# CellPPD Manual

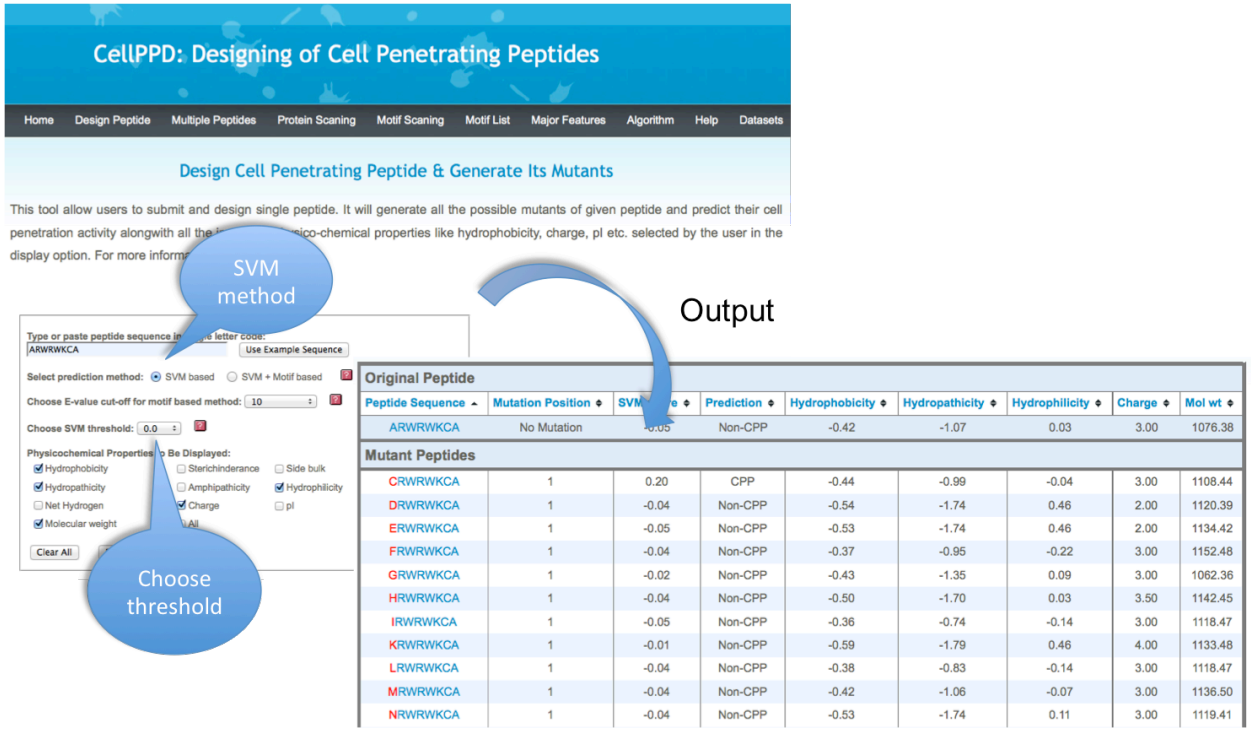
Designed at Dr. G.P.S Raghava group

CSIR – Institute of Microbial Technology

## 1. Design Peptide

This tool allows users to submit and design single cell penetrating peptide (CPP). It will generate all the possible mutants of given peptide and predict their cell penetration activity along with all the important physico-chemical properties e.g. hydrophobicity, charge, pI etc. selected by the user in the display option. It has two major options SVM based and SVM plus motif based.

**1.a.** SVM based method predicts the cell penetration efficiency on the basis SVM score, which uses binary profile of the peptide as input. User has to choose SVM threshold on the basis of which CPP and non-CPP will be classified (Figure 1).




