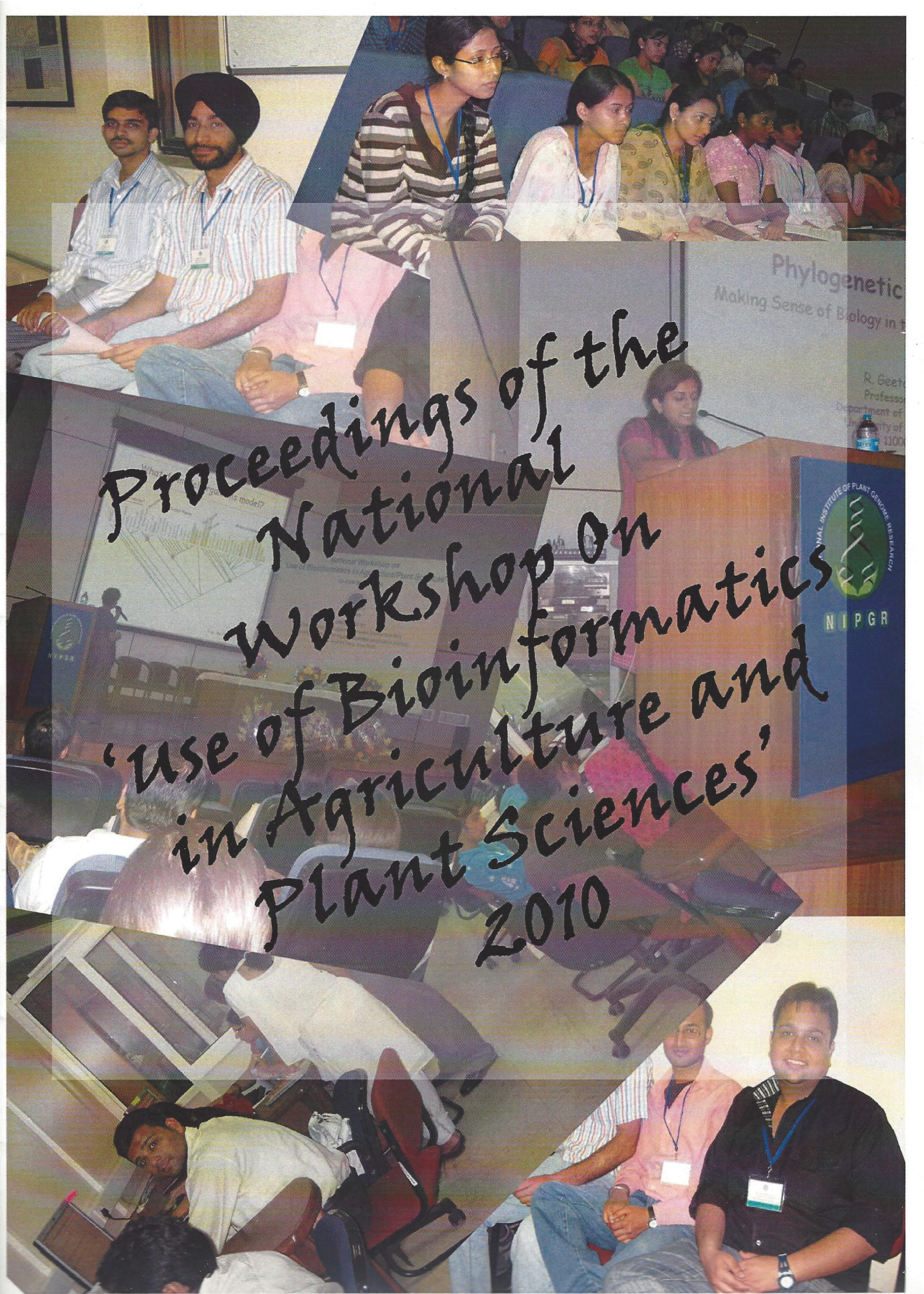


Proceedings of the
National
Workshop on
'Use of Bioinformatics
in Agriculture and
Plant Sciences'
2010



The BTIS-NET Sub-DIC at NIPGR organized the second Bioinformatics Workshop titled 'Use of Bioinformatics in Agriculture/ Plant Sciences' on 12-13th March 2010.

