BHARATHIDASAN UNIVERSITY

Tiruchirappalli- 620024,
Tamil Nadu, India.

Programme : M.Sc., Biomedical Science

Course Title : Bioinformatics

Course Code : BM35S1BI

Unit - II

TOPIC: PROSITE and BLOCKS

Dr. P. JEGANATHAN
Guest Lecturer
Department of Biomedical Science

PATTERN AND MOTIF SEARCHES:PROSITE AND BLOCKS

BIO INFORMATICS-DEFINITION

1. THE SCIENCE OF COLLECTING AND ANALYSING COMPLEX BIOLOGICAL DATA SUCH AS GENETIC CODES.

BIOLOGICAL DATABASE

BIOLOGICAL DATABASES ARE LIBRARIES OF BIOLOGICAL SCIENCES, COLLECTED FROM SCIENTIFIC EXPERIMENTS, PUBLISHED LITERATURE, HIGH-THROUGHPUT EXPERIMENT TECHNOLOGY, AND COMPUTATIONAL ANALYSIS

BIOLOGICAL DATABASE -TYPES

BIOLOGICAL DATABASE CAN BE CLASSIFIED INTO TWO TYPES:

THEY ARE

1. PRIMARY DATABASE
2. SECONDARY DATABASE

PRIMARY DATABASE

- PRIMARY DATABASE IS ALSO CALLED AS ARCHIEVAL DATABASE
- THEY ARE EXTREMELY POPULATED WITH EXPERIMENTALLY DEREIVED DATA SUCH AS

1. NUCLEOTIDE SEQUENCE
2. PROTEIN SEQUENCE

SECONDARY DATABASE

- RESULTS FROM ENTRIES OF PRIMARY DATABASE
- MANUALLY CREATED OR AUTOMATICALLY ARRANGED
- IT CONTAINS SOME VALUABLE INFORMATION SUCH AS MUTATIONS

OR EVALUTIONARY RELATIONSHIPS

SECONDARY DATABASE

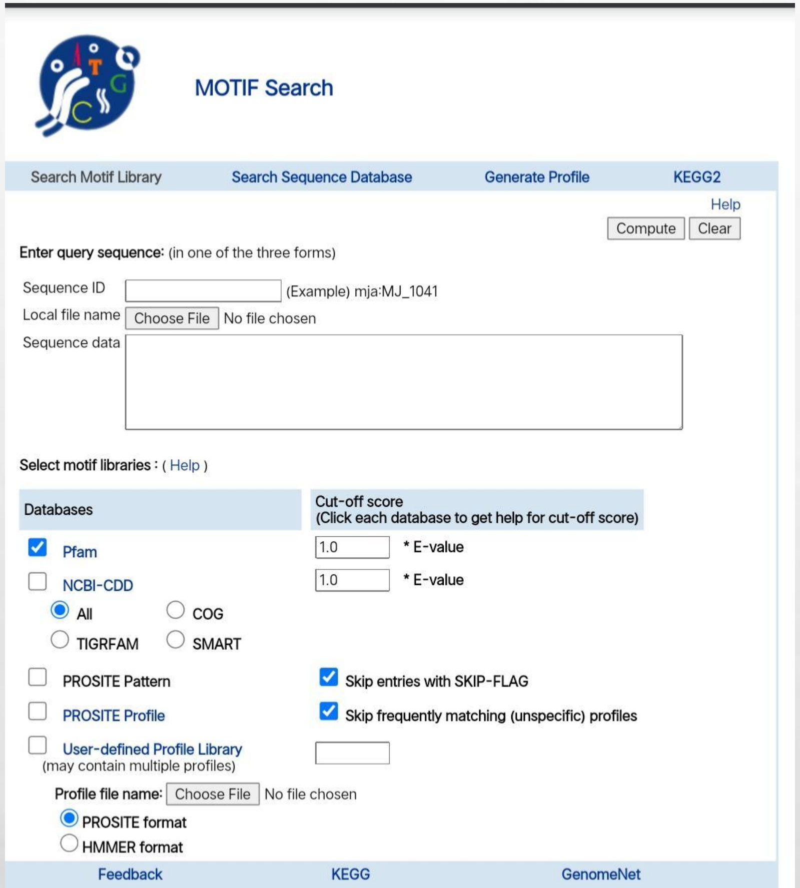
- THE EXAMPLES OF SECONDARY DATABASE INCLUDE

1. PROSITE
2. PFAM
3. BLOCKS
4. PRINTS

MOTIF SEARCHES

- MOTIF IS A SEARCH SERVICE PROVIDED BY GENOMENET TO SEARCH WITH A PROTEIN QUERY SUBSTANCE AGAINST MOTF LIBRARIES.
- SUPPORTS SEVERAL MOTIF DATABASES SUCH AS PROSITE,BLOCKS , PFAM, PRODOM,AND PRINTS.
- ALLOWS YOU TO SEARCH PROTEIN SEQUENCE LIBRARIESWITH YOUR PATTERNS.

