## **V.V.VANNIAPERUMAL COLLEGE FOR WOMEN**

(Belonging to Virudhunagar Hindu Nadars) An Autonomous Institution Affiliated to Madurai Kamaraj University, Madurai Re-accredited with 'A' Grade (3ª Cycle) by NAAC VIRUDHUNAGAR – 626 001



## (UNDER DBT STAR COLLEGE SCHEME)

No HRD-11011/163/2020-HRD-DBT Department of Biotechnology, Ministry of Science and Technology MHRD, New Delhi

K. Sudha Rameshwari

# BIOINFORMATICS LABORATORY MANUAL

## K.Sudha Rameshwari

## C April,2023@authors

All rights reserved. No part of this publication may be reproduced or transmitted any form or by any means without permission of the author. Any person who does any unauthorised act in relation to criminal prosecution and civil claims for damage.

## V.V.VANNIAPERUMAL COLLEGE FOR WOMEN



(Belonging to Virudhunagar Hindu Nadars) An Autonomous Institution Affiliated to Madurai Kamaraj University *Re-accredited with 'A' Grade (3<sup>rd</sup> cycle) by NAAC* **VIRUDHUNAGAR – 626 001** (TAMIL NADU)



## DBT STAR COLLEGE SCHEME Department of Biotechnology, Ministry of Science and Technology Government of India, New Delhi

Chairman & Principal : Dr.(Tmty.) S.M. MEENA RANI, M.S.C.M.Phil., P.G.D.C.A.Ph., Contact number: 9498088703 e-mail:smmeenarani@gmail.com Coordinator & Member Secretary : Dr.(Tmty.) M.TAMILSELVI, M.S.C.M.Phil., Ph.D., Contact Number: 9894883106 e-mail:tamilasc.selvi08@gmail.com

## FOREWORD

This Lab Manual on "**BIOINFORMATICS**" is prepared in accordance with the updated syllabus under DBT Star College Scheme sponsored by the Department of Biotechnology, Ministry of Science and Technology, MHRD, New Delhi to fulfill the needs of students.

This manual could enable the students to retrieve protein, nucleic acid sequence using different tools and to interpret the results. The protocols included in this manual elaborate the step by step procedure with URL link and illustrations.

We thank the **Department of Biotechnology**, **Ministry of Science and Technology**, **MHRD**, **New Delhi** for providing a good opportunity under Star College Scheme (No.HRD11011/163/2020-HRD-DBT Dt.24.8.2020).

We hope this manual will definitely meet out the student's needs to perform the experiments that enhance their research skills in the drug development.





Member Secretary/Coordinator

Chairman/Principal

#### PREFACE

Bioinformatics is an evergreen, emerging interdisciplinary field in life science. It provides a medium for the exchange of information in the fields of computational molecular biology and the post-genome era, with an emphasis on the documentation of large data sets and databases that allow the progress of biomedical research in a significant manner.

This lab manual presents a collection of twenty practical exercises, aimed at providing standard protocols to access nucleic acid and protein sequence databases, perform sequence alignments, predict secondary structure, and visualize proteins. This manual introduces the theory and provides a systematic procedure to facilitate students to carry out the practical exercises in an easy manner. Overall understanding NCBI resources, accessing biological sequences from GenBank, performing sequence alignments, and using protein visualization tools will become easy if the students and researchers use this manual effectively. I assure that this will become a handy tool and motivation factor for the basic analysis of sequences.

When the readers finish practicing the exercises in this manual, they will possess the knowledge to use the various online tools available for processing the biological data and will have a platform to develop their skills in handling bioinformatics techniques in the future.

> P. Sugapriya Menaga Reviewer Assistant Professor of Biotechnology, ANJAC. Sivakasi.

The field of bioinformatics, which applies both information technology and biological science make to prediction about biological processes using bioinformatics tools. Moreover bioinformatics is an essential part of biological research and has applications in different practical fields.

This bioinformatics manual helps the beginners to learn step-by-step procedure that will make the learners to follow the experiments easily. This manual covers the curriculum for the Biotechnology, Microbiology and Biochemistry students at the undergraduate level.

I would like to express my heartfelt gratitude to the reviewer, Mrs.P.Sugapriya Menaga, Assistant Professor, Department of Biotechnology, ANJAC, Sivakasi for her useful suggestions to improve the quality of this manual.

#### K.Sudha Rameshwari

Assistant Professor, Department of Biochemistry, V.V.Vanniaperumal College for Women, Virudhunagar-626001, Tamilandu, India.

#### FOREWORD

## BIOINFORMATICS LABORATORYMANUAL

## LIST OF PRACTICALS

S.No.	TITLE	Page No.
1.	Introduction to Bioinformatics	1
2.	Retrieval of Nucleotide sequence from GenBank	4
3.	Retrieval of Protein sequence from GenBank	8
4	Sequence Similarity Search using BLASTN	14
5.	Sequence Similarity Search using BLASTP	22
6.	Accessing Structural Database and Download the Protein Structure	28
7.	Working with Ensembl	32
8.	Multiple Sequence Alignment	36
9	Predicting Physiochemical properties of protein sequence	42
10	Predicting Peptide mass of protein sequence	48
11	Predicting cleavage site of protein sequence	54
12	Predicting secondary structure using SOPMA tool	58
13	Predicting secondary structure using CFSSP tool	62
14	Predicting transmembrane region of Protein sequence	66
15	Predicting hydrophilic region in the protein sequence	70
16	Detecting alignment of repeats in protein sequences	74
17	Predicting the peptide structure for the given protein sequence	76
18	Conversion of nucleotide sequence into Protein sequence	78
19	Molecular Visualization using RasMol	82
20	Measurement of bond length in protein structure using RasMol	94
21	Measurement of bond angle in protein structure using RasMol	99
22	Web links	101

#### INTRODUCTION TO BIOINFORMATICS

Bioinformatics is a newly emerging scientific discipline for the computational analysis and storage of biological data. Bioinformatics is the field in which biology, computer science and information technology merges into a single discipline for managing and analyzing biological data using advanced computing techniques.

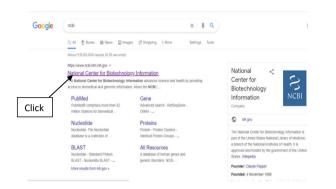
## **OBJECTIVES**

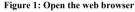
- To organize data, to access existing information and to submit new entries as they are produced.
- To develop tools and resources that aid in the analysis of data
- To analyze the data and interpret the results in a biologically meaningful manner

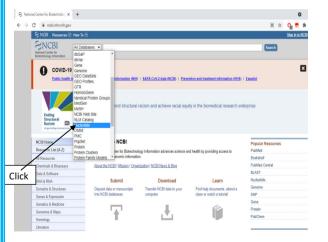
## SCOPE OF BIOINFORMATICS

Biological and medical labs use methods that produce extremely large data sets, which cannot be analyzed for instance sequencing human genomes. Thus, modern biological and medical research and development cannot be carried out without bioinformatics. In addition, bioinformatics plays an important role in biomedical research. Research in the area of genetic diseases and medical genomics is rapidly developing and the future of personalized medicine depends on the application of bioinformatics approaches.

This course focuses on employing existing bioinformatics resources - mainly web-based programs and databases - to access the wealth of data to answer questions relevant to the average biologist, and is highly hands-on. Different types of career opportunities are available for the students of bioinformatics like Scientific Curator, Gene Analyst, Protein Analyst, Phylogenitist, Research Scientist / Associate, Data Base programmer, Bioinformatics software developer, Computational biologist, Network Administrator / Analyst, Structural Analyst, Molecular Modeler, Biostatistician, Biomechanics, Database programmer, Cheminformatician, Pharmacogenetician, Pharmacogenomics etc.









## RETRIEVAL OF NUCLEOTIDE SEQUENCE FROM GENBANK

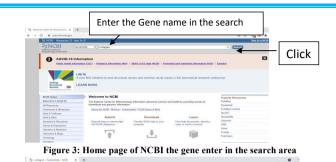
## AIM

To retrieve the gene from Genbank and to save the sequence in FASTA format

## **INTRODUCTION**

Entries of nucleotide sequence are stored in Genbank nucleotide database. Each entry contains complete information about the particular gene. Genbank website is accessible for anyone who is interested in getting the entry for research.

- Type NCBI in the web browser and click search, click National Center for Biotechnology Information, it directs to the URL : <u>https://www.ncbi.nlm.nih.gov/</u> (Figure 1) OR
- Type the URL <u>https://www.ncbi.nlm.nih.gov/</u> directly in the space for address in the address bar and press the enter key.
- NCBI homepage will appear.
- Click the All Databases drop –up menu and drag the bar and select nucleotide (Figure 2)
- Search list will be displayed, click the suitable accession number or any gene of interest (figure 4)



\* 0 \* \* COVID-19 Information whic health information (CDC) | Research information (NPD | SARS-CoV-2 data (NCB) | Prevention and treatmen COLIAA1 (COLLAGEN) collagen type II alots 1 chain in the Gene databas lagen inference sequences Transcript (7). Protein (7) Selected 1 The stars where I at \$555 March 1 had been Galus galus collagen a2 mRNA gene partial sequence; and precollagen alpha-2, collagen alphagen alona 2, collagen a2, and collagen alona 2 denes, partial co DNA 60.3 OF 1015525050 n collagen gang X and C elegans (permatode) collagen 2 (col-2) gene, complete ods Figure 4: search list page Click CLICK FAST Figure 5: Genebank entry format suum collagen gene, 3' end Figure 6: FASTA sequence in new window

- Click the gene of interest/accession number (Figure 4)
- Type Collagen in the search area and click (Figure 3)
- A new window will appear and shows the entry of the collagen gene in detail (Figure 5)
- Click the FASTA, FASTA sequence appears in the new window (Figure 6)
- Copy the FASTA sequence and paste it in note pad
- Save it for further investigation.

#### RESULT

The collagen gene was retrieved and saved as FASTA format in notepad.

## OUTCOME

Students acquire the knowledge to retrieve any gene from Genbank. From the retrieved sequence flat file, details of sequences, submitter's details, biological significance, and the scientific name and taxonomy of the organism is understood. A feature table shows characteristics that indicate coding regions, transcription units, mutation locations, etc. Retrieved gene sequence is useful for gene analysis. It can be compared with other sequences to determine which gene is mutant. It enlightens the diagnosis of hereditary disease.

	SNCBI National Center for Biolechnology Information	All Databases V NCBI Web Site NLM Catalog	•			Search
lick	Ending	Nucleotide OMIM PMC PopSet Protein Protein Clusters	end structural	racism and achieve racial equ	ity in the biomedical research en	terprise.
	Racism State	Protein Family Models PubChem BioAssay				
	NCBI Home	PubChem Compound PubChem Substance	NCBI			Popular Resources
	Resource List (A-Z)	PubMed	tter for Biotechnol	logy Information advances science ar	d health by providing access to	PubMed
	All Resources	SRA	enomic informatio			Bookshelf
	Chemicals & Bioassays	Structure	I Mission   Orga	nization   NCBI News & Blog		PubMed Central
	Data & Software	Taxonomy ToolKit				BLAST
	DNA 8 RNA	ToolKitAll	bmit	Download	Learn	Nucleotide
	Domains & Structures		* manuscripts	Transfer NCBI data to your	Find help documents, attend a	Genome
	Genes & Expression	into NC8I e	databases	computer	class or watch a tutorial	SNP
	Genetics & Medicine		_	~	2	Gene
	Genomes & Maps					Protein
	Homology		Τ	<u> </u>		PubChem
	Literature					
	Proteins					NCBI News & Blog
	Sequence Analysis		Develop	Analyze	Research	Codeathon from the Couch LINCOWS
	Taxonomy		APIs and code	Identify an NCBI tool for your	Explore NCBI research and	North Texas Workshaps and Codeathon e W 2021
	Tesising & Tetasisle		hulld and rations	data analusis task	callsbardin emiote	2021

Figure 7: Home page of NCBI, Selection of Protein

Protein Pro	tein V (COLLAGEN) AND "Homo sapiens" Create alert Advanced	C Search Hel
Species clear J Animals (3.854)	Summary + 20 per page + Sort by Default order + Send to	- Filters: <u>Manage Filters</u>
Customize Source databases	See the <u>results of this search (731 items</u> ) in our new identical Protein Groups database.	Results by taxon Top Organisms [Tree]
PDB (833) RefSeq (1,247) UniProtKB / Swits-Prot (323)	Items: 1 to 20 of 3854	Homo saplens (3854)
Customize	Fint - Page 1 of 193 Next+ Last-     Fint - Page 1 of 193 Next+ Last-	Find related data
Genetic compartments Mitochandrian (13)	collagen (Homo sapiens)	Database Select 🗸
Sequence length	1. 1678 aa protein Accession: BAAR409.1 Gr. 456538 Nacketelde PaaMed Texanomy	
Custom range Molecular weight	GenPast Identical Postelins EASTA Graphics	Search details
Custom range	collagen, parial (Homo sapiens) 2. 718 as protein	COLLAGEN[All Fields] AND "Homo sapiers" [porgn] AND animals[filter]
Custon range	Accession: AAA3958.1 GI: 292352 Nucleotice: PubMed Taxonomy	
Ck Revision date Custom range	GenPest Identical Proteins EASTA Graphics	Search See more
Clear at	<u>COL2A1.gental [Homo sapiens]</u> 1160 as protein	
Show additional filters	<ul> <li>Trolo da proteini Accession: CAA34683.1 (0: 30041 Naciedite PubMed Texanony</li> </ul>	Recent activity Activate Witer 01, Sea
	GenPest Identical Proteins EASTA Graphics	Q (COLLAGEN) AND "Homo sapiens" [porgn] AND (animate/Black) (1955)

## **RETRIEVAL OF PROTEIN SEQUENCE FROM GENBANK**

## AIM

To retrieve the protein from Genbank and to save the sequence in FASTA format

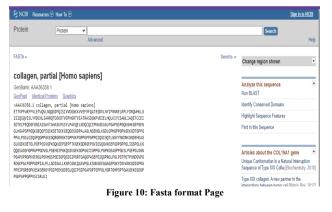
## **INTRODUCTION**

Entries of protein sequence are stored in Genbank database. Each entry contains complete information about a particular protein. Genbank website is accessible for anyone who is interested in getting the entry for research.

