conserved Residue that are likely to be functional (Spacefill)



Homo Sapien (PDB:4O9S) :

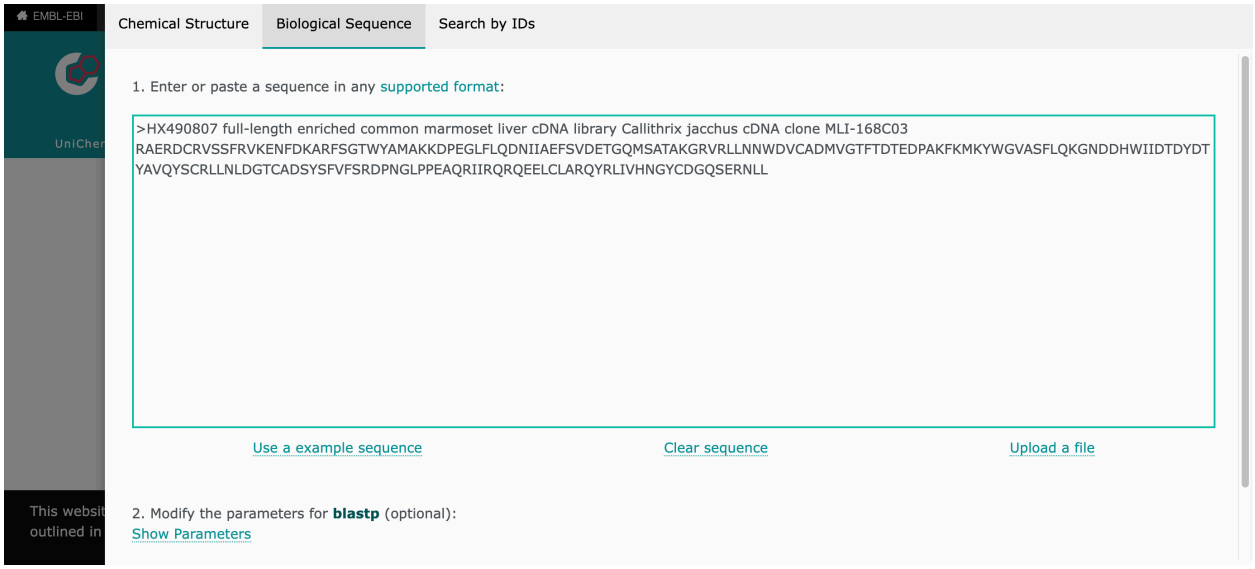
[Q10] Perform a “Target” search of ChEMBEL (https://www.ebi.ac.uk/chembl/) with your novel sequence. Are there any Target Associated Assays and ligand efficiency data reported that may be useful starting points for exploring potential inhibition of your novel protein? If there are no assays listed here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

Non available as of 02/27/2024 for Callithrix jacchus.
But I found human samples of RBP4 in the following:

CHEMBL details 20 Binding Assays and 3 Functional Assays.
https://www.ebi.ac.uk/chembl/web_components/explore/assays/STATE_ID:ybZpVu8MIbAPvsGS6y1aug==

Ligand efficiency data is in the following, All bio activities for target:

https://www.ebi.ac.uk/chembl/web_components/explore/activities/QUERYSTRING%3Atarget_chembl_id%3ACHEMBL3100%20AND%20standard_type%3A(IC50%20OR%20Ki%20OR%20EC50%20OR%20Kd)%20AND%20_exists_%3Astandard_value%20AND%20_exists_%3Aligand_efficiency



EBI > Databases > Chemical Biology > ChEMBL Database > CHEMBL3100

Target Report Card

Name And Classification

ID: CHEMBL3100

Type: SINGLE PROTEIN

Preferred Name: Plasma retinol-binding protein

Synonyms:

Plasma retinol-binding protein | Plasma retinol-binding protein(1-176) | Plasma retinol-binding protein(1-179) | Plasma retinol-binding protein(1-181) | Plasma retinol-binding protein(1-182)

ALL (4 MORE)

