

WEBLEM 8

(URL: <https://www.expasy.org/>)

Aim:

To explore the ExPasy Tools.

Introduction:

- **ExPasy:**

Expasy is the bioinformatics resource portal of the SIB Swiss Institute of Bioinformatics. It is an extensible and integrative portal which provides access to over 160 databases and software tools, developed by SIB Groups and supporting a range of life science and clinical research domains, from genomics, proteomics and structural biology, to evolution and phylogeny, systems biology and medical chemistry.

Thanks to a user-friendly search engine:

1. Expasy allows you to seamlessly query in parallel a subset of SIB databases through a single search, and to
2. Surface related information and knowledge from the complete set of >160 resources on the portal.

Expasy provides information that is automatically aligned with the most recent release of each resources, thereby ensuring up-to-date information.

1. String

STRING is the knowledgebase and software tool for known and predicted protein-protein interactions. It includes direct (physical) and indirect (functional) associations derived from various sources, such as genomic context, high-throughput experiments, (conserved) co-expression and the literature.

STRING, each protein-protein interaction is annotated with one or more 'scores'. Importantly, these scores do not indicate the strength or the specificity of the interaction. Instead, they are indicators of confidence, i.e. how likely STRING judges an interaction to be true, given the available evidence. All scores rank from 0 to 1, with 1 being the highest possible confidence. A score of 0.5 would indicate that roughly every second interaction might be erroneous (i.e., a false positive).

For most types of evidence, there are two types of scores: the 'normal' score, and the 'transferred' score. The latter is computed from data that is not originally observed in the organism of interest, but instead in some other organism and then transferred via homology/orthology. All potential source organisms are searched for evidence, but the actual transfers to the receiving organism are made non-redundant (according to 'clades' of closely related organisms in the tree of life).

2. UniProtKB/Swiss-Prot

Until 2002, the EBI/SIB Swiss-Prot + TrEMBL databases and the PIR Protein Sequence Database (PIR-PSD) coexisted as protein databases with differing protein sequence coverage and annotation priorities. In 2002, EBI, SIB, and PIR (at the Georgetown University Medical Center and National Biomedical Research Foundation) joined forces as the UniProt consortium. The primary mission of the consortium is to support biological research by maintaining a high quality database that serves as a stable, comprehensive, fully classified, richly and accurately annotated protein sequence knowledgebase, with extensive cross-references and querying interfaces freely accessible to the scientific community.

The UniProt Knowledgebase (UniProtKB) provides the central database of protein sequences with accurate, consistent, rich sequence and functional annotation. The UniProt Knowledgebase consists of two sections: Swiss-Prot - a section containing manually-annotated records with information extracted from literature and curator-evaluated computational analysis, and TrEMBL - a section with computationally analyzed records that await full manual annotation.

Swiss-Prot is an annotated protein sequence database. It was established in 1986 and maintained collaboratively, since 1987, by the group of Amos Bairoch first at the Department of Medical Biochemistry of the University of Geneva and now at the SIB Swiss Institute of Bioinformatics and the EMBL Data Library (now the EMBL Outstation - The European Bioinformatics Institute (EBI)). The Swiss-Prot Protein Knowledgebase consists of sequence entries. Sequence entries are composed of different line types, each with their own format. For standardization purposes the format of Swiss-Prot follows as closely as possible that of the EMBL Nucleotide Sequence Database.

3. EPD

This resource allows the access to several databases of experimentally validated promoters: EPD and EPDnew databases. They differ by the validation technique used and the coverage. EPD is a collection of eukaryotic promoters derived from published articles. Instead, the EPDnew databases (HT-EPD) are the result of merging EPD promoters with in-house analysis of promoter-specific high-throughput data for selected organisms only.

The Eukaryotic Promoter Database is an annotated non-redundant collection of eukaryotic POL II promoters, for which the transcription start site has been determined experimentally. Access to promoter sequences is provided by pointers to positions in nucleotide sequence entries. The annotation part of an entry includes description of the initiation site mapping data, cross-references to other databases, and bibliographic references. EPD is structured in a way that facilitates dynamic extraction of biologically meaningful promoter subsets for comparative sequence analysis. This database contains 4806 promoters from several species.

