

Ralph Goguanco

A15937787

BIMM-143: FIND-A-GENE PROJECT

[Q1] Name: Adenosine Deaminase

Accession: NP_000013

Species: Homo Sapiens

[Q2] Blast Method: TBLASTN search against Armadillidium

Database: Expressed Sequence Tags (est)

Organism: Armadillidium (taxid:13346)

Limits Applied:

BLAST® » tblastn Home Recent Results Saved Strategies Help

blastnblasttblastntblastn

Translated BLAST: tblastn

TBLASTN search translated nucleotide databases using a protein query. more... Reset page Bookmark

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) ? Clear

NP_000013

Query subrange ?

From

To

Or, upload file

No file chosen ?

Job Title

NP_000013:adenosine deaminase isoform 1 |Homo...

Enter a descriptive title for your BLAST search ?

☐ Align two or more sequences ?

Choose Search Set

Database *

Expressed sequence tags (est) ?

Organism Optional

Armadillidium (taxid:13346) ? ☐ exclude Add organism

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown ?

Exclude Optional

☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Limit to Optional

☐ Sequences from type material

Entrez Query Optional

YouTube Create custom database

Enter an Entrez query to limit search ?

BLAST

Search database est using Tblastn (search translated nucleotide databases using a protein query)

☐ Show results in a new window

Note: Parameter values that differ from the default are highlighted in yellow and marked with ♦ sign

i Your search is limited to records that include: Armadillidium (taxid:13346)

Job Title	NP_000013:adenosine deaminase isoform 1 [Homo...
RID	2U286TM0016 Search expires on 03-14 04:56 am Download All v
Program	TBLASTN Citation v
Database	est See details v
Query ID	NP_000013.2
Description	adenosine deaminase isoform 1 [Homo sapiens]
Molecule type	amino acid
Query Length	363
Other reports	

Filter Results

Organism only top 20 will appear ☐ exclude

+ Add organism

Percent Identity to

E value to

Query Coverage to

Filter **Reset**

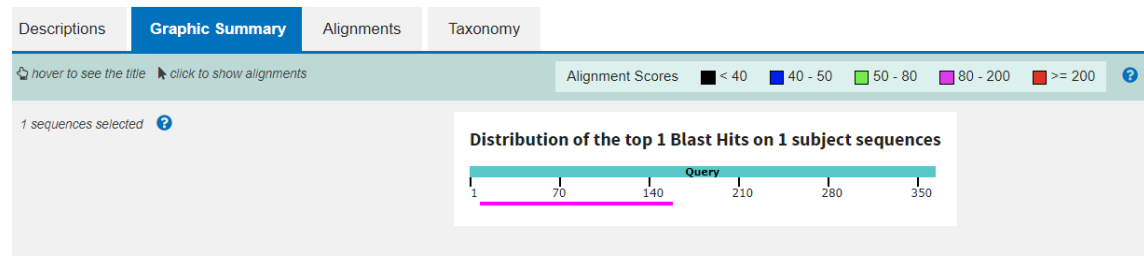
Descriptions	Graphic Summary	Alignments	Taxonomy
--------------	-----------------	------------	----------

Sequences producing significant alignments
Download v Select columns v Show 100 v ?

☒ select all 4 sequences selected [GenBank](#) [Graphics](#)

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	FQ906221 Armadillidium vulgare normalized library Arma...	Armadillidium vu...	108	108	41%	8e-29	40.59%	554	FQ906221.1
<input checked="" type="checkbox"/>	FQ893609 Armadillidium vulgare SSH library (adults challenged by S. typhimurium vs. not challenged) Arma...	Armadillidium vu...	92.4	142	40%	4e-22	49.48%	736	FQ893609.1
<input checked="" type="checkbox"/>	FQ893177 Armadillidium vulgare SSH library (asymbiotic vs. symbiotic) Arma...	Armadillidium vu...	59.7	59.7	20%	5e-11	40.79%	555	FQ893177.1
<input checked="" type="checkbox"/>	FQ896732 Armadillidium vulgare normalized library Arma...	Armadillidium vu...	49.7	49.7	26%	2e-07	29.47%	656	FQ896732.1

