MBPDB Users Guide

Table of contents:

Version 2.0 9/2023

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 YPFPGPIPN. If searching for 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 withing 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 selecting cow from the species drop down menu

Homology Seach 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).

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

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

Catagorical Search Options

These search features allow increased the specificity of a query or the ability to explore the database for peptides categorized by the search criteria of protein, bioactivity function and species of a peptide's origin. Each search option has a dropdown menu containing a list of available search parameters. Multiple search option can be selected. 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. Select 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. Peptides with a * in the table below are only displayed if at least one peptide in the query has matching data.

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.
Function	The bioactivity function associated with the matched peptide.
Additional details*	Additional details that complement the peptide's function.
IC50 (μM)*	Half maximal inhibitory concentration (IC50) is a measure of the potency of a substance in inhibiting a specific biological or biochemical function expressed in mico-Moles.
Inhibition type*	The method of determining the peptide's anti-microbial inhibition. Options are MIC (minimum inhibitory concentration), % colonies inhibited, and log(N0/Nf). Reference the article as the methods varies.
Inhibited microorganisms*	The species of microbe that is inhibited by the antimicrobial peptide.
PTM*	The Post translational modification (PTM) associated with the matched peptide.
Tittle	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.
* These columns will be absent in the results, if there is no data for the matched peptide	

<u>Column Descriptions (Blast Output):</u>

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	The reference protein sequence that has been found to be similar to the query sequence
Query start	The position in the query sequence where the alignment with the reference protein (Subject) begins.
Query end	The position in the query sequence where the alignment with the reference protein (Subject) ends.
Subject start	The position in the reference protein (Subject) where the alignment with the query sequence begins.
Subject end	The position in the reference protein (Subject) where the alignment with the query sequence ends.
e-value	A measure of the statistical significance of the match; lower e-values indicate more significant matches.
Alignment length	The total number of amino acid residues spanned by the alignment between the query and subject sequences.
Mismatches	The number of positions where the query and subject sequences do not match within the alignment.
gaps	The number of positions in the alignment where either the query or the subject sequence has a gap, indicating an insertion or deletion.

Administrative Tasks

The instructions for performing administrative tasks, such as adding new peptides, proteins, saving and exporting the database are in a separate private file found in the [Dallas Lab Box storage](#) (Dallas Lab Share > 3_Bioactive peptide database > MBPDB_Help Administrative.docx).

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.