# SciFinder<sup>N</sup> Biosequence Search

## BLAST, Nucleotides, Proteins, Commonly Determined Regions and MOTIFS

## **Table of Contents**

I.	Biosequence Search using a FASTA File-Nucleotide	3
II.	Biosequence Search using a FASTA File -Protein	5
III.	Biosequence -Influenza Virus using the following Protein Sequence (H1N1-2009)	11
IV.	Commonly Determined Region Search	21
V.	MOTIF Search	27

## I. Biosequence Search using a FASTA File-Nucleotide

FASTA File (Nucleotide File) for SARS-CoV-2 virus

NC\_000001:Homo sapiens chromosome 1, GRCh38.p13..

Copy the highlighted Genetic Nucelotide String and input it into the BLAST Sequence Upload location

GATTTGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCCATTTGTTCAACTCACAGTTT

Biosequences Settings Prior to Search in this screen

	ences			
ter a prote	in or nucleotide	string, or upload a .txt or .f	asta file. <u>Learn more about</u>	Biosequence Search.
BLAST	CDR	Motif	Upload Sequence	Clear Search
ATTTGGGG	TTCAAAGCAGTAT	CGATCAAATAGTAAATCCATTT	GTTCAACTCACAGTTT	Sequence Type:
				Nucleotide Protein
				Search Within:
				Nucleotides     Proteins
				Include NCBI Sequences
				Limit Total Sequence Results to
				10
				G Start Biosequence Searc
Advanced	Biosequence Se	<u>earch</u> ^ Adjust Parar	meters for Short Sequences	Reset All
<u>Advanced</u> Alignment	l Biosequence Se	earch Adjust Para Match with Gaps?	meters for Short Sequences Gap Costs <b>9</b>	Reset All
Advanced Alignment 80	Biosequence Set	earch_^ Adjust Paran Match with Gaps? Yes No	Gap Costs • Existence 5 Exter	Reset All
Advanced Alignment 80 Query Cor	Biosequence Set t Identity % • verage % •	earch  Adjust Parar Match with Gaps? Yes No Word Size	Gap Costs • Existence 5 Exter Reward for Match	Reset All
Advanced Alignment 80 Query Con 90	Biosequence Set	earch Adjust Paran Match with Gaps? Yes No Word Size •	Gap Costs • Existence 5 Exter Reward for Match Penalty for Mismat	Reset All
Advanced Alignmeni 80 Query Cor 90	Biosequence Set t Identity % • verage % •	Adjust Parar Match with Gaps? Yes No Word Size •	Gap Costs • Existence 5 Externation Reward for Match Penalty for Mismat 2, -3	Reset All
Advanced Alignment 80 Query Con 90 BLAST Alg	Biosequence Se t Identity % • verage % •	earch  Adjust Parar Match with Gaps? Yes No Word Size	Gap Costs • Existence 5 Exter Reward for Match Penalty for Mismat 2, -3 Exclude Low	Reset All

After entering the Nucleotide String, click "Start Biosequence Search" & results are generated.

January 11, 20
----------------

2:28 PM

Biosequences

nces	Sequence Type: Nucleotide Search Within: Nucleotides	GATTTGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCCATTTGTTC AACTCACAGTTT	View Results
	NCBI Included: Yes BLAST Algorithm: BLASTn Alignment Identity: 80%		Edit Search
	Query Coverage: 90%		Complete
	View All Y		

#### View results

BLAST Search Details	Siosequences (1) View:	Expanded
Sequence Type: Nucleotide Search Within: Nucleotides BLAST Algorithm: BLASTn	References	₹
NCBI Included: Yes Alignment Identity: 80% Ouery Coverage: 90%	Query Details GATTTGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCCATTTGTTCAACTCACAGTTT View More	
E-Value: 10 Match with Gaps?: No	1 Alignment Ident	ity: 100%
Gap Costs: Existence 5 Extension 2 Word Size: 11	Query (1) (60)	tches: 60
Bioscape Analysis	Mism	atches: 0
Visually explore sequence similarity with a new tool. Learn more about Bioscape.	Subject (1) 60) View Less ~	
Create Bioscape Analysis	Alignment Subject References	erences
Filter by	Alignment Data BLAST Score: 120 E-Value: 9.334e-23	
0 • to 10 <sup>6</sup> •	Q 1 GATTIGGGGT TCAAAGCAGT ATCGATCAAA TAGTAAATCC ATTIGTICAA CTCACAGTTI 60	
Query Coverage %	S 1 GATTIGGGGT TCAAAGCAGT ATCGAICAAA TAGTAAATCC ATTIGITCAA CTCACAGTIT 60	
0 to 100		
Subject Coverage %		
0 to 100		
Alignment Identity %		
0 to 100		

The results returned compares the inputted Genetic Nucleotide String against those found by the Biosequences Search. In the above case, there is 100% matching of the complete Genetic String (60 matches, 0 mismatches). The graphical representation shows identical corroboration of inputted versus "found" strings.

## II. Biosequence Search using a FASTA File -Protein

FASTA File (Protein Search) - Calcium-modulated protein

Sequence ID: <u>XP\_022517811.1</u>

Copy this File to the BLAST Sequence Upload location. Each letter highlighted letter in this protein string is an amino acid.

MADQLTEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNPTEAELQDMINEVDADGNGTIDFPEFLTMMARKM KDTDSEEEIREAFRVFDKDGNGYISAAELRHVMTNLGEKLTDEEVDEMIREADIDGDGQVNYEEFVQMMTAK

#### Settings Prior to Search

BLAST CDR	Motif	Upload Sequence	Clear Search	
ADQLTEEQIAEFKEAFSLFD	KDGDGTITTKELGTVMRSLGQNPTE	CAELQDMINEVDADGNGTIDF	Sequence Type:	
DTDSEEEIREAFRVFDKDGN	GYISAAELRHVMTNLGEKLTDEEVI	DEMIREADIDGDGQVNYEEFV	QMMTAK Nucleotide Prot	ein
			Search Within:	
			🔵 Nucleotides 🛛 🥥 Prote	ins
			Include NCBI Sequence	5
			Limit Total Sequence Resu	ts t
			10	
			<b>Q</b> Start Biosequence S	ear
Advanced Biosequence	Search Adjust Param	eters for Short Sequences	F   Reset All	
Advanced Biosequence Alignment Identity % @	Search Adjust Param Match with Gaps?	Gap Costs •	s   Reset All	
Advanced Biosequence Alignment Identity % @	Search Adjust Param Match with Gaps? Yes No	eters for Short Sequences Gap Costs <b>e</b> Existence 11 Ext	s   Reset All ension 1 •	
Advanced Biosequence Alignment Identity % @ - Query Coverage % @	Search Adjust Param Match with Gaps? Yes No Word Size O	eters for Short Sequences Gap Costs • Existence 11 Ext Scoring Matrix •	s   Reset All ension 1 -	
Advanced Biosequence Alignment Identity % - Query Coverage % 90	Search Adjust Param Match with Gaps? Yes No Word Size O 3 ~	eters for Short Sequences Gap Costs • Existence 11 Ext Scoring Matrix • BLOSUM62	s   Reset All ension 1 -	
Advanced Biosequence Alignment Identity % - Query Coverage % 90 BLAST Algorithm	Search Adjust Parame Match with Gaps? Yes No Word Size O 3 • E-Value O	Gap Costs Existence 11 Extr Scoring Matrix BLOSUM62 Exclude Low	s   Reset All ension 1 •	

Press "Start Biosequence Search"

Recent Search History



#### Biosequence Search is Completed

Recent Search History			
December 28, 2021			
2:18 PM			
Biosequences	Sequence Type: <b>Protein</b> Search Within: <b>Proteins</b> NCBI Included: <b>Yes</b>	MADQLTEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNFTEAELQDMI NEVDADGNGTIDFPEFLTMMARKMKDTDSEEEIREAFRVFDKDGNGYISAAEL RHVMTNLGEKLTDEEVDEMIREADIDGDGQVNYEEFVQMMTAK	View Results
	BLAST Algorithm: <b>BLASTp</b> Alignment Identity: - Query Coverage: <b>90%</b>		Edit Search
	Results will expire on Jan 28, 2022.		Complete

Choose "View Results". The following 3 screens are all for search results returned.

