## **Bioinformatics Protocols:** Antisense Peptide with Molecular Modelling



Section	Title	Pages
1	Background, aims and method overview	2-4
2	Target protein mRNA search	5-11
3	Antisense peptide generation using online Python	12-14
4	Confirmation of correct target mRNA	15-17
5	Antisense peptide BLAST searches	18-20
6	BLAST data extraction	21-31
7	Molecular recognition analysis	32-40
8	PDB files for protein-protein interaction modelling	41-52
9	ClusPro protein docking	53
10	ZDock protein-protein interaction 3D modelling	53-56
11	iCn3D protein-protein interaction data extraction	57-68
12	Protein-protein interaction images using EzMol	69-73
13	Interpretation of results	74-75
14	References and website links	76-79
15	Appendix 1 - Installing Python	80-83
16	Appendix 2 - Python antisense peptide generation	84-87
17	Appendix 3 - Manual antisense peptide generation	88-92
18	Acknowledgements	93



# 1: Background, aims and method overview

#### Background:

Computer based prediction of protein-protein interactions is a valuable *in silico* tool (Capone *et al.,* 2008; McGuire & Holmes 2005; Siemion *et al.,* 2004; Štambuk *et al.,* 2018; Štambuk *et al.,* 2021) in a biological setting. Antisense peptide (or Complementary Peptides) sequences are derived from the complementary strand of DNA encoding a given protein (Bost & Blalock 1989a & 1989b), read in the same open reading frame (ORF). Due to the presence of exons and introns within the genomic DNA sequence the mRNA sequences are ideal for generating antisense peptides. They can also be derived directly from the amino acid sequence of a protein, via reverse translation to produce a complementary DNA sequence. However, due to the degeneracy of the genetic code, there is typically more than one antisense sequence for any one protein (Bost & Blalock 1989a & 1989b). The basis of antisense peptides based on the observations of Blalock & Smith (1984) is illustrated below:





The Molecular Recognition Theory is based on a series of observations that protein sequences derived from the Sense Strand of DNA bind to protein sequences derived from the corresponding Antisense Strand of the DNA (Biro 2007; Blalock & Bost 1988; Blalock & Smith 1984; Hardison & Blalock 2012; Heal *et al.*, 2002; Root-Bernstein & Holsworth 1998; Štambuk *et al.*, 2005; Štambuk *et al.*, 2021).

The complementary DNA strand for each individual amino acid can be read in either the forward 3'-5'or reverse 5'-3' direction, adding further degeneracy to the potential antisense peptide sequences (Bost & Blalock 1989a & 1989b; Milton 2006). The antisense peptides have been shown to bind with high affinity to the given target protein due to hydropathic interactions (Illingworth *et al.*, 2012; Pullen *et al.*, 2013), which is basically an interaction between water loving and water hating amino acids (Kyte & Doolittle 1982).

This protocol details the generation of antisense peptide sequences against a selected target (Bost & Blalock 1989b) and then using them to identify possible protein-protein interactions (Miller 2015). The method uses a python script to generate the antisense peptides against the selected target based on the methods of Milton (2006) and then BLAST searches to identify sequences with similarity to the antisense sequences that may interact with the target protein. This is followed by *in silico* modelling of protein-protein interactions between the target and identified protein from the BLAST search (Pierce *et al.,* 2011 & 2014).

Binding of proteins to their antisense proteins has been demonstrated in a number of studies and the antisense peptides have also been shown to have sequence similarity to receptor binding sites plus compounds that specifically bind the sense peptides (Blalock & Bost 1986; Bost *et al.*, 1985; Clarke & Blalock 1990; Fassina *et al.*, 1989; Milton *et al.*, 2001; Mulchahey *et al.*, 1986; Štambuk *et al.*, 2019). The antisense peptides themselves have been used as binding peptides to modify the actions of the target protein (Bost *et al.*, 1985; Štambuk *et al.*, 2021). Antibodies raised against antisense peptide sequences have also been used to identify binding proteins *in vitro* or *in vivo*, for example LHRH as shown by Mulchahey *et al.*, (1986).

From a Bioinformatic point of view the DNA sequences of the Sense Strands that encode proteins are contained in many databases including the NCBI Nucleotide Database (<u>https://www.ncbi.nlm.nih.gov/nucleotide/</u>). As such it is possible to download the DNA sequence and use this as a source material. The amino acid sequences of the proteins from many species are likewise contained in many databases such as NCBI Protein Database (<u>https://www.ncbi.nlm.nih.gov/protein/</u>). If the sequences for proteins encoded by the Antisense strand of DNA for a given target protein are derived, they can be used to search protein databases for similar proteins using sequence comparison tools (<u>https://blast.ncbi.nlm.nih.gov/Blast.cgi</u>).

The nature of the identified interactions from this process is two-dimensional and does not fully take into account the three-dimensional structure (3D) of each protein. The final stage of this protocol is to use 3D modelling to predict the structure of complexes between the target protein and binding protein identified due to sequence similarity with the antisense peptide derived from the target protein mRNA. The original ZDock method (Chen & Weng 2002) used for this has been updated and is well described by



Pierce *et al.*, (2011 & 2014). The method uses an algorithm to predict the proteinprotein interactions with input of data from the identified interactions. The combination of the antisense methods with 3D modelling strengthens the *in silico* observations. Key confirmation of the validity of identified interactions is to determine that the *in silico* findings can be replicated *in vitro* and/or *in vivo*, for example catalase binding to amyloid- $\beta$  (Milton *et al.*, 2001) and the role of this interaction in preventing the toxicity of amyloid- $\beta$  combined with demonstration in samples from an Alzheimer's patient (Chilumuri *et al.*, 2013a & b).

#### Aims:

- (i) Generation of Antisense peptide sequences for a chosen target protein.
- (ii) Identification of potential protein binding partners for the chosen target protein.
- (iii) Identification of the regions of each interacting protein pair that are directly involved in the binding.
- (iv) 3D modelling of identified interactions

#### Method Overview:

The method uses either a PC or Mac (both Intel and Apple Silicon powered) based computers and can be run on Windows 10/11 or Mac OS. The protocols have been written using Microsoft Office 365 software with Microsoft Word and Microsoft Excel the main programs employed. The Python program for antisense peptide generation is compatible with Python 3 and has been tested on versions 3.7 upwards. Antisense peptide generation can also be carried out using Microsoft Word if online Python or Python installation is not possible.

- (i) Generation of antisense peptides (Milton 2006) that are model binding proteins for the chosen target.
- (ii) Searching of protein databases using a BLAST search for proteins that are similar to the antisense sequences and therefore may also be binding proteins for the chosen target.
- (iii) Extraction of data from the BLAST search and exclusion of results that are incompatible with the protein binding, for example sequences with Gaps that are identified in by the BLAST searching techniques, but which are would not indicate a potential binding interaction (Pullen *et al.*, 2013).
- (iv) Scoring of the binding interactions based on the Molecular Recognition Theory (Hardison & Blalock 2012) to identify the most relevant protein-protein interactions from the BLAST search.
- (v) 3D molecular modelling of the potential protein-protein interactions (Pierce *et al.*, 2011 & 2014) using the potentially interacting residues identified in BLAST searches as a basis for the contacting residues in the model.



(2a) Stage one is to find the mRNA coding sequence for the protein of interest. Best to use a normal browser such as Microsoft Edge, Safari, Google Chrome or Firefox for these searches. The mRNA sequences can be found on the <u>https://www.ncbi.nlm.nih.gov</u> databases – click on the nucleotide link (circled in red).

National Cent	ter for Biotechnol × +					×
← → C ☆ â ncbi	.nlm.nih.gov			ů 🕇	* * 🗆 🕯	. :
E Favourites						
An official website of the Ur	ited States government Here's how you know ~					
	Library of Medicine ar for Biotechnology Information				Log	in
All	Databases V			Search	L	
All	Databases  Welcome to NCBI			Search Popular Resour	Ces	_
All NCBI Home Resource List (A-Z)	Welcome to NCBI The National Center for Biotechnology Inform	nation advances science and h	ealth by providing access to biomedical	Search Popular Resour PubMed	ces	-
All . NCBI Home Resource List (A-Z) All Resources	Welcome to NCBI The National Center for Biotechnology Inform and genomic information.	nation advances science and h	ealth by providing access to biomedical	Search Popular Resour PubMed Bookshelf	rces	
All NCBI Home Resource List (A-Z) All Resources Chemicals & Bioassays	Welcome to NCBI           The National Center for Biotechnology Inform and genomic information.           About the NCBI   Mission   Organization   N	nation advances science and he	ealth by providing access to biomedical	Search Popular Resour PubMed Bookshelf PubMed Central	rces	
All NCBI Home Resource List (A-Z) All Resources Chemicals & Bioassays Data & Software	Welcome to NCBI           The National Center for Biotechnology Informand genomic information.           About the NCBI   Mission   Organization   Mission   Companying the second	nation advances science and he NCBI News & Blog	ealth by providing access to biomedical	Search Popular Resour PubMed Bookshelf PubMed Central BLAST	rces	

Which will go to where the name of protein of interest can be entered (as an example the Homo sapiens amyloid precursor protein has been used:

🔴 🔍 📄 🕑 Hom	e - Nucleotide - NCBI × +						~
< → C △ (	ncbi.nlm.nih.gov/nucleotide/	đ	☆	*		-	:
Favourites							
An official website	of the United States government Here's how you know						
NIH Nati Nationa	onal Library of Medicine al Center for Biotechnology Information				Lo	og in	)

This will then go to a list of nucleotide sequences that match the search term. By selecting Animals, mRNA and Nucleotide on the Left-hand side (marked in blue with a tick after selection – the sequence list can be limited to those of most use. Selecting RefSeq limits the search further to those sequences that have been reviewed by the NCBI and is therefore recommended. The main features of the RefSeq collection include:

- non-redundancy
- explicitly linked nucleotide and protein sequences
- updates to reflect current knowledge of sequence data and biology
- data validation and format consistency
- <u>distinct accession series</u> (all accessions include an underscore '\_' character)
- ongoing curation by NCBI staff and collaborators, with reviewed records indicated

🗧 😐 🌒 🔊 Hor	no sapien	s amyloid precurs × +		~
← → C ☆	🗎 ncbi	.nlm.nih.gov/nuccore	Ů ☆	* 🗉 😩 🗄
Favourites				
An official website	e of the Ur	nited States government Here's how you know Y		
NIH Nation	t <b>iona</b> l	Library of Medicine er for Biotechnology Information		Log in
Nucleotide	Nuc	eleotide	Search	
		Create alert Advanced		Help
Species	clear	Summary + 20 per page + Sort by Default order + Send to: +	Filter your results:	
✓ Animals (666) human (654)			All (666)	
Customize		Items: 1 to 20 of 666		Manage Filters
Molecule types	clear	<< First < Prev Page 1 of 34 Next > Last >>		
✓ mRNA (666) Customize		Filters activated: Animals, mRNA, RefSeq, Nucleotide. Clear all	Results by taxon	
Source databases	clear	Homo sapiens amyloid beta precursor protein (APP), transcript variant 10, mRNA     3,304 bp linear mRNA	Top Organisms [Tree] Homo sapiens (654)	
INSDC (GenBank) (0) RefSeq (666)		Accession: NM_001204303.2 GI: 1676441520 PubMed Taxonomy	Necator americanus (9) Pan troglodytes (3)	
Customize		GenBank FASTA Graphics		
Sequence Type	clear		Find related data	۲
Vucleotide (666)		Homo sapiens amyloid beta precursor protein (APP), transcript variant 3, mRNA     3 358 bp linear mRNA	Database: Select	~
Sequence length		Accession: NM_201414.3 GI: 1676319912		
Delever dete		PubMed Taxonomy		
Custom range		GenBank FASTA Graphics		
Pewision date		Homo sapiens amyloid beta precursor protein (APP), transcript variant 2, mRNA	Search details	۲
Custom range		3. 3,526 bp linear mRNA Accession: NM_201413.3 GI: 1675178653	(("Homo sapiens"[Organ sapiens[All Fields]) / Fields1 AND precursor	nism] OR Homo AND amyloid[All [All Fields]

The choice of sequence is dependent on several features. If the literature suggests a particular transcript is of interest, then that is the recommended sequence to choose. It is also important to check the names carefully, in the example below the search term "Homo sapiens amyloid precursor protein" will also identify related proteins such as the "amyloid precursor like proteins":

Homo sapiens amyloid precurs × +			~
← → C △ ê ncbi.nlm.nih.gov/nuccore Favourites	₫ ☆	*	<b>1</b>
<ul> <li>Homo sapiens amyloid beta precursor protein (APP). transcript variant 4. mRNA</li> <li>3,572 bp linear mRNA         Accession: NM_001136016.3 GI: 324021746         PubMed Taxonomy         GenBank FASTA Graphics     </li> <li>Homo sapiens amyloid beta precursor like protein 1 (APLP1), transcript variant 2, mRNA     </li> <li>2,367 bp linear mRNA         Accession: NM_005166.5 GI: 1890341900         Protein PubMed Taxonomy         GenBank FASTA Graphics     </li> </ul>			
Homo sapiens amyloid beta precursor like protein 2 (APLP2), transcript variant 1, mRNA     3,733 bp linear mRNA     Accession: NM_001642.3 GI: 1677484618     PubMed Taxonomy     GenBank FASTA Graphics			

The transcript of interest should be selected, generally select the longest transcript and run that through the antisense peptide generation first. Clicking on the link will go to:

Homo sapiens amyloid beta pri × +		~
← → C △ 🔒 ncbi.nlm.nih.gov/nuccore/NM_000484.4	₫ ☆ 🚖	: 😩 🗉
Favourites		
An official website of the United States government Here's how you know >		
NIH National Library of Medicine		Log in
Nucleotide Nucleotide   Advanced	Search	Help
GenBank + Send to: +	Change region shown	
Homo sapiens amyloid beta precursor protein (APP), transcript variant 1, mRNA NCBI Reference Sequence: NM_000484.4	Customize view	*
EASTA Graphics	Analyze this sequence Run BLAST	۲
LOCUS NM_000484 3583 bp mRNA linear PRI 03-JUL-2022	Pick Primers	
DEFINITION Homo sapiens amyloid beta precursor protein (APP), transcript variant 1, mRNA.	Highlight Sequence Features	
ACCESSION NM_000484	Find in this Sequence	
VERSION NM 000484 4	r ind in this bequence	
VERSION NN_000484.4 KEYWORDS RefSeg: MANE Select. SOURCE Homo sapiens (human) ORGANISM Homo sapiens	Show in Genome Data Viewer	

The next step is to scroll down this page until the CDS link (circled in red) is reached:



(2b) Clicking the protein id link (circled in blue, = NP\_000475.1 in this example) will go to the coded protein and this number (normally starting NP\_) can also be used in the BLAST check detailed in Section 4a (pages 15-16). Click on the CDS link (circled in red) will go to the following:



The part in Brown is the part required to generate antisense peptides. The FASTA tab (circled in red) will go to:

••• • Homo sapiens amyloid beta pr × +		~
← → C ☆ a ncbi.nlm.nih.gov/nuccore/NM_000484.4?from=151&to=2463&report=fasta	₫ ☆	* 🗉 😩 🗄
Favourites		
An official website of the United States government Here's how you know.		
NIH National Library of Medicine National Center for Biotechnology Information		Log in
Nucleotide  Nucleotide  Advanced	Search	Help
FASTA - Send to: -	Change region shown	
		÷.
Homo sapiens amyloid beta precursor protein (APP), transcript variant 1, mRNA NCBI Reference Sequence: NM_000484.4 GenBank Graphics	Whole sequence     Selected region     from: 151     to: 2463	Update View

(2c) The text starting at ATG and finishing at either TAA, TAG (in this example - <u>https://www.ncbi.nlm.nih.gov/nuccore/NM\_000484.4?from=151&to=2463&re</u> <u>port=fasta</u>) or TGA should be copied to a word file and save as mRNA. This is the sequence that will be used by the AntisensePeptide.py program to generate the antisense sequences after removal of the returns.

The returns at the end of each line are removed as follows. First in word select the view symbols tab (circled in red below):

	Aut	oSave Oorr	68	9.0	⊖ …			nRNA — S	wed to my Ma	o~					ର ୫
Home	Insert	Draw	Design	Layout	References	Mailings	Review	View	EndNote 20	V Tell me				🖻 Share	Comments
Paste		Calibri (Bo B I <u>U</u>	v 12 ∘	A A A	Aav Ao		• ₩ •		2↓ <b>¶</b> ∞ • ⊞ •	AaBbCcDdEe Normal	AaBbCcDdEe No Spacing	AaBbCcDc Heading 1	Styles	Dictate	Editor
D	iΞ	D Q	© I*	2	п. <del>1</del> . г. Д. 1	, I <sub>1</sub> 1, ₹.	(, <sup>1</sup> , )	(† . ( <del>.</del>		1, * . 1	10, 11, 11, 1	$(\mathbf{r}, \mathbf{r})$		5,1,36	17 18

This will show the returns as " $\P$ ", which need to be deleted:

	AutoSavi		68	9.0	<del>.</del>		APP n	nRNA — S	aved to my Ma	<b>a</b> ~						ଦ ୫
Home	Insert D	Draw [	Design	Layout	References	Mailings	Review	View	EndNote 20	0	Tell me				년 Share	Comments
Paste *	X Calib 3 B	ri (Bo I <u>U</u> ∽	v 12 v ⇔ x,	A A ×' A	Aa v   A v <u>ℓ</u> v <u>A</u> v		:•₩.• ≡≡	⊡ ⊡ ≣ •	2↓ ¶ ◇ • ⊞ •		AaBbCcDdEe Normal	AaBbCcDdEe No Spacing	AaBbCcDc Heading 1	Styles Pane	Dictate	Editor
	2 1	200	ATGCTG	CCCGG CTGGC	TTTGGCAG CTGCTGGG AGTGGGA	TGCTCCT CTGAACCC	GCTGGC CAGAT CATCAC	CCGCCT FGCCAT	GGACGGCT GTTCTGTG AAAACCTG	CGG GCA	GCGCTGG GACTGAP	AGGTACC CATGCAC	CACTGAT ATGAATGI GCATCCTG	15     9		17 18

In the file go to the EDIT tab, then Find and select Replace:

🧉 Word File Edit	t View Insert Format Tools	Table Window Help
😑 🔿 🕥 🛛 AutoSav Unde	do Paste # Z	APP mRNA — Saved to my Mac - Q
Home Insert C Can'	n't Repeat 🛛 🕷 Y	Mailings Review View EndNote 20 🔉 Tell me
Paste	₩ X by ೫ C	Harding 1 Styles Dictate Editor
Paste Paste	te \$\$¢ v	Pane Pane 10, 11, 12, 13, 14, 15, 16, 17, 18
Past	te and Match Formatting てるまV	GAACCCCAGATTGCCATGTTCTGTGGCAGACTGGAACATGCACATGAATGT
Clear Seler	ar > ect All % A	CAGATCCATCASGGACCAAAACCTGCATTGATACCAAGGAAGGCATCCTG CCCTGAACTGCAGATCACCAATGTGGTAGAAGCCAACCAA
Find	d >	Find ## F AAGTGCAAATTCTTACACCAGGAGAGGG
- Links	KS	Find Next         #G         AAGAGACATGCAGTGAGAAGAGTACCA           Find Previous         0#G         STTCCGAGGGGTAGAGTTTGTGTGTGTGTGTGTGTGTGTGT
Start	rt Word Dictation	Replace SAGGAGGATGACTCGGATGTCTGGTGG
selec	ect Data	Go to TAGTAGAAGTAGCAGGAGGAGGAAGAAG GAGGATGGTGATGAGGTAGAGGAAGAAG GGAGGATGGTGATGAGGTAGAGGAAGAAG
Togg	gle Drawing ^ # Z	ATTGCCACCACCACCACCACCACCACCAC

For Mac this shows as follows, when the arrow in the Find box is clicked (circled in red), then select Paragraph Mark (circled in blue) and leave the Replace blank:



÷J



🗯 Word File Edit View Insert Format Tools	Table Window Help	
● ● ● AutoSave ● • 師 合 曽 ウ ~ C 日 …	APP mRNA — Saved to my Mac 🗠	Q 8
Home Insert Draw Design Layout References	Mailings Review View EndNote 20 🖓 Tell me	🖻 Share 🛛 🖓 Comments
$\begin{array}{c c} & & & \\ & & \\ \hline & & \\ Paste & \\ \hline & \\ & \\ \hline & \\ & \\$	Image: Style state     Image: Style state     AaBbCcDdEe     AaBbCcDdEe     AaBbCcDdEe     AaBbCcDdEe     AaBbCcDdEe     Style state       Image: Style state     Image: Style state     Image: Style state     Image: Style state     Image: Style state     Image: Style state     Image: Style state       Image: Style state     Image: Style state     Image: Style state     Image: Style state     Image: Style state     Image: Style state       Image: Style state     Image: Style state     Image: Style state     Image: Style state     Image: Style state     Image: Style state	Dictate
	$1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\$	15 16 17 18
FIND AND REPLACE      Paragraph Mark     Find     Tab Character     Any Character     Any Digit     eplace     Any Digit     GTR	CTGCCCGGTTTGGCACTGCTCGCTGGCCGCCTGGACGGCTCGGGCGCTGGAGGTACCCACTGATG ATGCTGGCCTGCTGGCTGGACGCCCACGTTGCCATGTCTGTGGCAGACTGAACATGCACTGAATGAA	9 9

On a PC in Word running in Windows the trick is to select replace (crtl H - also circled in red) and then the More button (circled in blue):

	• •																V	/indov	vs 11		
🚾 A	kutoSave 🦲	Off	APP	mRNA •	Saved 1	to \\Mac\He	ome 🗸		,₽ Sea	arch (Alt-	PQ)			Ŷ	)		Nat Milto	n 🌘	¢	ä	- 1
File 9 Undo	Home Peste Clipboard	Insert کر الله ج	Draw Courier Ne B I	Design	<ul> <li>✓ ələ-</li> <li>▲ </li> <li>▲ </li> <li>Font</li> </ul>	Aa Aa AGAAG CCACC TAG¶	A A A A	Mailings Find and Find Find wha Find Kind Kind Kore	Review Replace Replace At:	go To	Help	Replace	r B	eplace All	Find Next	, ,	×	¥ 18	P Find C Find C Find C Find C Repli Selec Editing	Comr Q Dictate Voice	Editor

Then with the More window open click on the Special tab (circled in red) and then select the Paragraph Mark will then be displayed at the top (circled in blue):

<b>•</b>	lutoSave 🔘 D		9 mRNA +	Saved t	to \\Mac\						4				- 🕥	$\oplus$	0	
File	Home In	ert Draw	Design	Layo	out R	eferer	nces	Mailings	Review	View Help							Comm	ents
ි Unde	Paste	8		✓ ab ▲ ✓ Font	X, Aa Agaa Agaa Tag9 ¶	x <sup>2</sup> A <sup>^</sup> GGAAAA I	A A A S S CAGTH	Find and f Find and f Find Find Find what Replace w Search Oy Search Oy Search Oy Search Oy Search Oy Search Oy Search Oy Search Oy Search Oy Search Search	Replace Replace gitt: ss ss ss ss ss ss ss ss ss ss ss ss ss	Baragraph Mark Jab Character Any Character Any Digit Any Letter Caget Character Section Character Paragraph Character Olymn Break Em Dash En Dash En Dash Endonate Mark Frield Eootnote Mark Graphjic Manual Jine Break Manual Page Break Manual Page Break Nonbreaking Sypace Optional Hyphen Section Break	Replace All	Find Next th prefig th suffix re punctuation c re ghite-space c	? Ca	X V neel	•	D Editing	Uictate Voice	Edito



This will then show as ^P (circled in red) in the find box and blank in the replace box and the Replace All (circled in blue) should then be clicked to remove all the "¶" symbols at the end of each line:

● ● ● AutoSave ● 0FF 6 🛱 りょ 🖓 =	🖻 APP mRNA — Saved to my Mac 🗠	ଦ ୫
Home Insert Draw Design Layout References	Mailings Review View EndNote 20 🖓 Tell me	Share 🖓 Comments
$\begin{array}{c c} & & & \\ & & \\ & & \\ Paste & \\ & &$	Image: the second s	Dictate
	L 1 L 2 L 1 L 1 L 1 L 1 L 1 L 1 L 1 L 1	1 16 17 18
▼ FIND AND REPLACE GTA GTA GTA GTA GTA GTA GTA GTA	SCTGCCCGGTTTGGCACTGCTCGCGCGCCTGGACGCCTGGACGCCTGGAGGTACCCACTGATG AATCGTGGCCTGCTGGCTGAACCCCAGATTGCCATGTTCGTGGCGACGCAGAACATGCACAGAAGGCAATGT GCAATGGGAAGTGGGATTCGCAGTCGAGGACGAAACCAGCAGAAGGCAACCAGCGACGCACCAG GTATTGCCAAGAAGTCTACCCGGACGCAGCAGCAGCAAAGCCAATCGTAGAAGCCAACCAGTGACAA CAGAACTGGTGGAAGCGGGGCCGAAGCAGTGCAAGACCAACCA	

A single "¶" symbol at the end of the sequence after the TAA, TAG or TGA (in effect the whole sequence as a single word) will remain and the TAA, TAG or TGA (circled in red) triplet can be deleted as these stop codons are not needed:



(2d) Then save file and use this version of the mRNA sequence for Antisense peptide generation as described in either Section 3 below using the online Python 3 compiler (pages 12-14 below), using the downloaded Python 3 program (see Section 14 pages 80-83 and Section 15 pages 84-87) or manually as detailed in Section 16 (pages 88-92).

