



LEARN PYTHON & R FOR BIOINFORMATICS

Prerequisite Terminologies:

In order to have better understanding of the main topic, you should have the basic concept of the following term:

➤ Signal Peptide

Introduction:

SignalP is the currently most widely used tool for prediction of signal peptide from protein sequences. Signal peptides are short amino acid sequences in the amino terminus of many newly synthesized proteins that target proteins into or across the membranes. Among other Bioinformatics tools, SignalP is an efficient tool to improve the prediction of the presence of signal peptides and the location of their cleavage sites in proteins from Eukarya, Gram-positive bacteria, Gram-negative bacteria and Archaea. There are various versions of SignalP tool, the latest version is SignalP-5.0.

Steps:

- Open the SignalP-5.0 server which you can access from [here](#).
- Enter your query sequence in FASTA format in the given box or you can select the fasta file from your computer. You can also select the 'Example proteins'.

[In this video, we've utilized the fasta file of LCT protein which was retrieved from SwissProt, you can also do the same.]

Note: The best thing about SignalP server is that you can input various proteins at once. For example if you've multiple proteins, you can input them at once. SignalP can work with all of them efficiently.

- Select the '**Organism group**' based on the sequence you've entered.

[We've selected the '*Eukarya*' because the LCT protein is homo sapiens protein.]

- From the '**Output format**' option, select the '*Long output*'.

Note: You should always select the '*Long output*' option because it provides you the location of signal peptide and cleavage. '*Short output*' display results without figures.

- Click on '**submit**' to get the results.

Analyzing Results:

- You'll get the summary of your sequence where the following information is present.
 - Name and accession number of the protein.
 - It'll display you what sort of prediction is done like, 'Signal peptide (Sec/SPI)'.
 - The **cleavage site position** is provided.
 - **Protein type**
 - **Likelihood value**
- Since we've selected the long output, you'll also get a chart to analyze the signal peptide where 'probability' is given on y-axis and 'protein sequence' is given on x-axis.
- On the x-axis, you'll see the predicted sequence along with the protein sequence which starts from many 'S' and continues as many 'X'. Between them a 'C' is present.
 - The '**S**' region means that this is the signal peptide region.
 - '**C**' represents the cleavage site.
 - '**X**' region represents the sequence of protein which is left after the cleavage of the particular signal peptide.
- The cleavage site can also analyze through the spike of the green light which tells the position of the cleavage site.
- The signal peptide information of highly annotated proteins is also present on the NCBI genbank database. But in case of newly synthesized proteins you need to utilize the SignalP server or other tools to find out the signal peptide information.

Summary:

In this video, we have learnt about the SignalP server for the prediction of signal peptide within protein sequences. We also learned how to utilize this server to predict the signal peptide and how to analyze its results.