Find A Gene Project

[BIMM143] Xinyu Wen A17115443 x5wen@ucsd.edu

Q1

Name: Kinesin-like protein KIF11 (KIF11)
Accession: P52732
Species: Homo Sapiens

Q2

Method: TBLASTN, search against EST, exclude homo sapeins
Database: Expressed sequence tags (est)
Organism: Exclude Homo Sapeins

Chosen match: Accession DW624128.1. See below for alignment details.

Job Title	P52732:RecName: Full=Kinesin-like protein...
RID	TU73UHEU013 Search expires on 02-02 01:54 am Download All ⌄
Program	TBLASTN ❓ Citation ⌄
Database	est See details ⌄
Query ID	P52732.2
Description	RecName: Full=Kinesin-like protein KIF11; AltName: Full=K ...
Molecule type	amino acid
Query Length	1056
Other reports	❓

Filter Results

Organism *only top 20 will appear*   ☐ exclude

Type common name, binomial, taxid or group name

✚ Add organism

Percent Identity	E value	Query Coverage
☐ to ☐	☐ to ☐	☐ to ☐

Filter   Reset

| Descriptions | Graphic Summary | Alignments | Taxonomy |

Sequences producing significant alignments   Download ⌄   Select columns ⌄   Show 100 ⌄ ❓

☐ select all   1 sequences selected   GenBank   Graphics

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☐	Rhizoctonia solani EST, clone rsol_all_isotig14362, strain AG1-IB, mRNA sequence	Rhizoctonia solani	444	444	45%	1e-141	48.05%	1681	HG373950.1
☐	CGX41-C02.y1d-s SHGC-CGX Gasterosteus aculeatus cDNA clone CGX41-C02 5', mRNA sequence	Gasterosteus ac...	360	464	32%	2e-120	79.51%	1252	DN685345.1
☐	FS309663 Cp_aO Cynops pyrrhogaster cDNA clone Cp_aO_024_L08 3', mRNA sequence	Cynops pyrrhoga...	424	577	32%	1e-154	84.36%	1055	FS309663.1
☑	CLJ338-B02.y1d-s SHGC-CLJ2 Gasterosteus aculeatus cDNA clone CLJ338-B02 5', mRNA sequence	Gasterosteus ac...	446	446	31%	2e-145	67.79%	1043	DW624128.1
☐	FS310020 Cp_aO Cynops pyrrhogaster cDNA clone Cp_aO_025_K19 3', mRNA sequence	Cynops pyrrhoga...	422	575	31%	5e-154	86.32%	1027	FS310020.1
☐	AUF_IpOva_19_f11 Ovary cDNA library Ictalurus punctatus cDNA 5' similar to kinesin family member 11, mRNA s...	Ictalurus punctatus	417	417	29%	1e-134	68.71%	942	CK418843.1
☐	ypo4-4-G01 Yellow perch ovarian library 4 Perca flavescens cDNA, mRNA sequence	Perca flavescens	414	414	28%	2e-133	70.21%	965	GO660024.1
☐	CFW85-E12.y1d-s SHGC-CFW Gasterosteus aculeatus cDNA clone CFW85-E12 5', mRNA sequence	Gasterosteus ac...	448	448	27%	2e-145	75.35%	1208	DN675298.1
☐	JGI_CAAV7594.fwd CAAV Pimephales promelas testis 7-8 month adults, males and females pooled (H) Pimepha...	Pimephales pro...	544	544	27%	0.0	87.80%	865	DT277589.1

E-value: 2e-145
Percent identity: 67.79%
Query cover: 31%

Distribution of the top 107 Blast Hits on 100 subject sequences

CLJ338-B02.y1d-s SHGC-CLJ2 Gasterosteus aculeatus cDNA clone CLJ338-B02 5', mRNA sequence

