

# FIND-A-GENE PROJECT ASSIGNMENT

**Name:** Simon Midtvedt

**PID:** U07974645

**UCSD email:** [ax005991@acsmail.ucsd.edu](mailto:ax005991@acsmail.ucsd.edu)

**Regular email:** [s.midtvedt@gmail.com](mailto:s.midtvedt@gmail.com)

## [Q1]

**Name:** Immunoglobulin heavy constant mu (IgM)

**Accession:** P01871

**Species:** Homo sapiens

## [Q2]

**Method:** TBLASTN (2.8.1+) search against ESTs

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

**Organism:** Homo sapiens excluded



Descriptions

Sequences producing significant alignments:  
Select: All None Selected:0

Alignments Download GenBank Graphics

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	FS723360 full-length enriched mouse cDNA library Macaca mulatta cDNA clone 16_E7-FC1_055 5' mRNA sequence	543	605	95%	0.0	92.45%	FS723360.1
<input type="checkbox"/>	ILLUMIGEN_MCOQ_45084 Katze MMJ Macaca mulatta cDNA clone IBIUW:19769 5' similar to Bases 608 to 956 highly similar to human Unigene Hs.525648 mRNA sequence	540	601	95%	0.0	93.09%	COS83591.1
<input type="checkbox"/>	BW976705 full-length enriched swine cDNA library adult intestine Sus scrofa cDNA clone ITT010008C08 5' mRNA sequence	406	467	95%	2e-138	65.73%	BW976705.1
<input type="checkbox"/>	264930929 Bottlenose dolphin spleen cDNA library Tursiops truncatus cDNA clone TBC16#1003002773-B10 5' mRNA sequence	404	468	95%	8e-138	75.30%	GT115761.1
<input type="checkbox"/>	FM080331 etnoaor Rattus norvegicus cDNA clone etnoaorP0037E04 5' mRNA sequence	360	477	95%	4e-121	74.89%	FM080331.1
<input checked="" type="checkbox"/>	FQ182199 Rattus norvegicus spleen Sprague-Dawley Rattus norvegicus cDNA clone TL0AE65YJ12 3' mRNA sequence	343	456	95%	3e-114	74.19%	FQ182199.1
<input type="checkbox"/>	HX229255 full-length enriched swine cDNA library adult inguinal lymph node Sus scrofa cDNA clone ILNT10033A09 mRNA sequence	401	461	95%	2e-138	65.96%	HX229255.1
<input type="checkbox"/>	AGENCOURT_118846718 NIH_MGC_248 Rattus norvegicus cDNA clone IMAGE:9112822 5' mRNA sequence	382	422	95%	9e-129	61.71%	EX488007.1
<input type="checkbox"/>	Ferret_009_G13 Ferret library from spleen liver and thymus Mustela putorius furo cDNA mRNA sequence	446	571	94%	3e-152	59.06%	GD182759.1
<input type="checkbox"/>	PP0ADA95YG10FM1 Chlorocebus sabaeus PBMCs EST Library Chlorocebus sabaeus cDNA mRNA sequence	431	560	94%	4e-149	95.81%	JK124580.1
<input type="checkbox"/>	LIB30321_012_A09_SP6_1 LIB30321 Canis lupus familiaris cDNA clone LIB30321_012_A09 mRNA sequence	369	488	94%	1e-124	80.28%	DN272880.1

Chosen match: Accession FQ182199.1, a 765 base pair cDNA clone from Rattus Norvegicus (Norwegian rat).

Alignment details:

Download GenBank Graphics Sort by: E value

FQ182199 Rattus norvegicus spleen Sprague-Dawley Rattus norvegicus cDNA clone TL0AE65YJ12 3' mRNA sequence  
Sequence ID: FQ182199.1 Length: 765 Number of Matches: 3

