Xuqian Tan A13642946 xut006@ucsd.edu

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.

If you do not have a favorite protein, select human RBP4 or KIF11. Do not use beta globin as this is in the worked example report that I provide you with online.

Name: Cellular tumor antigen p53 isoform b

Accession: NP_001119586 Species: Homo Sapiens

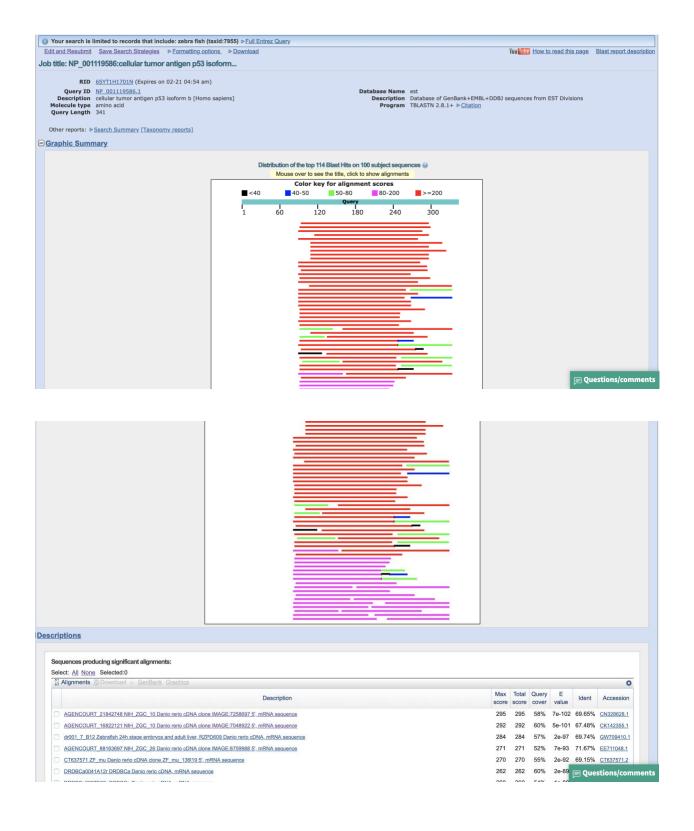
[Q2] Perform a BLAST search against a DNA database, such as a database consisting of genomic DNA or ESTs. The BLAST server can be at NCBI or elsewhere. Include details of the BLAST method used, database searched and any limits applied (e.g. Organism).

Method: TBLASTN (2.7.1) search against nematode ESTs

Database: Expressed Sequence Tags (est) Organism: Zebra Fish (taxid:7955)



Chosen match: Accession CK142355, a 785 base pair clone from Danio rerio. See below for alignment details.



■Download ▼ GenBank Graphics

AGENCOURT_16822121 NIH_ZGC_10 Danio rerio cDNA clone IMAGE:7048922 5', mRNA sequence Sequence ID: CK142355.1 Length: 785 Number of Matches: 1

Range 1: 23 to 640 GenBank Graphics W Next Match A Previous Match								
Score		Expect Method	Identities	Positives	Gaps	Frame		
292 bits(747)) 5e-101 Compositional matrix adjust.	139/206(67%)	158/206(76%)	0/206(0%)	+2		
Query	92	PLSSSVPSQKTYQGSYGFRLGFLHSGTAKSVT P +S+VP Y G++GFRL F SGTAKSVT						
Sbjct	23	PTTSTVPETSDYPGAHGFRLRFPQSGTAKSV						
Query	152	PPGTRVRAMAIYKQSQHMTEVVRRCPHHERCS P G+ VRA AIYK+S+H+ EVVRRCPHHER		/EGNLRVEYLDDRN /EGN R Y +D				
Sbjct	203	PQGSVVRATAIYKKSEHVAEVVRRCPHHERTI						
Query	212	FRHSVVVPYEPPEVGSDCTTIHYNYMCNSSCN RHSV VPYE P++G++ TT+ NYMCNSSCN	MGGMNRRPILTIITI MGGMNRRPILTIITI					
Sbjct	383	LRHSVFVPYEAPQLGAEWTTVLLNYMCNSSC			_			
Query	272	VRVCACPGRDRRTEEENLRKKGEPHH 297 VRVCACPGRDR+T E+ +K P H						
Sbjct	563	VRVCACPGRDRTT ET TR P H VRVCACPGRDRKTXGEHFQKT*RPKH 640						