Ref	erences								
Query [	Details MADQLTEEC	)IAEFKEAFSLFDF	OGDGTITTKE	LGTVMRSLGQNP	TEAELQDMIN	EVDADGNGTIDF	PEFLTMMARKI	MKDT V	/iew L
MADQI LRHVM	JTEEQIAEFKEAFSLFD MINLGEKLIDEEVDEMI	KDGDGTITTKELG READIDGDGQVNY	IVMRSLGQNPI EEFVQMMTAK	'EAELQDMINEVD	ADGNGTIDFF	EFLTMMARKMKI	DTDSEEEIREA	FRVFDKDGN	GYIS
1	>						Align	ment Identi	ity: 10
Query					(149)	072		Mato Mism	ches: ' atche
-	0					(3/3)			
View Le	ment Subject	References				(13)		🗐 Refe	erence
View Le Align Align BLAS E-Val	ment Data T Score: 759 ue: 1.12043e-93	References						Refe	erence
Align Align BLAS E-Val	ment Data T Score: <b>759</b> ue: <b>1.12043e-93</b>	References	DKDGDGTITT	KELGTVMRSL (	5QNPTEAELQ	DMINEVDADG I	NGTIDFPEFL	Refe	erence
Align Align BLAS E-Val Q S	ment Data T Score: 759 ue: 1.12043e-93 1 MADQLTEE 1111111 39 MADQLTEE	References	DKDGDGTITT	KELGTVMRSL (          KELGTVMRSL (	GQNPTEAELQ           GQNPTEAELQ	DMINEVDADG 1	NGTIDFPEFL            NGTIDFPEFL	<b>Refe</b> 70 108	erence
View Lee Align Align BLAS E-Val Q S Q	ment Data T Score: <b>759</b> ue: <b>1.12043e-93</b> 1 MADQLTEE 1111111 39 MADQLTEE 71 TMMARKMK 1111111	References       QI     AEFKEAFSLF       IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	DKDGDGTITT   DKDGDGTITT RVFDKDGNGY 	KELGTVMRSL (           KELGTVMRSL (  SAAELRHVM 2 	GQNPTEAELQ             GQNPTEAELQ TNLGEKLTDE 	DMINEVDADG I IIIIIII DMINEVDADG I EVDEMIREAD : IIIIIIIII	NGTIDFPEFL           NGTIDFPEFL IDGDGQVNYE 	70 108 140	erence
View Lee Align Align BLAS E-Val Q S Q S	ment Data T Score: <b>759</b> ue: <b>1.12043e-93</b> 1 MADQLTEE 1111111 39 MADQLTEE 71 TMMARKMK 1111111 109 TMMARKMK	References         QI       AEFKEAFSLF         IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	DKDGDGTITT IIIIII DKDGDGTITT RVFDKDGNGY IIIIIIII RVFDKDGNGY	KELGTVMRSL (          KELGTVMRSL ( ISAAELRHVM ?          ISAAELRHVM ?	GQNPTEAELQ           GQNPTEAELQ TNLGEKLTDE           TNLGEKLTDE	DMINEVDADG I IIIIIII DMINEVDADG I EVDEMIREAD : IIIIIIII EVDEMIREAD :	NGTIDFPEFL IIIINGTIDFPEFL IDGDGQVNYE IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	<ul> <li>Refe</li> <li>70</li> <li>108</li> <li>140</li> <li>178</li> </ul>	rence
View Lee Align Align BLAS E-Val Q S Q S Q S Q	ment Data T Score: <b>759</b> ue: <b>1.12043e-93</b> 1 MADQLTEE 1111111 39 MADQLTEE 71 TMMARKMK 11111111 109 TMMARKMK 141 EFVQMMTA 11111111	References	DKDGDGTITT IIIIII DKDGDGTITT RVFDKDGNGY IIIIIIII RVFDKDGNGY	KELGTVMRSL (          KELGTVMRSL ( ISAAELRHVM ?           ISAAELRHVM ?	GQNPTEAELQ             GQNPTEAELQ TNLGEKLTDE            TNLGEKLTDE	DMINEVDADG I IIIIII DMINEVDADG I EVDEMIREAD : IIIIIIIII EVDEMIREAD :	NGTIDFPEFL IIIII NGTIDFPEFL IDGDGQVNYE IIIIIIIII IDGDGQVNYE	70 108 140 178	rence

Click on "Subject". In this screen and the one above, 149 amino acids of the query 973 length Genetic String match the returned results. In this case, the Protein string (subject) found is a small portion of the queried string.

Biosequences (10)	Sort: Alignment Identity - View: Expande
References	
Query Details MADQLTEEQIAEFKEAFSLFDKDG	SDGTITTKELGTVMRSLGQNPTEAELQDMINEVDADGNGTIDFPEFLTMMARKMKDT View Mo
	Alignment Identity: 1009
Query 1	(149) Matches: 14 Mismatches: 19 973
Alignmen Subject References	References
Sequence	
1 MHHHHHHSSG LVPRGSGMKE TAAAK 81 NPTEAELQDM INEVDADGNG TIDFP 161 DEMIREADID GDGQVNYEEF VQMMT 241 LGEQIDNLQR VKQKLEKEKS EMKME 321 ESGEYSRQLD EKDSLVSQLS RGKQA 401 KANSEVAQWR TKYETDAIQR TEELE 481 DKKQRNFDKI LAEWKQKYEE THAEL 561 IHELEKIKKQ IEQEKSELQA ALEEA	FERQH MDSPDLGTMA DQLTEEQIAE FKEAFSLFDK DGDGTITTKE LGTVMRSLGQ PEFLTM MARKMKDTDS EEEIREAFRV FDKDGNGYIS AAELRHVMTN LGEKLTDEEV PAKGRP MGGATSAQIE MNKKREAEFQ KMRRDLEEAT LQHEATAATL RKKHADSVAE PIDDLA SNMEVISKSK GNLEKMCRTL EDQVSELKTK EEEQQRLINE LTAQRGRLQT AFTQQI EELKRQLEEE IKAKSALAHA LQSSRHDCDL LREQYEEEQE AKAELQRAMS PEAKKK LAQRLQDAEE HVEAVNAKCA SLEKTKQRLQ NEVEDLMIDV ERTNAACAAL JEASQK ESRSLSTELF KIKNAYEESL DHLETLKREN KNLQQEISDL TEQIAEGGKR AEASLE HEEGKILRIQ LELNQVKSEI DRKIAEKDEE IDQLKRNHIR VVESMQSTLD
641 AEIRSRNDAI RLKKKMEGDL NEMEI 721 AFIFFLDATI, FOTFDSDKIA FOFLI	QLNHS NRMAAEALRN YRNTQGILKD TQLHLDDALR GQEDLKEQLA MVERRANLLQ

Click on "References"

## Biosequences (10)

Sort: Alignment Identity - View: Expanded -

References	⊥
Query Details MADQLTEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNPTEAELQDMINEVDADGNGTIDFPEFLTMMARKMKDT	. View More
1 Alignment	Identity: 100%
Query 1 (149) Subject 1 (149)	Matches: <b>149</b> Mismatches: <b>0</b>
View Less  Alignment Subject References	References
モータータンパク質、人工筋肉、マイクロアクチュエータおよびこれらの製造方法 Assignee: 国立大学法人北陸先端科学技術大学院大学 JP2020050606 A   Seq ID No: 12	

Scroll to the next Biosequence found in the original search.

