**SelBioDB: A Comprehensive Database of Selenoproteins**

Debarup Sen and Vijay B. Baladhye

Bioinformatics Centre, Savitribai Phule Pune University, Pune-411007 (MS), INDIA.

**ABSTRACT**

Selenoproteins are a special class of proteins, in that they contain the non-standard amino acid selenocysteine. They are vital constituents of any organism’s proteome, with established roles for antioxidant defense and immune responses; including humans in whom it is important for hormone metabolism and muscle developmentas well. They have been known to have strong connection with important human diseases like cancer, diabetes, viral infections, Keshan disease and male infertility In spite of the known significance of selenoproteins, and their multifarious contribution in homeostasis, there has been inadequate experimental work relating to these special proteins. Similarly, or rather as a result of it, there is a dearth of databases that document selenoprotein-specific information.

Therefore, in the present study, we have designed a database of selenoproteins, that extracts data from UniProtKB and developed a tool for the prediction of SECIS element present in selenoprotein transcripts. We have extracted, formatted, added and uploaded data through a process of extensive automation. From the backend to the frontend of our database, including all intervening processing, has been fully automated. It is one of the unique features of our database. Hence, it will be a valuable resource for researchers involved in selenoproteomics or selenobiology and non-standard amino acids.

**Keywords:** Selenium, Selenocysteine, selenoprotein, SECIS element, SelBioDB, automation etc.

1. **INTRODUCTION**

Selenium, one of the essential micronutrients for humans, is of fundamental importance to human health. The predominant form of selenium in humans (as well as other domains of life) is selenocysteine, the 21st naturally occurring amino acid, which is the defining entity of selenoproteins. It is represented as ‘Sec’ in three letter code and as ‘U’ in single letter code. In the genetic code, the UGA codon serves not only as a termination codon, but also as the codon for incorporation of selenocysteines into proteins during translation. Selenocysteine is actually an analogue of the cysteine (Cys) residue which replaces the thiol (SOH) group with the selenol (SeOH) group. At physiological pH, this selenol group of Sec is mainly in its anionic selenolate form, while the thiol group of Cys residue is typically in protonated form and subsequently the pKa value of Sec is lower and nucleophilicity is higher as compare to Cys[1-3]. Due to this small change at atom level (selenium versus sulfur) makes Sec as much more reactive than Cys in selenoproteins. Thus, the presence of Sec residues in selenoproteins exhibit different characteristics compared to a Cys residue and give unique selenium-derived properties to selenoproteins.

As compared to other naturally occurring amino acids, Sec is one of the unique amino acid that have different mechanism of biosynthesis and incorporation into the proteins during translation. In eukaryotes, Sec is the only known amino acid whose biosynthesis occurs on its own tRNA and is designated as Sec tRNA[Ser]Sec. This tRNA[Ser]Sec first aminoacylated with serine in a reaction catalyzed by seryl-tRNA synthetase (SerRS) to form seryl-tRNA[Ser]Sec, which provides the backbone for Sec biosynthesis. Generally all canonical tRNAs have 7/5 secondary structures, in which the acceptor stem contains 7 bp and the T stem 5 bp. Whereas eukaryotic and archaeal tRNA[Ser]Sec have a 9/4 secondary structure and bacterial tRNA[Ser]Sec has a 8/5 cloverleaf form [4-6]. While mechanism of incorporation requires a specific translational incorporation machinery also called as selenoprotein synthesis machinery to insert Sec into the nascent polypeptide chain. This selenoprotein synthesis machinery requires some additional factors or components as compare to traditional canonical translational machinery. This includes a ribosome, selenoproteins transcript with Selenocysteine Insertion Sequence Element (SECIS Element), Sec-specific elongation factor, Sec-tRNASec, SECIS-binding protein 2 (SBP2), ribosomal protein L30, 43-kDa RNA-binding protein, soluble liver antigen protein and selenophosphate synthetase 1 (SPS1)[7-8].

Out of these components, SECIS elements is one of the key decider factors of selenoprotein synthesis. It is a unique nucleotide sequence on the 3’-untranslated regions (3′-UTR) of the selenoprotein transcript/mRNA that folds into very specific secondary structures which influence the downstream processes of Sec incorporation into the selenoproteins. In eukaryotic and archaea, these elements are cis-acting stem-loop RNA structures that are found in the 3′-UTR of selenoprotein mRNAs, although they have different sequences, motifs, and structures [9-10]. While in bacteria, these elements are of different from those in eukaryotes and archaea and they are located immediately downstream of the UGA codon, within the coding region of selenoprotein genes [11]. Several features distinguish SECIS elements from other functional mRNA stem-loop structures. Eukaryotic SECIS elements are formed by two helixes separated by an internal loop, a GA Quartet (SECIS core) structure and an apical loop or bulge. The GA Quartet is the main functional element of the SECIS and is required for interaction with SBP2. In some SECIS elements, the apical loop forms an additional ministem, which is used to classify SECIS elements into two different types. SECIS elements whose apical loop lacks the ministem are classified as type 1, and those containing the ministem (bulge) belong to type 2 SECIS elements [12-13]. In addition to the SECIS core, a conserved AAR motif in the apical region of SECIS is required for Sec incorporation [14]. In some selenoproteins, such as mammalian selenoproteins such as SelM and SelO contain SECIS elements with CC nucleotides instead of AA in this motif and involved in assisting with binding of eEFSec to the ribosome or accommodation of Sec-tRNA[Ser]Sec [15-16].

In brief, in eukaryotes, when a ribosome encounters the UAG codon then SECIS elements interacts with the standard translation machinery to increase the coding potential of UGA codons and prevent premature termination. In addition to SECIS elements, two more proteins such as SECIS binding protein 2 (SBP2) and Sec-specific translation elongation factor (eEFSec) are required for efficient recoding of UGA as Sec. This SECIS element-SBP2-eEFSec complex also bind to ribosome which recruits Sec-tRNA[Ser]Sec and facilitates incorporation of Sec into the nascent growing polypeptide chain [17] While in bacteria, instead of SBP2 and eEFSec, a Sec-specific translation elongation factor (SelB in *E. coli*) that directly recognizes SECIS and is required for binding and delivery of SECIS elements to the ribosome [18]. In such a way that varieties of selenoproteins or selenoenzymes are synthesized through selenoprotein translational machinery. There are one or more Sec containing selenoproteins have been identified among three domains of life. But there are some exceptions like yeast and higher plants do not have selenoproteins, instead, they express cysteine-containing homologues [19-20]. The function of all these selenoproteins depends on the number and position of Sec in the polypeptide chain(s). In mammalians, there are two groups of selenoproteins such as type-1 and type-2. Type-1 group of selenoproteins possesses Sec in a site very close to the C terminus of protein, such as TrxRs and selenoproteins SelS, SelR, SelO, SelI, and SelK. While type-2 group possesses Sec in a site close to the N terminus such as GPxs; DIOs; selenoproteins SelH, SelM, SelN, SelT, SelV, and SelW; SPS2; and Sep15 [21]. In human, all these types are involved in regulating cellular oxidative stress, ER stress, antioxidant defense, immune response, inflammatory response and other biological processes including antioxidant, anti-inflammation, anti-apoptosis, and regulating immune response and other biological functions [22-25]. They have been known to have strong connection with important human diseases like cancer, diabetes, viral infections, Keshan disease and male infertility[26-27]. In spite of the known significance of selenoproteins, and their multifarious contribution in homeostasis, there has been inadequate experimental work relating to these special proteins. Similarly, or rather as a result of it, there is a dearth of information regarding the biology of selenoproteins and roles of Se and selenoproteins in human health. Therefore, in the present study we have focused on the collection of all types of information about selenoproteins found in all three domains of life. In addition, some mechanisms of Se and selenoproteins in the regulation of diseases have not been clarified. Therefore, future research should focus on the specific mechanism of Se’s participation in regulating diseases after entering the body.