## **3: Antisense peptide generation using online Python**

- (3a) An online Python compiler (<u>https://trinket.io/python3</u>) can be used to generate antisense peptide sequences using the Antisense-Peptide.py file. This can be run in most browsers on a Mac, PC, Chromebook, iPad etc and has been tested using both Safari and Google Chrome.
- (3b) The text of the Antisense-Peptide.py (available from as a either Python script <u>https://www.neurodelta.uk/resources/BioinformaticsProtocolScript.py</u> or as a Word file from <u>https://www.neurodelta.uk/Protocols/</u>) can be copied and pasted into the compiler:



(3c) The Antisense-Peptide.py Python Script can also be downloaded and saved to an appropriate location on a computer hard drive and then uploaded using the compiler upload function (circled in red) followed by selection of the Antisense-Peptide.py file:

E Strinket Python3 Run main.py	≮Share ▼		( <u></u> )°	wered by 🕥 trinket
	Favourites V Dropbox	<> = •	🗖 Desktop 😒	Q Search
	🔁 Website	Name	Date Modified	✓ Size Kind
	Precents	AntisensePeptide.py	Today at 21:44	4 KB Python scri
	Applications			
	🚍 Desktop			
	Documents			
	Ownloads			
	OneDrive			
	iCloud			
	Cloud Drive			
	Tags			
	Purple			
	• Yellow			
	Red		Cancel	Choose for Upload

(3d) The compiler can then be run using the run function (circled in red) which will bring up the "Input Name:" command (circled in blue), the name or abbreviation of the target protein without spaces should be typed here followed by a return:



This will bring up the "Input coding mRNA:" command, the mRNA sequence for the target protein prepared in section 2d (page 10 above), in this example the sequence TCAGCTGATGCACAATCGTTTTTAAACGGGTTTGCGGTG is used:

	\star Plans 🞓 Learn 🛛 Help
+ ±	Powered by Itrinket
	Input coding mRNA: TCAGCTGATGCACAATCGTTTTTAAACGGGTTTGCGGTG
	SNC - nsp11 = SADAQSFLNGFAV
	SCN - nsp11 = VAFGNLFSQADAS
	AS35NC - nsp11 = SRLRVSKNLPKRH
	AS35CN - nsp11 = HRKPLNKSVRLRS
	AS53NC - nsp11 = XSICLRKXVPKRH
	AS53CN - nsp11 = HRKPVXKRLCISX

In a very rare number of cases the a, t, c or g residues in the mRNA sequence could be replaced by an "n". This will cause an UNKNOWN to show in the peptide sequences which should be replaced by an X.

Where there is an \* at the start (SCN, AS35CN and AS53CN) or end (SNC, AS35NC, AS53NC) of a sequence this is where the STOP codon was in the mRNA

\*/



and can be deleted form the sequences used to run BLAST searches. If there is an \* or an UNKNOWN in the middle of a sequence this indicates a problem with the mRNA used as these should only be at the end of coding sequences. Suggests a need to repeat section 2a-2d (see pages 5-10 above) to get the correct CDS mRNA component, particularly check section 2c and 2d (see pages 8-10 above) to create an mRNA sequence that is a single word has been completed properly.

(3e) Copy the text from Input name down to the end of the AS53CN sequence and paste into a word document:

Input name: nsp11

Input coding mRNA: TCAGCTGATGCACAATCGTTTTTAAACGGGTTTGCGGTG

SNC - nsp11 = SADAQSFLNGFAV

SCN - nsp11 = VAFGNLFSQADAS

AS35NC - nsp11 = SRLRVSKNLPKRH

AS35CN - nsp11 = HRKPLNKSVRLRS

AS53NC - nsp11 = XSICLRKXVPKRH

AS53CN - nsp11 = HRKPVXKRLCISX

Save the Python outputs word file with a suitable name. These are the sequences that will be used for BLAST searches in section 4 (see pages 14-16 below) and section 5 (pages 17-19 below):

# 4: Confirmation of correct target mRNA

(4a) The SNC sequence from the running AntisensePeptide.py on Python is the sense sequence of the protein obtained from the mRNA sequence, to check that the correct mRNA sequence has been used in AntisensePeptide.py a BLAST search using the <u>https://blast.ncbi.nlm.nih.gov/Blast.cgi</u> website can be run in the browser allowing a search for sequence identities with the SNC sequence. Select Protein Blast (circled in red):



From within the Protein BLAST paste the SNC sequence into the Enter Query Sequence. There are a number of options to select from. If the protein id for the sequence for the mRNA used to generate the antisense peptides is available (see section 2a, page 6) the Align two or more sequences checkbox can be selected and the protein id number pasted into the second box. For the purposes of the example used in the AntisensePeptide.py program above the sequence needed to BLAST is SNC – nsp11 = SADAQSFLNGFAV. The protein sequence id for nsp11 is YP\_009725312.1 (circled in red) from the search of the NCBI protein database (https://www.ncbi.nlm.nih.gov/protein):

1	S NCBI Resources 🖸	How To 🗹			
	Protein	Protein V Advanced			Search
	Ũ		COVID-19 is an emerging, rapidly evolving situation. Get the latest public health information from CDC: <u>https://www.coronavirus.gov</u> Get the latest research from NIH: <u>https://www.nih.gov/coronavirus</u> .		
	GenPept -			Send to: -	Change region shown
	NCBI Reference Sequer	acute respiratory s	syndrome coronavirus 2]		Analyze this sequence Run BLAST

\*



In the blastp search box enter the sequence, make sure the Align two or more sequences checkbox is ticked, paste the SNC sequence in the first box and enter the Protein id in the second box:

🖲 🗧 🗧 S Proteir	n BLAST: Align two or me 🗴 🖇 nsp11 [Severe acute r	espirator ×   +	
← → C	last.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins&P	ROGRAM=blastp&BLAST_PROGRAMS=blastp&PAGE_TYPE=BlastSear	ch&BLAST_SPEC=blast2seq&DATABA 🟠 🎇 🥝 🕒 🗄
0		COVID-19 is an emerging, rapidly evolving situation. Get the latest public health information from CDC: <u>https://www.coronavirus.gov</u> . Get the latest research from NIH: <u>https://www.nih.gov/coronavirus</u> .	
		Align Sequences Protein BLAST	
blastn blastp blastx	tblastn tblastx		
Enter Query Se	equence	BLASTP programs search protein subjects using a protein query. more	Reset page Bookmark
Enter accession no SADAQSFLNGFAV	umber(s), gi(s), or FASTA sequence(s) 🤢	Clear Query subrange 😖 From To	BLAST results will be displayed in a new format by default
Or, upload file Job Title ☑ Align two or mo	Choose file No file chosen		Traditional Results page.
Enter Subject S	Sequence		
Enter accession no	umber(s), gi(s), or FASTA sequence(s) 🦦	Clear Subject subrange 🖗 From To	
Or, upload file	Choose file No file chosen		
Program Selec	tion		
Algorithm	blastp (protein-protein BLAST)     Choose a BLAST algorithm		
BLAST	Search protein sequence using Blastp (protein-p	rotein BLAST)	Screenshot

Clicking on the BLAST button (circled in red) will bring up the following:

Your searce	h parameters were adjusted to search for a short input sequence.	
lob Title	Protein Sequence	Filter Results
RID	CWZDG0CC11N Search expires on 05-28 22:48 pm Download All >	Percent Identity E value Ouery Coverage
Program	Blast 2 sequences Citation 💙	
Juery ID	lcl Query_44365 (amino acid)	
Juery Descr	None	Filter Reset
Query Length	13	
iubject ID	YP 009725312.1 (amino acid)	
Subject Descr	nsp11 [Severe acute respiratory syndrome coronavirus 2]	
ubject Length	13	
	Multiple alignment MSA viewer	
other reports		
Descriptions	Graphic Summary Alignments Dot Plot	
Descriptions Sequences	Graphic Summary Alignments Dot Plot	Download × Manage Columns × Show 100 ×
Descriptions Sequences Select all	Graphic Summary Alignments Dot Plot Producing significant alignments I sequences selected	Download × Manage Columns × Show 100 <b>×</b> ( <u>GenPept Graphics Multiple alignme</u>

(4b) Selecting the Alignments tab (circled in red) shows the comparison of the amino acids, in this case confirming the SNC and the original protein are identical and therefore that the correct mRNA was used in the Antisensepeptide.py program the key number is the Positives, which should be 100% (circled in blue):

escriptions	Graphic Summa	ary Alignmen	ts Dot Plot		
ignment viev	v Pairwise		✓ Ø R	estore defaults	
sequences sele	ected 😮				
🛓 Downlo	ad v GenPept Gr	aphics			
nsp11 [S Sequence	evere acute respir	catory syndrome of Length: 13 Number	oronavirus 2] of Matches: 1		
Range 1:	1 to 13 GenPept Gra	phics	$\sim$	Vext Match 🔺 Previous Mat	ch
Score	Expect	Identities	Positives	Gaps	
42.2 bits	(92) 4e-14	13/13(100%)	13/13(100%)	0/13(0%)	
Query 3	SADAQSFLNGFAV	13			
Sbjct 3	SADAQSFLNGFAV	13			

ż



## **5: Antisense peptide BLAST searches**

(5a) The outputs form the AntisensePeptide.py program include the sense protein in normal N-terminus to C-terminus orientation (SNC) and also the sense protein in the reverse C-terminus to N-terminus orientation. The protein databases are always N-terminus to C-terminus orientation; however, proteins may interact with the binding site having one protein in the N-terminus to C-terminus orientation. Hence the need to search the C-terminus to N-terminus orientation antisense peptides.

The antisense sequences generated are:

AS35NC = Antisense peptide (3'-5' mRNA reading) N-terminus to C-terminus orientation

AS35CN = Antisense peptide (3'-5' mRNA reading) C-terminus to N-terminus orientation

AS53NC = Antisense peptide (5'-3' mRNA reading) N-terminus to C-terminus orientation

AS53CN = Antisense peptide (5'-3' mRNA reading) C-terminus to N-terminus orientation

(5b) Separate BLAST searches should be carried out for each antisense sequence (AS35NC, AS35CN, AS53NC & AS53CN) using the <u>https://blast.ncbi.nlm.nih.gov/Blast.cgi</u> website to run a Protein BLAST:





Under the header "Enter accession number, gi or FASTA sequence" paste in the antisense peptide sequence to be screened. Also enter the protein name into the Job title box.

Under the "Choose Search Set", "Database" select RefSeq Select proteins (refseq\_select), under the "Organism" type in homo sapiens and select "Homo sapiens (taxid:9606)", also select the tick boxes for Exclude "Models (XM/XP)", "Non-redundant RefSeq proteins (WP)" and "Uncultured/environmental sample sequences" plus select Algorithm "blastp (protein-protein BLAST)" – see red \* marks below:

← → C ● 1	blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&P	AGE_TYPE=BlastSearch&LINK_LOC=blasthome	: 😌   😒 :
		Standard Protein BLAST	
blastn blastp blast	s thiastn thiastn		
Enter Query S	anuanca	BLASTP programs search protein databases using a protein query. more	Reset page Bookmark
Enter accession of	number(e) gi(e) or EASTA sequence(e) ()	Circuit O una understande (1)	
VDERF2QVKQTLNDH NEILRLIKETYXLTHL XSLINVDMSXVNSCLXF	Initial (4, 10, 0, 0, 1907) a addation (4) PARTARAGIJANANCTGVISBORITSKOPERENGVISBITYTYMEPPET BRICONSTRUCTOREGONIC (1997) BRICONSTRUCTOREGONIC (1997) APPETER (1997) APPET	From To	BLAST results will be displayed in a new format by default You can always switch back to the
Or, upload file	Choose file No file chosen		Traditional Results page.
🛨 Job Title	orf3a(G251V)		
	Enter a descriptive title for your BLAST search 🤬		
Align two or me	ore sequences 🚱		
Choose Searc	h Set		
* Database	RefSeq Select proteins (refseq_select)	0	
Organism     Optional	Homo sapiens (taxid:9606) Enter organism common name, binomial, or tax id. Only 20 to	p taxa will be shown.	
Exclude Optional	Models (XM/XP) Non-redundant RefSeq proteins	(WP) Z Uncultured/environmental sample sequences	
Program Sele	ction		
🗙 Algorithm	blastp (protein-protein BLAST)     PSI-BLAST (Position-Specific Iterated BLAST)     PHI-BLAST (Pattern Hit Initiated BLAST)     PHI-BLAST (Pattern Hit Initiated BLAST)     DELTA-RELAST (Domain Enhanced Lookup Time Ac Choose a BLAST algorithm	celerated BLAST)	
BLAST	Search database refseq_protein using Blastp (prote Show results in a new window	in-protein BLAST)	Screenshot

Next click on "Algorithm parameters" (circled in red) which will go to:

			🗟 blast.ncbi.nlm.nih.gov	60
	Ngorithm parameter	Note: Parameter values that differ from the	he default are highlighted in yellow and marked with + sign	(Restore default search parameters)
	General Paran	neters		
*	Max target sequences	5000     Solect the maximum number of aligned sequences to display		
	Short queries	Automatically adjust parameters for short input sequences 😜		
*	Expect threshold	200001		
	Word size	0 0 0		
	Max matches in a query range	0 0		
	Scoring Param	neters		
	Matrix	BLOSUM62 🔯 🤬		
	Gap Costs	Existence: 11 Extension: 1 📴 🥹		
	Compositional adjustments	Conditional compositional score matrix adjustment		
	Filters and Ma	sking		
	Filter	Low complexity regions 🧕		
	Mask	Mask for lookup table only Q Mask lower case letters Q		
1	BLAST	Search database nr using Blastp (protein-protein BLAST) Bhow results in a new window		

Change "Max target sequences" to 5,000 from the dropdown and "Expect threshold" to 20,000 – leave all other settings as the defaults (the Short

queries box "Automatically adjust parameters for short input sequences" is checked). The default Matrix under Scoring Parameters is BLOSUM62 (for information about this and other options have a look at Pearson 2013). Then click on BLAST (circled in red) and wait for results to appear.

Note that any sequence with 30 or less amino acids will be automatically searched with adjusted parameters. The short sequence parameters automatically changed for short sequences like this will be the Expected threshold (which will be increased to 200,000), the Word size (which will reduce to 2), the Matrix (which will change to PAM30) and the Compositional adjustments (which will be set to no adjustment).

(5c) After clicking BLAST the screen will initially show:

Request ID	CXW1XC8J016
Status	Searching
Submitted at	Wed May 27 18:57:01 2020
Current time	Wed May 27 18:57:22 2020
Time since submission	00:00:20

If the sequence searched comprises 30 or less amino acids the following warning will then show:

Your search parameters were adjusted to search for	a short input sequence.
equest ID	GYCJ84F0016
tatus	Searching
ubmitted at	Wed Jul 15 12:43:52 2020
urrent time	Wed Jul 15 12:44:14 2020
ime since submission	00:00:21

This page will be automatically updated in 12 seconds

After the search has finished the screen will show the following. Click on Download All dropdown and select Text. Save txt file to an appropriate folder (this file will be used in sections 6 (pages 20-30) and 7 (pages 31-39). A screenshot or save as pdfs file should be created after selecting each of the Descriptions, Graphic Summary and Alignment tabs. Saving as pdf's preserves the links that can be used in data analysis more easily later, for an example and other files that can be downloaded see section 6 (pages 20-30):

< Edit Sear	Save Search Search Summary 💙	How to read this report? BLAST Help Videos DBack to Traditional Results Page
Your se	arch is limited to records that include: Homo sapiens (	axid:9606) ; and exclude: models (XM/XP), uncultured/environmental sample sequences, non-redundant
Reised	proteins (WP)	
Job Title	proteins (WP) orf3a(G251V)	Filter Results
Job Title RID	proteins (WP) orf3a(G251V) CXW1XC8J016 Search expires on 05-29 06:57 am	Filter Results
Job Title RID Program	proteins (WP) orf3a(G251V) CXW1XC8.J016 Search expires on 05-29 06:57 am BLASTP C Citation ~	ism only top 20 will appear exclude



## 6: BLAST data extraction

(6a) Data should be extracted from each of the BLAST searches carried out, giving a total of four sets of data for each target protein (corresponding to the BLAST searches for AS35NC, AS35CN, AS53NC and AS53CN.

Since the AS35NC and AS53NC peptides will theoretically bind to the SNC peptide this suggests that anything similar to the AS35NC or AS53NC peptides would bind to the SNC peptide. In the BLAST search Alignments, the Query corresponds to the protein searched (AS35NC or AS53NC peptide) and the Sbjct corresponds to the named protein that is similar. The Query residue numbers are identical to the SNC numbers so for example if the query numbers were 4-9 and the Sbjct residues were 321-326 this would suggest that Sbjct protein 321-326 theoretically binds residues 4-9 of the SNC protein.

(6b) The situation with the AS35CN and AS53CN peptides is similar in that they will theoretically bind to the SCN peptide and this suggests that anything similar to the AS35CN or AS53CN peptides would bind to the SCN peptide. The SCN peptide corresponds to the SNC peptide in reverse. If the SNC was a 14 amino acid peptide the N-C direction numbering would be 1-14, therefore the C-N numbering would be 14-1. Blast searches always number sequences 1-14, which means the results for AS35CN and AS53CN peptides searched using BLAST will always be numbered in the wrong direction and the numbers will need to be converted.

The easiest way to do this is to create a table with the numbers from start to finish of the sequence in a column in ascending order, in this case 1-14, and a second column with the numbers in descending order – starting from the total number of amino acids of the SNC protein, in this case 14-1:

	BLAST Query numbering	NC numbering
	AS35CN/AS53CN	
1 <sup>st</sup> residue	1	14
	2	13
	3	12
	4	11
	5	10
	6	9
	7	8
	8	7
	9	6
	10	5
	11	4
	12	3
	13	2
Last residue	14	1



In the BLAST search Alignments, the Query corresponds to the protein searched and the Sbjct corresponds to the named protein that is similar. If an alignment were found where the Query region on the BLAST search results from an AS35CN and AS53CN peptide was 4-9 (circled in red) using the table above the NC numbering would be 11-6 (circled in blue). The 11-6 region is therefore the 6-11 region of the SNC, which is the protein originally used to generate the antisense peptides.

(6c) For the purposes of illustrating how to analyse the data the Amyloid-ß 1-40 peptide (Aß) will be used as an example:

Coding mRNA for Aß	GATGCAGAATTCCGACATGACTCAGGATATGAAGTTCATCAT CAAAAATTGGTGTTCTTTGCAGAAGATGTGGGTTCAAACAAA
SNC-Aß	DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVV
SCN-Aß	VVGGVMLGIIAGKNSGVDEAFFVLKQHHVEYGSDHRFEAD
AS35NC-Aß	LRLKAVLSPILQVVVFNHKKRLLHPSLFPRXXPEYHPPQQ
AS35CN-Aß	QQPPHYEPXXRPFLSPHLLRKKHNFVVVQLIPSLVAKLRL
AS53NC-Aß	ICFESMVXSIFNMMLFQHEKCFIHTXVFTCDNSEHHATND
AS53CN-Aß	DNTAHHESNDCTFVXTHIFCKEHQFLMMNFISXVMSEFCI

- (6d) Running a BLAST search, as described in section 5 above (pages 17-19) for the AS53CN-AB peptide gave the set of results:
- (i) Descriptions:

Job Title	AS53CN-Abeta	Filter Results
RID	G70W30DT016 Search expires on 07-08 04:03 am Download All ↔	
Program	BLASTP 😮 Citation 💙	Organism only top 20 will appear
Database	refseq_protein <u>See details</u> ♥	Type common name, binomial, taxid or group name
Query ID	lcl Query_11240	+ Add organism
Description	None	Percent Identity E value Query Coverage
Molecule type	amino acid	
Query Length	40	
Other reports	Distance tree of results Multiple alignment MSA viewer	Filter
Descriptions	Graphic Summary Alignments Taxonomy	
Sequences	producing significant alignments	Download Y Manage Columns Y Show 20000 V
🗹 select all	3 sequences selected	GenPept Graphics Distance tree of results Multiple alignment
	Description	Max         Total         Query         E         Per.           Score         Score         Cover         value         Ident         Accession
3'-5' RNA h	elicase YTHDC2 isoform 3 [Homo sapiens]	20.4 20.4 65% 114 34.62% <u>NP_001332905.1</u>
3'-5' RNA h	elicase YTHDC2 isoform 2 [Homo sapiens]	20.4 20.4 65% 114 34.62% <u>NP_001332904.1</u>
3'-5' RNA h	elicase YTHDC2 isoform 1 [Homo sapiens]	20.4 20.4 65% 114 34.62% <u>NP_073739.3</u>



### (ii) Graphic Summary:

Job Title	AS53CN-Abeta	Filter Results
RID	G70W30DT016 Search expires on 07-08 04:03 am Download All 💙	
Program	BLASTP 3 Citation >	Organism only top 20 will appear exclude
Database	refseq_protein See details 💙	Type common name, binomial, taxid or group name
Query ID	lcl Query_11240	+ Add organism
Description	None	Percent Identity E value Query Coverage
Molecule type	amino acid	to to to
Query Length	40	
Other reports	Distance tree of results Multiple alignment MSA viewer <b>@</b>	Filter Reset
Descriptions	Graphic Summary Alignments Taxonomy	
hover to see to	he title 🗼 click to show alignments 🗹 Show Conserved Domains	Alignment Scores ■<40 ■40-50 □50-80 ■80-200 ■>=200
3 sequences sel	ected 😯 No putative conserved	domains have been detected
	Distribution of the top 3 Blast	Hits on 3 subject sequences
	1 8 16	40

### (iii) Alignments:

ment view Pairwise	Restore defaults	Downloa
uences selected 🕜		
Download - GenPept Graphics		▼ <u>Next</u> ▲ <u>Previous</u> ≪ <u>Descriptions</u>
3'-5' RNA helicase YTHDC2 isoform 3 [Hon Sequence ID: <u>NP_001332905.1</u> Length: <b>1130</b> Nu	no sapiens] mber of Matches: 1	
Bange 1: 398 to 423         GenPept Graphics           Score         Expect         Method           20.4 bits(41)         114         Composition-based stats           Query         9         NDCTFVXTHIFCKEH0FLMMNFISXV ND           FV         KE         F           Sbjct         398         NDVFVIDSGKVKEKSFALLFYML	▼ Next Match ▲ Previous Match           Identities         Positives         Gaps           9/26(35%)         13/26(50%)         0/26(0%)           34         423	Related Information Gene - associated gene details Genome Data Viewer - aligned genomic context
Download      GenPept Graphics 3'-5' RNA helicase YTHDC2 isoform 2 [Hon	no sapiens]	▼ <u>Next</u> ▲ <u>Previous</u> ≪ <u>Descriptions</u>
Lownload      GenPept Graphics      GenPept Graphics      Graphics	to sapiens] mber of Matches: 1 Vext Match & Previous Match Identities Positives Gaps 9/26(35%) 13/26(50%) 0/26(0%)	▼ Next ▲ Previous ≪Description: Related Information Gene - associated gene details Genome Data Viewer - aligned
Lownload         GenPept Graphics           3'-5' RNA helicase YTHDC2 isoform 2 [Hon Sequence ID: NP_001332904.1 Length: 1268 Nu Range 1: 536 to 561 GenPept Graphics           Score         Expect Method           20.4 bits(41)         114         Composition-based stats           Query 9         NDCTFVXTHIFCKEHQFLMWFISXV ND FV         KE F +NF++ + Sbjct 536	to sapiens] mber of Matches: 1 Wext Match & Previous Match Identities Positives Gaps 9/26(35%) 13/26(50%) 0/26(0%) 34 561	▼ Next ▲ Previous ≪Description: Related Information Gene - associated gene details Genome Data Viewer - aligned genomic context
L Download ~       GenPept Graphics         3'-5' RNA helicase YTHDC2 isoform 2 [Hon         Sequence ID: NP_001332904.1       Length: 1268 Nu         Range 1: 536 to 561 GenPept Graphics         Score       Expect Method         20.4 bits(41)       114 Composition-based stats         Duery 9       NDCTFVXTHIFCKEHOFLMMNFISXV ND FV         ND FV       KE F +NF++         Sbjct 536       NDVVFVIDSGKVKEKSFDALNFVTML         Lownload ~       GenPept Graphics	no sapiens] mber of Matches: 1 <u>Vext Match</u> A Previous Match <u>Jdentities</u> Positives Gaps 9/26(35%) 13/26(50%) 0/26(0%) 34 561	✓ Next ▲ Previous ≪Descriptions
Download ~ GenPept Graphics 3'-5' RNA helicase YTHDC2 isoform 2 [Hon Sequence ID: NP_001332904.1 Length: 1268 Nu Range 1: 536 to 561 GenPept Graphics Score Expect Method 20.4 bits(41) 114 Composition-based stats Query 9 NDCTFVXTHIFCKEHQFLMMNFISXV ND FV Sbjct 536 NDVVFVIDSGKVKEKSFDALNFVTML  Download ~ GenPept Graphics 3'-5' RNA helicase YTHDC2 isoform 1 [Hon Sequence ID: NP_073739.3 Length: 1430 Numbe	no sapiens] mber of Matches: 1	Next ▲ Previous ≪Description:     Related Information     Gene - associated gene details     Genome Data Viewer - aligned     genomic context      Next ▲ Previous ≪Description:
Download      GenPept Graphics 3'-5' RNA helicase YTHDC2 isoform 2 [Hon Sequence ID: NP_001332904.1 Length: 1268 Nu Range 1: 536 to 561 GenPept Graphics Score Expect Method 20.4 bits(41) 114 Composition-based stats Query 9 NDCTFVXTHIFCKEHQFLMMNFISXV ND FV KE F +NF++ + Sbjct 536 NDVVFVIDSGKVKEKSFDALNFVTML Download      GenPept Graphics 3'-5' RNA helicase YTHDC2 isoform 1 [Hon Sequence ID: NP_073739.3 Length: 1430 Numbe Range 1: 698 to 723 GenPept Graphics Score Expect Method	to sapiens] mber of Matches: 1	✓ Next ▲ Previous ≪Description