MOTIF SEARCHES



Enter query sequence: (in one of the three forms)

Sequence ID [ ] (Example) mja:MJ_1041
Local file name Choose File No file chosen
Sequence data [ ]

Select motif libraries : (Help)

Databases	Cut-off score (Click each database to get help for cut-off score)
☑ Pfam	1.0 * E-value
☐ NCBI-CDD	1.0 * E-value
● All ○ COG ○ TIGRFAM ○ SMART	
☐ PROSITE Pattern	☑ Skip entries with SKIP-FLAG
☐ PROSITE Profile	☑ Skip frequently matching (unspecific) profiles
☐ User-defined Profile Library (may contain multiple profiles)	[ ]

Profile file name: Choose File No file chosen
● PROSITE format
○ HMMER format

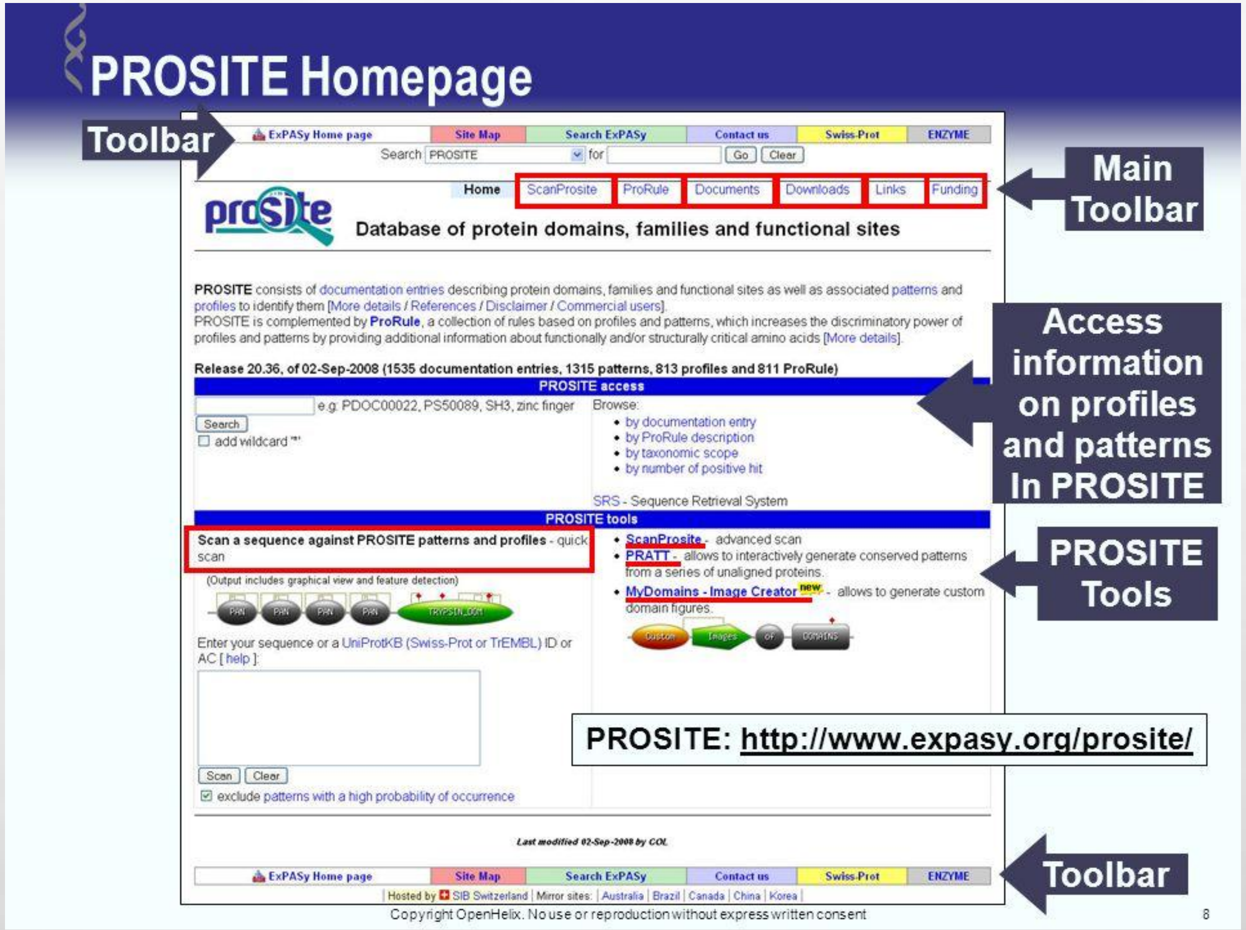
MOTIVE SEARCHES IN SEQUENCE DATABASES

- SEVERAL DATABASES EXIST WHICH PROVIDE INFORMATION ABOUT CONSERVED REGIONS WITHIN PROTEINS. ONE CAN QUERY THESE DATABASES WITH OWN SEQUENCES OR SEQUENCE BLOCKS. THE FOLLOWING WEBSITES CAN BE USED TO ACCESS SEVERAL MOTIF SEARCH ALGORITHMS:
- OVERVIEW ABOUT DATABASE WITH PROTEIN SEQUENCE MOTIFS
- OVERVIEW ABOUT DATABASE WITH PROTEIN DOMAINS
- OVERVIEW ABOUT DATABASE OF INDIVIDUAL PROTEIN FAMILIES

PROSITE

- PROSITE IS A PROTEIN DATABASE.IT CONSISTS OF ENTRIES DESCRIBING THE PROTEIN FAMILIES DOMAINS AND FUNCTIONAL SITES AS WELL AS AMINO ACID PATTERNS AND PROFILES IN THEM.
- THEY ARE MUTUALLY CURATED BY THE TEAM SWISS INSTITUTE OF BIOINFORMATICS AND TIGHTLY INTEGRATED INTO SWISS PROT PROTEIN ANNOTATION.

PROSITE –HOMEPAGE



PROSITE Homepage

Toolbar

Main Toolbar

Database of protein domains, families and functional sites