Methodology:

1. Open the homepage of ExPasy tools.
2. Explore the tools like UniProtKB/Swiss-Prot, String, EPD

Observation:

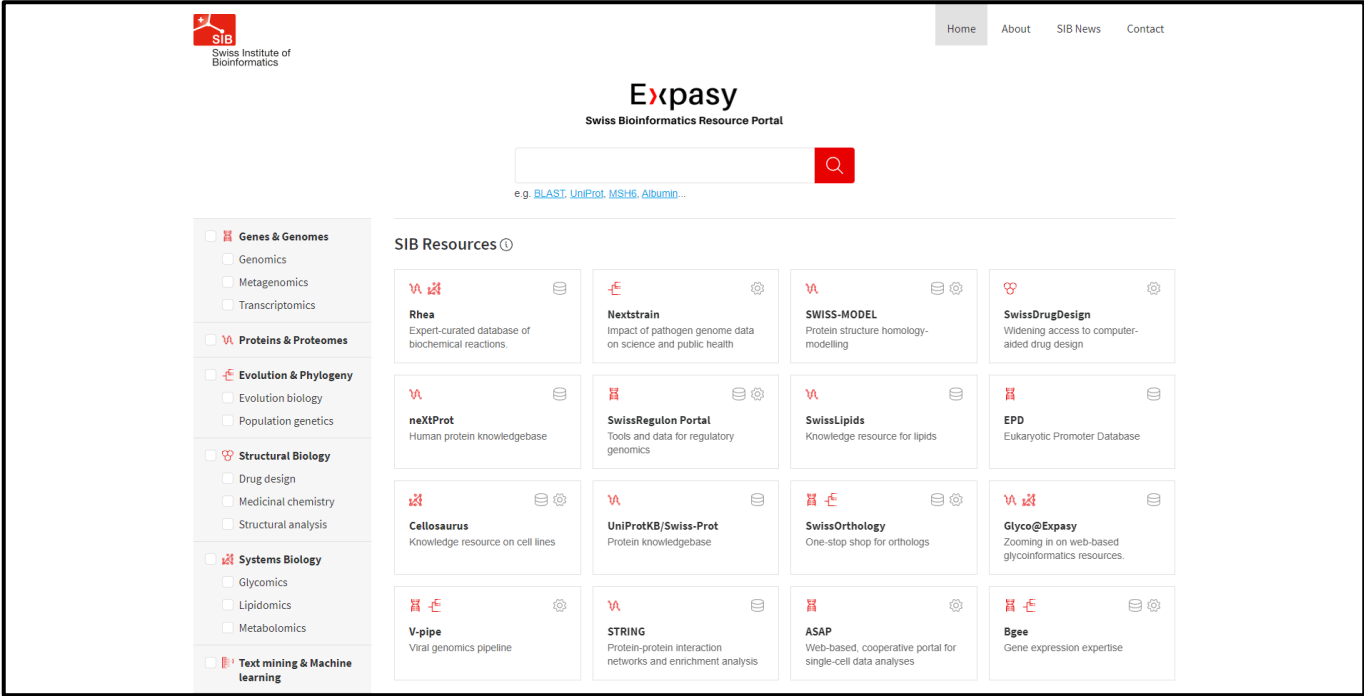


Fig1: Homepage of ExPasy Tools.