Results for Search result #2 - Choose "Alignment"

	2								Align	ment Identity: 100%
	Query (	1					(14	(9		Matches: <b>149</b> Mismatches: <b>0</b>
	Subject (	1)=						174	)	
,	View Less 🤊	,								
	Alignmer	nt	Subject	References						References
	Alignmer BLAST Sc E-Value:	nt Dat ore: 7 <b>1.120</b>	a 759 43e-93							
	Q	1	MADQLTEEQI	AEFKEAFSLF	DKDGDGTITT	KELGTVMRSL	GQNPTEAELQ	DMINEVDADG	NGTIDFPEFL	70
	S	15	MADQLTEEQI	AEFKEAFSLF	DKDGDGTITT	KELGTVMRSL	GQNPTEAELQ	DMINEVDADG	NGTIDFPEFL	84
	Q	71	TMMARKMKDT	DSEEEIREAF	RVFDKDGNGY	ISAAELRHVM	TNLGEKLTDE	EVDEMIREAD	IDGDGQVNYE	140
	S	85	TMMARKMKDT	DSEEEIREAF	RVFDKDGNGY	ISAAELRHVM	TNLGEKLTDE	EVDEMIREAD	IDGDGQVNYE	154
	Q	141	EFVQMMTAK :	149						
	S	155	EFVQMMTAK (	163						

#### Choose "Subject"

2			Alignment Identity: 100%
Query 1	)		149
			Matches: 149
			Mismatches: 0
Subject	)		174
View Less 🗸			
Alignment	Subject	References	References
CAS Registry Length: <b>174</b>	/ Number: 11458 aa	807-38-0	
Sequence			
1 1	AGSSHHHHHH HSS	GMADQLT EEQIAE	FKEA FSLFDKDGDG TITTKELGTV MRSLGQNPTE AELQDMINEV DADGNGTIDF
	PEFTTMMARK MKT	TDSEEEI REAFRV	FDKD GNGYISAAEL RHVMTNLGEK LTDEEVDEMI READIDGDGQ VNYEEFVQMM
81 1			

#### Choose "References"

2			Alignr	nent Identity: 100%
Query 1			(149)	
				Matches: <b>149</b> Mismatches: <b>0</b>
Subject			174	
View Less 🗸				
Alignment	Subject	References		References
A fusion tag of Assignees: UN WO20090465	comprising an NIVERSITY HEAI 20 A1   Seq ID	<b>affinity tag and</b> LTH NETWORK; No: -	d an EF-hand motif containing polypeptide and methods of use t MCCLUSKEY, ANDREW, J.; POON, GREGORY, M. K.; GARIÉPY, JEAN	hereof

### III. Biosequence -Influenza Virus using the following Protein Sequence (H1N1-2009)

https://www.ncbi.nlm.nih.gov/nuccore/CY065935

(Influenza A virus (A/Guam/NHRC0001/2009(H1N1)) segment 1, complete sequence

Note - (Copy the sequence below completely...the Sequence Search will disregard blank spaces)

MERIKELRDLMSQSRTREILTKTTVDHMAIIKKYTSGRQEKNPA

LRMKWMMAMRYPITADKRIMDMIPERNEQGQTLWSKTNDAGSDRVMVSPLAVTWWNRN

GPTTSTVHYPKVYKTYFEKVERLKHGTFGPVHFRNQVKIRRRVDTNPGHADLSAKEAQ

DVIMEVVFPNEVGARILTSESQLAITKEKKEELQDCKIAPLMVAYMLERELVRKTRFL

PVAGGTGSVYIEVLHLTQGTCWEQMYTPGGEVRNDDVDQSLIIAARNIVRRAAVSADP

LASLLEMCHSTQIGGVRMVDILRQNPTEEQAVDICKAAIGLRISSSFSFGGFTFKRTS

GSSVKKEEEVLTGNLQTLKIRVHEGYEEFTMVGRRATAILRKATRRLIQLIVSGRDEQ

SIAEAIIVAMVFSQEDCMIKAVRGDLNFVNRANQRLNPMHQLLRHFQKDAKVLFQNWG

IESIDNVMGMIGILPDMTPSTEMSLRGIRISKMGVDEYSSTERVVVSIDRFLRVRDQR

GNVLLSPEEVSETQGTEKLTITYSSSMMWEINGPESVLVNTYQWIIRNWEIVKIQWSQ

DPTMLYNKMEFEPFQSLVPKATRSRYSGFVRTLFQQMRDVLGTFDTVQIIKLLPFAAA

PPEQSRMQFSSLTVNVRGSGLRILVRGNSPVFNYNKATKRLTVLGKDAGALTEDPDEG

TSGVESAVLRGFLILGKEDKRYGPALSINELSNLAKGEKANVLIGQGDVVLVMKRKRD

**SSILTDSQTATKRIRMAIN** 

Select "Protein" as the Sequence Type and "Proteins" as the Search Type.

Searching for	Biosequ	ences					
& All	Enter a protei	in or nucleo	tide string, or	upload a .txt or .fasta f	ile. <u>Learn more about</u>	t Biosequence Se	arch.
☑ Substances	BLAST	CDR	Motif		Upload Sequence	Clear Search	
A Reactions	MERIKELRDI WSKTNDAGSI	LMSQSRTREI DRVMVSPLAV	LTKTTVDHMAII TWWNRNGPTTS1	IKKYTSGRQEKNPALRMKW IVHYPKVYKTYFEKVERLK	MMAMRYPITADKRIMDM KHGTFGPVHFRNQVKIRR	II PERNEQGQTL RVDTNPGHADL	Sequence Type: Nucleotide Protein
References	SAKEAQDVIN LHLTQGTCWP DICKAAIGLI	MEVVFPNEVG EQMYTPGGEV RISSSFSFGG	ARILTSESQLAI RNDDVDQSLII FTFKRTSGSSVI	ITKEKKEELQDCKIAPLMV AARNIVRRAAVSADPLASI KKEEEVLTGNLQTLKIRVH	/AYMLERELVRKTRFLPV LEMCHSTQIGGVRMVDI HEGYEEFTMVGRRATAIL	AGGTGSVYIEV LRQNPTEEQAV RKATRRLIQLI	Search Within:
📜 Suppliers	VSGRDEQSIA ILPDMTPSTH NGPESVLVN'	AEAIIVAMVF EMSLRGIRIS FYQWIIRNWE	SQEDCMIKAVRO KMGVDEYSSTER IVKIQWSDPTMI	GDLNFVNRANQRLNPMHQI RVVVSIDRFLRVRDQRGNV LYNKMEFEPFQSLVPKATF	LRHFQKDAKVLFQNWGI /LLSPEEVSETQGTEKLT RSRYSGFVRTLFQQMRDV	ESIDNVMGMIG ITYSSSMMWEI LGTFDTVQIIK	<ul> <li>Nucleotides ● Proteins</li> <li>✓ Include NCBI Sequences</li> </ul>
A Biosequences	LLPFAAAPPE LILGKEDKR)	EQSRMQFSSL YGPALSINEL	TVNVRGSGLRII SNLAKGEKANVI	LVRGNSPVFNYNKATKRLI LIGQGDVVLVMKRKRDSSI	IVLGKDAGALTEDPDEGT ILTDSQTATKRIRMAIN	SGVESAVLRGF	Limit Total Sequence Results to:
Retrosynthesis							1000     Q     Start Biosequence Search
	Advanced	Biosequend	te Search 🗸				

#### Start the Biosequence search



+

The Biosequence Search is completed. Click on the "View results" to view the results.