Currently, there are only two selenoprotein databases are existed—the most celebrated database is SelenoDB[28] and a recently developed Selenoprotein Database. Though these two databases seem to be general selenoprotein databases from nomenclature, they are, in reality, restricted in terms of the extent of information they provide. Both these databases are reserved for only eukaryotes; in case of Selenoprotein Database[ii], the data provided is almost wholly predicted data (at the time of writing this manuscript). Simultaneously, it is intriguing to notice that the general protein databases, for example UniProt[7 29], are increasingly depositing new and updated information along with reliable and high-quality annotation. This creates a pool of unused potential that is, in reality, waiting to be extracted, curated, and presented in a more focused fashion, that would be contextually relevant for selenoproteins. Hence, a database is needed, that will do the above-mentioned steps and give rise to a more verbose and extensive knowledge platform to serve the scientific community. Therein lies the need of SelBioDB and the promise it brings to the table. In addition to the existing annotation already derived from UniProtKB, other features usually not found in most databases, namely physicochemical properties, QSAR descriptors, CTD descriptors and other descriptors have been provided, as a value addition to SelBioDB entries. SelBioDB also features sequence alignment (database search using BLASTp and multiple sequence alignment) and descriptor calculation tools, to aid researchers in consumption of the available data.

**2. MATERIALS AND METHODS**

**2.1: Data retrieval and processing**

The data required for the development of SelBioDB is extracted from UniProtKB database. UniProtKB is queried for all proteins which are annotated for the non-standard amino acid selenocysteine. A list of selenocysteine containing proteins (i.e. selenoproteins) with its information is obtained. All relevant columns are selected from the list and the API query is triggered to fetch that selected information. This process is automated using Python requests module. The fetched data, which is in tab separated format, is directly written into a file (“main dataset”), before further processing is done. This data thus obtained is not suitable for direct upload to a backend database, since it is not in proper formatting. Hence, punctuations have to be corrected and empty spaces filled with Null values, for better formatting of data which would ultimately smoothen the upload process. This processed data was nomenclature by the addition of serial numbers to each entry of selenoproteins in the processed data. That serial number in this database is henceforth called SelBioDB accession IDs (SBDB\_IDs) or accession number that required to enhance sorting and recognition of entries. All this is done using Python and the columns are processed and added using pandas library.

**2.2: Descriptor and Properties Calculations:**

Simply extracting data from a major database and uploading it as one dataset, is neither enough, nor is it desirable. Therefore, we have added a wealth of information to each entry in order to make our database more relevant and informative, in the form of descriptors. Basic physicochemical properties of selenoproteins, for example, aliphatic index, instability index, hydropathicity, etc. are important bits of information that required for wet and dry lab experiments. Since such information is not directly available in UniProtKB entries, we use Python libraries for calculation of these parameters. Python libraries – ProtParam and peptides – where used to calculate multiple important features of the proteins.

The ProtParam module of the Biopython package (version 1.79) is a widely known and used program amidst the bioinformatics community for calculating metrics of protein sequences. It was used to calculate certain physicochemical properties like amino acid percentage information, total number of positively and negatively charged residues, extinction coefficient, aliphatic index, instability index, and grand average of hydropathicity.

Further calculations were performed using the proteinAnalysis (version 1) library, which is another useful library for protein basic composition feature calculation; it was implemented to compute basic information like most occurring residue, least occurring residue, percentage of non-standard residues, etc. All this descriptor data was compiled into a separate file (“descriptor dataset”).

**2.3 Classification of Selenoproteins:**

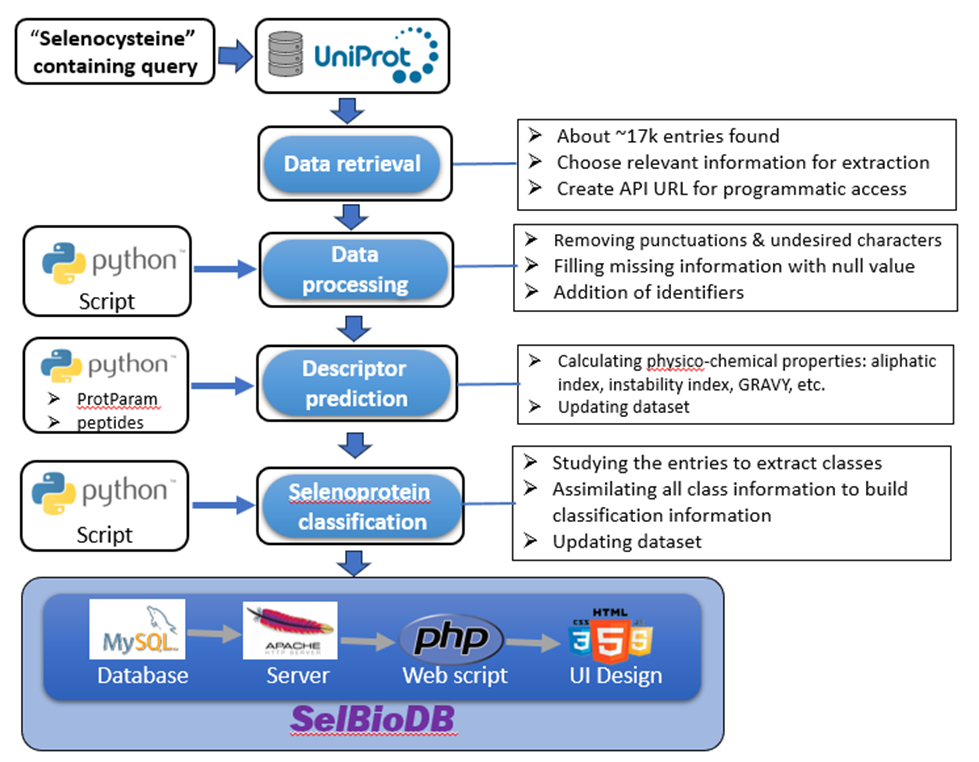
The above extracted data from UniProtKB was in the form of dump and it needs to be classified into various groups. The Data is extracted from UniProtKB as a dump. To improve its relevance, we have classified it into 113 classes based on function, enzymatic activity, cellular location, and other factors. The 113 separate datasets were generated for the different classes to facilitate searching within each class, as well as performing sequence alignment operations against single or multiple datasets. These datasets and a master dataset, amounting to a total of 114 datasets, were uploaded to the backend using ??.

**2.4 Database design and implementation**

The resulting datasets thus obtained need to be uploaded to a MySQL database in order to be presented to the frontend. For the same purpose, a Python script is written, which creates a connection with MySQL database on an APACHE HTTPS server using the Python library mysql.connector, checks if the database exists, and creates tables for upload of data. All the lines of data are uploaded to the MySQL database using SQL queries in a loop.

**2.5 Workflow**

The complete workflow implemented in current study is represented in Fig. 1.



**2.6 Automation:**

**3. RESULTS AND DISCUSSION**

**3.1 Development of Web User Interface**

The above classified datasets with a master dataset were first uploaded to the backend using ..... for the development of a comprehensive database of selenoproteins. There was a total of 114 different datasets including a master dataset for the different classes to facilitate searching within each class, as well as performing sequence alignment operations against single or multiple datasets. Simultaneously the frontend of SelBioDB database was created using HTML/CSS and is made interactive using JavaScript to make it simple yet user-friendly, with as minimal sophistication as possible. The above uploaded datasets were then uploaded to a MySQL database in order to be presented to the frontend of this database. All the lines of data are uploaded to the MySQL database using SQL queries in a loop and the home page of SelBioDB was ready to use. A screenshot with all the details is shown in Fig. 1.