- Type NCBI in the web browser and click search, it shows National Center for Biotechnology Information, click National Center for Biotechnology Information, it directs to the url : <u>https://www.ncbi.nlm.nih.gov/</u> or home page of NCBI (Figure 7) Or
- Type the url: <u>https://www.ncbi.nlm.nih.gov/</u> web/Genbank/ directly in the address bar and press the enter key
- NCBI homepage will appear
- Click the All Databases drop –up menu, and drag the bar to select protein (Figure 8)

Protein	Protein V Advanced		Search Help
GenPept +		Send to: +	Change region shown
<b>collage</b> GenBank: A	n, partial [Homo sapiens] AA36358.1		Customize view 💌
Identical Prot	eins FASTA Graphics		Analyze this sequence
ACCESSION	AAA36358 718 as linear PRI 12-JUN-1993 collagen, partial [Homo sapiens]. AAA36358		Identify Conserved Domains Highlight Sequence Features
VERSION DBSOURCE KEYNORDS SOURCE	AAA36358.1 locus HAMMCOLL accession <u>112347.1</u> Homo speciens (human)		Find in this Sequence
	Homo_saciens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutzleostomi; Nammalia; Euternia; Euarchontogilre; Primates; Haplornini; Catarhini; Monifidae; Homo.		Articles about the COL19A1 gene  Unique Conformation in a Natural Interruption Sequence of Type XIX Collar [Elochemistry. 2018]
	1 (residues 1 to 718) Myers,J.C., Sun,M.J., D'Ippolito,J.A., Jabs,E.W., Neilson,E.G. and Dion,A.S.		Type XIX collagen: A new partner in the interactions between tumor cel [Matrix Biol. 2017]
TITLE JOURNAL	Human CDHA clones transcribed from an unusually high-molecular-weight RNA encode a new collagen chain Gene 123 (2), 211-217 (1993)		Type XIX collagen purified from human umbilical cord is characterized by multi [J Biol Chem. 2003]
PUBHED COMMENT FEATURES	7915783 Method: conceptual translation. Location/Qualifiers		See al
source			Activate Windows Pathways for the COL 19A1 gene activate
	/organise Homo sapiers' /db_wref="taxon: <u>9606"</u> /map+"chronosceme 6"		Collagen chain trimerization

#### Figure 9: Collgen gene Genbank entry page



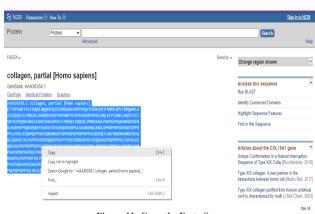
- Type collagen in the search area and press enter key or click search
- Search list will be displayed (Figure 8)
- Select the suitable accession number that describes collagen
- Click the accession number of collagen
- A new window will appear and shows the entry of the collagen gene (Figure 9)
- In display format, in place of summary choose "FASTA"
- Only sequences will be shown (Figure 10)
- Copy the sequence (Figure 11) and paste it on note pad (Figure 12) and save it for the further investigation

### RESULT

The Protein sequence was retrieved and saved as FASTA format in notepad.

## OUTCOME

Students master the technique of sequence retrieval which is the basic step in bioinformatics. From the retrieved sequence flat file, details of sequences, submitter's details, biological significance, and the scientific name and taxonomy of the organism are understood. Sequence retrieval is essential for the analysis of primary, secondary and tertiary structure of any protein sequence.



#### Figure 11: Copy the Fasta Sequence

- 0

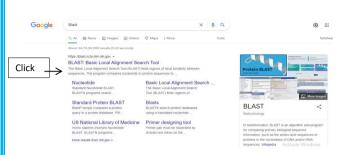
Notepad For Windows 10

#### >AAA36358.1 collagen, partial [Homo sapiens]

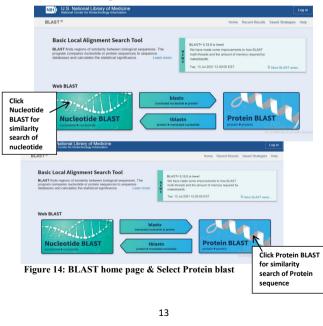
ETTVPFWRFFVLETWQVLNQQNIPQISIVVDGGKK/VEFMFQATEGDVLNYIFRNRELRPLFDRQWHKLG ISIQQQVISIYMDCNLIARRQTDEKDTVDFHGRTVATRASDGKPVDIELHQLKYCSANLIAGETCCEI SDTKPEDQDFGKNLSSWYTAHAKMSSYLPAKCURQQCQFINKBCAGLPGAPSEPGCKGHKEBEGA GLHGAPGFPGQKGEQGFEGSKGETGEKGEQGEKGDPALAGLNGENGLKGDLGPHGPPGFKGEKGDTGPPG PPALPCSLGIQCPCPPGKEGQKRGRKTCFPGPKPCPGPGPGIQHIQTLGGYMLKDNKGNDEHEAG GLKGDKGETGLPGFPGSVGPKGQKGPGEPFEKGEKGDFGPGVGSGGVKGEPGDPGPGIGSFGLKG QQCSAASMGPRCPPGDVGLPGEHGIPGKQGIKGEKGDPGGIPGPGLPGFKGEKGPGCPGDVG PGAPGPRGPKGERGLPGVHGSPGGIGPGKGLKGEKGDPGGIPGPGLPGFKGEKGDFGPGD PGAFGPGPFDDVGLPGEHGIPGKQGIKGEKGDFGJPGFLGPKKELGREFGLDGPR PGKGPGPCPFVSCSRLK

## B | ■ ↑↓ □ × B / U to an After Win Q to a state the second of the

Figure 12: Paste the Fasta Sequence in the note pad



#### Figure 13: Search list of BLAST



## SIMILARITY SEQUENCE SEARCH USING BLASTN

To find the similarity of sequence for the given nucleotide or protein sequence

#### **INTRODUCTION**

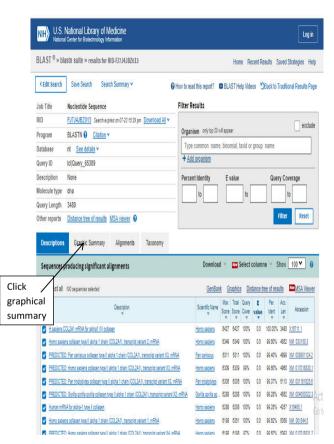
BLAST is **B**asic Local Alignment Search Tool. It is a technique for finding homology and similarity. It is a tool for searching of sequences that are similar to one another in databases. By matching the novel sequence with previously defined genes, it compares novel gene sequences with nucleotide databases. This tool focuses on identifying areas of sequence similarity. It will provide information on the structure and function of the novel sequence. Instead of using the best alignments, it looks for areas of sequence similarity. It produces ungapped alignments. It reports multiple local alignments between the query and database. It is based on an explicit statistical theory.

- Type BLAST in the web browser and click search
- Click the first link, BLAST (Figure 13)
- Go to Blast Home page (Figure 14)
- Click nucleotide blast BLASTN (Figure 15)

[		7		
Check bl		U.S. National Library on National Center for Scienchroby		Losin
<u></u>	,	BLAST <sup>®</sup> » blastn suite	Hame Recent Results Saved Strategi	s Help
۷ ۱	blastn blas	to blastr tblastr	Standard Nucleotide BLAST	
	Enter Query Se	equence	8.4 MV program worth numbers deletene unity a number party. <u>Term</u>	fastpp (totrat)
	Enter accession nu Gooccoccotosoc GA Oc. upload file Job Title	nterij); (i) (i 1931 separato) () () () 2004al-11300000000000000000000000000000000000		New columns added to the Description Table Citiz funct clurrer or Manage Columne:
	Alge two or non	Enter a descriptive title for your BLAST search 🖗	Paste the Fasta nucleotide	
	Choose Search	h Set	sequence in the search box	
		Bandard databases (m etc.) () RNA/TS databases () Genomic + transcr Nucleotice collector (mind) v	l	
	Coloral	Enter organise connece norme, binomial, or tao id. Only 22 top lace will be shown 🖗	dala (Litt oppnar)	
	Entrez Query	Sequences from type material	Courte custom database	
	Program Selec	Enter an Enter, gany in less search 🖗		
	Optimize for	Pigny similar sequences (magabilat)     Vitre desimilar sequences (bacontipuos magabilat)     Somewinat similar sequences (bacontipuos     Consent at Sinilar sequences (bacontipuos		
	BLAST	Search database Nucleotide collection (nrht) using Wegablast (Optimize for	bahi umar unumeni	Activate Windows Go to Settings to activate Windows.
		The second	And some strength	
Click	][1	Fick the box	Figure 15: BlastN Page	
			15	

Depending on the type of sequence, the programs in BLAST differs. They are

- Blastp compares an aminoacid query sequence against a protein sequence database
- Blastn compares an nucleotide query sequence against a nucleotide sequence database
- Blastx compares a nucleotide query sequence translated in all reading frames against a protein sequence database.
- **tblastn** compares a protein query sequence against a nucleotide sequence database dynamically translated in all reading frames.
- **tblastx** compares the six frame translation of a nucleotide query sequence against the six frame translation of a nucleotide sequence database.
- Paste the sequence (FASTA format) in the given box of BLASTN or paste the accession number (Figure 15)
- Tick the box "results in new window" in the last line of BLASTN Home page (Figure 15)
- Click BLAST (Figure 15)
- In the new window, format request display is shown, wait for few seconds, it leads to result page



#### Figure 16: Result window of BLASTN

- In the result page, the blast results for sequence similarity search appear as graphically and as text (descriptions, alignment)
- Click each text and find the similarity between the two sequences
- Click graphic summary (Figure 16)
- View the graphical representation of sequence alignment
- Lines in pink, red colors represent the sequence in a set of score values.
- Length of the line indicates the length of the local alignment with the query sequences. (Figure 17)

## RESULT

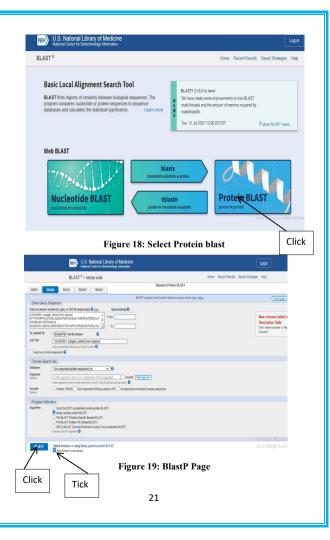
The most significant and similar sequences fetched by the blast are <u>Homo sapiens collagen type II alpha 1 chain (COL2A1),</u> transcript variant 2, mRNA,

## OUTCOME

By learning BLASTN, students obtain the skill to compare the new gene sequence with the nucleotide database by aligning the novel sequence with previously characterized gene/protein. From the comparison, students gather the functional and evolutionary clues about the structure and function of the novel sequence.

< Edit Search	Save Search Search Summary *	How to read this report? 🛛 BLAST Help Videos 🖒 Back to Traditional Results Page
Job Title	Nucleotide Sequence	Filter Results
RID	FJ7J4JBZ013 Search expires on 07-23 15:29 pm DoWnload All ♥	exclude
Program	BLASTN () Citation *	Organism only top 20 will appear
Database	nt See details *	Type common name, binomial, taxid or group name
Query ID	Id Query_65389	+ <u>Add organism</u>
Description	None	Percent Identity E value Query Coverage
Molecule type	dna	to to to
Query Length	3480	
Other reports	Distance tree of results MSA viewer 0	- Filter Reset
û hover to see ti	të të 🖡 click to show alignments	Alignment Scores <a></a> < 40 <a>40 - 50 <a>50 - 80 <a>80 - 200 <a>&gt;= 200 </a></a></a></a>
100	selected 💡	
rou deponice i		ion of the top 144 Blast Hits on 100 subject sequences





## SIMILARITY SEQUENCE SEARCH USING BLASTP

#### AIM

To find the similarity of sequence for the given protein sequence **INTRODUCTION** 

BLAST is **B**asic Local Alignment Search Tool. It is a technique for finding homology and similarity of protein sequences. It is made for the Windows platform and is employed to carry out DNA or protein similarity searches. The source is NCBI. It is a tool for searching for sequences that are similar to one another in databases. This tool focuses on identifying areas of sequence similarity. It is used to compare a novel protein sequence against protein database by aligning the novel sequence with previously characterized protein. The emphasis of this tool is to find regions of sequence similarity. It will yield functional and evolutionary clues about the structure and function of the novel sequence. It finds out patches of sequences similarity rather than best alignments. It produces un gapped alignments. It reports multiple local alignments between the query and database. It is based on an explicit statistical theory.