Alignment Statistics: E value = 8e-29, Total Score = 108, Percent Identity = 40.59%, Length 554



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

[Next](#)
[Previous](#)
[Descriptions](#)

FQ906221 Armadillidium vulgare normalized library Armadillidium vulgare cDNA, mRNA sequence

Sequence ID: [FQ906221.1](#) Length: 554 Number of Matches: 1

Range 1: 56 to 550 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method		Identities	Positives	Gaps	Frame
108 bits(271)	8e-29	Compositional matrix adjust.		69/170(41%)	96/170(56%)	25/170(14%)	+2
Query 9	KPKVELHVHLDGSIK	PETILYYGRRRGIALPAN-	TAEGLINVGMDKPLTLPDFLAKFDY	67			
Sbjct 56	K+VELHVHLDG+I+PETI	R++ +ALP + + E L + +++P L FL F	KCRVELHVHLDGAIRPETIWEILRKQALPGDGSLEALKKSLIVNEPRDLLFLFEPFSL	235			
Query 68	YMPAIAAGCREAIIKRIAYEFVEMKAKEGVVYEVRYSPHLL---	ANSKVEPIPWIIQAEGDL	124				
Sbjct 236	Y PA G EAI+R A+EFVE ++KEGV Y E R PHLL + S +V + I N E	YHPAFKGDLEAIERFAFEFVEDQSGEKGAVCEARLCPHLLLPDSPSTVDII--NNVE---	400				
Query 125	TPD-----	EVVALVGQGLQGEERDGVGKARSILCCMRHQ	158				
Sbjct 401	TP+ +V V G++ GE + K R ILCC+R +	TPNTVSHQTTTNNSTVDARKIVEAVLRGIERGEEKYKTVRLILCCRISK	550				

Alignment details:

>FQ906221 Armadillidium vulgare normalized library Armadillidium vulgare cDNA, mRNA sequence

Sequence ID: FQ906221.1 Length: 554

Range 1: 56 to 550

Score:108 bits(271), Expect:8e-29,

Method:Compositional matrix adjust.,

Identities:69/170(41%), Positives:96/170(56%), Gaps:25/170(14%)

Query 9 KPKVELHVHLDGSIKPETILYYGRRRGIALPAN-TAEGLLNVIGMDKPLTLPDFLAKFDY 67

K +VELHVHLDG+I+PETI R++ +ALP++EL +++++P L FL F

Sbjct 56 KCRVELHVHLDGAIRPETIWEILRKKQLALPGDGSLEALKKSLIVNEPRDLLLFLEPFSL 235

Query 68 YMPAIAAGCREAIKRIAYEFVEMKAKEGVVYVEVRYSPHLL---ANSKVEPIPWNQAEGDL 124

Y PA G EAI+R A+EFVE ++KEGV Y E R PHLL +S V+I N E

Sbjct 236 YHPAFKGDLEAIERFAFEFVEDQSKEGVAYCEARLCPHLLLPDSPSTVDII--NNVE--- 400

Query 125 TPD-----EVVALVGQGLQEGERDFGVKARSILCCMRHQ 158

TP+ ++V V +G++ GE + K R ILCC+R +

Sbjct 401 TPNTVSHQTTTNNSTVDARKIVEAVLRGIERGEEKYSTKVRLILCCIRSK 550

[Q3] Chosen sequence:

```
>56-550_1 Armadillidium vulgare normalized library Armadillidium vulgare cDNA, mRNA
sequence
KCRVELHVLHDGAIRPETIWEILRKKQLALPGDGSLEALKKSLIVNEPRDLLLFLEPFSL
YHPAFKGDLEAIERFAFEFVEDQSKEGVAYCEARLCPHLLLPDSPSTVDIINNVEPTNTV
SHQTTNNSTVDARKIVEAVLRGIERGEEKYSTKVRLILCCIRSK
```

Name: Adenosine deaminase

Species: Armadillidium vulgare

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Crustacea; Multicrustacea; Malacostraca; Eumalacostraca; Peracarida; Isopoda; Oniscidea; Crinocheta; Armadillidiidae; Armadillidium

[Q4] Details:

After A BLASTP search against the NR database of the protein we have produced from translating our results from the TBLASTN search of NP_000013 on the Armadillidium organisms as shown on the figure below, we got a top results of only 98.18% Percent Identity.

