HIVP-SAE Instruction

HIVP-SAE is a Flask framework-based web server that can be used for drug resistance prediction in human immunodeficiency virus (HIV) for personalized therapies. HIV-1 is known for its high mutation rate in its protease (HIVP), and the protease susceptibility testing is essential to develop new antiretroviral drugs or optimize the use of existing drugs. HIVP-SAE is built to predict the HIV susceptibility to drug treatment defined by the IC_{50} of a mutant vs. the wild-type control. To this end, we curated a large dataset including 776 unique HIVP mutants against 5 drugs with experimentally determined positive (susceptible to drug treatment) and negative (resistant to drug treatment) testing results. Based on the data, we developed 5 deep learning models to classify HIVP mutants with respect to five protease inhibitors (IDV: Indinavir, SQV: Saquinavir, NFV: Nelfinavir, APV: Amprenavir, LPV: Lopinavir). The models are derived from CNN-based supervised autoencoder (SAE) with new Chaos Game Representation (CGR) scheme to characterize HIVP sequences, and the web server is implemented with Python 3.8.10, Torch 1.11.0, and Flask 2.4.

1. User Input

The input of **HIVP-SAE** is simple and straightforward. There are two ways as shown below:

Input of HIVP Mutant FASTA Sequence(s)

1. directly input protein sequences in the fasta format

2. upload a file (tmp_HIVP.fasta) of protein sequences in the fasta format

```
>88_APV
PQITLWQRPIVPIRIEGQLTEALLDTGADDTVLEEINLSGRWKPKMIGGIGGFVKVRQYDQVTIEIC
DHKVIGTVLVGPTPANIIGRNLLTQLGCTLNF
>0_LPV
PQITLWQRPVVTIRIGGQLKEALLDTGADDTVLEEINLPGRWKPKMIGGLGGFIKVRQYDQILIEIC
GHKVAIGTVLVGPTPVNIIGRNMLTQIGCTLNF
>99_LPV
PQITLWQRPLVTIKIGGQLKEALLDTGADNTVLEEMNLPGRWKPKMIGGIGGFIKVRQYDQIPIEIC
GHKAIGTVLIGPTPVNIIGRDLMTQLGCTLNF
```

or Upload a File (Example):
Choose File | tmp_HIVP.fasta

SUBMIT   HELP

The users' input of protein sequences must consist of amino acid characters as single uppercase letters 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 of incorrect input:
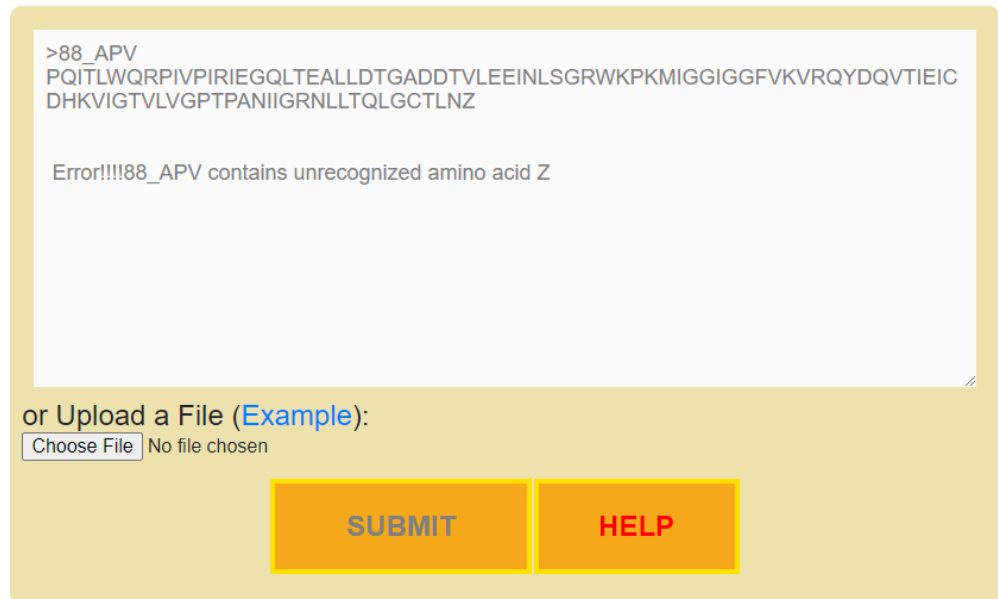
1). Unrecognized amino acid characters

