# Seq2Feature

Seq2Feature is a Python based toolkit to compute the numerical values associated to the DNA and protein sequences which can be useful for extracting the valuable descriptors for several prediction purposes.

To use these codes, you need following packages

\*\*\*these codes will run on linux Only!!!!!!!!!

Dependencies are:

1. python2.7 or above

2. pandas

You can download and extract these codes from

<https://www.iitm.ac.in/bioinfo/SBFE/help.html>

and go to Seq2Feature directory. In this directory following codes are kept:

1. sequence\_based\_features1.py: To calculate change upon mutation in a protein sequence

2. sequence\_based\_features2.py: To calculate average property values for protein sequence

3. seq\_dna1.py: To calculate change upon mutation in DNA sequence

4. seq\_dna2.py: To calculate average property values for DNA sequence

Steps:

1 Install required python packages

1. python2.7 or above

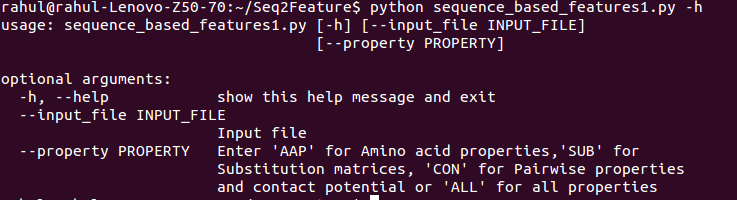
2. pandas

2. Go to Seq2Feature directory

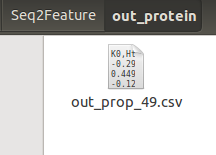
**Protein**

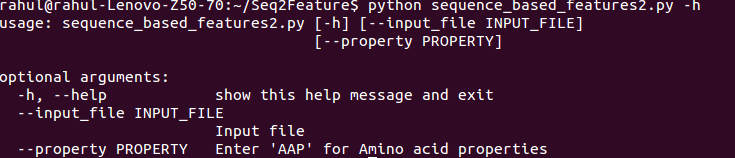
3. To calculate change upon mutation, run (sequence\_based\_features1.py) the following command:

Here input\_file parameter accepts tab separated file as given protein\_sample\_input.tsv. To calculate any required property, you can choose either of AAP, SUB, CON, or ALL.



4. After executing the code, output file (.csv) will be generated in out\_protein directory.



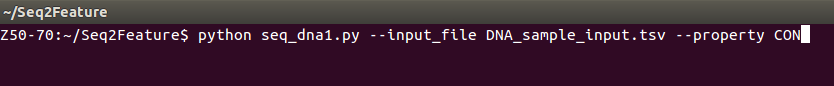
Similarly, to calculate average property values for protein sequence you can use sequence\_based\_features2.py

**Steps**

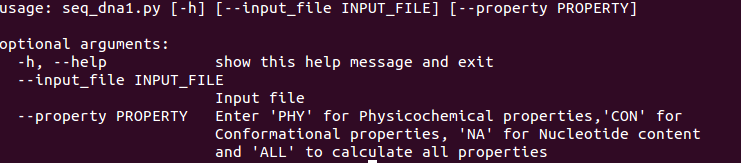
**DNA**

For DNA you can follow above mentioned steps 1 and 2 same as proteins.

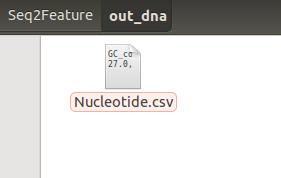
3. To calculate change upon mutation, run (seq\_dna1.py) the following command:

****

Here, input\_file parameter accepts tab saperated file as given protein\_sample\_input.tsv. You can calculate properties using any one of PHY, CON, or NA



4. After execution of the code output file (.csv) will be generated in out\_dna directory.



Similarly, to calculate average property values for DNA sequence, you can use seq\_dna2.py with the following command:

