

Name : Tianhao Qiu
PID: A92066783
Email : t6qiu@ucsd.edu
BIMM143 Project

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.

Name: Potassium voltage-gated channel subfamily A member 3
Accession: NP_002223.3
Species: Homo sapiens

[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 (2.9.0) **Database: Expressed Sequence Tags (est)**
Organism: Nematodes(Taxid: 6239)

BLAST » tblastnHomeRecent ResultsSaved StrategiesHelp

Translated BLAST: tblastn

blastnblasttblasttblastx

Enter Query Sequence

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

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

>NP_002223.3 potassium voltage-gated channel subfamily A member 3 [Homo sapiens]
MDERLSLLRSPPPPSARHRAHPPQRPASSGGGHTLVNHGYAEPAAGRELPPDMTVVPGDHLLPE
VADGG
GAPPQGGCGGGGCDRYEPLPPLPAAGEQDCGERVVINISGLRFETQLKTLQCFPETLLGDPKRR
MRYF

Query subrange

From

To

Or, upload file

Choose Fileno file selected

Job Title

NP_002223.3 potassium voltage-gated channel...
Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database

Expressed sequence tags (est)

Organism

Optional

nematode (taxid:6239)☐ exclude

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

YouTubeCreate custom database

Enter an Entrez query to limit search

BLAST

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

☒ Show results in a new window

[Algorithm parameters](#)

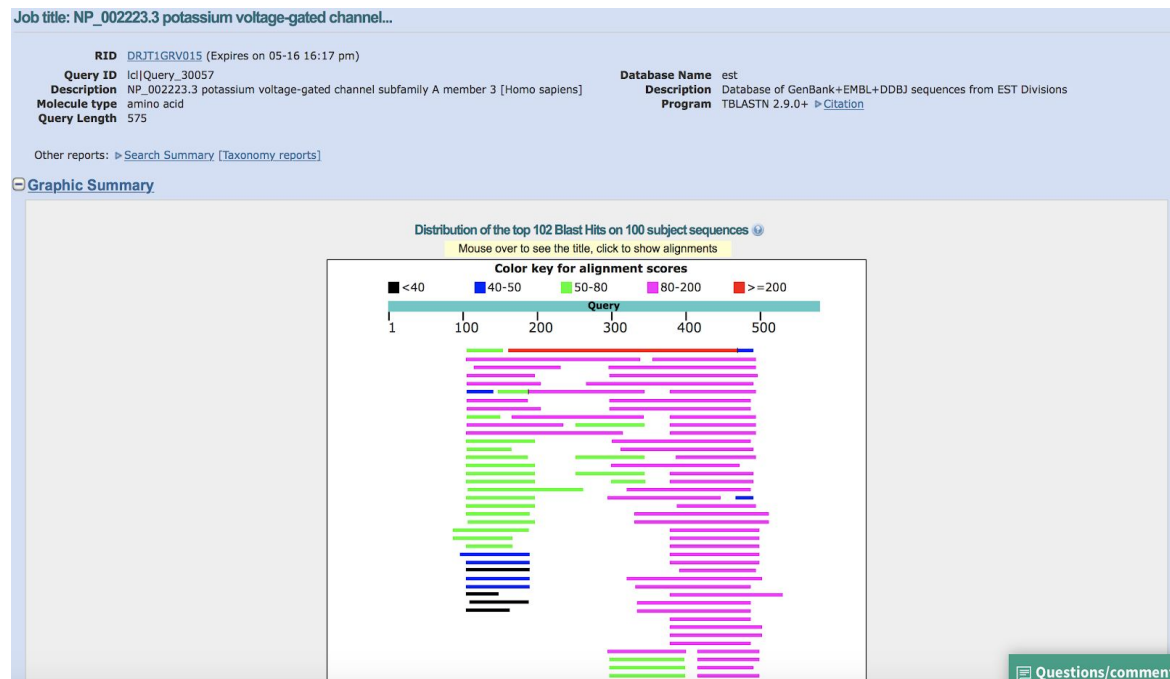
Note: Parameter values that differ from the default are highlighted in yellow and marked with + sign

We are beta testing a New Results page

☐ Click here if you would like to see your results in the new format. You can always switch back to the Traditional Results page.

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: Accession FN892626.1, a 1311 base pair clone from *C.Elegans* See below for alignment details.



Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input type="checkbox"/> FN892626 Caenorhabditis elegans whole worm N2 all stages Caenorhabditis elegans cDNA clone ab3100_c2_20070602.E06_9758164_5_GR3QTR_070531.f.mRNA sequence	211	258	56%	9e-62	37.38%	FN892626.1
<input type="checkbox"/> FN892628 Caenorhabditis elegans whole worm N2 all stages Caenorhabditis elegans cDNA clone ab3100_c2_20030813.E06_9758164_SL1_3_030813.r.mRNA sequence	191	191	40%	2e-56	40.34%	FN892628.1
<input type="checkbox"/> QSTR30005E07 pDONR201 Entry vector Caenorhabditis elegans cDNA, mRNA sequence	174	174	34%	2e-49	44.16%	CV123089.1
<input type="checkbox"/> QSTR30005F06 pDONR201 Entry vector Caenorhabditis elegans cDNA, mRNA sequence	163	163	34%	4e-45	41.21%	CV123100.1
<input type="checkbox"/> FN881339 Caenorhabditis elegans whole worm N2 all stages Caenorhabditis elegans cDNA clone ab3100_c2_20060527.B12_8305806_5_GR3QTR_060526.f.mRNA sequence	145	145	38%	1e-37	39.66%	FN881339.1
<input type="checkbox"/> BJ793488 unpublished oligo-capped cDNA library Caenorhabditis elegans cDNA clone yk1676g02 3', mRNA sequence	134	134	24%	5e-35	46.38%	BJ793488.1
<input type="checkbox"/> BJ141226 unpublished oligo-capped cDNA library, C. elegans L1 stage Caenorhabditis elegans cDNA clone yk1168a01 3', mRNA sequence	133	133	20%	6e-35	53.04%	BJ141226.1
<input type="checkbox"/> BJ781564 unpublished oligo-capped cDNA library Caenorhabditis elegans cDNA clone yk1500f10 3', mRNA sequence	134	134	32%	6e-35	37.17%	BJ781564.1
<input type="checkbox"/> BJ781447 unpublished oligo-capped cDNA library Caenorhabditis elegans cDNA clone yk1499c12 3', mRNA sequence	134	134	32%	6e-35	37.17%	BJ781447.1
<input type="checkbox"/> BJ784946 unpublished oligo-capped cDNA library Caenorhabditis elegans cDNA clone yk1540h10 3', mRNA sequence	134	134	20%	7e-35	53.04%	BJ784946.1
<input type="checkbox"/> BJ140991 unpublished oligo-capped cDNA library, C. elegans L1 stage Caenorhabditis elegans cDNA clone yk1164g03 3', mRNA sequence	134	134	20%	9e-35	53.04%	BJ140991.1
<input type="checkbox"/> BJ133820 unpublished oligo-capped cDNA library, C. elegans L1 stage Caenorhabditis elegans cDNA clone yk1081d09 3', mRNA sequence	134	134	20%	9e-35	53.04%	BJ133820.1
<input type="checkbox"/> BJ783576 unpublished oligo-capped cDNA library Caenorhabditis elegans cDNA clone yk1524c12 3', mRNA sequence	132	132	32%	3e-34	37.43%	BJ783576.1
<input type="checkbox"/> FN892625 Caenorhabditis elegans whole worm N2 all stages Caenorhabditis elegans cDNA clone ab3100_c2_20021105.E06_9758164_021105.r.mRNA sequence	102	159	34%	3e-32	33.33%	FN892625.1
<input type="checkbox"/> FN881340 Caenorhabditis elegans whole worm N2 all stages Caenorhabditis elegans cDNA clone ab3100_c2_20051114.B12_8305806f_5_dT35V_051114.f.mRNA sequence	128	128	30%	2e-31	40.66%	FN881340.1
<input type="checkbox"/> FN892624 Caenorhabditis elegans whole worm N2 all stages Caenorhabditis elegans cDNA clone ab3100_c2_20021105.E06_9758164_021105.f.mRNA sequence	123	123	30%	6e-31	36.93%	FN892624.1
<input type="checkbox"/> BJ793280 unpublished oligo-capped cDNA library Caenorhabditis elegans cDNA clone yk1674c01 3', mRNA sequence	119	119	18%	4e-30	50.47%	BJ793280.1
<input type="checkbox"/> BJ147254 unpublished oligo-capped cDNA library, C. elegans L1 stage Caenorhabditis elegans cDNA clone yk1244g01 3', mRNA sequence	119	119	29%	7e-30	37.93%	BJ147254.1
<input type="checkbox"/> QSTR185D3_2 AD-wrmcDNA Caenorhabditis elegans cDNA, mRNA sequence	118	118	19%	8e-30	50.88%	CB397258.1
<input type="checkbox"/> QSTR185D3_1 AD-wrmcDNA Caenorhabditis elegans cDNA, mRNA sequence	118	118	19%	8e-30	50.88%	CB397221.1
<input type="checkbox"/> QSTR143H5_1 AD-wrmcDNA Caenorhabditis elegans cDNA, mRNA sequence	117	117	28%	8e-30	35.71%	CB394726.1
<input type="checkbox"/> BJ106154 unpublished oligo-capped cDNA library, C. elegans L1 stage Caenorhabditis elegans cDNA clone yk1081d09 5', mRNA sequence	115	115	26%	3e-29	37.75%	BJ106154.2
<input type="checkbox"/> AU214472 unpublished oligo-capped cDNA library, stage L2 Caenorhabditis elegans cDNA clone yk813e07 3', mRNA sequence	116	116	18%	4e-29	50.00%	AU214472.1
<input type="checkbox"/> C71245 Yuiji Kohara unpublished cDNA Caenorhabditis elegans cDNA clone yk442g4 5', mRNA sequence	112	112	20%	2e-28	45.69%	C71245.1
<input type="checkbox"/> QSTR182G8_2 AD-wrmcDNA Caenorhabditis elegans cDNA, mRNA sequence	112	112	31%	1e-28	45.69%	C71245.1

