

## BIMM-143: INTRODUCTION TO BIOINFORMATICS

The find-a-gene project assignment  
[https://bioboot.github.io/bimm143\\_S20/](https://bioboot.github.io/bimm143_S20/)

Dr. Barry Grant

Email: [zashir@ucsd.edu](mailto:zashir@ucsd.edu)

A16125522

### **Questions:**

**[Q1]**

**Name:** Actin beta

**Accession:** NP\_001092

**Species:** Homo Sapiens

**Function Known:** Involved in intracellular signaling, cell motility and integrity. It is expressed everywhere since it is one of the two non-muscle cytoskeletal actins.

**[Q2]**

**Method:** TBLASTIN search against nematode ESTs

**Database:** Expressed Sequence Tags (est)

**Organism:** Nematodes (Taxid: 6231)

Translated BLAST: tblastn

blastn

blastp

blastx

**tblastn**

tblastx

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](#)

ref|NP\_001092

Query subrange [?](#)

From

To

Or, upload file

Choose File

no file selected [?](#)

Job Title

NP\_001092:actin, cytoplasmic 1 [Homo sapiens]

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

☐

Align two or more sequences [?](#)

Choose Search Set

Database

Expressed sequence tags (est) [?](#)

Organism

Optional

nematodes (taxid:6231)

☐

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

**Chosen match:** Accession JN376807.1, a 1363 base pair clone from *Meloidogyne incognita*. Score=773 bits(1995), E-Value=0.0, See below for alignment details.



[← Edit Search](#)

[Save Search](#)

[Search Summary](#) ▾

[How to read this report?](#)

[BLAST Help Videos](#)

[Back to Traditional Results Page](#)

**i** Your search is limited to records that include: nematodes (taxid:6231)

Job Title **NP\_001092:actin, cytoplasmic 1 [Homo sapiens]**

RID [4W2PSPBB013](#) Search expires on 05-01 14:29 pm  
[Download All](#) ▾

Program TBLASTN [Citation](#) ▾

Database est [See details](#) ▾

Query ID [NP\\_001092.1](#)

Description actin, cytoplasmic 1 [Homo sapiens]

Molecule type amino acid

Query Length 375

Other reports [?](#)

### Filter Results

Organism only top 20 will appear

☐ exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity

to

E value

to

Query Coverage

to

[Filter](#)

[Reset](#)

[Descriptions](#)

**[Graphic Summary](#)**

[Alignments](#)

[Taxonomy](#)

[hover to see the title](#) [click to show alignments](#)

Alignment Scores

☐ < 40

☐ 40 - 50

☐ 50 - 80

☐ 80 - 200

☐ >= 200

[?](#)

3 sequences selected [?](#)

### Distribution of the top 1 Blast Hits on 3 subject sequences



[Feedback](#)

Descriptions

Graphic Summary

Alignments

Taxonomy

Alignment view

Pairwise

Restore defaults

Download

3 sequences selected

Download

GenBank

Graphics

NextPreviousDescriptions

JN376807 Meloidogyne incognita J2 cDNA library Meloidogyne incognita cDNA clone mij2mn003, mRNA sequence

Sequence ID: JN376807.1Length: 1363Number of Matches: 1

Range 1: 180 to 1304GenBankGraphicsNext MatchPrevious Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
773 bits(1995)	0.0	Compositional matrix adjust.	365/375(97%)	374/375(99%)	0/375(0%)	-3
Query 1	MDDDIAALVVDNGSGMCKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEAQS					60
Sbjct 1304	MD+++AALVVDNGSGMCKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEAQS					1125
Query 61	KRGILTLPYIEHGIVTNWDDMEKIWHHTFYNELRVAPEEHPVLLTEAPLNPKANREKMT					120
Sbjct 1124	KRGILTLPYIEHGIVTNWDDMEKIWHHTFYNELRVAPEEHPVLLTEAPLNPKANREKMT					945
Query 121	QIMFETFNTPMYVAIQAVLSLYASGRITGIVMDSGDGVTHTVPIYEGYALPHAILRLDL					180
Sbjct 944	QIMFETFNTPMYVAIQAVLSLYASGRITGIVMDSGDGVTHTVPIYEGYALPHAILRLDL					765
Query 181	AGRDLDYLMKILTERGYSFTTTAEREIVRDIKEKLCYVALDFEQEMATAASSSSLEKSY					240
Sbjct 764	AGRDLDYLMKILTERGYSFTTTAEREIVRDIKEKLCYVALDFEQEMATAASSSSLEKSY					585
Query 241	ELPDGQVITIGNERFRCPEALFQPSFLGMESCGIHETTFNSIMKCDVDIRKDLANTVLS					300
Sbjct 584	ELPDGQVITIGNERFRCPEALFQPSFLGMESGIHETTFNSIMKCDVDIRKDLANTVLS					405
Query 301	GGTTMPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILASLSTFQQMWISKQ					360
Sbjct 404	GGTTMPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILASLSTFQQMWISKQ					225
Query 361	EYDESGPSIVHRKCF	375				
Sbjct 224	EYDESGPSIVHRKCF	180				