The endeavor of the organizing committee is to make this workshop an annual event, as per the BTIS-NET scheme, for Ph.D. scholars, postdoctoral researchers and teachers from various colleges and universities, in order to provide an opportunity to all participants to benefit from the rich experience and expertise available in bioinformatics in India.

Twenty two participants were selected for the current workshop based upon abstracts received from all candidates (page X). The Inaugural session of the event was held at 10 am on the 12th of March, 2010. The welcome address was delivered by the Sub-DIC coordinator, while Prof. R. Geeta, renowned plant biologist and phylogeneticist from Univ of Delhi, Dept of Botany, delivered the Inaugural lecture titled "Phylogenetic Biology: Making Sense of Biology in the Light of Evolution". She stressed upon the need for young scientific minds to hypothesize and to remain aware of past research when designing experiments, giving examples from her wide experience and her work in the evolution of the monocot leaf and the Angiosperm endosperm.

The remaining sessions of the workshop included, alternatively, three lecture sessions and five tutorial sessions (Page X). Young scientists working in the area of plant computational biology were invited for the lecture sessions, whereas previously designed hands-on practicals were imparted in the tutorial sessions. These included tutorials on Sequence Analysis, Data Interpretation, EST data analysis and structure prediction as well as modeling exercises.

During the First lecture session, Dr. G. Yadav, NIPGR, gave a lecture titled "Bioinformatics in Agriculture", describing the varied applications of computational biology in emerging areas like precision agriculture, crop improvement and evolutionary studies. Dr. S. M. Leighton, Univ of Delhi, in her lecture entitled "Bioinformatics and the pedagogy of Plant Sciences" described with examples, various steps in the design and implementation of a bioinformatics problem and its solution using freely available online programs and databases.

The first tutorial session gave the participants insights into sequence data retrieval and analysis, using web based tools. The next practical session included an exercise to enable the participants to carry out this analysis, using the example of wheat protein agglutinin.

On the second day of the workshop, participants were given a tutorial on EST data analysis and the methods involved therein. Later they were introduced to the concept of and protein structure. Exercises were designed using the example of wheat protein. The last hands-on session involved protein family analysis and phylogenetics. Summary slides from all lectures have been provided on Page X.

The second lecture session held on this day included a lecture by Dr. D. Chattopadhyay, NIPGR. His talk was titled 'Next generation sequencing and its applications in genomics and epigenomics'. Dr. Chattopadhyay elaborated upon the history of genome sequencing, the methods involved and described in details the applications of the new and futuristic methods with specific examples from plant genomes.

The valedictory session included an informal interaction with all participants who appreciated all the sessions and gave positive feedback on the workshop. On behalf of the Sub-DIC, I thank all participants as well as the NIPGR administration for making this event a success.

*Gitanjali Yadav
(Staff Scientist & Workshop Coordinator)*

Protein structure prediction and visualization

This exercise is concerned with various online tools available to study a protein structure. It is shown where you can find these tools, how to use them and most importantly how to interpret the results.

Part I: Predict the secondary structure of a protein sequence

Part 2: From the Primary Structure to the 3-D Structure

Part III: Visualization of protein 3D structure

Part IV(a): Gene and Protein sequences

Part IV(b): Translation of the nucleotide sequence


Part V: View the overall information of PDB

1.)

When the sequence of amino acids are linked by hydrogen bond is called secondary protein structure.

There are three types of local segments

- ✓ Helices
- ✓ Extended or Beta-strands
- ✓ Random coils



Helix

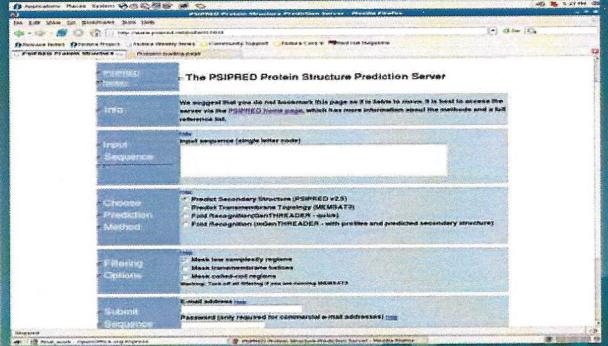
Beta-sheet

Loop

2.)