Standard Protein BLAST

blastn **blastp** blastx tblastn tblastx

BLASTP programs search protein databases using a protein query. more...

Reset page
Bookmark

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#) Clear

Query subrange [?](#)

From

To

Or, upload file No file chosen [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

Choose Search Set

Databases ☒ Standard databases (nr etc.): **New** ☐ Experimental databases

[Try experimental clustered nr database](#) [?](#)
For more info see [What is clustered nr?](#)

Standard

Database ?

Organism Optional ☐ exclude [Add organism](#)

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown. [?](#)

Exclude Optional ☐ Models (XM/XP) ☐ Non-redundant RefSeq proteins (WP) ☐ Uncultured/environmental sample sequences

Compare ☐ Select to compare standard and experimental database [?](#)

Program Selection

Algorithm ☐ Quick BLASTP (Accelerated protein-protein BLAST) ☒ **blastp (protein-protein BLAST)** ☐ PSI-BLAST (Position-Specific Iterated BLAST) ☐ PHI-BLAST (Pattern Hit Initiated BLAST) ☐ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm [?](#)

BLAST Search database nr using **Blastp (protein-protein BLAST)**

☐ Show results in a new window

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignmentsDownloadSelect columnsShow100

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"/>	Adenosine deaminase [Armadillidium vulgare]	Armadillidium vulgare	333	333	100%	9e-112	98.18%	404	RXG70787.1
<input checked="" type="checkbox"/>	Adenosine deaminase [Armadillidium nasatum]	Armadillidium nasatum	326	326	98%	1e-108	96.93%	416	KAB7506781.1
<input checked="" type="checkbox"/>	Adenosine deaminase [Armadillidium nasatum]	Armadillidium nasatum	268	268	100%	4e-86	77.58%	414	KAB7506780.1
<input checked="" type="checkbox"/>	Adenosine deaminase [Armadillidium vulgare]	Armadillidium vulgare	268	268	100%	1e-85	77.58%	414	RXG70786.1
<input checked="" type="checkbox"/>	adenosine deaminase-like isoform X2 [Penaeus japonicus]	Penaeus japonicus	201	201	99%	5e-60	58.54%	375	XP_042867791.1
<input checked="" type="checkbox"/>	adenosine deaminase-like isoform X1 [Penaeus japonicus]	Penaeus japonicus	200	200	99%	1e-59	58.54%	410	XP_042867782.1
<input checked="" type="checkbox"/>	adenosine deaminase-like [Penaeus monodon]	Penaeus monodon	198	198	99%	3e-59	57.93%	375	XP_037803349.1
<input checked="" type="checkbox"/>	adenosine deaminase-like [Penaeus vannamei]	Penaeus vannamei	196	196	99%	5e-58	56.55%	416	XP_027238533.1
<input checked="" type="checkbox"/>	adenosine deaminase-like isoform X2 [Homarus americanus]	Homarus americanus	184	184	99%	2e-53	53.80%	390	XP_042229091.1
<input checked="" type="checkbox"/>	adenosine deaminase-like isoform X1 [Homarus americanus]	Homarus americanus	183	183	99%	9e-53	53.80%	424	XP_042229090.1

Download

GenPept

Graphics

Next

Previous

Descriptions

Adenosine deaminase [Armadillidium vulgare]