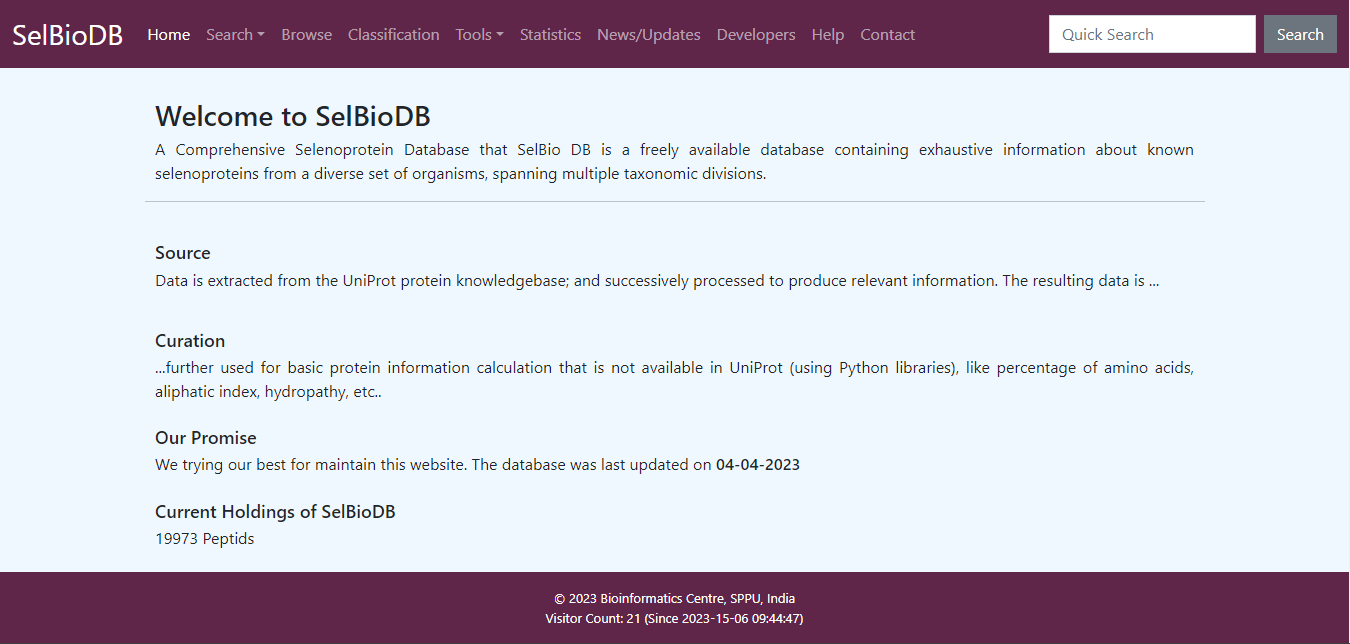


Fig 1: Home Page of SelBioDB.

The above created SelBioDB home page includes the information in the forms of various options such as home option about the information about the SelBioDB, search option is for the finding query about the selenoproteins or Sec congaing proteins, browse option to browse the information about the selenoproteins, classification button is for display the classes of selenoproteins, tools options is given to put required tools on the database, database statistics information is put under the statistics option, the information about the addition and deletion of data into the database is kept under the news or updates option, whereas information about the database developers as well as contact details of developers you will find under the heading contacts and help option contains the some solved tutorials about how to use this database.

**3.2 Model query for basic search**

The basic search is developed on this database in order to enable user to obtain information on a specific selenoprotein. Thus, basic search can be conducted using selenoprotein name or its abbreviation code. Upon successful execution of the query, this text search page accepts any character as text and queries it against the database to find relevant results. The resulting page tabulates the results in five major sections like ‘SBDB number’, ‘Protein name’, ‘Length’, ‘Gene names’ and ‘Source organism’. In addition to that this search option enables the user to restrict their search to desired dataset by clicking or selecting an option ‘Choose a dataset’. So that the submitted query searching inside the selected dataset and returning matching results containing the query string. The result for query made for SeP is illustrated in Fig. 4. This search option also enables user to reset the basic search option. Additional help option is provided with basic search and contains some solved tutorials about how to use this option. The resulting page tabulates with results containing number of records or entries that contains “SeP” text anywhere in their record in a selected dataset.

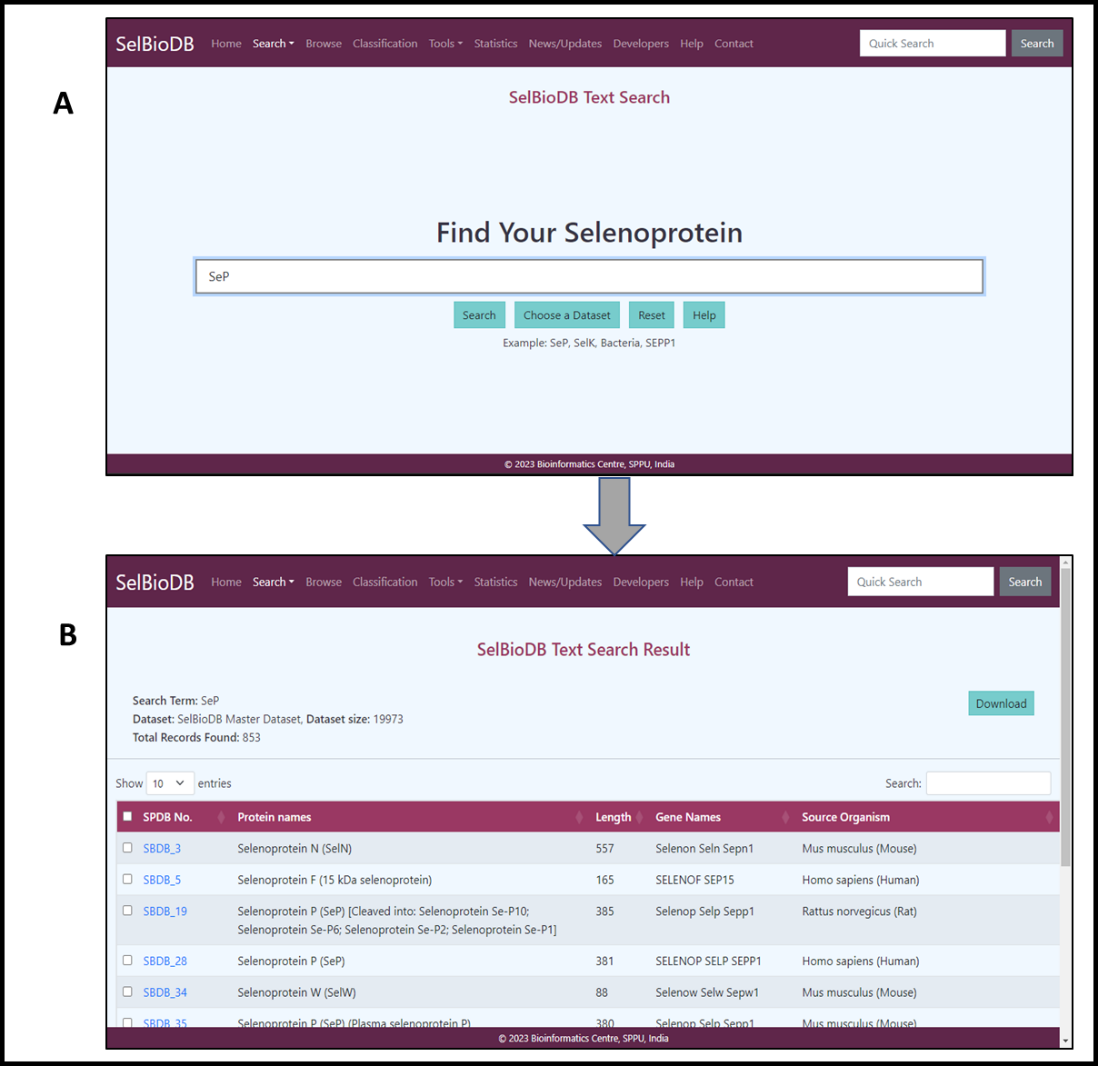


Fig. 4. Performing Basic Search on SelBioDB. (A) Illustration demonstrating the basic text search for finding a selenoprotein. (B) The result page of basic text search tabulating multiple hits that containing query text.