- Go to NCBI Blast page
- Chose the Blast program for proteins BLASTP (Figure 18)

	NIH) National Library of Medicine National Center for Biotechnology Information							Lo	g in	
	$BLAST^{\circledast} \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \$			ł	tome Ra	ecent Result	Saved S	ralegies	Help	
	< Edit Search Save Search Search Summary ◄	£0⊦	łow to read this rep	wrt? 🗖 BLAST	r Halo Vid	ns "PRari	to Tradition	al Results	Pana	
					1.20					
	Job Title gb[AAA29796.1]		Filter Results							
	RID <u>SHENYBXW016</u> Search expires on 05-09 17:19 pm D0	Milled All	Ornanism and	y top 20 will appear				en	lude	
	Program BLASTP O Citation Y		· · · ·	n name, binomia	o hivet le			0	7	
	Database nr <u>See details</u> *		+ Add organis		a, avu u	group nan	-			
	Query ID AAA29796.1		- Con voldilis							
	Description hemoglobin [Pseudoterranova decipiens]		Percent Identi	ty E val	ue		Query Cov	erage		
	Molecule type amino acid		to		to		to			
	Query Length 333								-	
	Other reports Distance tree of results Multiple alignment MSA	viewer 0					Filter	Rese	<u> </u>	
	Compare these results ag	ainst the nev	v Clustered nr d	latabase 😧	BLAST				X	
										J.
	Descriptions Graphic Summary Alignments Taxo	onomy								
	uences producing significant alignments select all 100 sequences selected		<u>GenPept</u>	Downloa Graphics			column: sults <u>M</u>		how Ignm	
			<u>GenPept</u>		Distance Max	e tree of re Total Que	sults M	ultiple al Per.	gnm Acc.	ent MSAV
☑ s	select all 100 sequences selected			Graphics Scientific Name	Distance Max Score	a tree of re Total Que Score Cow	suits M y E y value	Per. Ident	Acc.	ent <u>MSA V</u> Accessi
☑ s	Select all 100 sequences selected  Description hemostable (Pseudoteranova decident)			Graphics Scientific Name Pseudoterranov	Distance Max Score 692	e tree of re Total Que	suits M y E y value 6 0.0	Per. Ident	Acc. Len 333	ent MSAV
✓ s	select all 100 sequences selected	mi		Graphics Scientific Name	Distance Max Score 692 635	Total Que Score Cow 692 100	sults M y E y value 6 0.0	Per. Ident 100.00%	Acc. Len 333	Accessi Accessi
✓ s	Select all 100 sequences selected  Description  hemotobin Pseudoteratory decisiend  hemotobin Pseudoteratory adocisiend	<u>ani</u> ]		Graphics Scientific Name Pseudoterranov	Distance Max Score 692 635	a tree of re Total Que Score Cow 692 100' 635 969	suits M v E value 6 0.0 0.0 0.0	Ultiple al Per Ident 100.00% 95.34% 95.03%	Acc. Len 333 333 333	ent MSA V Accessi AAA29796. CAA77743. P26914.1
2 s 2 s 2 s 2 s	Select all 100 seconos selected	mal		Graphics Scientific Name Pseudoterranov. Pseudoterranov.	Distance Max Score 692 635 633 550 537	tree of re           Total         Que           Score         Cov           692         100'           633         96'           550         100'           537         95'	suits M y E t value 6 0.0 0.0 0.0 0.0 0.0 0.0	Ultiple al Per Ident 100.00% 95.34% 95.03% 79.04% 79.69%	ignm Acc. 1933 333 333 333 332	ent MSAV Accessi AAA29796. CAA77743. P26914.1 BAX39003: ASL68918.2
	select all no aquences andore Description Introductor Description Introductor Description Robins, Full-Electricity addres Plant Research Privademanou designed Introductor Research Strends Introductor Research Strends Introductor Research Strends	2021		Graphics Scientific Name Pseudoterranov Pseudoterranov Anisakis simplex Anisakis simplex Anisakis pegrefiji	Distance Max Score 692 635 633 550 537 537	tree of re           Total         Que           Score         Cov           692         100'           635         969           633         969           550         100'           537         959           537         959	suits M value 6 0.0 0.0 0.0 0.0 0.0 0.0 0.0	Uttiple al Per. Ident 100.00% 95.03% 79.04% 79.69%	333 333 333 332 332 332	ent MSAV Accessi AAA29796 CAA77743. P26914.1 BAX39003 ASL689182 BAX39002
	estek tali 100 sepanas sekend Desigtion henrotika: Panokhenana distemat henrotika: Panokhenana distemat Rotima: Edita Sekend Sekenary. Pendekerana diste henrotika: Kenaka sekend henrotika: Kenaka sekend henrotika: Kenaka sekend	203		Graphics Scientific Name Pseudoterranoy.	Distance Score 692 635 633 550 537 533	Iree of re           Total         Que           Score         Cov           692         100'           633         969           550         100'           537         959           533         959	suits M value 6 0.0 6 0.0 6 0.0 6 0.0 6 0.0 7 0.0 7 0.0 7 0.0 7 0.0	Ultiple al Per Ident 100.00% 95.34% 95.03% 79.04% 79.69% 79.06%	gmm Acc. Len 333 333 333 332 332 332 332	ent MSAV Accessi AAA29796 CAA77743 226914.1 BAX39003 BAX39002 BAX39004
	Select all 100 seconous selected Description hemosphile: Description hemosphile: Description description Becham: En-En-Extended select. Paral. Percenter: Penudistraneous descrip- hemosphile: Homalia servectal hemosphile: Homalia servectal hemosphile: Homalia servectal hemosphile: Homalia servectal hemosphile: Servectal servectal	2021		Graphics Scientific Name Pseudolemanoy - Pseudolemanoy - Pseudolemanoy - Anisatis simolex Anisatis simolex Anisatis semelix Anisatis semelix Anisatis semelix Anisatis semelix	Distance Score 692 635 633 550 537 533 502	Tree of re           Total         Que           692         100'           635         963           633         963           550         100'           537         953           533         953           533         953           659         893	suits M y E value 6 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1e-176	Utiple al Per. Ident 100.00% 95.34% 95.03% 79.04% 79.69% 79.06% 79.06% 79.46%	ann Acc. Len 333 333 333 333 332 332 332 332 332	ent MSAV Accessi AAA29796 CAA77743. P26914.1 BAX39003 ASL68918.2 BAX39004 AFY38825
	Select all no sepannes andres Designen Interdata Pandetamona decigned Rollams, Fall-Elazohlar adas, Fina, Pesanar Pandetamona decig Rollams, Fall-Elazohlar adas, Fina, Pesanar Pandetamona decig Interdata Interda Interda Interdata Interda Interda Interdata Interda Interda Interdata Interda Interda Interdata Interda Interda Interdata Interda Interda	mi		Graphics Scientific Name Pseudolemanoy - Pseudolemanoy - Pseud	Distance Score 692 635 633 550 537 533 502 445	Trees of res           Total         Que           Score         Covy           692         100'           633         963           550         100'           537         953           533         953           559         899           703         901	suits         M           y         E           value         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         0.0	Utiple al Per. Ident 100.00% 95.34% 95.03% 79.69% 79.69% 79.69% 79.65% 79.46% 78.33%	gnm Acc. Len 333 333 333 332 332 332 332 332 332 33	ent MSA V Accessi AAA29796. CAA77743. P26914.1 BAX39002. BAX39004. AFY98826: AVY54521.
	Select all 100 segurances selected Cessigner https://doi.log.org/selected https://doi.log.org	2021		Graphics Scientific Name Pseudoterranoy. Pseudoterranoy. Pseudoterranoy. Pseudoterranoy. Pseudoterranoy. Pseudoterranoy. Pseudoterranoy. Pseudoterranoy. Pseudoterranoy. Pseudoterranoy. Pseudoterranoy.	Distanc: Max Score 692 635 537 533 550 537 533 502 445 443	Tree of re           Total         Que           692         100'           635         963           633         963           550         100'           537         953           533         953           533         953           659         893	suits         M           v         E           value         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         1e-176           6e-155         3e-154	Ultiple al Per Ident 100.00% 95.34% 95.03% 79.04% 79.69% 79.69% 79.69% 79.65% 79.46% 78.33%	ignm Acc. Len 333 333 332 332 332 332 332 332 332 33	ent MSA V Accessi CAA29796. CAA77743. P26914.1 BAX39003. ASL68918.2 BAX39004. AFY98826. AVY54521. AVY54521.
	Select all no sepannes andres Designen Interdata Pandetamona decigned Rollams, Fall-Elazohlar adas, Fina, Pesanar Pandetamona decig Rollams, Fall-Elazohlar adas, Fina, Pesanar Pandetamona decig Interdata Interda Interda Interdata Interda Interda Interdata Interda Interda Interdata Interda Interda Interdata Interda Interda Interdata Interda Interda	2021		Graphics Scientific Name Pseudolemanoy - Pseudolemanoy - Pseud	Distanc: Max Score 692 635 537 533 550 537 533 502 445 443	Trees of res           Total         Que           Score         Cov           692         100'           635         969           633         969           550         100'           537         959           533         955           659         899           703         909           701         909	suits         M           y         E           y         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         0.0           4e-153         4e-153	Ultiple al Per Ident 100.00% 95.34% 95.03% 79.04% 79.69% 79.69% 79.69% 79.65% 79.46% 78.33%	gmm Acc. Len 333 333 333 332 332 332 332 332 332 33	ent <u>MSA V</u> Accessi <u>AAA29796.</u> <u>CAA77743.</u>
	Select all 100 sequences avecned Compton Internation Providements deviced Internation Providements deviced Internation Providements deviced Internation Providements deviced International Providements International Provi			Graphics Scientific Name Pseudoterranoy. Pseud	Distance Max Score 692 635 537 537 533 502 445 443 440	Total         Oue           Score         Cove           692         100'           635         963           633         963           550         100'           537         953           533         953           659         893           703         903           701         909           588         859	suits         M           y         E           r         value           s         0.0	ultiple al Per. Ident 100.00% 95.34% 95.03% 79.69% 79.69% 79.69% 79.69% 79.69% 79.69% 79.69% 79.33% 78.33% 77.57%	gmm Acc. Len 333 333 333 333 332 332 332 332 332 33	ent MSAV Accessi AAA29796. CAA77743. P26914.1 BAX39003. ASL68918.2 BAX39004. AFY98826. AVY54521.1 AVY54522.1
	Select all 100 segurous selected		E E E E E E E E E E E E E E E E E E E	Graphics Scientific Name Pseudoteranov Pseudoteranov Pseudoteranov Anisakis simplex Anisakis simplex Anisakis secretifi Anisakis corretti Anisakis corretti	Distance Max Score 692 635 537 537 537 533 502 445 443 440 414	Total         Oue           Score         Cove           692         100'           635         963           633         963           550         100'           537         955           533         959           659         899           703         909           701         901           588         855           414         959	suits M r r suits M c suits M c c c c c c c c c c c c c c c	utiple all Per. Ident 100.00% 95.03% 95.03% 79.06% 79.06% 79.06% 78.06% 78.33% 78.33% 63.41%	gnm Acc. Len 333 333 333 332 332 332 332 332 332 33	ent MSAV Accessi AAA29796. CAA77743. P26914.1 BAX39003. ASL69918.2 BAX39004. AFY98826. AVY54521.1 AVY54522.1 P28316.2
	Select all 100 seconos selected     Description     Interviewed      Description     Interviewed      Description     Interviewed      Description     Interviewed      Description     D		8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	Graphics Scientific Name Pseudoteranov Pseudoteranov Pseudoteranov Anisakis simolex Anisakis simolex Anisakis severifi Anisakis severifi Anisakis severifi Anisakis ceverifi Anisakis ceverifi	Distance Score 682 635 550 537 533 550 533 502 445 443 440 414 239	Image         Image <th< td=""><td>suits M r r suits M c suits M c c c c c c c c c c c c c</td><td>Utiple all Per. Ident 100.00% 95.34% 95.03% 79.04% 79.06% 79.06% 79.06% 79.06% 79.06% 78.33% 78.33% 78.33% 77.57% 63.41% 74.00%</td><td>gnm Acc. Len 333 333 332 332 332 332 332 332 332 33</td><td>ent MSA V Accessi AAA29796 CAA77743 P26914.1 BAX39003 ASL68918.7 BAX39002 BAX39004 AFY18828.1 AVY54521.1 AVY54521.1 AVY54521.1 AVY54522.1 P28316.2 1ASH.A</td></th<>	suits M r r suits M c suits M c c c c c c c c c c c c c	Utiple all Per. Ident 100.00% 95.34% 95.03% 79.04% 79.06% 79.06% 79.06% 79.06% 79.06% 78.33% 78.33% 78.33% 77.57% 63.41% 74.00%	gnm Acc. Len 333 333 332 332 332 332 332 332 332 33	ent MSA V Accessi AAA29796 CAA77743 P26914.1 BAX39003 ASL68918.7 BAX39002 BAX39004 AFY18828.1 AVY54521.1 AVY54521.1 AVY54521.1 AVY54522.1 P28316.2 1ASH.A
	estelet all 100 seguences and ord Description Immodules Descriptions descend Beckman Chel-Description descend Beckman Chel-Description descend Beckman Chel-Description descend Henrodes Henrichte andre desc. Description descend Henrodes Henrichte serettel Henrodes Henrichte serettel Anterna Angel 20 Annahle serettel Henrodes Henrichte serettel Beckman Chel-Description descend Beckman Chel-Description descend Beckman Chel-Description descend Beckman Chel-Description descend The Standard Merchanis Henrodes Homes (Adaptions Resources to Beckman Chel-Description Comman (Adaptions Resources to Merchanis		2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	Graphics Scientific Name Pseudoterranoy Pseud	Distance Score 682 635 550 537 533 502 445 443 440 414 239 212	Total         Ourse           Score         Converting           Score         Converting           Score         Converting           Score         Converting           Score         Score           Score         Converting           Score         Score	suits         M           y         E           s         0.0           0.0         0.0           0.0         0.0           0.0         0.0           0.0         1e-176           6e-155         3e-154           4e-153         3e-154           2e-141         3e-75           2e-244         4e-59	Per. Ident 100.00% 95.34% 95.03% 79.04% 79.06% 79.06% 79.06% 79.06% 79.06% 79.46% 79.06% 79.06% 79.06% 79.06% 79.06% 79.06% 79.06% 79.06% 79.06% 79.06% 79.06% 70.05%	gnm Acc. Len 333 333 332 332 332 332 332 332 332 33	ent MSA V Accessi AAA29796. CAA77743. P26914.1 BAX39003. ASL699187 BAX39002. BAX39004. AFY98826. AV754521. AV754521. AV754522. P28316.2 1ASH.A AAL56430.
	Select all 100 sequences and one Description Intersolution Description Restance The Endeddemicrosol description Restance The Restance Sector II Restance The Restance Restance II Restance Restance Restance Restance II Restance Restance Restance Restance II Restance Restance Restance Restance II Restance Restance Restance Restance Restance Restance II Restance Restance Rest		2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	Graphics Scientific Name Pseudobernaroz- Pseudobernaroz- Pseudobernaroz- Anisabis semoles eleisabis serentifi eleisabis serent	Distance Max Score 682 635 550 537 533 550 537 533 550 537 533 550 537 533 550 2445 443 440 414 239 212 197	Itrae of res           Total         Ouese           Score         Cov           682         100'           633         963           550         100'           537         953           553         955           554         959           555         958           558         859           701         90'           558         533           300         90'           300         93'           328         87''           316         71''	withs         M         E           r         value         0.0           6         0.0         0.0           6         0.0         0.0           6         0.0         0.0           6         0.0         0.0           6         0.0         0.0           6         0.0         0.0           6         0.0         0.0           0         0.0         0.0           0         0.0         0.0           0.0         0.0         0.0           0.0         0.0         0.0           0.0         0.0         0.0           0.0         0.0         0.0           0.0         0.0         0.0           0.0         0.0         0.0           0.0         0.0         0.0           0.0         0.0         0.0           0.0         0.0         0.0           0.0         0.0         0.0           0.0         0.0         0.0           0.0         0.0         0.0           0.0         0.0         0.0           0.0         0.0         0.0	ultiple al Per ident 100.00% 95.34% 95.03% 79.04% 79.06% 79.06% 79.06% 79.06% 78.33% 77.57% 63.41% 74.00% 61.21% 64.58% 75.00% 63.89%	gnm Acc. Len 333 333 333 332 332 332 332 332 332 33	AAA29766 AAA29766 CAA77431 BAX38006 BAX38006 BAX38006 AX38006
	Select all 100 segurance selected Description Intervention Encodemicance description Intervention Encodemicance description		E E E E E E E E E E E E E E E E E E E	Graphics Scientific Name Preudoterranov Preud	Distance 0 0 0 0 0 0 0 0 0 0 0 0 0	Itree of res           Total         Quee           Score         Cave           682         100''           653         955           653         965           550         100''           551         955           553         955           553         955           554         899           703         90%           558         859           580         953           581         859           583         953           584         859           535         953           541         959           553         963           554         963           555         963           556         969           550         963           550         930           550         933           550         933           550         933           550         93           550         93           500         93           516         75           517         93           516	withs         M         E           r         value         0.0           6         0.0         0.0           6         0.0         0.0           6         0.0         0.0           6         0.0         0.0           6         0.0         0.0           6         0.0         0.0           6         0.0         0.0           7         10-176         6e-155           2         2e-141         3e-75           2         2e-44         4e-59           4         4e-57         7e-57           5         3e-48         5e-48	Utiple al Per. Net. 100.00% 95.34% 95.03% 79.06% 70.06% 70	gmm Acc. Len 333 333 332 332 332 332 332 332 332 263 338 150 171 146 132 146 358	AAA23756 AAA23756 CAA77351 BAX35002 BAX35002 BAX35002 BAX35002 AAY34521 AAY34521 AAY34521 AAY34521 AAY34521 AAY34521 AAX35002 AAX
	Select all 100 sequences and one Description Intersolution Description Intersolution Description Recharan End-Extended actions description Recharan End-Extended actions. Description of description Intersolution Intersolution Control Providence on description Intersolution Intersolution Section Intersolution Intersolution Intersolution Intersolution Intersolution Intersolution Intersolution Intersolution Intersolution Intersolution Intersolution Intersolution Intersolution Intersolution Intersolution Intersolution IntersolutionIntersolution Intersolution Intersolution Intersolution IntersolutionIntersolution Intersolution Intersolution Intersolution Intersolution IntersolutionIntersolution Intersolution Intersolution Intersolution Intersolution IntersolutionIntersolution Intersolution Intersoluti		2 2 2 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4	Graphics Scientife Name Preudoleranov	Distance 0 0 0 0 0 0 0 0 0 0 0 0 0	Itrae of res           Total         Ouese           Score         Cov           682         100'           633         963           550         100'           551         953           553         955           554         959           555         958           558         859           701         90'           558         533           300         90'           300         93'           328         87''           316         71''	suits         M           r         e           s         0.0           s <t< td=""><td>ultiple al Per. lient 100.00% 95.34% 95.03% 79.05% 70.05% 70.0</td><td>ignm Acc. Len 333 333 332 332 332 332 332 332 332 33</td><td>AAA29766 AAA29766 CAA77431 BAX38006 BAX38006 BAX38006 AX38006</td></t<>	ultiple al Per. lient 100.00% 95.34% 95.03% 79.05% 70.05% 70.0	ignm Acc. Len 333 333 332 332 332 332 332 332 332 33	AAA29766 AAA29766 CAA77431 BAX38006 BAX38006 BAX38006 AX38006

Figure 19a: Result page of Blastp

- Copy the sequence from the notepad. Avoid the first FASTA line. Paste the protein sequence in the search window of BLASTP. (Figure 19)
- Tick the box "results in new window", in the last line of blast Home page
- Click BLAST
- In the new window, format request display is shown, wait for few seconds, it leads to result page (Figure 19a)
- In the result page, the blast results for sequence similarity search appear as graphically and as text
- Click each text and find the similarity between sequence
- Click graphic summary
- View the graphical representation of sequence alignment (Figure 19b)
- Lines in pink, red colors represent the sequence in a set of score values.
- Length of the line indicates the length of the local alignment with the query sequences.

## RESULT

The most significant and similar sequences fetched by the blastp are hemoglobin [*Pseudoterranovadecipiens*],CAA77743.1

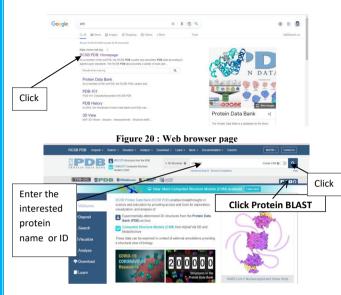
Descriptions	Graphic Summary	Alignments	Taxonomy							
) hover to see the t	tle 🖡 click to show alignments	Show Conser	ved Domains	Alignment Scores	■ < 40	40 - 50	50 - 80	80 - 200	>= 200	0
100 sequences sek				ected, click on the	Ť.,					
		Distribution o		ast Hits on 100 su	ıbject sei	quences				
		1 60	Que I 120	180 240	<b> </b> 300					
		_								
		_								
		=								
		Ξ								
									Activat	
										ttings t
		Figure	19b: Gra	phical su	mma	ry pa	ge			

25

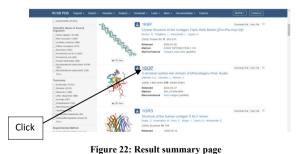
## OUTCOME

By learning BLASTP, students gain the knowledge to compare the new protein sequence with the protein database by aligning the novel sequence with previously characterized protein. From the comparison, students can obtain functional and evolutionary clues about the structure and function of the novel sequence.

26



#### Figure 21: PDB Home page



# ACCESSING STRUCTURAL DATABASE AND DOWNLOADING THE PROTEIN STRUCTURE

## AIM

To access the PDB (Protein Data Bank) structural database and to download the protein structure.

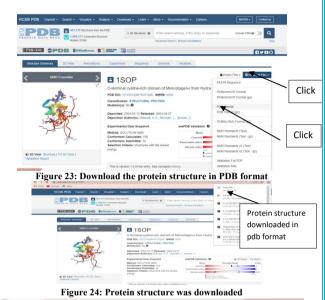
## **INTRODUCTION**

The three dimensional structure of biomolecules plays an important role in the functions and maintenance of the structural features of an organism. These structures were deciphered by research scientists and deposited in the databases, specifically designed for structural submission, for worldwide use. Research Collaborator for Structural Bioinformatics (RCSB), manages of the PDB. It provides free resources to assist the fields of biology and bioinformatics. It provides detailed information about sequence, atomic coordinates, structure factors, and crystallization conditions etc.

## PROCEDURE

- Type PDB in the web browser search field (Figure 20)
- Click the RCSB page (Figure 20)
- Go to the <u>WWW.rcsb.org</u> (Figure 21)
- Enter the protein name of interest in the search box
- Click Go

## 27



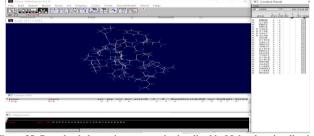


Figure 25: Downloaded protein structure is visualized in Molecular visualization tool (Rasmol and Swisspdbv)

- Choose the protein name of interest (Figure 22)
- Click the ID
- Click the download file (Figure 23)
- Click PDB format (Figure 23)
- Download file as PDB format (Figure 24)
- Save file
- View the protein structure in the molecular visualization tool like RASmol, Swisspdbv (Figure 25)

## RESULT

The protein structure was downloaded and their descriptions were observed.

## OUTCOME

Students get trained to retrieve any protein structure from structural protein database and can understand the characteristics of the protein structure.

#### WORKING WITH ENSEMBL

#### ۰ 🗉 🧌 X J O Q Ins Images I Videos I Books I More Tools Type ensembl 1 81 00 000 results /0 31 seconds my ensembliorn Ensembl genome browser 109 is a genome browser for vertebrate genomes that supports research in comparativ Ensembl volution sequence variation and transcriptional Human (Homo sapiens) Search Human (Homo sapiens) - Genome assembly, GRCh38 Ensemblinenome database project is a scientific nenier1 at the European Bioinformatics Institute which Click Ensembl project produces genome databases for provides a centralized resource for geneticists molecular biologists and other researchers studyin lants the nenomes of our own species and other sembl Plants is a genome-centric portal for plant species ertebrates and model organisms. Wikipedia Mouse Description: Ensembl Search Mouse (Mus musculus) - Genome assembly: GRCm39 . References allowing and states and states

#### Figure 26: Google search page

← → C ☆ # ensemblang/index.html?redirect=no 8 x x I I 2 . Search al species Ensemblies a penome howser for vertebrate penomes that support research in comparative penomics, evolution, sequence variation and RioMart > BLASTIRI AT > Variant Effect Predictor > transcriptional regulation. Ensembl annotate genes, computes multiple alconnents, credicts regulatory function and collects disease data Search our genomes for your DNA Analyse your own variants and Ensemblitude include RI AST RI AT RinMart and the Variant Effect reachi with this data-mining tool or protein sequence medict the functional consequence Prodictor (VEP) for all cummented spacies Ensembl Release 109 (Feb 2023) · New nene sets for deploy and horse Search . Updated SIFT and PolyPhen-2 missense variant nationenicit · New VEP plugins for UTR apportation · New ATAC-see tracks (peaks and signal) for fish species (Adamfic Go Salmon Furnean Seahass Rainboy Trust and Turbot e o BRCA2 or pet 54 97383-63627669 or rs699 or coronary heart disease Uno place naver? on our blog embl Rapid Release Favourite genomes / New assemblies with gene and protein annotation every two Select Type your interested Note: species that already exist on this site will continue to be species undated with the full range of apportations. gene or disease Go The Ensembl Rapid Release website provides annotation for recently produced, publicly available vertebrate and nor vertebrate genomes from biodivestev initiatives such as Darwin

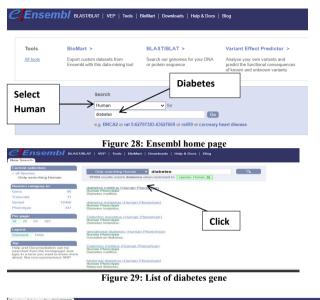
Figure 27: Ensembl home page

#### AIM

To retrieve the vertebrate genomic information **INTRODUCTION** 

Ensembl is one of the several well-known genome browsers for the retrieval of genomic information. It is considered to be the universal information source for the human genome. Data available in Ensembl include genes, SNPs, repeats and homologies. Genes may either be known experimentally or deducted from the sequence; because the experimental support for annotation of the human genome is so variable. It presents the evidence for identification of every gene. Extensive linking to other databases containing related information such as OMIM or expression databases is also available. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.

- · Type Ensembl in the web browser search field
- Click Ensembl genome browser 109 (Figure 26)
- Go to the home page, First choose the species of interest and type gene of interest or disease name (Figure 27 & 28) and click go



enotype-based displays Associated loci Related conditions	Loci assoc	iated with	n diabetes mellit	us (EFO:000	0400) 🥹				
🌣 Configure this page	Filter T Fo	ature type: All	Annotation source	r Al 🛡 Phen	chipo/Disease/Trait: All				
👔 Custom tracks									
🛃 Export data					Showhide columns			Search	
< Share this page	Name(s)	Туре	Genomic location (strand)	Reported gene(s)	Phenotype/Disease/Trait	Annotation source	Submitter	External reference	Supporting evidence
🖗 Bookmark this page	m8015373	Variant	14:98278970 (+)	C14orf54 C14orf177	Diabetes Melitus	dbGaP 67			
	rs1187501881	Variant	19:7141829 (+)	INSR	Insulin-resistant diabetes mellitus AND acanthosis regricans	ClinVar@			
	rs145397235	Variant	4:72492529 (+)		Cardio-cerebrovascular disease in diabetes melitus	NHGRI-EBI GWAS catalog-9		PMID 336322 3869	
	rs577444458	Variant	19:7114304 (+)	INSR	Insulin-resistant dabetes mellitus AND acanthosis regricars	Cirl/ar@			
	rs189233549	Variant	9:117365402 (+)		Ischemic stroke in diabetes mellitus	NHGRIEBI GWAS catalog-9		PMID 336322 3849	
	rs754015320	Variant	19:7115033 (+)	INSR	Insulin-resistant diabetes mellitus AND acanthosis nigricans	CinVar#			
	rs6415788	Veriant	9:4118111 (+)	GUS3	Disbetes Melitus	Cirl/ar#			PMD.314155 7549
	rs3758650	Variant	11.616855 (+)	IRE7	Diabetes Melitus	dbGaP#			
	rs587783675	Variant	11:17387104 (+)	KCNJ11	Diabetes Melitus	ClinVar@			
	rs1875084584	Variant	12:120978824 (+)	HNE1A	Diabetes Melitus	Cirilia:9			
	rs61742093	Variant	6.27912204 (+)	<u>0R262</u>	Estimated glomerular filtration rate in	NHGRI-EBI		PMID 314517	

Figure 30: Details of diabetes were listed

- List of genes were displayed (Figure 29)
- Select one and click
- Details of diabetes were shown (Figure 30)
- Click hyperlink one by one; the details of variant, genomic location, reported genes, phenotype/genotype trait, annotation source, submitter, and external reference were known. (Figure 30)

## RESULT

The diabetic gene annotations were retrieved from the Ensembl genome browser.

## **OUTCOME**

Students can retrieve any vertebrate gene and its details from the Ensembl genome browser.

#### **MULTIPLE SEQUENCE ALIGNMENT**

#### AIM

To study the closely related genes or proteins

## **INTRODUCTION**

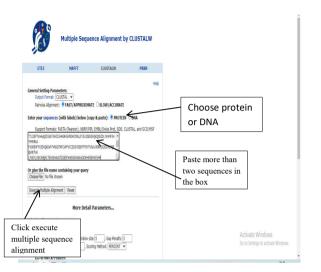
ClustalW tool is used for aligning multiple nucleotide or protein sequences. It uses progressive alignment methods, which align the most similar sequences first and work their way down to the least similar sequences until a global alignment is created. A multiple sequence alignment tools called Clustal Omega creates alignments between three or more sequences by using HMM profile-profile algorithms and seeded guide trees. Multiple sequence alignment is a tool used to study closely related genes or proteins in order to find the evolutionary relationships between genes and to identify shared patterns among functionally or structurally related genes.