Alignment details:

AGENCOURT_16822121 NIH_ZGC_10 Danio rerio cDNA clone IMAGE:7048922 5', mRNA sequence

Sequence ID: CK142355.1Length: 785Number of Matches: 1

Related Information

<u>UniGene</u>-clustered expressed sequence tags

Range 1: 23 to 640GenBankGraphics

Score		Expect	Method	Identities	Positives	Gaps	Frame
292 bits	s(747)	5e-101	Compositional matrix adjust.	139/206(67%)	158/206(76%)	0/206(0%).	+2
Query	92	PLSSSVPSQK	KTYQGSYGFRLGFLHSGTA	KSVTCTYSPALNK	MFCQLAKTCPVQ	LWVDSTP 15	51
		P +S+VP	Y G++GFRL F SGTA	KSVTCTYSP LNK	+ CQLAKTCPVQ	+ VD P	
Sbjct	23	PTTSTVPETS	SDYPGAHGFRLRFPQSGTA	KSVTCTYSPDLNK	LLCQLAKTCPVQ	MVVDVAP 20	02
Query	152	PPGTRVRAMA	AIYKQSQHMTEVVRRCPHH	IERCSDSDGLAPPQ	HLIRVEGNLRVE	YLDDRNT 2	11
		P G+ VRA A	AIYK+S+H+ EVVRRCPHH	IER D D LAP	HLIRVEGN R	Y +D T	
Sbjct	203	PQGSVVRATA	AIYKKSEHVAEVVRRCPHH	IERTPDGDNLAPAG	HLIRVEGNQRAN	YREDNIT 38	32
Query	212	FRHSVVVPYE	PPEVGSDCTTIHYNYMCN	ISSCMGGMNRRPIL	TIITLEDSSGNL	LGRNSFE 2	71
		RHSV VPYE	E P++G++ TT+ NYMCN	ISSCMGGMNRRPIL	TIITLE G L	LGR SFE	
Sbjct	383	LRHSVFVPYE	CAPQLGAEWTTVLLNYMCN	ISSCMGGMNRRPIL	TIITLETXEGQL	LGRRSFE 56	52
Query	272			297			
			DR+T E+ +K P H				
Sbjct	563	VRVCACPGRD	ORKTXGEHFQKT*RPKH	640			

[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:

>A. Danio rerio protein (sequence taken from BLAST result)

PTTSTVPETSDYPGAHGFRLRFPQSGTAKSVTCTYSPDLNKLLCQLAKTCPVQMVVDVAPPQGSVVRATAIYKKSEHVAEVVRRCP HHERTPDGDNLAPAGHLIRVEGNQRANYREDNITLRHSVFVPYEAPQLGAEWTTVLLNYMCNSSCMGGMNRRPILTIITLETXEGQ LLGRRSFEVRVCACPGRDRKTXGEHFQKT*RPKH

Name: Danio rerio cDNA clone

Species: Danio rerio

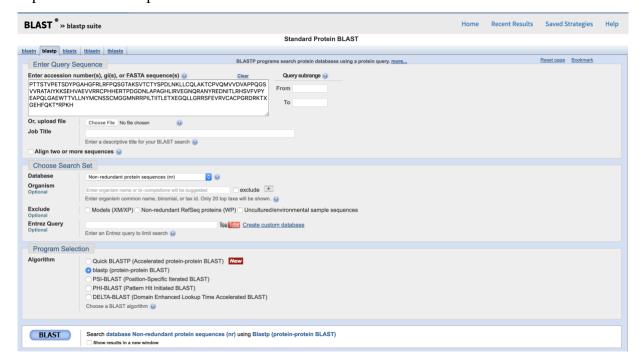
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