PROSITE consists of documentation entries describing protein domains, families and functional sites as well as associated patterns and profiles to identify them [More details / References / Disclaimer / Commercial users].
PROSITE is complemented by **ProRule**, a collection of rules based on profiles and patterns, which increases the discriminatory power of profiles and patterns by providing additional information about functionally and/or structurally critical amino acids [More details].

Release 20.36, of 02-Sep-2008 (1535 documentation entries, 1315 patterns, 813 profiles and 811 ProRule)

PROSITE access

e.g. PDOC00022, PS50089, SH3, zinc finger

☐ add wildcard '*'

Browse:

- by documentation entry
- by ProRule description
- by taxonomic scope
- by number of positive hit

SRS - Sequence Retrieval System

Access information on profiles and patterns In PROSITE

PROSITE tools

Scan a sequence against PROSITE patterns and profiles - quick scan

(Output includes graphical view and feature detection)

Enter your sequence or a UniProtKB (Swiss-Prot or TrEMBL) ID or AC [help]:

☑ exclude patterns with a high probability of occurrence

- **ScanProsite** - advanced scan
- **PRATT** - allows to interactively generate conserved patterns from a series of unaligned proteins.
- **MyDomains - Image Creator** new - allows to generate custom domain figures.

PROSITE Tools

PROSITE: http://www.expasy.org/prosite/

Last modified 02-Sep-2008 by COL

Toolbar

| Hosted by SIB Switzerland | Mirror sites: | Australia | Brazil | Canada | China | Korea |