**3.3 Model query for advanced search**

This is a comprehensive search feature that brings four multiple types of searches in a single platform and enables the user to club multiple parameters and build complex queries to search the database. Various logical operators (e.g. AND, OR and NOT) are implemented for constructing the query with complex logic to obtain user defined results. This advanced search utility contains multiple sections to obtain desired information from the database. Following attributes enables the user to construct and execute complex queries.

* **Non-range attributes:** This section allows the user to select multiple nominal attributes, for example, protein name, gene name, organism name, etc. and club them together with AND, OR and NOT operators to build highly specific query.
* **ID attributes:** This section can be used to build query using ID attributes, for example, SBDB accession ID, UniProt accession ID, ChEMBL ID, and lots more.
* **Range attributes:** This section contains all continuous and numerical attributes, for example, sequence length, hydrophobicity, instability index, etc. to build advanced queries.
* **Choose Class:** This option enables the user to select the class(es) of proteins that he/she wants to search for. There are multiple classes like, taxonomic class, biological activity, cellular location, prosthetic group, etc., which further contain the sub-classes. For example, taxonomic class contains the sub-classes, animal, plant, fungi, etc. Biological activity contains the sub-classes, antioxidant, antimicrobial, RNA-binding, etc. Cellular location contains the sub-classes cell membrane, microsome, endosome, etc.

All these search functions come together to provide an immensely powerful platform for the users to aptly build queries. This advanced search functionality provides multiple parameters to specify and narrow down the search, to be able to look for selenoproteins that match their requirements.

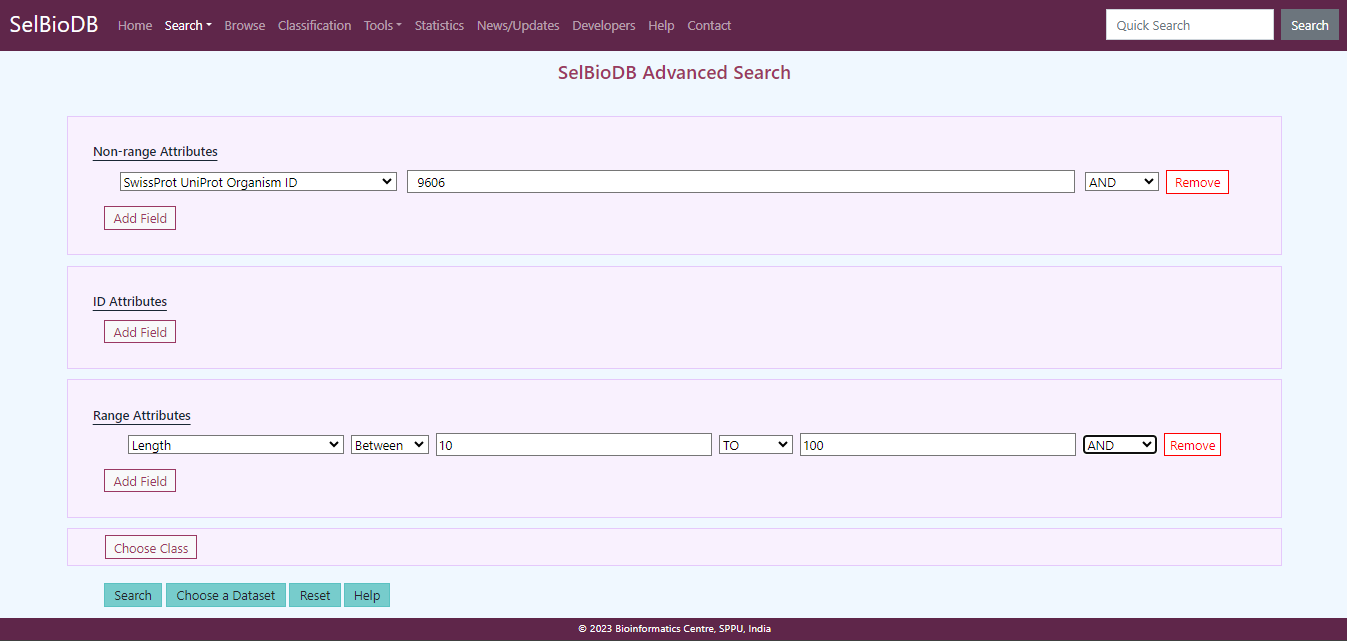


Fig 6: Advanced search functionality

For example, if we have selected ‘SwissProt UniProt Organism ID from the ‘Non-range Attributes’ dropdown menu and entered ‘9606’. As well as we have selected ‘AND” operator and added another filter ‘Length’ from the ‘Range Attributes’, searching for all selenoproteins that have sequence length between 10 and 100. In addition to that there is an option to ‘Choose Class’, which allows the user to choose a class from amongst classes and subclasses, to narrow down the search further. The entries are broadly categorized into the following classes such as Taxonomic class, biological activity, biological process, cellular location, enzymatic activity, prosthetic group, protein type and relation with disease. A screenshot with all the classes and subclasses is shown in Fig.7.

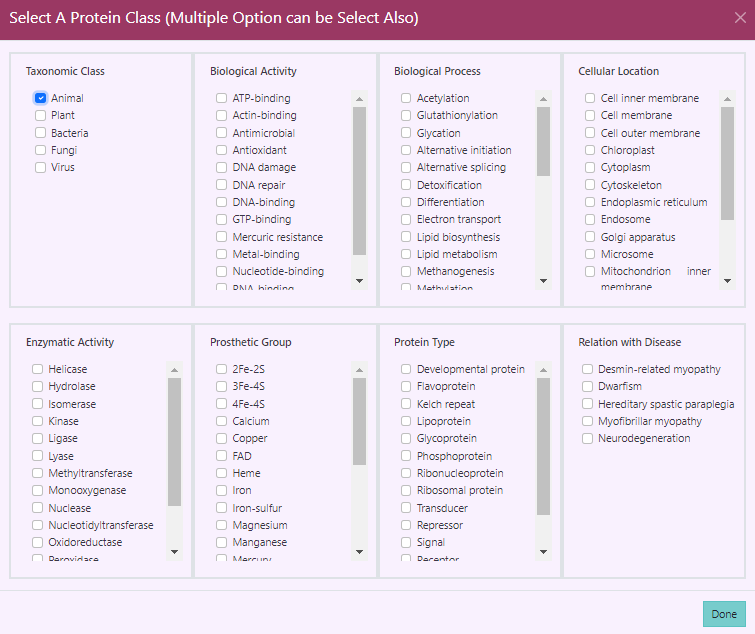


Fig 7: Classes available in Advanced search

Further if we have chosen the taxonomic classes ‘Animal’, since taxonomic ID ‘9606’ is for Human, which falls in taxonomic class ‘Animal’. We proceed to choose a dataset from the 114 available datasets. This option is for search restriction by choosing a specific dataset, user can restrict the search in only that dataset. Currently there are 114 datasets are available. We keep the default dataset, and proceed for search. Intermediate page tabulating all the names of datasets as shown in the Fig.8. Finally, this advanced search option narrows down the query that match the criteria set for the search and give matched results in the form of result page shown in the Fis 10.

