

MBPDB Users Guide

Table of contents:

Searching the Milk Bioactive Peptide Database	2
Simple Peptide Search.....	2
Advanced Search Options	2
Homology Seach Options	2
Sequence search options	2
Similarity Threshold	3
Scoring Matrix.....	3
Catagorical Search Options	4
Protein ID	4
Function.....	4
Species.....	4
Search Results.....	5
Column Descriptions (Database Matches):	5
Column Descriptions (Blast Output):	6
Administrative Tasks	7
Accessing Administrative Features	7
Procedure:	7
Adding new peptides	7
Procedure:	7
Exporting the entire dataset.....	7
Procedure:	7
MBPDB add proteins	8
Procedure:	8
References:	8

Searching the Milk Bioactive Peptide Database

The Milk Bioactive Peptide Database (MBPDB) includes all known bioactive peptides derived from milk proteins from any species. This program allows inputting peptide sequences of interest, including the results of peptidomics analyses, to determine which have homology with known bioactive peptides.

Simple Peptide Search

Input the IUPAC one letter amino acid sequence of your peptide(s) such as YPFPGPIP. If searching multiple peptides enter each peptide on a separate line in the search box or use the “load from a text file” link to upload a basic .csv, .TSV, or .txt file. Click search DB to query the MBPDB against your peptides or the add additional search parameters by clicking the advanced search options.

Advanced Search Options

These options allow for customization of the criteria for the peptide’s homology search, along with the ability to search based on categorical fields within the database. Several search options are available, and they can be used individually or in combination. Categorical fields such as species, protein or function can be searched independently without inputting peptides. For example, one could search the database for all cow derived peptides by exclusively inputting “cow” or “bos taurus” into species.

Homology Search Options

Sequence search options

Select a search option of your desire. The search functions are based on the protein blast search algorithm.

Sequence

Matches the entire input to all database entries and returns matches that has a equal or higher than the threshold value set.

Truncated

Searches for any of the bioactive database entries containing the with an equal or higher similarity than the set . The input amino acid sequence is indicated with **bold** text in the result sequences.

Precursor

Searches for bioactive peptides within the database which is contained within the with equal or higher similarity than the set. You can also input part of a protein sequence to identify bioactive peptides in that region of the protein. To do this you would have to select precursor in the .

Similarity Threshold

This option allows for searching database entries with similarities to the input single peptide sequence. The search is based on a scoring matrix which is either or . A restriction of the similarity search is that the peptides is at least four amino acids long. Sequences below four amino acids will automatically only search sequences that has a 100% similarity even though the indicated value is set lower than this.

Scoring Matrix

The scoring matrix assigns an alignment score for all pair of amino acid residue matches to calculate the between the and database entries.

Currently the database contains the and scoring matrix. If both scoring matrix are selected for a search, then the results will show each matching peptide with associated with the scoring matrix selected displayed in a new column.

Identity

This scoring matrix is based on exact pairwise amino acid residues matches.

BLOSUM62

The BLOSUM62 scoring matrix uses an amino acid substitution matrix for the alignment score of all pairs of amino acid residue matches to calculate the (Figure 1).

S	9
S -1	4
T -1	1 5
P -3	-1 -1 7
A 0	1 0 -1 4
G -3	0 -2 -2 0 6
N -3	1 0 -2 -2 0 6
D -3	0 -1 -1 -2 -1 1 6
E -4	0 -1 -1 -1 -2 0 2
Q -3	0 -1 -1 -1 -2 0 0 2 5
H -3	-1 -2 -2 -2 -2 1 -1 0 0 8
R -3	-1 -1 -2 -1 -2 0 -2 0 1 0 5
K -3	0 -1 -1 -1 -1 -2 0 -1 1 1 -1 2 5
M -1	-1 -1 -2 -1 -3 -2 -3 -2 0 -2 -1 -1 5
I -1	-2 -1 -3 -1 -4 -3 -3 -3 -3 -3 -3 1 4
L -1	-2 -1 -3 -1 -4 -3 -4 -3 -2 -3 -2 -2 2
V -1	-2 -2 -2 -2 -3 -3 -3 -2 -2 -3 -3 -2 1 3 1 4
F -2	-2 -2 -4 -2 -3 -3 -3 -3 -3 -1 -3 -3 0 0 -1 6
Y -2	-2 -2 -3 -2 -3 -2 -3 -2 -1 2 -2 -2 -1 -1 -1 3 7
W -2	-3 -2 -4 -3 -4 -3 -4 -3 -2 -3 -3 -3 -2 -3 1 2 11
C S T P A G N D E O H R K M I L V F Y W	

Figure 1: BLOSUM62 substitution matrix (&, 1992).

Catagorical Search Options

These search features allow the user further ability to increase the specificity of a query or explore the database for peptides categorized by the search criteria of protein, bioactivity function and species of a peptides origin.

- Each search option has a dropdown menu containing a list of available search parameters.
- Either the common name or scientific name/ID are acceptable inputs:
 - Protein: Beta-casein or P02666
 - Species: cow or Bos taurus
- Multiple of each search option can be entered if separated by a comma:
 - Species: Cow, sheep, homo sapiens
 - Protein: P02662, P47710, P04653, P09115, P18626, O97943
 - Function: Anti-microbial, ACE-inhibitory
- If no peptide(s) are inputted, the search will return all peptide matches for the provided categorical search options
- If peptide(s) are inputted, the results will be further limited to the inputted categorical search options

Protein ID

Searches for bioactive peptides derived from a specific protein. Input the protein ID also known as protein entry on uniprot.org or common name.

Example: Beta-casein or P02666 for exclusively bovine beta-casein.

Function

Searches for bioactive peptides with a specific bioactivity function.

Example: ACE-inhibitory, antimicrobial, etc.

Species

Searches for bioactive peptides with a specific species of origin.

Example: cow, sheep, homo sapiens

Search Results

The results of the query are displayed on the website under the search inputs. A hyperlink “Download results as a TSV file” is generated directly under the Search DB button which will download the results for the query displayed below along with a link to for citation of the database. Certain columns are only displayed when a relevant search is preformed such as

Column Descriptions (Database Matches):

A brief description of each column is below

Column	Description
Search Peptide	The peptide inputted into the search.
Protein ID	The Uniport protein ID associated with the searched peptide or category.
Peptide	The peptide matched against the search parameters.
Species	The scientific name for the matched species.
Intervals	The peptide’s amino acid location within the parent protein sequence.
Function	The primary bioactivity function associated with the matched peptide
Secondary function	TBD
PTM	The Post translational modification (PTM) associated with the matched peptide.
Title	The title of the source article for the match peptide. The results in this column might be nested to the matched peptide and its function.
Authors	The authors of the source article for the match peptide. The results in this column might be nested to the matched peptide and its function.
Abstract	The abstract of the source article for the match peptide. The results in this column might be nested to the matched peptide and its function.
DOI	The DOI of the source article for the match peptide. The results in this column might be nested to the matched peptide and its function.
Similarity Threshold	The similarity threshold used for the search. This column won’t populate if an irrelevant search was performed.
Scoring matrix	The scoring matrix used for the search. This column won’t populate if an irrelevant search was performed.
Sequence Search Options	The sequence search options used for the search. This column won’t populate if an irrelevant search was performed.

Column Descriptions (Blast Output):

A brief description of each column associated with the protein Blast search.
Blast information is not obtained for peptides below four amino acids long or if
the query is exclusively a similarity homology search with a threshold of
100%

Column	Description
Subject	
Start	
Stop	
e-value	
Alignment length	
Mismatches	
gaps	

Administrative Tasks

Accessing Administrative Features

Certain tasks are restricted and require accessing sites admin portal

Procedure:

1. Contact the existing site admin to be granted administrative access.
2. Log in as the database admin mbpdb.nws.oregonstate.edu/admin/ or add “/admin/ to the end of the homepage’s url using your user credentials

Adding new peptides

This function is for adding multiple entries into the database by uploading the information as a .TSV file. An example of the file format is available on the online database and an example is shown below. Before the entry is added, it needs to be accepted by a database administrator.

Procedure:

1. Consolidate your new entries into a tab separated (.TSV) file with the correct column names and formatting. Reference the [example link](#)
2. While logged in as the site administrator click on the following links:
 - a. Submissions (under PEPTIDE list)
 - i. ADD SUBMISSION (top right button)
 1. MBPDB add multiple entries using TSV file hyperlink (bottom middle)
3. Now on the new page uploaded your .TSV file ensuring no errors were found and the correct number of entries were submitted
4. Navigate back to the Submissions tab of the admin/peptide portal
 - a. Click the check box in the header to select all submissions or individually check these
 - b. In the actions menu select either approve or delete the submission and then click go
 - c. Acknowledge you made the correct decision, by clicking the “yes, I’m sure” button

Exporting the entire dataset

This is an option to export the entire dataset into a .TSV file for access outside of the web application

Procedure:

1. While logged in as the site administrator click on the following links:
 - a. Peptide infos (under PEPTIDE list)
 - i. Select all peptides, using the check box next to PEPTIDE
 - ii. In the Action dropdown menu, select “Export selected to TSV”
 - iii. Click the Go button to complete the download

MBPDB add proteins

This function is made for adding protein information into the database in fasta format (example shown below). Select one or multiple fasta files to be uploaded and press run. This function should be used when adding entries to the database which gives the error that the is not in the database. This function could also be used to add genetic variant of proteins by adding their sequence information in fasta format. Several proteins are milk proteins are already added to the database and can be view through the “View current protein fasta headers (IDs)” link.

```
>sp|P02666|CASB_BOVIN Beta-casein OS=Bos taurus GN=CSN2 PE=1 SV=2
MKVLILACLVALALARELEELNVPGEIVESLSSSEESITRINKKIEKFQSEEQQQTEDEL
QDKIHPFAQTQSLVYPFPGPPIPNSLPQNIPPLTQTPVVVPPFLQPEVMGVSKVKEAMAPK
HKEMPPPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVLSL
SQSKVLPVPQKAVPYQPQRDMPIQAFLLYQEPVLCPVRGPFPIIV
```

Procedure:

1. While logged in as the site administrator click on the following links:
 - a. Protein infos (under PEPTIDE list)
 - i. ADD PROTEIN INFO (top right button)
 1. Add Proteins (bottom middle)
2. Now on the new page uploaded your .fasta file ensuring no errors were found and the correct number of entries were submitted
3. Navigate back to the Protein info tab of the admin/peptide portal and look for the new proteins on the top of the list

References:

Henikoff S, Henikoff J.G. 1992. Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci U S A. 1992 Nov 15; 89(22): 10915–10919.