Sequence ID: [RXG70787.1](#) Length: 404 Number of Matches: 1

Range 1: 13 to 177

GenPept

Graphics

Next Match

Previous Match

Score	Expect	Method	Identities	Positives	Gaps
333 bits(855)	9e-112	Compositional matrix adjust.	162/165(98%)	163/165(98%)	0/165(0%)
Query 1	KCRVELHVHLDGAIRPETIWEILRKKQLALPGDGSLEALKKSLIVNEPRDLLLFLEPFSL				60
Sbjct 13	KCRVELHVHLDGAIRPETIWEILRKKQL+LPGDGSLEALKKSLIVNEPRDLLLFLEPFSL				72
Query 61	YHPAFKGDLEAIERFAFEFVEDQSKGVAYCEARLCPHLLLPDSPSTVDIINNVPNTV				120
Sbjct 73	YHPAFKGDLEAIERFAFEFVEDQSKGVAYCEARLCPHLLLPDSPSTVDIINNVPNTV				132
Query 121	SHQTTNNSTVDARKIVEAVLRGIERGEEKYSTKVRILILCCIRSK			165	
Sbjct 133	SHQTTNNSTVDARKTVEAVLRGIERGEEKYSIKVRILILCCIRSK			177	

Download

GenPept

Graphics

Next

Previous

Descriptions

Adenosine deaminase [Armadillidium nasatum]

Sequence ID: [KAB7506781.1](#) Length: 416 Number of Matches: 1

Range 1: 16 to 178

GenPept

Graphics

Next Match

Previous Match

Score	Expect	Method	Identities	Positives	Gaps
326 bits(835)	1e-108	Compositional matrix adjust.	158/163(97%)	160/163(98%)	0/163(0%)
Query 3	RVELHVHLDGAIRPETIWEILRKKQLALPGDGSLEALKKSLIVNEPRDLLLFLEPFSLYH				62
Sbjct 16	+VELHVHLDGAIRPETIWEILRKKQ ALPGDGSLEALKKSLIVNEPRDLLLFL+PFSLYH				75
Query 63	PAFKGDLEAIERFAFEFVEDQSKGVAYCEARLCPHLLLPDSPSTVDIINNVPNTVSH				122
Sbjct 76	PAFKGDLEAIERFAFEFVEDQSKGVAYCEARLCPHLLLPDSPSTVDIINNVPNTVSH				135
Query 123	QTTNNSTVDARKIVEAVLRGIERGEEKYSTKVRILILCCIRSK			165	
Sbjct 136	QTTNNNS VD RKIVEAVLRGIERGEEKYSTKVRILILCCIRSK			178	

Download

GenPept

Graphics

Next

Previous

Descriptions

This isn't 100% identity, so results indicate that our match is a novel protein.

[Q5] Sequence for alignment:

>56-550_1 Armadillidium vulgare normalized library Armadillidium vulgare cDNA, mRNA sequence

KCRVELHVHLDGAIRPETIWEILRKKQLALPGDGSLEALKKSLIVNEPRDLLLFLEPFSL
YHPAFKGDLEAIERFAFEFVEDQSKEGVAYCEARLCPHLLLPDSPSTVDIINNVPNTV
SHQTTNNSTVDARKIVEAVLRGIERGEEKYSTKVRLLILCCIRSK

>RXG70787.1:13-177 Adenosine deaminase [Armadillidium vulgare]

KCRVELHVHLDGAIRPETIWEILRKKQLSLPGDGSLEALKKSLIVNEPRDLLLFLEPFSLYHPAFKGDLEAIERFAFEFV
EDQSKEGVAYCEARLCPHLLLPDSPSTVDIINNVPNTVSHQTTNNSTVDARKTVEAVLRGIERGEEKYSIKVRLLILC
CIRSK

Re-labeled sequences for alignment:

>Caridean Shrimp | XP_042867791.1:13-171 adenosine deaminase-like isoform X2 [Penaeus japonicus]

KCRVQLHIHLDGAIRHETIWELLRKKKLPLPGSGSLEDLKNALKVQEPQDLGFFLHGFTIFMPAFVGDLEAIERISHEFV
EDQAKESVAYCETRFCPHLFLPDSAHPDYLTSEVNGTAEGNGGNVTIDDLKAVLKGLKRGEEEFGTKVRVILCCIRG

>Giant tiger prawn | XP_037803349.1:13-171 adenosine deaminase-like [Penaeus monodon]