Download ▾ GenBank Graphics Sort by: E value ▾

▼ Next ▲ Previous ▲ Descriptions

FN892626 *Caenorhabditis elegans* whole worm N2 all stages *Caenorhabditis elegans* cDNA clone abi3100_ce2_20070602.E06_9758164_5_GR3OTR_070531.f, mRNA sequence

Sequence ID: [FN892626.1](#) Length: 1311 Number of Matches: 2

Range 1: 4 to 876 [GenBank](#) [Graphics](#) ▾ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
211 bits(538)	9e-62	Compositional matrix adjust.	140/305(46%)	204/305(66%)	14/305(4%)	+1
Query	161	YYYQSGGRIRRPVNPVIDIFSEEIRFYQLGEEAMEKFREDEGFlreerplprDFQRQV	220			
		YY GR++RP VPIDIF +E+RF+Q+G++ +E+F EG+ + +E +P QR++				
Sbjct	4	YYSYVCGRVKRPEIVPIDIFLKEMRFFQMGDDLLBEFWIAEGYEKPEVMMPNKTRKI	183			
Query	221	WLLFEYPSSGPGARGIAivsvlvilislviFCLETLPFRDEKDYPASTSQDSFEAAGNS	280			
		W L EYP+SS AR IA +S+ VI +SI+ FC ET+P +EK S + + +				
Sbjct	184	WELMEYPPDSSLARIAFIASIAVIALSIISFCWETVPDSIEEKPINNSATAELLDENDEK	363			
Query	281	TSGSRAGASSFSDDPFVVFETLCIIWFSEFELLVRFACPSKATFSRNIMNLIDIVAIPIYF	340			
		PFF +E +CI+WF+ EL++RF +CP K TF+ +++N+ID VAI P+P				
Sbjct	364	HYS-----PFFWIELMCILWFTEILRLRFISCPCKVTFATSVLNIIDFVAIAPFF	513			
Query	341	ITLGTLEAERQNGQQAMSLailrvlrvfrifKLSRHSKGLQLGQTLKASMRElgI	400			
		+ +MS A+LRV+RLRVRF+FKLSRHS GLQLLG+T ++S++E L				
Sbjct	514	VNF----FFADTSKSNSSMSFAVLRLVLRVRFVFKLSRHSVGLQLGKTPRSSVQEFCL	684			
Query	401	liffllfigvllfSSAVYFAEADDPSTSGFSSIPDAFWNAVVTMTTVGYGDMHPVTIGGKIV	460			
		LIFF+ I ++L+F+S +YFAE +P S F+SIP +FW+ +VTMTTVGYGD+ P++ GK V				
Sbjct	685	LIFFMAIALVLFASGMFYFAEQGEPSKFTSIPASFWFVLVMTTVGYGDLVPLSPFGK-V	861			
Query	461	GSLCA 465				
		++CA				
Sbjct	862	WAVCA 876				

Related Information

[UniGene](#) - clustered expressed sequence tags

Range 2: 855 to 944 [GenBank](#) [Graphics](#) ▾ Next Match ▲ Previous Match ▲ First Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
46.2 bits(108)	9e-62	Compositional matrix adjust.	19/30(63%)	25/30(83%)	0/30(0%)	+3
Query	458	KIVGSLCAIAGVLTIALPVPVIVSNFNFFY	487			
		K VG +CA+ GVLT+ALPVP+IV+ F +FY				
Sbjct	855	KSVGGMCAIMIGVLTIALPVPVIVAKFKHFF	944			

Alignment Details :

FN892626 *Caenorhabditis elegans* whole worm N2 all stages
Caenorhabditis elegans cDNA clone
abi3100_ce2_20070602.E06_9758164_5_GR3OTR_070531.f, mRNA
sequence

Length = 1311,

Score = 211 bits (538) for Range 1 and 46.2 bits (108) for Range 2

Expect = 9e-62

Method = Compositional matrix adjust.

Identities = 140/305(46%) for Range 1 and 19/30(62%) for Range 2

Positives = 204/305(66%) for Range 1 and 25/30(83%) for Range 2

Gaps = 14/3055(3%) for Range 1 and 0/30(0%) for Range 2

Frame = +1 for Range 1 and +3 for Range 2

Range 1:

Query 161 YYYQSGGRIRRPVNPIDIFSEEIRFYQLGEEAMEKFREDEGFfreeerplprDFQRQV 220
YY GR++RP VPIDIF +E+RF+Q+G++ +E+F EG+ + +E +P QR++
Sbjct 4 YYSYVCGRVKRPEIVPIDIFLKEMRFFQMGGDLLLEEFWIAEGYEKPKEVMMPNNKTQRKI 183

Query 221 WLLFEYPESSGPARGIAivsvlivilisiviFCLETLPEFRDEKDYPASTSQDSFEAAGNS 280
W L EYP+SS AR IA +S+VI+SI+ FC ET+P +EK S + + +
Sbjct 184 WELMEYPDSSLSARIIAFISIAVIALSIISFCWETVPSDIEEKPINNSATAELLDEMDEK 363

Query 281 TSGSRAGASSFSDPFFVETLCIWFSEFELLVRFFACPSKATFSRNIMNLIDIVAIIPYF 340
PFF +E +CI+WF+ EL++RF +CP K TF+ +++N+ID VAI P+F
Sbjct 364 HYS-----PFFWIELMCILWFTIELILRFISCPCKVTFATSVLNIIDFVAIAPFF 513

Query 341 ITLGTELAERQGNQQAMSLAilrvirlrvfrifKLSRHSKGLQILGQTLKASMRElgl 400
+ +MS A+LRV+RLVRVFR+FKLSRHS GLQILG+T ++S++E L
Sbjct 514 VNF---FFADTSKSNSSMSFAVLRVLRVFRVFKLSRHSVGLQILGKTRSSVQEFCL 684

Query 401 lifffigvilfSSAVYFAEADDPTSGFSSIPDAFWWAVVTMTTVGYGDMHPVTIGGKIV 460
LIFF+ I ++LF+S +YFAE +P S F+SIP +FW+ +VTMTTVGYGD+ P++ GK V
Sbjct 685 LIFFMAIALVLFASGMYFAEQGEPNSKFTSIPASFWVLVTMTTVGYGDLVPLSPFGK-V 861

Query 461 GSLCA 465
++CA
Sbjct 862 WAVCA 876

Range 2:

Query 458 KIVGSLCAIAGVLTIALPVPVIVSNFNIFY 487
K VG +CA+ GVLT+ALPVP+IV+ F +FY
Sbjct 855 KSVGGMCMAMIGVLTALPVPVIAKFKHFY 944

[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.