Sequence ID: DW624128.1 Length: 1043 Number of Matches: 1

Score	Expect	Method	Identities	Positives	Gaps	Frame
446 bits(1146)	2e-145	Compositional matrix adjust.	232/326(71%)	273/326(83%)	0/326(0%)	+1

```
Query 206  NKDEVYQILEKGAAKRTTAATLMNAYSSRSHSVFSVTIHMKETTIDGEELVKIGKLNLVD 265
           ++DEVYQILE+G+AKR TA+TLMNAYSSRSHSVFSVTIHMKE T+DGEELVKIGKLNLVD
Sbjct 43   DQDEVYQILERGSAKRRTASTLMNAYSSRSHSVFSVTIHMKEITMDGEELVKIGKLNLVD 222

Query 266  LAGSENIGRSGAVDKRAREAGNINQSLLTLGRVITALVERTPHVPYRESKLTRILQDSLG 325
           LAGSENIGRSGAVDKRAREAGNINQSLLTLGRVITALVE+ PHVPYRESKLTRILQDSLG
Sbjct 223  LAGSENIGRSGAVDKRAREAGNINQSLLTLGRVITALVEKRPHVPYRESKLTRILQDSLG 402

Query 326  GRTRTSIIATISPAslnleetlstleYAHRAKNILNKPEVNQKLTKKALIKEYTEEIERL 385
           GRT+TSIIAT+SP+S NLEETLSTLEYA RAKNI+NKPEVNQKLTK+ LIKEYTEEIERL
Sbjct 403  GRTKTSIIATVSPSSSNLEETLSTLEYASRAKNIMNKPEVNQKLTKRTLIKEYTEEIERL 582

Query 386  KRDLAAAREKNGVYISEENFRVMSGKLTVQEEQIVELIEKIGAVEEELNRVTELFMDNKN 445
           KRDLAA R+KNG+Y+S EN+ M G++T E  VE ++I A+EEE+ +VTELF+D+K
Sbjct 583  KRDLAATRDKNGIYLSAENYESMMGQITSHEVHTVEYSDRIAAMEEEIKKVTELFVDSKT 762

Query 446  ELDQCKSDLQNKTQELETTQKHLQETKLQLVKEEYITSALESTEEKLHDAASKLLNTVEE 505
           L+ C DL K Q LE T +LQ TK +L++ E++ S L  +E L+D A +LL+TV+
Sbjct 763  RLELCAVDLDEKQQRLEETSRDLQHTKEKLMEXEFVCSELTLVQESLYDTAGRLLSTVDA 942

Query 506  TTKDVSGLHSKLDRKKAVDQHNAEAQ 531
           +T DV GL +LDR K V+QH + Q
Sbjct 943  STGDVCGLPGGQLDRXKXVEQHYSGVQ 1020
```

Q3

Chosen sequence: (translated by EMBOSS Transeq, frame +1)

>DW624128.1_1 CLJ338-B02.y1d-s SHGC-CLJ2 Gasterosteus aculeatus cDNA clone CLJ338-B02 5', mRNA sequence
FXXXXXXXFGPESRDQDEVYQILERGSAKRRTASTLMNAYSSRSHSVFSVTIHMKEITMD
GEELVKIGKLNLVDLAGSENIGRSGAVDKRAREAGNINQSLLTLGRVITALVEKRPHVPY
RESKLTRILQDSLGGRTKTSIIATVSPSSSNLEETLSTLEYASRAKNIMNKPEVNQKLTK
RTLIKEYTEEIERLKRDLAATRDKNGIYLSAENYESMMGQITSHEVHTVEYSDRIAAMEE
EIKKVTELFVDSKTRLELCAVDLDEKQQRLEETSRDLQHTKEKLMEXEFVCSELTLVQES
LYDTAGRLLSTVDASTGDVCGLPGQLDRXKXVEQHYSGVQQSSLSAWX

NAME: Gasterosteus aculeatus
ORGANISM: Gasterosteus aculeatus
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Actinopterygii; Neopterygii; Teleostei; Neoteleostei;
        Acanthomorphata; Eupercaria; Perciformes; Cottioidei;
        Gasterosteales; Gasterosteidae; Gasterosteus.

Q4

Details:

A BLASTP search against nr database (see setup in first screen-shot below). The top match is from Gasterosteus aculeatus aculeatus.