(iv) Text File, with the key elements that should be extracted highlighted in red are best put into a table format and saved:

```
RID: G70W30DT016
Job Title:AS53CN-Abeta
Program: BLASTP
Database: refseq_protein NCBI Protein Reference Sequences
Query #1: Query ID: lcl|Query_11240 Length: 40
Sequences producing significant alignments:

        Total Query
        E
        Per.

        Score
        cover
        Value
        Ident

        20.4
        65%
        114
        34.62

        20.4
        65%
        114
        34.62

                                                                                                                  Max
                                                                                                                  Score
20.4
3'-5' RNA helicase YTHDC2 isoform 3 [Homo sapiens]
3'-5' RNA helicase YTHDC2 isoform 2 [Homo sapiens]
3'-5' RNA helicase YTHDC2 isoform 1 [Homo sapiens]
                                                                                                                  20.4
                                                                                                                  20.4
                                                                                                                             20.4 65%
Alignments:
>3'-5' RNA helicase YTHDC2 isoform 3 [Homo sapiens]
Sequence ID: NP 001332905.1 Length: 1130
Range 1: 398 to 423
Score:20.4 bits(41), Expect:114,
Method:Composition-based stats.,
Identities:9/26(35%), Positives:13/26(50%), Gaps:0/26(0%)
                    NDCTFVXTHIFCKEHQFLMMNFISXV 34
Query 9
ND FV KE F +NF++ +
Sbjct 398 NDVVFVIDSGKVKEKSFDALNFVTML 423
>3'-5' RNA helicase YTHDC2 isoform 2 [Homo sapiens]
Sequence ID: NP_001332904.1 Length: 1268
Range 1: 536 to 561
Score:20.4 bits(41), Expect:114,
Method:Composition-based stats.,
Identities:9/26(35%), Positives:13/26(50%), Gaps:0/26(0%)
Query 9 NDCTFVXTHIFCKEHQFLMMNFISXV 34
                     ND FV
                                         KE
                                                    +NF++
Sbjct 536 NDVVFVIDSGKVKEKSFDALNFVTML 561
>3'-5' RNA helicase YTHDC2 isoform 1 [Homo sapiens]
Sequence ID: NP_073739.3 Length: 1430
Range 1: 698 to 723
Score:20.4 bits(41), Expect:114,
Method:Composition-based stats.,
Identities:9/26(35%), Positives:13/26(50%), Gaps:0/26(0%)
Query 9
                    NDCTFVXTHIFCKEHQFLMMNFISXV 34
                     ND FV
                                         KE
                                                      +NF++
Sbjct 698 NDVVFVIDSGKVKEKSFDALNFVTML 723
```

#### (6e) Any results with a % gaps greater than 0 should be reviewed as these may not be compatible with protein-protein binding.

Gaps are where a space is introduced between two amino acids in either the Query or Sbjct sequence to achieve better alignment, they are indicated with one or more "-" in between two amino acid residues (circled in red).

```
>melanoma-associated antigen 10 [Homo sapiens]
Sequence ID: NP 066386.3 Length: 369
Range 1: 170 to 199
Score:18.5 bits(36), Expect:12367,
Method: Compositional matrix adjust.,
Identities:11/43(26%), Positives:16/43(37%), Gaps:17/43(39%)
           DHYP ---- KDKECTXTEDVKRIRLSLSKNIYXFNXKETDNTGH
      22
                                                       60
Query
           DH+P
                   +
                      EC
                                       ++
                                          + KE D TGH
           DHFPLLFSEASECMLI
Sbjct
      170
                                                       199
```

\*

 Ident
 Accession

 34.62
 NP\_001332905.1

 34.62
 NP\_001332904.1

 34.62
 NP\_073739.3

114

The key is that gaps are not present in protein sequences but are artifacts of the BLAST search methods. This means that gaps could create an artificial alignment that in reality would not lead to a protein-protein interaction. The higher the % gaps the more likely the result would not be compatible with binding.

The length of ungapped alignment within the BLAST results needs to be sufficient to suggest possible binding and as such needs to be accounted for when taking results with gaps forward. In this example the selection circled in red could still potentially be used:

```
>peptidyl-prolyl cis-trans isomerase-like 4 [Homo sapiens]
Sequence ID: NP 624311.1 Length: 492
Range 1: 126 to 156
Score:19.2 bits(38), Expect:7508,
Method:Compositional matrix adjust.,
Identities:8/32(25%), Positives:18/32(56%), Gaps:1/32(3%)
     173
            FXYINKSLSLFIRKVDETXTCEKDKPYHDLTV
                                               204
Query
               + + + + I+K++ET
            F
                                 +
                                      PY D+ +
Sbjct
            FGEVTEGMDI-IKKINETFVDKDFVPYQDIRI
       126
                                               156
```

The resultant alignment would be used along with recalculated % identity and % positives for the shorter segment. The E value and the statistical analysis would no longer be valid.

```
>peptidyl-prolyl cis-trans isomerase-like 4 [Homo sapiens]
Sequence ID: NP 624311.1 Length: 492
Range 1: 136 to 156
Score:19.2 bits(38), Expect:7508,
Method: Compositional matrix adjust
Identities: 7/21(33%) Positives 13/21(62%)
                                             Gaps: 0/21 (0%
       184
            IRKVDETXTCEKDKPYHDLTV
                                    204
Query
            I+K++ET
                       +
                           PY D+ +
            IKKINETFVDKDFVPYQDIRI
       136
Sbjct
                                    156
```

For the purpose of antisense binding 0% gaps is best, and generally the majority of sequences with >0% gaps are discarded.

(6f) The easiest method to extract the data is to open the txt file in Word, select the txt file and then right-click to bring up the options. Select Open With (circled in red) and then select Word (circled in red), just click OK for any conversion settings are suggested and then save document as Protein Name Results.docx. Keep the original txt file as it contains the aligned sequences that are needed in section 7c (page 32 below) and is a record of the results of the BLAST search:



^	Name	^	Date modified	Туре	Size
	📙 My Stationer	/	09/05/2016 00:15	File folder	
	G70W30DT01	6-Alignment <b>Open</b> Print Edit	<u>06/07/2020 23:33</u>	Text Document	2 KB
		<ul> <li>Share with Skype</li> <li>Move to OneDrive</li> <li>Send with Transfer</li> <li>Scan with Windows Defender</li> <li>Share</li> </ul>	er		
		Open with Parallels Toolbox Open on Mac Show in Finder Share on	> > >	Notepad Word WordPad Search the Microsoft S Choose another app	tore

Once saved delete the initial parts of the file down to alignments (highlighted in blue):

S Word File Edit View Insert Format Tools Table	le Window Help				04439	100% (42) Fri 21:13	
🛢 🥥 💿 AutoSave 🚥 🏹 🏠 🕄 ්ර ් 🖨 🗸		Abeta results — Saved –					۹ 🌌
Home Insert Draw Design Layout References Mailin	ngs Review View EndNote X9 Acr	robat 💡 Tell me				년 Share	Comments
Paste $\checkmark$ Courier New $\checkmark$ 10.5 $\checkmark$ A <sup>*</sup> A <sup>*</sup> Aa $\checkmark$ Aa $\checkmark$ Ac ${\models}$ $\checkmark$ B I $\checkmark$ $\checkmark$ ab x, x <sup>i</sup> $\land$ $\land$ $\checkmark$ Aa $\checkmark$ $\checkmark$	· ⊞ • ₩ •   型 型   \$↓   ¶   , 돌 Ξ ≡   1≣ •   ◊ • ⊞ •	Aa8bCcDdEe Aa8bCcDdEe Aa8bCcDc Aa8 Normal No Spacing Heading 1 He	ading 2 Title AaBbCcOdEc Subtitie	Ao8bCcDdEe Ao8bCcDdEe Subtle Emph Emphasis	AgibCcDdEe AaBbCcDdEe Intense Emp Strong	AoBbCcDdEe , J	Dictate
		Englis Parily Barrages	1 10	$\eta_{i} = 1^{-1}, \ \eta_{i} = 1^{-1}, \ j$	5	s 17	18
RID: G70W30 Job Title:A Program: BL Database: r Query #1: (	DT016 S53CN-Abeta ASTP efseq_protein NC Query ID: 1cl Qu	BI Protein Refe ery_11240 Lengt	erence Seque ch: 40	nces			
Sequences p	roducing signifi	cant alignments	3:				
		2			Max		
Total Ouerv	E Per.						
Description					Score		
Score cover	Value Ident Ac	cession					
3'-5' BNA b	elicase YTHDC2 i	soform 3 [Homo	sapiensl		20.4		
20.4 65%	114 34.62 NP	001332905.1	odbroupl		2011		
3'-5' RNA he	elicase YTHDC2 i	soform 2 [Homo	sapiens]		20.4		
20.4 65% 3'-5' RNA he 20.4 65%	elicase YTHDC2 i 114 34.62 NP	soform 1 [Homo 073739.3	sapiens]		20.4		
Alignments:		_					



Then select complete text and using the Table option convert text to table (highlighted in blue), make sure converts to one column table:

🔹 Word File Edit View Insert Format Tools 🕇	able Window Help	O 👯 🌰 🖇 🎅 100% 🖗	Fri 21:19	■ ⅲ ≔
🔹 😳 🔹 AutoSave 🚥 🎧 🏠 🎧 🖓 🖓 🖓 🖓	Draw Table	Abeta results Saving		۹ 🛓
Home Insert Draw Design Layout References	Insert >	dNote X9 Acrobat 🖓 Tell me	년 Share	Comments
$\begin{array}{c c} & & & \\ & & \\ & & \\ & \\ & \\ & \\ & \\ & $	Select  Merge Cells Split Cells	1 T AMERGERIE	Ee , Styles Pane	Dictate
* 1. * *	AutoFit and Distribute	$ \{ \frac{1}{2}, 1, \frac{1}{2}, \frac{1}$	U .	18
>3'-5' RN Sequence Bange 1:	Repeat Header Rows Convert Sort Formula View Gridlines	Convertion to Table. )rm 3 [Homo sapiens] Converting th: 1130		

Then add a row at the top and type in Name of protein similar to AS53CNprotein name in that row. Then select add columns to the right and add a total of seven columns to give an eight-column table. Then put the headings in table below:

S Word File Edit View Insert Format Tools	Table Window Help		0 7 🖷 🕯 🤅	100% 🚮 Fri 21:28 🗐 🛗 😑
● ● ● AutoSave @ へ 合 忌 ち ~ み 告 =	Draw Table		Abeta results — Saving	۹ 🛓
Home Insert Draw Design Layout References	Insert >	Table	ble Design Layout 🛇 Tell me	🖻 Share 💭 Comments
✓ Header Row ✓ First Column Total Row Last Column	Select  Merge Cells	Columns to the Left Columns to the Right	· · · · · · · · · · · · · · · · · · ·	<b>2</b> • <b>1</b> • <b>1</b>
Banded Rows Banded Columns     Split Cells     Split Cells	Split Cells	Rows Above	Shiding Border 1/2 pt 0	Pen Borders Border Colour Painter
2 1 Split Table	Split Table	Rows Below	"你看你们就是你不能是你开心想你们的现在,你能知道你想你们的梦心下的感觉不	21日 17 18
	AutoFit and Distribute Repeat Header Rows	Cells		
N	Convert Sort Formula View Gridlines			
	Table Properties			
*			<i>c c c c c c c c c c</i>	_
>3'-5' R	NA helicase	YTHDC2 1se	oform 3 [Homo sapiens]	

In the table delete the > symbol at the start of the protein name. Then copy the NP number (circled in red) and paste in column 2; the length number (circled in red) and paste in column 3; the range numbers (circled in red) and paste in column 4; the Query start and end numbers (circled in red) and paste in column 5; the % ID (circled in red) and paste in column 6; the % +ve (circled in red) and paste in column 8. This will give a table like this:

Name of protein similar to AS53CN-Aß	Protein ID	Size	Residues	Query residues	% ID	% +ve	% Gaps
3'-5' RNA helicase YTHDC2 isoform 3 [Homo sapiens]	NP_001332905.1	1130	398-423	9-34	35	50	0
Sequence ID: NP 001332905.1 Length: 1130		1					
Range 1: 398 to 423							)
Score:20.4 bits(41), Expect:114,							
Method:Composition-based stats.,							
Identities:9/26 358, Positives:13/26 508, Gaps:0/26 08							
		1					
Query 9 NDCTFVXTHIFCKEHQFLMMNFISXV (34)							
ND FV KE F +NF++ +							
Sbjct 398 NDVVFVIDSGKVKEKSFDALNFVTML 423							
							-
>3'-5' RNA helicase YTHDC2 isoform 2 [Homo							

Delete the highlighted rows in blue above, down to the next protein name and repeat the same process above, until all of the proteins have been processed to give a table like this:



Name of protein similar to AS53CN-Aß	Protein ID	Size	Residues	Query residues	% ID	% +ve	% Gaps
3'-5' RNA helicase YTHDC2 isoform 1	NP_073739.3	1430	698-723	9-34	35	50	0
3'-5' RNA helicase YTHDC2 isoform 2	NP_001332904.1	1268	536-561	9-34	35	50	0
3'-5' RNA helicase YTHDC2 isoform 3	NP_001332905.1	1130	398-423	9-34	35	50	0

- (6g) Make sure that none of the proteins have a % Gaps > 0 as this data should have been discarded in section 6e above (page 23), if there are any rows with a % Gaps > 0 they should be reviewed as described.
- (6h) If the antisense peptide sequence screened is an AS35NC or AS53NC the Query residue numbers correspond directly to the SNC residue numbering (see 6a, page 20 above). However, in this case the AS53CN peptide was used, which means the numbers need to be converted as described in 6b above (pages 20-21). Using a table with 1-40 ascending in Column A and 40-1 descending in Column B (40 is the number of residues in the SNC-Aß [see Table 6c, page 21 above], can be determined using word count for the sequence):



The 9-34 Query residues (circled in red) correspond to 32-7 (circled in blue). The Query residues with the revised numbering corresponding to the Aß residues 32-7 and the results table can be modified to show this:

#### Table 6h

Human Protein that theoretically binds human Aß	Protein ID	Size	Residues	Aß residues	% ID	% +ve	% Gaps
3'-5' RNA helicase YTHDC2 isoform 1	NP_073739.3	1430	698-723	32-7	35	50	0
3'-5' RNA helicase YTHDC2 isoform 2	NP_001332904.1	1268	536-561	32-7	35	50	0
3'-5' RNA helicase YTHDC2 isoform 3	NP_001332905.1	1130	398-423	32-7	35	50	0

Thus 3'-5' RNA helicase YTHDC2 isoform 1 residues 698-723 theoretically binds to AB residues 7-32.

(6i) Within the results files there are other potentially useful pieces of information that can be used when writing up the results. There are also other useful sets of data that can be downloaded from the Blast search while online:

Job Title	AS53CN-Abeta	Filter Results										
RID	G70W30DT016 Search expires on 07-08 04:03 am Download All V	Organism only top 20 will appear										
Program	BLASIP 😮 <u>Citation</u> 🗸	Type common name, binomial, taxid or group name										
Database	refseq_protein See details	- i jpe common name	, on on our of the other	or Broad in								
Query ID	lcl Query_11240	+ Add organism										
Description	None	Percent Identity	E value		Query Cov	erage						
Molecule type	amino acid	to		· 0	] [	·•						
Query Length	40											
Other reports	Distance tree of results Multiple alignment MSA viewer				Filter	Reset						
Sequences												
	producing significant alignments	Downloa	d 🌱 Manag	e Columns	✓ Show	20000 💙 🔞						
select all	producing significant alignments 3 sequences selected	<b>Downloa</b> Genf	d 🌱 Manag Pept <u>Graphics</u>	e Columns	Show Show	20000 💙 🕜						
select all	producing significant alignments 3 sequences selected Description	Downloa Genf	d <sup>V</sup> Manag <u>Pept Graphics</u> Max Score	e Columns Distance tre Total Query Score Cover	<ul> <li>✓ Show</li> <li>← of results</li> <li>M</li> <li>E</li> <li>Per.</li> <li>value</li> <li>Ident</li> </ul>	20000 ♥ ② ultiple alignment Accession						
Select all	producing significant alignments 3 sequences selected Description helicase YTHDC2 isoform 3 [Homo sapiens]	Downloa Genf	d × Manag Pept Graphics Max Score 20.4	Total Query Score Cover 20.4 65%	Show Show Ee of results M E Per. value Ident 114 34.62%	20000 V ? ultiple alignment Accession NP_001332905.1						
<ul> <li>select all</li> <li>3'-5' RNA</li> <li>3'-5' RNA</li> </ul>	producing significant alignments 3 sequences selected Description helicase YTHDC2 isoform 3 [Homo sapiens] helicase YTHDC2 isoform 2 [Homo sapiens]	Downloa Genf	d Y Manag Pept Graphics Max Score 20.4 20.4	Distance tree Distance tree Total Query Score Cover 20.4 65%	Show Show E of results M E Per. Ident 114 34.62% 114 34.62%	20000 ♥ eventorial eventorial eventoris eventorial eventorial eventorial eventorial eventorial e						

The distance tree of results (click link circled in red) can be useful and also the MSA viewer (again click link circled in red). From these links the files can be downloaded as pdfs. For the Distance tree of results: after clicking on link go to Tools (circled in red), then Download (circled in red), then select PDF:



NIH U.S. National Library of Medicine NCBI	National Center for Biotechnology Information	Sign in to NCBI
BLAST <sup>®</sup>		
E	Blast Tree View	Home Recent Results Saved Strategies Help
This tree was produced	using BLAST pairwise alignments. more	
Reset Tree		
BLAST RID GHFH6U9Y016	Query ID lcl Query_62681	Database refseq_protein
Tree method Max Seq Difference	Distance Sequence Label	
Fast Minimum Evolution 📀 😔 0.85 😒 😔	Grishin (protein) 😧 😡 Sequence Title (if a 😒 😡	
	Mouse over an internal node for	or a subtree or alignment. Click on tree label to select sequence to download Hide legend
Find:		Tools 🗾 💽 Upload 🛛 🧟 🖓 🕶 Label color map
		Download     ASN text file     rom type material
		Layout  ASN binary file t names color map
		Sort  Newick file unknown
		NEXUS file primates
		Zoom behavior P PDF file

This will give the following where the unnamed protein product is the AS53CN-AB and the results indicate how closely related the sequences are:



Multiple sequence alignment (from MSA viewer) results: after clicking on link go to Download (circled in blue), then select Printer-Friendly PDF/SVG, which will bring up a second box:

••• <>				0	6	2							ncbi.r	ılm.nih	.gov						Ċ						0	₫	Ø
			NCBI E	Blast:AS	653CN-	Abeta													N	BI Mul	tiple Se	quence	Alignm	ent View	ver 1.15.0				+
S NCBI	Home P	ubMed	l GenB	ank	BLAS	T														м	ultip	ole Se	que	nce /	lignm	ient \	liewer	1.15	.0 了
Alignment																											Link To V	iew   Fe	edback
9 10.1		13	.1 13.1 14.	1 14.1 1	15.1 15.	1 16.1 1	6.1 17.1	1 17.1	18.1 18	1 19.1	19.1 2	0.1 20.	1 21.1	21.1 22	.1 22.1	23.1 2	3.1 24.	1											34
4								1											4									4	PVIS-up
9 - 34	(26r shown)				13		-	+ 41	6			-		1	- (							X Too	ls •   4	🗘 Rows	Dov	vnload •	Color	ring •   /	27.
Sequence ID	Start	9	10 11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30 3	31 3	2 33		FASTA Ali	ignment		
Query_62681	\$ 9	N	D C	т	F	v	x	T	н	I	F	с	ĸ	Ε	н	Q	F	L	м	м	N	F	I	s x		Printer-Fr	riendly PDF/	SVG	
NP 001332905.1	¥ 398	1.0	• V	٧	- 2	- 200	I	D	S	G	K	V	- <u>6</u> -		K	S	- *	D	λ	L	<u> </u>	1	V :	T M		12.3	Homo sap	liens	
NP 001332904.1 NP 073739.3	* 536 * 698		· v	v		(0)	I	D	s	G	ĸ	v	-		K	S		D	A	L			v	T M	L	723	Homo sap Homo sap	viens	-

In the second box the Possible range formats box (circled in blue) should have the 9-34 for the fragment of AS53CN-AB that is aligned, if there are other alignments may have a range covering all of them. Select PDF (circled in blue) and click Download (circled in blue):

PROTEIN: 9 - 34 (26r shown) - anchor Query_62681	Download Image	💉 🔹 Rows shown: 4/4
	Enter Sequence Range Possible range formats include 10K-20K, 10:20, 20000-30000, 5 to 515, 1246	
	9:34	
	Simplified color shading (allows greater compatibility with image editors)	
	File type: (PDF and SVG contain vector graphics for high quality images)	
	@ PDF	
	Preview Download Cancel	
	Ready	

This will give the following where the query is AS53CN-AB and the alignments are for the proteins from table at end of 6f above (page 26), note in this case the numbering is AS53CN-AB sequence numbering and not the corrected AB SNC sense numbering:



(6j) Within the saved Alignments files there are a number of components that are also useful: NEW TABLE HERE – Alignments, Expect, plus bits from 6h



- (6k) The Expect (E-value) is a statistical indication of similarity between the Query and Sbjct, the smaller the E-value, the better the match. Within the extracted data files created in above the Identities, Positives and Gaps data has been extracted.
- (i) The Identities is the % of amino acids in the Query sequence that are identical to those in the Sbjct sequence these are shown in the middle row between the Query and Sbjct rows as letters.
- (ii) The number of Positives includes the identical amino acids. In the context of alignments displayed in the BLAST output, the Positives are those non-identical substitutions that receive a Positives score in the underlying scoring matrix, BLOSUM62 by default. Most often, Positives indicate a conservative substitution or substitutions that are often observed in related proteins. In effect the Positives is the % of amino acids in the Query sequence that are similar to those in the Sbjct sequence these are shown in the middle row between the Query and Sbjct rows as a + symbol. Similarity of amino acids in this case is often based on similar structural features in the R group of the amino acids.
- (iii) Gaps are where a space is introduced between two amino acids in either the Query or Sbjct sequence to achieve better alignment, for the purpose of antisense binding a 0% Gaps is essential, and sequence alignments with Gaps are normally discarded (see 6e above, page 23).



## 7: Molecular recognition analysis

(7a) The basis for the molecular recognition theory is that amino acids encoded by the sense strand of DNA will bind the corresponding amino acids encoded by the antisense strand. Using the BLAST search with a scoring matrix to identify similarities has limitations and may identify sequences that do not have a good potential to actually bind the target protein. The % identity from the BLAST search data represent identity with the antisense peptide and is much more useful than the % positives, which may give an inflated indication of potential binding.