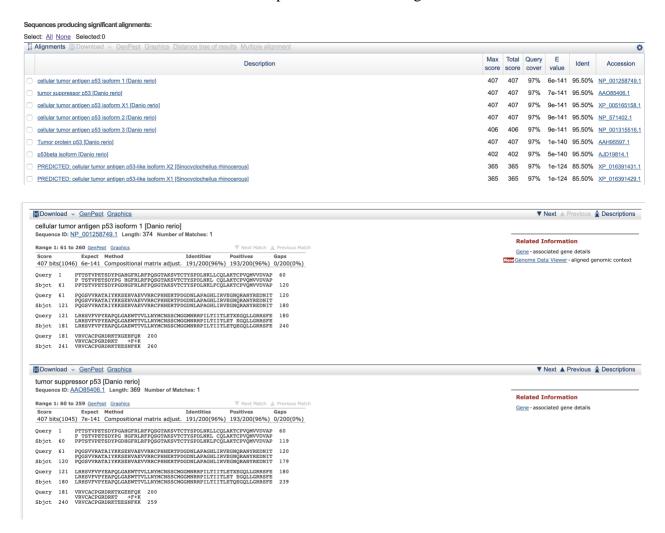
Cypriniformes; Cyprinidae; Danio.

[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. Details:

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



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



[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.

Re-labeled sequences for alignment:

>Human_p53 gi|9606|ref|NP_001119586.1| cellular tumor antigen p53 isoform b [Homo sapiens]

PLSSSVPSQKTYQGSYGFRLGFLHSGTAKSVTCTYSPALNKMFCQLAKTCPVQLWVDSTPPPGTRVRAMAIYKQSQH MTEVVRRCPHHERCSDSDGLAPPOHLIRVEGNLRVEYLDDRNTFRHSVVVPYEPPEVGSDCTTIHYNYMCNSSCMGG MNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGRDRRTEEENLRKKGEPHH

>Danio rerio protein (sequence taken from BLAST result) PTTSTVPETSDYPGAHGFRLRFPQSGTAKSVTCTYSPDLNKLLCQLAKTCPVQMVVDVAPPQGSVVRATAIYKKSEH VAEVVRRCPHHERTPDGDNLAPAGHLIRVEGNQRANYREDNITLRHSVFVPYEAPQLGAEWTTVLLNYMCNSSCMGG MNRRPILTIITLETXEGOLLGRRSFEVRVCACPGRDRKTXGEHFOKT*RPKH

>Ray-finned fish gi|307959|ref|XP 016391431.1| PREDICTED: cellular tumor antigen p53-like isoform X2 [Sinocyclocheilus rhinocerous] PPTASVPVATDYPGEHGFKLGFPQSGTAKSVTCTYSADLNKLFCQLAKTCPVQMVVDVAPPQGSVIRATAIYKKSEH VAEVVRRCPHHERTPDGDALAPPAHLIRMEGNSRAVYREDDITYRHSVVVPYEAPQLGAEWTTVLYNFMCNSSCMGG MNRRPILTIITLETHEGQLLGRRSFEVRVCACPGRDRKTEESNFRK

>Golden-line barbel qi|1608454|ref|XP 016321636.1| PREDICTED: cellular tumor antigen p53-like isoform X2 [Sinocyclocheilus anshuiensis] PPTASVPVATDYPGEHGFKLGFPOSGTAKSVTCTYSADLNKLFCOLAKTCPVOMVVDVAPPOGSVIRATAIYKKSEH VAEVVRRCPHHERTPDGDALAPPAHLIRMEGNSRAVYREDDITYRHSVVVPYEAPQLGAEWTTVLYNFMCNSSCMGG MNRRPILTIITLETHEGQLLGRRSFEVRVCACPGRDRKTEESNFRK

