

Find a Gene Assignment – BGGN213

Jason Hsiao (A15871650)

1) Protein: Human RBP4 (Retinol Binding Protein 4 Isoform a Precursor)

Accession Number: [NP_001310446.1](#)

Species: Human

Function: "This protein belongs to the lipocalin family and is the specific carrier for retinol (vitamin A alcohol) in the blood. It delivers retinol from the liver stores to the peripheral tissues. In plasma, the RBP-retinol complex interacts with transthyretin which prevents its loss by filtration through the kidney glomeruli. A deficiency of vitamin A blocks secretion of the binding protein post translationally and results in defective delivery and supply to the epidermal cells." [provided by RefSeq, Jul 2008].

2) BLAST Method Used: tblastn 2.14.1+

Database: Expressed sequence tags (est)

Job title	refNP_001310446.1
Full query description	
RID	R03WM1JS016 Search expires on 12-07 13:07 pm Download All
Program	TBLASTN Citation
Database	est See details
Query ID	NP_001310446.1
Description	retinol-binding protein 4 isoform a precursor [Homo sapiens]
Molecule type	amino acid
Query Length	201
Other reports	?

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	603191445F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5262803 5'. mRNA sequence	Homo sapiens	404	404	100%	2e-142	96.02%	703	BI546698.1
<input checked="" type="checkbox"/>	DC643333 macaque kidney cDNA library QreA Macaca fascicularis cDNA clone QreA-00085 5'. mRNA sequence	Macaca fascicularis	402	402	100%	1e-141	95.02%	753	DC643333.1
<input checked="" type="checkbox"/>	AGENCOURT_8351270 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6286887 5'. mRNA sequence	Homo sapiens	404	404	100%	1e-141	96.02%	902	BQ645928.1
<input checked="" type="checkbox"/>	AGENCOURT_8511748 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6296828 5'. mRNA sequence	Homo sapiens	404	404	100%	2e-141	96.02%	916	BQ650963.1
<input checked="" type="checkbox"/>	603246016F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5288666 5'. mRNA sequence	Homo sapiens	404	404	100%	2e-141	96.02%	910	BI599359.1
<input checked="" type="checkbox"/>	DC629429 macaque liver cDNA library QivC Macaca fascicularis cDNA clone QivC-30299 5'. mRNA sequence	Macaca fascicularis	402	402	100%	3e-141	95.02%	845	DC629429.1
<input checked="" type="checkbox"/>	ILLUMIGEN_MCO_54653 Katze_MNLV Macaca nemestrina cDNA clone IBIUW:31443 5' similar to Bases 5 to 85...	Macaca nemestrina	403	403	100%	3e-141	95.02%	905	DR772668.1
<input checked="" type="checkbox"/>	DC629059 macaque liver cDNA library QivC Macaca fascicularis cDNA clone QivC-29199 5'. mRNA sequence	Macaca fascicularis	402	402	100%	3e-141	95.02%	850	DC629059.1
<input checked="" type="checkbox"/>	ILLUMIGEN_MCO_53369 Katze_MNLV Macaca nemestrina cDNA clone IBIUW:32883 5' similar to Bases 5 to 83...	Macaca nemestrina	403	403	100%	4e-141	95.02%	934	DR771938.1

3)

[Download](#) [GenBank](#) [Graphics](#)

603191445F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5262803 5', mRNA sequence

Sequence ID: [BI546698.1](#) Length: 703 Number of Matches: 1

Range 1: 52 to 654 [GenBank](#) [Graphics](#)

[Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps	Frame
404 bits(1037)	2e-142	Compositional matrix adjust.	201/201(100%)	201/201(100%)	0/201(0%)	+1
Query 1	MKWVWAl111aa1GSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV					60
Sbjct 52	MKWVWALLLLAALGSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIV					231
Query 61	AEFSVDETGQMSATAKGRVRLNNWDVCADMGVTFDTEDPAKFKMKYWGVSFLQKGND					120
Sbjct 232	AEFSVDETGQMSATAKGRVRLNNWDVCADMGVTFDTEDPAKFKMKYWGVSFLQKGND					411
Query 121	DHWIVDTYDITYAVQYSCRLNLDGTCADSYSFVFSRDPNGLPPEAQKIVRQRQEELCLA					180
Sbjct 412	DHWIVDTYDITYAVQYSCRLNLDGTCADSYSFVFSRDPNGLPPEAQKIVRQRQEELCLA					591
Query 181	RQYRLIVHNGYCDGRSERNLL		201			
Sbjct 592	RQYRLIVHNGYCDGRSERNLL		654			

```
>BI546698.1 603191445F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5262803 5',
mRNA sequence
AGCGGAGGCTTGC GCGCGGTTCCCTCCCGGTGGGCGGATTCTTGGGCAAGATGAAGTGGGTGTGGGCGC
TCTTGCTGTTGGCGGCGCTGGGCAGCGGCCGCGCGGAGCGCGACTGCCGAGTGAGCAGCTTCCGAGTCAA
GGAGAACTTCGACAAGGCTCGCTTCTCTGGGACCTGGTACGCCATGGCCAAGAAGGACCCCGAGGGCCTC
TTTCTGCAGGACAACATCGTCGCGGAGTTCTCCGTGGACGAGACCGGCCAGATGAGCGCCACAGCCAAGG
GCCGAGTCCGTCTTTTGAATAACTGGGACGTGTGCGCAGACATGGTGGGCACCTTCACAGACACCGAGGA
CCCTGCCAAGTTCAAGATGAAGTACTGGGGCGTAGCCTCCTTTCTCCAGAAAGGAAATGATGACCACTGG
ATCGTCGACACAGACTACGACACGTATGCCGTGCAGTACTCCTGCCGTCTCCTGAACCTCGATGGCACCT
GTGCTGACAGCTACTCCTTCGTGTTTTCCCGGGACCCCAACGGCCTGCCCCCAGAAGCGCAGAAGATTGT
AAGGCAGCGGCAGGAGGAGCTGTGCCTGGCCAGGCAGTACAGGCTGATCGTCCACAACGGTTACTGCGAT
GGCAGATCAGAAAGAAACCTTTTGTAGCAATATCAAGAATCTAGTTTCATCTGAGAACTTCTGATTAGCT
CTC
```

Translated Protein Sequence via Emboss Transeq

```
>BI546698.1_1 603191445F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5262803 5', mRNA
sequence
```

```
SGGLRAVPLPVGGFLGKMKVWVALLLLAALGSGRAERDCRVSSFRVKENFDKARFSGTWY
AMAKKDPEGLFLQDNIVAIEFSVDETGQMSATAKGRVRLNNWDVCADMVGTFTDTEPAK
FKMKYWGVAFLQKGNDDHWIIDTDYDTYAVQYSCRLLNLDGTCADSYSFVFSRDPNGLP
PEAQKIVRQRQEELCLARQYRLIVHNGYCDGRSERNLI*QYQESSFI*ELLISX
```

Longest sequence without stop * of the 6 reading frames.

4)

Descriptions	Graphic Summary	Alignments	Taxonomy					
Sequences producing significant alignments								
Download Select columns Show 100 ?								
<input checked="" type="checkbox"/> select all 100 sequences selected								
GenPept Graphics Distance tree of results Multiple alignment MSA Viewer								
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> retinol-binding protein 4 isoform X1 (Gorilla gorilla gorilla)	Gorilla gorilla go...	454	454	91%	4e-160	100.00%	265	XP_018890983.3
<input checked="" type="checkbox"/> retinol-binding protein 4 isoform X1 (Hylobates moloch)	Hylobates moloch	438	438	89%	2e-154	99.05%	214	XP_031992464.1
<input checked="" type="checkbox"/> retinol-binding protein 4 precursor (Pan troglodytes)	Pan troglodytes	421	421	85%	3e-148	100.00%	201	NP_001038960.1
<input checked="" type="checkbox"/> retinol-binding protein 4 (Papio anubis)	Papio anubis	421	421	85%	6e-148	99.50%	201	XP_003904062.1
<input checked="" type="checkbox"/> retinol-binding protein 4 (Nomascus leucogenys)	Nomascus leuco...	420	420	85%	1e-147	99.00%	201	XP_003255281.1

No 100% identity with Homo sapiens, top hit is from gorilla.