The Molecular Recognition (MR) scoring is an alternative system that determines the potential of the target protein residues binding to the identified interacting protein residues based on antisense/sense interactions. The method determines the quantity of antisense/sense pairs within the identified target/interacting protein regions identified by the BLAST search. Using aligned target/interacting protein regions and the following table a score for the potential interaction is calculated. For example, if the first residue of the target is an A, then a score of 1 would be given if the first residue of the binding protein was either a C, G, R or S. If the first residue of the binding protein was any other amino acid, a score of 0 would be given. Then the process is repeated for each residue of the target/interacting protein regions:

Target Residue		Bindin	g Protein R	Residue		MR Score
Ala (A)	Cys (C)	Gly ( <b>G</b> )	Arg (R)	Ser (S)		+1
Cys ( <b>C</b> )	Ala (A)	Thr (T)				+1
Asp (D)	Ile (I)	Leu (L)	Val (V)			+1
Glu ( <b>E</b> )	Phe (F)	Leu (L)				+1
Phe (F)	Glu (E)	Lys (K)				+1
Gly ( <b>G</b> )	Ala (A)	Pro (P)	Ser (S)	Thr (T)		+1
His (H)	Met (M)	Val ( <b>V</b> )				+1
Ile (I)	Asp (D)	Asn (N)	Tyr (Y)			+1
Leu (L)	Asp (D)	Glu (E)	Lys (K)	Asn (N)	Gln (Q)	+1
Lys (K)	Leu (L)	Phe (F)				+1
Met (M)	His (H)	Tyr (Y)				+1
Asn (N)	Ile (I)	Leu (L)	Val (V)			+1
Pro (P)	Gly (G)	Arg (R)	Trp (W)			+1
Gln ( <b>Q</b> )	Leu (L)	Val (V)				+1
Arg (R)	Ala (A)	Pro (P)	Ser (S)	Thr (T)		+1
Ser (S)	Ala (A)	Gly ( <b>G</b> )	Arg (R)	Ser (S)	Thr (T)	+1
Thr (T)	Cys (C)	Gly ( <b>G</b> )	Arg (R)	Ser (S)	Trp (W)	+1
Val (V)	Asp (D)	His (H)	Asn (N)	Gln (Q)	Tyr (Y)	+1
Trp (W)	Pro (P)	Thr (T)	1998 - To			+1
Tyr (Y)	Ile (I)	Met (M)	Val (V)			+1

(7b) Table for Molecular Recognition Scoring:



(7c) For AS35NC and AS53NC sequences the Query residues from a BLAST search should align with the SNC residues for comparison. For AS35CN and AS53CN sequences the Query residues from a BLAST search should align with the SCN residues for comparison. The key is to use the Query residue numbers to identify the SNC or SCN residues to use.

Copy the alignments from the txt file (see section 6d, (iv) above, page 23) and paste into a word document (for ease always use a monospaced font such as Courier New). Save the document as a word file.

Open the Python results file for the target protein and then copy the SNC or SCN residues corresponding to the Query residues for each protein and paste then above the Query – see examples below. Then align the SNC or SNC residues with the Query residues. The SNC or SCN should then be aligned with the Sbjct residues.

Using Table 7b above (page 31) the Molecular Recognition (MR) score for the protein can be determined, using the SNC or SCN residues as the Target protein residues in the table and the Sbjct residues as the Binding Protein residues.

The total score is best expressed as a % of the total number of residues in the target sequence. The higher this % the more likely a binding interaction is likely to occur. This number will always be equal to or greater than the % identity score from the BLAST search.

(7d) The following efficient process of calculating the MR score for alignments obtained from antisense peptide Blast results using AS35CN, AS53CN, AS35NC, AS53NC employs Microsoft Excel to eliminate human error: As an example, using data from the AS53CN-Aß Blast results below, plus the Aß SCN sequence detailed in section 6f (pages 24-27 above) the MR score will be determined for the BLAST alignment that predicted an interaction between Aß and the 3'-5' RNA helicase YTHDC2 isoform 1:

```
>3'-5' RNA helicase YTHDC2 isoform 1 [Homo sapiens]
Sequence ID: NP 073739.3 Length: 1430
Range 1: 698 to 723
Score:20.4 bits(41), Expect:114,
Method:Composition-based stats.,
Identities: 9/26(35%), Positives: 13/26(50%), Gaps: 0/26(0%)
            NDCTFVXTHIFCKEHOFLMMNFISXV
Query
      9
                                         34
            ND
                FV
                        KE
                            F
                                +NF++ +
Sbjct
       698
            NDVVFVIDSGKVKEKSFDALNFVTML
                                         723
```



(i) The whole SCN sequence was copied and pasted from the python outputs file into a single cell in an excel spreadsheet.

```
Input name: Aß
Input coding mRNA:
GATGCAGAATTCCGACATGACTCAGGATATGAAGTTCATCATCAAAAATTGGTGTTCTTTGC
AGAAGATGTGGGTTCAAACAAAGGTGCAATCATTGGACTCATGGTGGGCGGTGTTGTC
SNC - AB = DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVV
SCN - AB = VVGGVMLGIIAGKNSGVDEAFFVLKQHHVEYGSDHRFEAD
        AutoSave OFF A B P 2 C ...
                                                                      Book1
• •
Home
                   Page Layout
                                                        Q Tell me
      Insert
                             Formulas
                                      Data
                                            Review
                                                   View
             Draw
 Ô
      X
                             ~ A^ A
                                        ΞΞ=
            Courier New
                        v 12
                                                 87 ~
                                                         ab Wrap Text ∨
                                                                        Genera
     [] ·
 Paste
            B I U → H → 🖉 → <u>A</u> →
                                        🔁 Merge & Centre 🗸
                                                                       3
            √ fx VVGGVMLGIIAGKNSGVDEAFFVLKQHHVEYGSDHRFEAD
A1
         ×
    A
           В
                  С
                         D
                                E
                                                                         Κ
1 VVGGVMLGIIAGKNSGVDEAFFVLKQHHVEYGSDHRFEAD
```

(ii) In cell B1, the following formula was then entered: =MID(\$A1, COLUMNS(\$A\$1:A\$1), 1), where the first amino acid of the sequence appeared. To separate each amino acid of the sequence into one cell per column, cell B1 was dragged to the right across row one until the last amino acid had a position.

🖲 🔵 🔵 Auto	oSave OFF	6 B B	ッ・ペ …	-					Book1
Home Insert	Draw P	age Layout	Formulas	Data	Review	View	♀ Tell me		
	Tahoma	v 11	✓ A <sup>^</sup> A <sup>×</sup>	Ξ		87 v	ce Wrap Te	xt v	Genera
Paste 🞸	ΒΙυ	•   표 •	<u>∽ ~ A</u> ~	Ξ	≡ ≡	<u>∓</u> ≡ <u>→</u> Ξ	Merge &	Centre 🗸	<b>(</b>
B1 * ×	<i>√ f</i> x =	MID(\$A1, COL	JMNS(\$A\$1:A\$:	1), 1)					
A 1 VVGGVMLGI V	B C	D	E	F	G	Н	- 1-	J	К
🖲 🔵 🔵 Auto	oSave OFF	6 B B	ッ・C …						Book1
Home Insert	Draw P	age Layout	Formulas	Data	Review	View	♀ Tell me		
	Tahoma	~ 11	✓ A <sup>^</sup> A <sup>×</sup>	Ξ		87 ~	ee Wrap Te	xt v	Genera
Paste	ΒΙυ	•   🗄 •	<u>⊘</u> ~ <u>A</u> ~	Ξ	ΞΞΙ	<u>₹</u> Ξ <u>₹</u> Ξ	Merge &	Centre 🗸	<b>()</b> ~
B1 🛔 🗙	<i>√ f</i> x =	MID(\$A1, COL	JMNS(\$A\$1:A\$:	1), 1)					
AG	A11 A		AV	A1	414		40	4.0	40
AG	AH A	AJ AJ	AK	AL	AIVI	AN	AU	AP	AQ

(iii) The amino acid sequenced was then copied and pasted in a separate sheet by using the Paste Special options 'Values' and 'Transpose'. By pasting the first amino acid in position A1 the rest amino acids were arranged to the positions



underneath so that the number of the row corresponded to the number of the amino acid residue.



(iv) The interacting residue section of the SCN peptide was then isolated, in this example from the BLAST result detailed in section 7d above this was residues 9-34 (IIAGKNSGVDEAFFVLKQHHVEYGSD), pasted into a new excel file and saved as MR comparison.



A9	*	$\times \checkmark f_{x}$ I													
	A	В	С	D	E	F	G	н	I	J	К				
9	I	1													
10	I														
11	A														
12	G									1	1				
13	к														
14	N										1				
15	S	1													
16	G									1					
17	v														
18	D														
19	E														
20	A														
21	F	ji iji													
22	F														
23	v	1													
24	L														
25	к														
26	Q														
27	H														
28	н														
29	v														
30	E									ř	Ĩ.				
31	Y														
32	G														
33	s									1					
34	D														
25															

(v) The residue section of the Sbjct sequence, in this case 3'-5' RNA helicase YTHDC2 isoform 1 residues 698-723 (NDVVFVIDSGKVKEKSFDALNFVTML) was also copied and pasted in an excel spreadsheet in a similar manner to the SCN sequence and steps (i) – (iii) above were repeated.

A	1 🗘	$\times \checkmark f_{\mathbf{x}} \mid N$												
2	A	В	С	D	E	F	G	н	1	J	К			
1	N													
2	D													
3	v													
4	v													
5	F													
6	V													
7	1													
8	D													
9	S				1									
10	G													
11	к													
12	V													
13	к													
14	E													
15	к													
16	S													
17	F													
18	D					ſ								
19	Α													
20	L													
21	N													
22	F													
23	V													
24	T													
25	M													
26	L													
27														

(vi) The above section was then copied and pasted next to the SCN sequence in the saved MR comparison excel file as below:


B1	. Å	X V	<i>f</i> x ℕ								
	A	В	С	D	E	F	G	н	1	J	К
1	I	N	1								
2	I	D	0								
3	A	v									
4	G	v									
5	к	F									
6	N	v									
7	S	1									
8	G	D									
9	v	S									
10	D	G									
11	E	к									
12	A	v									
13	F	к									
14	F	E									
15	V	к				1					
16	L	S									
17	К	F									
18	Q	D									
19	н	A									
20	н	L									
21	V	N									
22	E	F	ji li								
23	Y	V									
24	G	Т									
25	S	м									
26	D	L									

(vii) The following formula was subsequently copied and pasted in cell C1, to facilitate merging of columns A+B: **=CONCATENATE(A1, " ",B1)**. To complete merging of all the rows, cell C1 was dragged all the way down to last row containing an amino acid code (in this example row 26).

	A	В	С	D	E	F	G	н	 J	К
1	I	N	IN	7						
2	I	D	ID							
3	A	v	AV							
4	G	v	GV							
5	к	F	KF							
6	N	v	NV							
7	S	1	SI							
8	G	D	GD							
9	v	S	VS							
10	D	G	DG							
11	E	к	EK							
12	A	v	AV							
13	F	к	FK							
14	F	E	FE							
15	v	к	VK							
16	L	S	LS							
17	к	F	KF							
18	Q	D	QD							
19	н	A	HA							
20	н	L	HL							
21	٧	N	VN							
22	E	F	EF							
23	Y	v	YV							
24	G	Т	GT							
25	S	м	SM	1						
26	D	L	DL							

(viii) The following formula was inserted into cell D1:

=SUM(COUNTIFS(\$C1:C26, {"A C","A G","A R","A S","C A","C T","D I","D L","D V","E F","E L","F E","F K","G A","G P","G S","G T","H M","H V","I D","I N","I Y","L D","L E","L K","L N","L Q","K L","K F","M H","M Y","N I","N L","N V","P G","P R","P W","Q L","Q V","R A","R P","R S","R T","S A","S G","S R","S S","S T","T C","T G","T R","T S","T W","V D","V H","V N","V Q","V Y","W P","W T","Y I","Y M","Y V"})).

The C26 in the formula should be altered to match the last row containing amino acid pairs in column C.

Н	ome I	nsert D	Draw Page	e Layout	Formulas	Data	Review	View	♀ Tell m	е	
		Tah	ioma	<b>v</b> 11	• A^	A ⊂ =	= =	87 -	ab 🗸	General	
	Paste	3 B	I <u>U</u> ∽	🖽 •   🖉	• <u>A</u> •	E	≡ ≡	<u>+=</u> <u>+=</u>	<b>.</b>	<b>@ *</b> %	9 500
D	1	XV	fx =SUN	A(COUNTIFS(	\$C1:C26, {	'A C","A G",	"A R","A S",	"C A","C	T","D I","D	L","D V","E f	F","E L","F
	A	В	С	D	E	F	G	Н	1	J	К
1	I	N	IN	12							
2	I	D	ID	The second secon						_	
3	A	V	AV								
4	G	V	GV								1
5	K	F	KF								
6	N	V	NV								
7	S	1	SI							_	
8	G	D	GD								
9	V	S	VS	-							
10	D	G	DG								
11	E	ĸ	EK								
12	A .	V	AV						-		
13	-	r.	FK								
14	r.	E V									
15	1	C C	V K			-					
10	L V	5								-	
18	0	D	OD							-	
10	H	Δ	HA							-	
20	н	L L	HL							-	
21	v	N	VN								
22	E	F	EF								
23	Y	v	YV								
24	G	T	GT								
25	S	M	SM								
26	D	L	DL								
27	1										1

- (ix) The final score calculated by this formula equals the amount of all the +1 amino acid pairs out of the total amino acid pairs within the alignment section between the SNC sequence and the subject protein. In this case 12/26 amino acid pairs were awarded a score of +1 which converts into an MR score of 46%.
- (x) The amino acid sequence of AB residues 32-7 corresponds to IIAGKNSGVDEAFFVLKQHHVEYGSD and the amino acid sequence of 3'-5' RNA helicase YTHDC2 isoform 1 residues 698-723 corresponds to NDVVFVIDSGKVKEKSFDALNFVTML. Aligning these in a table manually also allows scoring:

AßIIAGKNSGVDEAFFVLKQHHVEYGSDYTHDC2NDVVFVIDSGKVKEKSFDALNFVTMLScore11001100000011001000111101

\*



- (xi) The same process was followed for all residue alignments, however depending on the source of the subject protein, AS35CN and AS53CN or AS35NC and AS53NC, the corresponding SCN or SNC peptide was used initially in steps (i) -(iii) detailed above on pages 33-34 respectively.
- (7e) The following examples have alignments copied from txt files from BLAST searches (see section 6d, (iv) page 23) with SNC or SCN sequences inserted. Example (a) has been taken from an AS35NC BLAST search and Example (b) has been taken from an AS35CN BLAST search. The scoring has used table 7b above (page 31):

Example (a)

```
>pancreatic triacylglycerol lipase precursor [Homo sapiens]
Sequence ID: NP_000927.1 Length: 465
Range 1: 240 to 256
```

Score:25.8 bits(55), Expect:35, Method:Compositional matrix adjust., Identities:10/17(59%), Positives:12/17(70%), Gaps:0/17(0%)

SNC 54 VELEKGVLPQLEQPYVF 70

Query 54 HLDLFPHDGVELVGIHK 70 HLD FP+ GVE+ G K Sbjct 240 HLDFFPNGGVEMPGCKK 256

#### Scoring (a):

SNC VELEKGVLPQLEQPYVF Sbjct HLDFFPNGGVEMPGCKK Score 1111110111001001

#### Example (b)

>sterile alpha and TIR motif-containing protein 1 precursor [Homo sapiens]
Sequence ID: NP\_055892.2 Length: 724
Range 1: 326 to 341

Score:23.5 bits(49), Expect:214, Method:Compositional matrix adjust., Identities:11/16(69%), Positives:13/16(81%), Gaps:0/16(0%)

SCN 93 LEVVKHGHNTSLADSR 108

Query	93	DLQQFVPVLWSNRLRA	108
		DLQ+ VP+L SNRL A	
Sbjct	326	DLQRLVPLLDSNRLEA	341

Scoring (b):

SCN LEVVKHGHNTSLADSR Sbjct DLQRLVPLLDSNRLEA Score 1110111010111101

The total MR score is best expressed as a % of the total number of residues in the target sequence, in this case %MR = 71% (12/17) for Example (a) and 75% (12/16) for Example (b). The higher this %MR the more likely a binding



interaction could potentially to occur. This number will always be equal to or greater than the % identity score from the BLAST search. Within the results table (see section 6h, page 27) a ninth column can be inserted with the heading % MR and the values determined from this scoring added.

Human Protein that theoretically binds human AB	Protein ID	Size	Residues	Aß residues	% ID	% + <u>ve</u>	% Gaps	% MR
3'-5' RNA helicase YTHDC2 isoform 1	NP_073739.3	1430	698-723	32-7	35	50	0	46
3'-5' RNA helicase YTHDC2 isoform 2	NP_001332904.1	1268	536-561	32-7	35	50	0	46
3'-5' RNA helicase YTHDC2 isoform 3	NP_001332905.1	1130	398-423	32-7	35	50	0	46

(7f) Where comparing multiple potential protein interactions the % MR is a useful indicator of potential interactions. The % scoring accounts for the size of the interacting region without consideration for the overall size of each protein.

The size of each interacting region will be variable, and a larger interacting region may not be more important than a very short one. It is important to remember that protein interactions can involve both small and large sequences binding together. For example, the tripeptide thyrotropin-releasing hormone (TRH) binds to a 398 amino acid receptor but only interacts with a very limited number of the TRH receptor amino acids. As such it is important to consider short interacting regions as well as long interacting regions.

The nature of the molecular recognition theory does not directly take into account secondary or tertiary structures of the proteins that might interact and the %MR should always be considered as a predictive tool with these limitations in mind.



## 8: PDB files for protein-protein interaction modelling

- (8a) For molecular modelling using the ZDock (<u>http://zdock.umassmed.edu</u>, section 9 pages 52-55 below) the software requires a PDB file from the target protein and a PDB file from the suggested interacting protein. From the table of results from the Molecular Recognition analysis above (section 7e, page 39) it is possible to select results in terms of the regions of proteins that interact. Choosing which interactions to study is detailed in Section 12a-e (pages 73-74).
- (8b) Since the target protein will have a protein ID (NP\_\*\*\* number: see section 2b, page 7 above) and the interacting proteins from the BLAST search results will also have a protein ID (NP\_\*\*\* number: see results tables generated in section 7e, page 39 above) it is possible to run a Protein BLAST search to identify similar protein structures. Under the header "Enter accession number, gi or FASTA sequence" paste in the protein ID for the target or interacting protein, in this example the NP\_000475.1 protein ID has been used (circled in red). Also enter the protein name into the Job title box (circled in red). Under the "Choose Search Set", "Database" select Protein Data Bank proteins(pdb), under the "Organism" type in homo sapiens and select "Homo sapiens (taxid:9606)", leave the tick boxes for Exclude "Models (XM/XP)", "Non-redundant RefSeq proteins (WP)" and "Uncultured/environmental sample sequences" unchecked plus select Algorithm "blastp (protein-protein BLAST)" see red \* marks below:

blastn	blastp	blastx	tblastn	tblastx	
Enter Or	and Company				BLASTP programs search protein databases using a protein query. more
Enter access	sion number(s	s), gi(s), or FAS	TA sequence(s)	Clear	Query subrange 🕜
					To
Or, upload fi	ile Cho	ose File no file s	elected	0	
Job Title	Enter	get a descriptive title	o for your BLAST s	earch 😮	
Align two	or more seque	ences 🕜			
Choose	Search Set				
Database	* • Pro	otein Data Bank	proteins(pdb)		<ul> <li>♥</li> </ul>
Organism Optional	* Ho Enter	mo sapiens (taxi	id:9606) on name, binomial	, or tax id. Only	20 top taxa will be shown.
Exclude Optional	*	Models (XM/XP)	Non-redunda	ant RefSeq prot	eins (WP) 🗌 Uncultured/environmental sample sequences
Program	Selection				
Algorithm	*	ollastp (protein-pr PSI-BLAST (Posi PHI-BLAST (Patt DELTA-BLAST (E Disse a BLAST algorithmedia	otein BLAST) ition-Specific Iten ern Hit Initiated E Domain Enhance rithm <b>?</b>	ated BLAST) BLAST) d Lookup Time	Accelerated BLAST)
BLAST		rch database po Show results in a ne	b using Blastp (	(protein-protei	n BLAST)



After clicking BLAST (circled in red above) a series of results will appear. Using the "Alignments" tab in the results it is possible to find structures similar to regions of interest:

escriptions	Graphic Summary	Alignments Tax	conomy		
gnment vie	w Pairwise		Restore defaults		Download 🗠
sequences s	elected 😮				
Ł Downl	oad - GenPept Graphic	25		•	Next  A Previous  A Descriptions
A recep	tor molecule [Homo sap	piens]			
Sequence See 1	ID <u>5BUO_A</u> Length: 342 I more title(s) ✓ See all Ider	Number of Matches: 1 ntical Proteins(IPG)			
		······································			
Range 1:	2 to 342 GenPept Graphics		Viext Match		Related Information
Score 708 bits	Expect Method 1827) 0.0 Compositional	Identities matrix adjust. 341/341(1	Positives Gaps 00%) 341/341(100%) 0/341	(0%)	Structure - 3D structure displays Identical Proteins - Identical
Query				429	proteins to 5BUO_A
Query	STPDAVDKYLETPGDEN	EHAHFQKAKERLEAKHRERMS	QVMREWEEAERQAKNLPKADKK	425	
Sbjct	2 STPDAVDKYLETPGDEN	EHAHFQKAKERLEAKHRERMS	QVMREWEEAERQAKNLPKADKK	61	
Query	430 AVIQHFQEKVESLEQEA	ANERQQLVETHMARVEAMLND	RRRLALENYITALQAVPPRPRH	489	
Sbjct	62 AVIQHFQEKVESLEQEA/	ANERQQLVETHMARVEAMLND ANERQQLVETHMARVEAMLND	RRLALENYITALQAVPPRPRH	121	
Query				549	
Query	VFNMLKKYVRAEQKDRQ	HTLKHFEHVRMVDPKKAAQIR	SQVMTHLRVIYERMNQSLSLLY	545	
Sbjct	122 VFNMLKKYVRAEQKDRQH	HTLKHFEHVRMVDPKKAAQIR	SQVMTHLRVIYERMNQSLSLLY	181	
Query	550 NVPAVAEEIQDEVDELLO	QKEQNYSDDVLANMISEPRIS	YGNDALMPSLTETKTTVELLPV	609	
Sbict	NVPAVAEEIQDEVDELLO	QKEQNYSDDVLANMISEPRIS OKEONYSDDVLANMISEPRIS	YGNDALMPSLTETKTTVELLPV YGNDALMPSLTETKTTVELLPV	241	
0					
query	NGEFSLDDLQPWHSFGAL	DSVPANTENEVEPVDARPAAD	RGLTTRPGSGLTNIKTEEISEV	609	
Sbjct	242 NGEFSLDDLQPWHSFGA	DSVPANTENEVEPVDARPAAD	RGLTTRPGSGLTNIKTEEISEV	301	
Query	670 KMDAEFRHDSGYEVHHQ	KLVFFAEDVGSNKGAIIGLMV	GGV 710		
Shict	KMDAEFRHDSGYEVHHQ	KLVFFAEDVGSNKGAIIGLMV	GGV 342		

Checking through the alignments allows identification of the regions of the searched protein that are present within a structure and therefore choice of a structure that will cover the region of interest from the antisense peptide BLAST results. The protein encoded by NP\_000475.1 is the 770 amino acid amyloid precursor protein. From the alignment's information above the A receptor molecule with a sequence ID 5BUO\_A (circled in red) has a region similar to the NP\_000475.1. The results show that the NP\_000475.1 residues 370 – 710 (circled in blue) are 100% identical (Positives circled in blue) to the 5BUO\_A residues 2-342.