>Grass carp gi|7959|.|ANP93608.1| tumor antigen p53 [Ctenopharyngodon idella] PPTSTVPIATDYPGEHGFKLQFPQSGTAKSVTCTYSPDLNKLFCQLAKTCPVQMVVDVAPPQGSVLRATAIYKKSEH VAEVVRRCPHHERTPDTDGLAPAAHLIRVEGNLRATYKEDDVTSRHSVVVPYEAPOLGAGFTTVLYNYMCNSSCMGG MNRRPILTIITLETODGOMLGRRSFEVRVCACPGRDRKTEESNFRK

>Gold fish gi|7957|ref|XP 026112491.1| cellular tumor antigen p53-like isoform X2 [Carassius auratus]

PPTASVPVATDYPGEHGFKLGFPQSGTAKSVTCTYSSDLNKLFCQLAKTCPVQMLVDVAPPQGSVIRATAIYKKSEH VAEVVRRCPHHERTPDGDVLAPPAHLIRVEGNSLAVYREDDITHRHSVAVPYEAPQLGAEWTTVLYNFMCNSSCMGG MNRRPILTIITLETHEGQLLGRRSFEVRVCACPGRDRKTEESNFGK

>Common carp qi|7962|ref|XP 018953153.1| PREDICTED: cellular tumor antiqen p53-like isoform X2 [Cyprinus carpio]

PPTSSVPVATDYPGEHGFKLGFPQSGTAKSVTCTYSSDLNKLFCQLAKTCPVQMLVDVAPPQGSVIRATAIYKKSEH VADVVRRCPHHERTLDGDGLAPPAHLIRVEGNSRAVYREDDMTYRHSVVVPYEAPOLGAEWTTVLYNFMCNSSCMGG MNRRPILTIITLETHDGQLLGRRSFEVRVCACPGRDRKTEESNFRK

Alignment:

Obtained using MUSCLE (version 3.8) at EBI:

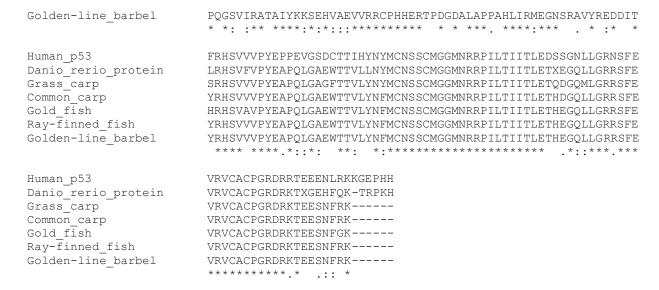
CLUSTAL multiple sequence alignment by MUSCLE (3.8)

Human p53 Common carp Gold fish

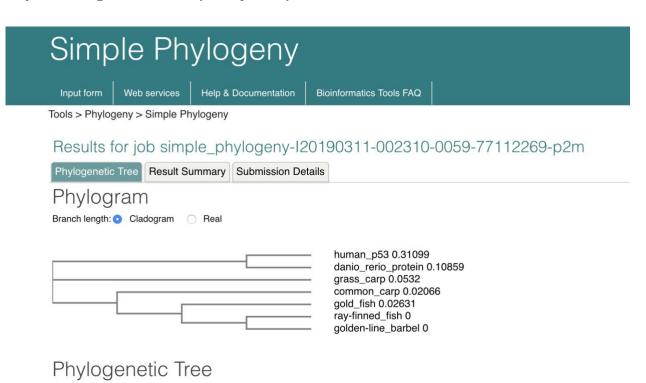
PLSSSVPSOKTYOGSYGFRLGFLHSGTAKSVTCTYSPALNKMFCOLAKTCPVOLWVDSTP Danio_rerio_protein PTTSTVPETSDYPGAHGFRLGFLGFLGFSGTAKSVTCTYSPALNKMFCQLAKTCPVQLWVDSTP
Grass_carp PTTSTVPETSDYPGHGFKLQFPQSGTAKSVTCTYSPDLNKLLCQLAKTCPVQMVVDVAP PPTSSVPVATDYPGEHGFKLGFPQSGTAKSVTCTYSSDLNKLFCQLAKTCPVQMLVDVAP PPTASVPVATDYPGEHGFKLGFPQSGTAKSVTCTYSSDLNKLFCQLAKTCPVQMLVDVAP Ray-finned_fish PPTASVPVATDYPGEHGFKLGFPQSGTAKSVTCTYSADLNKLFCQLAKTCPVQMVVDVAP
Golden-line_barbel PPTASVPVATDYPGEHGFKLGFPQSGTAKSVTCTYSADLNKLFCQLAKTCPVQMVVDVAP