KCRVQLHIHLDGAIRHETIWELLRKKKLPLPGSGSLEDLKNALKVQEPQDLGFFLHGFTIFMPAFAGDLEAIERISHEFV
EDQAKESVAYCETRFCPHLFLPDIAHQPDYLTSEVNGTAEGNSGSVTIDDLKAVLKGLRRGEEEFGTKVRVILCCIRG

>American Lobster | XP_042229091.1:13-183 adenosine deaminase-like isoform X2 [Homarus americanus]

KCRVQLHVHLDGCIRHETIWEVMRRKGLKPLPGSGSLADLKLALQVQEPEDLLFFLSGFKIFLPAIKGDLAVIERIAHEFV
EDQAKEYVAYCEARFCPHLLLPNDTTQFSVVKCLKGKSRVNGTTDTTDEDGPVKDGEVTVDSILNAVKGFSRGEEDF
GT
KVRVILCCIHG

>Louisiana crawfish | XP_045625303.1:13-179 adenosine deaminase-like isoform X3 [Procambarus clarkii]

KCRVQLHVHLDGAIRHETLWEVMRHKGMKLPGRGSLADLKTAVQVQEPKDLGLFLRGFQIFLPAIVGDLAVIERIAYEF
V
EDQANDSVAYCEARFSPHLLLPSEQSQPNLHNEAEVQLNGTVVGSNGDSINNEVTVDLSILAVLKGLARGEEDFGTKVR
V
IICCIRG

>>Gooseneck Barnacle | XP_037080455.1:12-151 adenosine deaminase-like [Pollicipes pollicipes]

KCRVHLHVHLDGAVRPSTVWELSRQKKLLLPNGSLKDLEEALQIHSPRDLAHFLQKFRWISPAIVGDLAAIERIAYEFV
EDEARQAVLYTEARFSPHLLVGGPGQVTARQVTEAVMRGMARGERQFGVTARLLLCIRG

Alignment:

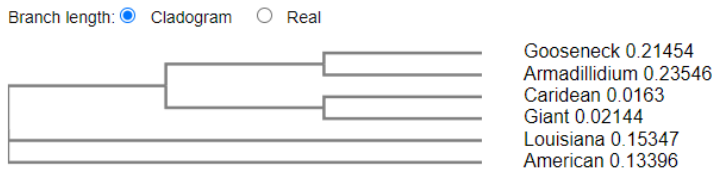
Obtained using MUSCLE (version 3.8) at EBI:

Gooseneck	KCRVHLHVHLDGAVRPSTVWELSRQKKLLLPNGSLKDLEEALQIHSPRDLAHFLQKFRW
Armadillidium	KCRVELHVHLDGAIRPETIWEILRKKQLSLPGDGSLEALKKSLIVNEPRDLLLFLEPFSL
Caridean	KCRVQLHIHLDGAIRHETIWELLRKKKLPLPGSGSLEDLKNALKVQEPQDLCFFLHGFTI
Giant	KCRVQLHIHLDGAIRHETIWELLRKKKLPLPGSGSLEDLKNALKVQEPQDLGFFLHGFTI
Louisiana	KCRVQLHVHLDGAIRHETLWEVMRHKGMKLPGRGSLADLKTAVQVQEPKDLGLFLRGFQI
American	KCRVQLHVHLDGCIRHETIWEVMRRKGLKLPGSGSLADLKLALQVQEPEDLLFFLSGFKI
	**** *:****.:* .*:**:*.*:*** ***:*::::.* ** *
Gooseneck	ISPAIVGDLAAIERIAYEFVEDEARQAVLYTEARFSPHLLV-----
Armadillidium	YHPAFKGDLEAIERFAFEFVEDQSKEGVAYCEARLCPHLLLPDSPSTVDIINNVEPNTV
Caridean	FMPAFVGDLEAIERISHEFVEDQAKESVAYCETRFCPHLFLPDSAHQPDYLT-----SEV
Giant	FMPAFAGDLEAIERISHEFVEDQAKESVAYCETRFCPHLFLPDIAHQPDYLT-----SEV
Louisiana	FLPAIVGDLAVIERIAYEFVEDQANDSVAYCEARFSPHLLLPSEQSQPNLHN--EAEVQL
American	FLPAIKGDLAVIERIAHEFVEDQAKKEYVAYCEARFCPHLLLPNDTTQFSVVKCLKGKSRV
	:* * .****:..*****::: * * *:*:..****:
Gooseneck	-----GP-----GGQVTARQVTEAVMRGMARGERQFGVTARLLLC CIRG-
Armadillidium	SHQTTTN-----NSTVDARKTVEAVLRGIERGEEKYSIKVRLILCCIRSK
Caridean	NGTAEGN-----GGNVTIDDVLKAVLKGLKRGEFEFGTKVRVILCCIRG-
Giant	NGTAEGN-----SGSVTIDDVLKAVLKGLRRGEFEFGTKVRVILCCIRG-
Louisiana	NGTVVGS--NGDSINNEVTVDASILIAVLKGLARGEEDFGTKVRVILCCIRG-
American	NGTTDTTTEDGPVKDGEVTVDSILNAVLKGFSRGEEDFGTKVRVILCCIHG-
	.. * . ***:*: *** ..:..*:::****..