Clicking on the 5BUO\_A link (circled in red) will go to information about the protein, which also contains links to the PDB files (circled in red below):

Protein	Protein	6			Search	
		Advanced				Help
GenPept +				Send to: -	Change region shown	
Chain A	, Amyloid beta A	4 protein				
PDB: 5BUO_	A				Analyze this sequence	
Identical Prote	ins FASTA Graphics				Run BLAST	
Go to: 🖂					Identify Conserved Domains	
LOCUE		242 22	Linear DRT 01 DEC 2020		Highlight Sequence Features	
DEFINITION ACCESSION	Chain A, Amyloid beta 5BUO_A 5BUO_A	A4 protein.	linear PRI 01-DEC-2020		Find in this Sequence	
DBSOURCE	pdb: molecule 5BUO, o	chain A, release	Jul 28, 2016;		Protein 3D Structure	-
KEYWORDS SOURCE	deposition: Jun 4, 20 class: Metal Transpor source: Mmdb_id: <u>141</u> Exp. method: X-Ray Di Homo sapiens (human)	ols; rt; <u>l10</u> , Pdb_id 1: 5B lffraction.	JO;		A Receptor Molec PDB: 5BUO Source: Homo s Method: X-Ray Resolution: 2.31	cule apiens Diffraction Å
UKGANISM	Homo sapiens Eukaryota; Metazoa; ( Mammalia; Eutheria; E Catarrhini; Hominidae	Chordata; Craniat Euarchontoglires; e; Homo.	a; Vertebrata; Euteleostomi; Primates; Haplorrhini;			

For many of the protein entries similar links to structures are available in the NCBI protein database (<u>https://www.ncbi.nlm.nih.gov/protein/</u>), often with links to multiple structures. The BLAST search detailed above is a simple way of selecting an appropriate structure that covers a region of interest.

Clicking on the link for the Protein 3D Structure (circled in red above) will go to the structure file. Within this file there is an option to download the PDB file (see link circled in red below) which can then be saved for use in protein modelling as detailed in sections 9 pages 52-55 below:



(8c) The ZDock molecular modelling software uses PDB files from an individual target protein and suggested interacting protein. The key pieces of information from the antisense peptide BLAST searches required to aid the molecular modelling are the residue number information contained within the results tables generated in section 7e (page 39). The residues of the target protein (in this example amyloid-β residues 32-7) involved in binding and the residues of the interacting protein (in this example NP\_073739.3 - 3'-5' RNA helicase YTHDC2 isoform 1 residues 698-723) from the BLAST search results:

Human Protein that	Protein ID	Size	Residues	Aß	% ID	% + <u>ve</u>	% Gaps	% MR
theoretically				residues				
binds human AB								
3'-5' RNA helicase	NP_073739.3	1430	698-723	32-7	35	50	0	46
YTHDC2 isoform 1				-	· · · · ·			
3'-5' RNA helicase	NP_001332904.1	1268	536-561	32-7	35	50	0	46
YTHDC2 isoform 2								
3'-5' RNA helicase	NP_001332905.1	1130	398-423	32-7	35	50	0	46
YTHDC2 isoform 3								



A BLAST search as described in section 8b (pages 40-42 above) for the NP\_073739.3 sequence corresponding to isoform 1 of the 3'-5' RNA helicase YTHDC2 sequence which interacts with Aß (see table in 8c, page 42 above) identified two structures derived from the YTHDC2 sequence. These structures 6K6U (https://www.ncbi.nlm.nih.gov/Structure/pdb/6K6U) and 2YU6 (https://www.ncbi.nlm.nih.gov/Structure/pdb/2YU6) which both cover a structure representing the YTH domain which corresponds to residues 1288-1421 of isoform 1 of the 3'-5' RNA helicase YTHDC2 sequence. This region of 3'-5' RNA helicase YTHDC2 is outside the proposed 698-723 region of interaction with Aß and therefore modelling using these models is not possible. There are PDB structures for other proteins which do show similarity to the appropriate region of YTHDC2 (698-723), but they are not identical proteins.

- (8d) For protein sequences, like the 3'-5' RNA helicase YTHDC2 698-723 region, where no PDB file is available a predicted protein structure can be created, using the I-Tasser website (<u>https://zhanglab.ccmb.med.umich.edu/I-TASSER/</u>). Protein sequences of 10-1500 amino acids can be entered and have a predicted structure created. There are also options to specify related known structure files as templates if these are available, with details in the dropdown Option menus. The site requires registration as a user, using an academic email address, and download links to the created files will be sent via email.
- (8e) PDB file information can be obtained from a number of sources, the RSCB Protein Databank (<u>https://www.rcsb.org</u>) is the preferred choice and contains validated structural information. Structures can also be obtained from the NCBI protein database (<u>https://www.ncbi.nlm.nih.gov/protein/</u>), the UniProt protein database (<u>https://www.uniprot.org</u>), or the Model Archive (<u>https://www.modelarchive.org</u>).
- (8f) Where structures for a protein are only available in the crystallographic information file (CIF) or macromolecular CIF (mmCIF) formats, which download as "name.cif" files these can be converted to PDB format using the PDBx/mmCIF conversion service (https://mmcif.pdbj.org/converter/index.php?l=en).
- (8g) For many structural models the PDB files contain multiple protein chains from one or more proteins. The PDB file for human catalase (PDB 1DGH; <u>https://www.rcsb.org/structure/1DGH</u>) has four catalase chains as the biologically active enzyme exists as a tetramer of the molecule (Putnam *et al.*, 2000).

An example of a structure with more than one protein type is the structure of the human interferon alpha-2 in complex with human interferon alpha/beta receptor 2 PDB file 3S9D, <u>http://www.rcsb.org/structure/3s9d</u>). The structure comprises two human interferon alpha-2 chains (A & C) in complex with two human interferon alpha/beta receptor 2 chains (B & D):



ind similar proteins by	Sequence -	(by identity cutoff)   Structu	ure		
Entity ID: 1					
Molecule	Chains	Sequence Length	Organism	Details	Image
Interferon alpha-2	A, C	168	<u>Homo sapiens</u>	Mutation(s): 3 <b>3</b> Gene Names: <u>IFNA2</u> , <u>IFNA2A</u> , <u>IFNA2B</u> , <u>I</u> <u>FNA2C</u>	
Find proteins for P018	563 (Homo sapien	s)	Explore P01563	<b>0</b>	Go to UniProt
NIH Common Fund I	Data Resources				
PHAROS: P01563		GTEx: ENSG0000	00188379		
Protein Feature View	1				Exp
PDB ENTITY 3S9D_ UNIPROT ALIGN P UNMODELED A UNMODELED C	3S9D_1		80 AGI PALPERIOSSAVOETI LOKEVITEI VOELA OGI IPALPERIOSSAVOETI LOKEVITEI VOELA		140 160 Wolstower Metro Structure Strukture Metro Structure Strukture
PDB ENTITY 3S9D_' UNIPROT ALIGN P UNMODELED A UNMODELED C ARTIFAC MUTATION	3S9D_1 ACCESACIONELECTICAL CONTRACTOR	(by identity cutoff)   Structu			140 160 WKOSTOWENARE MARPIN STRUCTUR STOWENARE MARPINES 1
PDB ENTITY 3S9D_ UNIPROT ALIGN P. UNIMODELED A UNMODELED A UNMODELED A ARTIFAC MUTATION	3S9D_1	(by identity cutoff)   Structu			
PDB ENTITY 3S9D_ UNIPROT ALIGN P. UNIPROT ALIGN P. UNMODELED Q ARTIFAC MUTATION	3S9D_1	(by identity cutoff)   Structu	Naineusenvaariusevraavaa anneusenvaariusevraavaa I J J J Organism		
PDB ENTITY 3S9D_ UNIPROT ALIGN P. UNMODELED A UNMODELED C ARTIFAC MUTATION	3S9D_1 XCCCAPCIPECENTIAL XC	(by identity cutoff)   Structu	Adite Periode Water Liber Metal Address Adite Periode Water Liber Metal Address and Periode Water Liber Metal Address Jire Organism Homo sapiens	Details Mutation(s): 0 C Gene Names: IFNA B	
PDB ENTITY <b>3S9D</b> UNIPROT ALIGN P UNMODELED A UNMODELED C ARTIFAC MUTATION Find similar proteins by Entity ID: 2 Molecule Interferon alpha/beta receptor 2 Find proteins for P483	3S9D_1	(by identity cutoff)   Structu Sequence Length 199	Addited.settingsweetill.devine.com	Details Mutation(s): 0 Gene Names: IFNA B B B B B Details	Image Go to UniProt
PDB ENTITY <b>3S9D</b> UNIPROT ALIGN P. UNMODELED A UNMODELED A ARTIFAC MUTATION Find similar proteins by Entity ID: 2 Molecule Interferon alpha/beta receptor 2 Find proteins for P483	3S9D_1	(by identity cutoff)   Structu Sequence Length 199 (s)	ADIPA PERCASA VACIONAL DE VICA A ADIPA PERCASA VACIONAL DE VICA A INC. Organism Homo sapiens Explore P48551	Details Mutation(s): 0 O Gene Names: IFNA R2, IFNABR, IFNAR B	Image Go to UniProt P485

(8h) An important feature of PDB files that must be checked is the numbering of residues. PDB files are derived from structural information obtained using purified proteins, which may represent fragments of the whole molecule or posttranslationally modified proteins and may therefore lack regions cleaved as part of this process. Details of how PDB files are derived is available from https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/methods-fordetermining-structure. Often the PDB residue numbering differs from the NCBI or Uniprot residue numbering. Structures are often analysed after removal of signal sequences and some have other regions modified. In the following example the 3S9D\_A chain described above will be used. It is important to check the sequence numbering of the PDB chain against the NCBI chain identified in BLAST searches described above (see section 6 pages 20-30 and section 7 pages 31-39 above). In this example the A chain information can be found at http://www.rcsb.org/structure/3s9d.



Selecting the Interferon alpha-2 A chain (circled in red) will access information about that specific chain:

B PDB Deposit -	Search - Visua	ilize - Analyze - Downloa	ad - Learn - More	- Documentation -	My
Macromolecules					
Find similar proteins t	by: Sequence -	(by identity cutoff)   Struct	ure		
Molecule	Chains	Sequence Length	Organism	Details	Image
Interferon alpha-2	A,C	168	Homo sapiens	Mutation(s): 3 Gene Names: IFNA2 , IFNA2A, IFNA2B, I	

In this example the sequence comparison is between the Interferon alpha-2 A chain and the Uniprot P01563 file (https://www.uniprot.org/uniprot/P01563), circled in red. This sequence corresponds to the NCBI P01563 file (https://www.ncbi.nlm.nih.gov/protein/P01563/). Selecting the K residue highlighted with a red dot (see blue arrow on diagram) on the Uniprot P01563 line of the information brings up numbering information for that specific residue (circled in blue) and compares the Uniprot numbering with the PDB model numbering, selecting any residue in the sequence would have the same effect:

Structure Summary	3D View	Annot	ations	1	Experim	ent	Seq	luence	G	enome		Vers	lions		_		_	_		_	_	_	_	_
S9D																		Dis	play F	les •	٩	Dow	nload	Files -
nary complex betw	ween IFNa2	and IFN	NAR2																				ł	lelp
TANCE	rferon alpha-2	- Homo sa	piens	View F	eatures	s in 3D					ſ	RESID	IUE [UN	IIPRO	T] P	osition	26 [	auth: :	23]   [U	NIPRO	DT] PC	01563	: 46	
	r	1	10		15		20	25		30	-	3	5		40		1 45			1		1	7	
		5																						
PDB INSTANCE	A A D P	CDLP	ОТН	SLO	SRR	TLN	LL	AQMR	RIS	LF	sc	кс	RH	DF	GF	ΡQ	EE	FG	NQ	FQI	K A I	ET	I P	(
PDB INSTANCE UNIPROT ALIGN P015	EA ADP		отн отн	SLO	SRR	R T L M		AQMR	R   8	LF	s c	L K C	RH	D F D F	G F G F	P Q P Q	EE	F G F G	N Q N Q	FQI	( A )	ET	I P	(
PDB INSTANCE UNIPROT ALIGN P015 HEI	EA ADP 163	CDLP	отн отн	SLO	S R R	RTLN		AQMR	RIU		S C	LKC	RH	D F	G F	P Q	EE	F G	NQ	FQI	( A )	ET	I P	۱
PDB INSTANCE UNIPROT ALIGN P015 HEI UNMODELL	EAADP 163 LIX	C D L P	Q T H	SLO	S R A	R T L N			R I !		S C	LKC	RH	D F	G F	PQ	EE	F G	NQ	FQI	< A		I P	۲
PDB INSTANCE UNIPROT ALIGN P015 HEI UNMODEL ZERO OCCUPANCY ATC	EA ADP i63 LIX ED DM	C D L P	Q T H	S L (	SSRA SSRR	RTLN RTLM	I L L		R I S		S C		RH	D F	G F	PQ	E E	FG	NQ	FQI	K A I		I P	۲

The numbering in the blue circled box shows that residue 26 of the model corresponds to the Uniprot P01563 number of 46 (the author of the model used a slightly different numbering scheme starting at -2, which results in an author residue number of 23 for this residue and this numbering shows when using structure viewers (see section 8i, page 46 below). As such if a BLAST search had suggested that residues 50-65 of the Interferon alpha-2 interacted with another protein when selecting these residues on this model they would correspond to residues 30-45 (or author residues 27-42). It often easier to select residues based on a sequence rather than numbers. In this example the sequence would correspond to FSCLKDRHDFGFPQEE. Such sequence information can be found in the results downloaded from BLAST searches if they show similarity to either the AS35CN, AS35NC, AS53CN or AS53NC sequences (see section 6d (iii) Alignments and (iv) Text file on pages 22-23 above, labelled as Sbjct) or they can be found in the SNC sequence of the target protein using the protein id link detailed in section 2b, page 7 above).



(8i) Within the ZDock modelling software (see section 9 pages 52-55 below) it is not possible to select different chains from uploaded files and therefore all protein chains within a model will be modelled. As part of the modelling, it may be preferable to create a PDB file with only a single chain within it (or even part of a chain). The PDB files can be viewed and edited using the iCn3D protein structure viewer (<u>https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html</u>) as described by Wang *et al.*, (2020). See Help Doc (circled in red) and Selection Hints (circled in blue).

File	Select	View	Style	Color	Analysis	Help
			<u>, , , , , , , , , , , , , , , , , , , </u>	2	A	About iCn3D 2.19.0
						Gallery
						Web APIs >
						Source Code
						Transform Hints >
						Selection Hints

(8j) Initially close the "Please input MMDB or PDB" window if it is open and go to File on the main window. From the File menu select Open file from the dropdown and then PDB from the second dropdown to upload a saved PDB file:

••• <>				0			<b>e</b>	ncbi.nlm.nih.gov	
File	Select	View	Style	Color	Analysis	Help	All atoms	Show Toolbar	one-letter see Search Seq. ?
Please input N	IMDB or P.A.								
MMDB or PDB ID:	ITUP Load								
	E			0			ini i	ncbi.nlm.nih.gov	
File	Select	View	Style	Color	Analysis	Help	All atoms	Show Toolbar	one-letter sei Search Seq. ?
Retrieve by ID	PDB File								
Align	TIMOIP PEG								
Realign Selection 1	Mol2 File								
	-						_	_	
Please input	t PDB File			_	_	2 8 8			
Note: Several PD	B files could be co	ncatenated	Favourites	$\langle \rangle$	· ·	Desktor	p	🔕 Q s	Search
PDB File: Choo	se File 3s9d.pdb		<ul> <li>Recents</li> </ul>						
			A Applicati	📄 3s9d.p	odb			-	
			Desktop						
			iCloud						
			iCloud Dri						
			Media						
			🎵 Music						
			O Photos						
			H Movies						
			Tags			3	s9d.pdb		
			Purple			D	ocument - 4	92 KB	
			Yellow			lr c	reated		Show More Today, 10:47
			Red				-		
			Orange				Ca	ncel	Choose for Upload



(8k) Having chosen the PDB file (3s9d.pdb in this example) click on the load (circled in red):

File	Select	View	Style	Color	Analysis	Help	All atoms Show Toolbar   one-letter sec Search Seq. ?
Please inp	ut PDB File				2 B B		
Note: Several P PDB File: Cho	DB files could be conc ose File 3s9d.pdb	catenated into a sing	e PDB file. Use the I	ine "ENDMDL" to se	parate PDB files.		

The initial view of the uploaded structure will look as follows:



Selecting Analysis (circled in red) and "Defined Sets" (circled in blue) from the dropdown menu will result in the appearance of a box listing the chains of the PDB file (outlined in blue on the RHS).



The Defined Sets correspond to the different chains of the structure. In this example the 3S9D\_A and 3S9D\_C correspond to the Interferon alpha-2 A and C chains, whilst the 3S9D\_B and 3S9D\_D correspond to the Interferon alpha/beta receptor 2 B and D chains (see section 8g; page 44 above).

For this example, the 3S9D\_A chain will be selected (corresponding to the one of the Interferon alpha-2 molecules, represented by the A chain in the 3S9D model - <u>http://www.rcsb.org/structure/3s9d</u>). By clicking on the selected chain



(circled in red) followed by the View menu (circled in red) and then the "View only selection" option in the dropdown menu (circled in blue) only the A chain, will be visible – the pink chain in this model:



Files uploaded into the iCn3D protein structure viewer (<u>https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html</u>) can be further modified by selecting specific parts of chains to create a representation of a specific region. If this option is not required proceed straight to section 8m on page 49 below.

(81) After closing the "Select Sets" window selecting the Analysis (circled in red) followed by "View Sequences & Annotations" from the dropdown menu (circled in blue) will bring up a box showing the chain information. Selecting the "Details" tab in this box (circled in red) will bring up the sequence details:



The protein sequence for the model is highlighted in Yellow at the bottom of the "Details" window. Scrolling across this allows the area of interest to be found, in this case the "FSCLKDRHDFGFPQEE" sequence detailed in section 8h above on page 45. Clicking on the F at the start of this sequence and dragging across to the E at the end allows that specific sequence to be selected and the residues will also be highlighted in yellow on the pink image to the left:



File Select View Style Color Analysis Help	Selection Show Toolbar one-letter sec Search Seq. ?
	Sequences and Annotations
BINARY COMPLEX BETWEEN IFNA2 AND IFNAR2	Details
	Conserved Domains ClinVar Functional Sites 3D Domains SNPs Interactions Bonds Cross-Linkages
	: Name: seq_9 Save Clear
	f359D_A: Add Track Custom Color/Tube Hetix Sets Sheet Sets Coil Sets H1 10 20 H2 30 40 H3 55 V Protein 359D_A -2 hSLOSRRTLMLLAQMRRISLEEDENDRHDFGFHQEEFgngfgka

Selecting "View" (circled in red) in the main window followed by "View Selection Only" (circled in blue) in the dropdown menu will result in an image of just the selected structure components:

File	Select	View	Style	Color	Analysis	Help	Selection Show Toolbar one-letter set Search Seq. ?	
		View Only Selection					Sequences and Annotations	×
BINARY C	OMPLEX BET	Hide Selection Zoom in Selection Center Selection View Full Structure	2 AND IFN	IAR2			Details : Conserved Domains ClinVar Functional Sites	
		Side by Side Rotate > Camera > Fog for Selection Slab for Selection XYZ-axes >	2				SD Domains SNPs Interactions Bonds Cross-Linkapes ins : Name: seq_tC Save Clear	
		Reset > Undo Redo Full Screen					1359D_A: Add Track Cuttom Color/Tube Helix Sets Sheet Sets Coll Sets H1 10 20 H2 30 40 H3 V Protein 359D_A -2 hBLGSRRTLMLLAGMRRISLESCLKDHHDFGFRGEEFgngfg	5C ka
					5			

(8m) Clicking on the File (circled in red) followed by "Save Files" (circled in blue) in the first dropdown menu, followed by PDB (circled in blue) in the second dropdown menu will result in the download of a new PDB with just the selected chain or parts of a chain information within the file:





The downloaded file can be renamed from "custom\_icn3d.pdb" to an appropriate name, for example 3S9D\_A\_chain or 3S9D\_A\_binding\_site. The saved PDB file can be used for modelling interactions in the ZDock program as described in section 9 (pages 52-55 below).

(8n) The ZDock program uses two PDB files and has a specific requirement that the chains of each protein within the PDB file have different labels. If there are two A chains these will be merged in the resultant comparison model and this will affect the analysis and production of images.

For an example, the catalase PDB 1F4J (<u>https://www.rcsb.org/structure/1F4J</u>) and the amyloid-ß PDB 5AEF (<u>https://www.rcsb.org/structure/5AEF</u>) could be used as source PDB files, based on the interaction between human catalase and amyloid-ß (Milton *et al.*, 2001). The catalase PDB 1F4J file represents a tetramer of 4 chains labelled A, B, C and D. The amyloid-ß PDB 5AEF file represents a dimer of 2 chains labelled A and B. If when the PDB files were prepared as detailed above resulting in one PDB containing the catalase A chain (1F4J\_chain\_A) and one PDB containing the amyloid-ß A chain (5AEF\_chain\_A) this would create problems in the ZDock modelling (see section 9 pages 52-55 below).

To overcome these problems of PDB files containing the same chain labels the renaming chains website (<u>http://www.canoz.com/sdh/renamepdbchain.pl</u>) can be used. After opening the site files can be directly loaded (circled in red), for this example the 1F4J\_chain\_A is used to rename the chain A:

	0 0	Not Secure —	canoz.com	9 <b>1</b> 1
RENAME CHAIN IN PDB FILE				-200
	Favourites <ul> <li>Recents</li> </ul>	< > 📖 🗸 🧱 🗸 🛅 Dow	nloads 📀	Q Search
Please provide a PDB file with the chain id in column 2 Upload the PDB file from your computer. Please note that this PDB file is not validated, so make Choose File to file selected or Enter the PDB file contents here. Please note that these PDB file contents are not	<ul> <li>▲ Applicati</li> <li>➡ Desktop</li> <li>ICloud</li> <li>➡ iCloud Dri</li> <li>Media</li> <li>➡ Music</li> <li>➡ Photos</li> <li>➡ Movies</li> </ul>	1f4j_chain_A.pdb 1f4j.pdb 5aef_chain_A.pdb 5aef.pdb		
(Provide either a file or enter the contents here, If both provided, then contents here will be proce Rename chain to : (mandatory field)	Tags Purple Yellow Red Orange		5aef_chain_A.pdb Document - 44 KB Information Created Cancel	Show More Today, 12:45 Choose for Upload

In the "Rename chain to" option box the new chain label name is entered, in this case B (circled in blue), the "Write results to output file" option (circled in



# red) is selected, and the "Upload" button (circled in red) then clicked to download a file called "renamepdbchain.pl":

Please provide a PDB file with the chain id in column 22. Upload the PDB file from your computer.
Please note that this PDB file is not validated, so make sure that it is a valid PDB file. <u>Choose File</u> 1f4j_chain_A.pdb
or
Enter the PDB file contents here. Please note that these PDB file contents are not validated, so make sure that they are valid PDB file contents.
(Provide either a file or enter the contents here, not both. If both provided, then contents here will be processed instead of uploaded file.)
Rename chain to : B (mandatory field)
Only rename chain : (optional field)
Line from : (optional field)
Line to : (optional field)
Oisplay results on screen Write results to output file
Upload

The "renamepdbchain.pl" file can be renamed, for example 1F4J\_chain\_B, and then be used in ZDock analysis (see section 9 pages 52-55 below).

## 9: ZDock protein-protein interaction modelling

(9a) The ZDock protein docking program (http://zdock.umassmed.edu), see Pierce et al., (2011 & 2014), uses PDB structure files from two proteins to determine a model structure for two proteins binding each other. Launching the ZDock protein docking program (http://zdock.umassmed.edu) goes to a window where the PDB Files can be uploaded. The ZDock system requires an academic or registered non-profit email address for non-commercial access. The email address should be entered and a link to the files generated by ZDock will be sent to this address. Leave the select ZDock version as the default that is showing (in this case ZDock 3.02 – which is the latest version). The dropdown menu (circled in red) next to the Input Protein 1 or 2 is used to select PDB files that can be uploaded (make sure the PDB File is ticked and highlighted blue. The Choose File tab (circled in blue) should then be clicked:

ZDOCK M-ZDOCK	K Help Tools References
Input Protein 1 Input Protein 2 Enter your email:	PDB File Choose File no file selected
Optional: Select ZDOCK version	ZDOCK 3.0.2

For this example, the catalase PDB file for 1F4J\_chain\_B and the amyloid-ß PDB 5AEF\_chain\_A have been uploaded:

 Favourites	OOCK M-7DOCK	Heln Tools Refere	arch	
Applicati Cloud	1f4j_chain_B.pdb	-		
<ul> <li>iCloud Dri</li> <li>Tags</li> <li>Purple</li> <li>Yellow</li> </ul>				
 Red     Orange     Important		5aef_chain_A.pdb Document - 44 KB		
Blue     Green		Information Created	Show More Today, 12:45	
O Work		Cancel	hoose for Upload	

Once both files have been uploaded click Submit (circled in red):

<b>0</b>	Not Secure zdock.umassmed.edu さ
	CK SERVER
ZDOCK M-	-ZDOCK Help Tools References
Input Proteir	1 PDB File 6 Choose File Saef_chA.pdb
Input Proteir	12 PDB File Choose File 1f4j_chB.pdb
<u>Enter your em</u>	ail: info@neurodelta.uk
Optional:	
Select ZDOCK	version ZDOCK 3.0.2

The next stage is to select the residues that are to be part of the binding site. These are the residues from the BLAST search results file plus the additional sequence information see Table in section 7e above (page 39; how these results can be obtained is fully detailed in sections 6 and 7 pages 20-39 above).