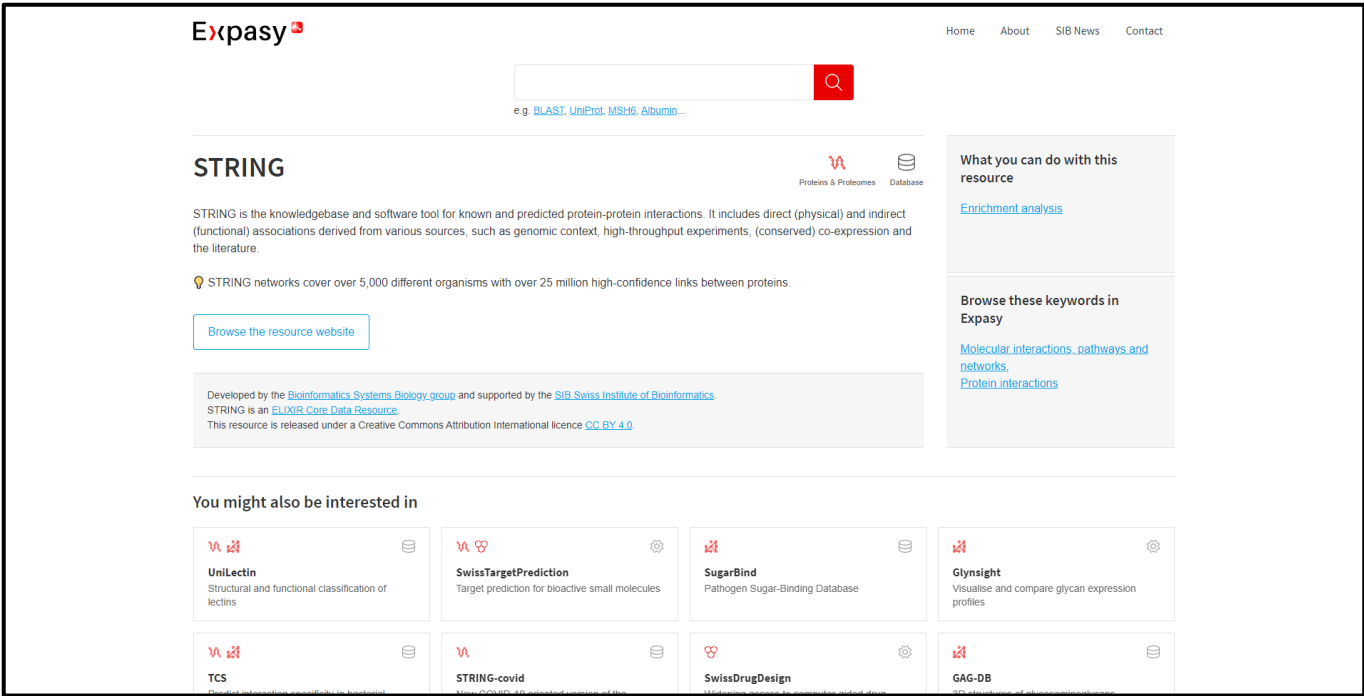


Fig2. Hit page of String

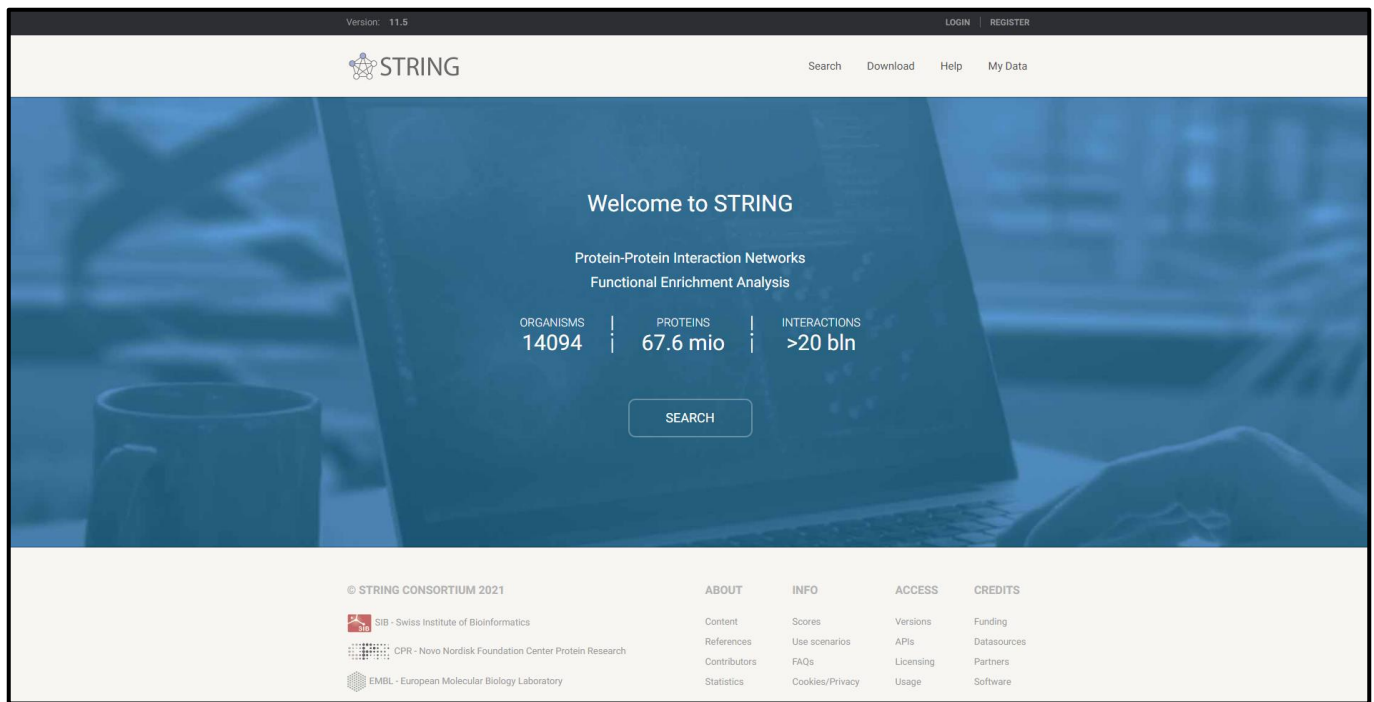


Fig3: Homepage for String

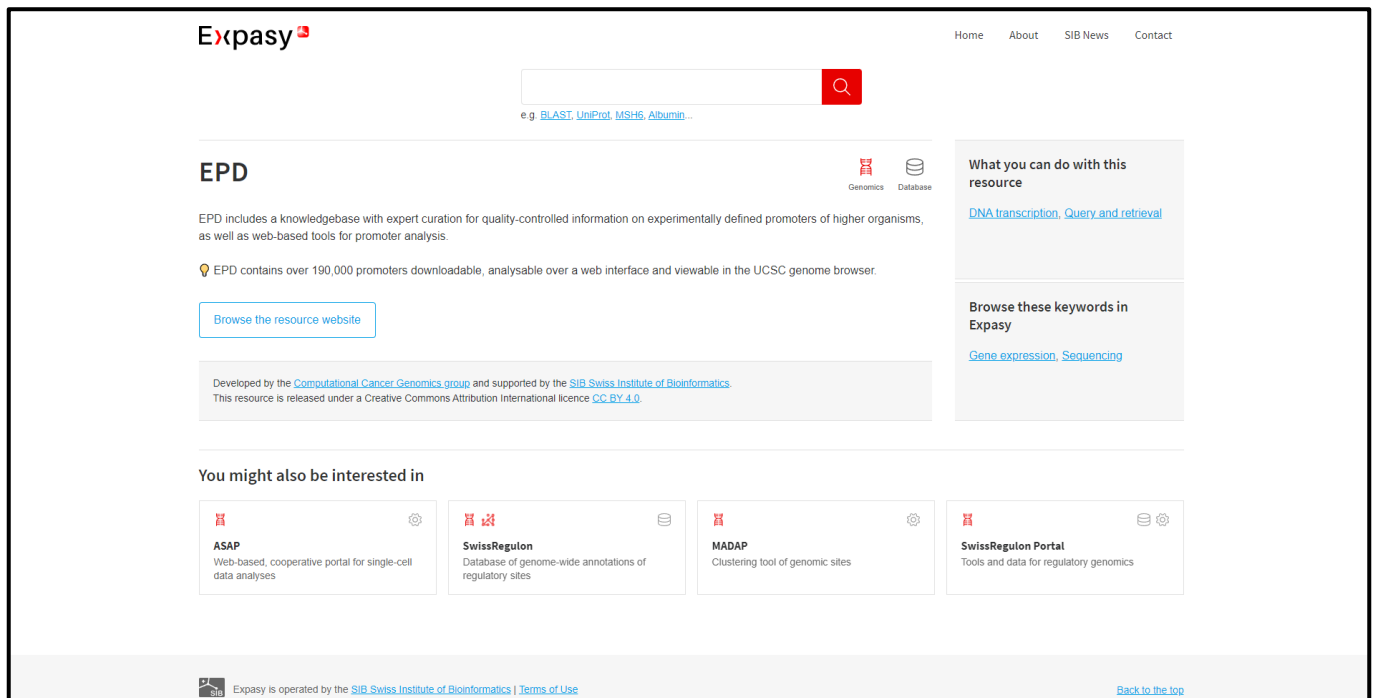




Fig4: Hit page for EPD

EUKARYOTIC PROMOTER DATABASE

Computational Cancer Genomics | ExPASy | EPFL

[Check out our short survey](#)
[Follow @EPD_SIB](#)

Access EPDnew

H. sapiens

H. sapiens non-coding

M. mulatta

M. musculus

M. musculus non-coding

R. norvegicus

C. familiaris

G. gallus

D. melanogaster

A. mellifera

D. rerio

C. elegans

A. thaliana

Z. mays

S. cerevisiae

S. pombe

P. falciparum

Select / Download

Promoter analysis tools

FTP site

News: 2019-04-26 -- First EPDnew release of non-coding RNA promoter collection for H. sapiens

This resource allows the access to several databases of experimentally validated promoters: EPD and EPDnew databases. They differ by the validation technique used and the coverage. EPD is a collection of eukaryotic promoters derived