January 10, 2022			
2:42 PM			
♦ Biosequences	Sequence Type: <b>Protein</b> Search Within: <b>Proteins</b> NCBLIncluded: <b>Yes</b>	MERIKELRDLMSQSRTREILTKTTVDHMAIIKKYTSGRQEKNPALRMKWMMAM RYPITADKRIMDMIPERNEQGQTLWSKTNDAGSDRVMVSPLAVTWWNRNGPTT STVHYPKVYKTYFEKVERLKHGTFGPVHFRNOVKIRRRVDTNPGHADLSAKEA	View Results
	BLAST Algorithm: BLASTp	QDVIMEVVFPNEVGARILTSESQLAITKEKKEELQDCKIAPIMVAYMLERELV	Edit Search
	Alignment Identity: - Query Coverage: 90% View All ~	RKTRFLFVAGGTGSVYIEVLHLTQGTCWEQWYTFGGEVRNDDVDQSLIIAAR	Complete

Results returned from Search - zero Mismatches

BLAST Search Details	Sort: Alignment Identity - View:	Expanded
Sequence Type: Protein Search Within: Proteins BLAST Algorithm: BLASTp	References	4
NCBI Included: Yes Alignment Identity: - Query Coverage: 90%	Query Details merikelrölmsösrtreiltktivdemaiikkytsörörennalrengömmanrypitadkriminipernegögtlasktindags. V	/iew More
Match with Gaps?: No Gap Costs: Existence 11 Extension 1 Word Size: 3	Query 1 (759)	ty: 100%
Bioscape Analysis		atches: u
Visually explore sequence similarity with a new tool. Learn more about Bioscape.	Subject (1) (759)	
Create Bioscape Analysis	Alignment Subject References	rences
Filter by	Alignment Data BLAST Score: 3857 E-Value: 0	
0 - to	Q 1 MERIKELADI MSQSRTREII TKTYUDHMAI IKKYTSGRQE KNPALRMKHM MAMRYPITAD KRIMIMIPER 70	^
Query Coverage %	Q 71 NEQQQTLMSK INDAGSDRUM VSPLAVINAM ENGETISTVA YEKVIKTYFE KVERLKAGTF GFVHFRAQVK 140	
	S 71 NEQGQTLNSK TNDAGSDRVM VSPLAVTWAR RNGPTISTVH YPKVYKIYPE KVERLKHGTF GFVHFENQVK 140 0 141 IRRRVDTNPG HADLSAKEAO DVIMEVVFPN EVGARILISE SOLAITKEKK EELODCKIAF LMVAYMLERE 210	
100	S 141 IRRRUDING HADISAKEAQ DVIMEVVFPN EVGARILTSE SQLAITKEKK EELQDCKIAP LMVAYMLERE 210	~

View the 2<sup>nd</sup> Search returned with 1 Mismatch

There is one amino acid (red line in the graphical representation) different between the sequence search and the #2 Query Result returned. There is one mismatch.

The graphic shows the location where the mismatch occurs.

The Alignment Data shows the exact point in the sequence where the mismatch occurs between the two amino acids.

)uery (1)			_			(	759		
									Matches: Mismatche
bject 1						(	759		
w Less 🐱									-
Alignment	Subject	Reference	s						Reference
Alignment Alignment D 3LAST Score 5-Value: 0 2 35	Subject ata : 3856 1 TLKIRVHEGY	Reference	TAILRKATRR	LIQLIVSGRD	EQSIAEAIIV	AMVFSQEDCM	IKAVRGDLNF	420	Reference
Alignment Alignment D BLAST Score S-Value: 0 Q 35 3 35	Subject ata : 3856 1 TLKIRVHEGY 1 ILKIRVHEGY 1 TLKIRVHEGY	Reference EEFTMVGRRA	TAILRKATRR           TAILRKATRR	LIQLIV3GRD            LIQLIV3GRD	EQSIAEAIIV          EQSIAEAIIV	AMVFSQEDCM	IKAVRGDLNF           IKAVRGDLNF	420	Reference
Alignment Alignment D BLAST Score S-Value: 0 Q 35 3 35 Q 42	Subject ata : 3856 1 TLKIRVHEGY !	Reference EEFTMVGRRA IIIIIII EEFTMVGRRA MHQLLRHPQK	TAILRKATRR           TAILRKATRR DAKVLFQNWG	LIQLIVSGRD             LIQLIVSGRD IESIDNVMGM	EQSIAEAIIV          EQSIAEAIIV IGILPDMTPS 	AMVFSQEDCM AMVFSQEDCM TEMSIKGIRI	IKAVRGDLNF          KAVRGDLNF SPOLVDEYSS 	420 420 490	Reference
Alignment Alignment D BLAST Score E-Value: 0 Q 35 S 35 Q 42 S 42	Subject ata : 3856 1 TLKIRVHEGY 1 TLKIRVHEGY 1 VNRANQRLNP 1 VNRANQRLNP	Reference EEFTMVGRRA HULLRHFQK HULLRHFQK	TAILRKATRR IIIIIIIII TAILRKATRR DAKVLFQNWG DAKVLFQNWG	LIQLIVSGRD            LIQLIVSGRD  ESIDNVMGM            IESIDNVMGM	EQSIAEAIIV IIIIIII EQSIAEAIIV IGILPDMTPS IIIIIIII IGILPDMTPS	AMVFSQEDCM IIIIIII AMVFSQEDCM TEMSTAGIRI IIIIIIII TEMSTRGIRV	IKAVRGDLNP IIIIIII KAVRGDLNP SPALVDEVSS IIIIIIII SPALVDEVSS	420 420 490	Reference
Alignment Alignment D BLAST Score E-Value: 0 Q 35 S 35 Q 42 S 42 Q 49	Subject ata : 3856 1 TLKIRVHEGY 1 TLKIRVHEGY 1 VNRANQRLNP 1 UNRANQRLNP 1 TERVVVSIDR	Reference EEFTMVGRRA IIIIIIII EEFTMVGRRA MHQLLRHFQK IIIIIIII MHQLLRHFQK FLRVRDQRGM	TAILRKATRR IIIIIIII TAILRKATRR DAKVLFQNWG IIIIIIII DAKVLFQNWG VLLSPEEVSE	LIQLIV3GRD           LIQLIV3GRD  ESIDNVMGM           IESIDNVMGM TQGTEKLIIT	EQSIAEAIIV EQSIAEAIIV IGILPDMTPS IIIIIIII IGILPDMTPS YSSSMAMMEIN	AMVFSQEDCM AMVFSQEDCM TEMSTAGIRI IIIIIIII TEMSIRGIRU GPESVLVNI	IKAVRGDLNP IIIIIII KAVRGDLNP SPOM VDEYSS IIIIIIII SPOM VDEYSS SWIIRNWEIV	420 420 490 490 560	Reference

The "Subject" Tab shows cross references to other similar proteins

Substance Detail	
Reference (1)     A Reactions (0)     Suppliers (0)	上 📓 💄 Save
CAS Registry Number 1639314-78-5	
Image Not Available	
Unspecified RNA-dependent RNA polymerase (Influenza A virus strain A/California/07/09 gene PB2) (ACI) Protein/Peptide Sequence Sequence Length: 759 Related Sequences (20)	
	Expand All   Collapse A
<ul> <li>Other Names and Identifiers</li> </ul>	
<ul> <li>Sequence Details</li> </ul>	
<ul> <li>Additional Details</li> </ul>	

#### Choosing the "Reference" Tab brings up the documentation where the sequence is found

					4	Alignment Identity: 9	9.87%
Query 1					759	Matche	s: 758
Subject 1					759		
Alignment	Subject	References				Reference	nces
<b>流感病毒重配</b> Assignees: 港 CN10572297(	华股份有限公司 6 A   Seq ID No	司; 合成基因组疫苗) o: -	股份有限公司				^
Influenza vire	us reassortme	ent					
Assignees: SE US201703262	QIRUS UK LIM 227 A1   Seq II	IITED; SYNTHETIC D No: 17	GENOMICS, INC.				
17-44	トインフルエン	/ザAウイルス					
119 72							
Assignees: /	バルティス アー D A   Seq ID No	ーゲー: シンセティ o: 17	ック ジェノミクス ヴ	アクシンズ,インコ	ーポレーテッド		