| Peptide Sequence | Mutation Position | SVM Score | Prediction | Hydrophobicity | Hydropathicity | Hydrophilicity | Charge | Mol wt |
|---|---|---|---|---|---|---|---|---|
| ARWRWKCA | No Mutation | -0.05 | Non-CPP | -0.42 | -1.07 | 0.03 | 3.00 | 1076.38 |
| CRWRWKCA | 1 | 0.20 | CPP | -0.44 | -0.99 | -0.04 | 3.00 | 1108.44 |
| DRWRWKCA | 1 | -0.04 | Non-CPP | -0.54 | -1.74 | 0.46 | 2.00 | 1120.39 |
| ERWRWKCA | 1 | -0.05 | Non-CPP | -0.53 | -1.74 | 0.46 | 2.00 | 1134.42 |
| FRWRWKCA | 1 | -0.04 | Non-CPP | -0.37 | -0.95 | -0.22 | 3.00 | 1152.48 |
| GRWRWKCA | 1 | -0.02 | Non-CPP | -0.43 | -1.35 | 0.09 | 3.00 | 1062.36 |
| HRWRWKCA | 1 | -0.04 | Non-CPP | -0.50 | -1.70 | 0.03 | 3.50 | 1142.45 |
| IRWRWKCA | 1 | -0.05 | Non-CPP | -0.36 | -0.74 | -0.14 | 3.00 | 1118.47 |
| KRWRWKCA | 1 | -0.01 | Non-CPP | -0.59 | -1.79 | 0.46 | 4.00 | 1133.48 |
| LRWRWKCA | 1 | -0.04 | Non-CPP | -0.38 | -0.83 | -0.14 | 3.00 | 1118.47 |
| MRWRWKCA | 1 | -0.04 | Non-CPP | -0.42 | -1.06 | -0.07 | 3.00 | 1136.50 |
| NRWRWKCA | 1 | -0.04 | Non-CPP | -0.53 | -1.74 | 0.11 | 3.00 | 1119.41 |

**Figure 1.** Output of design peptide module by using SVM based method.

**1.b.** In SVM plus motif based method, motif information is used for the prediction of CPP and non-CPP. In this method, first motif is searched in the query peptide and that information is used for the further prediction (Figure 2). Here user has to choose E-value

cut off for the motif search in query sequence (lesser the cut off more will be the stringency).

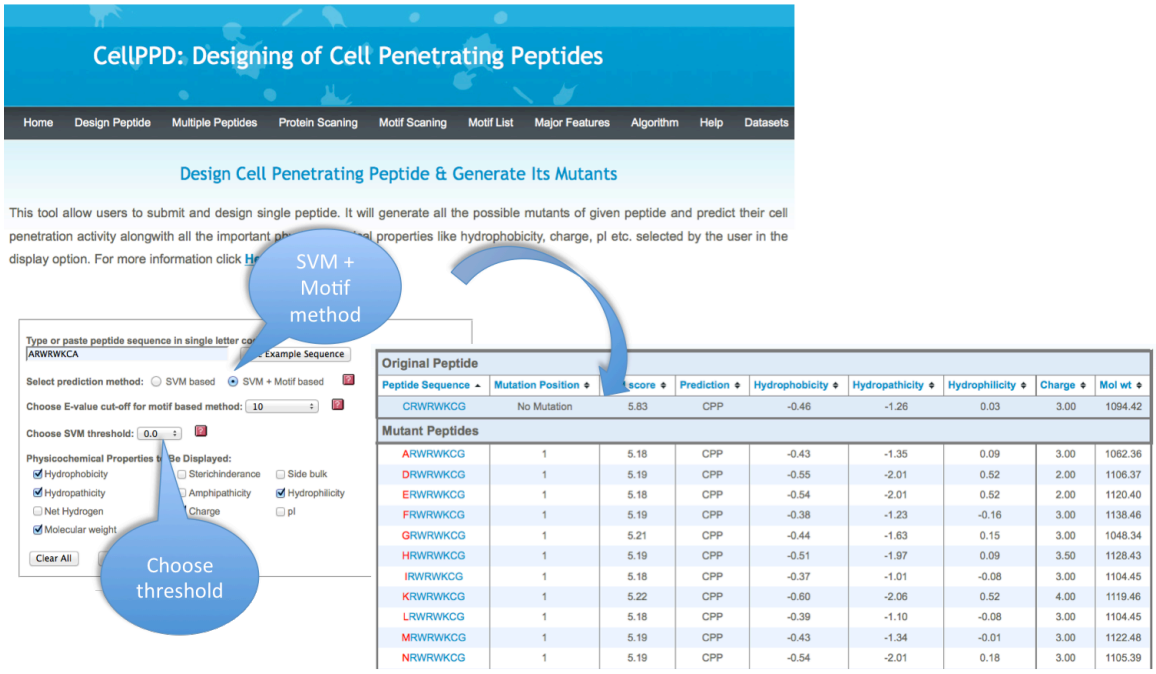



**Figure 2.** SVM plus motif based method to predict CPP

## 2. Multiple Peptides

This module allows the users to submit more than one peptide for the designing by either SVM or by SVM plus motif based method. Functioning and other requirements are same as the design peptides (Figure 3).