Range 1: 114 to 764 GenBank Graphics

▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
343 bits(881)	3e-114	Compositional matrix adjust.	161/217(74%)	186/217(85%)	0/217(0%)	-2
Query	237	TKSKTKLTCVLTDLTTYDSVTISWTRQNGEAVKTHNTISESHPNATFSAVGEASICEDDDWN			296	
		TK KL+CLVT+L TYD++ ISW+ ++GE ++T+T I ESHPN TFSAVG AS+C +DW+				
Sbjct	764	TKFAKLSCLVTNLATYDTLNISSWSSKSGEPLETNTKIMESHNPNGTFSAVGVASVCMEDWD			585	
Query	297	SGERFTCTVTHDLPSPKQTISRPKGVALHRPDVYLLPPAREQLNLRESATITCLVTGF			356	
		+ + F CTVTH DLPSP K+ IS+P VA H P VYLLPPAREQL LRESAT+TCLV GF				
Sbjct	584	NKKEFVCTVTHRDLPSPQKKFISKPNVAKHPPAVYLLPPAREQLILRESATVTCVLKGF			405	
Query	357	SPADIFVQWLQRGQPLSSDKYVTSAPMPEPGAPGLYFTHSILTVTEEEWNSGETYTCVV			416	
		SPAD+FYQW+QRGQPLS +KYVTSAPMPEP APC YF HSILTV+EEEW+GETYTCVV				
Sbjct	404	SPADIFVQWLQRGQPLSSDKYVTSAPMPEPGAPGLYFTHSILTVTEEEWNSGETYTCVV			225	
Query	417	HEALPHMVTERTVDKSTGKPTLYNVSLIMSDTGTCY 453				
		HEALP+ VTERTVDKSTGKPTLYNVSL+MSDT GTCY				
Sbjct	224	HEALPHMVTERTVDKSTGKPTLYNVSLIMSDTGTCY 114				

Range 2: 156 to 755 GenBank Graphics

▼ Next Match ▲ Previous Match ▲ First Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
59.3 bits(142)	9e-07	Compositional matrix adjust.	52/212(25%)	96/212(45%)	21/212(9%)	-2
Query	130	SKLICOATGFSRQ-IQVSWLREGKQVSGVTTDQVQAEAKESGPTTYKVTSTLTIKESD			188	
		+KL C T + + +SW + G + T+ E+ +G T+ ++ D				
Sbjct	755	AKLSCLVTNLATYDTLNISSWSSKSGEPLETNTKIMESHNPNG--TFSAVGVASVCMED			591	
Query	189	WLGQSMFTCRVDHRLTFQGNASSMCVPDQDTAI--RVFAIPPSFASIFLTKSTKLTCV			246	
		W + F C V HR L O + P++ V+ +PP+ + L+ S +TCLV				
Sbjct	590	WONRKEFVCTVTHRDLPSPQKKFISKPNVAKHPPAVYLLPPAREQLILRESATVTCV			414	
Query	247	TDLTTYDSVTISWTRQNGEAVKTHNTISES---HPNA--TFSAVGEASICEDDWNNGERF			301	
		+ D + + W O G+ + + ++ + P A + E++WNSGE +				
Sbjct	413	KGFSPAD-IFVQWL-QRGQPLSSDKYVTSAPMPEPGAPGLYFTHSILTVTEEEWNSGETY			240	
Query	302	TCVTHDLPSPK-QTISRPKGVALHRPDVY 332				
		TC V H LP + +P+ + G +P+Y				
Sbjct	239	TCVVGHEALPHMVTERTVDKSTG----KPTLY 156				

Range 3: 213 to 764 GenBank Graphics

▼ Next Match ▲ Previous Match ▲ First Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
53.3 bits(127)	7e-05	Compositional matrix adjust.	48/189(25%)	76/189(40%)	10/189(5%)	-2
Query	21	TSSVAVGCLAQDFLP-DSITFSWKYKNNDSISSTRGFPVLRGGKYAATSQVLLPSKDVM			79	
		T + CL + D++ SW K+ + + G ++A + +D				
Sbjct	764	TKFAKLSCLVTNLATYDTLNISSWSSKSGEPLETNTKIMESHNPNGTFSAVGVASVCMED--			591	
Query	80	QGTDEHVVCVKVQHPN--GNKEKNVPLP-VIAELPPKVSFVFPDAGFFGNPKSKLICQA			136	
		+ VC V H + ++K + P +A+ PP V + P R+ + +C				
Sbjct	590	WONRKEFVCTVTHRDLPSPQKKFISKPNVAKHPPAVYLLPPAREQLILR-ESATVTCV			414	
Query	137	TGFSRQIQVSWLREGKQVSGVTTDQVQAEAKESG-PTTYKVTSTLTIKESDNLGQSMF			195	
		GFSP I V WL+ G+ + S A E G D Y S LT+ E +W +				
Sbjct	413	KGFSPADIFVQWLQRGQPLSSDKYV--TSAPMPEPGAPGLYFTHSILTVTEEEWNSGETY			240	
Query	196	TCRVDHRL 204				
		TC V H L				
Sbjct	239	TCVVGHEAL 213				

[Q3]

>Norwegian\_rat protein (taken from BLAST result alignment range 1)  
TKFAKLSCLVTNLATYDTLNISSWSSKSGEPLETNTKIMESHNPNGTFSAVGVASVCMEDWDNRKEFVCTVTHRDLP  
SPQKKFISKPNVAKHPPAVYLLPPAREQLILRESATVTCVLKGFSPADIFVQWLQRGQPLSSDKYVTSAPMPEP  
GAPGLYFTHSILTVTEEEWNSGETYTCVVGHEALPHMVTERTVDKSTGKPTLYNVSLIMSDTGTCY

**Organism:** Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria;  
Euarchontoglires; Glires; Rodentia; Myomorpha; Muroidea; Muridae; Murinae; Rattus.

## [Q4]

The results from blastp are ordered in descending identities; **no 100 % identity**

**Descriptions**

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	Ident	Accession
<input type="checkbox"/>	Ig mu chain C region - rat (fragment)	449	496	100%	1e-157	99.54%	<a href="#">S25644</a>
<input type="checkbox"/>	Igh-6 protein [Rattus norvegicus]	448	496	100%	6e-157	99.54%	<a href="#">AAH86386.1</a>
<input type="checkbox"/>	Igh-6 protein [Rattus norvegicus]	451	500	100%	7e-155	99.54%	<a href="#">AAH92580.1</a>
<input type="checkbox"/>	Igh-6 protein [Rattus norvegicus]	451	499	100%	8e-155	99.54%	<a href="#">AAH97958.1</a>
<input type="checkbox"/>	Igh-6 protein [Rattus norvegicus]	451	500	100%	9e-155	99.54%	<a href="#">AAI05826.1</a>
<input type="checkbox"/>	Igh-6 protein [Rattus norvegicus]	408	457	90%	1e-137	99.49%	<a href="#">AAH92586.1</a>
<input type="checkbox"/>	Ig mu-chain [Mus musculus]	292	292	69%	2e-98	92.05%	<a href="#">AAA37908.1</a>
<input type="checkbox"/>	mCG147612, isoform CRA_a [Mus musculus]	408	408	100%	7e-140	89.86%	<a href="#">EDL18541.1</a>
<input type="checkbox"/>	Ig mu chain C region (allele b) - mouse	408	408	100%	7e-140	89.86%	<a href="#">A24976</a>
<input type="checkbox"/>	RecName: Full=Immunoglobulin heavy constant mu; Contains: RecName: Full=Mu' chain; AltName: Full=55 kDa mu' chain	408	408	100%	8e-140	89.86%	<a href="#">P01872.2</a>
<input type="checkbox"/>	mu-immunoglobulin [Mus musculus]	408	408	100%	9e-140	89.86%	<a href="#">CAA24199.1</a>
<input type="checkbox"/>	immunoglobulin heavy chain [Mus musculus]	372	372	91%	3e-124	88.94%	<a href="#">AAC08349.1</a>
<input type="checkbox"/>	immunoglobulin heavy chain [Mus musculus]	372	372	91%	4e-124	88.94%	<a href="#">AAC08348.1</a>
<input type="checkbox"/>	immunoglobulin heavy chain constant region mu [Mus musculus]	403	403	100%	5e-138	88.94%	<a href="#">CAC20701.1</a>
<input type="checkbox"/>	immunoglobulin mu chain [Mus musculus domesticus]	367	367	90%	3e-123	88.83%	<a href="#">AAB59651.1</a>
<input type="checkbox"/>	mCG147612, isoform CRA_b [Mus musculus]	367	367	90%	3e-123	88.83%	<a href="#">EDL18542.1</a>
<input type="checkbox"/>	immunoglobulin heavy chain mu constant region [synthetic construct]	366	366	90%	5e-123	88.83%	<a href="#">ABP35934.1</a>
<input type="checkbox"/>	unnamed protein product [Mus musculus]	366	366	90%	5e-123	88.83%	<a href="#">BAC40490.1</a>
<input type="checkbox"/>	RecName: Full=Ig mu chain C region	393	442	99%	6e-134	87.50%	<a href="#">P06337.2</a>

## [Q5]

### SEQUENCES FOR ALIGNMENT:

```
>Human_IGHM |sp|P01871.4|Immunoglobulin heavy constant mu [Homo sapiens]
TKSTKLTCLVTDLTITYDSVTISWTRQNGEAVKTHTNISESHPNATFSAVGESICEDDWNNGERFTCTVTHTDLP
SPLKQTI SRPKGVALHRPDVYLLPPAREQLNLRESATITCLVTGFS PADVFVQWMQRGQPLSPEKYVTSAPMPEP
QAPGRYFAHSILTVSEEEWNTGETYTCVVAHEALPNRVTERTVDKSTGKPTLYNVSLVMSDTAGTCY
```

```
>Norwegian_rat protein (taken from BLAST result alignment range 1)
TKFAKLSCLVTNLATYDTLNISSSKSGEPLNTNKIMESHNPNGTFSAVGVASVCMEDWDNRKEFVCTVTHRDLP
SPQKKFISKPNVAKHPPAVYLLPPAREQLILRESATVTCLVKGFSPADIFVQWLQRGQPLSSDKYVTSAPMPEP
GAPGLYFTHSILTVTEEEWNSGETYTCVVGHEALPHMVTERTVDKSTGKPTLYNVSLIMSDTGGTCY
```

```
>House_mouse |AAC08349.1|immunoglobulin heavy chain, partial [Mus musculus]
SKSANLTCLVSNLATYETLNISSWASQSGEPLTKIKIMESHNPNGTFSAGVASVCVEDWNNRKEFVCTVTHRDLP
SPQKKFISKPNVHKHPPAVYLLPPAREQLNLRESATVTCLVKGFSPADISVQWLQRGQPLLPQEKYVTSAPMPEP
GAPGFYFTHSILTVTEEEWNSGETYTCVVGHEALPHLVTERTVDKSTGK
```

```
>Chinese_hamster |EGW12804.1|Ig mu chain C region [Cricetulus griseus]
KSAKLTCLVTNLATYDTLNISSWASQSGEPLNTKLSESHSNGTFSAGQANVCVADWESGKEFVCTVAHRDLPS
PQKKIISKPREVVKQPPAVYLLPPAREQLILREAAATVTCLVKGFSPADIFVQWLQRGQPLSPDKYVTSAPMPEPQ
VPGLYFAHSVLTVTEEEWNSGETYTCVVGHEALPHMVTERTVDKST
```

```

>Mongolian_gerbil |BAL70395.1|immunoglobulin mu heavy chain, partial
[Meriones unguiculatus]
TKSAKLTCCLVTLNLSYDTLNISWSSQSAGEALETNMKLIESHTNGTFSAEVVAHVCAEDWNSRKEFVCTVTHRDLP
SPQKKTISKPNEMNR-
PPAVYVLPSPACEQLILRESVTVTCLVKGFSPADVLVQWLQQRGQPLPSDKYVTSAPMPEPQAQGLYFAYSTLTVAE
EEWNSGETYTCVVVHEALPHTVTERTVTDQSTGKPTLYNVT

>Rabbit |AAA64251.1|Ig mu heavy chain secreted form prepeptide [Oryctolagus
cuniculus]
SKSARLICLVTDLTITYGSLNLSWASHNGKALDTHMNITESHNPATFSAMGEASVCAEDWESGEQFTCTVTHADLP
FPLKHTISKSREVAKHPPAVYVLPAREQLVLRESATVTCLVKGFSPADVFVQWQQRGQPLSSDKYVTSAPAPEP
QAPGLYFTHSTLTVTTEEDWNSGETFTCVVGHEALPHMVTERTVDKSTGKPTLYNVSLIMSDTASTCY

>Horse |AAU09792.1|immunoglobulin mu heavy chain constant chain secreted
form, partial [Equus caballus]
TKSAKLSCQVTNLGTYDSLISWTRQNGEILKHTNISESHPNGTFSALGEATICVEDWESGDDYICTVTHTDLP
FPLKQAI SRPDAVAKHPPSVYVLPPTREQLSLRESASVTCLVKGFSPPDVFVQWLQKGQPLSSDKYVTSAPMPEP
QAPGLYFVHSILTVSEEDWSSGETYTCVVGHEALPHVVTERTVTDKSTGKPTLYNVSLVMSDMASTCY

>Bottlenose_dolphin |AAG40853.1|IgM heavy chain [Tursiops truncatus]
TKSAKLSCCLVTDLTITYDSLISRTRQNGEALQHTNVSESHNGTFSAVGEASVCVEEWESGERFTCTVTHTDLP
SPLKRDISRPEVAKHMP SVYVLPPTREQLSLRESASVTCLVKGFSPADVFVQWLQKGQPMSSDKYVTSAPAPEP
QVPGAYFVYSILTVSEEDWNSGETYTCVVGHEALPHMVTERTVDKSTGKPTLYNVSLVMSDTASTCY

>Giant_panda |AAX73309.1|immunoglobulin mu heavy chain, partial [Ailuropoda
melanoleuca]
TKSAKLSCCLVTDLATYDSLISWTRQNGEPLKHTNISESHPNITFSAMGEATVCVEDWESGEQFTCTVTHTDLP
SPLKKTISRPEVAKHMP SVYVLPSPREQLSLRESASVTCLVKGFSPPDVFVQWLQKGQPMSPDSYVTSAPTPEP
QAPGLYFVHSTLTVTSEEDWSAGETYTCVVGHEALPHMVTERSVDKSTGKPTLYNVSLVLSDTASTCY