In an example of amyloid-ß and catalase the residues selected for the ZDock predictions are taken from Figure 1a of Milton *et al.*, 2001. The residues thought to interact based on a BLAST search are amyloid-ß 31-40 (circled in red for 5AEF\_chain\_A) and human erythrocyte catalase 400-409 (circled in red for 1F4J\_chain\_B). These residues are selected under "Select Binding Site Residues" on the ZDock server:

2		OCK M-ZDOCK	Help Tools Re	ferences
		Both files have bee	en successfully uploaded.	
		Pick Contact an	d Blocking Residues	
		Select Residues to B	lock from the Binding Site:	
5aef chain Andh	5	aef_chain_A.pdb	1f4j_chain_B.pdb.pl	1f4i chain Bndhnl
to not have Java applets enabled in your browser, or your browser is blocking this applet. the warning message from your browser and/or enable Java applets in eb browser preferences, or install the Java time Environment from <u>www.java.com</u>	11 11 11 11 11 11 12 22 22 22 22 22 22 2	5 Chain A GLN 6 Chain A LYS 7 Chain A LEU 8 Chain A VAL 9 Chain A PHE 0 Chain A PHE 1 Chain A ALA 2 Chain A AL 2 Chain A ASP 4 Chain A VAL	24 Chain B ALA 25 Chain B ASP 26 Chain B VAL 27 Chain B VAL 28 Chain B THR 29 Chain B THR 30 Chain B GLY 31 Chain B ALA 32 Chain B ALN 33 Chain B ASN	You do not have Java applets enabled in your web browser, or your browser is blocking this applet. Check the warning message from your browser and/or enable Java applets in your web browser preferences, or install the Java Runtime Environment from <u>www.java.com</u>
	✓Spin 5/3 3/3 3/3 3/3	Select Bind	ing Site Residues: 1f4j_Chain_B,pdb,pl 400 Chain B GIY 401 Chain B ALA 402 Chain B ARA 402 Chain B ARA 403 Chain B ARA 404 Chain B TYR 405 Chain B TYR 405 Chain B TYR	
	3334	8 Chain A GLY 9 Chain A VAL 0 Chain A VAL	406 Chain B PRO 407 Chain B ASN 408 Chain B SER 409 Chain B PHF	

After clicking the "Submit" button at the bottom of the page the option to check the selected residues before final submission to ZDock is available.



C	Not Secure -	zdock.umassmed.edu C					
	ZDOCK	SERVER					
	ZDOCK M-ZDOCK	Help Tools References					
	You have selected these residues to block and contact						
	To Contact in 5aef_chain_A.pdb:	To Block in 5aef_chain_A.pdb:					
	31 Chain A ILE 32 Chain A ILE 33 Chain A GLY 34 Chain A LEU 35 Chain A MET 36 Chain A VAL 37 Chain A GLY 38 Chain A GLY 39 Chain A VAL 40 Chain A VAL	none					
	To Contact in 1f4j_chain_B.pdb.pl:	To Block in 1f4j_chain_B.pdb.pl:					
	400 Chain B GLY 401 Chain B ALA 402 Chain B PRO 403 Chain B ASN 404 Chain B TYR 405 Chain B TYR	none					
	406 Chain B PRO 407 Chain B ASN 408 Chain B SER						

After final submission to ZDock a notification will appea	r detailing the sending
of results via email and the average wait time for receipt	of the results.

OK

409 Chain B PHE



A receipt email from ZDock Server, followed by second email with a link to the results will then be sent to the address provided (check junk email folder if not received). The job number (in this example 227329 should be noted and linked to the submitted PDB file information for future review):



(9b) The link in the email is for the website results where the ZDock Output, Receptor PDB, Ligand PDB and Top 10 Predictions files should all be downloaded. In this example the Receptor PDB corresponds to amyloid-β (5AEF chain A) and the Ligand PDB corresponds to catalase (1F4J chain B). The predictions are for the interaction between these two proteins.



The top 10 predictions file will be downloaded as a Zip file which should be extracted and will give a folder labelled top\_preds containing up to 10 files labelled complex1.pdb, complex2.pdb etc, which should be saved along with the other downloaded files in a suitable folder.



## 10: iCn3D protein-protein interaction data extraction

(10a) The top 10 prediction PDB files from the ZDock analysis (see section 9b; page 55 above) can also be viewed and modified using the iCn3D protein structure viewer (https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html) as described in section 8i (above; page 46). The complex1.pdb file should be selected at first from the folder containing the downloaded ZDock files. This complex1.pdb contains the predicted complex between the two proteins with the highest and References ZDock score (see Help on the ZDock site, http://zdock.umassmed.edu).

000				0	⊜ ncbi.nlm.nih.;	gov	
File	Select	View	Style		÷	top_preds	\$
Please inpu Note: Several P PDB File: Cho	It PDB File DB files could be conc ose File complex.1.p	catenated into a sing	le PDB file. Use the line "EN	Favourites Dropbox Website Recents Applications Desktop	complex.1.pdb complex.2.pdb complex.3.pdb complex.4.pdb complex.5.pdb complex.6.pdb complex.7.pdb complex.8.pdb		

After selecting load a structure image should appear (these can take time to load).





(10b) Selecting the Analysis tab followed by the View sequences and annotations in the dropdown allows viewing of some of the information about the displayed structure.



Within the Sequences and Annotations window that appears selecting the details tab (circled in red) allows the residues of interest to be highlighted by clicking and dragging over the region from the first to the last residue of interest – in this example derived from Figure 1a of Milton *et al.*, 2001 the IIGLMGGVV sequence that corresponds to the 31-40 residues of amyloid-ß in Protein stru\_A has been selected which is then highlighted in **yellow**. The region selected also shows as a **yellow** selection on the model. By doing the same for each protein, in this case selecting the GAPNYYPNSF sequence that corresponds to the residues 400-409 of catalase in Protein stru\_B, it is possible to view how close the BLAST search sequence interactions from the results (for example see Table in 9a page 42 above). Selecting the Interactions box (circled in blue) shows the residues within the displayed model that interact.



For comparison the image for complex3.pdb shows much closer alignment of the amyloid- $\beta$  31-40 and catalase 400-409 highlighted sequences but also shows considerably fewer residues interacting (see green text in the Interact.B and Interact.A lines in the image above (complex1.pdb) and image below (complex3.pdb). In these examples for amyloid- $\beta$  (11 residues for complex3.pdb) and 20 residues for complex1.pdb) and catalase (8 residues for complex3.pdb and 25 residues for complex1.pdb). The lower the complex number the more valid the model is in terms of the ZDock algorithm scoring, even if this may not



be as close to the predicted alignment from the antisense similarities used to suggest the interaction.

	Sequences and	Annotation	ns				
	Summary Deta	ails				_	
	Annotations: All Custom Disulfide Bonds	Conserved 3D Domair Cross-Link	l Domains ns ages	ClinVar SNPs	C	Functional S	Sites
	Show All Chains						
$\sim$	+ Selection: Name:	seq_1 Sa	ive	Clear			
	Proteins:						
	Annotations of stru_A	A: Add Track	Custor	Color/Tube	Helix Sets	Sheet Sets	Coil Sets
a boan	Protein stru_A Interact .B (prot	15 QKLV 11 Res	FFAED	GSNKGA <mark>1</mark>	IGLMVG IGLMVG	GVVIA GVVIA 11	42 Residues
	Annotations of stru_E	3: Add Track	Custom	Color/Tube	Helix Sets	Sheet Sets	Coil Sets
			30	40		50	60
2 Statistica	Protein stru_B Interact .A (prot	25 DVLT 8 Res	TGAGN	VGDKLNV	ITVGPR	GPLLVQ	VVFTDEM

(10c) By selecting the download feature in the Sequences and Annotations window (circled in red) an html file containing the interactions information is downloaded and can be saved to the folder with the downloaded ZDock files.

Summary D	etails			
Annotations:				
	Conserved Domains	ClinVar	Functional Sites	
Custom	3D Domains	SNPs	Interactions	
Disulfide Bonds	Cross-Linkages			

(10d) Selecting the Analysis tab followed by the H-Bonds & Interactions in the dropdown allows viewing of some of the information about the interactions between the two proteins in the displayed structure.

File	Select	View	Style	Color	Analysis	Help	All atoms	Show Toolbar	one-letter set	Search S
					View Sequences & Annotations					
					Defined Sets					
					H-Bonds & Interactions					
					Bring to Front >					

This will bring up a new window entitled "Hydrogen Bonds/Interactions between two sets of atoms:



File	Selec	ct Viev	v	Style	Color	Analysis	Help All atoms	Show Toolbar one-letter set Search S
Hydrogen b	onds/inte	eractions betw	een two	sets of ator	ns	2 8 8		
1. Choose intera	ction types a	nd their thresholds:						
Hydrogen	Bonds 📒 3.	8 📀 Å 🛛 Salt Br	idge/lonic	6 🖸 A 💽	Contacts/Interaction	ns 📕 4 💿 A		
Halogen B	onds 3.	8 🖸 A 🛛 π-Cati	on 📕	6 🖸 A 🛛	π-Stacking	5.5 🜍 A	0000	
2. Select the fi	rst set:	3. Select the s	econd set:				( CODS	
proteins stru		proteins						
stru_A		stru_A stru_B						
h.			h.					200
4 3D Display I	nteractions	)					ALA	Bar
Highlight Int	eractions in	Table Sort Inte	eractions or	Set 1	et 2			V A Sal
2D Interacti	on Network	show interactions	s between tv	vo lines of residu	e nodes			TRA
2D Interacti	on Map to s	how interactions as it	man				1	A >h
		-					TXIS	
2D Graph (F	orce-Directed	d) to show interacti	ons with str	ength parameter	s in 0-200:			X JAK
Helix or Shee	t: 100	Coil or Nucleotide:	50	Disulfide Bonds:	50			DACK F
Hydrogen Bo	nds: 50	Salt Bridge/Ionic:	50	Contacts:	25			
Halogen Bon	ds: 50	π-Cation:	50	π-Stacking:	50			
(Note: you car	n also adjust i	thresholds at #1 to a	dd/remove i	nteractions.)				
Buried Surfa	ce Area							
5. Reset and	select new s	ets				h		

Within this window all the options are selected by default in part 1 "Choose interaction types and their thresholds". If for example only the Contacts/Interactions option had been ticked in the following parts the data for the other interactions would not be included.

In part 2 the option to select the first set is given, for this example stru\_B could be selected, which corresponds to the part of the model showing as blue lines and represents the catalase structure derived from catalase 1F4J PDB file used in the ZDock comparison (see section 9b, page 55 above).

In part 3 the option to select the second set is given, for this example stru\_A could be selected, which corresponds to the part of the model showing as pink lines and represents the amyloid-ß structure derived from amyloid-ß 5AEF PDB file used in the ZDock comparison (see section 9b, page 55 above).

Part 4 allows the interactions to be views in a number of ways. The "3D Display Interactions" highlights the residues from catalase (blue) and amyloid-ß (pink) that interact and shows lines between the interacting amino acids:



10(e) The "Highlight Interactions in Table" will generate a table of the interactions in a separate window:

Sh	ow interactions				
	hbonds, salt bridge, intera	actions, halogen, pi-cation, p	bi-stacking betw	een Two Sets:	-
		Set 2: stru_A Highlight in 3	D		
		The interfaces are:			
		interface_1 Highlight in 30	0		
		interface_2 Highlight in 30	0		
Note Selection	Each checkbox below selects the correst ction and click on "Highlight" button to clear in the click on "Highlight" button to clear	oonding residue. You can click the checkboxes. 3 hydrogen bond pairs:	"Save Selection"	in the "Select" menu	to save the
	Atom 1	Atom 2	Distance (Å)	Highlight in 3D	
LAND 61	PRO \$stru.B:70@N	GLY \$stru.A:37@0	3.1	Highlight	
ME ( LET	ASP \$stru.B:396@O	ASN \$stru.A:27@N	3.6	Highlight	
TO A STA	SER \$stru.B:408@OG	PHE \$stru.A:19@N	3.4	Highlight	
	0	salt bridge/ionic interaction	pairs:		

By selecting the download feature in the Show Interactions window (circled in red) an html file containing the interactions information is downloaded and can be saved to the folder with the downloaded ZDock files.

If the "Sort Interactions on: Set 1" is selected this table will be sorted based on the stru\_B numbering, which corresponds to the part of the model showing as blue lines and represents the catalase structure derived from catalase 1F4J PDB, rather than sorted into separate tables for each interaction type. If "Sort Interactions on: Set 2" is selected the table will be sorted based on the stru\_A numbering, which corresponds to the part of the model showing as pink lines and represents the amyloid-ß structure derived from 5AEF PDB.



Show so	rted intera								
PRO391	# Hydrogen Bond	# Salt Bridge /lonic	# Contact	# Halogen Bond	#π- Cation	#π- Stacking	Hydrogen Bond	Salt Bridge/Ionic Interaction	Contact
MET392	0	0	4	0	0	0			MET Satru.B.302@N         LE Satru.A.31@CB         1         2.8         4.0         Highlight           MET Satru.B.302@CG         ILE Satru.A.32@N         1         3.5         6.1         Highlight           MET Satru.B.302@CG         ILE Satru.A.32@N         1         3.5         6.1         Highlight           MET Satru.B.302@CG         ILY Satru.A.32@N         1         3.7         6.2         Highlight           MET Satru.B.302@CG         ILY Satru.A.33@CG         1         3.7         6.2         Highlight
CYS393	0	0	1	0	0	0			CYS \$stru.B:393@O
GLN395	0	0	1	0	0	0			GLN \$stru.B:395@0 ASN \$stru.A:27@C   1 3.9 6.6 Highlight
ASP396	1	0	2	0	0	0	ASP Sstru B.396@O		ASP Sstru.B.396@O         ASN Sstru.A.27@N         1         3.6         5.2         Highlight           ASP Sstru.B.396@O         SER Sstru.A.26@CA         1         3.7         5.6         Highlight
ASN397	0	0	2	o	0	0			ASN Setru.B:397@N         ASN Setru.A:27@ND2         1 3.7         5.4         Highlight           ASN Setru.B:397@ND2         GLY Setru.A:25@C         1 3.9         7.4         Highlight

Within a given table the information given the entries detail which atom of a given amino acid in the chain for amyloid-ß interacts with which atom of a given amino acid in the chain for catalase. In this example the MET \$stru.A:35@C and MET \$stru.A:35@N are different atoms from the amyloid-ß methionine 35 residue. The ARG \$stru.B:363@SD, HIS \$stru.B:364@SD, GLY \$stru.B:367@CE and PRO \$stru.B:368@CE are different atoms from catalase residues arginine 363, histidine 364, glycine 367 and proline 368. The Min Distance (Å) represents the minimum distance between the atoms and the C-alpha Distance (Å) represents the distance between the C-alpha atoms of each amino acid. The identified residues can be compared with the BLAST search identified residues (see section 7e, page 39) and also the table of Molecular Recognition scoring table (see section 7b, page 31), in this example the amyloid-ß Met35 and catalase His364 would be predicted to interact based on the Molecular Recognition scoring but not the BLAST search (Milton *et al.*, 2001).

Atom1 A	tom2 # Contacts Min Dis	stance (Å) C-alpha Distance	e (Å)  Highlight in 3D
	MET \$stru.A:35@C	ARG \$stru.B:363@SD	1 2.8 7.9 Highlight
MET25	MET \$stru.A:35@N	☐ HIS \$stru.B:364@SD	1 3.2 5.1 Highlight
	MET \$stru.A:35@N	GLY \$stru.B:367@CE	1 3.7 7.0 Highlight
	MET \$stru.A:35@N	PRO \$stru.B:368@CE	1 3.8 9.1 Highlight

10(f) Diagrams of the interactions can also be manipulated and downloaded. By selecting the Style option for Proteins and or Side Chains the style of the selected components (in this case the interacting residues can be modified:





Resulting in this type of image:



By selecting the View option and View Only Selection, the other parts of the structure are removed from the image.



Using the File, Save Files, iCn3D PNG Image options it is possible to download an image of the currently displayed screen.



Examples of downloaded images from the Catalase (blue) Amyloid-ß (pink) interaction showing the full molecular structure plus interacting molecules of the structure:



- (10g) Other outputs included in the viewer are detailed in the Help Menu (see 10a, page 56 above). All of these features come with default settings, which can be modified as required. Links to each of these specific web pages can be downloaded as described in section 10c on page 58 above.
  - Show interactions between two lines of residue nodes Hold Ctri key to select multiple nodes/lines. – Green: H-Bonds; Cyan: Salt Bridge/lonic; Grey: contacts Magenta: Halogen Bonds; Red: π-Cation; Blue: π-Stacking SVG PNG JSON Scale: 1 📀 169 R363 L366 P368 M392 Q395 N397 G400 P406 S408 D65 E67 F64 R66 R68 P70 H364 G367 P391 C393 D396 G399 Y404 N407 F409  $\mathbf{0}$  $\mathbf{O}$ ( 000 0000000 L17 F19 A21 G25 N27 131 G33 G37 V39 141 V18 F20 D23 S26 A30 132 M35 G38 V40 A42
  - (i) 2D Interaction Network:

(ii) 2D Interaction Map:

Show int	terac	tions	as s	catt	erpl	lot													 _						2	3
Hold Ctrl key SVG P	y to sel	JSON	ltiple no	odes. ale:	+ 1	0																				
F409	0				1																					
S408	0																									
N407	0																									
P406	0																									
Y404	0				1																					
G400	0																									
G399	0																									
N397	0																									
D396	0																									
Q395	0																									
C393	0																									
M392	•																									
P391	0								1																	
P368	0																									
G367	0											62														
L366	0																									
H364	0																									
R363	0																									
P70	0																-									
169	0																									
R68	0															_										
E67	0																									
R66	•																									
D65	0																									
F64	0	-		• •	-	-	~	-	~			-	-	-	-		~	-								
		L17	F19	A2		G25		N27		31	G3	3	G37	,0,	V39		41	0								
		V1	8 F	20	D23	3	S26	F	30	13	2	M35	5	G38	1	V40	1	A42								



(iii) 2D Graph (Force-Directed), which by selecting the Force on Nodes options (circled in red on the Random version below) can be viewed in different formats:

#### Random:





### Y-axis:



Circle:





#### (iv) Buried Surface Area:



Set 1: stru\_B, Surface: 24313.62 Å<sup>2</sup> Set 2: stru\_A, Surface: 3803.79 Å<sup>2</sup> Total Surface: 25953.61 Å<sup>2</sup> **Buried Surface**: 2163.80 Å<sup>2</sup>

## 11: Protein-protein interaction images using EzMol

(11a) Image files can be created by uploading PDB files using the EzMol structure display (<u>http://www.sbg.bio.ic.ac.uk/ezmol/</u>) as described by Reynolds *et al.,* (2018). In Step 1 the PDB file is uploaded.

		0		Not Secure — sbg	g.bio.ic.ac.uk
<b>FZMOI</b>				top_preds	٥
2.1		Favourites	complex.1.pdb complex.2.pdb complex.3.pdb		
Step 1. Upload	Step :	Website  Recents	complex.4.pdb complex.5.pdb		
Step 5. Side-chains	Step	Applications	complex.6.pdb complex.7.pdb		_
You can browse for a <u>PDB</u> file (Brows Or open an EZM file you previously saved in EZMOI Brows	se	Desktop	complex.8.pdb complex.9.pdb complex.10.pdb		
Or enter a 4-letter PDB ID code (e.g. 1CDT)		Cloud		COM PDB F	plex.1.pdb ile - 339 KB
By continuing, you agree that you accept the Disclaimer and Terms and Conditions.		Locations Wetwork		Infor Create Modifi	mation ed ied

(11b) In Step 2 the type of display can be selected for each chain (in this example corresponding to amyloid- $\beta$  for chain A and catalase for chain B. The options are cartoon or stick with the ability to display the surface of the molecules.

Step 1. Upload	Step 2.	Style					
Step 5. Side-chains	Step 6.	Surfaces					
	Which chair	is do you wa	ant to displa	y?			
	Structure	Moloculo	Backbone	e style		Display	
COM .	Structure	wolecule	Cartoon	Stick	Hide	Surface	
and the second	Chain A		$\odot$	0	0		
The second	Chain B		$\odot$	0	0		
E Colores	Colo	ur stick he	teroatom	s by elen	nent		

ż





(11c) In Step 3 the colours of the displayed structures can be selected for each chain (in this example corresponding to amyloid-β for chain A and catalase for chain B. The options allow colour selection for the whole of each chain. In the first example amyloid-β is selected dark grey for both cartoon and stick colour and catalase is selected white.



a. Select a background colour	
b. Select chain or surface colours	
Colour structures Advanced: Cartoon gradients	
Select a single solid colour for a chain ribbon or surface, or use a	rair
StructureCartoon colour@Stick colour@Surface colou	
Chain A	
Chain B	



In the second example the surface structure is chosen with amyloid- $\beta$  selected as dark grey and catalase selected as white.

	a. Select a background colour
	b. Select chain or surface colours
5 5 5 5	Colour structures Advanced: Cartoon gradients
	Select a single solid colour for a chain ribbon or surface, or use a rair
	StructureCartoon colour Stick colour Surface colour
	Chain A
	Chain B
ASA	
The second	

(11d) Step 4 allows selection of different colours for specific residues of the main chain and cartoon. In the example amyloid-ß 31-40 (IIGLMGGVV) have been selected as red and catalase 400-409 (GAPNYYPNSF) selected as green.

Carlo and a second	This tab allows you to apply colours to individual or multiple residues on the chains. Select a colour to paint residues Select the eraser to hide residues	Show 2° structure on sequence below
	Apply by individual residues Apply by secondary structure type Apply by chemical typ Apply the palette colour to the preview image by expanding one of the chains below and then c	e icking-and-dragging on the target residues.
-AQ	15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 Q K L V F F A E D V G S N K G A I I G L	35 36 37 38 39 40 41 42 M V G G V V I A

(11e) Step 5 allows the selection different colours for the specific side chains of residues.

	This tab allows	you to	pick out	locat	tions	and c	olour	s for s	ide c	hains	5.															
	Select a colou	ir for ad	ded sid	e cha	ins	Sele	ct the	erase	r to h	nide s	side cl	hains														
		ur stic	k hete	roat	oms	by e	lem	ent																		
Contraction of the second seco	Apply by ind	ividual r	esidues		App	ly by	chem	ical ty	pe																	
	Apply the p	alette co	lour to	the p	revie	w ima	ige by	expa	ndinį	g one	of th	e cha	ins be	low a	ind ti	nen o	licki	ng-ai	nd-d	raggi	ng oi	n the	e targ	et re	sidue	s.
ATT ATT																	Cha	in A								
-300																- 57	Cha	in B								
	25 2 D	6 27 / L	28 29 T T	30 G	31 A	32 3 G I	13 34 N P	35 V	36 G	37 D	38 3 K	39 4 L M	0 41 I V	42 1	43 T	44 V	45 G	46 P	47 R	48 G	49 P	50 L	51 8 L	52 5 V 4	53 5 Q [	4 55 V
Rock Roll Reset camera Enlarge Shrink	66 6 R	7 68 R	69 70	71 E	72 R	73 7 V 1	4 75 V H	5 76 A	77 K	78 G	79 8 A	30 8 G /	1 82	83 G	84 Y	85 F	86 E	87 V	88 T	89 H	90 D	91	92 9 T	93 9 K	94 9 Y 5	5 96 5 K
	107 1	08 109	110 111	112	113	114 1	15 11	6 117	118	119	120 1	21 12	2 123	124	125	126	127	128	129	130	131	132	133 1	34 1	35 13	6 13
	148 1	19 150	A V	R 153	F 154	5 1	I V	A 7 158	G 150	E 160	S 161 1	G 8	5 A	D 165	166	V 167	R 168	D 169	P 170	R 171	G 172	F 173	A 174 1	V I	K 1	Y 17
Translate Zoom Rotate	N	T	PI	F	F	1	RD	P	1	L	F	PS	F	1	H	S	Q	ĸ	R	N	P	Q	T	H	L	D
	189 1	90 191	192 193	194	195	196 1	97 19	8 199	200	201	202 2	03 20	4 205	206	207	208	209	210	211	212	213	214	215 2	16 2	17 2	8 21
	230 2	21 232	S L	H 235	236 3	V 37 2	38 23	9 240	F 241	5 242	243 2	R C	5 246	247	248	G 249	н 250	R 251	H 252	M 253	N 254	G 255	¥ 256 0	G 3	58 2	9 26
	v	C	KF	H	Y	K	T D	Q	G	1	K	NL	. S	V	E	D	A	A	R	L	S	Q	E	D	P	Y
	271 2	2 273	274 275	276	277 2	278 2	79 28	0 281	282	283	284 2	85 28	6 287	288	289	290	291	292	293	294	295	296	297 2	98 2	99 30	0 30
	312 3	5 K	Y P	317	W 318 1	T 1	F Y	1 322	323	324	M 325 3	T F	N 328	320	A 330	E 331	T 332	F 333	P 334	335	N 336	P 337	338 7	D 30 3	40 3/	K
	P 1	/ G	K L	V	L	NI	RN	P	V	N	Y	FA	E	V	E	Q	1	A	F	D	P	S	N	M	P	G
	353 3	54 355 3	356 357	358	359 3	360 3	61 36	2 363	364	365	366 3	67 36	8 369	370	371	372	373	374	375	376	377 3	378	379 3	80 3	81 38	2 38
	G 304 3	< L	F A	300	P 400	01 4		R 3 404	H 405	406	L /		N 410	Y 411	L 412	H 413	1	P 415	V 416	N 417	418	P 410	¥ 420 /	R /	A F	V 12
	M (	2 D	N Q	G	G	A	P N	Y	Y	P	N	S F	G	A	P	E	Q	Q	P	S	A	L	E	H	S 4	Q



(11f) Step 6 allows the selection different colours for the surface structure of residues; this is preferable to selecting surface colour in step 3 as it shows the chain detail within the model. In the example again the amyloid-β 31-40 (IIGLMGGVV) have been selected as red and the catalase 400-409 (GAPNYYPNSF) selected as green.