<gb |JN376807.1| Meloidogyne incognita J2 cDNA library Meloidogyne incognita cDNA clone mij2mn003, mRNA sequence, Sequence ID: JN376807.1, Length= 1363

Score=773 bits(1995), Expect =0.0, Method: Compositional matrix, Identities =365/375(97%), Positives = 374/375 (99%), Gaps = 0/375(0%) Frame = -3

Query 1  
MDDDIAALVVDNGSGMCKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEAQS 60  
MD+++AALVVDNGSGMCKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEAQS  
Sbjct 1304  
MDEEVAALVVDNGSGMCKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEAQS 1125

Query 61 KRGILTLKYPIEHGIVTNWDDMEKIWHHTFYNELRVAPEEHPVLLTEAPLNPKANREKMT 120  
KRGILTLKYPIEHGIVTNWDDMEKIWHHTFYNELRVAPEEHPVLLTEAPLNPKANREKMT

Sbjct 1124 KRGILTLKYPIEHGIVTNWDDMEKIWHHTFYNELRVAPEEHPVLLTEAPLNPKANREKMT  
945

Query 121 QIMFETFNTPAMYVAIQAVLSLYASGRTTGIVMDSGDGVTHTVPIYEGYALPHAILRLDL 180  
QIMFETFNTPAMYVAIQAVLSLYASGRTTGIV+DSGDGVTHTVPIYEGYALPHAILRLDL

Sbjct 944 QIMFETFNTPAMYVAIQAVLSLYASGRTTGIVLDSGDGVTHTVPIYEGYALPHAILRLDL 765

Query 181 AGRDLTDYLMKILTERGYSFTTTAEREIVRDIKEKLCYVALDFEQEMATAASSSSLEKSY 240  
AGRDLTDYLMKILTERGYSFTTTAEREIVRDIKEKLCYVALDFEQEMATAASSSSLEKSY

Sbjct 764 AGRDLTDYLMKILTERGYSFTTTAEREIVRDIKEKLCYVALDFEQEMATAASSSSLEKSY 585

Query 241 ELPDGQVITIGNERFRCPEALFQPSFLGMESCGIHETTFNSIMKCDVDIRKDLYANTVLS 300  
ELPDGQVIT+GNERFRCPE+LFQPSFLGMES GIHET++NSIMKCD+DIRKDLYANTVLS

Sbjct 584 ELPDGQVITVGNERFRCPESLFQPSFLGMESAGIHETSYNSIMKCDIDIRKDLYANTVLS 405

Query 301 GGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILASLSTFQQMWISKQ 360  
GGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILASLSTFQQMWISKQ

Sbjct 404 GGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILASLSTFQQMWISKQ 225

Query 361 EYDESGPSIVHRKCF 375  
EYDESGPSIVHRKCF

Sbjct 224 EYDESGPSIVHRKCF 180

**[Q3]**

Chosen sequence

>180-1304\_4 Meloidogyne incognita J2 cDNA library Meloidogyne incognita cDNA clone  
mij2mn003, mRNA sequence  
MDEEVAALVVDNGSGMCKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEAQS  
KRGILTLKYPIEHGIVTNWDDMEKIWHHTFYNELRVAPEEHPVLLTEAPLNPKANREKMT  
QIMFETFNTPAMYVAIQAVLSLYASGRRTGIVLDSGDGVTHTVPIYEGYALPHAILRLDL  
AGRDLTDYLMKILTERGYSFTTTAEREIVRDIKEKLCYVALDFEQEMATAASSSSLEKSY  
ELPDGQVITVGNERFRCPESLFQPSFLGMESAGIHETSYNSIMKCDIDIRKDLYANTVLS  
GGTTMYPGIADRMQKEITALAPSTMKIKIIPPERKYSVWIGGSILASLSTFQQMWISKQ  
EYDESGPSIVHRKCF

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:** *Strongyloides ratti*

**Species:** Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Nematoda; Chromadorea; Rhabditida; Tylenchina; Panagrolaimomorpha; Strongyloidea; Strongyloidea; Strongyloidea

[Q4]

A BLASTP search against NR database (see setup in first screen-shot below) yielded a top hit result is to a protein from *Meloidogyne enterolobii*