See screenshots below for alignment details:

Job Title	DW624128.1_1 CLJ338-B02.y1d-s SHGC-CLJ2 Gasterosteus..		
RID	TUCH7CM7016 *Search expires on 02-02 03:27 am* Download All ⌄		
Program	BLASTP ❓ Citation ⌄		
Database	nr See details ⌄		
Query ID	lcl|Query_6479531		
Description	DW624128.1_1 CLJ338-B02.y1d-s SHGC-CLJ2 Gasteroste ...		
Molecule type	amino acid		
Query Length	348		
Other reports	Distance tree of results Multiple alignment MSA viewer ❓		

Filter Results

Organism *only top 20 will appear*   ☐ exclude

Type common name, binomial, taxid or group name

➕ Add organism

Percent Identity   **E value**   **Query Coverage**

[ ] to [ ]   [ ] to [ ]   [ ] to [ ]

Filter   Reset

Compare these results against the new Clustered nr database ❓   **BLAST**   ✕

| **Descriptions** | Graphic Summary | Alignments | Taxonomy |

Sequences producing significant alignments

Download ⌄   Select columns ⌄   Show 100 ⌄   ❓

☑ 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
☑	kinesin-like protein KIF11 [Gasterosteus aculeatus aculeatus]	Gasterosteus aculeatus aculeatus	637	637	94%	0.0	96.65%	1060	XP_040035136.1
☑	kinesin-like protein KIF11 [Pungitius pungitius]	Pungitius pungitius	617	617	94%	0.0	93.60%	1060	XP_037320999.2
☑	kinesin-like protein KIF11 isoform X1 [Anarrhichthys ocellatus]	Anarrhichthys ocellatus	613	613	94%	0.0	93.29%	1060	XP_031728199.1
☑	kinesin-like protein KIF11 isoform X2 [Anarrhichthys ocellatus]	Anarrhichthys ocellatus	613	613	94%	0.0	93.29%	1054	XP_031728209.1
☑	kinesin-like protein KIF11 isoform X1 [Anoplopoma fimbria]	Anoplopoma fimbria	609	609	94%	0.0	91.77%	1062	XP_054456823.1
☑	kinesin-like protein KIF11 [Cebidichthys violaceus]	Cebidichthys violaceus	607	607	94%	0.0	92.68%	1062	XP_068587987.1
☑	hypothetical protein VZT92_001339 [Zoarces viviparus]	Zoarces viviparus	600	600	94%	0.0	90.55%	1056	KAK9541282.1
☑	kinesin-like protein KIF11 [Sander lucioperca]	Sander lucioperca	599	599	94%	0.0	90.55%	1057	XP_031164435.1

Alignment view	Pairwise	⌄	❓	**Restore defaults**

100 sequences selected ❓

⬇ **Download** ⌄   GenPept Graphics

kinesin-like protein KIF11 [Gasterosteus aculeatus aculeatus]

Sequence ID: XP_040035136.1   Length: **1060**   Number of Matches: **1**

See 2 more title(s) ⌄   See all Identical Proteins(IPG)