Chosen Sequence (From Emboss TransSeqResult):

>FN892626.1_1 Caenorhabditis elegans whole worm N2 all stages Caenorhabditis elegans cDNA clone
abi3100_ce2_20070602.E06_9758164_5_GR3OTR_070531.f, mRNA sequence
HYYSYVCGRVKRPEIVPIDIFLKEMRFFQMGGDLLLEEFWIAEGYEKPKEVMMPNNKTQRK
IWELMEYPDSSLSARIIAFISIAVIALSIISFCWETVPSDIEEKPINNSATAELLDEMDE
KHYSPPFFWIELMCILWFTIELILRFISCPCKVTFATSVLNIIDFVAIAPFFVNFFADTS

KSNSSMSFAVLRLVRLVRFVFKLSRHSVGLQILGKTRSSVQEFCLLIFFMAIALVLF
ASGMFYFAEQGEPNSKFTSIPASFWFVLVTMTTVGYGDLVPLSPFGKVWAVCAQ*LVS*HW
HFQFLLLLPNSTSTDQKNGLASMKSKGDDANND*FSGEGL*NCETERTPFAPKARHP
LGGTFGLLSSIKPFQPLTISRALQLCFILKVIFYPTPFSPNFFFLKKRGKTPCGPPGN
SPPPPPPKKILFSPSP

>FN892626.1_2 Caenorhabditis elegans whole worm N2 all stages Caenorhabditis elegans cDNA clone
abi3100_ce2_20070602.E06_9758164_5_GR3OTR_070531.f, mRNA sequence

IIHMCAGE*NDLKLCLTYF*KKCDFSKWVMIFWKNSG*LRVMKNQKK**CQIIKPKEK
YGN*WNIRIVHFPRELLHSFQLLLLHYPSSHSAGKQYHLTSKKNQSIIQQLRNYWMKWTK
NIIPHFSGLSSCVFCGLLLN*FFGLFHVLVK*HLPRLC*ILLTLWQLLRFLTFSRTQA
NPIHRCRLQCSVCSDWFEYSEFNLVDIPLVFKFLEKLSDPLSKSFVFSFSSWQLLWFSL
HQECILRSKGSQTQSSLQFQQASGLYWSQ*QQSDMEIWFR*VHLEKCGRYVRNDWCLDIG
TSSSYCCQIQTLLPTKKTVLLQ*NQRVTMQTTTSHDGVKAFETSVKPKEPHLPRKLAIP
WGEPSCD*APSSPSPSPSPGPSNFVFQF*KFFIFQPLFPPIFFFLKKGGGKPLVAPPEI
PPPPPLQKKFFSLPLPP

>FN892626.1_3 Caenorhabditis elegans whole worm N2 all stages Caenorhabditis elegans cDNA clone
abi3100_ce2_20070602.E06_9758164_5_GR3OTR_070531.f, mRNA sequence

LLFICVRKSETT*NCAD*HIFERNAIFPNG**SSGRILDS*GL*KTKRSDDAK**NPKKN
MGTNGISG*FTFRENYCIHFNCCYCIHHLLGNSTI*HRRKTNQ*FSNCGTIG*NGRK
TLFPIFLD*AHVYFVYYY*INSSVYFMSL*SDICHVCVEYY*LCGNCSVFC*LFLRGHKQ
IQFIDVVCASAPCAPIGSSISQSFQT*STFRWSSNSWKNFPILCPRVLSSHFLHGNCSSGLC
IRNVFCGARGAKLVHFNSSKLLVCTGHNDNSRIWRFGSAESIWKSVGGMCAMIGVLTLA
LPVPIIVAKFKHFYRPPKKRSCFNEIKG*RCKQRLMTG*RPLKQV*NRKNPICPESSPSP
GGNLRIVKLHQALSAPHHLQGPPTLFSNFKSFLFSNPFFPQFFFS*KKGGENPLWPPRKF
PPPPPSKKNSFLSLSPX

>FN892626.1_4 Caenorhabditis elegans whole worm N2 all stages Caenorhabditis elegans cDNA clone
abi3100_ce2_20070602.E06_9758164_5_GR3OTR_070531.f, mRNA sequence

GGRGREKNFFWRGGGGGSGGATRGFPFFFKKKKIGGKRGWKIKNF*NWTKLEGPGDG
EGLKGLDGA*QSEGSPQGMASFRGKWGSFGFTLVSKAFTPS*EVVVCIVTL*FH*SKTVF
LVGRSV*IWQQ**ELEVPMSRHSRLTYRPHFSKWTQRNQISISDCCHCDQYKPEACWN*
SEL*VWLPLLRIHS*CKENQSNCHEENEKTKLLDRGSESFKNLKTNGMSTKFENSEYS
NQSEHTEHCKRHR*IGFACVREEKVNKKRSNCHKVNNIQHRRGKCHFTRT*NKPN*FNS
KPQNTHELNPEKWGIMFFVHFIQ*FRSC*IIDWFFFVVRWYCFPAE*DDG*CNNSN*NEC
NNSRGK*TIRIFH*FPYFSLGFIWHHHFFWFFITLSYPEFFQKIITHLEKSHFFQKYVN
RHNFRSFHSSAH*IIM

>FN892626.1_5 Caenorhabditis elegans whole worm N2 all stages Caenorhabditis elegans cDNA clone
abi3100_ce2_20070602.E06_9758164_5_GR3OTR_070531.f, mRNA sequence

GERERKEFFLEGGGGGNFRGGHKGFSPPFF*EKKNWGKKGLENKLLKLENKVGGPWRW*
GAERA*WSLTIRRFPPGDGELSGQMGGFRFHTCFKGLHPVMRSRCLHRHPLISLKQDRFF
GR*KCLNLATIIGTGSANVKTPIIAHIPPTLFQMDSAEPNLHIRLLSL*PVQTRSLLLELK
*TSLAPLAPQNTFLMQREPEQLP*RK*EDKTLGQRIGKFFQEFEDQRNVD*V*KL*ILE
PIGAHGALQTTSMNWICLCPRRKS*QKTEQLPQSQ*YSTQTWQMSLYKDMK*TEELIQ**

TTKYT*AQSRKMGNVFRPFHPIVQLLNY*LVFLRCQMVLFP SRMR*WIMQ*QQLK*MQ
*FSRKVNYPDIPLVPIFFFGFYFLASSLLL VFHNPQLSRILPEDHHPFGKIAFLSKICQS
AQFQVVSLFRTHMNNX

>FN892626.1_6 *Caenorhabditis elegans* whole worm N2 all stages *Caenorhabditis elegans* cDNA clone
abi3100_ce2_20070602.E06_9758164_5_GR3OTR_070531.f, mRNA sequence

GGEGEKRIFFGGGGGGEFPGGPQGVFPPLFLRKKKLGEKGVGK*KTFKIGKQSWRALEMV
RG*KGLMELNNPKVPPRGWRAFGANGVLSVSHLFQRPSRHEKSLFASSPFDIEARPFF
WSVEVFEFGNNNRNWKCQCQDTNHCAHTAHTFPNGLSGTKSPYPTVVIVTSTNQKLAGIE
VNFEFGSPCSAKYIPDAKRTRAIAMKKMRRQNSWTEDRKVFPRI*RPTECRLSLKTLNTR
TNRSTRSTANDIDELDLLVS AKKKLTKN GAIATKSIIFNTDVANVTLQGHEINRRINSIV
NHKIH MSSIQKNGE*CFSSISSNSSAVAELLIGFSSMSDGTVSQQNEMMDNAITAIEMNA
IILAESELSGYSISSHIFLWVLLFGIITSFGFS*PSAIQNSSRRSSPIWKNRISFKNMSI
GTISGRFTLPHTYE**X

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 of Protein: Not found

Species *Caenorhabditis elegans*

**Eukaryota; Metazoa; Ecdysozoa; Nematoda; Chromadorea;
Rhabditida; Rhabditina; Rhabditomorpha;
Rhabditoidea; Rhabditidae; Peloderinae; *Caenorhabditis*.**

[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 NR database (see setup in the first screen-shot below) yielded a top hit result is to a protein from C.Elegans.

See additional screenshots below for top hits and selected alignment details

The screenshot displays the NCBI BLASTP search interface. At the top, the BLAST logo and navigation links (Home, Recent Results, Saved Strategies, Help) are visible. The main heading is "Standard Protein BLAST". Below this, the "Enter Query Sequence" section contains a text area with a FASTA sequence of a Caenorhabditis elegans cDNA clone. The "Job Title" field is populated with "6 sequences (FN892626.1_1 Caenorhabditis elegans...)". The "Choose Search Set" section shows the "Database" set to "Non-redundant protein sequences (nr)". The "Program Selection" section has "blastp (protein-protein BLAST)" selected as the algorithm. At the bottom, the "BLAST" button is prominent, and the "Show results in a new window" checkbox is checked. A beta testing notice is also present on the right side of the interface.