[Q6] Phylogenetic Tree

Phylogenetic Tree

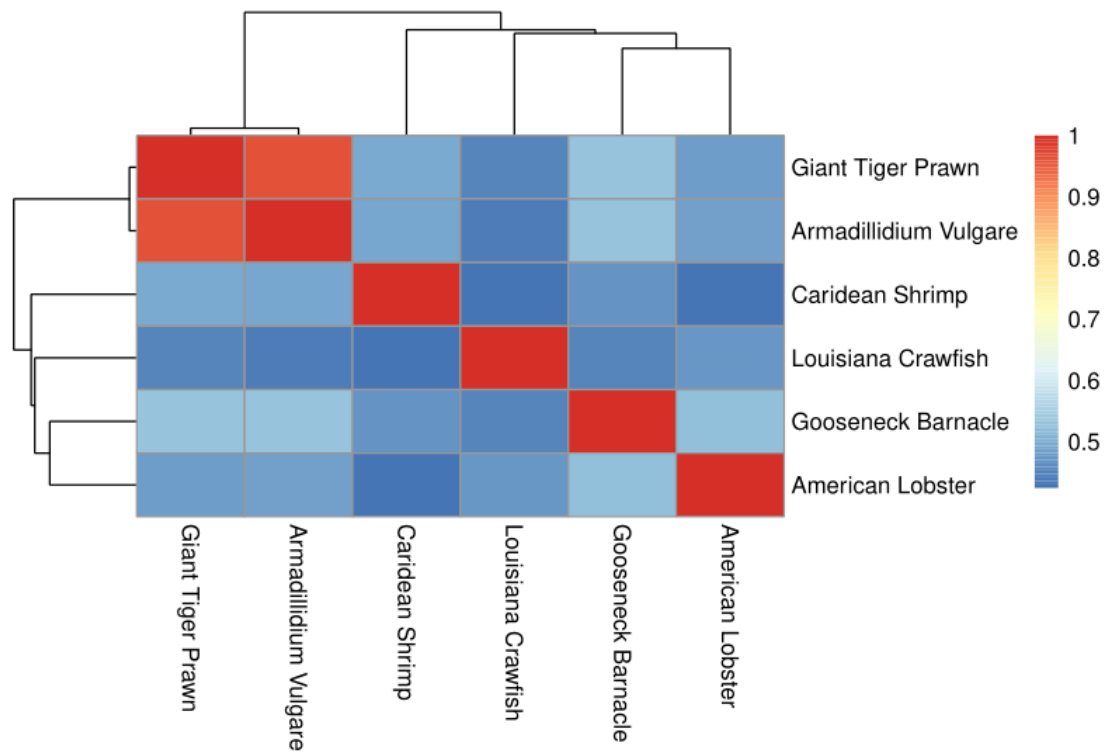
This is a Neighbour-joining tree without distance corrections.