Range 1: 204 to 531 GenPept   Graphics   ▽ Next Match △ Previo

Score	Expect	Method	Identities	Positives	Gaps
637 bits(1644)	0.0	Compositional matrix adjust.	317/328(97%)	320/328(97%)	0/328(0%)

```
Query   15 DQDEVYQILERGSAKRRTASTLMNAYSSRSHSVFSVTIHMKEITMDGEELVKIGKLNLVD  74
              ++DEVYQILERGSAKRRTASTLMNAYSSRSHSVFSVTIHMKEITMDGEELVKIGKLNLVD
Sbjct  204 NKDEVYQILERGSAKRRTASTLMNAYSSRSHSVFSVTIHMKEITMDGEELVKIGKLNLVD  263

Query   75 LAGSENIGRSGAVDKRAREAGNINQSLLTLGRVITALVEKRPHVPYRESKLTRILQDSLG  134
              LAGSENIGRSGAVDKRAREAGNINQSLLTLGRVITALVEKRPHVPYRESKLTRILQDSLG
Sbjct  264 LAGSENIGRSGAVDKRAREAGNINQSLLTLGRVITALVEKRPHVPYRESKLTRILQDSLG  323

Query  135 GRTKTSIIATVSPSSSNLEETLSTLEYASRAKNIMNKPEVNQKLTKRTLIKEYTEEIERL  194
              GRTKTSIIATVSPSSSNLEETLSTLEYASRAKNIMNKPEVNQKLTKRTLIKEYTEEIERL
Sbjct  324 GRTKTSIIATVSPSSSNLEETLSTLEYASRAKNIMNKPEVNQKLTKRTLIKEYTEEIERL  383

Query  195 KRDLAATRDKNGIYLSAENYESMMGQITSHEVHTVEYSDRIAAMEEEIKKVTELFVDSKT  254
              KRDLAATRDKNGIYLSAENYESMMGQITSHEVHTVEYSDRIAAMEEEIKKVTELFVDSKT
Sbjct  384 KRDLAATRDKNGIYLSAENYESMMGQITSHEVHTVEYSDRIAAMEEEIKKVTELFVDSKT  443

Query  255 RLELCAVDLDEKQQRLEETSRDLQHTKEKLMEXEFVCSELTLVQESLYDTAGRLLSTVDA  314
              RLELCAVDLDEKQQRLEETSRDLQHTKEKLME EFVCSELT VQESLYDTAGRLLSTVDA
Sbjct  444 RLELCAVDLDEKQQRLEETSRDLQHTKEKLMEEEFVCSELTSVQESLYDTAGRLLSTVDA  503

Query  315 STGDVCGLPGQLDRXKXVEQHYSGVQQS  342
              STGDVCGL +LDR K VEQH S VQQS
Sbjct  504 STGDVCGLQDKLDRKKKVEQHNSEVQQS  531
```

Q5

Re-labeled sequences for alignment (NEW, gaps removed):

Same as MSA alignment, obtained using MUSCLE (version 3.8) at EBI:

In FASTA format
Q5NewMSAmuscle-I20250225.aln-fasta

>Homo sapiens
EKGAAKRTTAATLMNAYSSRSHSVFSVTIHMKETTIDGEELVKIGKLNLVDLAGSENIGRSGAVDK
RAREAGNINQSLLTLGRVITALVERTPHVPYRESKLTRILQDSLGGRTRTSIIATISPASLNLEETLSTL
EYAHRAKNILNKPEVNQKLTKKALIKEYTEEIERLKRDLAAAREKNGVYISEENFRVMSGKLTVQ

EEQIVELIEKIGAVEEELNRVTELFMDNKNELDQCKSDLQNKTQELETTQKHLQETKLQLVKEEYI
TSALESTEEKLHDAASKLLNTVEETTKDVSGLHSKLDRKKAVDQHNAEAQDI

>Gasterosteus aculeatus (Novel protein)
ERGSAKRRTASTLMNAYSSRSHSVFSVTIHMKEITMDGEELVKIGKLNLVDLAGSENIGRSGAVDK
RAREAGNINQSLLTLGRVITALVEKRPHVPYRESKLTRILQDSLGGRTKTSIIATVSPSSSNLEETLST
LEYASRAKNIMNKPEVNQKLTKRTLIKEYTEEIERLKRDLAATRDKNGIYLSAENYESMMGQITS
HEVHTVEYSDRIAAMEEEIKKVTELFVDSKTRLELCAVDLDEKQQRLEETSRDLQHTKEKLMEX
EFVCSELTLVQESLYDTAGRLLSTVDASTGDVCGLPGQLDRXKXVEQHYSGVQQS

>Pungitius pungitius
ERGSAKRRTASTLMNAYSSRSHSVFSVTIHMKEITVDGEELVKIGKLNLVDLAGSENIGR
SGAVDKRAREAGNINQSLLTLGRVITALVEKRPHVPYRESKLTRILQDSLGGRTKTSIIA
TVSPSSSNLEETLSTLEYASRAKNIMNKPEVNQKLTKRTLIKEYTEEIERLKRDLAATRD
KNGIFLSAENYESMMGQITSHEMHSVEYSDRIAAMEEEIKKVTELFVDSKTRLELCTVDL
DQKQQRLQETSRDLQHTKEKLMEEEFVCSELTSVQESLYDTAGRLLSTVDASTGDVWGLH
DKLDRKNKVEQHNSEVQQS