BLAST[®] » blastp suite [Home](#) [Recent Results](#) [Saved Strategies](#) [Help](#)

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](#)

>FN892626.1 Caenorhabditis elegans whole worm N2 all stages Caenorhabditis elegans cDNA clone abi3100_ce2_20070602.E06_9758164_5_GR3OTR_070531.f, mRNA sequence
HYYSYVCGRVKRPETVPIIDIFLKEMRFFQMGDDLLLEEFWIAEGYEKPKVEMMPNNKTQK
IWELMEYDSSLSARIIFISIAVIALSIISFCWETVPSDIEEKPINNSATAELDEMD
KHYSPPFWIELMCLWFTIELILRFISCPCKVTFATSVLNIIDFVAIAPFFVNIFFADTS

Or, upload file [Choose File](#) no file selected [Upload](#)

Job Title
6 sequences (FN892626.1_1 Caenorhabditis elegans...)
Enter a descriptive title for your BLAST search [Help](#)

☐ Align two or more sequences [Help](#)

Choose Search Set

Database
Non-redundant protein sequences (nr) [Help](#)

Organism
Optional
☐ exclude [+](#)
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown. [Help](#)

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

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 [Help](#)

BLAST Search database nr using Blastp (protein-protein BLAST)
☒ Show results in a new window

We are beta testing a New Results page
☐ Click here if you would like to see your results in the new format. You can always switch back to the Traditional Results page. [New](#)

The top result is to a protein from *Caenorhabditis elegans* with identity lower than 100%, see second screenshot below for alignment details:

Sequences producing significant alignments:

Select: All None Selected:0

Alignments

Download

GenPept

Graphics

Distance tree of results

Multiple alignment

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input type="checkbox"/>	Potassium voltage-gated channel protein shk-1 [Caenorhabditis elegans]	555	555	75%	0.0	89.58%	NP_871936.1
<input type="checkbox"/>	Potassium voltage-gated channel protein shk-1 [Caenorhabditis elegans]	555	555	75%	0.0	89.58%	NP_871935.1
<input type="checkbox"/>	Potassium voltage-gated channel protein shk-1 [Caenorhabditis elegans]	555	555	75%	0.0	89.58%	NP_001040829.1
<input type="checkbox"/>	Potassium voltage-gated channel protein shk-1 [Caenorhabditis elegans]	555	555	75%	0.0	89.58%	NP_496104.2
<input type="checkbox"/>	Potassium voltage-gated channel protein shk-1 [Caenorhabditis elegans]	554	554	75%	0.0	89.58%	NP_001022516.1
<input checked="" type="checkbox"/>	Potassium voltage-gated channel protein shk-1 [Caenorhabditis elegans]	554	554	76%	0.0	88.30%	NP_871934.1
<input type="checkbox"/>	hypothetical protein FL81_09050 [Caenorhabditis remanei]	507	507	75%	3e-175	82.54%	POM48488.1
<input type="checkbox"/>	Hypothetical protein CBG01037 [Caenorhabditis briggsae]	506	506	75%	4e-175	82.48%	XP_002629795.1
<input type="checkbox"/>	hypothetical protein CRE_05172 [Caenorhabditis remanei]	508	508	75%	8e-175	82.54%	XP_003094464.1
<input type="checkbox"/>	hypothetical protein FL82_22701 [Caenorhabditis remanei]	509	509	75%	9e-175	82.54%	OZF88426.1
<input type="checkbox"/>	Protein CBG01037 [Caenorhabditis briggsae]	507	507	75%	1e-174	82.48%	CAP22224.2
<input type="checkbox"/>	hypothetical protein B9Z55_004005 [Caenorhabditis nigoni]	507	507	75%	1e-174	82.48%	PIC43188.1
<input type="checkbox"/>	hypothetical protein Y032_0001g50 [Ansylostoma ceylanicum]	423	423	74%	3e-142	71.86%	EYC34606.1
<input type="checkbox"/>	Potassium channel and ion transport domain containing protein [Haemonchus contortus]	422	422	73%	4e-142	69.97%	CDJ84709.1
<input type="checkbox"/>	hypothetical protein WR25_12724 [Diploscapter pachys]	421	421	74%	4e-141	67.89%	PAV87002.1
<input type="checkbox"/>	hypothetical protein WR25_24211 [Diploscapter pachys]	415	415	74%	9e-139	67.07%	PAV86974.1
<input type="checkbox"/>	unnamed protein product [Nippostrongylus brasiliensis]	409	409	65%	7e-138	78.97%	VDL66569.1
<input type="checkbox"/>	Potassium voltage-gated channel protein Shaker [Toxocara canis]	400	400	75%	6e-134	65.77%	KHN86087.1
<input type="checkbox"/>	unnamed protein product [Toxocara canis]	401	401	75%	9e-134	65.77%	VDM38739.1
<input type="checkbox"/>	unnamed protein product [Angiostrongylus costaricensis]	398	398	74%	1e-132	59.95%	VDM52259.1

Potassium voltage-gated channel protein shk-1 [Caenorhabditis elegans]

Sequence ID: [NP_871936.1](#) Length: 501 Number of Matches: 1

[See 1 more title\(s\)](#)

Range 1: 173 to 501

GenPept

Graphics

▼ Next Match

▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
555 bits(1431)	0.0	Compositional matrix adjust.	301/336(90%)	306/336(91%)	14/336(4%)
Query 8	GRVKRPEIVPIDIFLKEMRFFQMGDDLLFEFWAEGYEKPEKVMPPNNKTQRKIWELMEY	67			
Sbjct 173	GRVKRPEIVPIDIFLKEMRFFQMGDDLLFEFWAEGYEKPEKVMPPNNKTQRKIWELMEY	232			
Query 68	PDSSLSARIIAFISIAVIALSIISFCWETVPSDIEEKPINNSATAELLDDEMDERKHSYFFF	127			
Sbjct 233	PDSSLSARIIAFISIAVIALSIISFCWETVPSDIEEKPINNSATAELLDDEMDERKHSYFFF	292			
Query 128	WIELMCLWFTIELLRPISCPCKVTFATSVLNIIDFVAIAFFVNFVFADTSKNSNSMS	187			
Sbjct 293	WIELMCLWFTIELLRPISCPCKVTFATSVLNIIDFVAIAFFVNFVFADTSKNSNSMS	352			
Query 188	FAVLRLVLRVRFVKLSRHSVGLQILGKTRSSVQEFCLLIFFMAIALVLFASGMFYA	247			
Sbjct 353	FAVLRLVLRVRFVKLSRHSVGLQILGKTRSSVQEFCLLIFFMAIALVLFASGMFYA	412			
Query 248	EQGEFNSKFTSIPASFVFLVITMTTVGVDLVLSPFGK-VWAVCAQ*IVS*HWHFQILL	306			
Sbjct 413	EQGEFNSKFTSIPASFVFLVITMTTVGVDLVLSPFGK V +CA V L	465			
Query 307	LIP-----NSNTSDQKRLASMKSGGDANDDFS 336				
Sbjct 466	ALPVPPIIVANFKHYQENRLASMKSGGDADDIA 501				

Download

GenPept

Graphics

▼ Next

▲ Previous

Descriptions

Potassium voltage-gated channel protein shk-1 [Caenorhabditis elegans]

Sequence ID: [NP_871935.1](#) Length: 536 Number of Matches: 1

[See 2 more title\(s\)](#)

Range 1: 208 to 536

GenPept

Graphics

▼ Next Match

▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
555 bits(1430)	0.0	Compositional matrix adjust.	301/336(90%)	306/336(91%)	14/336(4%)
Query 8	GRVKRPEIVPIDIFLKEMRFFQMGDDLLFEFWAEGYEKPEKVMPPNNKTQRKIWELMEY	67			
Sbjct 208	GRVKRPEIVPIDIFLKEMRFFQMGDDLLFEFWAEGYEKPEKVMPPNNKTQRKIWELMEY	267			
Query 68	PDSSLSARIIAFISIAVIALSIISFCWETVPSDIEEKPINNSATAELLDDEMDERKHSYFFF	127			
Sbjct 268	PDSSLSARIIAFISIAVIALSIISFCWETVPSDIEEKPINNSATAELLDDEMDERKHSYFFF	327			
Query 128	WIELMCLWFTIELLRPISCPCKVTFATSVLNIIDFVAIAFFVNFVFADTSKNSNSMS	187			
Sbjct 328	WIELMCLWFTIELLRPISCPCKVTFATSVLNIIDFVAIAFFVNFVFADTSKNSNSMS	387			
Query 188	FAVLRLVLRVRFVKLSRHSVGLQILGKTRSSVQEFCLLIFFMAIALVLFASGMFYA	247			
Sbjct 348	FAVLRLVLRVRFVKLSRHSVGLQILGKTRSSVQEFCLLIFFMAIALVLFASGMFYA	447			
Query 248	EQGEFNSKFTSIPASFVFLVITMTTVGVDLVLSPFGK-VWAVCAQ*IVS*HWHFQILL	306			
Sbjct 448	EQGEFNSKFTSIPASFVFLVITMTTVGVDLVLSPFGK V +CA V L	500			

Questions/comments

[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.