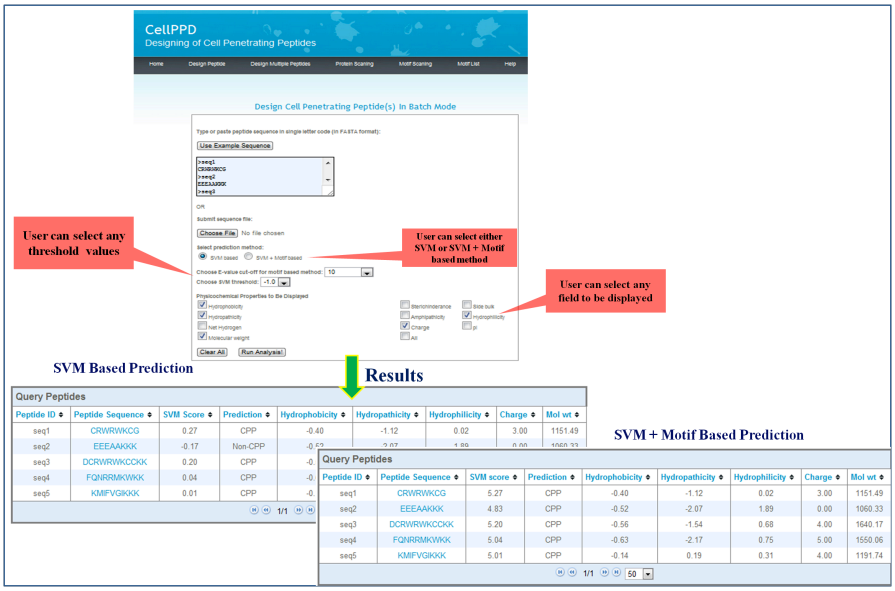


**Figure 3.** Multiple sequence designing.

## 3. Protein Scanning

Here user can dig out a protein sequence for the search of CPP peptide sequences within the protein sequence by either SVM or by SVM plus motif based method. Users have to select the length of the fragment peptide, so that only peptide with specific length will be generated (Figure 4).