Getting Started :: PSIPRED

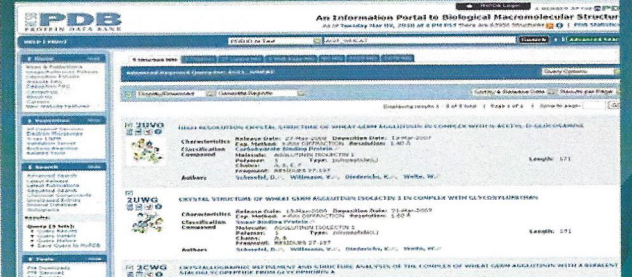
- Open the PSIPRED Server from <http://bioinf4.cs.ucl.ac.uk:3000/psipred/>



3.)

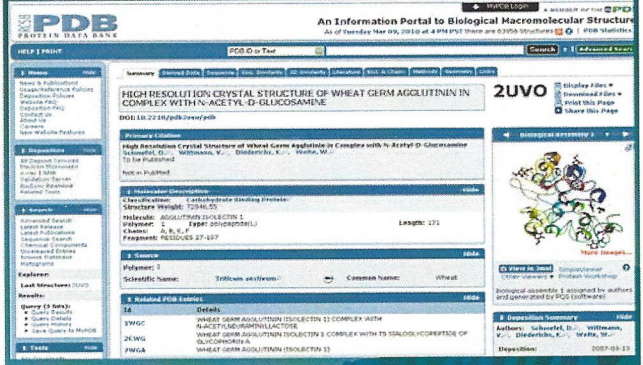
Retrieve protein from PDB

- Open Protein Data Bank (PDB) www.rcsb.org
- Enter a "Keyword" search by typing a keyword in the text box on the search bar at the top of the search box: keyword like "AGH1_WHEAT"
- Click the adjacent SEARCH button
- The result page will contain a list of PDB's entry name i.e. of four letters associated with this protein name called AGH1_WHEAT" eg 2UVO, 2UWG,



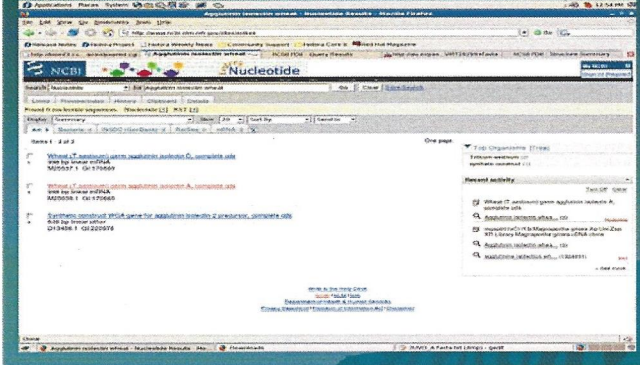
4.)

- Click on the 2UVO for further analysis of its structure
- Download PDB file save it on your computer



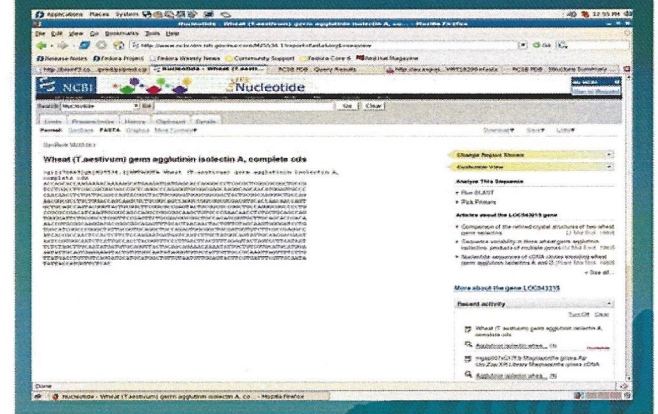
5.)

- Point your web browser to the following URL: <http://www.ncbi.nlm.nih.gov/>
- In the top search box select Nucleotide and enter the text "Agglutinin isolectin wheat"
- Click on the Go button



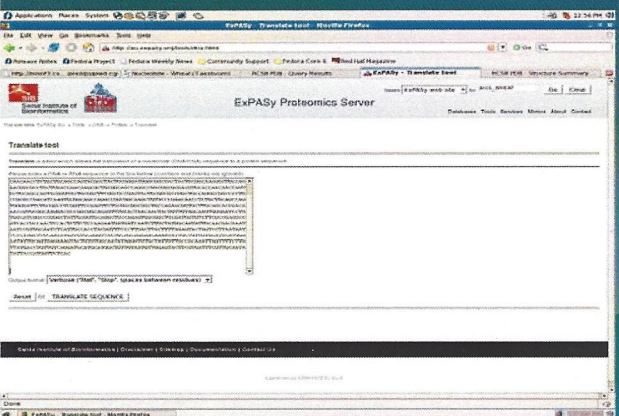
6.)

Obtain the FASTA sequence of Agglutinin isolectin wheat



7.)

Translate gene to protein sequence by expasy tool.



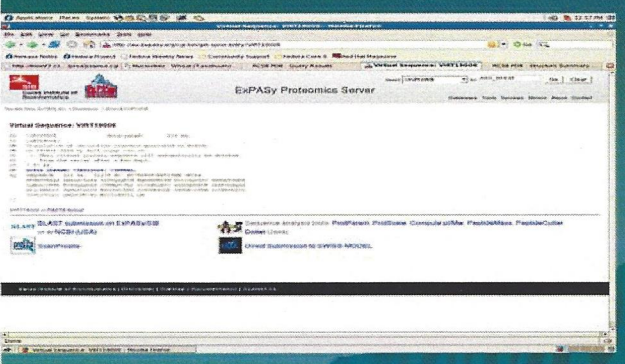
8.)

Translated protein sequence in six Open Reading Frame

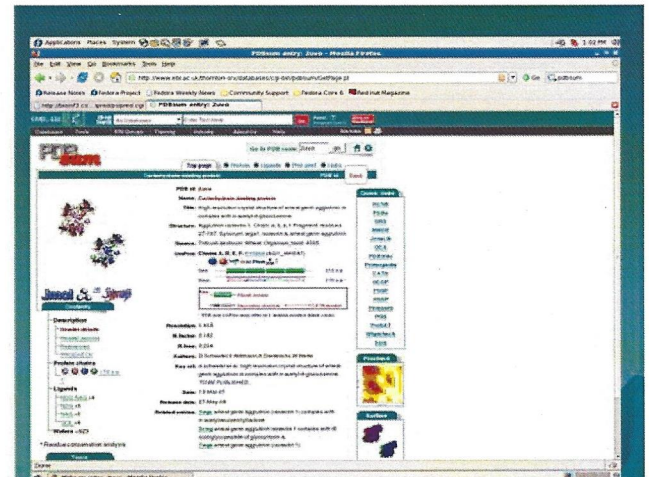


9.)

- Click on any frame six of them.
- The result page highlights the Methionine residues, or the starting point of the protein sequence.
- Click the first 'M' in the sequence.



10.)



Biological Data Retrieval & Analysis

1.)

NCBI Entrez Gene

Search Gene for calnexin AND Cytochrome

Enter terms and click Preview to see only the number of search results.

To save search indefinitely, click query # and select Save in My NCBI.

To combine searches use search, e.g., #2 AND #3 or click query # for more options.

Search: #1 Search calnexin

Add Term(s) to Query or View Index:

Enter a term in the text box; use the pull-down menu to specify a search field.

Click Preview to add terms to the query box and see the number of search results, or click Index to view terms within a field.

2.)

NCBI HomoloGene

HomoloGene:1324. Gene conserved in Eukaryota

Genes identified as putative homologs of one another during the construction of HomoloGene.

Accession	Protein
NP_001737.1	Calnexin
NP_00191434.1	Calnexin
NP_01060222.1	Calnexin
NP_031623.1	Calnexin
NP_742095.1	Calnexin
NP_011007571.1	Calnexin
NP_006613.1	Calnexin
NP_430170.1	Calnexin
NP_520812.1	Calnexin
NP_000343.1	Calnexin
XP_014001.1	Calnexin

3.)

Pfam

Family: Calreticulin (PF00262)

Domain organization: Calreticulin family

Structure: 3D model of Calreticulin

Gene Ontology: Molecular function, Biological process, Cellular component

External database links: UniProt, PDB, TrEMBL, etc.

4.)

NCBI Nucleotide

Search results for Glycine max calnexin mRNA, complete cds

Accession numbers: U00006, U00007, etc.

Links: Change Region Sizes, Custom View, etc.

5.)

BLAST Basic Local Alignment Search Tool

Enter Query Sequence: g1660003.gn[AAAGNSA 1] calnexin [Cytochrome]

Database: Reference proteins [refseq_protein]

Program Selection: blastp (protein-protein BLAST)

BLAST Search database Reference proteins [refseq_protein] using blastp (protein-protein BLAST)

6.)