from published articles. Instead, the EPDnew databases (HT-EPD) are the result of merging EPD promoters with in-house analysis of promoter-specific high-throughput data for selected organisms only. This process gives EPDnew [high precision and high coverage](#).

EPDnew is a collection of databases of experimentally validated promoters for selected model organisms. Evidence comes from TSS-mapping from high-throughput experiments such as CAGE and Oligocapping. The resulting databases are the following:

- Animals:
 - Homo sapiens*: 29598 promoters,
 - Macaca mulatta*: 9575 promoters,
 - Mus musculus*: 25111 promoters,
 - Rattus norvegicus*: 12601 promoters,
 - Gallus gallus*: 6127 promoters,
 - Canis familiaris*: 7545 promoters,
 - Drosophila melanogaster*: 16972 promoters,
 - Apis mellifera*: 6493 promoters,
 - Danio rerio*: 10728 promoters,
 - Caenorhabditis elegans*: 7120 promoters;
- Plants:
 - Arabidopsis thaliana*: 22703 promoters;
 - Zea mays*: 17081 promoters;
- Fungi:
 - Saccharomyces cerevisiae*: 5117 promoters,
 - Schizosaccharomyces pombe*: 4802 promoters;
- Invertebrates:
 - Plasmodium falciparum*: 5597 promoters.

Access EPD

Promoter elements

Select / Download

FTP site

Access MGA Database

MGA-Search

MGA Data Overview

MGA FTP site

Documents

Other Resources

References


The **Eukaryotic Promoter Database** is an annotated non-redundant collection of eukaryotic POL II promoters, for which the transcription start site has been determined experimentally. Access to promoter sequences is provided by pointers to positions in nucleotide sequence entries. The annotation part of an entry includes description of the initiation site mapping data, cross-references to other databases, and bibliographic references. EPD is structured in a way that facilitates dynamic extraction of biologically meaningful promoter subsets for comparative sequence analysis. This database contains **4806 promoters from several species**.