5) Trimmed Alignment:

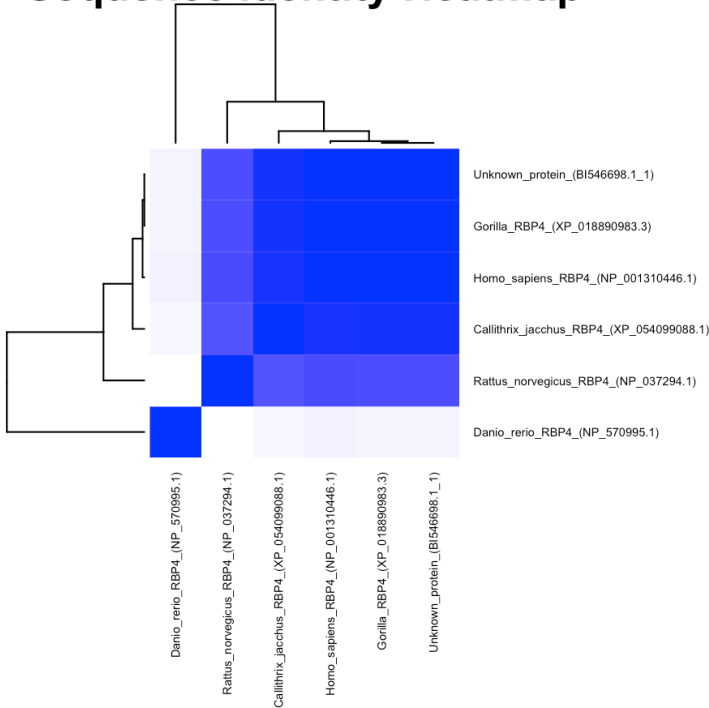
```
>Danio_rerio_RBP4_(NP_570995.1)
XM--LRLCIAVCVLATCWAQDCQVS NFAVQQDFNRTRYQGTWYAVAKKDPVGL
FXLLDNIVANFKVEEDGTM TATAIGRVII LNNWEMCANMFGTFEDTEDPAKFKMKYWGAA
AYXLQGTGYDDHWIIDTDYDNYA I HYS CRELDEDDGTCLDGYSFIFSRHPDGLRPEDQAI VT
QKKXQDICFLGKYRRVAHTGFCEAA
>Rattus_norvegicus_RBP4_(NP_037294.1)
XMEWVWALVLLAALGGGSAERDCRVSSFRVKENFDKARFSGLWYAI AKKDPEGL
FXLQDNIIAEFSVDEKGHMSATAKGRVRLLSNWEVCADMVGTFTDTEPAKFKMKYWGVA
SFXLQQRGNDDHWIIDTDYDTFALQYSCRLLNLDGTCADSYSFVFSRDPNGLTPETRLVR
QRQXEELCLERQYRWIEHNGYCQSR
>Callithrix_jacchus_RBP4_(XP_054099088.1)
KMKVWVALLLLAVLGISRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGL
FXLQDNIIAEFSVDETGQMSATAKGRVRLNNWDVCADMVGTFTDTEPAKFKMKYWGVA
```

SFXLQKGNDHWHIIDTDYDTYAVQYSCRLNLNLDGTCADSYSFVFSRDPNGLPPEAQRIIR
QRQXEELCLARQYRLIVHNGYCDGK
>Gorilla_RBP4_(XP_018890983.3)
KMKWVWALLLLAALGSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGL
FXLQDNIVAEFSVDETQMSATAKGRVRLNNDVDCADMVGTFTDTEPAKFKMKYWGVA
SFXLQKGNDHWHIVDTDYDTYAVQYSCRLNLNLDGTCADSYSFVFSRDPNGLPPEAQKIVR
QRQXEELCLARQYRLIVHNGYCDGR
>Unknown_protein_(BI546698.1_1)
KMKWVWALLLLAALGSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGL
FXLQDNIVAEFSVDETQMSATAKGRVRLNNDVDCADMVGTFTDTEPAKFKMKYWGVA
SFXLQKGNDHWHIVDTDYDTYAVQYSCRLNLNLDGTCADSYSFVFSRDPNGLPPEAQKIVR
QRQXEELCLARQYRLIVHNGYCDGR
>Homo_sapiens_RBP4_(NP_001310446.1)
XMKWVWALLLLAALGSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGL
FXLQDNIVAEFSVDETQMSATAKGRVRLNNDVDCADMVGTFTDTEPAKFKMKYWGVA
SFXLQKGNDHWHIVDTDYDTYAVQYSCRLNLNLDGTCADSYSFVFSRDPNGLPPEAQKIVR
QRQXEELCLARQYRLIVHNGYCDGR

6)