Human p53 Danio_rerio_protein Grass carp Common carp Gold fish Ray-finned fish

PPGTRVRAMAIYKQSQHMTEVVRRCPHHERCSDSDGLAPPQHLIRVEGNLRVEYLDDRNT PQGSVVRATAIYKKSEHVAEVVRRCPHHERTPDGDNLAPAGHLIRVEGNQRANYREDNIT PQGSVLRATAIYKKSEHVAEVVRRCPHHERTPDTDGLAPAAHLIRVEGNLRATYKEDDVT PQGSVIRATAIYKKSEHVADVVRRCPHHERTLDGDGLAPPAHLIRVEGNSRAVYREDDMT PQGSVIRATAIYKKSEHVAEVVRRCPHHERTPDGDVLAPPAHLIRVEGNSLAVYREDDIT POGSVIRATAIYKKSEHVAEVVRRCPHHERTPDGDALAPPAHLIRMEGNSRAVYREDDIT

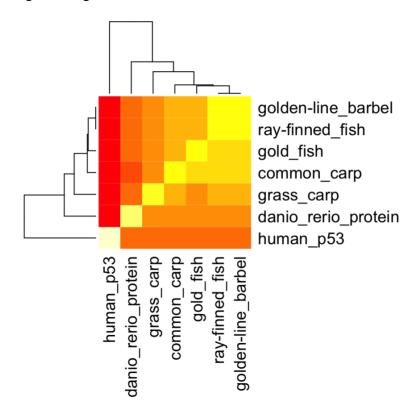


[Q6] Create a phylogenetic tree, using either a parsimony or distance-based approach. Bootstrapping and tree rooting are optional. Use "simple phylogeny" online from the EBI or any respected phylogeny program (such as MEGA, PAUP, or Phylip). Paste an image of your Cladogram or tree output in your report.



[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.

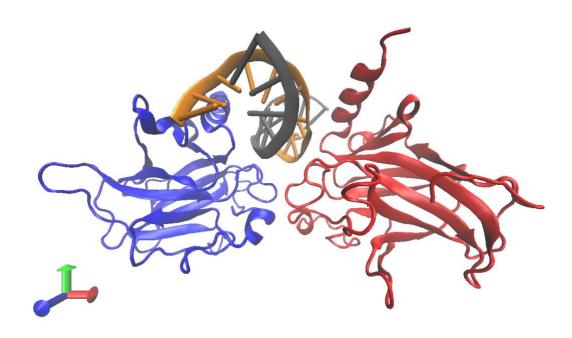


[Q8] Using R/Bio3D (or an online blast server if you prefer), search the main protein structure database for the most similar atomic resolution structures to your aligned sequences. List the top 3 unique hits (i.e. not hits representing different chains from the same structure) along with their Evalue 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).

ID	Technique	Resolution	Source	Evalue	Identity
3EXJ	X-RAY	2.0	Mus	8.99e-89	74.603
	DIFFRACTION		musculus		
1GZH	X-RAY	2.6	Homo	9.26e-88	71.574
	DIFFRACTION		sapiens		
3LX5	X-RAY	1.9	Streptococcus	5.50e+00	29.577
	DIFFRACTION		thermophilus		

[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?