Repeat the same search for the Nucleotides you have previously, but this time remove part of the sequence

Searching for	Bioseque	nces			
& All	Enter a protein	or nucleotide string, or upload a	.txt or .fasta file. <u>Learn more abou</u>	<u>t Biosequence Search.</u>	
O Substances	BLAST	CDR Motif	Upload Sequence	Clear Search	
A Reactions	GATTTGGGGTTC	AAAGCAGTATCGATCAAATA <mark>GTAAATC</mark>	CATTTGTTCAACTCACAGTTT	Sequence Type Nucleotide	e: Protein
References				Search Within:	
🚆 Suppliers				<ul> <li>Nucleotides</li> <li>Include NCE</li> </ul>	s OProteins BI Sequences
A Biosequences				Limit Total Sec	uence Results to:
Retrosynthesis				1000	•
				Q Start Bio	osequence Search
	Advanced Bi	osequence Search 🔺 🛛 Adju	ust Parameters for Short Sequence	rs   Reset All	
	Alignment Id	lentity % 🛛 Match with Ga	aps? Gap Costs 🧕		
	80	🔾 Yes 🔍 N	o Existence 5 Exte	ension 2 🔻	

#### Now Rerun the search

Searching for	Bioseque	ences					
& All	Enter a protei	n or nucleotid	e string, or upload a	.txt or .fasta fi	le. <u>Learn more abou</u> t	<u>t Biosequence Se</u>	earch.
G Substances	BLAST	CDR	Motif		Upload Sequence	Clear Search	
A Reactions	GATTTGGGGT	TCAAAGCAGTAT	CGATCAAATA				Sequence Type:           Nucleotide         Protein
References     Suppliers							Search Within:
* Biosequences							Limit Total Sequence Results to:
Retrosynthesis							<b>Q</b> Start Biosequence Search
	Advanced	Biosequence	Search 🔺 🛛 Adji	ust Parameters	for Short Sequence	s   Reset All	
	Alignment	Identity % 🧕	Match with Ga	aps? Io	Gap Costs 🛛	nsion 2 -	
	80			-	Existence 5 Exte	nsion 2 🔻	

The Updated search is found. Now the number of mismatches is much more pronounced and is shown in the highlighted red vertical lines shown in the graphics for the 2<sup>nd</sup> search returned.

BLAST Search Details	S Biosequences (11)	Sort: Alignment Identity 👻 View: Expanded 🕶
Sequence Type: Nucleotide Search Within: Nucleotides BLAST Algorithm: BLASTn	References	Ŧ
NCBI Included: Yes Alignment Identity: 80%	Query Details GATTTGGGGTTCAAAGCAGTATCGATCAAATA View More	
E-Value: <b>10</b> Match with Gaps?: <b>No</b>	1	Alignment Identity: 100%
Gap Costs: Existence 5 Extension 2 Word Size: 11	Query 1	32) Matches: <b>32</b>
Bioscape Analysis		Mismatches: 0
Visually explore sequence similarity with a new tool. Learn more about Bioscape.	Subject (1)	(60)
Create Bioscape Analysis	Alignment Subject References	References
Filter by A E-Value	Alignment Data BLAST Score: 64 E-Value: 5.68136e-8	
0 • to 10 <sup>6</sup> •	Q 1 GATTTGGGGT TCAAAGCAGT ATCGATCAAA TA 32	
Query Coverage %	S 1 GATTTGGGGT TCAAAGCAGT ATCGATCAAA TA 32	
2		Alignment Identity: 87 1%
2		Alignment identity. 07.170
Query 1	32	
		Matches: 27 Mismatches: 4
Subject		1,368
View Less 🗸		
Alignment Subject	References	References
Alignment Data BLAST Score: <b>42</b> E-Value: <b>0.0384269</b>		
Q 2 ATTTGGGGT	T CAAAGCAGTA TCGATCAAAT A 32	
S 422 ATTTGGGGT	C TGAAGCAGTG TCGATCAAAT A 452	

The #3 search returned show more mismatched (red vertical lines) nucleotides than does search #2.



Now Plot the Analysis of the Sequences by choosing "Create Bioscape Analysis"

The database searches for content



Results returned and plotted

Scroll with your mouse to Zoom in or Out of the Plotted variable screen

Sequence	e Similarity		
Less Simila	r		More Similar
	<b>C</b> • •		



#### Click on one of the Silos to see the results returned – 3 Patents are returned

**.** 





#### Zoomed in View of the Patents returned

321092-3	57-9	X
Delevent Patente	Converse Langth	
nelevani. Faterits	ooz	
3	957	
Domain Sequence		
catacettta	caatgageteegtgttgeteeegaggageaeeeeg	tgct
tttgaccgaag	geteetttgaaccccaagtccaaccgtgaaaagatg	accc
agateatgtte	.gagacetteaacgeeceageettetatgtetetat	tcag
getgtettgte	actttacgtttccggtcgtaccactggtattgttc	ttga
ttetggegatg	gtgteteteacacegtecetatetaegagggtae	tete
tecccaegee	atcotgogtottgatatggcoggcogtgacttgac	tgac
tacctcatgog	cateettgeegagegegeeactegtteaceacea	ctgc
tgagogtgaaa	attgtccgcgacatcaaggagaagctctgctacgtc	geee
tggattttgag	jeaggagatgeagaetgeegeaeagteeteegeeet	cgag
aagtootatga	agetteeegaeggteaggteateacegteggtaatg	ageg
gtgagtataat	.gttetettaegatttettetteagetttegetaaa	acga
tgegttattet	agttteegeacteeegaggetetetteeaaceete	gete
ttgggtettga	ateteetggtateeatgagaceacetaeaaeteta	tcat
gaagtgtgatg	yttgatatoogcaaggaottgtactcaaacattgto	atgg
taattatcaac		cgca
teggatttgte	etgtagtetggtggtaccaccatgtaccetggtatt	getg
accgtatgcag	gaaggaaatcactgotttggcccctagetccatgaa	gate
aagattottoo	etecteccgagegaagtactetgtetggateggtg	gtte
	conttionanottionanatataa	

#### Click on the Patent Number(s) found (in this case #3)

#### **References** (1)

View: Partial Abstract 👻

□ 😡 Substances 🗕 🖪 Reaction	ons - 66 Citing -	L Save
1		
Use of short activating RNA and therapeutic applications By: Saetrom, Paal; Stovner, Endre Bak World Intellectual Property Organizati	for modulating target gene express s ken ion, WO2016170348 A2 2016-10-27   Language	ion for diagnostic, prognostic
FIG. 1 interview interview intervie	The invention relates to oligonucleotides, e.g of a target gene and therapeutic composition of using the oligonucleotides and the therap	g., saRNAs useful in upregulating the expression ns comprising such oligonucleotides. Methods eutic compositions are also provided.
PatentPak •     Full Text •	Substances (0)	ons (0) 66 Citing (3) Ø Citation Map