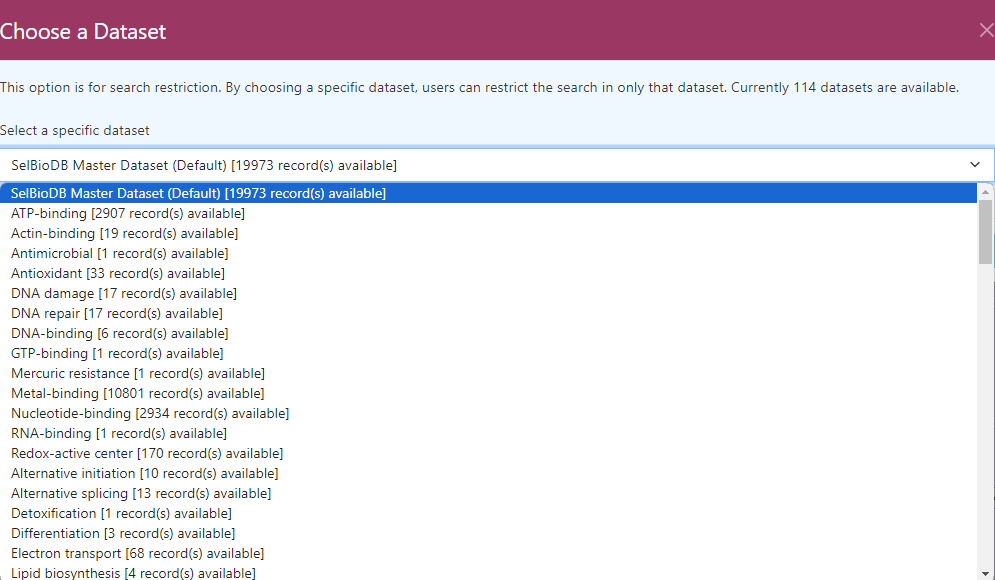


Fig 8: Datasets available in Advanced search

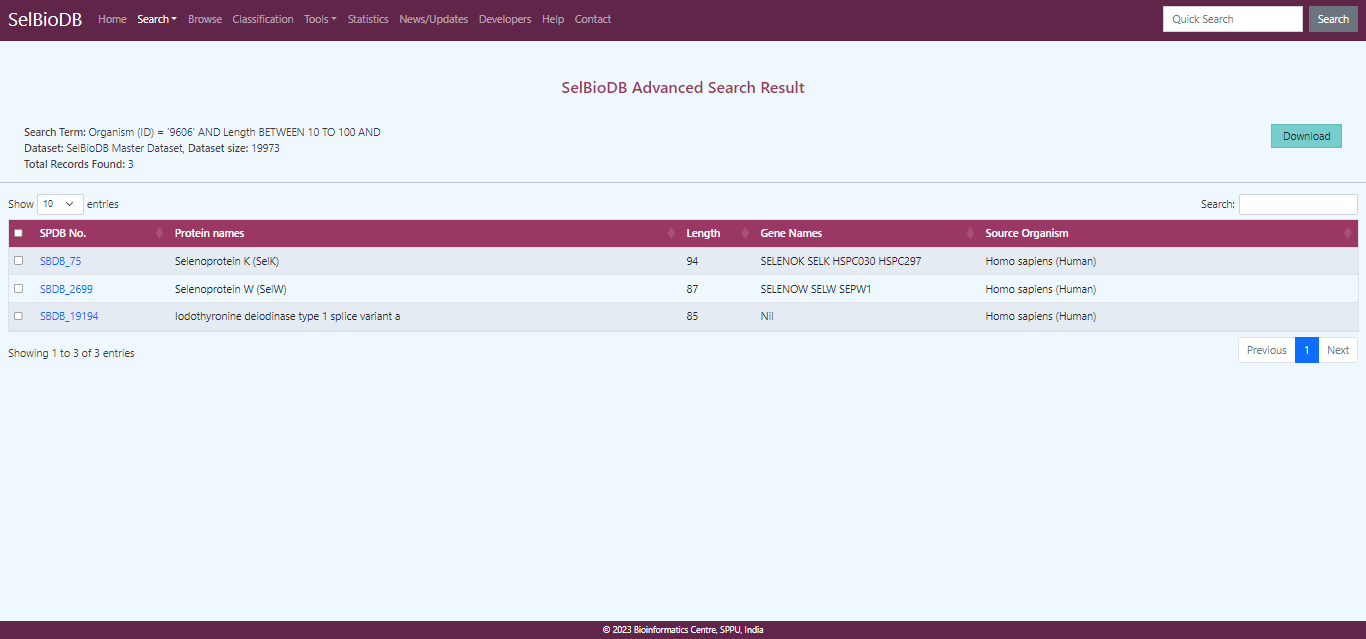


Fig. 9: The Advanced search results

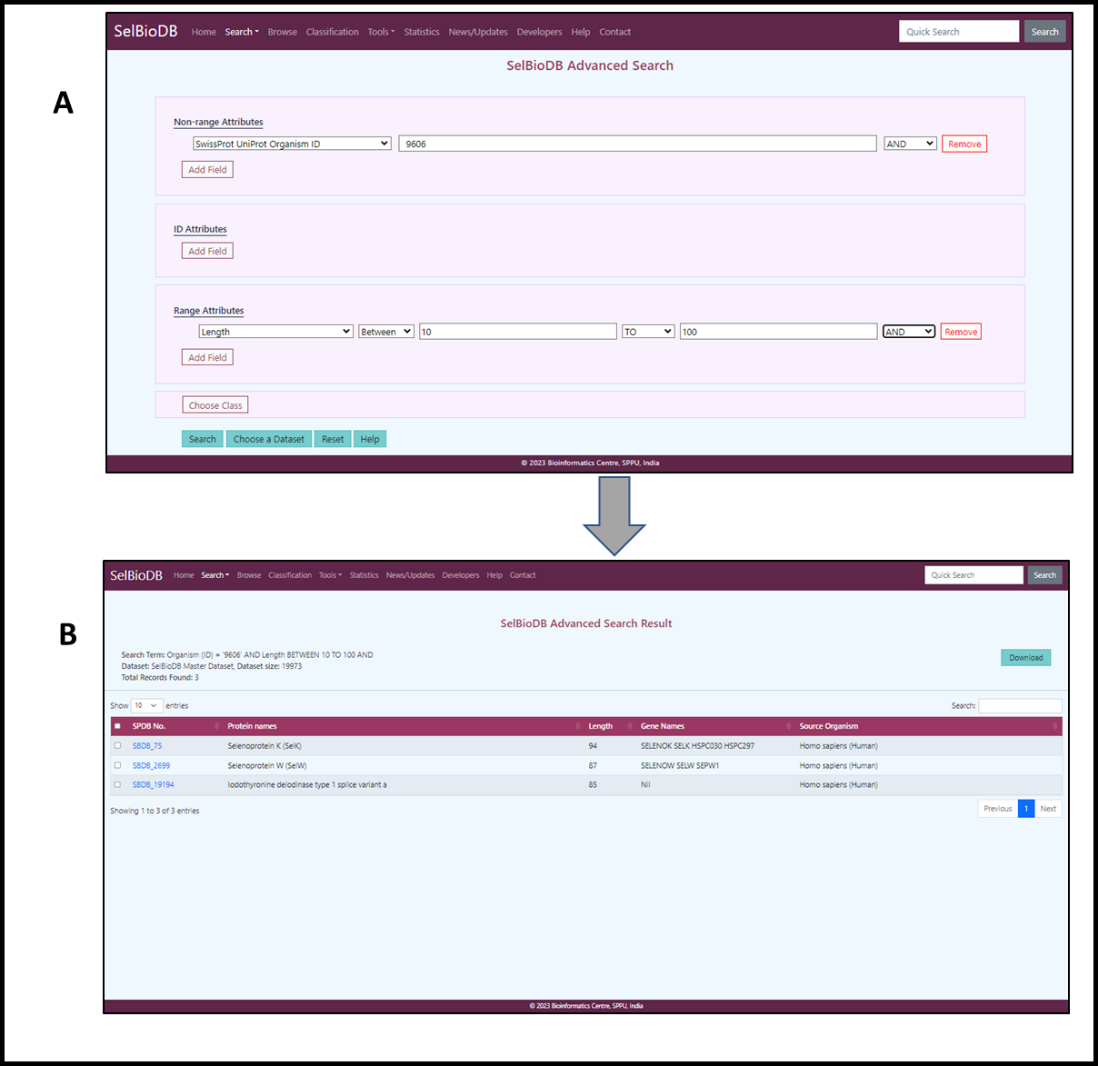


Fig. 7. Performing Advanced Search on SelBioDB. (A) Illustration demonstrating the advanced search for finding a selenoprotein. (B) The result page of advanced search tabulating multiple hits that containing query text.

In addition to that we have created one more facility with advanced search option to download the search results in multiple formats, including FASTA, Text, TSV, JSON, and as list of accession numbers. This option also creates a custom report that user can selectively download the data which is relevant to their purpose.

**3.4 Model query for Browsing the database**

The browsing option is developed on this database in order to enable user to enlist the information about the total number of selenoproteins. User can search selenoproteins according to their classes, subclasses, datasets, ect. The resulting browse page shows all the entries currently present in the dataset. Where the users can get a bird’s eye view of the data and an idea of what all information they can expect from the database.

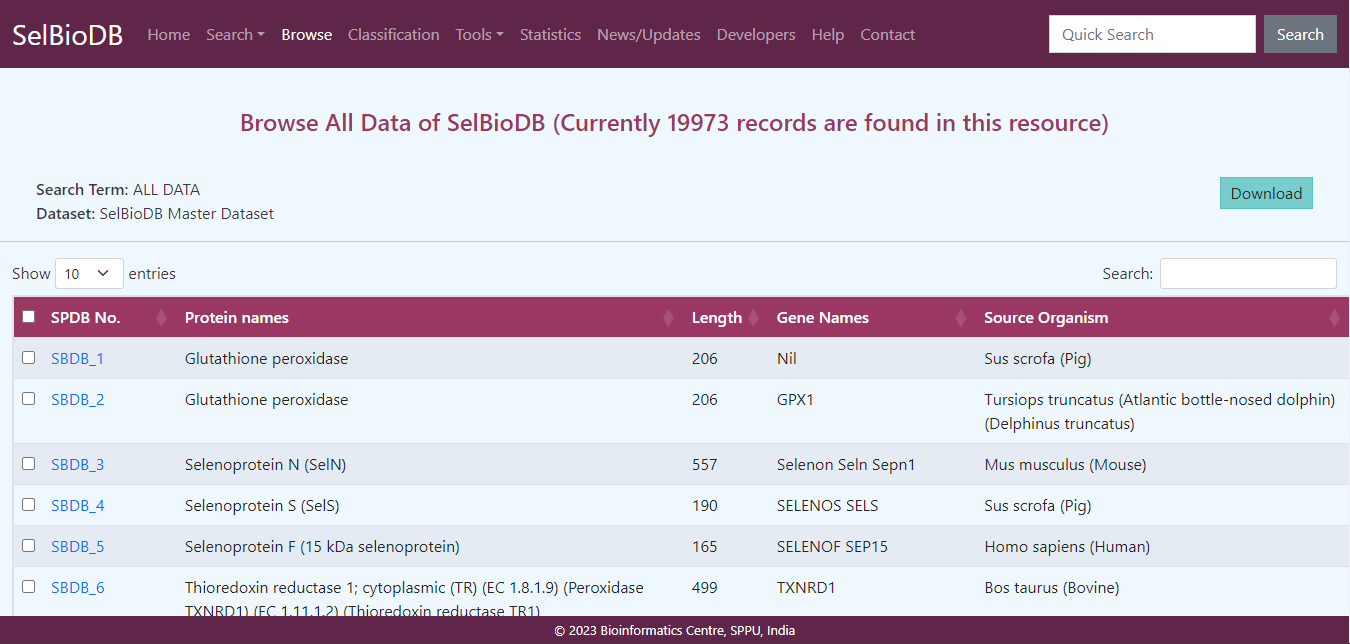


Fig. 2: Browsing the database

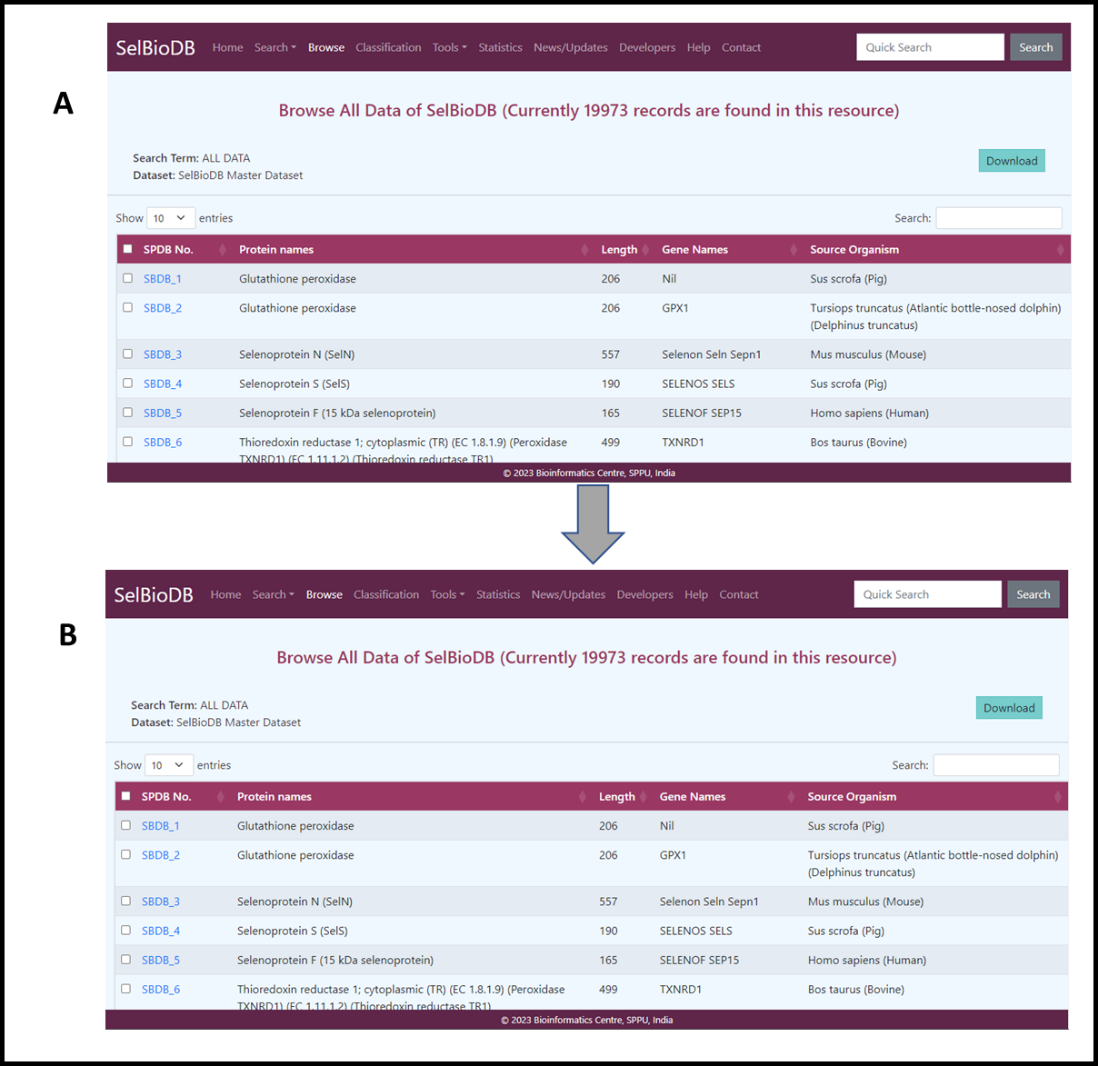


Fig. 9. Performing query for Browsing the database. (A) Illustration demonstrating the browsing for finding a selenoprotein. (B) The browsing result page tabulating total number of selenoprotein related data.

This browsing option also enables users to browse the data on the basis of their classes. The master dataset of this database contains 113 different classes of selenoproteins based on multiple factors, as mentioned before. Thisoption enables users to browse the classification list and directly land up on the entries of the class they are interested in. The resulting browsing page display all 113 classes of data in the following screenshot of browse option.shown in Fig. 10.

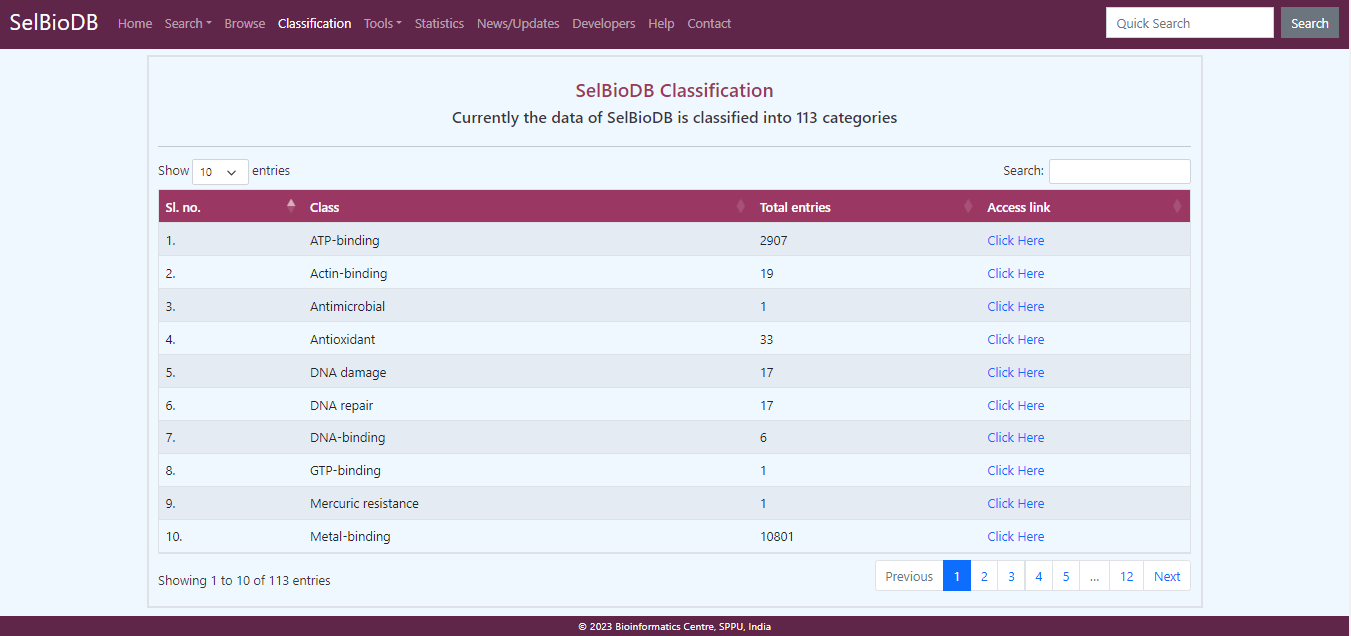


Fig. 10: The Classification page

**3.6 Database Tools**

SelBioDB comes with two toolboxes such as Sequence Alignment toolbox and the Protein Feature Calculation toolbox. These toolboxes aim to enrich the user experience by providing with platforms to perform initial analysis, by the means of sequence similarity searches and alignments and protein descriptor calculation.

**3.6.1 Sequence Alignment Toolbox:**

**3.6.1.1 BLASTp:** It is one of the most reliable and accurate sequence alignment tools that searches the protein databases using an amino acid sequence as the query. This tool have been pre-selected in the toolbox as the first and default option. It comes with a wide range of tuneable parameters, so that when performing a database using BLASTp, the user can customise the search as much as possible, to suit their needs. In addition to that there is provision to choose from over 10 types of formats ranging from the familiar Pairwise, Tabular and Flat query-anchored formats, to the machine-friendly XML, JSON and Binary ASN formats. The user also has the choice to set desired e-value, word size, gap penalties, query coverage, as well as what type of matrix to use – BLOSUM, PAM or Identity.