## PROCEDURE

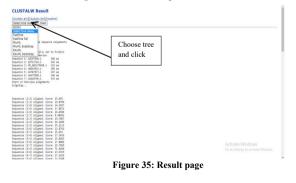
- Type multiple sequence alignment tool in the web browser search field (Figure 31)
- Click Clustal Omega < Multiple Sequence Alignment < EMBL-EBI Or
- Click Multiple Sequence Alignment-CLUSTALW Genome Net ((Figure 31)



X L D Q



#### Figure 34: Paste the sequences in the box



- Go to home page (Figure 32)
- Copy the sequences from the notepad (Figure 33)
- Paste the more than two sequences in input sequences box (Figure 34)
- Choose the sequence as protein or DNA (Figure 34)
- Click submit or execute multiple sequence alignment (Figure 34)
- Result page is shown (Figure 35)
- Choose fast tree and click
- Phylogenetic tree was displayed (Figure 36)

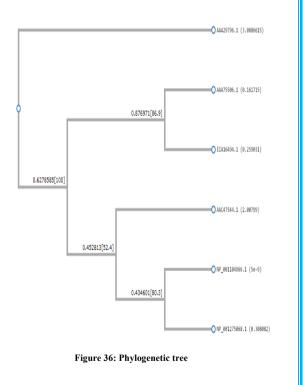
## RESULT

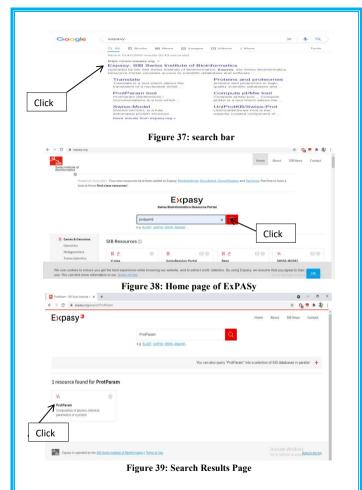
- The close relationship for this <u>NP\_001275068.1</u> (any one Id which sequences are pasted in input sequence box) is <u>NP\_001104966.1.</u>
- The far relationship for this <u>NP\_001275068.1</u> (any one Id which sequences are pasted in input sequence box) is <u>AAA29796.1</u>.

## OUTCOME

Using this multiple sequence tool, students attain the ability to infer the evolutionary relationships between the sequences under study.







## PREDICTING PHYSIOCHEMICAL PROPERTIES OF PROTEIN SEQUENCE

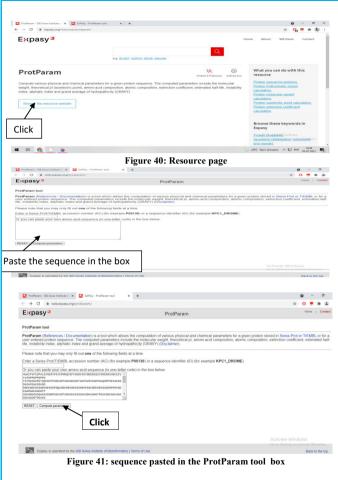
## AIM

To predict the physiochemical properties of protein sequence using ExPASy resources.

## **INTRODUCTION**

With the help of the tool ProtParam, users may compute a number of physical and chemical parameters for a specific protein that is contained in Swiss-Prot or TrEMBL, as well as for a user-entered protein sequence. Molecular weight, theoretical pI (Isoelectric point), amino acid composition, atomic composition, extinction coefficient, assumed half-life, instability index, aliphatic index, and grand average of hydropathicity (GRAVY) are among the calculated characteristics. **PROCEDURE** 

- Type ExPASy in the web browser search box (Figure 37)
- Click ExPASy
- Go to ExPASy page, enter specific proteomic tool in search
- box (Figure 38) Or
- Enter the url: https://web.ExPASy.org/ProtParam/
- For predicting physiochemical properties, type ProtParam and enter



- Select protoparm tool (Figure 39)
- Click "browse the resource website", protoparm tool page is opened in new window (Figure 40)
- Paste the protein sequence / accession number in the given box (Figure 41)
- Click compute parameters (Figure 41)

## RESULT

Using ProtParam tool, the physicochemical properties of protein sequence were predicted as follows (Figure 42):

- Number of aminoacids:718
- Molecular weight: 72106.72
- Theoretical pI: 7.31
- Total number of negatively charged residues (Asp + Glu): 77
- Total number of positively charged residues (Arg + Lys): 77

## OUTCOME

Students learnt the art of analysing the physical and chemical properties for a specific protein.

## ← → C 🔒 web.expasy.org/cgi-bin/protparam/protparam

Atomic composition:

Carbon	C	3150
Hydrogen	н	4974
Nitrogen	N	926
Oxygen	0	986
Sulfur	S	16

Formula:  $C_{3150}H_{4974}N_{926}O_{986}S_{16}$ Total number of atoms: 10052

Extinction coefficients:

Extinction coefficients are in units of  $\,\,{\rm M}^{-1}\,\,{\rm cm}^{-1},$  at 280 nm measured in water.

Ext. coefficient 34545 Abs 0.1% (=1 g/l) 0.479, assuming all pairs of Cys residues form cystines

Ext. coefficient 33920 Abs 0.1% (=1 g/l) 0.470, assuming all Cys residues are reduced

Estimated half-life:

The N-terminal of the sequence considered is E (Glu).

The estimated half-life is: 1 hours (mammalian reticulocytes, in vitro). 30 min (yeast, in vivo). >10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 24.32 This classifies the protein as stable.

Aliphatic index: 55.43

Grand average of hydropathicity (GRAVY): -0.777

Figure 42: Result page of Protoparam tool

45

Expasy <sup>3</sup>			Home	About	SIB News	Contac
	PeptideMass e.g. BLAST, UniProt. MSH6	×Q				
	eği <u>bi.kor</u> , unimu, isono	AND -				
		You can also query "PeptideMass	" into a selection	n of SIB dat	abases in par	ralel <mark>+</mark>
1 resource found for Peptide	Mass					
W ©						
PeptideMass Precedual protein cleavage by a given						
enzime						
lick						
lick						
lick	stade of Bioinformatics ( Terms of Use				e Windows tings to active	
Expansy is operated by the <u>SIR Swiss In</u>		oteomic tool sear	ch site			
Equary is operated by the <u>550 Series</u> for Figure 4.		oteomic tool sear	ch site			
Equary is operated by the <u>550 Series</u> for Figure 4.		oteomic tool sear		Go to Set	tings to activa	<u>lack in the</u>
Equasy is operated by the <u>SIR Swiss In</u>				Go to Set	tings to activa	<u>lack to the</u>
Equary is operated by the <u>SB Seni, the</u> Figure 4.	3: ExPASy Pro	Tool opens in Q M ©	n new	Go to Set	tings to activa	<u>lack in the</u>
Equary is operated by the <u>BB Bank in</u> Figure 4. Equary Sees Boortematics Resource Pottel PeptideMass area a protein sequences with a chosen enzyme/pro-	3: ExPASy Pro	Tool opens in Q Bus Diversion Statement egenerated operforts. The tool also	n new What yr resource Protein s	Go to Set	ndov o with this	<u>lack in the</u>
Equary is operated by the <u>SS Section</u> <b>Figure 4</b> . Equary Sector Boordematics Resource Post PeptideMass even a protein sequence with a chosen enzymelyre men showcad a sectoric port and mass values to and to carry post-interation medications, and the	3: ExPASy Pro	Tool opens in Q These is here as permeted perfiles. The tot also defaulties can refer the tot also	M new	Go to Set	ndov o with this	<u>v</u>
Every is updated by the <u>BES seek in</u> <b>Figure 4</b> . Expany: Seek Biorkomatics Resource Portel <b>CopplicatedMasss</b> wave a porter advancement of the other entrymology methodered an observation and relations, and call proprietance or space waves.	3: ExPASy Pro	Tool opens in Q These is here as permeted perfiles. The tot also defaulties can refer the tot also	What yu resource Protein s Protein s Protein s Browse	Go to Set 2 2 2 2 2 2 2 2 2 2 2 2 2	ndov o with this nelysis.	<u>v</u>
Equary is operated by the <u>550 Series</u> for	3: ExPASy Pro	Tool opens in Q These is here as permeted perfiles. The tot also defaulties can refer the tot also	M NEW What yy resource Protein a Protein a Protein a Browse Expasy Protein a	Go to Set 2 2 2 2 2 2 2 2 2 2 2 2 2	ndov with this network	<u>v</u>



## PREDICTING PEPTIDE MASS OF PROTEIN SEQUENCE

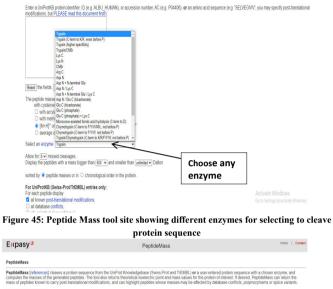
## AIM

To predict the peptide mass of protein sequence using ExPASy resource

## **INTRODUCTION**

In the analytical method of protein identification known as peptide mass fingerprinting (PMF), the unknown protein of interest is first broken up into smaller peptides whose absolute masses may be precisely determined with a mass spectrometer like the MALDI-TOF or ESI-TOF.

- Type ExPASy in the web browser search field
- Click ExPASy
- Go to ExPASy page, Enter Peptide Mass in the search field (Figure 43) Or
- Type url : https://web.ExPASy.org/peptide\_mass/
- Click peptide mass tool
- It leads to peptide mass resource page
- Click the "Browse the resource website", Peptide mass tools site opens in new window (Figure 44)
- Paste the protein sequence of interest



#### Instructions are available.

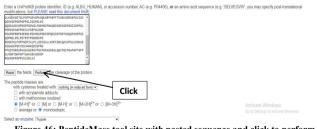


Figure 46: PeptideMass tool site with pasted sequence and click to perform

- Move the cursor to choose which type of enzyme to cleave the interested protein sequence (Figure 45)
- Select "trypsin" from the drop-down box under "Select enzyme: in the tool," if we want other enzyme use drop menu and select the desired enzyme (Figure 45)
- Click perform (Figure 46)
- Result page shown (Figure 47)

## RESULT

- The high molecular weight peptide mass sequence is <u>TGPPGKPGPPGPPGPGIQGIHQTLGGYYNK</u>, position <u>309-</u> <u>339</u>and its molecular mass is <u>3036.5689</u>
- The low molecular weight peptide mass sequence is <u>QELK</u>, position <u>174-177</u> and its molecular weight is <u>517.2980</u>.

## **OUTCOME**

Students can identify the molecular weight of peptide after cleavage of protein using different enzymes or chemicals.

The selected enzyme is: Trypsin

Maximum number of missed cleavages (MC): 0

All cysteines in reduced form.

Methionines have not been oxidized.

Displaying peptides with a mass bigger than 500 Dalton.

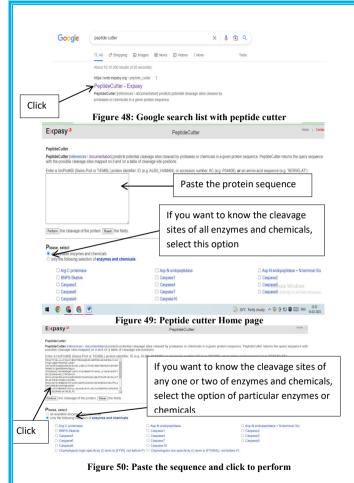
Using monoisotopic masses of the occurring amino acid residues and giving peptide masses as [M+H]\*.

#### The peptide masses from your sequence are:

			s): 72106.72 / Mw (monoisotopic mass): 72062.31
mass	position	#MC modifications	peptide sequence
3036.5689	309-339	0	TGPPGKPGPPGPPGPPGIQG IHQTLGGYYNK
2972 5880	9-34	0	FFVLETWQVLNQQNIPQISI VVDGGK
2591,4806	562-588	0	DGKPGLPGPPGDPIALPLLG DIGALLK
2398 2183			GDEGLOGIPGIPGAPGPTGP PGLMGR
2349 2196			GDTGPPGPPAI PGSI GIOGP OGPPGK
2345.2160			I GISIOSOVISI YMDONI IA R
2278,1212	36-54	0	WEFMFQATEGDVLNYIFR
2205.0029	145-165	0	CPEQDGFGNIASSWVTAHAS K
2204,9872	125-144	0	IYCSANLIAQETCCEISDTK
2038.0828	504-525	0	GLPGVHGSPGDIGPQGIGIP GR
1845 8911	606,715	0	GSDGPPGKPGPPGPPVSCSR
1815.9347			GLPGLPGTPGTPGNDGVPGR
1748.8350	205-222	0	GEPGENGLHGAPGFPGQK
1687.8398	480-497	0	SLPGEPGLDGNPGAPGPR
1492,7502	528,541	0	TGAQGPAGEPGIQGPR
1437.6613			NECGNCQASVPGLK
1428.7845			GDPGGIIGPPGLPGPK
1425.7332	244-258	0	GDPALAGLNGENGLK
1399.7215	357-371	0	GETGLPGFPGSVGPK
1322,6698			GEAGLPGAPGSPGQK
1132,5164			GQQGSAGSMGPR
1128.5796			GDLGPHGPPGPK
1127,6055			GEPGVIGSQGVK GNDEHEAGGLK
1126.5123			GNUEHEAGGLK DQCQCIPNK
1045 5789			ELRPLEDR
1035 5258			ETTVPEWR
987.4377	606-616	0	GEEGGAGEPGK
961.4625	375-383	0	GEPGEPFTK
946.4377	96-103	0	DTVDFHGR
938.4214			GEQGFEGSK
922.4741			TGHPGPTGAK
917.4727			GYPGIPGEK
896.4546			MSSYLPAK
750.4257 742.3188			GPPGIPGR YDSMAR
712 3624			GEAGPPGK
660,4039			TVIATR
647,2995			GEQGEK
620,2886			QTDEK
620.2886	232-237	0	GETGEK
614.3256	624-629	0	GDIGPR
598.3096	65-68	0	QWHK
517.2980	174-177	0	QELK

90.7% of sequence covered (you may modify the input parameters to display also peptides < 500 Da or > 10000000000 Da):

## Figure 47: Result page of peptide mass



## PREDICTING CLEAVAGE SITE OF PROTEIN SEQUENCE

#### AIM

To predict the cleavage site of protein sequence using ExPASy resource (Peptide cutter).

## **INTRODUCTION**

Peptide Cutter explores for protease cleavage sites in a protein sequence given by the user, a protein sequence from the SWISS-PROT and/or TrEMBL databases, or both. Users can use a protease, a group of proteases, or the entire list of proteases and compounds. Different forms of output of the results are available: Tables of cleavage sites that are either ordered alphabetically by the names of the enzymes or sequentially by the number of amino acids are available. A map of the cleavage locations is a third possible result. The user can choose the block size to print out the sequence and the cleavage sites that have been mapped onto it.

- Type Peptide cutter in the web browser search field and press enter
- Click Peptide cutter tool (Figure 48) Or Type url : <u>https://web.ExPASy.org/peptide\_cutter/</u>

Exp	asy ª	PeptideCutter
Peptide	Cutter	
The surgest	10	ан 20 така илика нарка санантакан витака или тара орад 19 така илика нарка санантакан витака или тара орад
		100 500 500 500 500 500 500 500 500 500
VACEVICE	TT ECCEAND	200 CLTWV0110 ENLANDER BLCCALOF DEMORALITY
-	TRA SPASING	240 - 270 - 100 -
RCCORPVI		120 LINE VOLVEDAVE RINGRADOV LITEVIN
RHPEYSVE	178	THE INE AND FROM TRANSFORMED VEHICLES
		440 440 440 440 440 440 440 440 440 440
		500 510 520 520 540 540 540 540 540 540 540 540 540 54
PEAEKQI	CKQ BALVELS	100 PARTECUT VIEDTOTING ECCAREDER CHEEDER
****		
The sec		600 amino acids long.
-	of all that you have	
	roteinase k ve chosen	K to disclay di possible cleaving enzymes.
Name of	No. of	
enzyme	cleavages	Positions of cleavage sites
oteinase	328	112 D4 105 110 112 16 119 C1 214 26 127 03 139 440 144 143 144 460 446 156 115 155 155 156 159 167 150 157 156 159 177 157 15
ase are the clea	avage sites of th	is chosen engines and chemicals mapped onto the entered portions sequence:
<ul> <li>You have</li> </ul>	ave chosen	n a block size of 60 for the map.
• Please	e note that	the cleavage occurs at the right side (C-terminal direction) of the marked amino acid.
	a moto unat	en startige occurs al intergent way to terminal another of the mainted anne and
. Vou ha	ave the new	ssibility to display the results of a single enzyme by mouseclicking on the respective enzyme name in the map.
* Touris	ave use pue	sound, to display the results of a single enzyme of mouseoneoning on the respective enzyme hame in the map.
		Pretts.
		Press Press

- Go to the peptide cutter tool home page (Figure 49)
- Paste the protein sequence of interest in the given box
- Move the cursor to choose the enzyme type to cleave the sequence of interest or select all available enzymes and chemicals; Click perform, (Figure 50)
- Result page is displayed (Figure 51)

## RESULT

<u>Proteinase K</u> enzyme (Specific enzyme choose) cleaves 328 sites in a given protein sequence.

## **OUTCOME**

Students can predict the potential cleavage sites cleaved by proteases or chemicals in a given protein sequence. Peptide Cutter returns the query sequence with the possible cleavage sites mapped on it and /or a table of cleavage site positions.

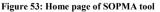


## Figure 51: Result page

Proti ProtX ProtK ProtK) ProtK Protit ProtK Prote ProtK Proti

Google	sopma tool	X I Q
	Q, All Ø Shopping D Videos 🗟 News 🕻 Images I More	Tools
	About 23,600 results (0.43 seconds)	
	https://npsa-prabi.ibcp.fr > NPSA > npsa_sopma *	
	NPS@ : SOPMA secondary structure prediction	
Click	SOPMA SECONDARY STRUCTURE PREDICTION METHOD. [Abstract] [NPS@ hel	] [Original
	server). Sequence name (optional) : Paste a protein sequence below : help.	
	People also search for	×
	sopma tool uses expasy sopma tools	
	sopma full form sopma pdf	
	sopma slideshare swiss-model	
	https://npsa-prabi.ibcp.fr> npsahlp_secpredsopma *	
	NPS@ help : Help on SOPMA tool - IBCP	
	SOPIMA (Self-Optimized Prediction Method with Alignment) is an improvement of SOF	M method.
	These methods are based on the homologue method of Levin et al The	
Figure	52: Google search list with seconda	ry structure prediction

prabi			PRABI-GERLAND BIOINFORMATIC POLE			
filione	<b>V</b> Services	🗯 Teaching	@Publications	ΦLinks	<b>≜</b> .kbs	ÇıÇontact
SOPMA SECONI [Astract][NSS bid][Otginal se Sequence name (optional): Prette a protein sequence below: b	rver]	IRE PREDICTION	IMETHOD			
Output width: 70 SUBMIT CLEAR						
Parameters						
Number of conformational states : Similarity threshold : 8	4 (Helix, Sheet, Turn, Coil) 🛩					
Window width: 17					Activate V Go to Settin	Windows as to activate Window



## PREDICTING SECONDARY STRUCTURE OF PROTEIN SEQUENCE USING SOPMA TOOL

## AIM

To predict the secondary structure of protein sequence using ExPASy resources

## **INTRODUCTION**

Secondary structure prediction is a group of techniques in bioinformatics that aim to calculate the secondary structures of proteins and nucleic acid sequences based on the information from their basic structures (primary structures). Through base pairing and base stacking interactions, it predicts the formation of nucleic acid structures like helixes and stem-loop structures while predicting the formation of protein structures like alpha helices and beta strands for proteins. The Self-Optimized Prediction Method with Alignment (SOPMA) is a tool to predict the secondary structure of a protein. Based on the query (primary sequence of a protein), SOPMA will predict its secondary structure. Protein secondary structure prediction offers insight into the activity, interactions, and functions of proteins as well as serving as an important initial step toward tertiary structure prediction. The polypeptide backbone of the local conformation proteins is referred to as the protein's secondary structure.

# Home	<b> </b>	🛱 Teaching	(2) Publications	φLinks	💼 Jobs	0Q
OPMA SECON stract.] (NPS@ help] [Original se		RE PREDICTION	METHOD			
puence name (optional): date a protein sequences below: T TVPF/WFFVLETWQ/LNQQL NQ	elp PQISIVUDGGKKVVEFMFQA + HKLG DTVDFHGRTVLATRASDGKPV DTVDFHGRTVLATRASDGKPV SKMSSTVARKQELKDQCQCIP PGEN TGEKKEQGEKKDPALAGLN +					
ameters nber of conformational state larity threshold : 8 dow width : 17	Click					
		igure 54: se	quence paste	d in box		
	ne Chrokes PAA.(2): 5:370-530-20 in survival on NFIQ and Results will be phone lation. Place will	in Standa on 2011/17 (92218). In Adamsk on 2011/17 (92218). Ind daving as facts.	(haliation)	45 UMBS	mikke CPCertail	
	n 1 deguerron Analysia 1 fail (1991) (1993) (1993) 1 fail (1994) (1993) (1993) 1 fail (1994) (1993) (1993) 1 fail (1994) (1993)					
SOPMA re Manufact Gamerice, 6 View SOPMA in Lan	esult for : UNK_10614 Classifier C. Stiffet: Spokest Systems of Control (Control (Co	) prinké akonéky dructura pratictich by conaces	a pradiction from multiple alignments, Cabine (1993) b			
		40     7				
Sequence 1	ength : 718				Go to Settings to activate Windo	0
SOPMA : Alpha h 3 <sub>10</sub> hel Pi heli Beta br Extende Beta tu Bend re Random Ambiguo Other s	idge (Db): 0 d strand (Ee): 65	is 13.51% is 0.00% is 0.00% is 0.00% is 0.00% is 6.55% is 0.00% is 0.00% is 0.00%				
			400 500			
Althe	Alababatella	Allan		Marking A	A	
Parameters Window Similar	: width : 17 ity threshold : 8 of states : 4					
Number			esult page of	SODMA		

## PROCEDURE

- Type SOPMA tool in the web browser search field and press
- enter
- Click the NPS@SOPMA secondary structure prediction (Figure 52)
- Open the SOPMA secondary structure prediction tool home page (Figure 53)
- Paste a protein sequence in the given box (Figure 54)
- Click submit (Figure 54)
- Result page (Figure 55) is displayed.

## RESULT

- The secondary structure location present in the protein sequence is predicted.
- The secondary structures present in the given sequence are alpha helix, extended strand, beta turn and Random coil.
- Random coil is the most prominent secondary structure (70.89%) in the given sequence.
- The lowest percentage of secondary structure in the given sequence is Beta turn.

## **OUTCOME**

Students learn how to predict the regions of different forms of secondary structure from the protein sequence.

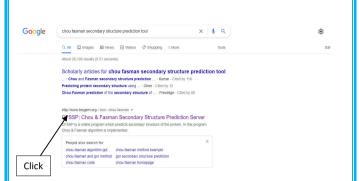
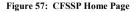


Figure 56: Search list of Secondary structure prediction tool

III CFSSP: Chou & Fasman Secondal × +	
biogem.org/tool/chou-fasman/	
CFSSP: Chou & Fasman Secondary Structure Prediction Server	
HURE DRAF LUXS ACARCINE, CONTACT FROM	
This server predicts secondary structure of protein from the amino acid sequence. In this server, Chou & Fasman algorithm has been implemented.	
Enter the protein sequence (in fasta format)	
CLEAR PREDICT	
Citation:	
<ol> <li>Ashok Kumar, T. (2013). CFSSP: Chou and Fasman Secondary Structure Prediction server. WIDE SPECTRUM: Research Journal. 1(9):15-19.</li> </ol>	
Reference:	
<ol> <li>Peter Y. Chou, and Gerald D. Fasman. Prediction of protein conformation. <i>Biochemistry</i>, 1974 Jan; 13(2), pp 222-245.</li> </ol>	
<ol> <li>Peter Y. Chou, and Gerald D. Fasman. Conformational parameters for amino acids in helical, β-sheet, and random coil regions calculated from proteins. <i>Biochemistry</i>. 1974 Jan; 13(2): pp 211-222.</li> </ol>	Activ: Go to S



## PREDICTING SECONDARY STRUCTURE OF PROTEIN SEQUENCE USING CFSSP TOOL

## AIM

To predict the secondary structure of the given protein sequence through CFSSP tool.

### **INTRODUCTION**

CFSSP (Chou and Fasman Secondary Structure Prediction Server) is an online protein secondary structure prediction server. The output predicts regions of secondary structure from the protein sequence such as alpha helix, beta sheet, and turns from the amino acid sequence. The method implemented in CFSSP is Chou and Fasman algorithm, which is based on analyses of the relative frequencies of each amino acid in alpha helices, beta sheets, and turns based on known protein structures solved with X-ray crystallography. CFSSP is freely accessible via ExPASy server or directly from BioGem tools at http://www.biogem.org/tool/chou-fasman.

- In web browser, type Chou Fasman secondary structure prediction tool, search list will be displayed and click CFFSSP (Figure 56)
- Click CFSSP or <u>http://www.biogem.org > tool > chou-fasman</u>



This server predicts secondary structure of protein from the amino acid sequence. In this server, Chou & Fasman algorithm has been implemented.

----- Enter the protein sequence (in fasta format) -----



Reference: 1. Peter Y. Chou, and Gerald D. Fasman. Prediction of protein conformation. Biochemistry Click 13(2), pp 222-245.

- (2), pp 222-245.
   Peter Y, Chou, and Gerald D. Fasman. Conformational parameters for amino acids in helical, β-she and random coil regions calculated from proteins. *Biochemistry*. 1974 Jan; 13(2): pp 211-222.
  - Figure 58: Sequence pasted in the box

		-	Contraction of the	-				and the second second	200
	3.60	2.40	144		a reaction and	100	7.0		
					a source tet				
THEFT		POPARasio				LPGAPGEFE			
morni	226	2 ho	2.44	ILEN DUPOLO	A LANDARD COM	1. mercanet 1. 7 40	and and		
ALPOSI		aaroorrore	oononno218		E correcto STR		monoene28		
		Learnandan	Kagkarraif						
asaasi		narranval	CARTING & PROPERTY.		a				
araral		aenai.rath	anaoranost		nanerox 300				
animali	955	POPPEAL	ODIGALLENP	CONCOALOR	LEAPHONE SA	ADDEPORTO SER	ARKODIGRAG		
areas					E energeest.				
- norm			or particular	CONFERENCE	a Postionality	Garterinerite	ALGEROSOGP		
at meaning	CA PT 271	VACABLES							
be differ					Read and the second			and the second second second	- 1001
									] = 7.m
cond		STORE THE COL							
			-	-	-	-	-		
		TTOPPUMPPN	LETHOVLNOON			EGDVLNVIPPO	RELAPLFDRO	21513 72	
222 3		eeeeeeeee	eeeeeeee	eeeeeee ,	T eccecce	Tececce	eeeee		
						-	-		
	1 1		HOCHLIABBOT	DEKOTVOFHO	TYLATRASPR	OVDIELHQLK	LACENNE INGE.	TCCEX 140	
221 Z	: :			** *	Tecconorter			140	
								10000	
222 E:	11 1	DTKEPEQOGP	GNIABBOATAN	ABREADYLES	restreet	PARGEAGLPGA	-asraqkarea	10101 230	
	11 · · ·	T T T	T		+ ceeccee		т т т	T 518	
	** **	*******							
778 B	ee - 0	LINGADGED	Kargareaska	exancanĝan	KODPALAGLING	INGLIGGEGON	зеевекаенае	rappa 200	
	1								
rue a						*****	************	1 1 2 0 0	
		PALPANLAT	argarrasiag	nannastärr	assanan	-argariation	างงานเฉียนเฉยาอา	PREASE AND	
			*			COLCECCE		222	
R.	6X +	<del>*</del>	************	e e e e tre e tr			eess frances fr	111-111 35-5	
NAME OF A	353		-	-	-	-	-	PPOLIOSPOLK	0 420
		HH	Turrur Puse	*********	1000001	inder de Lang	**	1000	
heet	353				T TT				1 420
true	351	HINCTOCTO	acceccetee	c Transmic	CTHERITHCCC	eccreace r	HECTECCEC	CTCCCCTTCHH	1 420
			and the second second		on galaxies and the	and the second s	a ser a s		
wery	423	ODGSAGSH	OPROPPODVO	LPGEMOTEO	KOGIKGEKGOR	GOLIGPPOL	OPKOLAGPPE	KSLPGEPGLOG	-
	423								400
1417111	423	100	· · · · ·	(m) m	T T	T. TT.	T T T		
TOUC	-12.3	ARTCOCCO	CTOCCCTCCC	COTCECET	HTHRH THECCE	TECCETTCE	TOUTCER TOU	TORCTOTORET	< -1114E
			-			-	-	-	
alix	493	POAPOPRO	PROERGLPGV	HOSPODIOP	QOIGIPORTON	QOPAGEPGI	GPRGLPGLPG	TPOTPONDOVP	0 560
DOD OT	49.8								1000.00
ures.	493	I. I.			Leeve Veren	Y	* *	correct of the second	1 560
erue:	491	Therefore	unu recece	concreteer	TALLACTOOCC	A PROPERTY OF THE OWNER		rerece	1 5554
			-	-	-	-			
wary with	563	RDGKPGLP				PGLKSNKGEI	GGAGEPGKYE	SMARKGDIGPR	
			EEEE	EE	DECEE				
Truc	5453	encerece	TCCC TCRMER		EXPERENCE	encerecent	receterces	seccificacie	C 630
uery		PROFESSOR	GPROSEDE DO	VPRIPRERG	DEGLOSTPOTE	GAPOPTOPP	A PROPERTY OFFICE	TGARGERGSDG	0 700
					1.0.1			1.01.0	200
PLAT AT T	223					Carlo Sana		1.00	
trun		erceree	cocceccie	CCCCC TCCC		CCC TCCC TT	ceccecete	COMPRESSES	C 700
	701	PAKPAPPA	PPVACARLKI	73.8					
wery	701			738					
			τ						
elis heet	7603								
TIME T	701		receeeeee	N 27 88					
ells heet urns truc	701	Advent He			120				

Figure 59: Result page

- Open the CFSSP secondary structure prediction tool home page (Figure 57)
- Paste a protein sequence in the given box (Figure 58)
- Click predict (Figure 58)
- Result page (Figure 59) is displayed.