<input type="checkbox"/> Potassium channel and ion transport domain containing protein [Haemonchus contortus]	422	422	73%	4e-142	69.97%	CDJ84709.1
<input type="checkbox"/> hypothetical protein WR25_24211 [Diploscapter pachys]	415	415	74%	9e-139	67.07%	PAV86974.1
<input type="checkbox"/> unnamed protein product [Nippostrongylus brasiliensis]	409	409	65%	7e-138	78.97%	VDL66569.1
<input type="checkbox"/> Potassium voltage-gated channel protein Shaker [Toxocara canis]	400	400	75%	6e-134	65.77%	KHN86087.1
<input type="checkbox"/> unnamed protein product [Toxocara canis]	401	401	75%	9e-134	65.77%	VDM38739.1
<input type="checkbox"/> unnamed protein product [Angiostrongylus costaricensis]	398	398	74%	1e-132	59.95%	VDM52259.1
<input type="checkbox"/> transporter, cation channel family protein [Ancylostoma caninum]	388	388	70%	3e-130	70.35%	RCN37415.1
<input type="checkbox"/> unnamed protein product [Anisakis simplex]	395	395	65%	1e-129	73.43%	VDK46358.1
<input type="checkbox"/> unnamed protein product [Enterobius vermicularis]	380	380	64%	7e-127	70.71%	VDD95068.1
<input checked="" type="checkbox"/> unnamed protein product [Thelazia callipaeda]	379	379	65%	2e-124	71.43%	VDN08188.1
<input type="checkbox"/> K+ channel tetramerization domain protein [Onchocerca flexuosa]	368	368	66%	2e-120	67.58%	OZC11969.1
<input type="checkbox"/> unnamed protein product [Onchocerca flexuosa]	368	368	66%	3e-120	67.92%	VDP12905.1
<input type="checkbox"/> unnamed protein product [Haemonchus placei]	355	355	73%	1e-116	63.66%	VDO32825.1
<input type="checkbox"/> unnamed protein product [Onchocerca ochengi]	353	353	66%	2e-115	68.94%	VDK62583.1
<input type="checkbox"/> unnamed protein product [Brugia timoti]	346	346	66%	2e-112	69.62%	VDO32192.1
<input type="checkbox"/> BMA-SHK-1, isoform e [Brugia malayi]	346	346	66%	3e-112	69.62%	CDP90796.1
<input type="checkbox"/> unnamed protein product [Wuchereria bancrofti]	342	342	62%	7e-112	70.80%	VDM12676.1
<input type="checkbox"/> BMA-SHK-1, isoform g [Brugia malayi]	347	347	66%	9e-112	69.62%	CDP90793.1
<input type="checkbox"/> Potassium voltage-gated channel protein Shaker [Strongyloides ratti]	346	346	65%	7e-111	62.50%	XP_024509009.1
<input type="checkbox"/> hypothetical protein LOAG_01893 [Loa loa]	342	342	66%	7e-111	68.94%	XP_020303708.1
<input type="checkbox"/> unnamed protein product [Dracunculus medinensis]	337	337	62%	9e-110	68.50%	VDN55395.1
<input type="checkbox"/> BMA-SHK-1, isoform c [Brugia malayi]	340	340	66%	1e-109	68.00%	CDP90795.1

(Figure shows the clicked links corresponding to the different species selected)

```
>Human_K+_Channel_Protein | NP_002223.3 | potassium voltage-gated channel
subfamily A member 3 | [Homo sapiens]
MDERLSLLRSPPPPSARHRAHPPQRPASSGGAHTLVNHGYAEPAAGRELPPDMTVVPGDHLLPEVADGG
GAPPQGGCGGGGCDRYEPLPPSLPAAGEQDCCGERVVINISGLRFETQLKTLQCQFPETLLGDPKRRMRYP
DPLRNEYFFDRNRPSFDAILYYYQSGGRIRRPVNPVIDIFSEEIRFYQLGEEAMEKFREDEGFLREEERP
LPRRDFQRQVWLLFEYPSSGPARGIAIVSVLVILISIVIFCLETLPEFRDEKDYPASTSQDSFEAAGNS
TSGSRAGASSFSDPFFVETLCIIWFSFELLVRFFACPSKATFSRNIMNLIDIVAIIPYFITLGTALAER
QGNGQQAMSLAILRVIRLVRFRIKLSRHSKGLQILGQTLKASMRELGLLIFFLFIGVILFSSAVYFAE
ADDPTSGFSSIPDAFWAVVTMTTVGYGDMHPVTIGGKIVGSLCAIAGVLTIALPVPVIVSNFNFYFHRE
TEGEEQSQYMHVGCQHLSSSAEELRKARSNSTLSKSEYMVIEEGGMNHSAFPQTPFKTGNSTATCTTNN
NPNSCVNIKKIFTDV
```

>C.Elegans K+_Protein(Shaker-1) | NP_871936.1 | Potassium voltage-gated channel protein shk-1 | [Caenorhabditis elegans]

MIDSLIRRLEQVLNVHPEKYQNGCSSQRPSRQSPVEQLAPIRSHSADFGRAAEMASGNSEEGKMLLRN
GDDRIRLPSPQRGTPDTSSTQGHPTYDQIVTINVSGMRFTQFESTLSRYPNSLLGDRNKRQHFFVSDTNE
FFFDRHRTTSSSFTFEIRNYLFESILYIYQSGGRVKRPEIVPIDIFLKEMRFFQMGGDLLLEEFWIAEGYE
KPKEVMMPPNNKTQRKIWELMEYPDSSLSARIIAFISIAVIALSIISFCWETVPSDIEEKPINNSATAELL
DEMDEKHYSPPFFWIELMCILWFTIELILRFISCPCKVTFATSVLNIIDFVAIAPFFVNFADTSKSNSS
MSFAVLRVLRVFRVFRVFKLSRHSVGLQILGKTFRSSVQEFCLLIFFMAIALVLFASGMYFAEQGEPNSK
FTSIPASFVFLVMTTMTVGYGDLVPLSPFGKVVGGMICAMIGVLTALPVPIIVANFKHFYRQENRLASMK
SKGDDADDDIA

>Barber's_pole_worm K+_Channel_Protein | CDJ84709.1 | Potassium channel and Ion transport domain containing protein |[Haemonchus contortus]

MWRLRRYRGPPILDLEANGNLLVDDSTRNASPQRGTPQGDERLRSSDLVTINVSGLRFQTFERTLARFP
NTLLGCKSKRERYMQDTNEYFFDRHRSTFESILYIYQSGGRVKRPESVPIDIFLREMRFQMGEQLIEE
FWIAEGYEKPEEAILPKNLSQRRLWELMEYPDSSLAARIIAFISVIVIVISIVSFCWETVPSGDVRDTNG
TTGVTSPDKDEGRVWNPFFWLELVCIWFTIELTLRFISCPKVTFMMSFLNIIDFVAIAPFFVNLIIWA
DANKSSGSMFAVLRVLRVFRIFKLSRHSAGLQILGKTFRASVQEFCLLIFFMAIALVLFSSGMYFA
EQGEANSKFTSIPASFVFLVMTTMTVGYGDLVPLSTQGKIVGSMCALIGVLTALPVPIIVANFKHFYRQ
ENRLATMRSAGKDLETAESIDDS

>Dog_roundworm K+_Channel_Protein | KHN86087.1 | Potassium voltage-gated channel protein Shaker [Toxocara canis]

MIPEEEAVNVIPPTDGISSNMSENNSIVTINVSGLKFTFNSTLSRYPETLLGDPVKRAREFWNGHDGEYF
FDRHRGSFESILYIYQSHGIVKRPEAVPIDIFLKEMKFFEMTDDILEAFWISEGYEKPKAKMPQNIHQ
RIWELMEYPDSSLSARILAFVSIIVITVSIVSFCLETLPKPADIEGVRDWSNPFLWVEFVCIWFTIE
LLLRFASCPCKLTFMRSIPNIIDFIAIAPFFVNLVWSDVSKSNSSMSFAVLRVLRVFRIFKLSRHSV
GLQILGKTFRASIQEFCLLIFFMAIALVLFASGMYFAEQGEPDTKFTSIPASFVFLVMTTMTVGYGDLTP
TGVYGLVGSVCALIGVLTALPVPIIVANFKHFYRQETRLATMRANADEEDEECSDSSRT