(11g) Step 7 allows the insertion of labels where the colour of the background and label together with the format applies to all the added labels.

Step 5. Side-chains	Step 6. Surfaces	Step 7. Labels
	This tab allows you to label the chain residues	Labels can either be off (white cell) or on (black cell). Click, drag and release to add labels
	I Always show labels on top	sequence below
	Label style settings Apply to individual	residues
E Posto Star	O Three letter code with number	e.g. ILE32
and the second	One letter code with number	e.g. 132
-AD -	O Number only <i>e.g. 32</i> O Three letter code without num	iber <i>e.g. ILE</i>
Q	O One letter code without numb	er e.g. l
Rock Roll Reset camera Enlarge Shrin	Select the foreground colour for the label	Select the background colour for the label
• • • •	Apply labels to the preview image by expan	ding one of the chains below and then clicking-and-dragging on the target residues.
Step 5. Side-chains	Step 6. Surfaces	Step 7. Labels
	This tab allows you to label the chain residues	Labels can either be off (white cell) or on (black cell). Click, drag and release to add labels
	I Always show labels on top	sequence below
	Label style settings Apply to individual	residues
E Portos		- Chain A
A42	15 16 17 18 19 20 21 22 2 Q K L V F F A E I	3 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 <b>V G S N K G A I I G L M V G G V V I A</b>
-AD -		
an and		
Rock Roll Reset camera Enlarge Shrin	k	
Q ? ?		


(11h) The final Step 8 generates a rendered image that can then be downloaded as a png file. Using the controls below the image the model can be rotated to achieve the desired version for presentation.



Example of downloaded png file showing the interaction between amyloid-ß and catalase with specific residues from Milton *et al.*, 2001 labelled.





## **12: Interpretation of results**

- (12a) The target protein originally chosen (see Section 2 pages 5-10) will have a protein id (see Section 2(b) page 7) that can be searched to obtain further information from the NCBI Proteins (https://www.ncbi.nlm.nih.gov/protein/) or UniProt (https://www.uniprot.org) websites. The SNC sequence can also be BLAST searched (see Section 4(a) page 14) to obtain information about related proteins and isoforms of the target protein. Key features to identify for the target protein include modified residues (for example phosphorylated residues); residues that bind co-factors, ligands, allosteric ligands, substates or other proteins; regions linked to protein activity (for example the active sites of enzymes or ligand binding domains of receptors); regions that play a role in localisation (for example the extracellular, transmembrane and intracellular regions of receptors); regions where a protein undergoes post-translational cleavage (for example signal peptides or hormone pre- and pro- forms); regions with structural details (for example β-turns).
- (12b) These key features also need to be identified for each of potential binding proteins listed in the results tables (see Section 7(e) page 39 above) from the BLAST searches for alignments with AS35NC, AS35CN, AS53NC and AS53CN sequences.
- (12c) A graphical map of potential interaction domains can be created from the data obtained and is particularly useful to look at the potential effects of interactions.



A graphical view of the domains and interacting regions within VEGFA. The black boxes within the chain represent the protein residues in which VEGFA binds to cubilin precursor, gametogenetin binding protein 2, and FGFR2 isoform 7 precursor. PDGF = platelet-derived growth factor.

Similar graphical views can be prepared for each interacting protein, in this example the cubulin precursor, gametogenetin binding protein 2 and FGFR2 isoform 7 precursor.

(12d) Searches of publications using PubMed (<u>https://pubmed.ncbi.nlm.nih.gov</u>), Google Scholar (<u>https://scholar.google.com</u>) and Science Direct (<u>https://www.sciencedirect.com</u>) can be used to identify published links or processes that are linked to each protein.



- (12e) Tissue localisation of the proteins is also very useful to help determine if an interaction is likely, this information may be provided in publications and can also be checked using the online Human Protein Atlas (https://www.proteinatlas.org).
- (12f) From the interaction modelling results tables from section 10(e) (pages 60-61) it is possible to determine which amino acids are predicted to interact. This information can be combined with the information from the results tables (see Section 7(e) page 39 above) from the BLAST searches, for alignments with AS35NC, AS35CN, AS53NC and AS53CN sequences, to determine if the two techniques identify similar regions of proteins involved in interactions. Information about published interactions for each protein can be found at the IntAct Molecular Interaction Database (https://www.ebi.ac.uk/intact/) and structural information can also be found at the RSCB Protein Databank (https://www.rcsb.org). From these data sets it is possible to suggest which residues may interact based on the Bioinformatic computer predictions.
- (12g) Images of the suggested interacting structures from section 10 (pages 56-67) and section 11 (pages 68-72) can be generated to illustrate the suggested interactions.
- (12h) From these results practical experiments can be designed to prove or disprove the theoretical interactions identified using the Antisense Peptide Bioinformatics and Molecular Docking protocols.
- (12i) The antisense peptide sequences can also be used to generate synthetic peptides for use in experimental settings.



## **13: References and websites**

- [1] Bioinformatics Protocols (<u>https://www.neurodelta.uk/Protocols/</u>)
- Biro, J.C. (2007) The Proteomic Code: a molecular recognition code for proteins. *Theor. Biol. Med. Model.*, **4**, 45. (<u>https://doi.org/10.1186/1742-4682-4-45</u>).
- [3] Blalock, J.E. & Bost, K.L. (1986) Binding of peptides that are specified by complementary RNAs. *Biochem. J.*, **234(3)**, 679-683. (<u>https://doi.org/10.1042/bj2340679</u>).
- [4] Blalock, J.E. & Bost, K.L. (1988) Ligand receptor characteristics of peptides encoded by complementary nucleic acids: implications for a molecular recognition code. *Recent Prog. Horm. Res.*, 44, 199-222. (<u>https://doi.org/10.1016/b978-0-12-571144-9.50011-9</u>).
- Blalock, J.E. & Smith, E.M. (1984) Hydropathic anti-complementarity of amino acids based on the genetic code. *Biochem. Biophys. Res. Commun.*, **121(1)**, 203-207. (<u>https://doi.org/10.1016/0006-291x(84)90707-1</u>).
- [6] Bost, K.L. & Blalock, J.E. (1989a) Complementary Peptides as Interactive Sites for Protein Binding. *Viral Immunol.*, 2(4), 229-238. (<u>https://doi.org/10.1089/vim.1989.2.229</u>).
- [7] Bost, K.L. & Blalock, J.E. (1989b) Preparation and use of complementary peptides. *Methods Enzymol.*, **168**, 16-28.
   (<u>https://doi.org/10.1016/0076-6879(89)68005-6</u>).
- [8] Bost, K.L., Smith, E.M. & Blalock, J.E. (1985) Similarity between the corticotropin (ACTH) receptor and a peptide encoded by an RNA that is complementary to ACTH mRNA. *Proc. Natl. Acad. Sci. U S A.*, 82(5), 1372-1375. (https://doi.org/10.1073/pnas.82.5.1372).
- [9] Capone, G., De Marinis, A., Simone, S., Kusalik, A. & Kanduc, D. (2008) Mapping the human proteome for non-redundant peptide islands. *Amino Acids.*, **35(1)**, 209-216. (<u>https://doi.org/10.1007/s00726-007-0563-7</u>).
- [9a] Chen, R., & Weng, Z. (2002). Docking unbound proteins using shape complementarity, desolvation, and electrostatics. *Proteins*, **47(3)**, 281–294. (<u>https://doi.org/10.1002/prot.10092</u>).
- [10] Chilumuri, A., Odell, M. & Milton, N.G. (2013a) Benzothiazole aniline tetra(ethylene glycol) and 3-amino-1,2,4-triazole inhibit neuroprotection against amyloid peptides by catalase overexpression in vitro. ACS Chem. Neurosci., 4(11), 1501-1512. (https://doi.org/10.1021/cn400146a).
- [11] Chilumuri, A., Ashioti, M., Nercessian, A. N., & Milton, N. G. (2013b). Immunolocalization of Kisspeptin Associated with Amyloid-β Deposits in the Pons of an Alzheimer's Disease Patient. *Journal of neurodegenerative diseases*, 2013, 879710. (https://doi.org/10.1155/2013/879710).



- [12] Clarke, B.L. & Blalock, J.E. (1990) Steroidogenic activity of a peptide specified by the reversed sequence of corticotropin mRNA. *Proc. Natl. Acad. Sci. U S A.*, 87(24), 9708-9711. (https://doi.org/10.1073/pnas.87.24.9708).
- [13] EzMol structure display (<u>http://www.sbg.bio.ic.ac.uk/ezmol/</u>)
- [14] Fassina, G., Roller, P.P., Olson, A.D., Thorgeirsson, S.S. & Omichinski, J.G. (1989) Recognition properties of peptides hydropathically complementary to residues 356-375 of the c-raf protein. *J. Biol. Chem.*, **264(19)**, 11252-11257. (https://www.jbc.org/content/264/19/11252.long)
- [15] Google Scholar (<u>https://scholar.google.com</u>)
- [16] Hardison, M.T. & Blalock, J.E. (2012) Molecular recognition theory and senseantisense interaction: therapeutic applications in autoimmunity. *Front. Biosci. (Elite Ed).*, **4**, 1864-1870. (<u>https://doi.org/10.2741/508</u>).
- [17] Heal, J.R., Roberts, G.W., Raynes, J.G., Bhakoo, A. & Miller A.D. (2002) Specific interactions between sense and complementary peptides: the basis for the proteomic code. *Chembiochem.*, **3(2-3)**, 136-151. (<u>https://doi.org/10.1002/1439-7633(20020301)3:2/3%3c136::aidcbic136%3e3.0.co;2-7</u>).
- [18] iCn3D struct viewer (<u>https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html</u>)
- [19] Illingworth, C.J., Chintipalli, S.V., Serapian, S.A., Miller, A.D., Veverka, V., Carr, M.D. & Reynolds, C.A. (2012) The statistical significance of selected senseantisense peptide interactions. *J. Comput. Chem.*, **33(16)**, 1440-1447. (<u>https://doi.org/10.1002/jcc.22977</u>).
- [20] IntAct Molecular Interaction Database (<u>https://www.ebi.ac.uk/intact/</u>)
- [21] I-Tasser PDB Prediction (<u>https://zhanglab.ccmb.med.umich.edu/I-TASSER/</u>)
- [22] Kyte, J. & Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, **157(1)**, 105-132. (<u>https://doi.org/10.1016/0022-2836(82)90515-0</u>).
- [23] McGuire, K.L. & Holmes, D.S. (2005) Role of complementary proteins in autoimmunity: an old idea re-emerges with new twists. *Trends Immunol.*, 26(7), 367-372. (<u>https://doi.org/10.1016/j.it.2005.05.001</u>).
- [24] Miller, A.D. (2015) Sense-antisense (complementary) peptide interactions and the proteomic code; potential opportunities in biology and pharmaceutical science. *Expert Opin. Biol. Ther.*, **15(2)**, 245-267. (https://doi.org/10.1517/14712598.2015.983069).
- [25] Milton, N.G.N., Mayor, N.P. & Rawlinson, J. (2001) Identification of amyloidbeta binding sites using an antisense peptide approach. *Neuroreport*, **12(11)**, 2561-2566. (<u>https://doi.org/10.1097/00001756-200108080-00054</u>).
- [26] Milton, N.G.N. (2006) Anti-sense Peptides. In *Cell Biology Protocols*, Eds D. Rickwood, J. Graham & J.R. Harris, Wiley, London, pp 353-358. (<u>https://doi.org/10.1002/0470033487.ch6</u>).

- [27] Model Archive (<u>https://www.modelarchive.org</u>)
- [28] Mulchahey, J.J., Neill, J.D., Dion, L.D., Bost, K.L. & Blalock, J.E. (1986) Antibodies to the binding site of the receptor for luteinizing hormone-releasing hormone (LHRH): generation with a synthetic decapeptide encoded by an RNA complementary to LHRH mRNA. *Proc. Natl. Acad. Sci. U S A.*, 83(24), 9714-9718. (https://doi.org/10.1073/pnas.83.24.9714).
- [29] NCBI Nucleotide Database (https://www.ncbi.nlm.nih.gov/nucleotide/)
- [30] NCBI Protein Database (https://www.ncbi.nlm.nih.gov/protein/)
- [31] NCBI BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi)
- [32] PDBx/mmCIF conversion (<u>https://mmcif.pdbj.org/converter/index.php?l=en</u>)
- [33] PDB chain renaming (<u>http://www.canoz.com/sdh/renamepdbchain.pl</u>)
- [34] PDB information (<u>https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/methods-for-determining-structure</u>)
- [35] Pearson W. R. (2013). Selecting the Right Similarity-Scoring Matrix. *Curr. Prot. Bioinf.*, **43**, 3.5.1–3.5.9. (https://doi.org/10.1002/0471250953.bi0305s43).
- [36] Pierce, B. G., Hourai, Y., & Weng, Z. (2011). Accelerating protein docking in ZDock using an advanced 3D convolution library. *PloS One*, 6(9), e24657. (<u>http://doi:10.1371/journal.pone.0024657</u>).
- [37] Pierce, B.G., Wiehe, K., Hwang, H., Kim, B.H., Vreven, T. & Weng, Z. (2014) ZDock server: interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics.*, **30(12)**, 1771-1773. (<u>https://doi.org/10.1093/bioinformatics/btu097</u>).
- [38] PubMed (<u>https://pubmed.ncbi.nlm.nih.gov</u>)
- [39] Pullen, J.R., Dalmaris, J., Serapian, S.A. & Miller, A.D. (2013) Assessing the preferred solution conformation of an interacting sense-antisense (complementary) peptide pair. *Bioorg. Med. Chem. Lett.*, **23(2)**, 496-502. (<u>https://doi.org/10.1016/j.bmcl.2012.11.038</u>).
- [40] Putnam, C.D., Arvai, A.S., Bourne, Y., & Tainer, J. A. (2000). Active and inhibited human catalase structures: ligand and NADPH binding and catalytic mechanism. *Journal of molecular biology*, **296(1)**, 295–309. (<u>https://doi.org/10.1006/jmbi.1999.3458</u>)
- [41] Python (<u>https://www.python.org</u>)
- [42] Python complier (<u>https://trinket.io/python3</u>)
- [43] Reynolds, C. R., Islam, S. A., & Sternberg, M. (2018). EzMol: A Web Server Wizard for the Rapid Visualization and Image Production of Protein and Nucleic Acid Structures. *Journal of molecular biology*, **430(15)**, 2244–2248. (https://doi.org/10.1016/j.jmb.2018.01.013).
- [44] Root-Bernstein R.S. & Holsworth, D.D. (1998) Antisense peptides: a critical minireview. *J. Theor. Biol.*, **190(2)**, 107-119.

÷J



(https://doi.org/10.1006/jtbi.1997.0544).

- [45] RSCB Protein Databank (<u>https://www.rcsb.org</u>)
- [46] Science Direct (<u>https://www.sciencedirect.com</u>)
- [47] Siemion, I.Z., Cebrat, M. & Kluczyk, A. (2004) The problem of amino acid complementarity and antisense peptides. *Curr. Protein Pept. Sci.*, **5(6)**, 507-527. (<u>https://doi.org/10.2174/1389203043379413</u>).
- [48] Štambuk, N., Konjevoda, P., Boban-Blagaić, A. & Pokrić, B. (2005) Molecular Recognition Theory of the complementary (antisense) peptide interactions. *Theory Biosci.*, **123(4)**, 265-275. (https://doi.org/10.1016/j.thbio.2005.02.001).
- [49] Štambuk, N., Konjevoda, P., Turčić, P., Kövér, K., Kujundžić, R.N., Manojlović, Z. & Gabričević, M. (2018). Genetic coding algorithm for sense and antisense peptide interactions. *Biosystems.*, **164**, 199-216. (<u>https://doi.org/10.1016/j.biosystems.2017.10.009</u>).
- [50] Štambuk, N., Konjevoda, P., Turčić, P., Šošić, H., Aralica, G., Babić, D., Seiwerth, S., Kaštelan, Ž., Kujundžić, R.N., Wardega, P., Žutelija, J.B., Gračanin, A.G. & Gabričević, M. (2019). Targeting Tumor Markers with Antisense Peptides: An Example of Human Prostate Specific Antigen. *Int. J. Mol. Sci.*, **20(9)**, 2090. (<u>https://doi.org/10.3390/ijms20092090</u>).
- [50a] Štambuk, N., Konjevoda, P., & Pavan, J. (2021). Antisense Peptide Technology for Diagnostic Tests and Bioengineering Research. *International journal of molecular sciences*, **22(17)**, 9106. (<u>https://doi.org/10.3390/ijms22179106</u>).
- [51] The Human Protein Atlas (<u>https://www.proteinatlas.org</u>)
- [52] UniProt (<u>https://www.uniprot.org</u>)
- [53] Upsidedown text (<u>http://www.upsidedowntext.com/</u>)
- [54] Wang, J., Youkharibache, P., Zhang, D., Lanczycki, C. J., Geer, R. C., Madej, T., Phan, L., Ward, M., Lu, S., Marchler, G. H., Wang, Y., Bryant, S. H., Geer, L. Y., & Marchler-Bauer, A. (2020). iCn3D, a web-based 3D viewer for sharing 1D/2D/3D representations of biomolecular structures. *Bioinformatics (Oxford, England)*, 36(1), 131–135. (https://doi.org/10.1093/bioinformatics/btz502).
- [55] ZDock server (<u>http://zdock.umassmed.edu</u>)



## 14: Appendix 1 - Installing Python

- (14a) This protocol uses Python 3 software to run a program. Recommended to download the latest stable version, the program used has been tested on versions 3.6.0 and above (current stable version is 3.9.4). If Python cannot be installed either use an online version of Python (Section 3 above, pages 11-13) or an alternative method to generate antisense peptides manually is detailed in Section 15 (pages 83-86 below).
- (14b) On a Mac to download Python go to <u>https://www.python.org</u> and follow instructions:



On a PC to download Python go to <u>https://www.python.org</u> and follow instructions:

	Windows 10 File Edit View Actions	Devices Window Help			<b>0 7 6</b> 8 🤅	100% 🚱 Thu 19:16 📑 🔐 😑
	••			Windows 10	EE @ .	• • • • • • • • • • • • • • • •
	📌 Welcome to Python.org 🛛 🗙 🕂					- ø ×
م	$\leftrightarrow$ $\rightarrow$ C $\hat{\mathbf{a}}$ python.org					<b>☆ 0</b> :
0		Python				Í
0		authoo"				
1		- pymon		Donate Search:	GO Socialize	
		About	Downloads Documer	ntation Community Success Stories News	s Events	
۲						
		# Python 3: Lis	All releases	Download for Windows		
		>>> loud_fruits	Source code	Python 3.8.3	e one of the hds. Lists can be	
		fruits]	Windows	Note that Python 3.5+ cannot be used on Windows XP	puilt-in	-
2		['BANANA', 'APP	Mac OS X	or earlier.		
۲		# List and the	Other Platforms	Not the OS you are looking for? Python can be used on many operating systems and environments.		
<u>ک</u>		<pre>&gt;&gt;&gt; list(enumer [(0, 'Banana').</pre>	License	View the full list of downloads.		
			Alternative Implementations			

(14c) Once installed on a Mac there will be a Python folder in the Applications (Mac), to open double click on the IDLE icon (highlighted in blue):

Recents	🔻 🛃 Python 3.8	1 Apr 2020 at 01:02		Folder
Applications	Icon?	24 Feb 2020 at 22:54	317 KB	TextEdit
Desktop	🥺 IDLE	1 Apr 2020 at 01:02	188 KB	Application
	Install Certificates.command	24 Feb 2020 at 22:54	1 KB	Terminll script
Documents	License.rtf	24 Feb 2020 at 22:54	13 KB	RTF Document
O Downloads	Python Documentation.html	1 Apr 2020 at 01:02	98 bytes	Alias
	🎍 Python Launcher	1 Apr 2020 at 01:02	269 KB	Application
III Hathinton	🔓 ReadMe.rtf	24 Feb 2020 at 22:54	3 KB	RTF Document
Pictures	<ul> <li>Update Shell Profile.command</li> </ul>	24 Feb 2020 at 22:54	3 KB	Terminll script
-	· · · · · · ·			



On a PC the Python folder should be on the C drive (PC) in the Programs folder and to open double click on the IDLE icon (highlighted in blue):

		A				
Videos	^	Name	Date modified	Туре	Size	
his PC		A IDLE (Python 3.7 64-bit)	28/03/2020 19:56	Shortcut	3 KB	
3D Objects		Python 3.7 (64-bit)	28/03/2020 19:56	Shortcut	2 KB	
Desktop		Python 3.7 Manuals (64-bit)	28/03/2020 19:56	Shortcut	1 KB	

(14d) Download the Python script (AntisensePeptide.py) file from Antisense-Peptide.py (available from as a either Python script <u>https://www.neurodelta.uk/resources/BioinformaticsProtocolScript.py</u> save to a suitable folder on the hard drive. A copy the complete text is shown below:

```
##Original code by Jonathan C Goulding, Adapted to Py3 and Modified by
Harrison R S Milton, based on Milton, N.G.N. (2006) Anti-sense Peptides.
Protocols 6.39 In Cell Biology Protocols, Eds D. Rickwood, J. Graham &
J.R. Harris, Wiley, London, pp 353–358. 🖓 J.C. Goulding, H.R.S. Milton &
N.G.N. Milton; School of Clinical & Applied Sciences, Leeds Beckett
University; Neurodelta Ltd.
g=input("Input name: ")
s=input("Input coding mRNA: ")
i=s.replace(' ', '').replace('0', '').replace('1', '').replace('2',
'').replace('3', '').replace('4', '').replace('5', '').replace('6',
'').replace('7', '').replace('8', '').replace('9', '').replace('A',
'a').replace('C', 'c').replace('G', 'g').replace('T', 't').replace('U',
't').replace('u', 't').replace('\n', '')
def breakdown(data):
     array=[]
     for i in range(0,len(data),3):
         if (i+3>len(data)):
              upper = len(data)
         else:
              upper =i+3
         seq=data[i:upper]
         try:
              array.append(amino_acids[seq])
         except KeyError :
              array.append('unknown')
     return array
def flip(x):
  return x[::-1]
amino acids =
{'aaa<sup>'</sup>:'K','aac':'N','aag':'K','aat':'N','aca':'T','acc':'T','acg':'T','a
ct':'T','aga':'R','agc':'S','agg':'R','agt':'S','ata':'I','atc':'I','atg'
:'M', 'att':'I', 'caa':'Q', 'cac':'H', 'cag':'Q', 'cat':'H', 'cca':'P', 'ccc':'P
','ccg':'P','cct':'P','cga':'R','cgc':'R','cgg':'R','cgt':'R','cta':'L','
ctc':'L','ctg':'L','ctt':'L','gaa':'E','gac':'D','gag':'E','gat':'D','gca
':'A','gcc':'A','gcg':'A','gct':'A','gga':'G','ggc':'G','ggg':'G','ggt':'
G', 'gta': 'V', 'gtc': 'V', 'gtg': 'V', 'gtt': 'V', 'tac': 'Y', 'tat': 'Y', 'tca': 'S',
'tcc':'S','tcg':'S','tct':'S','tgc':'C','tgg':'W','tgt':'C','tta':'L','tt
c':'F','ttg':'L','ttt':'F','taa':'*','tga':'*','tag':'*'}
output=breakdown(i)
combined=''
for acid in output:
     combined =combined+acid
```

```
print ("")
print("SNC -",g,"=",combined)
d=flip(combined)
print ("")
print("SCN -",g,"=",d)
amino_acids =
{'aaa':'F','aac':'L','aag':'F','aat':'L','aca':'C','acc':'W','acg':'C','a
ct':'X','aga':'S','agc':'S','agg':'S','agt':'S','ata':'Y','atc':'X','atg'
:'Y', 'att':'X', 'caa':'V', 'cac':'V', 'cag':'V', 'cat':'V', 'cca':'G', 'ccc':'G
', 'ccg':'G', 'cct':'G', 'cga':'A', 'cgc':'A', 'cgg':'A', 'cgt':'A', 'cta':'D', '
ctc':'E','ctg':'D','ctt':'E','gaa':'L','gac':'L','gag':'L','gat':'L','gca
':'R','gcc':'R','gcg':'R','gct':'R','gga':'P','ggc':'P','ggg':'P','ggt':'
P','gta':'H','gtc':'Q','gtg':'H','gtt':'Q','tac':'M','tat':'I','tca':'S',
'tcc':'R','tcg':'S','tct':'R','tgc':'T','tgg':'T','tgt':'T','tta':'N','tt
c':'K','ttg':'N','ttt':'K','taa':'*','tga':'*','tag':'*'}
output=breakdown(i)
combined=''
for acid in output:
     combined =combined+acid
print ("")
print("AS35NC -",g,"=",combined)
e=flip(combined)
print ("")
print("AS35CN -",g,"=",e)
amino_acids =
{'aaa':'F','aac':'V','aag':'L','aat':'I','aca':'C','acc':'G','acg':'R','a
ct':'S','aga':'S','agc':'A','agg':'P','agt':'T','ata':'Y','atc':'D','atg'
:'H','att':'N','caa':'L','cac':'V','cag':'L','cat':'M','cca':'W','ccc':'G
','ccg':'R','cct':'R','cga':'S','cgc':'A','cgg':'P','cgt':'T','cta':'X','
ctc':'E','ctg':'Q','ctt':'K','gaa':'F','gac':'V','gag':'L','gat':'I','gca
':'C','gcc':'G','gcg':'R','gct':'S','gga':'S','ggc':'A','ggg':'P','ggt':'
T','gta':'Y','gtc':'D','gtg':'H','gtt':'N','tac':'V','tat':'I','tca':'X',
'tcc':'G','tcg':'R','tct':'R','tgc':'A','tgg':'P','tgt':'T','tta':'X','tt
c':'E','ttg':'Q','ttt':'K','taa':'*','tga':'*','tag':'*'}
output=breakdown(i)
combined=''
for acid in output:
     combined =combined+acid
print ("")
print("AS53NC -",g,"=",combined)
d=flip(combined)
print ("")
print("AS53CN -",g,"=",d)
```

(14e) Alternatively in the Python Shell after clicking in File will give a pull down on which the New File (Ctrl+N) can be selected:

	IDLE	File	Edit	Shell	Debug	Options	Window	Help	
•	•				Python	3.8.2 Shell			
Python [Clang Type >>>	n 3.8.2 g 6.0 (d "help",	(v3.8 clang- "copy	.2:7b3 600.0. right"	ab5921f 57)] on , "cred:	, Feb 24 darwin its" or "	2020, 17:5 license()"	2:18) for more	information.	