```

## ALIGNMENT OBTAINED FROM MUSCLE:

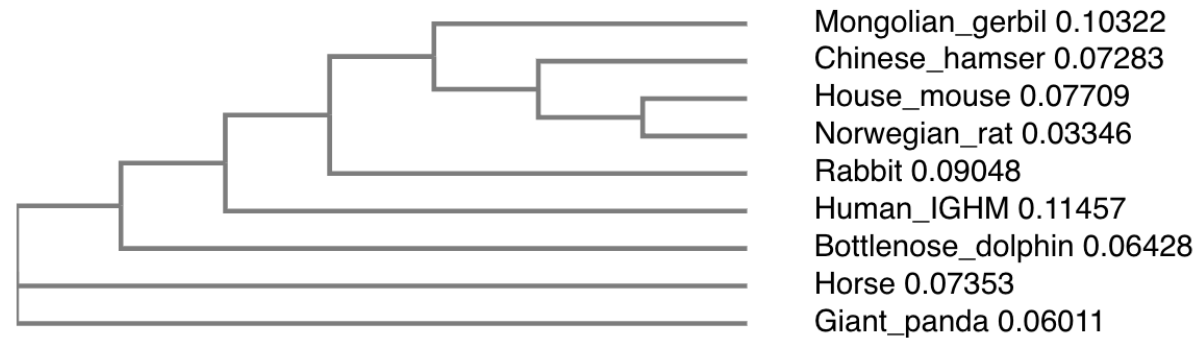
### CLUSTAL multiple sequence alignment by MUSCLE (3.8)

