

PepHemo-LLM Instruction

PepHemo-LLM is a user-friendly, easy-to-access web server that can be used to predict the hemolytic nature of peptides with sequences as the input. The hemolytic activity of a peptide indicates its ability to destroy red blood cells, leading to the release of hemoglobin. Numerous therapeutic peptides failed in clinical trials just because of their high hemolytic activity. Since experimental determination of the hemolytic potency of a large number of peptides is a labor-intensive, time-consuming, and costly, we developed **PepHemo-LLM** which can serve the scientific community to screen hemolytic peptides in an ultrafast manner. The model was constructed (using experimentally determined 1,140 positive and 713 negatives along with 442 randomly generated negative peptides) based on a protein large language model ProteinBERT to tokenize peptide sequences. The web server is implemented based on the Flask framework, with Apache 2.4.58, Python 3.10.20, and Flask 2.0.3.

1. Input to PepHemo

The input from users is simple and straightforward: just the sequences of peptides to be predicted. There are two ways to input peptide sequences, as shown below:

Input of Peptide FASTA Sequences

The screenshot displays the web interface for inputting peptide sequences. It features a text input area, a file upload option, and two buttons: 'SUBMIT' and 'HELP'.

1. directly input peptide sequences in the fasta format

```
>peptide_43
RLYRKVYGRLYRKVYGRLYRKVYGRLYRKVYGRLYRKVYGRLYRKV
YGRLYRKVYGRLYRKVYGRLYRKVYGRLYRKVYGRLYRKVYGRLYRKV
>peptide_126
DGVRYSPRLRIVQELNAAAGAHG
```

2. upload a file (tmp_peptide.fasta) of peptide sequences in the fasta format

or Upload a File ([Example](#)):
Choose File tmp_peptide.fasta

SUBMIT **HELP**

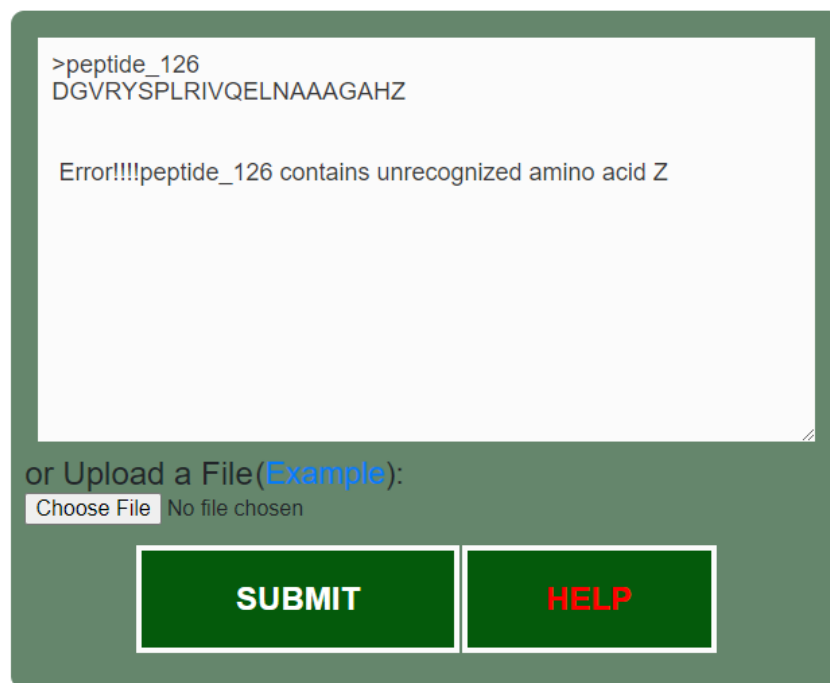
The input peptide sequences must consist of amino acids in single uppercase characters in the fasta format. Otherwise, the input will be considered as “illegal”. If this happens, an error message will show up in the input Form. Here are some examples with incorrect input:

1). Unrecognized amino acid characters

```
>peptide_126  
DGVRYSPRLRIVQELNAAAGAHZ
```

An error message will show up to indicate there is an unrecognized amino acid in the sequence (“Z” here).

Input of Peptide FASTA Sequences



The screenshot shows a web form with a text input area containing the following text:

```
>peptide_126  
DGVRYSPRLRIVQELNAAAGAHZ
```

Below the input area, an error message is displayed: "Error!!!!peptide_126 contains unrecognized amino acid Z".