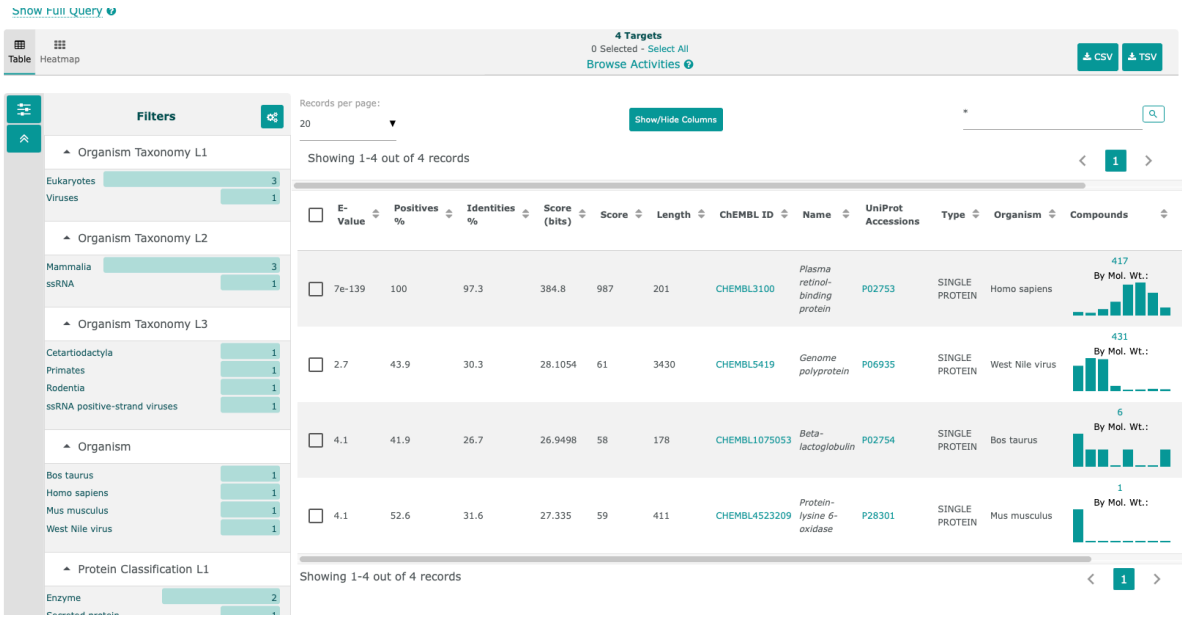
Synonyms: ---

organism: Homo sapiens

Species Group: No

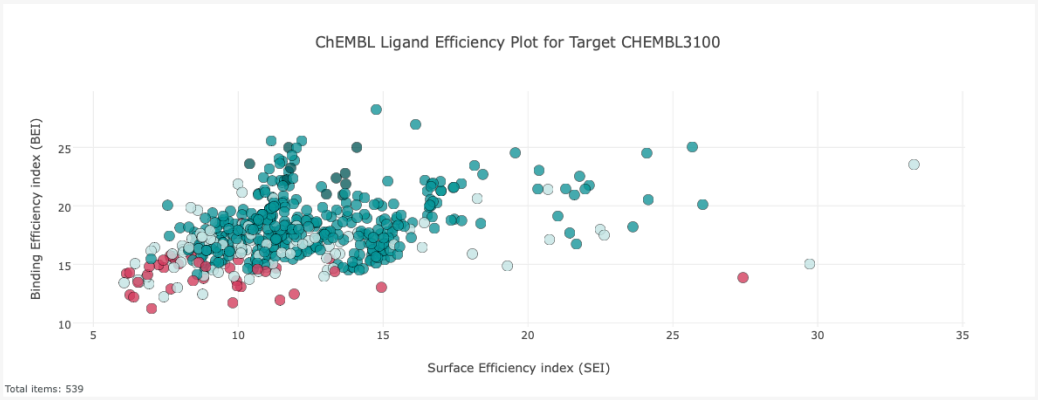
Protein Target Classification:

1. Secreted protein



Ligand Efficiencies

See all bioactivities for target CHEMBL3100 used in this visualisation

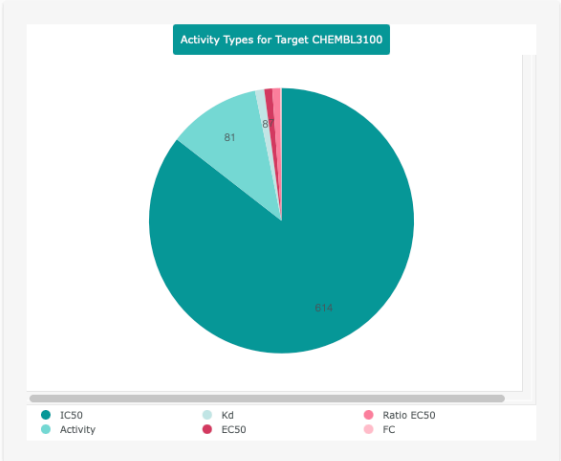


The Ligand Efficiency chart plots Binding Efficiency Index (BEI) against Surface Efficiency Index (SEI), where:

SEI = (-log10(Standard Value*10^-9))*100/PSA
BEI = (-log10(Standard Value*10^-9))*1000/MWT

Activity Charts

Associated Bioactivities



Associated Assays