Sequences producing significant alignments

Accession	Description	Hit score	Total score	Query coverage	E value
U00006	Calnexin (Homo sapiens)	911	911	92%	0.0
U00007	Calnexin (Mus musculus)	856	856	88%	0.0
U00008	Calnexin (Rattus norvegicus)	856	856	88%	0.0
U00009	Calnexin (Sus scrofa)	856	856	88%	0.0

Working with a Single Protein Sequence

This exercise is concerned with various online tools available to study a protein before designing a new experiment in the lab. It is shown where you can find these tools, how to use them and most importantly how to interpret the results.

Part I: Predicting the main physico-chemical properties of a protein

Part II: Digesting a protein in a computer

Protease digestions — where you use an enzyme to cut your protein in specific ways — can be useful if you're only interested in carrying out experiments on a portion of your protein.

Part III: Primary Structure Analysis

A primary structure analysis is conducted to find segments in a protein that display a special composition. These segments can reveal some interesting properties of your protein.

Note: For the following analysis, obtain the protein sequence Q39817 (CALX_S04BN) from SwissProt.

Part IV: Predicting post-translational modifications in your protein.

Part V: Finding Known Domains in your Protein

Note: In many interesting domains, a particular type of amino acid is overrepresented. For instance, there are more leucines than expected by chance in the leucine zippers, or more glycines than expected by chance in the glycine-rich domains — and so on — for many domains. Because repeated residues make a sequence simpler, sequences that contain them are described as low-complexity. If you keep the Apply Low Complexity Filter box checked, you may lose these domains.

1.)

Enter the SwissProt ID for WGA protein at: <http://www.expasy.ch/sprot/>

Predicting the main physico-chemical properties of a protein

UniProt Knowledgebase - Swiss Prot and TrEMBL

Swiss Institute of Bioinformatics

ExPASy Proteomics Server

UniProt Protein knowledgebase

TrEMBL

Computer-annotated supplement to Swiss-Prot

The UniProt Knowledgebase consists of:

- Multi-FASTA Swiss-Prot, a curated protein sequence database which strives to provide a high level of annotation (such as the description of the function of a protein, its domain structure, post-translational modifications, variants, etc.), a minimal level of redundancy and high level of integration with other databases (More details)
- UniProtKB TrEMBL, a computer-annotated supplement of Swiss-Prot that contains all the translations of EMBL nucleotide sequence entries not yet integrated in Swiss-Prot

These databases are developed by the Swiss-Prot groups at SIB and at EBI

2.)

Retrieve sequence for WGA in Fasta format

UniProtKB

Search in: Protein Knowledgebase (UniProtKB)

Reviewed, UniProtKB/Swiss-Prot P10968 (AG11_WHEAT)

Last modified March 2, 2010, Version 99, History...

Names and origin	
Protein names	Recommended name: Agglutinin-like lectin 1
	Alternative name(s): WGA1
	Isolated from: Triticum A.
Digestion	Triticum aestivum (Wheat)
Taxonomic identifier	4562 [NCBI]

9.)

Finding Known Domains in your Protein

The Main Domain Collection

Name	Web Address	Number of Domains	Generation
PROSITE-Profile (IP)	www.expasy.org/prosite	618	Manual
PlainA (IP)	www.sanger.ac.uk/Software/Pfam	7973	Manual
PRINTs (IP)	www.bioinf.man.ac.uk/dbbrowser/PRINTs	1900	Manual
PRODOM (IP)	protein.toulouse.inra.fr/prodom/current/html/home.php	736000	Automatic
SMART (IP)	smart.embl-heidelberg.de	685	Manual
COGS	www.ncbi.nlm.nih.gov/COG/new/	4852	Manual
TIGRFAM (IP)	www.tigr.org/TEIGRFAMs	2453	Manual
BLOCKS	blocks.fhcrc.org/	12542	Automatic

10.)

11.)

12.)

13.)

14.)

More Protein Analysis for Free over the Internet

Name	Site	Description
ExPASy	www.expasy.org/tools	Proteins
Pbl	npca-pbl1.lbcp.fr	Proteins
PIR	pir.georgetown.edu	Proteins
CES	www.cbe.stu.dk/services	Proteins
Hits	hits.isb-sib.ch/	Proteins
InterPro	www.ebi.ac.uk/interpro/scan.html	Domains
CD search	http://www.ebi.ac.uk/InterProScan/	Domains

Protein Family Analysis & Phylogenetics

1.)