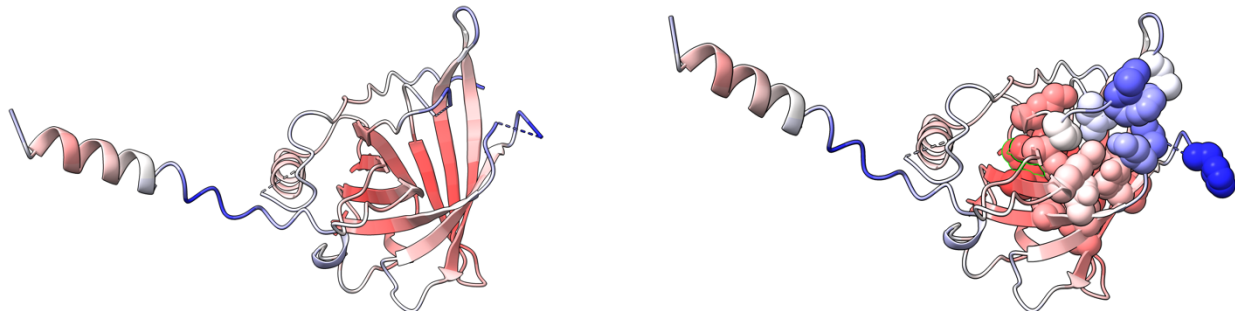
7) **Sequence Identity Heatmap**



8) Ran trimmed alignment of 'unknown protein' from (5) on NCBI's protein blast against pdb database. Top 3 non-RBP4 hits and information are shown as follows:

1. Sander vitreus Sander cyanin Fluorescent Protein (SFP), Chain A
E-Value = 4e-08, Sequence Identity = 29.41%, PDB Identifier = 5EZ2, Method = X-ray diffraction, Resolution = 1.85 angstroms, Source Organism = Sander vitreus
2. Human Glycodelin, Chain A
E-Value = 4e-06, Sequence Identity = 24.84%, PDB Identifier = 4R0B, Method = X-ray diffraction, Resolution = 2.45 angstroms, Source Organism = Homo sapiens
3. Human Apolipoprotein D, Chain A
E-Value = 8e-06, Sequence Identity = 27.85%, PDB Identifier = 2HZQ, Method = X-ray diffraction, Resolution = 1.8 angstroms, Source Organism = Homo sapiens

9) One of 5 of the resulting pdb files were chosen for illustration.
Using InterPro database for conserved domains and function, I found predicted retinol binding sites of my unknown protein (conserved residues between my MSA, boxed in red) and spacefilled with spheres (bottom right figure), and colored the sequence by pLDDT score (low = blue, high = red):



The spacefill clearly shows the retinol binding site within the beta barrel of the protein structure!

10) Yes, there are target associated assays and ligand efficiency data, as shown below for the top hit (RBP) in ChEMBL:

ChEMBL >Unknown_protein_(BI54669 . . . QRQXEELCLARQYRLVHNGYCDGR

Show Full Query

6 Targets
0 Selected - Select All
Browse Activities

Table Heatmap

Records per page: 20

Showing 1-6 out of 6 records

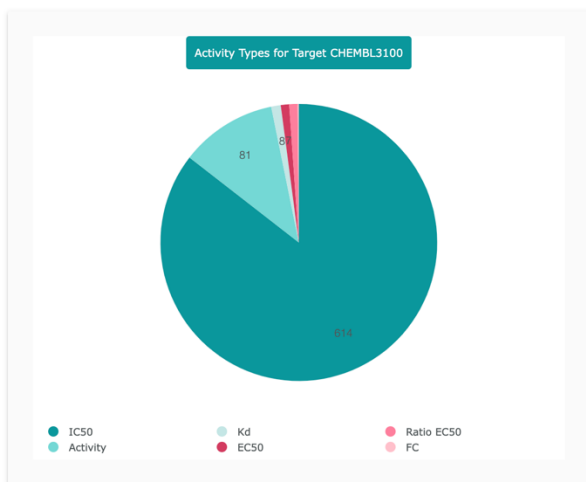
Filters

- Organism Taxonomy L1
 - Eukaryotes 4
 - Bacteria 2
- Organism Taxonomy L2
 - Mammalia 4
 - Gram-Negative 2

Showing 1-6 out of 6 records

	E-Value	Positives %	Identities %	Score (bits)	Score	Length	ChEMBL ID	Name	UniProt Accessions	Type	Organism
<input type="checkbox"/>	3.3e-143	98.5	98.5	396.356	1017	201	CHEMBL3100	Plasma retinol-binding protein	P02753	SINGLE PROTEIN	Homo sapiens

Associated Bioactivities



Associated Assays