Tree Data

```
(
(
(
Gooseneck:0.21454,
Armadillidium:0.23546)
:0.04329,
(
Caridean:0.01630,
Giant:0.02144)
:0.12639)
:0.02549,
Louisiana:0.15347,
American:0.13396);
```

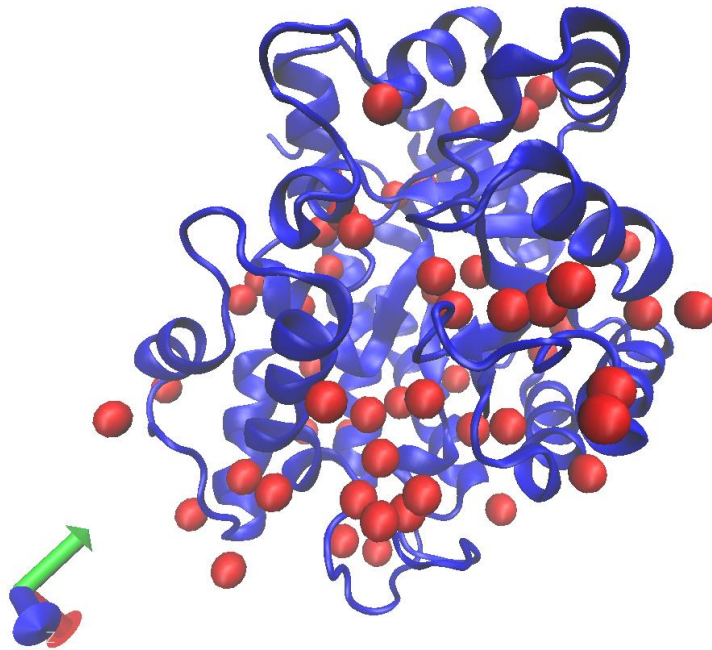
[Q7] Heatmap figure



[Q8]

ID	Technique	Resolution	Source	Evalue	Percent Identity
1FKX	X-RAY DIFFRACTION	2.4	Mus Musculus	3E-28	42.33
1NDV	X-RAY DIFFRACTION	2.3	Bos Taurus	2E-24	41.07
6N91_A	X-RAY DIFFRACTION	2.05	Vibrio Cholerae	2E-07	37.11

[Q9] Molecular Figure of One Identified PDB Structure Using VMD

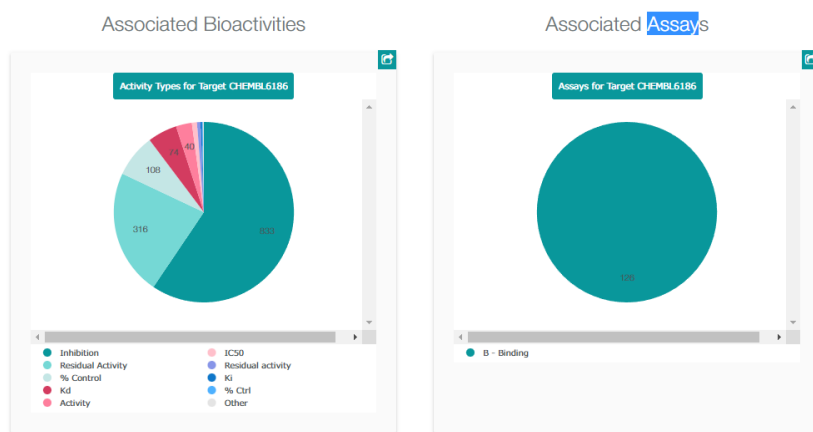


Not likely to be similar because the structure of the *Mus musculus* protein is only 42.33% similar to our novel protein.

[Q10]

A target search of our novel protein details one single protein type named Serine/Threonine-protein kinase Sgk3 that belongs to Homo sapiens. Activity Charts in https://www.ebi.ac.uk/chembl/target_report_card/CHEMBL6186/ states that the protein is mostly used for Inhibition, Residual Activity and Control. Associated Assays suggest that 100% of it is for B – Binding. Ligand Efficiency Plot for Target CHEMBL6186 is provided in a figure below

Activity Charts



Ligand Efficiencies