Gathering sequences with online BLAST servers
<http://www.expasy.ch/tools/blast/>

2.)

The guide tree is NOT a phylogenetic tree.

To obtain a phylogenetic tree from a ClustalW server, you must cut and paste an actual multiple sequence alignment — not a set of unaligned sequences.

3.)

Success Table

Sort by Sequence Number View Output File

Seq Name	Length	Depth	Score
1 #P109817 CAL2_S0038	546	3	16187762 STPLAQ_MSDTR 545 75
1 #P109817 CAL2_S0038	546	3	16187762 STPLAQ_MSDTR 545 75
1 #P109817 CAL2_S0038	546	3	16187762 STPLAQ_MSDTR 545 75
1 #P109817 CAL2_S0038	546	3	16187762 STPLAQ_MSDTR 545 75
1 #P109817 CAL2_S0038	546	3	16187762 STPLAQ_MSDTR 545 75

PLEASE NOTE: Some scores may be missing from the above table if the alignment was done using remote CPU nodes. Please check the log.

Sort by Sequence Number View Output File

Alignment

Hide Colors View Alignment File

ClustalW 3.0.10 multiple sequence alignment

```

16187762|STPLAQ_MSDTR 545 75 16187762|STPLAQ_MSDTR 545 75 16187762|STPLAQ_MSDTR 545 75
16187762|STPLAQ_MSDTR 545 75 16187762|STPLAQ_MSDTR 545 75 16187762|STPLAQ_MSDTR 545 75
16187762|STPLAQ_MSDTR 545 75 16187762|STPLAQ_MSDTR 545 75 16187762|STPLAQ_MSDTR 545 75
16187762|STPLAQ_MSDTR 545 75 16187762|STPLAQ_MSDTR 545 75 16187762|STPLAQ_MSDTR 545 75
16187762|STPLAQ_MSDTR 545 75 16187762|STPLAQ_MSDTR 545 75 16187762|STPLAQ_MSDTR 545 75

```

4.)

Aligning sequences & structures with Tcoffee

It uses a principle that's a bit similar to ClustalW, but it yields more accurate alignments at the cost of a slightly longer running time.

Tcoffee builds a progressive alignment like ClustalW, but it compares segments across the entire sequence set.

Usage	Description
TCOFFEE	Produce a multiple sequence alignment with Tcoffee.
CORE	Evaluate the reliability of an existing multiple alignment.
MCOFFEE	Run any requested Multiple Sequence Alignment package and combine all the output into one final alignment.
EXPRESSO	Incorporate all the available structural information in your alignment. Will produce the best sequence alignments if the structures are available.

5.)

http://www.tcoffee.org/

Network Support Shop Products Training

TCoffee

HOME | references | help

A collection of tools for Computing, Evaluating and Manipulating Multiple Alignments of DNA, RNA, Protein Sequences and Structures

Minor sites: [Tcoffee](#) [Tcoffee](#) [Tcoffee](#) [Tcoffee](#)

Tool	Regular	Advanced	File
TCOFFEE	Regular	Advanced	File
EXPRESSO(JCOFFEE)	Regular	Advanced	File
MCOFFEE	Regular	Advanced	File
KCOFFEE	Regular	Advanced	File
COMBINE	Regular	Advanced	File

Tool	Regular	Advanced	File
CORE	Regular	Advanced	File
RMBS-A-PDB	Regular	Advanced	File
PROTOGENE	Regular	Advanced	File

Minor sites: [Tcoffee](#) [Tcoffee](#) [Tcoffee](#) [Tcoffee](#)

6.)

TCoffee

HOME | references | help

TCOFFEE :: Regular

You may still remain available on this server over the next 9 days. It will then be deleted. Do not forget to bookmark the URL, or use it for further reference.

RESULTS

Multiple Alignment	chr26a_01a	chr26a_01b	chr26a_01c	chr26a_01d	chr26a_01e
Tree	chr26a_01a	chr26a_01b	chr26a_01c	chr26a_01d	chr26a_01e
System files	LSX1	Command line			
Input(s)	k(TCOFFEE)2765_1596.pdf				

SEND RESULTS

to ProtGene: testing amino acid alignments into fasta file CDF multiple alignments
 to MSA hub: MySite: a new interactive server for protein annotation and domain identification

Home Server: TCOFFEE :: Regular | TCOFFEE :: Advanced

© 2006 Institute of Bioinformatics | V. 3.1

7.)