>Anarrhichthys ocellatus | XP_031728199.1 KIF11
ERGSAKRRTASTLMNAYSSRSHSVFSVTIHMKEITMDGEELVKIGKLNLVDLAGSENIGR
SGAVDKRAREAGNINQSLLTLGRVITALVEKRPHVPYRESKLTRILQDSLGGRTKTSIIA
TVSPSSSNLEETLSTLEYASRAKNIMNKPEVNQKLTKRTLIKEYTEEIERLKRDLAATRD
KNGVYLSAENYESMLGQITSHEVHTVEYSDRIAAMEDEIKKVTELFVDSKTRLEQCTVVL
DEKQQRLEETSRDLQHTKEKLSEEEFVCTELTSVQESLYDTAGQLLSTVDASTGDVSGLH
EKLDRKKKVEQHNSEIQQS

>Anoplopoma fimbria | XP_054456823.1 KIF11
ERGSAKRRTASTLMNAYSSRSHSVFSVTIHMKEITMDGEELVKIGKLNLVDLAGSENIGR
SGAVDKRAREAGNINQSLLTLGRVITALVEKRPHIPYRESKLTRILQDSLGGRTKTSIIA
TVSPSSSNLEETLSTLEYASRAKNIMNKPEVNQKLTKRTLIKEYTEEIERLKRDLAATRD
KNGVYLSAENYETMMGQITSHEVHTGEYTDRIAAMEEEIKKVTELFVDSKTRLEQCAVDL
DEKQQRLEETSKDLQQTKEKLSQEEFVCSELTSVQETLYDTAGQLLSTVDASTSDVTGLH
DKLDRKKKVEQHNTEIQQS

>Cebidichthys violaceus | XP_068587987.1 KIF11
ERGSAKRRTASTLMNAYSSRSHSVFSVTIHMKEITMDGEELVKIGKLNLVDLAGSENIGR
SGAVDKRAREAGNINQSLLTLGRVITALVEKRPHVPYRESKLTRILQDSLGGRTKTSIIA
TVSPSSSNLEETLSTLEYASRAKNIMNKPEVNQKLTKRTLIKEYTEEIERLKRDLAATRD
KNGVYLSAENYESMMGQITSHEVHTVEYSDRIAAMEDEIKKVTELFVDSKTRLEQCAVVL
DEKQQRLEETSRALQHTKDKLSQEEFVSTELTSVQESLYDTAGQLLSTVDASTGDVSGLH
DKLDRKKKVEQHNSEIQQS

>Sander lucioperca | XP_031164435.1 KIF11
ERGSAKRRTASTLMNAYSSRSHSVFSVTIHMKEITMDGEELVKIGKLNLVDLAGSENIGR

SGAVDKRAREAGNINQSLLTLGRVITALVEKRPHIPYRESKLTRILQDSLGGRTKTSIIA
TVSPSSSNLEETLSTLEYASRAKNIMNKPEVNQKLTKRTLIKEYTEEIERLKRDLAATRD
KNGVYLSAENYESMMGQITSHEEHTSEYTDRIAAMEEEIKKVTELFTDSKTRLEQCTVDL
DQKQQMLEETSKDLQQTKEKLSQEEFVCSELTVVQETLYNTAGQLLSTVDASTSDVMGLH
DKLDRKKKVEQHNSQIQQS