Python 3.7.7 She	ell			_		×
File Edit Shell Deb New File Open Open Module Recent Files Module Browser Path Browser	Ctrl+N Ctrl+O Alt+M Alt+C	ndow Help ':d7c567b08f, Mar 10 2020, "credits" or "license()"	10:41:24) for more	[MSC v.1900	64	bit ^
Save Save As Save Copy As	Ctrl+S Ctrl+Shift+S Alt+Shift+S					
Close Exit	Alt+F4 Ctrl+Q					

(14f) Download the Python script (AntisensePeptide.docx) file from <u>https://www.neurodelta.uk/Protocols/</u>. Open the document in Microsoft Word and Copy the complete text. Paste into the blank untitled window on Python:



Then save the file on the Hard drive in a suitable folder as AntisensePeptide.py using the file Save (Ctrl+S) command.

# 15: Appendix 2 - Python antisense peptide generation

(15a) For the installed version of Python on a Mac there will be a Python folder in the Applications (Mac), double click the IDLE (highlighted in blue):

Recents	V 🛃 Python 3.8	1 Apr 2020 at 01:02		Folder
Applications	lcon?	24 Feb 2020 at 22:54	317 KB	TextEdit
Desktop	🕺 IDLE	1 Apr 2020 at 01:02	188 KB	Application
	<ul> <li>Install Certificates.command</li> </ul>	24 Feb 2020 at 22:54	1 KB	Terminll script
Documents	License.rtf	24 Feb 2020 at 22:54	13 KB	RTF Document
C Downloads	Python Documentation.html	1 Apr 2020 at 01:02	98 bytes	Alias
	🛔 Python Launcher	1 Apr 2020 at 01:02	269 KB	Application
	ReadMe.rtf	24 Feb 2020 at 22:54	3 KB	RTF Document
Pictures	Update Shell Profile.command	24 Feb 2020 at 22:54	3 KB	Terminll script
-				

For the installed version on a PC the Python folder should be on the C drive (PC) in the Programs folder, double click the IDLE (highlighted in blue):

Videos	^	Name	Date modified	Туре	Size
This PC		🧀 IDLE (Python 3.7 64-bit)	28/03/2020 19:56	Shortcut	3 KB
3D Objects		Python 3.7 (64-bit)	28/03/2020 19:56	Shortcut	2 KB
Desktop		🔗 Python 3.7 Manuals (64-bit)	28/03/2020 19:56	Shortcut	1 KB
Documents		Python 3.7 Module Docs (64-bit)	28/03/2020 19:56	Shortcut	3 KB

This will open a Python Shell and by clicking in File will give a pull down on which the Open (Ctrl+O) can be selected followed by selection of the AntisensePeptide.py file and then clicking open:

Ś.	IDLE	File	Edit	Shell	Debug	Options	Window	Help	
Python 3.8.2 Shell									
Pythor [Clang Type >>>	n 3.8.2 g 6.0 ( "help",	(v3.8 clang- "copy	.2:7b3; 600.0.! right"	ab5921f, 57)] on , "credi	Feb 24 darwin its" or "	2020, 17:5 license()"	2:18) for more	information.	

	Python 3.7.7 She	ell			×	
F	ile Edit Shell Deb	ug Options Wi	indow Help			
	New File	Ctrl+N	(:d7c567b08f, Mar 10 2020, 10:41:24) [MSC v.1900	64	bit	^
	Open Open Module Recent Files	Ctrl+O Alt+M	"credits" or "license()" for more information.			
;	Module Browser Path Browser	Alt+C	-			
	Save	Ctrl+S				
	Save As	Ctrl+Shift+S				
	Save Copy As	Alt+Shift+S	_			
2	Print Window	Ctrl+P				
d	Close	Alt+F4				
	Exit	Ctrl+Q				



### (i) Mac:

Ś	IDLE	File	Edit	Shell	Debug	Options	5 W	/indow	Help				
1		0				F	Pytho	n 3.8.3 s	Shell				
< :		• E	ť			Doc	umen	its		٥	Q Sear	ch	
Favorite Favorite Control Favorite R R R R R R R C C C C C C C C C C C C C	es Dropbox decents applicatic Desktop Documen Downloac DneDrive Cloud Dri uns letwork	ons ts Is		Antisens Custom Outlook Website	ePeptide.p Office Tem Files	y plates		##0rig Modifi (2006) Biolog Harris g=inpu s=inpu i=s.re '').re	nal code ed by Ha Anti-se y Protoc, wiley, t("Input place('2 place('2 place('2 eplace(' eplace(' eplace('4 eplace('4 eplace('4 eplace('4 eplace('4 eplace('4 eplace('4 eplace('4 eplace('4 eplace('4)	<pre>by Jor rrison nnse Pee cols, Ec Londor name: coding ', ''). ', ''). ', '') ', '') C', 'c' U', 't' ptide. KB</pre>	hathan C Gouldin S R Milton, bai otides. Protoco' ds D. Rickwood, n, pp 353-358. ") g mRNA: ") .replace('0', ' .replace('3', ' .replace('6', ' .replace('6', ' ).replace('4', ' ).replace('4', ' ).replace('4', ' ).replace('4', '	ng, Adaptec sed on Mill ls 6.39 In J. Graham ').replace( ').replac	d to Py3 and ton, N.G.N. Cell & J.R. ('1', ('4', ('7', ('A', ace('T', ace('\n', ace('\n',
					Filter:	Pytho	n files	s (.py, .py	/w)		3		
Opt	tions											Cancel	Open

### (ii) PC:





(15b) After opening AntisensePeptide.py a new window will appear, same for Mac and PC so from here onwards the protocol uses images from a Mac:



In the new window the Run function needs to be clicked and then the Run module F5 selected which will go to the following where the name on the target protein that the antisense peptides are generated against can be entered:



Once typed name hit return and will bring up following (nsp11 in this example):



The mRNA sequence used in the program needs to be in the format of a single word with no line breaks or paragraph marks (" $\P$ ") in the sequence (section 2d (page 10 above).

Then paste in mRNA sequence, for this example the sequence used is TCAGCTGATGCACAATCGTTTTTAAACGGGTTTGCGGTG, which is the nsp11 mRNA sequence and gives the following output:



In a very rare number of cases the a, t, c or g residues in the mRNA sequence could be replaced by an "n". This will cause an UNKNOWN to show in the peptide sequences which should be replaced by an X.

Where there is an \* at the start (SCN, AS35CN and AS53CN) or end (SNC, AS35NC, AS53NC) of a sequence this is where the STOP codon was in the mRNA and can be deleted form the sequences used to run BLAST searches. If there is an \* or an UNKNOWN in the middle of a sequence this indicates a problem with the mRNA used as these should only be at the end of coding sequences. Suggests a need to repeat section 2a-2c (see pages 5-10 above) to get the correct CDS mRNA component, particularly check that section 2c (see pages 8-10 above) to create an mRNA sequence that is a single word has been completed properly.

(15c) Copy the text from Input name down to the end of the AS53CN sequence and paste into a word document, save the Python outputs file with suitable name. These are the sequences that will be used for BLAST searches in sections 4 (see pages 14-16 above) and 5 (pages 17-19 above).



# 16: Appendix 3 - Manual antisense peptide generation

- (16a) The AntisensePeptide.py when run online in Python (see section 3 pages 11-13) performs a series of tasks with the mRNA sequence. As an alternative method where Python cannot be installed or run online the antisense sequences can be generated manually (Milton 2006).
- (16b) Initially, taking the mRNA sequence as a single string of text a space needs to be inserted between the 3<sup>rd</sup> and 4<sup>th</sup> base to convert the text into the coding triplet. This then needs to be repeated for the whole of the mRNA sequence. (in this example a short sequence peptide for nsp11 and its corresponding mRNA sequence has been used):

### mRNA: tcagctgatgcacaatcgtttttaaacgggtttgcggtg

mRNA triplets: tca gct gat gca caa tcg ttt tta aac ggg ttt gcg gtg

(16c) To do this for a long sequence this can alternatively be carried out using find and replace. The cursor should be moved to the start of the first codon (often ATG) and then find and replace used as follows. Using the find and replace command in word as above select Any Letter, then repeat until three Any Letters are selected.

#### PC

Find what:	s^ / / / July Digit	~	
	Any Letter		AaBbCcC AdD
Replace with:	Ca <u>r</u> et Character § Section Ch <u>a</u> racter 1 P <u>a</u> ragraph Character Col <u>u</u> mn Break	~	Heading 2 Title V
<< Less	E <u>m</u> Dash E <u>n</u> Dash	Replace All Eind Next Cancel	
Search; All	Fiel <u>d</u>		SC GTA CTT CTT
Match case Find whole w	<u>F</u> ootnote Mark Graph <u>i</u> c	Match prefig	CA ACG GTT TAC AT CTT CTG GTC
Use wildcard	ds Manual <u>L</u> ine Break (Er Manual Page Brea <u>k</u>	Ignore punctuation characters	
Find all word	d fi Nonbreaking <u>S</u> pace <u>O</u> ptional Hyphen Section <u>B</u> reak	☐ Ignore <u>w</u> hite-space characters	
Replace	White Space		
F <u>o</u> rmat *	Special • No Formatting		



### Mac

- 🖲 🔵 ● AutoSave ● 📭 🏠 🖨 🗲 🗸 🖑 🖶 ≑	Document1
Home Insert Draw Design Layout References	Mailings Review View EndNote X9 Acrobat 🖓 Tell me
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	ABBCCDdEe Title Sub
	<u>, , , </u> , <u>,</u> μ, <u>,</u> , <u>,</u> μ, μ, <u>,</u> μ, μ, <u>,</u> μ,
FIND AND REPLACE	
Paragraph Mark Tab Character Any Character Any Digit Any Letter	ATGTACTCATTCGTTTCGGAAGAAACAGGTACGTTAATAGTTAATAGCGTACTTCTTTTCTTGCTTTCG

(16d) In the replace box use the Special tab and select "Find What Text" and then put a space after that.

ind and Replace				?	×
Fin <u>d</u> Replace	<u>G</u> o To				
Find what: ^\$^\$^	\$				~
Replace with:	<u>P</u> aragraph Mark <u>T</u> ab Character				~
<< Less	Ca <u>r</u> et Character § Section Ch <u>a</u> racter	Replace <u>A</u> ll	Eind Next	Cancel	
Search Options Search: All Match case Find whole wo Use wildcards	1 P <u>a</u> ragraph Character <u>C</u> lipboard Contents Col <u>u</u> mn Break E <u>m</u> Dash E <u>n</u> Dash	□ Ma □ Ma	tch prefi <u>x</u> Ich suffix		
Sounds like (Er	Eind What Text Manual Line Break Manual Page Brea <u>k</u> Nonbreaking <u>H</u> yphen Nonbreaking <u>S</u> pace	Ign	ore punctuation ch ore <u>w</u> hite-space ch	naracter <u>s</u> naracters	
Replace	Optional Hyphen				
Format •	Special - No Formatting				

PC





(16e) This will generate a sequence as follows, which should be saved and then a copy saved:

 $\begin{array}{c} \texttt{ATG} \cdot \texttt{TAC} \cdot \texttt{TCA} \cdot \texttt{TTC} \cdot \texttt{GTT} \cdot \texttt{TCG} \cdot \texttt{GAA} \cdot \texttt{GAA} \cdot \texttt{ACA} \cdot \texttt{GGT} \cdot \texttt{ACG} \cdot \texttt{TTA} \cdot \texttt{ATA} \cdot \texttt{GTT} \cdot \texttt{AAT} \cdot \texttt{AGC} \cdot \texttt{GTA} \cdot \texttt{CTT} \cdot \texttt{CTT} \cdot \texttt{TTT} \cdot \texttt{CTT} \cdot \texttt{GCG} \cdot \texttt{GTA} \cdot \texttt{TTC} \cdot \texttt{TTG} \cdot \texttt{CTA} \cdot \texttt{GTC} \cdot \texttt{ACA} \cdot \texttt{CTA} \cdot \texttt{GCC} \cdot \texttt{ATC} \cdot \texttt{CTT} \cdot \texttt{ACT} \cdot \texttt{GCG} \cdot \texttt{CTT} \cdot \texttt{CGA} \cdot \texttt{TTG} \cdot \texttt{TGT} \cdot \texttt{GCG} \cdot \texttt{TTC} \cdot \texttt{TGC} \cdot \texttt{ACA} \cdot \texttt{CTA} \cdot \texttt{GCC} \cdot \texttt{ATC} \cdot \texttt{CTT} \cdot \texttt{ACT} \cdot \texttt{GCG} \cdot \texttt{CTT} \cdot \texttt{CGA} \cdot \texttt{TTG} \cdot \texttt{TGT} \cdot \texttt{GCG} \cdot \texttt{TAC} \cdot \texttt{TGC} \cdot \texttt{ACA} \cdot \texttt{CTA} \cdot \texttt{GCC} \cdot \texttt{ATC} \cdot \texttt{CTT} \cdot \texttt{ACT} \cdot \texttt{GCG} \cdot \texttt{CTT} \cdot \texttt{CGA} \cdot \texttt{CTG} \cdot \texttt{ACA} \cdot \texttt{CTA} \cdot \texttt{GCC} \cdot \texttt{ATC} \cdot \texttt{CTT} \cdot \texttt{ACT} \cdot \texttt{GCG} \cdot \texttt{CTT} \cdot \texttt{CGA} \cdot \texttt{CTC} \cdot \texttt{CTC} \cdot \texttt{ACA} \cdot \texttt{CTC} \cdot \texttt{CTC} \cdot \texttt{CTC} \cdot \texttt{ACA} \cdot \texttt{CCA} \cdot \texttt{ACG} \cdot \texttt{GTC} \cdot \texttt{TAC} \cdot \texttt{CCT} \cdot \texttt{CCT} \cdot \texttt{CCT} \cdot \texttt{ACA} \cdot \texttt{CCA} \cdot \texttt{CCT} \cdot \texttt{CTG} \cdot \texttt{CTT} \cdot \texttt{CTG} \cdot \texttt{CT} \cdot \texttt{CT}$ 

(16f) The sense coded target protein sequence should be available, however, if it needs to be generated from the mRNA this can be carried out using find and replace in word and replacing each triplet with the single letter code for the respective amino acids derived from the following tables for Sense strands:

RNA	Sense	RNA	Sense	RNA	Sense	RNA	Sense
triplet	Amino Acid						
AAA	K	CAA	Q	GAA	E	TAA	*
AAC	N	CAC	Н	GAC	D	TAC	Y
AAG	K	CAG	Q	GAG	E	TAG	*
AAT	N	CAT	Н	GAT	D	TAT	Y
ACA	Т	CCA	Р	GCA	A	TCA	S
ACC	Т	CCC	Р	GCC	A	TCC	S
ACG	_ <b>T</b>	CCG	P	GCG	A	TCG	S
ACT	Т	CCT	Р	GCT	A	TCT	S
AGA	R	CGA	R	GGA	G	TGA	*
AGC	S	CGC	R	GGC	G	TGC	С
AGG	R	CGG	R	GGG	G	TGG	W
AGT	S	CGT	R	GGT	G	TGT	С
ATA	I	CTA	L	GTA	V	TTA	L
ATC	I	CTC	L	GTC	V	πс	F
ATG	M	CTG	L	GTG	V	TTG	L
ATT	I	СТТ	L	GTT	V	TTT	F

The resultant sense sequence is the mRNA encoded peptide in the N-terminus to C-terminus direction (SNC). From the example of nsp11 above this would be:

SNC - nsp11 = SADAQSFLNGFAV



(16g) To generate the AS35NC antisense peptide sequence each triplet with the single letter code for the respective amino acids derived from the following tables for Antisense strand read in the 3'-5' for each amino acid:

RNA	Antisense	RNA	Antisense	RNA	Antisense	RNA	Antisense
triplet	Amino Acid						
AAA	F	CAA	V	GAA	L	TAA	*
AAC	L	CAC	V	GAC	L	TAC	М
AAG	F	CAG	V	GAG	L	TAG	*
AAT	L	CAT	V	GAT	L	TAT	I
ACA	С	CCA	G	GCA	R	TCA	S
ACC	W	CCC	G	GCC	R	TCC	R
ACG	С	CCG	G	GCG	R	TCG	S
ACT	Х	CCT	G	GCT	R	TCT	R
AGA	S	CGA	А	GGA	Р	TGA	*
AGC	S	CGC	А	GGC	Р	TGC	Т
AGG	S	CGG	А	GGG	Р	TGG	π
AGT	S	CGT	А	GGT	Р	TGT	Т
ATA	Y	CTA	D	GTA	Н	TTA	N
ATC	Х	CTC	E	GTC	Q	TTC	К
ATG	Y	CTG	D	GTG	Н	TTG	N
ATT	Х	СТТ	E	GTT	Q	Π	К

The resultant output is the mRNA encoded Antisense 3'-5' peptide in the N-terminus to C-terminus direction (AS35NC). From the example of nsp11 above this would be:

AS35NC - nsp11 = SRLRVSKNLPKRH

(16h) To generate the AS53NC antisense peptide sequence each triplet with the single letter code for the respective amino acids derived from the following tables for Antisense strand read in the 5'-3' for each amino acid:

RNA	Antisense	RNA	Antisense	RNA	Antisense	RNA	Antisense
triplet	Amino Acid						
AAA	F	CAA	L	GAA	F	TAA	*
AAC	V	CAC	V	GAC	V	TAC	V
AAG	L	CAG	L	GAG	L	TAG	*
AAT	I	CAT	М	GAT	I	TAT	I
ACA	С	CCA	W	GCA	С	TCA	Х
ACC	G	CCC	G	GCC	G	TCC	G
ACG	R	CCG	R	GCG	R	TCG	R
ACT	S	CCT	R	GCT	S	TCT	R
AGA	S	CGA	S	GGA	S	TGA	*
AGC	А	CGC	Α	GGC	Α	TGC	А
AGG	Р	CGG	Р	GGG	Р	TGG	Р
AGT	Т	CGT	Т	GGT	Т	TGT	Т
ATA	Y	CTA	Х	GTA	Y	TTA	Х
ATC	D	CTC	E	GTC	D	TTC	E
ATG	Н	CTG	Q	GTG	Н	TTG	Q
ATT	N	СТТ	K	GTT	Ν	Π	К



The resultant output is the mRNA encoded Antisense 5'-3' peptide in the N-terminus to C-terminus direction (AS53NC). From the example of nsp11 above this would be:

AS53NC - nsp11 = XSICLRKXVPKRH

(16i) The protein databases are always N-terminus to C-terminus orientation; however, proteins may interact with the binding site having one protein in the N-terminus to C-terminus orientation and the other in the C-terminus to Nterminus orientation. Hence the need to search the C-terminus to N-terminus orientation antisense peptides. To generate the SCN, AS35CN and AS53CN sequences they can be reversed using <u>http://www.upsidedowntext.com/</u>.

Type text, words, letters, or symbols here:	26
DSGYEVHHQKLVFFAEDVGSNKGAII	
Text Effects: Backwards Effect (Reverses text) Upside Down Effect (Flips text)	Post to:
Last Pass Never forget a passwork	

Using the sequence DSGYEVHHQKLVFFAEDVGSNKGAII as an example will generate an output of IIAGKNSGVDEAFFVLKQHHVEYGSD:

The overall output of sequences for the nsp 11 example should be:

SNC - nsp11 = SADAQSFLNGFAV SCN - nsp11 = VAFGNLFSQADAS AS35NC - nsp11 = SRLRVSKNLPKRH AS35CN - nsp11 = HRKPLNKSVRLRS AS53NC - nsp11 = XSICLRKXVPKRH AS53CN - nsp11 = HRKPVXKRLCISX



## **17: Acknowledgements**

The Antisense Peptide protocol has been constantly developed, initially from studies carried out by the author Dr Nat Milton during the original set up of NeuroDelta Ltd in 2003. The methods are based on his academic research carried out at Leeds Trinity University, Leeds Beckett University, the University of Westminster, the University of Roehampton, and University College London). The author would like to thank collaborators Dr Maria Ashioti, Tim Barnes, Dr Amrutha Chilimuri, Prof J. Robin Harris, Dr Farideh Javid, Dr Neema Mayor, Amanda Nercessian, Dr Mark Odell, Prof Jolanta Opacka-Juffry, Dr John Rawlinson, Dr Eridan Roche-Ferrera, Sabrina Werner, Prof Anthony F. Winder, and Dr Tamana Zemaryalai plus all of the other staff, students and collaborators who have supported this development and contributed to the research undertaken using these methods. I would also like to thank Jonny Goulding and Harrison Milton for the work on the Python Script that has speeded up the generation of antisense peptides dramatically plus Alexandra Fragkoulaki for the Excel sheets for Molecular Recognition.

The Antisense Peptide protocol methods have been tested by BSc, MSc and PhD students as part of their research projects plus taught Bioinformatics sessions at Leeds Beckett University, the University of Westminster, the University of Roehampton, and University College London. The author would like to thank all these students for their hard work and dedication during their studies, plus would like to congratulate them on their graduations and subsequent career progression. The author would particularly like to thank Dr Naghmeh Nikkheslat for encouraging the author to start running Bioinformatics undergraduate research projects and being the initial trainee on this programme whilst she was an undergraduate at the University of Roehampton.

The financial support by NeuroDelta Ltd and a U.K. Department of Trade and Industry Grant (LOT/0311684) that supported the initial work on the development and application of Antisense Peptides methods described in this protocol is gratefully acknowledged.