Mongolian_gerbil	TKSAKLTCCLVTLNLSYDTLNISWSSQSAGEALETNMKLIESHTNGTFSAEVVAHVCAEDWN
Chinese_hamster	-KSAKLTCCLVTLNLSYDTLNISWSSQSAGEPLETNTKLSHSHNGTFSAGQANVCVADWE
House_mouse	SKSANLTCCLVSNLATYETLNISWSSQSAGEPLETKIKIMESHNPNGTFSAGVASVCVEDWN
Norwegian_rat	TKFAKLSCCLVTLNLSYDTLNISWSSQSAGEPLETNTKIMESHNPNGTFSAGVASVCMEDWD
Human_IGHM	TKSTKLTCCLVTDLTITYDSVTISWTRQNGEAVKHTNISESHPNATFSAVGEASICEDDWN
Rabbit	SKSARLICLVTDLTITYGSLNLSWASHNGKALDTHMNITESHNPATFSAMGEASVCAEDWE
Horse	TKSAKLSCQVTNLGTYDSLISWTRQNGEILKHTNISESHPNGTFSALGEATICVEDWE
Bottlenose_dolphin	TKSAKLSCCLVTDLTITYDSLISRTRQNGEALQHTNVSESHNGTFSAVGEASVCVEEWE
Giant_panda	TKSAKLSCCLVTDLATYDSLISWTRQNGEPLKHTNISESHPNITFSAMGEATVCVEDWE
	* :.* * :*: * :* :*: :*: :*: :*: :*: :*: :*: :*: :*: :*
Mongolian_gerbil	SRKEFVCTVTHRDLPSPQKKTISKPNEMNR-PPAVYVLPSPACEQLILRESVTVTCLVKGF
Chinese_hamster	SGKEFVCTVAHRDLPSPQKKIISKPREVVQPPAVYLLPPAREQLILREAAVTCLVKGF
House_mouse	NRKEFVCTVTHRDLPSPQKKFISKPNVHKHPPAVYLLPPAREQLNLRESATVTCLVKGF
Norwegian_rat	NRKEFVCTVTHRDLPSPQKKFISKPNVAKHPPAVYLLPPAREQLILRESATVTCLVKGF
Human_IGHM	SGERFTCTVTHTDLPSPKQTISRPKGVALHRPDVYLLPPAREQLNLRESATITCLVTGF
Rabbit	SGEQFTCTVTHADLPFLKHTISKSREVAKHPPAVYVLPAREQLVLRESATVTCLVKGF
Horse	SGDDYICTVTHTDLPFLKQAI SRPDAVAKHPPSVYVLPPTREQLSLRESASVTCLVKGF
Bottlenose_dolphin	SGERFTCTVTHTDLPSPKRDISRPEVAKHMP SVYVLPPTREQLSLRESASVTCLVKGF
Giant_panda	SGEQFTCTVTHTDLPSPKKTISRPEVAKHMP SVYVLPSPREQLSLRESASVTCLVKGF
	. . : :*: :* * * * * : : : * :*: :* :* :*: :*: :*: :* :*