>88_APV

PQITLWQRPIVPIRIEGQLTEALLDTGADDTVLEEINLSGRWKPKMIGGIGGFVKVRQYD
QVTIEICDHKVIGTVLVGPTPANIIGRNLLTQLGCTLNF

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

Input of HIVP Mutant FASTA Sequence(s)

>88_APV
PQITLWQRPIVPIRIEGQLTEALLDTGADDTVLEEINLSGRWKPKMIGGIGGFVKVRQYDQVTIEIC
DHKVIGTVLVGPTPANIIGRNLLTQLGCTLNZ

Error!!!!88_APV contains unrecognized amino acid Z

or Upload a File (Example):
Choose File No file chosen

SUBMIT   HELP

2). Input without ">" or names for peptides

88_APV

PQITLWQRPIVPIRIEGQLTEALLDTGADDTVLEEINLSGRWKPKMIGGIGGFVKVRQYD
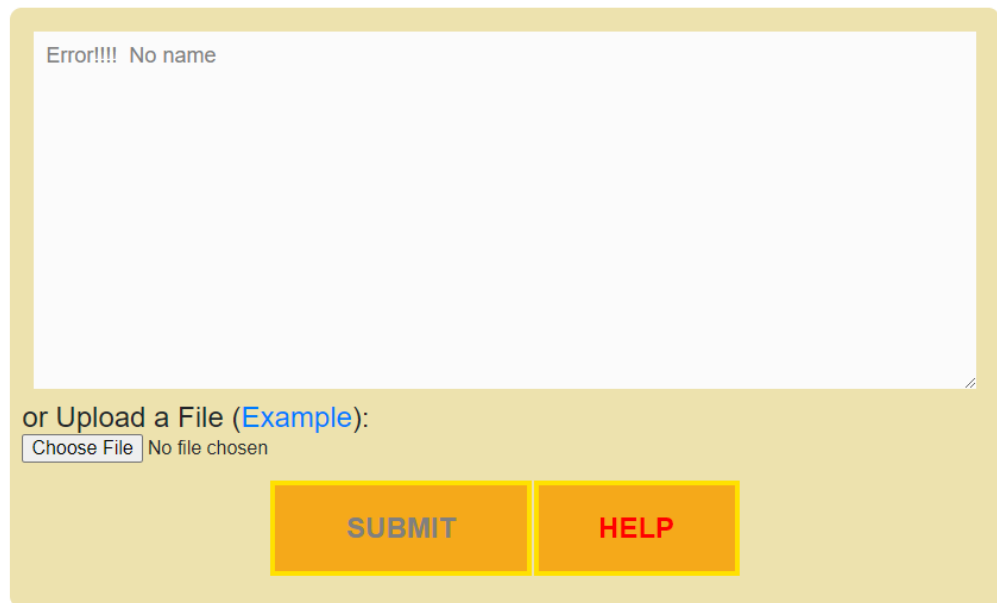QVTIEICDHKVIGTVLVGPTPANIIGRNLLTQLGCTLNF