It would be likely to have similar structure to my "novel protein" since protein have over 70% sequence similarity.



[Q10] Perform a "Target" search of ChEMBEL (https://www.ebi.ac.uk/chembl/) withyour 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?

CHEMBL details 9 Binding Assay (CHEMBL3883306) and 0 Functional Assays; No ligand efficiency data.

https://www.ebi.ac.uk/chembl/target/inspect/CHEMBL3883306

ChEMBL Assay	Assay Source	Assay Type	Assay Organism	Description	Activity Count	Reference	Z
CHEMBL3223544	Scientific Literature	В	Homo sapiens	Inhibition of MDM2/MDMX-p53 interaction in human SJSA-1 cells expressing amplified MDM2 and wild-type p53 assessed as induction of p21 protein expression at 5 uM after 4 hrs by western blot analysis	3	MedChemComm, (2013) 4:9:1297	
CHEMBL3223540	Scientific Literature	В	Homo sapiens	Inhibition of MDM2/MDMX-p53 interaction in human A2780 CP70 cells expressing p53 mutant gene assessed as induction of MDM2 protein expression at 0.2 to 5 uM after 4 hrs by western blot analysis	3	MedChemComm, (2013) 4:9:1297	~
CHEMBL3223543	Scientific Literature	В	Homo sapiens	Inhibition of MDM2/MDMX-p53 interaction in human A2780 cells expressing wild-type p53 assessed as induction of MDM2 protein expression at 0.2 to 5 uM after 4 hrs by western blot analysis	3	MedChemComm, (2013) 4:9:1297	~
CHEMBL3223542	Scientific Literature	В	Homo sapiens	Inhibition of MDM2/MDMX-p53 interaction in human A2780 cells expressing wild-type p53 assessed as induction of p53 protein expression at 0.2 to 5 uM after 4 hrs by western blot analysis	3	MedChemComm, (2013) 4:9:1297	2
CHEMBL3223538	Scientific Literature	В	Homo sapiens	Inhibition of MDM2/MDMX-p53 interaction in human A2780 CP70 cells expressing p53 mutant gene assessed as induction of p21 protein expression at 0.2 to 5 uM after 4 hrs by western blot analysis	3	MedChemComm, (2013) 4:9:1297	2
CHEMBL3223546	Scientific Literature	В	Homo sapiens	Inhibition of MDM2/MDMX-p53 interaction in human SJSA-1 cells expressing amplified MDM2 and wild-type p53 assessed as induction of MDM2 protein expression at 5 uM after 4 hrs by western blot analysis	3	MedChemComm, (2013) 4:9:1297	
CHEMBL3223541	Scientific Literature	В	Homo sapiens	Inhibition of MDM2/MDMX-p53 interaction in human A2780 cells expressing wild-type p53 assessed as induction of p21 protein expression at 0.2 to 5 uM after 4 hrs by western blot analysis	3	MedChemComm, (2013) 4:9:1297	
CHEMBL3223545	Scientific Literature	В	Homo sapiens	Inhibition of MDM2/MDMX-p53 interaction in human SJSA-1 cells expressing amplified MDM2 and wild-type p53 assessed as induction of p53 protein expression at 5 uM after 4 hrs by western blot analysis	3	MedChemComm, (2013) 4:9:1297	Ø
CHEMBL3223539	Scientific Literature	В	Homo sapiens	Inhibition of MDM2/MDMX-p53 interaction in human A2780 CP70 cells expressing p53 mutant gene assessed as induction of p53 protein expression at 0.2 to 5 uM after 4 hrs by western blot analysis	3	MedChemComm, (2013) 4:9:1297	2

Showing 1 to 9 of 9 entries

https://www.ebi.ac.uk/chembl/assay/results/1/chembl_id/asc/tab/keyword