#### Choose PatentPak to View the Patents.

									Vi	ew: Partial	Abstract
	5 <b>-</b>		•	Ge Citing -					4		🔔 Save
1											
Use of short ac and therapeut By: Saetrom, Paal; S World Intellectual P	<b>tivating</b> ic applica stovner, End roperty Org	RNA for ations Ire Bakken ganization,	wo201	ulating targ	et gene	e expres	sion for e: English,	<b>diagno</b> Database	stic, pr	ognostic	
FIG. 1	1	т	he inver	ntion relates to	oligonuc	lootidos o	σ saRNΔ	s useful in	upregula	ating the exp	pression
The second secon	new Mitterweite og Kil 70 100 100 100 100 100 100 100	Contraction of Contra	of a targe of using t	et gene and the the oligonucleo	erapeutic otides and	composition for the composition of the composition	peutic com	ising such npositions	a oligonuc are also	cleotides. M provided.	ethods
et en	The second secon	Company	f a targe f using t	© Substance	erapeutic otides and	composition I the thera	tions (0)	ising such apositions	ing (3)	cleotides. Mo provided.	ethods on Map
Patent	Full Text	• maximum o	e Paten	Substance	erapeutic otides and	compositi I the thera	tions (0)	ising such positions	ing (3)	cleotides. Mo provided.	ethods on Map
Patent MU2016170348	Full Text Language English	<ul> <li>Company of Company o</li></ul>	e Paten PDF	Substance	ces (0)	Composities, econoposities the thera	tions (0)	ising such apositions	ing (3)	cleotides. Mi provided.	ethods on Map
Patent W02016170348 W02016170348	Full Text Language English English	<ul> <li>Kind Code</li> <li>A2</li> <li>A3</li> </ul>	e Paten PDF PDF	Substance	ces (0)	Lithe thera	tions (0)	ising such	ing (3)	Cleotides. Mo provided.	on Map
Patent W02016170348 JP2018512876	Full Text Language English English Japanese	<ul> <li>Kind Code</li> <li>A2</li> <li>A3</li> <li>T</li> </ul>	e Paten PDF PDF PDF	Substance	ces (0)	Ithe thera	tions (0)	ising such	ing (3)	cleotides. Mo provided.	ethods on Map
PatentPak • Patent W02016170348 JP2018512876 US20180305689	Full Text Language English English Japanese English	<ul> <li>Kind Code</li> <li>A2</li> <li>A3</li> <li>T</li> <li>A1</li> </ul>	e Paten PDF PDF PDF PDF	Substance	ces (0)	A Reac	tions (0)	ising such	ing (3)	Cleotides. Mu provided.	ethods on Map
PatentPak - Patent W02016170348 JP2018512876 US20180305689 JP2021035360	Full Text Language English Japanese English Japanese	<ul> <li>Kind Code</li> <li>A2</li> <li>A3</li> <li>T</li> <li>A1</li> <li>A</li> </ul>	e Paten PDF PDF PDF PDF PDF PDF	Substance	ces (0)	A Reac	tions (0)	ising such	ing (3)	Cleotides. Mu provided.	on Map

Open any PatentPak PDF and Save it to your PC. Open the Saved PDF. Now Search with your original query.

GATTTGGGGTTCAAAGCAGTATCGATCAAATA

Choose View Full Text for the PDF document via EspaceNet

Manadawia Manadawi Manadawi Mila sayaka Mila sayaka Mila sayaka Mila sayaka Mila sayaka	acenet WO2016170348A2	×	G
ly Espacenet F	elp Classification search Results 🕕 Advanced search 🌑 Filters 🕦 Popup tips		
e > Results > V	VO2016170348A2		
WO2016170348			
Bibliographic data	Description Claims Drawings Original document Citations Legal events Patent family		
Register 7 0	Global Dossier 7		
Applicante			
Inventors			
Classifications			
CPC	A61P1/00 (EP); A61P11/00 (EP); A61P13/00 (EP); A61P17/00 (EP); A61P25/00 (EP); A61P25/02 (EP); A61P27/02 (EP); A61P3/00 (EP);		
Priorities	US201562150892P-2015-04-22; US201562150893P-2015-04-22; US201562150895P-2015-04-22; US201562150897P-2015-04-22; US201562150900P-2015-04-22; US20156200P-2015-04-22; US20156200P-2015-04-22; US2015000P-2015-04-22; US20156200P-2015-04-22; US20156200P-2015000P-20156200P-2015000P-2015000P-201500P-2015000P-201500P-201500P-201500P-201500P-201500P-201500P-201500P-201500P-201500P-201500P-201500P-201500P-201500P-20150	:2;	
Application	GB2016051116W-2016-04-21		
Publication	WO2016170348A2 2016-10-27		
Published as	EP3286318A2; JP2018512876A; JP2021035360A; US2018305689A1; US2021363525A1; WO2016170348A2; WO2016170348A3; WO2016170348A8		
EN ED			
SAKINA COMPOS	STIONS AND METHODS OF USE		
Abstract			

## **IV.** Commonly Determined Region Search

Commonly Determined Regions (CDR's) are used to find the complimentary regions found in antibodies (also monoclonal antibodies – aka mAb's). CDR's, Antibodies and mAB's bind to antibodies antigens. An antigens is a substance introduced within the host body such as a chemical, pollen, virus or bacteria. Monoclonal antibodies are made synthetically by pharmaceutical and biotech manufacturers and antibodies are found and formed naturally within organism's bodies.

Monoclonal antibodies help fight off infection the body cannot compensate and resolve itself such as during Chemotherapy and Radiation treatments.

Any of the following example CDR's can be inputted into the SciFinderN tool within the CDR field and will find antibodies based on the antigen search needed.

WASTRES DASNRAT PNPLKAM LVSNLES SYTFH GYYMH

Below are the examples with can be inputted into the CDR tabs.

For this Examples, input LVSNLES into the CDR input. It does not matter if the input is capitalized.

Searching for	Biosec	uences				
& All	Enter a pr	otein string, or uplo	ad a .txt or .fasta file	e. Learn more about Biosequence	Search.	
Substances	BLAST	CDR	Motif	Upload Sequence	Clear Search	
A Reactions		LVSNLES		>		Limit Total Sequence Results to:
References	CDR2	Enter CDR2				1000 -
■ Suppliers	CDR3	Enter CDR3				Q Start Biosequence Search
4 <sup>b</sup> Biosequences						
Retrosynthesis						

#### The system starts its search.

March 21, 2022			
Biosequences 9:29 AM	NCBI Included: <b>No</b> Results will expire on Apr 20, 2022.	CDR1: LVSNLES CDR2: - CDR3: -	View Results Edit Options Searching

View the results – The CDR found a perfect match against the database of possibilities. Over 1000 CDR matches were found.

CDR Segments	Siosequences (1,000)	Sort: Alignment Identity 🝷 View: Expanded 🗸
Select a segment below to view individual or intersecting CDR resctor	References	7
CDR1 CDR2 1,000 0 0	Query Details View Less	
000	LVSNLES > CDR2	
0 CDR3	1	Alignment Identity: 100%
Apply Reset Segments	1 CDR1 7	,
Bioscape Analysis Visually explore sequence	Subject 1	Matches: 7 Mismatches: 0
Learn more about Bioscape.	View Less V	B Deferencer
Filter by	Alignment Subject References Alignment Data BLAST Score: 49	Nerer ences
<ul> <li>►-Value</li> <li>0 • to 10<sup>6</sup> •</li> </ul>	E-Value: 71 5005 CDR1 1 LVSNLES 7	
Query Coverage %	S 54 LVSNLES 60	

Let's rerun the search with 2 CDR's inputted. Leave LVSNLES in the search and add WASTRES as the  $2^{nd}$  CDR search. Start the Biosequence.

Searching for	Bioseq	uences					
& All	Enter a pro	ein string, or u	pload a .txt or .	fasta file. <u>Learn mor</u>	e about Biosequence	Search.	
Substances	BLAST	CDR	Motif		Upload Sequence	Clear Search	
		_					Include NCBI Sequences
Z Reactions	CDR1	LVSNLES				×	Limit Total Sequence Results to:
References	CDR2	WASTRES		ノ		×	1000 -
📕 Suppliers	CDR3	Enter CDR3					<b>Q</b> Start Biosequence Search
Biosequences							
Retrosynthesis							

System Search Begins

March 21, 2022			
<b>% Biosequences</b> 9:41 AM	NCBI Included: No	CDR1: LVSNLES CDR2: WASTRES	View Results
	Apr 20, 2022.	Luns: -	Edit Options
			Searching.