## RESULT

- The secondary structures present in the given sequence are alpha helix, beta sheet and betas turn.
- Alpha helix is the most prominent secondary structure (32.9%) in the given sequence.
- The lowest percentage of secondary structure (16.7) in the given sequence is β turn.

## **OUTCOME**

Students understand how to predict the regions of different forms of secondary structure from the protein sequence.

	Expasy <sup>a</sup>			Home	About	SIB News	Conta
		Tmpred	×Q				
		e g. BLAST, UniPiol, MSH6, Aburn	0-				
			You can also query "Tmpred" in	ito a selection	n of SIB dat	labases in par	sliel •
	7						
Click	1 resource found for Tmpred						
	w Ф 💿						
	TMPred Prediction of trans-membrane regions in						
	protein sequences						
	Expansy is operated by the <u>Sill Swiss Instite</u>	ule of Boinformatics   Terms of Lise				e Windows	ack to Dr
	Expany is operated by the <b>SID Swins Instit</b>	ute of Bioinformatics ( Terms of Line				e Windows tings to activ	esta da
	_						en ta fi
	_	neethestenstallionsethe	ASy home page				ex ta R
	_		ASy home page				est.ta.th
	]	Figure 60: ExPA	ASy home page				<b>Home   Cont</b>
	_	Figure 60: ExPA		~			
	]	Figure 60: ExP					
	]	Figure 60: ExPA TM Predictor of Transmetor	pred ane Regions and Orientation e-spanning regions and her orientation. The algorithm				

For further information see the TMbase document

very title (optional) collagen Protein sequence Input sequence Plain Text or ID or AC or Gi (see above for valid

Click

# TMbase - A database of membrane spanning proteins segments Biol. Chem. Hoppe-Seyler 374,165 lisage: Paste your sequence in one of the supported formats into the sequence field bel or age in the sequence in one of the sequence field shows the correct format and press the "Run TMpred" botton. Make sure that the format button (next to the sequence field) shows the correct format boose the minimal and maximal length of the hydrophic part of the transmembrane ha Output format html v minimum 17 v maximum 33 v

#### PREDICTING TRANSMEMBRANE REGION OF **PROTEIN SEQUENCE**

#### AIM

To predict the transmembrane region in the given protein sequence/ID.

# **INTRODUCTION**

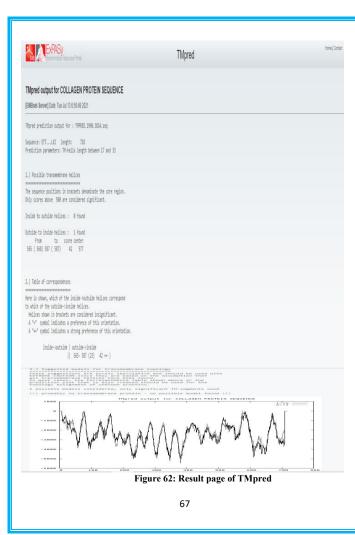
The orientation of membrane-spanning sections is predicted by the TMpred software. The technique is based on a statistical investigation of the transmembrane protein database known as TMbase. A combination of various weight matrices for scoring are used to make the prediction.

# PROCEDURE

- Go to ExPASy page, enter TMpred in search box (Figure 60)
- Click TMpred
- Go to the home page of TMpred tool (Figure 61)
- Paste the protein sequence in the search field
- Click Run TMpred (Figure 61)
- Results page is displayed (Figure 62)

Run TMpred Clear Input Figure 61: Home page of Mpred

PPGIPGRE6PKGSKGERGYPGIPGEKGDEGLOGIPGIPGAPGPTGPPGLHGRTG



# RESULT

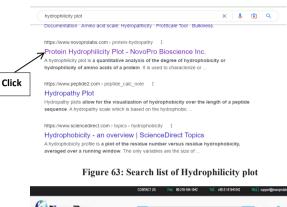
- Outside to inside helices : 1 found from to score center 565 (568) 587 (587) 42 577
- outside->inside

"++" symbol indicates a strong preference of orientation

(565-587 (23) 42 ++)

#### **OUTCOME**

Students comprehend the technique to predict the membranespanning regions and their orientation. They also understand that the transmembrane proteins act as gateways for transporting specific substances across the membrane.



NovoPro Products Services **Resources & Support** Company Protein Hydrophilicity Plot Quick Search Search keyword and Hit Enter A hydrophilicity plot is a quantitative analysis of the degree of hydrophobicity or hydrophilicity of amino acids of a protein It is used to characterize or identify possible structure or domains of a protein The plot has amino acid sequence of a protein on its x-axis, and degree of hydrophobicity and hydrophilicity on its y-axis There is a number of methods to measure the degree of interaction of polar solvents such as water with specific amin Services acids. For instance, the Kyte-Doolittle scale<sup>1</sup> indicates hydrophobic amino acids(employed in this tool), whereas the Hopp-Woods scale measures hydrophilic residues Gene Synthesis Site, Directed Mutanenesis 1. Sequence (Paste the raw sequence, not fasta format): Plasmid DNA Prenaration Full length () Input protein sequence here, Only 20 amino acids plus 'B', 'Z', 'X' can be accepted, others will be automatically removed. Protein Crystallography **Oustom Polycional Antibodie Oustom pentide Synthesis** E coli Everencion System Mammalian Cell Expression

Figure 64: Hydrophilicity home page

#### PREDICTING HYDROPHILICITYREGION IN THE PROTEIN SEQUENCE

# AIM

To predict the hydrophilic region in the given protein sequence

#### **INTRODUCTION**

The hydrophobicity or hydrophilicity of the amino acids in a protein is statistically analysed using a hydrophilicity plot. It is used to analyse or identify a protein's potential structure or domains. If the protein fragment is sufficiently hydrophobic to interact with or remain in a membrane can be predicted from the plot. The plot's xaxis represents the amino acid sequence of a protein, while the yaxis represents the degree of hydrophobicity and hydrophilicity. The degree of interaction between certain amino acids and polar solvents like water can be determined using a variety of techniques. For instance, the Hopp-Woods scale assesses hydrophilic residues while the Kyte-Doolittle scale identifies hydrophobic amino acids. Understanding the plot's shape reveals the details of the protein's partial structure. For instance, if a group of roughly 20 amino acids exhibits positive hydrophobicity, then it is possible that these amino acids are a portion of an alpha-helix that spans a lipid bilayer made up of hydrophobic fatty acids.

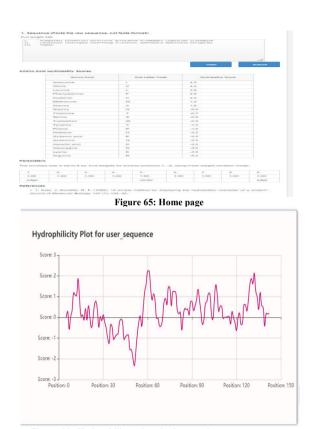


Figure 66: Hydrophilic regions in the protein sequence

On the other hand, amino acids with high hydrophilicity show that these residues are in touch with a solvent, such as water, and are therefore likely to be found on the protein's outer surface. First, a hydrophobicity rating between 4.6 and -4.6 is provided to each amino acid. The highest hydrophobic value is 4.6, and the least hydrophilic score is -4.6.

#### PROCEDURE

- Type hydrophilicity plot in the web browser and press enter
- Search list will be displayed, Click protein hydrophilic city plot –novoprolabs (Figure 63)
- Go to the home page (Figure 64)
- Paste the protein sequence in the given field
- Click submit (Figure 65)
- Results will be displayed (Figure 66)

#### RESULT

The hydrophilic residues in given Protein sequence was predicted.

# OUTCOME

Students learn to predict whether or not the protein segment has enough hydrophilicity to either interact with or reside in a membrane.



#### DETECTING ALIGNMENT OF REPEATS IN A PROTEIN SEQUENCES

#### AIM

To detect the alignment of repeats in a given protein sequence.

# **INTRODUCTION**

RADAR stands for Rapid Automatic Detection and Alignment of Repeats in protein sequences. RADAR identifies gapped approximate repeats and complex repeat architectures involving different types of repeats.

# PROCEDURE

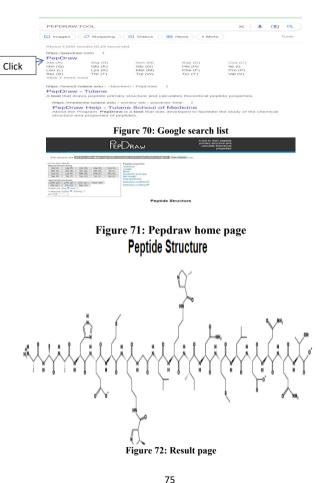
- Type Radar protein tool in the web browser and press enter
- Click the RADAR tool (ebi.uk)in the search list (Figure 67)
- Go to the RADAR home page (Figure 68)
- Paste the fasta protein sequence in the input box
- Click submit
- Result page was shown (Figure 69)

# RESULT

The given sequence contains  $\underline{2}$  alignment repeats.

#### **OUTCOME**

Students can able to predict the number of alignment repeats in the proteins sequences.



#### PREDICTING THE PEPTIDE STRUCTURE FOR THE **GIVEN PROTEIN SEQUENCE**

#### AIM

To predict the peptide structure of given protein sequence using Pepdraw tool

# INTRODUCTION

Pepdraw tool is used to draw primary peptide and also calculate the physico chemical properties.

# **PROCEDURE**

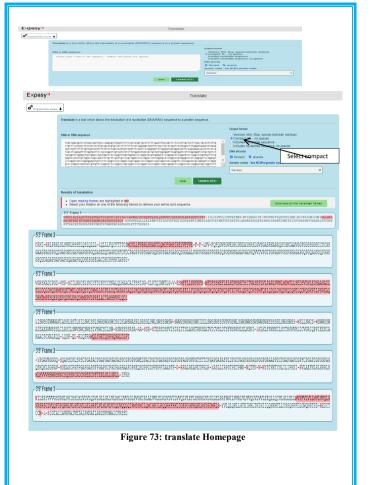
- Type Pepdraw tool in the web browser and press enter
- Click the Pepdraw in the search list (Figure 70) •
- Go to the Pepdraw home page (Figure 71)
- Paste the fasta protein sequence in the input box
- Click draw peptide •
- Result page will be displayed (Figure 72)

# RESULT

The protein sequence was converted into peptide chain.

# **OUTCOME**

Students acquire the skill to draw the primary chemical structure of an amino acid sequence and to predict chemical properties for any protein sequences.



#### CONVERSION OF NUCLEOTIDE SEQUENCES INTO PROTEIN SEQUENCES

#### AIM

To convert the nucleotide sequences into protein sequences and to identify the correct reading frame.

#### **INTRODUCTION**

Translate is a tool which allows the translation of a nucleotide (DNA/RNA) sequence into a protein sequence. Translate accepts a DNA sequence and converts it into a protein in the reading frame as specified. Translate supports the entire IUPAC alphabet and several genetic codes. A raw sequence or one or more FASTA sequences is pasted in the text area. Input limit is 200,000,000 characters. Determining is a complex process if a nucleic acid sequence actually codes for a protein. Because, generally it is not known which strand is the coding strand or which is the correct reading frame. Both these questions are resolved by translating both strands in all three reading frames and looking for the one that gives the longest amino acid sequence before a stop codon is encountered. A stop codon is expected to appear on average once for every 20 amino acids when reading a sequence in the incorrect frame. It is possible for an out of frame translation to extend over 100 amino acids before a stop codon is reached.

#### Translate is a tool which allows the translation of a nucleotide (DNA/RNA) sequence to a protein sequence Output format OVerbose: Met. Stop. spaces between residues DNA or RNA sequence O Compact: M - no spaces Includes nucleotide sequence atgtcggaagatctaaagcagattgcccaggagactgagtctctccgtaaagttgcattcttcggaattgcagtctctacaattgctactttgactgca Includes nucleotide sequence, no spaces attattgctgttccaatgctttacaactacatgcagcatgtgcaatcttctcttcaatcggaggttgaattctgccaacacagatcaaatggactttgg gatgagtataagagagtatgtttttttgttgaataattttaattttagttaaatgtttgatttcagttccaaggagtttctggagttgaaggacgtat DNA strands caagagagacgcatatcaccgtagcctcggaggtttctggtgcttcccgcaaggctcgtcgtcaatcttatggaaatgacgctgctgtcggaggattcgg Select nucleotide sequence with no tggatcatctggaggatcatgctgctgcaggatctggagctgctggaccagctggatcaccaggacaagatggagcaccaggaaacgatggagctcc 🛿 forward 🛛 reverse spaces aggagcaccaggaaaacccaggacaagatgcttctgaggatcaaactgctggaccagacagcttctgcttcgattgcccagctggaccaccaggaccatc aggageaccaggacaaaagggaccatcaggagettecaggageceeccaggacaatetggaggagetgetettettecaggaccagegaccagetggaccace Genetic codes - See NCBI's genetic codes aggaccagccegacaaccaggatccaacggaaacgccggaggctccaggagcaagtcgtcgatgttccaggaactccaggaccagctggacc Standard TRANSLATE! Results of translation Open reading frames are highlighted in red lowoload all the translated frames Select your initiator on one of the following frames to retrieve your amino acid sequence M S E D L K Q I A Q E T E S L R K V A F ttcggaattgcagtototacaattgctactttgactgcaattattgctgttcc tacaactacatgcagcatgtgcaatcttctcttccatcggaggttgaattctgccaaca Y N Y M Q H V Q S S L Q S E V E F C Q H agatcaaatggactttgggatgagtataagagagtatgtttttttg R S N G L W D E Y K R V C F F C attttagttaaatgtttgatttcagttccaaggagtttctggagttgaaggacgtatca I L V K C L I S V P R S F W S - R T Y ( gagagacgcatatcaccgtagcctcggagtttctggtgcttcccgcaaggctcgtcgtc E R R I S P - P R S F W C F P Q G S S S atottatggaaatgacgctgctgtoggaggattoggtggatoatotggaggatoatgo I L W K - R C C R R I R W I I W R I M I tcatgoggatotggagotgotggaccagotggatoacoaggacaagatggagoaco L M R I W S C W T S W I T R T R W S T aaacgatggagctocaggagcaccaggaaacccaggacaagatgcttotgagga K R M S S R S T R K P R T R C F - G tgctggaccagacagcttctgcttcgattgcccagctggaccaccaggaccatcagg C W T R Q L L L R L P S W T T R T I R accaggacaaaagggaccatcaggagctccaggagccccaggacaatctggaggag T R T K G T I R S S R S P R T I W R S tottocaggaccaccaggaccagotggaccaccaggaccagooggacaaccag S S R T T R T S W T T R T S R T T R cggaaacgcoggagetecaggageeccaggacaagtegtegatgttecaggaactees R K R R S S R S P R T S R R C S R N S accagotggaccaccaggatcaccaggaccagocggagotccaggacaaccagggcaa I S W T T R I T R T S R S S R T T R A

Figure 74: Result page of translate with nucleotide sequence with aminoacid

# PROCEDURE

- Select the nucleotide sequence, copy it and then paste it into the translate sequence window in the ExPASy translate tool
- Under Output format, select "Compact" or nucleotide sequence without space
- Click on Translate Sequence
- Result page is displayed (Figure 73 and 74).