**3.6.1.2 BLASTx:** Another BLAST implementation, which uses nucleotide sequence as input, translates it via six open reading frames (3 on same strand and 3 on the complementary strand) to produce six peptides, and uses the peptides to query the protein database. This tool is particularly useful when only coding sequence of a gene is available, and the aim is to find out proteins similar to the product of the CDS. This tool provides as many options to the user as the previous tool, making it highly malleable; the parameters can be altered to suit the users’ needs.

**3.6.1.3 MUSCLE:** In case a user wants to perform Multiple Sequence Alignment, that too is possible using the sequence alignment toolbox. MUSCLE has been integrated for the same purpose, and this tool also provides as many customisable parameters as possible, including the multiple formats, number of iterations, grouping sequences by similarity etc.

**3.6.2 Protein Feature Calculation toolbox:**

**3.6.2.1 Composition features:** These are basic features of the protein, which include residue counts and frequencies in a protein. For any input protein, two tables are returned – one that shows the counts and frequencies of all 25 residues (20 standard, 1 non-standard and 4 ambiguous residues), and the other shows the same information for only the residues present in the input sequence. This gives a side-by-side comparison of the content of the input sequence in terms of residues that are present and the quantity in which they are present.

Some other basic information like most common residues, least common residue, number hydrophilic residues, hydrophobic residues, acidic residues, and basic residues, as well as residues that are not present in the sequence and the secondary structure fraction.

**3.6.2.2 Physicochemical features:** These features include indices like aliphatic index, instability index, hydrophobicity, hydrophobic moment, isoelectric point, charge (at pH 7), aromaticity, flexibility, etc. The peptides library of Python is implemented for these calculations.

**3.6.2.3 CTD Descriptors:**  The toolbox gives the option to calculate composition, transition and distribution descriptors as well, since these descriptors are often used for prediction and analysis purposes. The propy2 Python library is used to calculate these descriptors. We are continuously striving to provide the best functionalities to the user. Therefore, we will definitely be adding more tools with time.