View the results. Now there is a cross reference with multiple CDR regions within the database. Only the WASTRES CDR is found in any of the searches run. Both CDR's are not found within the same sequences returned.

CDR Segments	♦ Biosequences (1,000)	Sort: Alignment Identity 👻 View: Expanded 🕶
Select a segment below to view includual or intersecting SDR results.	References	⊻ ₹
CDR1 CDR2	Query Details View Less  CDR1 LVSNLES  CDR2	-
CDR3 Apply Reset Segments	WASTRES > CDR3	
Bioscape Analysis	1	Alignment Identity: 50%
Visually explore sequence similarity with a new tool. Learn more about Bioscape. Create Bioscape Analysis	Subject 1	Matches: 7 Mismatches: 7
Filter by	Alignment Subject Defenses	References
∧ E-Value	Aignment Subject References	
0 • to 10 <sup>6</sup> •	Alignment Data BLAST Score: 27	
Query Coverage %     0     to 100	E-Value:3.387 CDR2 1 WASTRES 7         S 56 WASTRES 62	
Subject Coverage %		
<b></b> O	2	Alignment Identity: 50%
0 to 100	CDR2	
<ul> <li>Alignment Identity %</li> </ul>		Matches: 7
00	Subject 1	113 Mismatches: 7
0 to 100	View Less 🗸	

Try a different CDR to search with LVSNLES. In this case use PNPLKAM.

Recent Search History		View All Search History
March 21, 2022		
<b>% Biosequences</b> 10:07 AM	NCBI Included: No CDR1: LVSNLES CDR2: PNPLKAM Results will expire on CDR3:	
	Apr 20, 2022.	Edit Options
		Searching.

#### There were no searches return with both CDR's in the returned information

#### Rerun the search but use NCBI sequences

Searching for	Biosequ	iences				
& All	Enter a prote	ein string, or u	upload a .txt o	or .fasta file. <u>Learn more about Biosequence</u>	Search.	
© Substances	BLAST	CDR	Motif	Upload Sequence	Clear Search	
A Reactions						Include NCBI Sequences
References	CDR1	LVSNLES			×	Limit Total Sequence Results to:
	CDR2	PNPLKAM			×	1000 -
E Suppliers	CDR3	Enter CDR3.				<b>Q</b> Start Biosequence Search
√ <sup>b</sup> Biosequences						
Retrosynthesis						

Both CDR's were found, but no sequences where both CDR's are found in the same sequence.

CDR Segments	Sort A	lignment Identity 👻 View: Expanded 🕶
Select a segment below to view individual or intersecting CDR results.	References	T
CDR1 CDR2 957 0 43	Query Details View Less	-
	> CDR1 LVSNLES > CDR2 PNPLKAM	
Apply	> CDR3	
	1	Alignment Identity: 50%
Bioscape Analysis Visually explore sequence similarity with a new tool. Learn more about Bioscape. Create Bioscape Analysis	Subject 1	Matches: 7 Mismatches: 7
Filter by	Alignment Subject	References
	Alignment Data BLAST Score: 27.5 E-Value: 3.14668 CDR2 1 PNPLKAM 7 IIIIIII S 99 PNPLKAM 105	
Subject Coverage %	2	Alignment Identity: 50%
0 to 100 Alignment Identity %	1 CDR1 7 Subject 1 112	Matches: 7

Try another CDR to compare against WATRES. Use GYYMH as the 2<sup>nd</sup> CDR input. Check the "Include NCBI Sequences" box.

Searching for	Biosequ	iences				
& All	Enter a prote	ein string, or up	pload a .txt or	.fasta file. <u>Learn more about Bioseque</u>	ence Search.	
Substances	BLAST	CDR	Motif	Upload Sequen	ce Clear Search	
						🗹 Include NCBI Sequences
A Reactions	CDR1	WASTRES			×	Limit Total Sequence Results to:
References	CDR2	GYYMH			×	1000 -
🐂 Suppliers	CDR3	Enter CDR3				<b>Q</b> Start Biosequence Search
√ <sup>b</sup> Biosequences						
Retrosynthesis						

Both CDR's are found within the search and data displayed as found.

CDR Segments	Siosequences (1,000)	Sort: Alignment Identity + View: Expanded +
Select a segment below to view individual or intersecting CDR results.	References	1
CDR1 CDR2	Query Details View Less	-
0 CDR3	WASTRES > CDR2 GYYMH	
	> CDR3	Aligoment Identity: 100%
Bioscape Analysis		Anghment identity. 100%
Visually explore sequence similarity with a new tool. Learn more about Bioscape. Create Bioscape Analysis	Subject 1	CDR2 1 5 Matches: 12 Mismatches: 0
	View Less 🗸	
Filter by           Filter by           • E-Value	Alignment Subject References	References
0 • to 10 <sup>6</sup> •	Alignment Data BLAST Score: 50 E-Value: 48.8762	
	CDR1 1 WASTRES 7	
0 to 100	 5 56 WASTRES 62	
	CDR2 1 GYYMH 5	
Subject Coverage %	 S 415 GYYMH 419	

## V. MOTIF Search

MOTIF is a protein genetic pattern that has chemical or biological activity/function. MOTIFS are typically 10 to 30 characters long but can be much larger. Zinc finger proteins are very common MOTIFS and can 1000 to 2000 characters in length. This MOTIF is a protein and is 1133 amino acids long.

Here is an example of a Homo Sapien Ring Finger to try.

**MPVQAPQWTDFLSCPICTQTFDETIRKPISLGCGHTVCKMCLNK** LHRKACPFDQTTINTDIELLPVNSALLQLVGAQVPEQQPITLCSGVEDTKHYEEAKKC VEELALYLKPLSSARGVGLNSTTQSVLSRPMQRKLVTLVHCQLVEEEGRIRAMRAARS LGERTVTELILQHQNPQQLSSNLWAAVRARGCQFLGPAMQEEALKLVLLALEDGSALS RKVLVLFVVQRLEPRFPQASKTSIGHVVQLLYRASCFKVTKRDEDSSLMQLKEEFRTY EALRREHDSQIVQIAMEAGLRIAPDQWSSLLYGDQSHKSHMQSIIDKLQTPASFAQSV QELTIALQRTGDPANLNRLRPHLELLANIDPSPDAPPPTWEQLENGLVAVRTVVHGLV DYIQNHSKKGADQQQPPQHSKYKTYMCRDMKQRGGCPRGASCTFAHSQEELEKFRKMN KRLVPRRPLSASLGQLNEVGLPSAAILPDEGAVDLPSRKPPALPNGIVSTGNTVTQLI PRGTDPSYDSSLKPGKIDHLSSSAPGSPPDLLESVPKSISALPVNPHSIPPRGPADLP PMPVTKPLQMVPRGSQLYPAQQTDVYYQDPRGAAPPFEPAPYQQGMYYTPPPQCVSRF VRPPPSAPEPAPPYLDHYPPYLQERVVNSQYGTQPQQYPPIYPSHYDGRRVYPAPSYT REEIFRESPIPIEIPPAAVPSYVPESRERYQQIESYYPVAPHPTQIRPSYLREPPYSR LPPPPQPHPSLDELHRRRKEIMAQLEERKVISPPPFAPSPTLPPTFHPEEFLDEDLKV AGKYKGNDYSQYSPWSCDTIGSYIGTKDAKPKDVVAAGSVEMMNVESKGMRDQRLDLQ RRAAETSDDDLIPFGDRPTVSRFGAISRTSKTIYQGAGPMQAMAPQGAPTKSINISDY SPYGTHGGWGASPYSPHQNIPSQGHFSERERISMSEVASHGKPLPSAEREQLRLELQQ LNHQISQQTQLRGLEAVSNRLVLQREANTLAGQSQPPPPPPKWPGMISSEQLSLELH QVEREIGKRTRELSMENQCSLDMKSKLNTSKQAENGQPEPQNKVPAEDLTLTFSDVPNGSALTQENISLLSNKTSSLNLS **EDPEGGGDNNDSQRSGVTPSSAP** 