Latest research from us that uses EPD:
[Influence of Rotational Nucleosome Positioning on Transcription Start Site Selection in Animal Promoters](#), Dreos, R., Ambrosini, G., Bucher, P. PLOS Comp Biol (2016) doi: 10.1371/journal.pcbi.1005144, PMID: [27716823](#)

Please cite us using the following references:

EPDnew database description:
[The eukaryotic promoter database in its 30th year: focus on non-vertebrate organisms](#), Dreos, R., Ambrosini, G., Groux, R., Périer, R., Bucher, P. Nucleic Acids Res. (2017). doi: 10.1093/nar/gkx1069; PMID: [27859657](#)

Promoters analysis tools:
[The Eukaryotic Promoter Database: expansion of EPDnew and new promoter analysis tools](#), Dreos, R., Ambrosini, G., Périer, R., Bucher, P. Nucleic Acids Res. (2014) doi: 10.1093/nar/gku1111; PMID: [25378343](#)





[Home](#)
[About](#)
[SIB News](#)
[Contact](#)


Q

e.g. BLAST UniProt MSHE Albumin...

UniProtKB/Swiss-Prot

 Proteins & Proteomes
 Database

UniProtKB/Swiss-Prot is the expertly curated component of UniProtKB (produced by the UniProt consortium). It contains hundreds of thousands of protein descriptions, including function, domain structure, subcellular location, post-translational modifications and functionally characterized variants.

 UniProt is one of the most widely used protein information resources in the world.

[▶ Watch the video](#)

[Browse the resource website](#)

Developed by the [Swiss-Prot group](#) and UniProt partners at EMBL-EBI and PIR, and supported by the [SIB Swiss Institute of Bioinformatics](#).
 UniProt is an [ELIXIR Core Data Resource](#).
 This resource is released under a Creative Commons Attribution International licence [CC BY 4.0](#).


What you can do with this resource

- [Query and retrieval](#)
- [Protein feature detection](#)


Browse these keywords in ExPasy

- [Protein sites, features and motifs](#)
- [Data submission, annotation and curation](#)
- [Proteomics](#)
- [Molecular interactions, pathways and networks](#)
- [Function analysis](#)
- [Protein structure analysis, Taxonomy](#)
- [Keyword, UniProt accession](#)
- [Protein sequence record, Text](#)


You might also be interested in




ViralZone
Fact sheets about viruses; linked to sequence databases.