Q6

Align the sequences with MUSCLE, import into Simple Phylogeny at EBI, and create a
neighbor-joining tree:
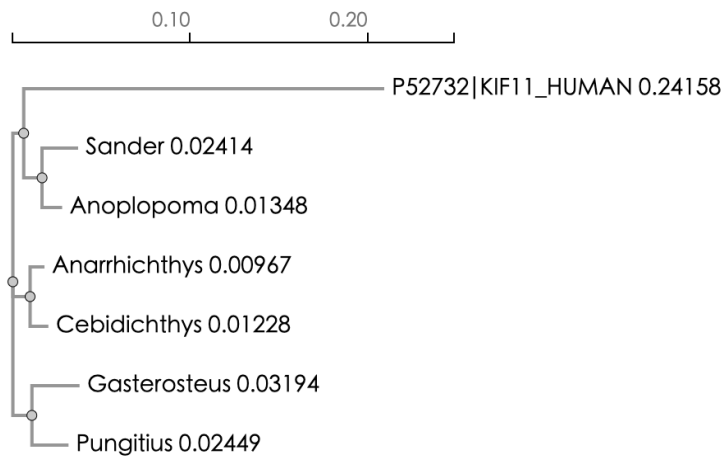
Q7
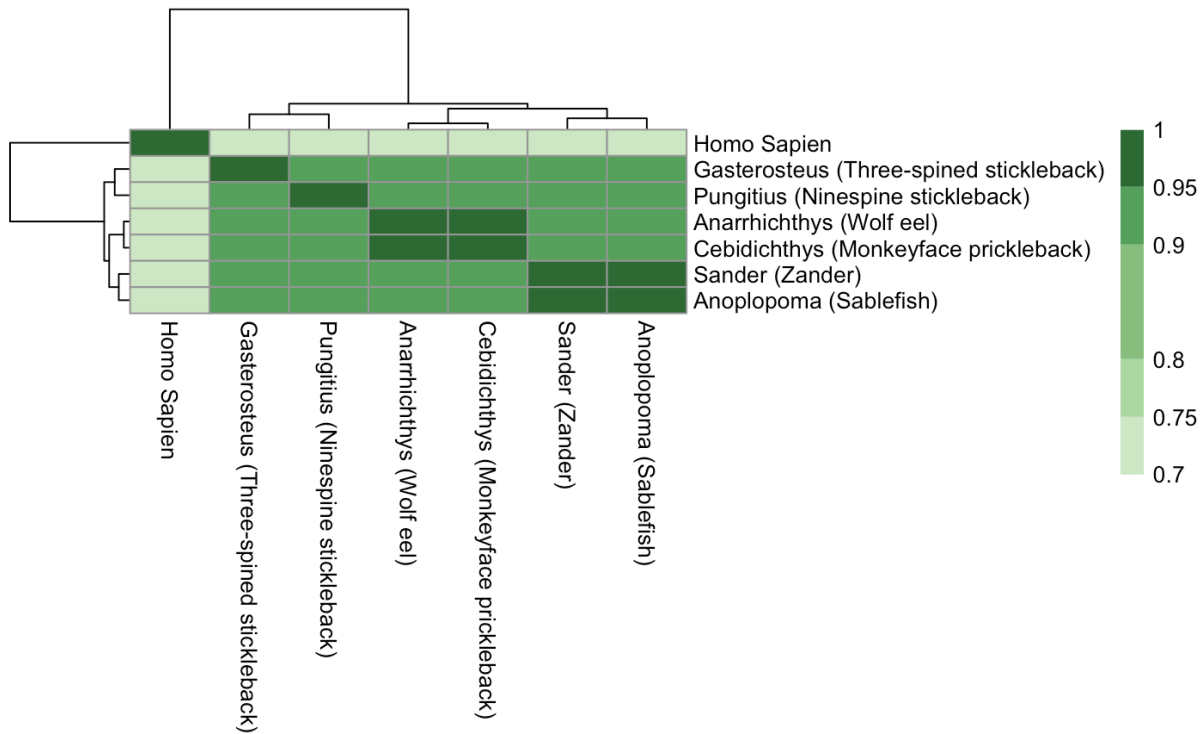
Q8

Using R/Bio3d and Gasterosteus aculeatus sequences. Also searched on PDB for source organism.