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



National Library of Medicine  
National Center for Biotechnology Information

Log in

BLAST® » blastp suite

HomeRecent ResultsSaved StrategiesHelp

blastnblastpblastxtblastntblastx

Standard Protein BLAST

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

Reset pageBookmark

Enter Query Sequence

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

>180-1304\_4 Meloidogyne incognita J2 cDNA library Meloidogyne incognita cDNA clone  
mij2mn003, mRNA sequence  
MDEEVAALVVDNGSGMCKAGFAGDDAPRAVFPISIVGRPRHQGVMMVGQKDSYVGDEAQSQR  
GILTLKYP  
IEHGIVTNWDDMEKIWHHTFYNELRVAPEEHPVLLTEAPLNPKANREKMTQIMFETNTPAMYVAI  
GAM

Query subrange [?](#)  
From   
To

Or, upload file  no file selected [?](#)

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?](#)

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

Standard

Database ?

Organism Optional  ☐ exclude   
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

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. [?](#)

The top result is to a protein from *Meloidogyne enterolobii*, see second screen shot below for alignment details:



[← Edit Search](#)

[Save Search](#)

[Search Summary ▾](#)

[? How to read this report?](#)

[▶ BLAST Help Videos](#)

[↶ Back to Traditional Results Page](#)

Job Title	180-1304_4 Meloidogyne incognita J2 cDNA library...
RID	<a href="#">4W7A1SE9013</a> Search expires on 05-01 15:48 pm <a href="#">Download All ▾</a>
Program	BLASTP <a href="#">? Citation ▾</a>
Database	nr <a href="#">See details ▾</a>
Query ID	lcl Query_66506
Description	180-1304_4 Meloidogyne incognita J2 cDNA library Meloidc
Molecule type	amino acid
Query Length	375
Other reports	<a href="#">Distance tree of results</a> <a href="#">Multiple alignment</a> <a href="#">MSA viewer</a> <a href="#">?</a>

### Filter Results

**Organism** only top 20 will appear ☐ exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

---

**Percent Identity**  to  **E value**  to  **Query Coverage**  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 ▾



<input checked="" type="checkbox"/> select all 100 sequences selected		<a href="#">GenPept</a>	<a href="#">Graphics</a>	<a href="#">Distance tree of results</a>	<a href="#">Multiple alignment</a>	<a href="#">MSA Viewer</a>			
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	<a href="#">actin [Meloidogyne enterolobii]</a>	<a href="#">Meloidogyne e...</a>	784	784	100%	0.0	100.00%	375	<a href="#">AHB20148.1</a>
<input checked="" type="checkbox"/>	<a href="#">hypothetical protein AAVH_01235 [Aphelenchus avenae]</a>	<a href="#">Aphelenchus a...</a>	781	781	99%	0.0	99.73%	473	<a href="#">KAH7732336.1</a>
<input checked="" type="checkbox"/>	<a href="#">actin 1 [Ditylenchus destructor]</a>	<a href="#">Ditylenchus de...</a>	780	780	99%	0.0	99.73%	376	<a href="#">ACT78498.1</a>
<input checked="" type="checkbox"/>	<a href="#">hypothetical protein L596_004117 [Steinernema carpocapsae]</a>	<a href="#">Steinernema c...</a>	780	780	99%	0.0	99.20%	427	<a href="#">TMS37116.1</a>

Feedback



Alignment view Pairwise [Restore defaults](#) [Download](#)

100 sequences selected

[Download](#) [GenPept](#) [Graphics](#) [Next](#) [Previous](#) [Descriptions](#)

**actin [Melioidogone enterolobii]**  
Sequence ID: [AHB20148.1](#) Length: 375 Number of Matches: 1  
[See 2 more title\(s\)](#) [See all Identical Proteins\(IPG\)](#)

Range 1: 1 to 375 [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
784 bits(2024)	0.0	Compositional matrix adjust.	375/375(100%)	375/375(100%)	0/375(0%)

Query 1 MDEEVAALVDNNGSMCKAGFAGDDAPRAVFPISVGRPRHQGMVGMGQKDSYVGDEAQS  
Sbjct 1 MDEEVAALVDNNGSMCKAGFAGDDAPRAVFPISVGRPRHQGMVGMGQKDSYVGDEAQS

Query 61 KRGILTLYKPIEHGIVTNDDMEKIWHHTFYNELRVAPEEHPVLLTEAPLNPKANREKMT  
Sbjct 61 KRGILTLYKPIEHGIVTNDDMEKIWHHTFYNELRVAPEEHPVLLTEAPLNPKANREKMT