>Deer_roundworm K+_Channel_Protein | OZC11969.1 | K+ channel tetramerization domain protein, partial [Onchocerca flexuosa]

RQIEHLLNVHPEKYRSRSGQSIETSTSVHLNQMGCHRSRSDYLSPPMDASNKEINEEKSNNGSIFPMI
TEESSNIMYNDTSLTKRRERVTVINVSGLRFQTYESTLRRYPLTLLGNSFKRNRFDWPKNKEYFFDRHRT
SFESILYTYQSGGVVKRPESVPIDMFIKELKFFEMGEELLEKFWISEGYEKPKIOMPKNVLLRKLWELV
EYPDSSLFARIFALLSIFVISISIIISFCLETLPKMKVNPGRDVRDWSNPFFHIELFCIIWFTTELLLRFIS
CPNKLSFLRSALNIIDFIAIAPFFGNLLWMSAKSSMSFAILRVLRVLRVFRIFKLSRHSVGLQILGKTF
RASIQEFCLLIFFMIIALVLFSSGIYFAEQNEPNTKFTSIPASFVFLVMTTMTVGYGDLTPTGVYGLV
GLCALIGVLTALPVPIIVANFKHFYRQETRLAQMRASVEDDETASEKSSKSP

>Rat_roundworm K+Channel Protein | XP_024509009.1 | Potassium voltage-gated channel protein Shaker [Strongyloides ratti]

MNKKTFSCGNDDISFSFRQKIEQILHVHPEKYRARGGVGSRRSTPGKRGSYDDRLHVRASASEELRQRSL
QSSSDSPSINDSNNSKQLEELSIPIVPECINSVGNKSNDSYRNPLTSHHNHIDPNDGIKWKNEEVITI
NVSGLRFTYESTLARFPNSLLGHPAKRAKYWNNKTKEYFLDHSRYCFESILNIYQTNGDVRPELVPLE
IYLKELKFYEMNEEVMEFWLSEGYEKPIEHIMPNNFIQKYIWELMEYPDSSVPAVLAFISISIVTISI
VSFCLETLPDFREASETLTEIGYSKWNPFILLES LCIVWFSIELFLRFISCPKSLFMKSFLNIVDFIA
IAPFFINLILSEGNKNSSTTSFAVLRVLRVLRVFRIFKLSRHSVGLQILGKTFRASIQEFCLLVFFMVI
ALVLFSSAIFYEQGEADTRFTSIPASFVFLVMTTMTVGYGDLVPTGTGYGLVGSLLCALLGVLTALPVP
IIVANFKHFYRQECRLAAMRVSREIDDASRYNESDEEYTT

>Human_Lymphatic_Roundworm Shaker 1 Protein | CDP90795.1 | BMA-SHK-1, isoform c [Brugia malayi]
MDTPSNREVSDEKPNNGSIFPMITEEEPNNIMFNDTSSAKQLERNLITINVSGLRFQTYESTLERYPLTL
LGNSFKRNRFWDPKNEEYFFDRHRTSFESILYAYQSGGIMKRPEVSPIDMFVKELKFFEVPNPKLQKMGE
LLEKFWISEGYEKPKEIQMPQNVLQRLWELVEYPDSSLFARIFALLSIFVISISIIISFCLETLP
PGDVRDWSNPFFYIELFCIIWFTIELLLRFISCPNKFSLRSVLNIIDFVAIAPFFGNLMWMDSTKSSSS
MSFAVLRLVRLVRVFRIFKLSRHSVGLQILGKTFRASIQEFCLLIFFMVIALVLFSSGVYFAEQNEPNTK
FTSIPASFWFVLVMTTMTVGYGDLTPTGVYGLVGGLCALIGVLTALPVPPIIVANFKHFYRQETRLAQM
ASAEEDETNSEHSSKSP

Alignment: (Obtained using MUSCLE (version 3.8) at EBI:

Human_K+_Channel_Protein	-----MDERLSLLRSPPPPSARHRAHPPQRPASSGGAHTLVNHGYAEP
Rat_roundworm	MNKKTFSCGNDDISFSFRQKIEQILHVHPEKYRARGGVGSRSTPGKRGSYDDRLHVRS
Dog_roundworm	-----
Deer_roundworm	-----RQIEHLLNVHPEKYRSRSGQSIETSTSVHLNQMGCHHRSR
Human_Lymphatic_Roundworm	-----
C.Elegans	-----MIDSLIRRLQVLNVHPEKYQNGCSSQRPSPQSPVEQLAPIRSHS-
Barber's_pole_worm	-----

Human_K+_Channel_Protein	AAGRELPPDMTVVPGDHLLEPEVADGG-----GAPPQGGCGGGGCDR-----YEPL
Rat_roundworm	ASEELRQRSLQSSSDSPSINDSNNSKQLEELSIPIVPECINSGVNKSNDSSEYRNPLTSH
Dog_roundworm	-----MIPE-----EEAVN-----VIPP
Deer_roundworm	V---DYLSPMMDASNKEINEEKSNNGS-----IFPMITE-----EESN-----IMYN
Human_Lymphatic_Roundworm	-----MDTPSNREVSDEKPNNGS-----IFPMITE-----EPPNN-----IMFN
C.Elegans	-----ADFGRAEMASGNSEEGK-----GMLLRNG-----DDRIR-----LPSP
Barber's_pole_worm	-----MWRLRRYRGPPILDLEANGN-----LLV-----DDSTR-----NASP

Human_K+_Channel_Protein	PPSLPAAGE---QDCGERVVINISGLRFETQLKTLQFPETLLGDPKRRMRYFDPLRNE
Rat_roundworm	HNHIDPNGD--IKWKNEEVITINVSGLRFQTYESTLARFPNSLLGHPAKRAKYWNNKTKE
Dog_roundworm	TDGISSNMS-----ENNSIVTINVSGLKFQTFNSTLSRYPETLLGDPVKRARFWNGHDGE
Deer_roundworm	DTSLTKRRE-----RVTINVSGLRFQTYESTLRRYPLTLGNSFKRNRFWDPKNKE
Human_Lymphatic_Roundworm	DTSSAKQLE-----RNLTINVSGLRFQTYESTLERYPLTLGNSFKRNRFWDPKNEE
C.Elegans	QRGTPDTSSTQGHYPYTDQIVTINVSGLRFQTFESTLSRYPNSSLGDRNKRQHFFVSDTNE
Barber's_pole_worm	QRGTPQGDE---RLRSSDLVTINVSGLRFQTFERTLARFPNTLLGCKSKRERYMQDTNE
	. . :.*:***.*:* ** .* :*** .* ::: *