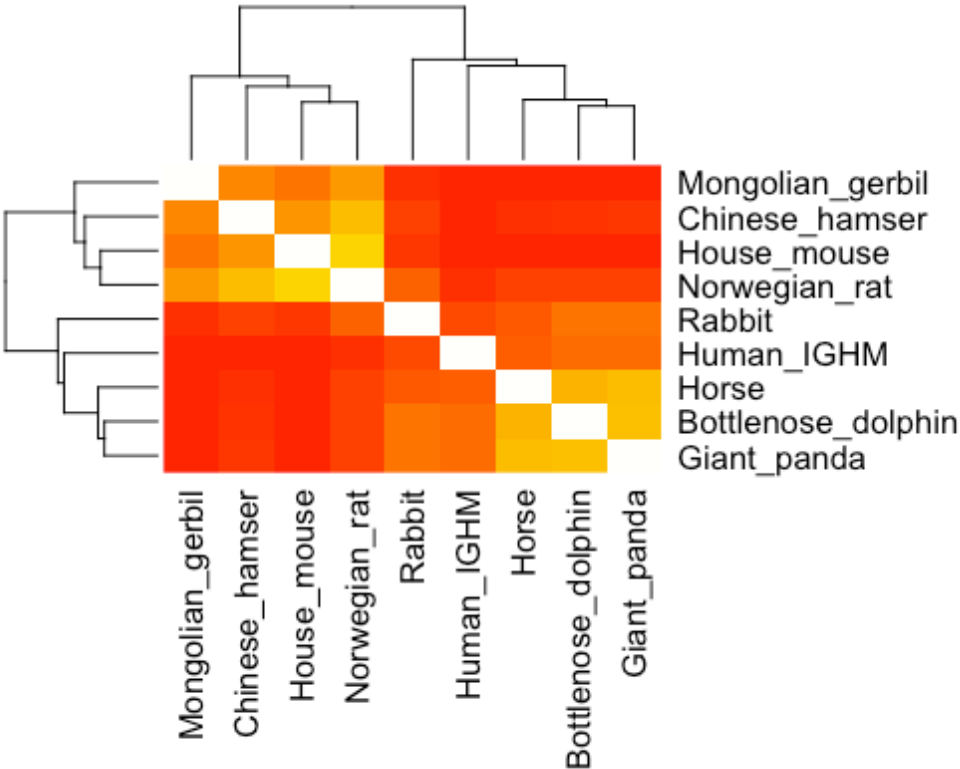
Mongolian_gerbil	SPADVLVQWLQRGQPLSPDKYVTSAPMPEPQAQGLYFAYSTLTVAEEEWNSGETYTCVVV
Chinese_hamser	SPADIFVQWLQRGQPLSPDKYVTSAPMPEPQVPGLYFAHSVLTVTEEEWNSGETYTCVVG
House_mouse	SPADISVQWLQRGQLLPQEKYVTSAPMPEPGAPGFYFTHSILTVTEEEWNSGETYTCVVG
Norwegian_rat	SPADIFVQWLQRGQPLSSDKYVTSAPMPEPGAPGLYFTHSILTVTEEEWNSGETYTCVVG
Human_IGHM	SPADVVFVQWMQRGQPLSPEKYVTSAPMPEPQAPGRYFAHSILTVSEEEWNTGETYTCVVA
Rabbit	SPADVVFVQWQRGQPLSSDKYVTSAPAPEPQAPGLYFTHSTLTVTEEDWNSGETFTCVVG
Horse	SPPDVVFVQWLQKGQPLSSDKYVTSAPMPEPQAPGLYFVHSILTVSEEDWSSGETYTCVVG
Bottlenose_dolphin	SPADVVFVQWLQKGQPMSSDKYVTSAPAPEPQVPGAYFVYSILTVSEEDWNSGETYTCVVG
Giant_panda	SPPDVVFVQWLQKGQPMSPDSYVTSAPTPEPQAPGLYFVHSTLTVSEEDWSAGETYTCVVG
	***.*: *** *.** :. :.***** *** . * **.*: * ***:**:*.:***:****
Mongolian_gerbil	HEALPHTVTERTVDQSTGKPTLYNVT-----
Chinese_hamser	HEALPHMVTERTVDKST-----
House_mouse	HEALPHLVTERTVDKSTGK-----
Norwegian_rat	HEALPHMVTERTVDKSTGKPTLYNVSLIMSDTGGTCY
Human_IGHM	HEALPNRVTERTVDKSTGKPTLYNVSLVMSDTAGTCY
Rabbit	HEALPHMVTERTVDKSTGKPTLYNVSLIMSDTASTCY
Horse	HEALPHVVTERTVDKSTGKPTLYNVSLVMSDMASTCY
Bottlenose_dolphin	HEALPHMVTERTVDKSTGKPTLYNVSLVMSDTASTCY
Giant_panda	HEALPHMVTERTVDKSTGKPTLYNVSLVLSDTASTCY
	*****: *****:**:**

[Q6]

Used “simple phylogeny” from EBI to create cladogram:



[Q7]