Query 121 QIMFETNTPAMYVAIQAVLSLYASGRTTGIVLSDGSGVTHTVPIYEGYALPHAILRLDL  
Sbjct 121 QIMFETNTPAMYVAIQAVLSLYASGRTTGIVLSDGSGVTHTVPIYEGYALPHAILRLDL

Query 181 AGRLTDYLMKILTERGYSFTTTAEREIVRDIKEKLCYVALDFEQEMATAASSSSLEKSY  
Sbjct 181 AGRLTDYLMKILTERGYSFTTTAEREIVRDIKEKLCYVALDFEQEMATAASSSSLEKSY

Query 241 ELPDGGVITVGNFRFCPELFPQSFGLMESAGIHETSYNSIMKCDIDIRKDYANTVLS  
Sbjct 241 ELPDGGVITVGNFRFCPELFPQSFGLMESAGIHETSYNSIMKCDIDIRKDYANTVLS

Query 381 GGTMYPGIADRMQKEITALAPSTMKIKIIPAPPERKYSVWIGGSILASLSTFQQMWISKQ  
Sbjct 381 GGTMYPGIADRMQKEITALAPSTMKIKIIPAPPERKYSVWIGGSILASLSTFQQMWISKQ

Query 361 EYDESGPSIVHRKCF 375  
Sbjct 361 EYDESGPSIVHRKCF 375

[Download](#) [GenPept](#) [Graphics](#) [Next](#) [Previous](#) [Descriptions](#)

**hypothetical protein AAVH\_01235 [Aphelenchus avenae]**  
Sequence ID: [KAH7732336.1](#) Length: 473 Number of Matches: 1

Range 1: 100 to 473 [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
781 bits(2017)	0.0	Compositional matrix adjust.	373/374(99%)	374/374(100%)	0/374(0%)

Query 2 DEEVAALVDNNGSMCKAGFAGDDAPRAVFPISVGRPRHQGMVGMGQKDSYVGDEAQS  
Sbjct 100 DEEVAALVDNNGSMCKAGFAGDDAPRAVFPISVGRPRHQGMVGMGQKDSYVGDEAQS

Query 62 RGILTLYKPIEHGIVTNDDMEKIWHHTFYNELRVAPEEHPVLLTEAPLNPKANREKMTQ  
Sbjct 160 RGILTLYKPIEHGIVTNDDMEKIWHHTFYNELRVAPEEHPVLLTEAPLNPKANREKMTQ

Query 122 IMFETNTPAMYVAIQAVLSLYASGRTTGIVLSDGSGVTHTVPIYEGYALPHAILRLDLA  
Sbjct 220 IMFETNTPAMYVAIQAVLSLYASGRTTGIVLSDGSGVTHTVPIYEGYALPHAILRLDLA

Query 182 GRDLTDYLMKILTERGYSFTTTAEREIVRDIKEKLCYVALDFEQEMATAASSSSLEKSYE  
Sbjct 280 GRDLTDYLMKILTERGYSFTTTAEREIVRDIKEKLCYVALDFEQEMATAASSSSLEKSYE

**[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.

**[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.

**[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.

**[Q9]** Generate a molecular figure of one of your identified PDB structures using the **NGL viewer** online (or **VMD/PyMol**). 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?

**[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?

**Scoring Rubric:**

[45 total points available]

**Q1 (4 points)**

Protein name	1
Species	1
Accession number	1
Function known	1

**Q2 (6 points)**

Blast method	1
Database searched	1
Limits applied	1
Search output list (top hits)	1
Alignment of choice	1
Evalue and other alignment stats	1

**Q3 (3 points)**

Protein sequence of choice matches Subject above	1
Name in header	1

Species	1
<b>Q4</b> (3 point)	
Blastp output list with identities & Evalue	1
Top alignment shown with alignment statistics	1
Results indicates a “novel” gene found	1
<b>Q5</b> (3 points)	
MSA labeled with useful names	1
MSA trimmed appropriately (i.e. no gap overhangs)	1
Pasted MSA fits report page width (i.e. font, format)	1
<b>Q6</b> (1 point)	
Figure illustrates sequence clustering pattern	1
<b>Q7</b> (10 points)	
Heatmap figure included in report	5
Heatmap is legible (i.e. no labels obscured)	5
<b>Q8</b> (10 points)	
PDB identifiers from multiple species reported	5
Annotation of PDB source, resolution and technique	4
Annotation of Evalue and Sequence Identity	1
<b>Q9</b> (4 points)	
Structure figure provided	2
Uses white background for molecular figure	1
Figure of high resolution (i.e. not just snapshot)	1
<b>Q10</b> (1 point)	
Evidence of ChEMBL searches	1