Searching for	Biosequences		
& All	Enter a protein or nucleotide string. Learn more about Biosequence Search.		
Substances	BLAST CDR Motif Clear Search		
	GDQSHKSHMQSIIDKLQTPASFAQSVQELTIALQRTGDPANLNRLRPHLELLANIDPSPDAPPTWEQLENGLVAVR A Sequence Type:		
A Reactions	TVVHGLVDYIQNHSKKGADQQQPPQHSKYKTYMCRDMKQRGGCPRGASCTFAHSQEELEKFRKMNKRLVPRRPLSAS LGQLNEVGLPSAAILPDEGAVDLPSRKPPALPNGIVSTGNTVTQLIPRGTDPSYDSSLKPGKIDHLSSSAPGSPPDL Nucleotide Protein		
References	LESVPKSISALPVNPHSIPPRGPADLPPMPVTKPLQMVPRGSQLYPAQQTDVYYQDPRGAAPPFEPAPYQQGMYYTP PPQCVSRFVRPPPSAPEPAPPYLDHYPPYLQERVVNSQYGTQPQQYPPIYPSHYDGRRVYPAPSYTREEIFRESPIP		
🐂 Suppliers	IEIPPAAVPSYVPESRERYQQIESYYPVAPHPTQIRPSYLREPPYSRLPPPPQPHPSLDELHRRRKEIMAQLEERKV ISPPPFAPSPTLPPTFHPEEFLDEDLKVAGKYKGNDYSQYSPWSCDTIGSYIGTKDAKPKDVVAAGSVEMMNVESKG Limit Total Sequence Results to:		
	MRDQRLDLQRRAAETSDDDLIPFGDRPTVSRFGAISRTSKTIYQGAGPMQAMAPQGAPTKSINISDYSPYGTHGGWG ASPYSPHQNIPSQGHFSERERISMSEVASHGKPLPSAEREQLRLELQQLNHQISQQTQLRGLEAVSNRLVLQREANT		
√ <sup>®</sup> Biosequences	LAGQSQPPPPPPKWPGMISSEQLSLELHQVEREIGKRTRELSMENQCSLDMKSKLNTSKQAENGQPEPQNKVPAED LTLTFSDVPNGSALTQENISLLSNKTSSLNLSEDPEGGGDNNDSQRSGVTPSSAP		
Retrosynthesis	Advanced Biosequence Search V		

72 Searches are returned with different degrees of Matching.

4 Searches are returned with 100% Matching including #1. The search also returns the Organism each MOTIF is found in. This shows similarities between different species for this protein.

Motif Search Details	Sort: Alignment Identity	View: Expande
Sequence Type: Protein NCBI Included: Yes Query Coverage: 90%	References	اً الح
-value. To	Query Details > Seq 1: 1 MPVQAPQWTDFLSCPICTQTFDETIRKPISLGCGHTVCKMCLNKLHRKACPFDQTTINTDIELLPVNS	ALLQL View Mor
Bioscape Analysis	1 Align	ment Identity: 100%
/isually explore sequence imilarity with a new tool. .earn more about Bioscape.	Query (1) (1,133)	Matches: 1.13
Create Bioscape Analysis	Subject (1)	Mismatches: 0
E-Value	View Less	
) • to	Alignment Subject References	References
Query Coverage %	Alignment Data BLAST Score: 8577 E-Value: 0	
) to	Q 1 MPVQAPQWTD FLSCPICTQT FDETIRKPIS LGCGHTVCKM CLNKLHRKAC PFDQTTINTD IELLPVNSAL 70 	<u>^</u>
Subject Coverage %	Q 71 LQLVGAQVPE QQPITLCSGV EDTKHYEEAK KCVEELALYL KPLSSARGVG LNSTTQSVLS RPMQRKLVTL 14 	0
to	Q       141 VHCQLVEEEG RIRAMRAARS LGERTVTELI LQHQNPQQLS SNLWAAVRAR GCQFLGPAMQ EEALKLVLLA 21         S       141 VHCQLVEEEG RIRAMRAARS LGERTVTELI LQHQNPQQLS SNLWAAVRAR GCQFLGPAMQ EEALKLVLLA 21	0 0 <b>-</b>
Alignment Identity %	2 Align	ment Identity: 100
to	Query (1) (1,133)	
00		Matches: 1,10 Mismatches:
Orgeniamo	subject (1,109)	
Macaca fascicularis (3)	View Less V	
Papio anubis (3)	Alignment Subject References	References

#### System Search included the NCBI DB.

arch 21, 2022			
√ <sup>b</sup> Biosequences	Sequence Type: Protein	MPVQAPQWTDFLSCPICTQTFDETIRKPISLGCGHTVCKMCLNKLHRKACPFDQTTINTDIELLPV	
11:08 AM	NCBI Included: Yes	NSALLQLVGAQVPEQQPITLCSGVEDTKHYEEAKKCVEELALYLKPLSSARGVGLNSTTQSVLSRP	
	Query Coverage: 90%	MQRKLVTLVHCQLVEEEGRIRAMRAARSLGERTVTELILQHQNPQQLSSNLWAAVRARGCQFLGPA	
	E-value: 10	MQEEALKLVLLALEDGSALSRKVLVLFVVQRLEPRFPQASKTSIGHVVQLLYRASCFKVTKRDEDS	Edit Options
		SLMQLKEEFRTYEALRREHDSQIVQIAMEAGLRIAPDQWSSLLYGDQSHKSHMQSIIDKLQTPASF	
	Results will expire on	AQSVQELTIALQRTGDPANLNRLRPHLELLANIDPSPDAPPPTWEQLENGLVAVRTVVHGLVDYIQ	Searching
	Apr 20, 2022.	NHSKKGADQQQPPQHSKYKTYMCRDMKQRGGCPRGASCTFAHSQEELEKFRKMNKRLVPRRPLSAS	20010110
		LGQLNEVGLPSAAILPDEGAVDLPSRKPPALPNGIVSTGNTVTQLIPRGTDPSYDSSLKPGKIDHL	
		SSSAPGSPPDLLESVPKSISALPVNPHSIPPRGPADLPPMPVTKPLQMVPRGSQLYPAQQTDVYYQ	
		DPRGAAPPFEPAPYQQGMYYTPPPQCVSRFVRPPPSAPEPAPPYLDHYPPYLQERVVNSQYGTQP	

Further in the research results the matches decrease and mismatched increase. Here is result #72.

The red vertical lines denote the mismatch locations between the Query and the Subject along the genetic string motif.

72 Alignm	ent Identity: 95.23%
Query         1         1,133           Subject         1         1,133	Matches: 1,079 Mismatches: 54
View Less 🗸	
Alignment Subject References	References
Alignment Data BLAST Score: 8523 E-Value: 0	
Q 1 MPVQAPQWTD FLSCPICTQT FDETIRKPIS LGCGHTVCKM CLNKLHRKAC PFDQTTINTD IELLPVNSAL 70	A
<ul> <li>MPVQAPQWXD FLSCPICXQX FDEXIRKPIS LGCGHXVCKM CLNKLHRKAC PFDQXXINXD IELLPVNSAL 70</li> <li>O 71 LOLVGA0VPE OOPITLCSGV EDTKHYEEAK KCVEELALYL KPLSSARGVG LNSTTOSVLS RPMORKLVTL 14</li> </ul>	0
	a
Q 141 VHCQLVEEEG RIRAMRAARS LGERTVTELI LQHQNPQQLS SNLWAAVRAR GCQFLGPAMQ EEALKLVLLA 21	0
S 141 VHCQLVEEEG RIRAMRAARS LGERXVXELI LQHQNPQQLS SNLWAAVRAR GCQFLGPAMQ EEALKLVLLA 210	0 🗸