**3.7 Statistics:**

This page displays a graphical representation of all the entries present in SelBioDB in the form of bar graphs. There are 8 bar graphs, showing data distribution with respect to taxonomy, biological activity, biological process, cellular location, enzymatic activity, prosthetic group, protein type and disease relation. These graphs are all equipped with re-scalable X- and Y- axes; the user can rescale the axes, by pressing the buttons provided alongside the graphs, to obtain a improve the context against which the graphs are plotted. Also worthwhile to note here, is that all the graphs update in real-time, so there is never any discrepancy between the data distribution displayed by the graphs and the actual distribution in the database.

**3.8 Other Database Features**

**3.8.1 News/Updates:**

This section is dedicated to all the latest news and updates regarding SelBioDB. Version updates, additions, and all other information will be displayed in this section as and when the changes take place.

**3.8.2 Developers:**

This page provides information about the developers of the database

**3.8.3 Contact:**

This is the contact page of the development team, for feedbacks, suggestions, and comments from the users.

**4. CONCLUSION**

SelBioDB is one of its kind in leveraging the already existing information about selenoproteins, which is both copious in amount and excellent in terms of reliability. On top of that, SelBio DB provides information, both calculated and predicted, from ProtParam as well as an in-house developed tool, the SECPred, a SECIS Element prediction tool. Along with that, SelBioDB brings the advantage of end-to-end automation, ensuring regular updates, minimal downtime for updates and little to no dependency on manual maintenance; therefore, promising a foolproof and seamless performance. Hence, SelBioDB will serve as a valuable database and add to the wealth of scientific knowledge about selenoproteins.

**5. REFERENCES**

1. R.E. Huber, R.S. Criddle, Comparison of the chemical properties of selenocysteine and selenocystine with their sulfur analogs, Arch.Biochem. Biophys. 122 (1967) 164 – 173.
2. Linda Johansson, Guro Gafvelin, Elias S.J. Arnér, Selenocysteine in proteins—properties and biotechnological use, Biochimica et Biophysica Acta (BBA) - General Subjects, Volume 1726, Issue 1, 2005, Pages 1-13.
3. Arner ES. Selenoproteins: what unique properties can arise with selenocysteine in place of cysteine? Exp Cell Res 316: 1296–1303, 2010.
4. Hubert N, Sturchler C, Westhof E, Carbon P, Krol A. The 9/4 secondary structure of eukaryotic selenocysteine tRNA: more pieces of evidence. RNA 4: 1029–1033, 1998.
5. Itoh Y, Chiba S, Sekine S, Yokoyama S. Crystal structure of human selenocysteine tRNA. Nucleic Acids Res 37: 6259–6268, 2009.
6. Itoh Y, Sekine S, Suetsugu S, Yokoyama S. Tertiary structure of bacterial selenocysteine tRNA. Nucleic Acids Res 41: 6729–6738, 2013.
7. Xu XM, Mix H, Carlson BA, Grabowski PJ, Gladyshev VN, Berry MJ, Hatfield DL. Evidence for direct roles of two additional factors, SECp43 and soluble liver antigen, in the selenoprotein synthesis machinery. J Biol Chem 280: 41568–41575, 2005.
8. Labunskyy VM, Hatfield DL, Gladyshev VN. Selenoproteins: molecular pathways and physiological roles. Physiol Rev. 2014 Jul;94(3):739-77.
9. Low SC, Berry MJ. Knowing when not to stop: selenocysteine incorporation in eukaryotes. Trends Biochem Sci 21: 203–208, 1996
10. Rother M, Resch A, Gardner WL, Whitman WB, Böck A. Heterologous expression of archaeal selenoprotein genes directed by the SECIS element located in the 3′ non-translated region. Mol Microbiol 40: 900–908, 2001.
11. (Böck A, Rother M, Leibundgut M, Ban N. Selenium metabolism in prokaryotes. In: Selenium: Its Molecular Biology and Role in Human Health, edited by Hatfield DL, Berry MJ, Gladyshev VN. New York: Springer, 2006, p. 9–28.
12. (Fagegaltier D, Lescure A, Walczak R, Carbon P, Krol A. Structural analysis of new local features in SECIS RNA hairpins. Nucleic Acids Res 28: 2679–2689, 2000.
13. Grundner-Culemann E, Martin GW, 3rd, Harney JW, Berry MJ. Two distinct SECIS structures capable of directing selenocysteine incorporation in eukaryotes. RNA 5: 625–635, 1999.
14. Martin GW, 3rd, Harney JW, Berry MJ. Functionality of mutations at conserved nucleotides in eukaryotic SECIS elements is determined by the identity of a single nonconserved nucleotide. RNA 4: 65–73, 1998.
15. Howard MT, Gonzales-Flores JN, Copeland PR. Molecular mechanism of eukaryotic selenocysteine insertion. In: Selenium: Its Molecular Biology and Role in Human Health, edited by Hatfield DL, Berry MJ, Gladyshev VN. New York: Springer, 2012, p. 33–46.
16. Labunskyy VM, Hatfield DL, Gladyshev VN. Selenoproteins: molecular pathways and physiological roles. Physiol Rev. 2014 Jul;94(3):739-77.146, 183, 186.
17. Tujebajeva RM, Copeland PR, Xu XM, Carlson BA, Harney JW, Driscoll DM, Hatfield DL, Berry MJ. Decoding apparatus for eukaryotic selenocysteine insertion. EMBO Reports 1: 158–163, 2000.
18. Fagegaltier D, Hubert N, Yamada K, Mizutani T, Carbon P, Krol A. Characterization of mSelB, a novel mammalian elongation factor for selenoprotein translation. EMBO J 19: 4796–4805, 2000.
19. Hoffmann Peter R. and Berry Marla J. Selenoprotein Synthesis: A Unique Translational Mechanism Used by a Diverse Family of Proteins.Thyroid® 2005 15:8, 769-775.
20. Linda Johansson, Guro Gafvelin, Elias S.J. Arnér, Selenocysteine in proteins—properties and biotechnological use, Biochimica et Biophysica Acta (BBA) - General Subjects, Volume 1726, Issue 1, 2005, Pages 1-13.
21. Jun Lu, Arne Holmgren, Selenoproteins\*, Journal of Biological Chemistry, Volume 284, Issue 2, 2009, Pages 723-727, ISSN 0021-9258, <https://doi.org/10.1074/jbc.R800045200>.
22. Rayman M.P. Selenium and human health. Lancet. 2012;379:1256–1268. doi: 10.1016/S0140-6736(11)61452-9.
23. Kieliszek M. Selenium(-)fascinating microelement, properties and sources in food. Molecules. 2019;24:1298. doi: 10.3390/molecules24071298.
24. Papp L.V., Lu J., Khanna K.K. From selenium to selenoproteins: Synthesis, identity, and their role in human health. Antioxid. Redox Signal. 2007;9:775–806. doi: 10.1089/ars.2007.1528.;
25. Ye R, Huang J, Wang Z, Chen Y, Dong Y. The Role and Mechanism of Essential Selenoproteins for Homeostasis. Antioxidants (Basel). 2022 May 15;11(5):973.
26. Hatfeld DL, Berry MJ, Gladyshev VN (2006) Selenium: its molecular biology and role in human health, 2nd edn. Springer, New York, pp 3–287.
27. Bellinger FP, Raman AV, Reeves MA, Berry MJ (2009) Regulation and function of selenoproteins in human disease. Biochem J 422:11–22.
28. Romagné F, Santesmasses D, White L, Sarangi GK, Mariotti M, Hübler R, Weihmann A, Parra G, Gladyshev VN, Guigó R and Castellano S SelenoDB 2.0: annotation of selenoprotein genes in animals and their genetic diversity in humans Nucleic Acids Res, 42, D437-D443 (2014).
29. The UniProt Consortium, UniProt: the universal protein knowledgebase in 2021, Nucleic Acids Res. 49:D1 (2021).