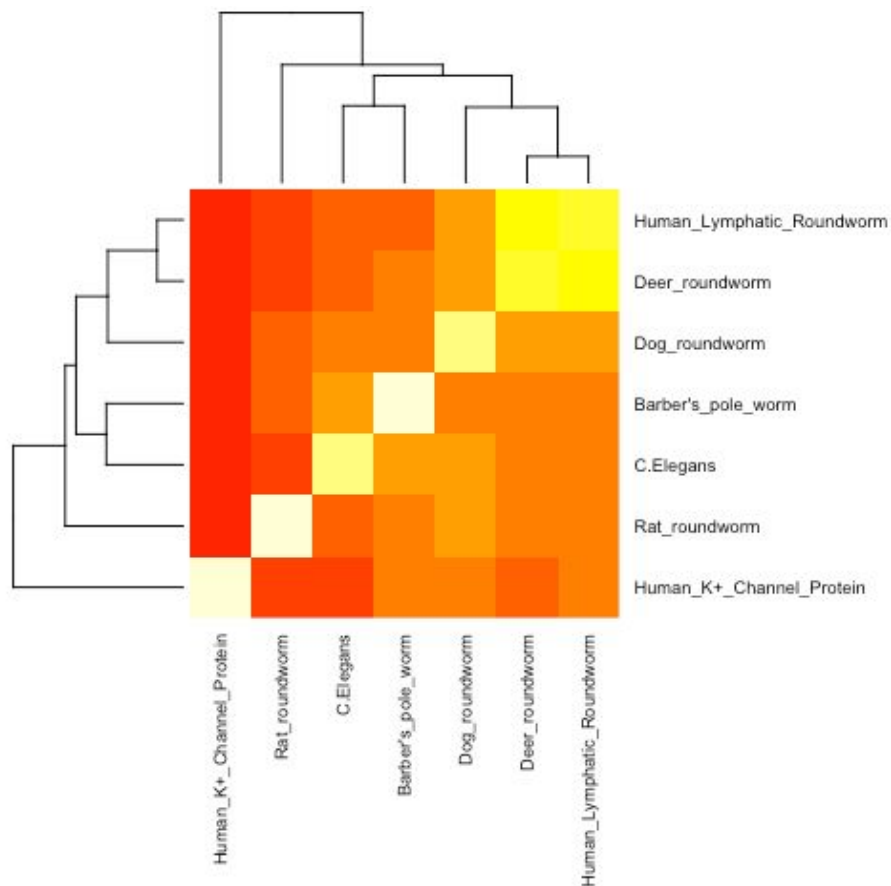
Human_K+_Channel_Protein	YFFDRNRPS-----FDAILYYYQSGGRIRRPVNPVIDIFSEEIRFYQ-----
Rat_roundworm	YFLDHSRYC-----FESILNIYQTNGDVRPELVPLEIYLKELKFYE-----
Dog_roundworm	YFFDRHRGS-----FESILYIYQSHGIVKRPEAVPIDIFLKEMKFFE-----
Deer_roundworm	YFFDRHRTS-----FESILYTYQSGGVVKRPEVSPIDMFVKELKFFE-----
Human_Lymphatic_Roundworm	YFFDRHRTS-----FESILYAYQSGGIMKRPEVSPIDMFVKELKFFEVPNPKLQ
C.Elegans	FFFDHRHTTSSSFTFEIRNYLFESILYIYQSGGRVKRPEIVPIDIFLKEMRFFQ-----
Barber's_pole_worm	YFFDRHRST-----FESILYIYQSGGRVKRPEVSPIDIFLREMRFQ-----
	:.*:.* **::** **: * :.* **::: *:.*::

Human_K+_Channel_Protein	-LGEEAMEKFREDEGFLREEERPLRRDFQRQVWLLFEYPSSGPARGIAIVSVLVILIS
Rat_roundworm	-MNEEVMEFNLSEGYEKPKEIHPNPNFIQKYIWELMEYPDSSVPARVLAFISISIVITS
Dog_roundworm	-MTDDILEAFWISSEGYEKPKEAKMPQNIHQRIWELMEYPDSSLSARILAFVSIIVITS
Deer_roundworm	-MGEELLEKFWISEGYEKPKEIQMPKNVLLRKLWELVEYPDSSLFARIFALLSIFVISIS
Human_Lymphatic_Roundworm	KMGEELEKFWISEGYEKPKEIQMPQNVLQRLWELVEYPDSSLFARIFALLSIFVISIS
C.Elegans	-MGDDLLEEFWIAEGYEKPKEVMPNNKTQRKIWELMEYPDSSLSARIIAFISIAVIALS
Barber's_pole_worm	-MGEQLIEEFWIAEGYEKPKEAAILPKNLSQRRLWELMEYPDSSLAARIIAFISVVFVIVIS
	: : : * *. **: . * :*.. . : * *.***:** ** :*:*: ** *

Human_K+_Channel_Protein	IVIFCLETLPFRDEKDYPASTSQDSFEAAGNSTSGSRAGASSFSDFPFVETLCIIWFS
--------------------------	--

Rat_roundworm	IVSFVCLETLPDPFRE-----ASETLEIGYSKWNPPFLILESLCIVWF5
Dog_roundworm	IVSFVCLETLP5LTK-----PADIEGVDRW5NPPFLWVEFVCI1WFT
Deer_roundworm	IISFCLETLP5MKV-----NPGD---VRDWSNPPFFHIELFC1WFT
Human_Lymphatic_Roundworm	IISFCLETLP5MK-----NPGD---VRDWSNPPFFYIELFC1WFT
C.Elegans	IISFCWETVPSDIE-----EKPINNSATAELLDEMDEKHYS-PFFWIELMCI1WFT
Barber's_pole_worm	IVSFVCWETVPSGDV-----RDTNGTGTGVTSPDKDEGR1VWN-PFFWLELVC1WFT
	<div> *: * * *: *. </div> <div> : . *: : * :*:*: </div>
Human_K+_Channel_Protein	FELLVRFFACPSKATFSRNIMNLIDIVAIIPYFITLGTLELAERQGNQGQAMSLAILRVIR
Rat_roundworm	IELFLRFISCP5KSLFMK5FLNIVDFIAIAPFFINL--ILSEGNKNNSS5TSFAVLRVLR
Dog_roundworm	IELLLRFASCPCKLTFMRSIPNIIDFIAIAPFFVNL--VWSDVSK5N5S-MSFAVLRVLR
Deer_roundworm	TELLRLFISCPNKL5FLRSALNIIDFIAIAPFFGNL--LWMD5AK5S---MSFAILRVLR
Human_Lymphatic_Roundworm	IELLLRFISCPNKF5FLRSVLNIIDFVAIAPFFGNL--MWMD5TK5SS5S-MSFAVLRVLR
C.Elegans	IELILRFISCPCKVTFATSVLNIIDFVAIAPFFVNF--FFAD5TK5N5S-MSFAVLRVLR
Barber's_pole_worm	IELTLRFISCP5K5VTFM55FLNIIDFVAIAPFFVNL--IWADANK5SG5-MSFAVLRVLR
	<div> * * :* :* * * . *::*: * *: * : : : . . *:*:*:*: </div>
Human_K+_Channel_Protein	LVRVFRIFKLSRHSKGLQILGQTLK5MRELGLLIFFFLFIGVILFSSAVYFAEADDPTSG
Rat_roundworm	LVRVFRIFKLSRHSVGLQILGKTFRASIQEFCLLVFFMVIALVL5SSAIFYVEQGEADTR
Dog_roundworm	LVRVFRIFKLSRHSVGLQILGKTFRASIQEFCLLIFFMAIALVL5FASGM5YFAEQGEPDTK
Deer_roundworm	LVRVFRIFKLSRHSVGLQILGKTFRASIQEFCLLIFFMIIALVL5SSGIYFAEQNEPNTK
Human_Lymphatic_Roundworm	LVRVFRIFKLSRHSVGLQILGKTFRASIQEFCLLIFFMVIALVL5SSGVYFAEQNEPNTK
C.Elegans	LVRVFRVFKLSRHSVGLQILGKTFR5SVQEFCLLIFFMAIALVL5FASGM5YFAEQGEPNSK
Barber's_pole_worm	LVRVFRIFKLSRHSAGLQILGKTFRASVQEFCLLIFFMAIALVL5SSGM5YFAEQGEANSK
	<div> *****:***** *:*:*:*: *:*:*: *:*:*:*:*:*:*: * :. : </div>
Human_K+_Channel_Protein	FSSIPDAFWWAVVTMTTVGYGDMHPVTIGGKIVG5SLCAIAGVLTIALPVPVIV5NFN5FY
Rat_roundworm	FTSIPASFWFV5VATMTTVGYGDLVPTGTYGKLVG5SLCALLGVLTLALPVPIIVANFKHFY
Dog_roundworm	FTSIPASFWFVLVTMTTVGYGDLTPTGVYGKLVG5SVCALIGVLTALPVPIIVANFKHFY
Deer_roundworm	FTSIPASFWFVLVTMTTVGYGDLTPTGVYGKLVGGLCALIGVLTALPVPIIVANFKHFY
Human_Lymphatic_Roundworm	FTSIPASFWFVLVTMTTVGYGDLTPTGVYGKLVGGLCALIGVLTALPVPIIVANFKHFY
C.Elegans	FTSIPASFWFVLVTMTTVGYGDLVPLSPFGKVVG5GMCAMIGVLTALPVPIIVANFKHFY
Barber's_pole_worm	FTSIPASFWFVLVTMTTVGYGDLVPLSTQGKIVG5SMCALIGVLTALPVPIIVANFKHFY
	<div> *:** *:*:*.*****: * **:*.*:*: *****:*****:*****: </div>
Human_K+_Channel_Protein	HRETEGEEQ5QYMHV5GCQHL5SSAEELRKARSNTLSKSEYMVIEEGGMNHS5AFPQTFP
Rat_roundworm	RQEC-----RLAAMRV5SREID5ASRYNESDEEYTT-----
Dog_roundworm	RQET-----RLATMRANADEEDEE5CQD5SRT-----
Deer_roundworm	RQET-----RLAQMRASVEDDE5ASEK5SK5P-----
Human_Lymphatic_Roundworm	RQET-----RLAQMRASAEEDETN5H5SK5P-----
C.Elegans	RQEN-----RL5MK5K5GDDADDDIA-----
Barber's_pole_worm	RQEN-----RLATMR5AGKDL5ETAESID5S-----
	<div> ..* : . </div>
Human_K+_Channel_Protein	KTGN5STATCTNNNPNSCVNIKKIFTDV
Rat_roundworm	-----
Dog_roundworm	-----
Deer_roundworm	-----
Human_Lymphatic_Roundworm	-----
C.Elegans	-----
Barber's pole worm	-----