OrthoDB
Orthology evolutionary and functional annotations



ImmunoDB
Insect immune-related genes and gene families



UniProtKB
Protein sequence database

Entry	Entry name	Protein names	Gene names	Organism	Length
B6J853	MIAB_COXB1	TRNA-2-methylthio-N(6)-dimethylallyl...	miaB CbuK_1268	Coxiella burnetii (strain CbuK_Q154) (Coxiella burnetii (strain Q154))	439
Q9D2V8	MFS10_MOUSE	Major facilitator superfamily domain...	Mfsd10 Tetran	Mus musculus (Mouse)	456
B5R2V6	MINC_SALEP	Probable septum site-determining pr...	minC SEN1223	Salmonella enteritidis PT4 (strain P125109)	235
Q2RJG3	MIAB_MOOTA	TRNA-2-methylthio-N(6)-dimethylallyl...	miaB Moth_1112	Moorella thermoacetica (strain ATCC 39073 / JCM 9320)	444
Q8VHK5	MLC1_MOUSE	Membrane protein MLC1	Mlc1	Mus musculus (Mouse)	382
Q5HRU5	METN1_STAEQ	Methionine import ATP-binding prote...	metN1 SERP0097	Staphylococcus epidermidis (strain ATCC 35984 / RP62A)	341
B6IVL6	METK_RHOCS	S-adenosylmethionine synthase	metK RC1_2972	Rhodospirillum centenum (strain ATCC 51521 / SW)	387
Q80T22	MFS4B_RAT	Sodium-dependent glucose transport...	Mfsd4b Naglt1	Rattus norvegicus (Rat)	484
C1CQE9	MIAA_STRZT	TRNA dimethylallyltransferase	miaA SPT_0695	Streptococcus pneumoniae (strain Taiwan19F-14)	294
Q62L22	METE_BURMA	5-methyltetrahydropteroylglutama...	metE BMA0467	Burkholderia mallei (strain ATCC 23344)	764
P0A820	METK_SHIFL	S-adenosylmethionine synthase	metK SF2933, S3137	Shigella flexneri	384
A7Z7Y9	METK_BACVZ	S-adenosylmethionine synthase	metK RBAM_027570	Bacillus velezensis (strain DSM 23117 / BGSC 10A6 / FZB42) (Bacillus amyloliquefaciens subsp. plantarum)	400
P58547	MER23_EUPRA	Mating pheromone Er-23	MAT23	Euplotes raikovi	51
A0A0B2WJN5	MLAC1_METAS	Laccase 1	lacc1 Abr2, MAM_08216	Metarhizium album (strain ARSEF 1941)	602
A8H6B7	METAS_SHEPA	Homoserine O-succinyltransferase	metAS Spea_2784	Shewanella pealeana (strain ATCC 700345 / ANG-SQ1)	314
B3PLH7	MLTF_CELJU	Membrane-bound lytic murein transgl...	mltF CJA_2597	Cellvibrio japonicus (strain Ueda107) (Pseudomonas fluorescens subsp. cellulosa)	495
B3QAC6	MIAB_RHOPT	TRNA-2-methylthio-N(6)-dimethylallyl...	miaB RpaL_0452	Rhodopseudomonas palustris (strain TIE-1)	463

Fig7: Homepage of UniProtKB

Results:

Expsay Tools:

It is an extensible and integrative portal which provides access to over 160 databases and software tools, developed by SIB Groups and supporting a range of life science and clinical research domains, from genomics, proteomics and structural biology, to evolution and phylogeny, systems biology and medical chemistry.

1. STRING:

The STRING database contains information from numerous sources, including experimental data, computational prediction methods and public text collections. It is freely accessible and it is regularly updated. The resource also serves to highlight functional enrichments in user- provided lists of proteins, using a number of functional classification systems such as GO, Pfam and KEGG. The latest version 11b contains information on about 24,5 million proteins from more than 5000 organisms.

2. EPD:

The Eukaryotic Promoter Database (EPD) is an annotated, non-redundant collection of eukaryotic Pol II promoters, for which the transcription start site has been determined experimentally. Access to promoter sequences is provided by pointers to positions in nucleotide sequence entries.

3. UniProtKB:

Uniprot is a freely accessible database of protein sequence and functional information, many entries being from genome sequencing projects. It contains a large amount of information about the biological function of proteins derived from the research literatures.

Conclusion:

Expasy is an online bioinformatics resource operated by the SIB Swiss Institute of Bioinformatics. The individual resources (databases, web-based and downloadable software tools) are hosted in a decentralised way by different groups of the SIB Swiss Institute of Bioinformatics and partner institutions.

References:

1. (n.d.). Retrieved from <https://epd.epfl.ch//index.php>
2. EPD. (n.d.). Retrieved from <https://www.expasy.org/resources/epd>
3. SIB Swiss Institute of Bioinformatics. (n.d.). Retrieved from <https://www.expasy.org/>
4. STRING. (n.d.). Retrieved from <https://www.expasy.org/resources/string>
5. UniProt ConsortiumEuropean Bioinformatics InstituteProtein Information ResourceSIB Swiss Institute of Bioinformatics. (n.d.). UniProt Consortium. Retrieved from <https://www.uniprot.org/uniprot/?query=reviewed:yes>
6. UniProtKB/Swiss-Prot. (n.d.). Retrieved from <https://www.expasy.org/resources/uniprotkb-swiss-prot>
7. Welcome to STRING. (n.d.). Retrieved from <https://string-db.org/>