Below the error message, there is a section for file upload: "or Upload a File(Example):". This section includes a "Choose File" button and the text "No file chosen".

At the bottom of the form, there are two buttons: "SUBMIT" and "HELP".

2). Input without “>” or names for peptides

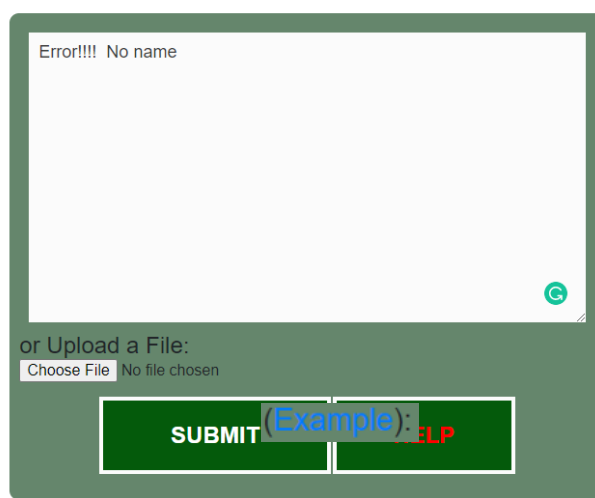
```
peptide_126  
DGVRYSPRLRIVQELNAAAGAHG
```

Or

```
>  
DGVRYSPRLRIVQELNAAAGAHG
```

An error message will show up to indicate there is no name for the peptide.

Input of Peptide FASTA Sequences



Error!!!! No name

or Upload a File:
Choose File No file chosen

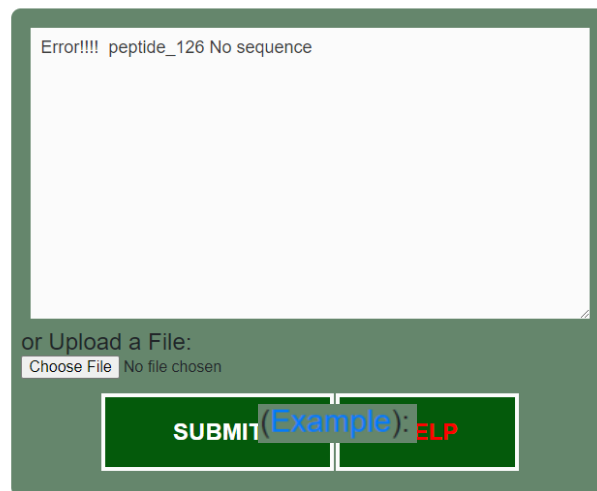
SUBMIT (Example): ELP

3). Peptide name and sequence in the same line

>peptide_126 DGVRYSPLRIVQELNAAAGAHG

An error message will show up to indicate peptide_126 has no sequence.

Input of Peptide FASTA Sequences



Error!!!! peptide_126 No sequence

or Upload a File:
Choose File No file chosen

SUBMIT (Example): ELP

2. Output of PepHemo-LLM

The output of **PepHemo-LLM** is also easy to understand, with three columns in a table. The first column is the peptide name or ID, as provided by users in the input; The second column is the predicted hemolytic properties, with 1 representing hemolytic and 0 representing non-hemolytic. Please refer to our manuscript for more details. The below shows some examples of prediction: the peptide test-n125 is predicted as non-

hemolytic (0) while the peptide test-p296 as hemolytic (1), both in agreement with experimental observations.

Output: Predicted Peptide Hemolyticity [\(Download Results\)](#)

Peptide Names	Hemolytic? (Yes=1, No=0)
test-n125	0
test-p296	1
test-p11	1
train-n16	0
test-n34	0
test-p271	1
train-p649	1

In addition to visualizing the result in a table, we also dynamically provide users with an option to download the data as a prediction.csv file. This is particularly useful if the input number of sequences is large (e.g., >50). The format of the downloadable prediction.csv is as follow:

```
Peptide Name,"Hemolytic? (Yes=1, No=0) "  
test-n125,0  
test-p296,1  
test-p11,1  
train-n16,0  
test-n34,0  
test-p271,1  
train-p649,1
```

3. Instruction and Help

This detailed instruction can be found by clicking on the red-yellow “Help” button next to the “Submit” button.