

HIVP-CGR Instruction

HIVP-CGR is a Flask framework-based web server that can be used for drug susceptibility and 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-CGR 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 supervised autoencoder (SAE) with new Chaos Game Representation (CGR) scheme to characterize HIVP sequences, and the web server is implemented with Apache 2.4.58, Python 3.12.3, and Flask 2.0.3.

1. User Input

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

Input of HIVP Mutant FASTA Sequence(s)

The screenshot shows a web form for inputting HIVP mutant FASTA sequences. It features two input methods:

- 1. directly input protein sequences in the fasta format:** A text area containing three FASTA entries:

```
>88_APV
PQITLWQRPVPIRIEQLTEALLDTGADDTVLEEINLSGRWPKMIGGIGGFVKVRQYDQVTIEIC
DHKVIGTVLVGPTPANIIGRNLLTQLGCTLNF
>0_LPV
PQITLWQRPVVTIRIGGQLKEALLDTGADDTVLEEINLPGRWPKMIGGLGGFIKVRQYDQILIEIC
GHKVAIGTVLVGPTPVNIIGRNMLTQIGCTLNF
>99_LPV
PQITLWQRPLVTIKIGGQLKEALLDTGADNTVLEEMNLPGRWPKMIGGIGGFVKVRQYDQIPIEIC
GHKAIGTVLIGPTPVNIIGRDLMTQLGCTLNF
```
- 2. upload a file (tmp_HIVP.fasta) of protein sequences in the fasta format:** A file upload button labeled "Choose File" with the filename "tmp_HIVP.fasta" next to it.

Below the input area are two buttons: "SUBMIT" and "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:

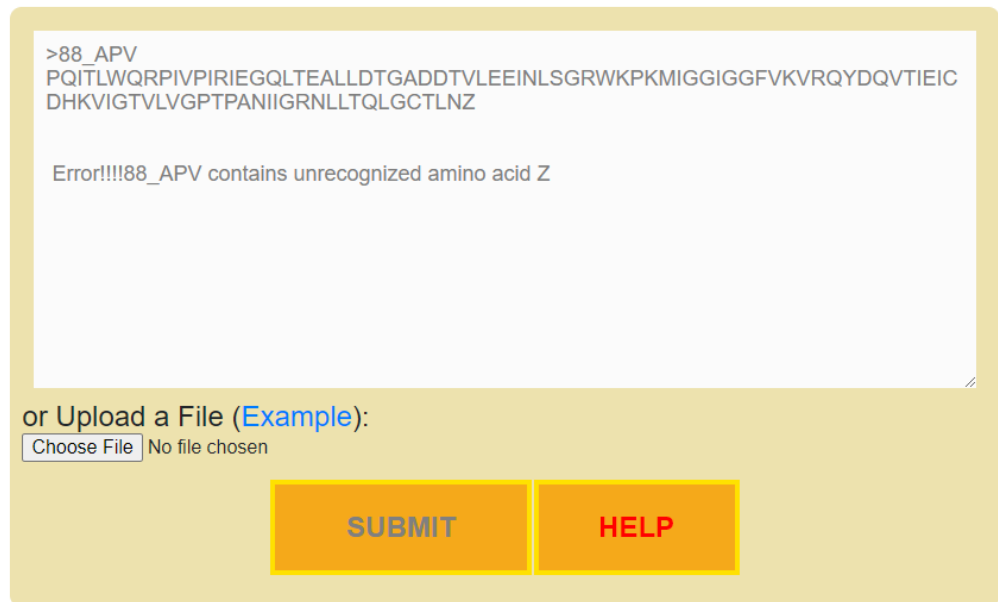
1). Unrecognized amino acid characters

>88_APV

PQITLWQRPIVPIRIEQLTEALLDTGADDTVLEEINLSGRWKPKMIGGIGGFVKVRQYD
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)



The screenshot shows a web form with a text input field containing the following FASTA sequence:

```
>88_APV
PQITLWQRPIVPIRIEQLTEALLDTGADDTVLEEINLSGRWKPKMIGGIGGFVKVRQYDQVTIEIC
DHKVIGTVLVGPTPANIIGRNLLTQLGCTLNZ
```

Below the input field, an error message is displayed: "Error!!!!88_APV contains unrecognized amino acid Z".

Below the error message, there is a section for file upload: "or Upload a File (Example):" with 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

88_APV

PQITLWQRPIVPIRIEQLTEALLDTGADDTVLEEINLSGRWKPKMIGGIGGFVKVRQYD
QVTIEICDHKVIGTVLVGPTPANIIGRNLLTQLGCTLNF

Or

>

PQITLWQRPIVPIRIEQLTEALLDTGADDTVLEEINLSGRWKPKMIGGIGGFVKVRQYD
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](#)):

No file chosen

3). Protein name and sequence in the same line

```
>88_APV QITLWQRPVPIPIRIEQLTEALLDTGADDTVLEEINLSGRWPKPKMIGGIGGFV  
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](#)):

No file chosen

It is worth noting the HIVP mutants generally have 99 amino acids in their monomer sequences, except some with insertions or deletions. However, HIVP-CGR 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-CGR

The output of **HIVP-CGR** 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 is 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 (APV_pos406, IDV_neg30, LPV_neg5, NFV_neg98, SQV_pos391) is susceptible to protease inhibitors APV and SQV, while resistant to the other three drugs (IDV, LPV and NFV). This means if a patient has this HIVP mutant, he/she can be treated with APV or SQV, but not IDV, LPV or NFV. Of note, the wild type HIVP is sensitive to all 5 drugs. Similar interpretations can be applied to other HIVP mutants with respect to the 5 HIVP inhibitors. All such predictions are in agreement with experimental observations.

Output: Predicted Drug Susceptibility ([Download Data](#))

HIVP Mutant	APV	IDV	LPV	NFV	SQV
APV_pos406, IDV_neg30, LPV_neg5, NFV_neg98, SQV_pos391	1	0	0	0	1
APV_neg201, IDV_pos262, LPV_neg100, NFV_pos302, SQV_pos359	0	1	0	1	1
APV_neg134, IDV_neg28, LPV_pos187, NFV_neg383, SQV_neg194	0	0	1	0	0
APV_neg139, IDV_neg122, LPV_neg141, NFV_neg345, SQV_neg275	0	0	0	0	0
Wild_Type	1	1	1	1	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:

Name	LPV	IDV	NFV	SQV	APV
APV_pos406, IDV_neg30, LPV_neg5, NFV_neg98, SQV_pos391	0	0	0	1	1
APV_neg201, IDV_pos262, LPV_neg100, NFV_pos302, SQV_pos359	0	1	1	1	0
APV_neg134, IDV_neg28, LPV_pos187, NFV_neg383, SQV_neg194	1	0	0	0	0
APV_neg139, IDV_neg122, LPV_neg141, NFV_neg345, SQV_neg275	0	0	0	0	0
Wild_Type	1	1	1	1	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.