# RESULT

- Output compact is selected; it gives the amino acid sequence as one letter code with stop codons indicated by a hyphen with different frames.
- Output nucleotide sequence without space is selected; it gives the nucleotide sequence with one letter code aminoacid with different frames.
- Red colour indicates the open reading frame.

# OUTCOME

Students gain knowledge to convert any unknown nucleotide sequence into protein sequence and to identify the most correct reading frame. They also acquire knowledge about the exons, pseudogenes, noncoding region of DNA and regulatory functions.

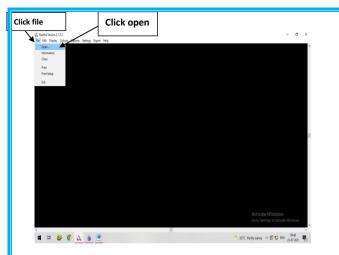


Figure 75 : Open the PDB file in RASMOL tool

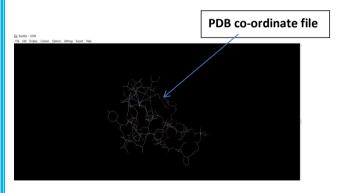


Figure 76: Protein structure open in molecular visualization tab

#### MOLECULAR VISUALIZATION USING RASMOL

#### AIM

To visualize the tertiary structure of protein molecule in graphic view and command line

#### **INTRODUCTION**

RasMol is free software for molecular visualization created by Roger Sayle. It is a molecular graphics programme intended for the visualization of proteins, nucleic acids and small molecules. The programme aims at display, teaching and generation of publication with quality images. The program reads in a molecule coordinate file and interactively displays the molecular screen in variety of colour schemes and molecular representations.

#### REQUIREMENT

RASMOL software, PDB molecule

# PROCEDURE

- Open a molecular visualization tool (Figure 75)
- From the file menu open a PDB atom co-ordinate file (Figure 76)
- Rotate the molecule

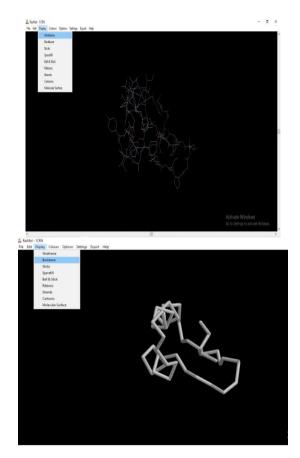


Figure 77:Visualise the different forms of protein structure

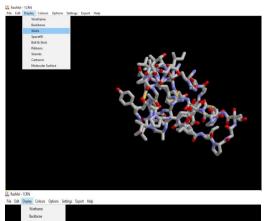
- Try various options (Figure 77,78,79)
- Try different commands in command line and visualize the changes in structure (Figure 80, 81)

84

- Save the required structural view
- Exit the application

# COMMANDS

- Select
- Colour
- Zoom on
- Zoom off
- Label on
- Label off
- Spacefill
- Star on
- Background pink
- Stereo on
- Stereo off
- Pick angle
- Pick distance
- Label 250
- Star
- Rotate



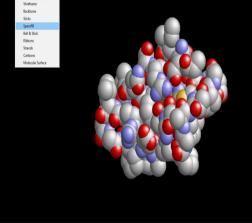


Figure 78: Visualise the different forms of protein structure

- Hbonds
- Wireframe
- Cartoon
- dots
- Quit

Other command reference:

Backbone	Background	Bond	Bulgarian	Cartoon	Centre	<u>Chinese</u>	Clipboard
<u>Colour</u>	ColourMode	Connect	<u>CPK</u>	<u>CPKnew</u>	Defer	<u>Define</u>	<u>Depth</u>
Dots	<u>Echo</u>	English	Execute	<u>Exit</u>	French	HBonds	<u>Help</u>
<u>Italian</u>	Japanese	Label	Load	<u>Map</u>	Molecule	Monitor	NoToggle
Pause	<u>Play</u>	Print	<u>Quit</u>	Record	Refresh	Renumber	Reset
Restrict	<u>Ribbons</u>	Rotate	Save	Script	Select	<u>Set</u>	Show
<u>Slab</u>	Source	<u>Spacefill</u>	<u>Spanish</u>	SSBonds	<u>Star</u>	Stereo	Strands
Structure	Surface	Trace	Translate	<u>UnBond</u>	Wireframe	<u>Write</u>	<u>Zap</u>
<u>Zoom</u>							

# RESULT

The tertiary structure of protein molecule is visualized in graphic view and command line

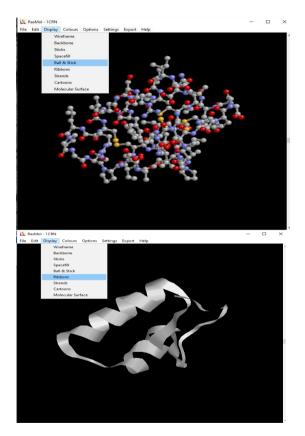


Figure 79: Visualise the different forms of protein structure

# OUTCOME

The students are imparted the ability to evaluate and interpret molecular models. Students can interpret the complicated molecule structure, properties, and interactions with the use of molecular visualisation tools. These resources aid their study in the fields of chemistry, pharmacology, biology, and bioinformatics.

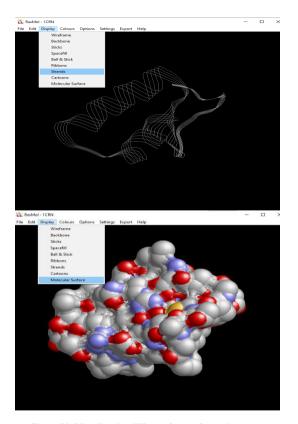


Figure 80: Visualise the different forms of protein structure

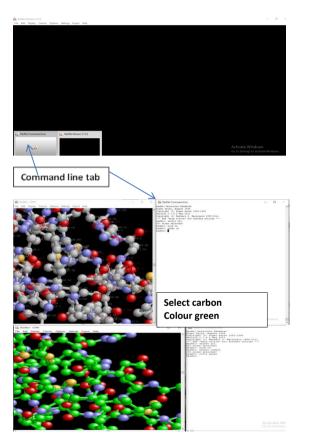


Figure 81: Various commands used in command line and visualize the changes in structure

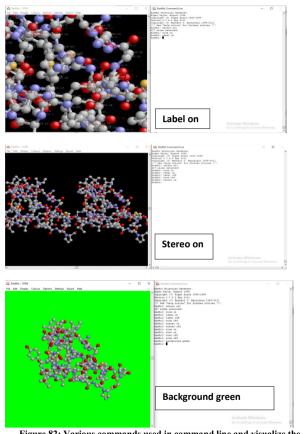


Figure 82: Various commands used in command line and visualize the changes in structure

# MEASUREMENT OF BOND LENGTH IN PROTEIN STRUCTURE USING RASMOL

# AIM

To measure the bond length of atoms in the protein structure and to visualize it in graphic view and command line.

#### **INTRODUCTION**

RasMol is a free software for molecular visualization created by Roger Sayle. It is a molecular graphics programme intended for the visualization of proteins, nucleic acids and small molecules. The program aims at display, teaching and generation of publication with quality images. The program reads in a molecule coordinate file and interactively displays the molecular screen in variety of colour schemes and molecular representations.

## PROCEDURE

- Open RasMol and import a file of Pdb atom coordinates
- Use the various menu options and get a composition of the molecule
- Set the display style to "ball and stick" (Preferable, but works with other displaying style as well)
- Use Shift+ mouse down to zoom in on the molecule to see the bonds more clearly.
- Open the command line window
- Type set picking distance and press enter key

- Open the display Window and select the two atoms participating in the bond formation by clicking on them successively.
- The command line window displays the bond length.
- Record the results
- Alternatively, to show bond and to measure the bond length between two atoms, type set picking monitor in the command line window.
- Now click on the two atoms again.
- A bond line appears (The bond is removed when a click on the atoms is made more than once.)
- Note the results from the command line window.

## RESULT

The distance between two atoms is 90.0Å.

# OUTCOME

Students learn how to find out the distance between two atoms using command line and directly picking atoms in structure.

#### MEASUREMENT OF BOND ANGLE IN PROTEIN STRUCTURE USING RASMOL

# AIM

To measure the bond angle between the atoms in the protein structure and to visualize it in graphic view and command line.

### **INTRODUCTION**

RasMol is a free and most popularly used software for molecular visualization created by Roger Sayle. It is a molecular graphics programme intended for the visualization of proteins, nucleic acids and small molecules. The program aims at display, teaching and generation of publication with quality images. The program reads in a molecule coordinate file and interactively displays the molecular screen in variety of colour schemes and molecular representations.

# PROCEDURE

- Open RasMol and load a file of Pdb atom coordinates
- Use the various menu options and get a feel of the molecule
- Set the display style to "ball and stick" (Preferable, but works with other displaying style as well)
- Use Shift + mouse down to zoom in on the molecule to see the bonds more clearly.
- Go to command line window.
- Type set picking angle and press enter key

- Go to display Window and select the three atoms forming the bond angle by clicking on them successively.
- The command line window displays the bond angle.
- Note the results.

## RESULT

The bond angle between atoms is 25.0°

## **OUTCOME**

Students acquire the skills necessary to find out the bond angle between atoms using command line and directly picking atoms in structure.

#### **WEBLINKS:**

- https://www.careerindia.com/courses/unique- courses/what-isbioinformatics-scope- career-opportunities-012034.html
- https://www.ncbi.nlm.nih.gov/
- https://www.ebi.ac.uk/Tools/sss/ncbiblast/
- https://en.wikipedia.org/wiki/Clustal
- https://www.ebi.ac.uk/Tools/msa/clustalo/
- https://www.bing.com/ck/a?!
- https://web.ExPASy.org/
- https://web.ExPASy.org/ProtParam/
- https://web.ExPASy.org/peptide\_mass/
- https://web.ExPASy.org/peptide\_cutter/
- https://www.academia.edu/3112992/CFSSP\_Chou\_and\_
   Fasman Secondary Structure Prediction Server/
- http://www.openrasmol.org/
- https://www.novoprolabs.com/tools/protein-hydropathy
- https://www.ebi.ac.uk/Tools/pfa/radar/
- https://pepdraw.com/
- https://ase.tufts.edu/biology/bioinformatics/exercise3.asp



#### V.V.VANNIAPERUMAL COLLEGE FOR WOMEN (Belonging to Virudhunagar Hindu Nadars)

(Belonging to Virudhunagar Hindu Nadars) An Autonomous Institution Affiliated to Madurai Kamaraj University Re-accredited with 'A' Grade (3<sup>ed</sup> cycle) by NAAC VIRUDHUNAGAR.



# About the Author

Sudha Rameshwari is a Biochemist and her native is Kingmaker Kamaraj birth town Virudhunagar, Tamilnadu, India, She did her B.Sc. and M.Sc. (Biochemistry) from V.V.Vanniaperumal College for Women. Virudhunagar, Tamilnadu, India. She obtained her M.Phil degree in Life Science (2003) from Manonmaniam Sundaranar University, Tirunelveli. She has completed her Post graduate diploma in Bioinformatics (2005) in Bharathiyar University, Coimbatore. She has completed 24 years' of teaching experience in Biochemistry and her Research interests include Microbiology, Pharmacology, Green nanotechnology and Bioinformatics. She submitted one sequence to GEN BANK on April 2018. She regularly teaches Techniques, Enzymology, Clinical biochemistry, Microbial biochemistry and Bioinformatics. She is well trained in Microbiology and Molecular Biology Techniques. She has published more than 30 research articles in well reputed International journals which are Scopus indexed (8), Web of Science (4) and UGC approved. She is interested in participating and to organize workshops. She guided 2 M.Phil students and 25 PG students. She has filed and published one patent (Indian) Publication. She is a reviewer in 15 reputed International journals. She also published 3 chapters in edited books. She received grants from TNSCST-DBT, autonomy fund and MRP grant from VVVCMB-MRP Scheme sponsored by VVVCollege Manage Board. She has also been awarded as Honorary Doctorate (D.Litt.(hc)) from University of Central America, CV Raman Prize 2022 from Institute of Researchers, Wayanad, Kerala and Bharat Excellence award 2022, Leading educationist of India Award 2022 from Friendship forum, New Delhi, Global Personalities of Asia 2022 from Global brotherhood forum, New Delhi and Outstanding Researcher in Microbial Biochemistry from VIHA, 2018. She is a life member (LM052202) in Institute of Researchers, Waynad and also SAS eminent Fellow Membership (SAS/SEFM/077/2021) in Scholars Academic and Scientific society. She has scored 8.7/10 in VIDWAN expert database and National Researchers network