ID	Method	Resolution	Source Organism	E-value	% Identity
3HQD	X-RAY DIFFRACTION	2.19 Å	Homo sapiens	1.11e-95	90.968
4ZHI	X-RAY DIFFRACTION	2.30 Å	Homo sapiens	1.11e-95	90.968
5ZO7	X-RAY DIFFRACTION	2.60 Å	Homo sapiens	1.19e-95	90.968

Q9

PDB files generated through Colabfold, using Gasterosteus aculeatus sequences. Then the pdb files were uploaded to Mol* to generate protein structures.
Conserved sequences cut off set as 0.995 in R.
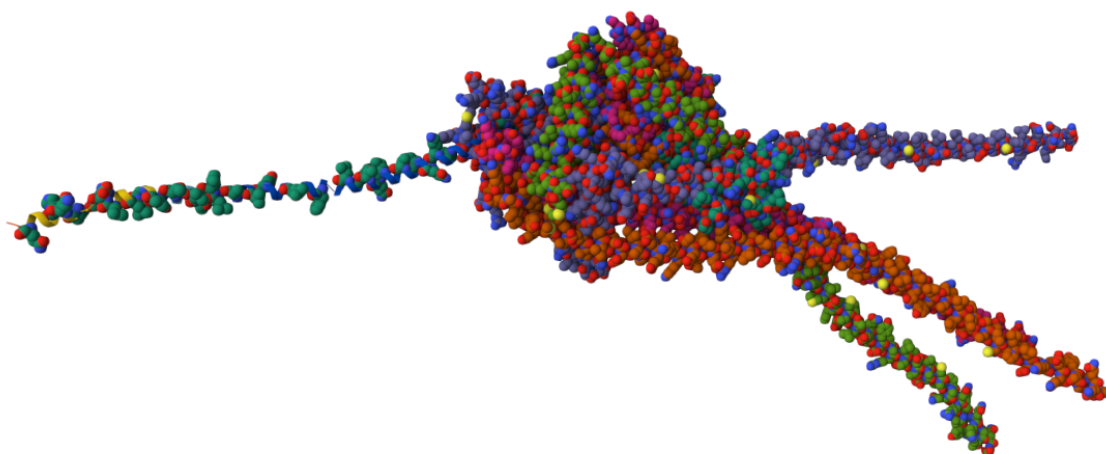
Q10

The top result of target BLAST is Kinesin-like protein 1, CHEMBL458, accession number P52732. It has 174 binding assays and 8 functional assays.

https://www.ebi.ac.uk/chembl/explore/assays/STATE_ID:ILpv0TNIersVv7xamqpxkw%3D%3D

Below is the ligand efficiency data.

Ligand Efficiencies for Target CHEMBL4581 (Kinesin-like protein 1)

Binding Efficiency Index (BEI)

Surface Efficiency Index (SEI)
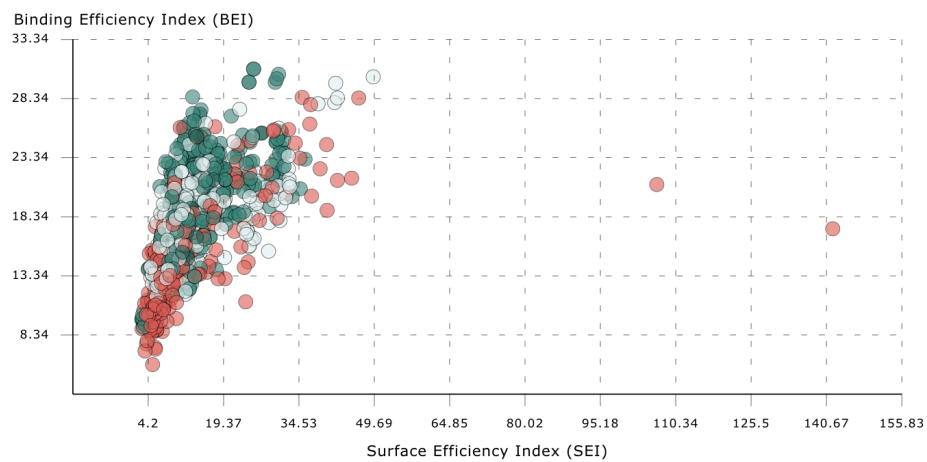
Standard Value nM

- ■ <1 (19 items)
- ■ [1,100) (334 items)
- ■ [100,1000) (336 items)
- ■ >=1000 (310 items)

Showing 999 data points.
Hover over a point to see the details.

Not showing 1 out of 1000 data points because of missing data. (Click to show details)

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