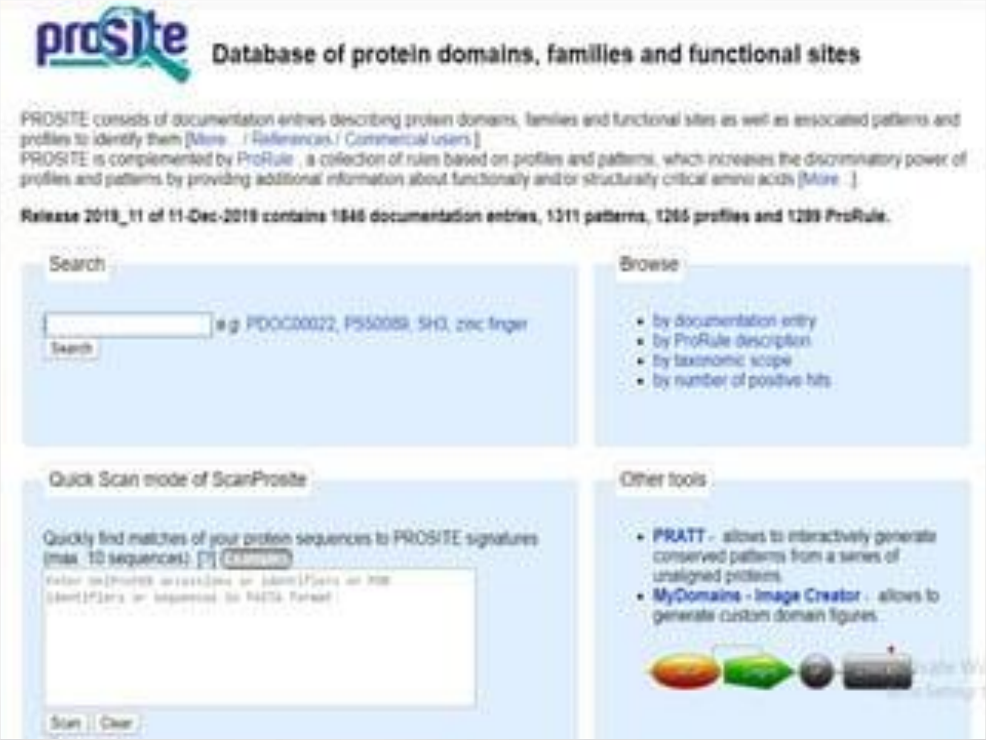
Copyright OpenHelix. No use or reproduction without express written consent

8

PROSITE DATABASE

- PROSITE,A PROTEIN DOMAIN DATABASE FOR FUNCTIONAL CHARECETERIZATION AND ANNOTATION
- PROSITE IS COMPLEMENTED FROM PRO-RULE, A COLLECTION OF RULES BASED ON PROFILES AND PATTERNS
- THE RULES CONTAIN INFORMATION ABOUT BIOLOGICALLY MEANINGFUL RESIDUES LIKE ACTIVE SITES, SUBSTRATE OR COFACTOR BINDING SITES TO HELP IN FUNCTION DETERMINATION.

PROSITE



Database of protein domains, families and functional sites

PROSITE consists of documentation entries describing protein domains, families and functional sites as well as associated patterns and profiles to identify them [More... / References / Commercial users]

PROSITE is complemented by ProRule , a collection of rules based on profiles and patterns, which increases the discriminatory power of profiles and patterns by providing additional information about functionally and/or structurally critical amino acids [More...]

Release 2019_11 of 11-Dec-2019 contains 1846 documentation entries, 1311 patterns, 1265 profiles and 1289 ProRule.

Search

e.g. PDOC00022, PS50089, SH3, zinc finger

Browse

- by documentation entry
- by ProRule description
- by taxonomic scope
- by number of positive hits

Quick Scan mode of ScanProsite

Quickly find matches of your protein sequences to PROSITE signatures (max. 10 sequences). [?]

Enter UniProtKB accessions or identifiers or PDB identifiers or sequences in FASTA format

Other tools

- **PRATT** - allows to interactively generate conserved patterns from a series of unaligned proteins.
- **MyDomains - Image Creator** - allows to generate custom domain figures.