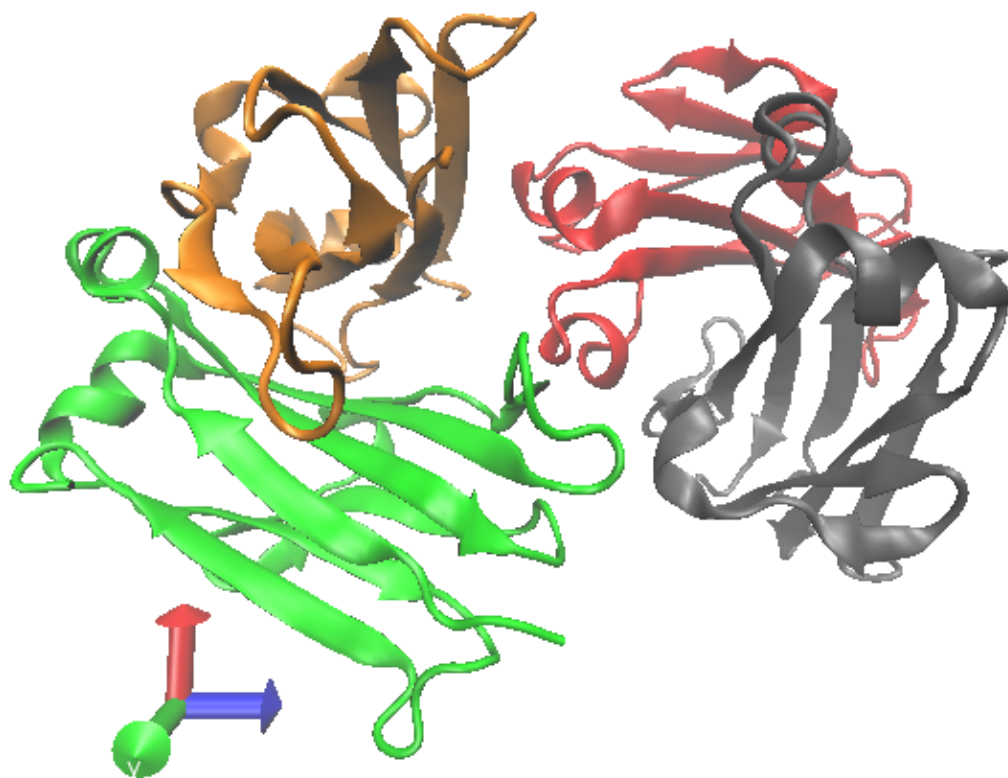


[Q8]

ID	Technique	Resolution	Source	Evalue	Identity
4JYW_A	X-RAY DIFFRACTION	2.0	Mus musculus	2.60e-52	83.2
1R70_B	SOLUTION SCATTERING	30.0	Homo sapiens	3.43e-23	41.1
1IGA_A	SOLUTION SCATTERING	NA	Homo sapiens	6.01e-23	32.2

[Q9]

Since the identity is >80 %, the structure of the Mus musculus IgM protein chain A (4JYW\_A) is highly likely to be similar in structure with my initial novel Norwegian rat protein. In the figure below, IgM chain A is colored in green.



### [Q10]

CHEMBL search gave the following results

(<https://www.ebi.ac.uk/chembl/target/inspect/CHEMBL3885503>):

- 3 binding assays
- 0 functional assays
- The binding assays tested an indazole acid lead compound that was “advanced to first-in-human trials for treatment of diabetic nephropathy”.

#### **Discovery and Preclinical Characterization of 6-Chloro-5-[4-(1-hydroxycyclobutyl)phenyl]-1H-indole-3-carboxylic Acid (PF-06409577), a Direct Activator of Adenosine Monophosphate-activated Protein Kinase (AMPK), for the Potential Treatment of Diabetic Nephropathy**

Kimberly O. Cameron, Daniel W. Kung, Amit S. Kalgutkar, Ravi G. Kurumbail, Russell Miller, Christopher T. Salatto, Jessica Ward, Jane M. Withka, Samit K. Bhattacharya, Markus Boehm, Kris A. Borzilleri, Janice A. Brown, Matthew Calabrese, Nicole L. Caspers, Emily Cokorinos, Edward L. Conn, Matthew S. Dowling, David J. Edmonds, Heather Eng, Dilinie P. Fernando, Richard Frisbie, David Hepworth, James Landro, Yuxia Mao, Francis Rajamohan, Allan R. Reyes, Colin R. Rose, Tim Ryder, Andre Shavnya, Aaron C. Smith, Meihua Tu, Angela C. Wolford, and Jun Xiao

*Journal of Medicinal Chemistry* **2016 59** (17), 8068-8081

DOI: 10.1021/acs.jmedchem.6b00866

(<http://europepmc.org/abstract/MED/27490827>)