Using “simple phylogeny” online from the EBI, we got this tree:

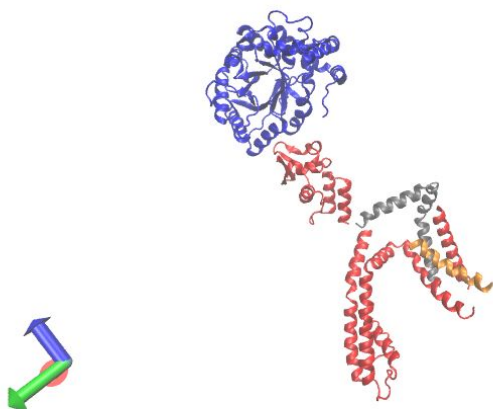


[Q8] Using R/Bio3D (or an online blast server if you prefer), search the main protein structure database for the most similar atomic resolution structures to your aligned sequences.

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).

ID	Technique	Resolution	Source	Evalue	Identity
2A79_B	X-RAY DIFFRACTION	2.9	Rattus norvegicus	6.64e-140	51.523
3LUT_B	X-RAY DIFFRACTION	2.9	Rattus norvegicus	7.08e-140	51.523
2R9R_B	X-RAY DIFFRACTION	2.4	Rattus norvegicus	3.98e-134	49.744

[Q9] Generate a molecular figure of one of your identified PDB structures using **VMD**. You can optionally highlight conserved residues that are likely to be functional. Please use a white or transparent background for your figure (i.e. not the default black). Based on sequence similarity. How likely is this structure to be similar to your “novel” protein?



2A79 is chosen to generate the VMD PDB structures (Rendered in internal Tachyon, saved as tga file and then converted to JPG in order to be put in the google doc). Based on sequence similarity (~50%) and the rule of thumb mentioned in the lecture, since they have more than 45% identity, they should have similar structures.

[Q10] Perform a “Target” search of ChEMBL (<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?

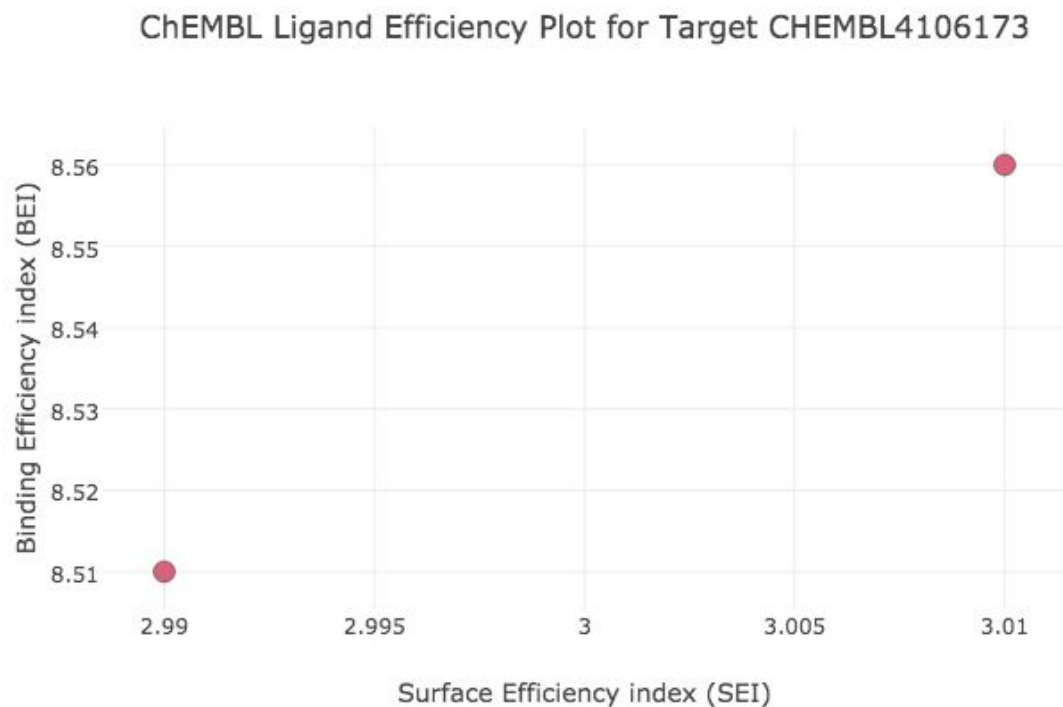
https://www.ebi.ac.uk/chembl/target_report_card/CHEMBL4106173/

CHEMBL Details 8 Binding Assays

<input type="checkbox"/> ChEMBL ID	Description	Organism	Compounds	Document	BAO Format	Source
<input type="checkbox"/> CHEMBL4009564	Inhibition of recombinant rat Kv1.1/1.1/1.2/1.2 expressed in HEK293 cells at 10 uM at -90 mV holding potential by whole cell patch clamp Qpatch method relative to control	Rattus norvegicus	1 By Mol. Wt.:	CHEMBL4007435	cell-based format	Scientific Literature
<input type="checkbox"/> CHEMBL4009568	Inhibition of recombinant rat Kv(1.1)3/Kv1.2 expressed in HEK293 cells assessed as time constant for activation of Kv1.1 channel current at 10 uM at -90 mV holding potential by whole cell patch clamp Qpatch method (Rvb = 3.1 +/- 0.2 ms)	Rattus norvegicus	1 By Mol. Wt.:	CHEMBL4007435	cell-based format	Scientific Literature
<input type="checkbox"/> CHEMBL4009563	Inhibition of recombinant rat Kv1.2/1.1/1.2/1.1 expressed in HEK293 cells at 10 uM at -90 mV holding potential by whole cell patch clamp Qpatch method relative to control	Rattus norvegicus	1 By Mol. Wt.:	CHEMBL4007435	cell-based format	Scientific Literature
<input type="checkbox"/> CHEMBL4009567	Inhibition of recombinant rat Kv(1.1)2/Kv(1.2)2 expressed in HEK293 cells assessed as time constant for activation of Kv1.1 channel current at 10 uM at -90 mV holding potential by whole cell patch clamp Qpatch method (Rvb = 2.8 +/- 0.4 ms)	Rattus norvegicus	1 By Mol. Wt.:	CHEMBL4007435	cell-based format	Scientific Literature
<input type="checkbox"/> CHEMBL4009561	Inhibition of recombinant rat Kv1.1/Kv(1.2)3 expressed in HEK293 cells at -90 mV holding potential by whole cell patch clamp Qpatch method	Rattus norvegicus	1 By Mol. Wt.:	CHEMBL4007435	cell-based format	Scientific Literature
<input type="checkbox"/> CHEMBL4009560	Inhibition of recombinant rat Kv(1.1)2/Kv(1.2)2 expressed in HEK293 cells at -90 mV holding potential by whole cell patch clamp Qpatch method	Rattus norvegicus	1 By Mol. Wt.:	CHEMBL4007435	cell-based format	Scientific Literature
<input type="checkbox"/> CHEMBL4009565	Inhibition of recombinant rat Kv1.2/1.2/1.1/1.1 expressed in HEK293 cells at 10 uM at -90 mV holding potential by whole cell patch clamp Qpatch method relative to control	Rattus norvegicus	1 By Mol. Wt.:	CHEMBL4007435	cell-based format	Scientific Literature
<input type="checkbox"/> CHEMBL4009559	Inhibition of recombinant rat Kv(1.1)3/Kv1.2 expressed in HEK293 cells at -90 mV holding potential by whole cell patch clamp Qpatch method	Rattus norvegicus	1 By Mol. Wt.:	CHEMBL4007435	cell-based format	Scientific Literature

Link to one of the binding assay linked articles:

<https://pubs.acs.org/doi/10.1021/acs.jmedchem.6b01262>



Ligand Efficiency session has the following plots. Both points are associated with molecule CHEMBL4079366 (Different standard values) that is possible to inhibit the novel protein.

Information about this molecule:

Name And Classification