Or

>

PQITLWQRPIVPIRIEGQLTEALLDTGADDTVLEEINLSGRWKPKMIGGIGGFVKVRQYD
QVTIEICDHKVIGTVLVGPTPANIIGRNLLTQLGCTLNF

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

Input of HIVP Mutant FASTA Sequence(s)
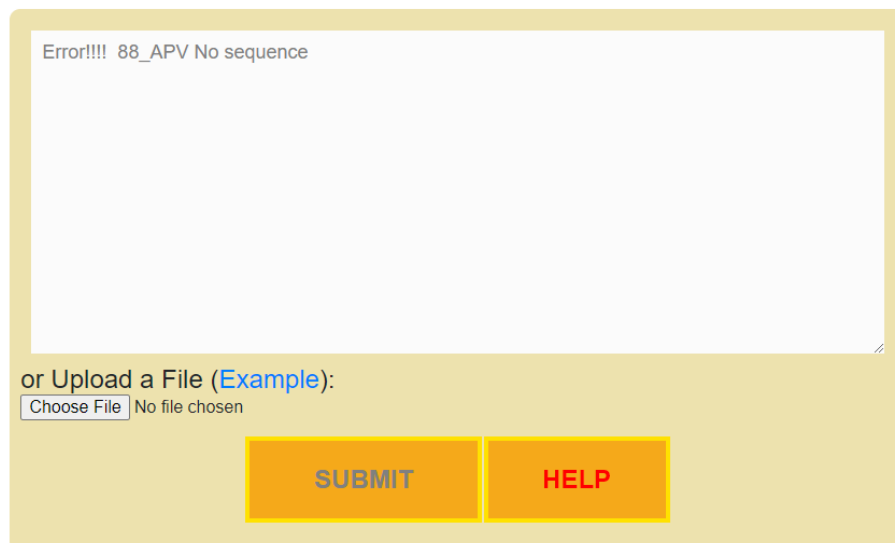
Error!!!! No name

or Upload a File (Example):
Choose File   No file chosen

SUBMIT   **HELP**

3). Peptide name and sequence in the same line

>88_APV   QITLWQRPIVPIRIEGQLTEALLDTGADDTVLEEINLSGRWKPKMIGGIGGFV
KVRQYDQVTIEICDHKVIGTVLVGPTPANIIGRNLLTQLGCTLNF

An error message will show up to indicate 88_APV has no sequence.

Input of HIVP Mutant FASTA Sequence(s)

Error!!!! 88_APV No sequence

or Upload a File (Example):
Choose File   No file chosen

SUBMIT   **HELP**

It is worth noting the HIVP mutants generally have 99 amino acids in their monomer sequences, except some with insertions or deletions. However, HIVP-SAE can accept any length of protein sequences and make predictions; please be advised that such

predictions may not be robust since our models were built mostly based on mutants with 99 residues.

2. Output of HIVP-SAE

The output of **HIVP-SAE** is also easy to understand, with six columns in a table. The 1st column is the HIV protein name or ID as provided by users. From the 2nd to the 6th column are the predicted susceptibility of the input sequences with respect to 5 HIVP inhibitors (drugs): LPV, IDV, NFV, SQV and APV. "1" represents the mutant is susceptible to the drug (the drug us effective to inhibit the virus), and "0" represents the mutant is resistant to the drug (the drug is ineffective to treat the virus). Please refer to our manuscript for more details. The below shows some examples of predictions: the mutant 88_APV is susceptible to protease inhibitors IDV, SQV, and APV, while resistant to the other two drugs (LPV and NFV). This means if a patient has 88_APV mutant, he/she can be treated with IDV, SQV, or APV, but not LPV or NFV. Similar interpretations can be applied to other HIVP mutants with respect to the 5 HIVP inhibitors.

Output of Predicted HIVP Susceptibility to Drugs (Download Data)

HIVP Mutant	Lopinavir (LPV)	Indinavir (IDV)	Nelfinavir (NFV)	Saquinavir (SQV)	Amprenavir (APV)
88_APV	0	1	0	1	1
0_LPV	1	0	1	1	0
99_LPV	1	1	0	0	1

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

```
HIVP Name    LPV      IDV      NFV      SQV      APV
88_APV       0        1        0        1        1
0_LPV        1        0        1        1        0
99_LPV       1        1        0        0        1
```

3. Frequently Asked Questions

This FAQ section is quite self-explanatory. It provides brief answers to the most frequently asked questions, along with some more information about the use of our tools and services.