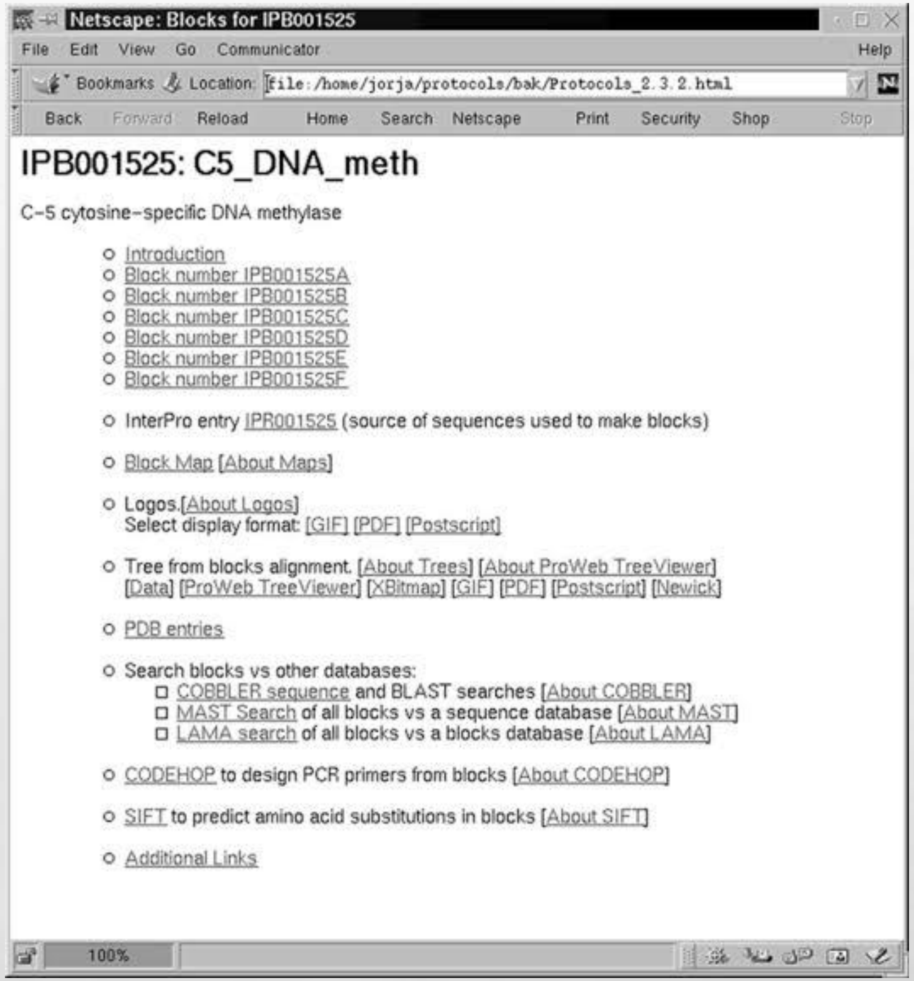
BLOCKS

- BLOCKS DATABASE IS BASED ON INTERPRO ENTRIES WITH SEQUENCES FROM SWISS-PROT AND TREMBL
- BLOCKS ARE MULTIPLE ALIGNED UNGAPPED SEGMENTS CORRESPONDING TO THE MOST HIGHLY CONSERVED REGIONS OF PROTEIN

BLOCKS

- BLOCKS CROSS REFERENCE TO PROSITE AND/OR PRINTS AND/OR SMART,AND/OR PFAM,AND/OR PRODOM DIARIES.
- BLOCKS DATABASE WAS CONSTRUCTED BY THE PROTOMAT SYSTEM USING THE MOTIF ALGORITHM.

BLOCKS



IPB001525: C5_DNA_meth

C-5 cytosine-specific DNA methylase

- Introduction
- Block number IPB001525A
- Block number IPB001525B
- Block number IPB001525C
- Block number IPB001525D
- Block number IPB001525E
- Block number IPB001525F

- InterPro entry IPR001525 (source of sequences used to make blocks)

- Block Map [About Maps]

- Logos.[About Logos]
  Select display format: [GIF] [PDF] [Postscript]

- Tree from blocks alignment. [About Trees] [About ProWeb TreeViewer]
  [Data] [ProWeb TreeViewer] [XBitmap] [GIF] [PDF] [Postscript] [Newick]

- PDB entries

- Search blocks vs other databases:
  - COBBLER sequence and BLAST searches [About COBBLER]
  - MAST Search of all blocks vs a sequence database [About MAST]
  - LAMA search of all blocks vs a blocks database [About LAMA]

- CODEHOP to design PCR primers from blocks [About CODEHOP]

- SIFT to predict amino acid substitutions in blocks [About SIFT]

- Additional Links

REFERENCES

- HTTPS://WWW.GENOME.JP/TOOLS/MOTIF/
- HTTPS://WWW.GOOGLE.COM/URL?SA=T&SOURCE=WEB&RCT=J&URL=HTTPS://PROSITE.EXPASY.ORG/&VED=2AHUKEWIK7OG4H-Z9AHWXS2WGHWPWCAWQFNOECAUQAQ&USG=AOVVAW2YSYV7B25SZFM7D4PADBFC

THANK YOU…!