8.)

9.)

10.)

11.)

12.)

PROGRAM DETAILS FOR THE NATIONAL WORKSHOP ON 'Use of Bioinformatics in Agriculture & Plant Sciences'

12-13th March 2010

DAY 01 (12th March 2010)

INAUGURAL SESSION

9:30 am	REGISTRATION	Reception, NIPGR
10:00 am	WELCOME ADDRESS	Dr. G. Yadav, Coordinator, Sub-Dic
10:05 am	KEYNOTE ADDRESS	Dr. T. Madhan Mohan, DBT
10:10 am	INAUGURAL ADDRESS	Prof. R Geeta, University of Delhi
11:00 am	VOTE OF THANKS	Coordinator, Sub-DIC
11:05 am	HIGH TEA	

WORKSHOP SESSION - I

11:30 am	BIOINFORMATICS IN AGRICULTURE	Dr. G. Yadav, NIPGR
12:15 pm	BIOLOGICAL DATA RETRIEVAL & ANALYSIS: DNA	Sachin Pundhir
14:00 pm	LUNCH	

WORKSHOP SESSION - II

14:30 pm	PROTEIN SEQUENCE ANALYSIS: WHEAT PROTEIN	Smriti Shridhar
15:45 pm	TEA	
16:00 pm	BIOINFORMATICS & THE PEDAGOGY OF PLANT SCIENCES	Dr. S.M. Leighton, University of Delhi

DAY 02 (13th March 2010)

WORKSHOP SESSION - III

9:30 am	METHODS FOR ANALYSIS OF EST DATA	Sachin Pundhir
11:00 am	TEA	
11:15 am	NEXT GEN SEQUENCING & APPLICATIONS	Dr. D. Chattopadhyay, NIPGR
13:00 pm	LUNCH	

WORKSHOP SESSION - IV

14:00 pm	3-D STRUCTURAL ANALYSIS OF PROTEINS	Daljit Singh
15:45 pm	TEA	
16:00 pm	PROTEIN FAMILY ANALYSIS & PHYLOGENETICS	Smriti Shridhar

VALEDICTORY FUNCTION

17:30 pm	CERTIFICATE DISTRIBUTION	Prof. A.K Tyagi, DIRECTOR, NIPGR
17:45 pm	CLOSING CEREMONY	

**PARTICIPANTS OF THE NATIONAL WORKSHOP ON
'Use of Bioinformatics in Agriculture & Plant Sciences'**

12-13th March 2010

<i>Name of Participant</i>	<i>Affiliation</i>
1 Ravi Kant	Haryana Agricultural University
2 Poonam Sharma	CCS Haryana Agricultural University
3 Harmeet Kaur	NIPGR
4 Dr. Neelam R. Yadav	CCS Haryana Agricultural University
5 Rehna Augustine	NIPGR
6 Pooja Verma	NIPGR
7 Rahul Gautam	National Institute of Technology
8 Renu Kumari	NIPGR
9 Swati Chaudhary	NIPGR
10 Ashutosh Pandey	Shobhit University
11 Anurag Srivastava	Shobhit University
12 Abhay Pratap	Shobhit University
13 Pritika Singh	GNDU
14 Dr Parveen Chhuneja	Punjab Agricultural University
15 Khagendra Kumar	National Institute of Technology
16 Praveena K.	NIPGR
17 Meenu	NIPGR
18 Dr. Sridhar Gutam	Indian Agricultural Research Institute
19 Bharu Prakash Petla	NIPGR
20 Tanima Datta	GNDU
21 Harsimran Kaur	GNDU

*THE NATIONAL WORKSHOP ON
'Use of Bioinformatics in Agriculture & Plant Sciences'*

12-13th March 2010

Organisers

Dr. Gitanjali Yadav

Coordinator, NIPGR

Sachin Pandhir

NIPGR

Smriti Shridhar

NIPGR

Subhasish Mondal

NIPGR

Speakers

Prof. R. Geeta

University of Delhi

Dr. Sadeshna M. Leighton

University of Delhi

Dr. Gitanjali Yadav

Coordinator, NIPGR

Dr. Debasis Chattopadhyay

NIPGR

NIPGR