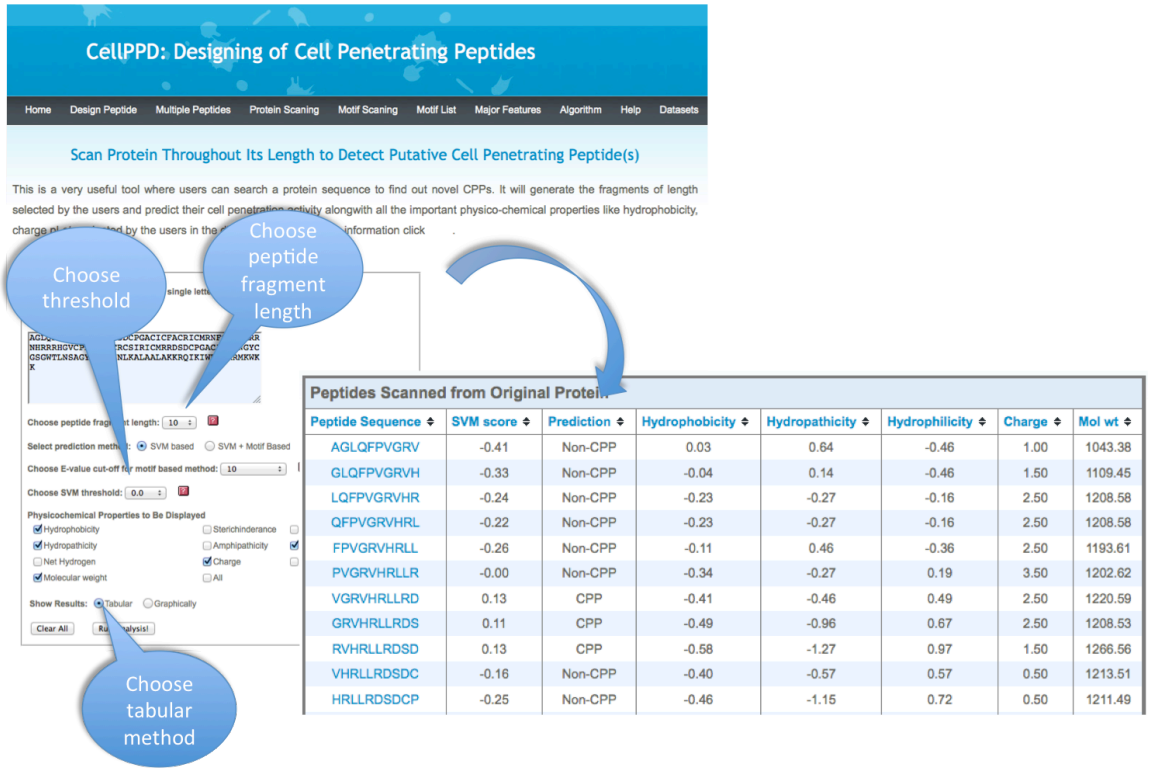



**Figure 4.** Protein scanning showing CPPs and non-CPPs in a protein sequence.

One major advantage of this module is to get the results in graphical format (Figure 5).

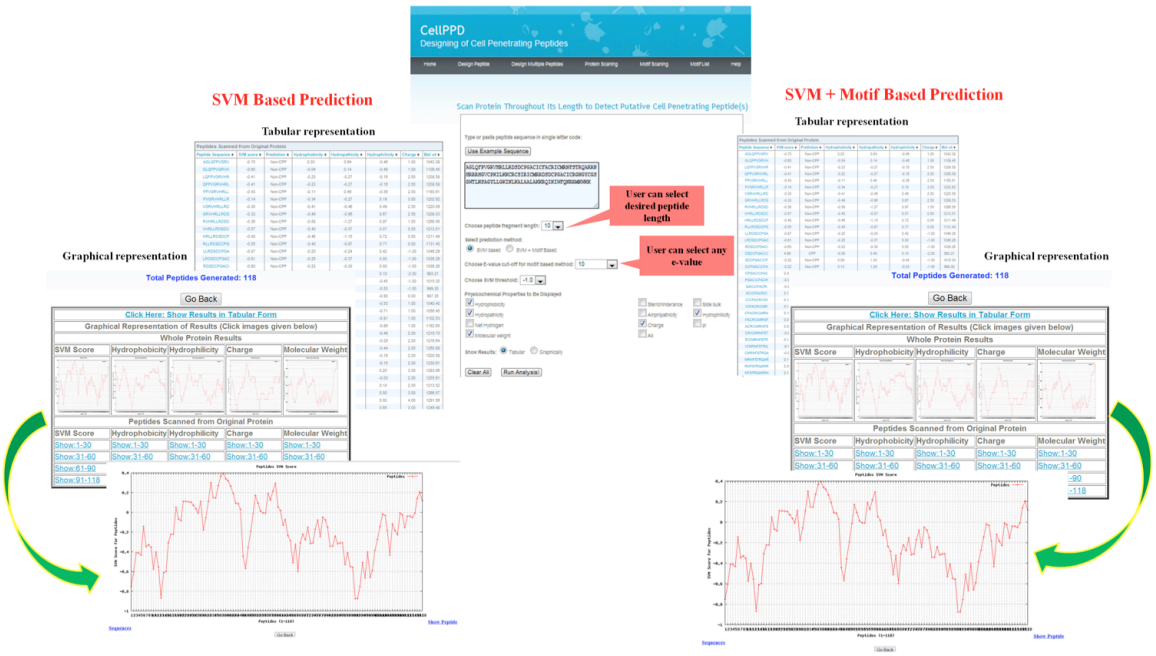



**Figure 5.** Protein scanning output in graphical format.

## 4. Motif Scanning

Here users can find out the CPP motif in any of the protein sequence given. These motifs are those motifs, which were found in the known CPPs (Figure 6).

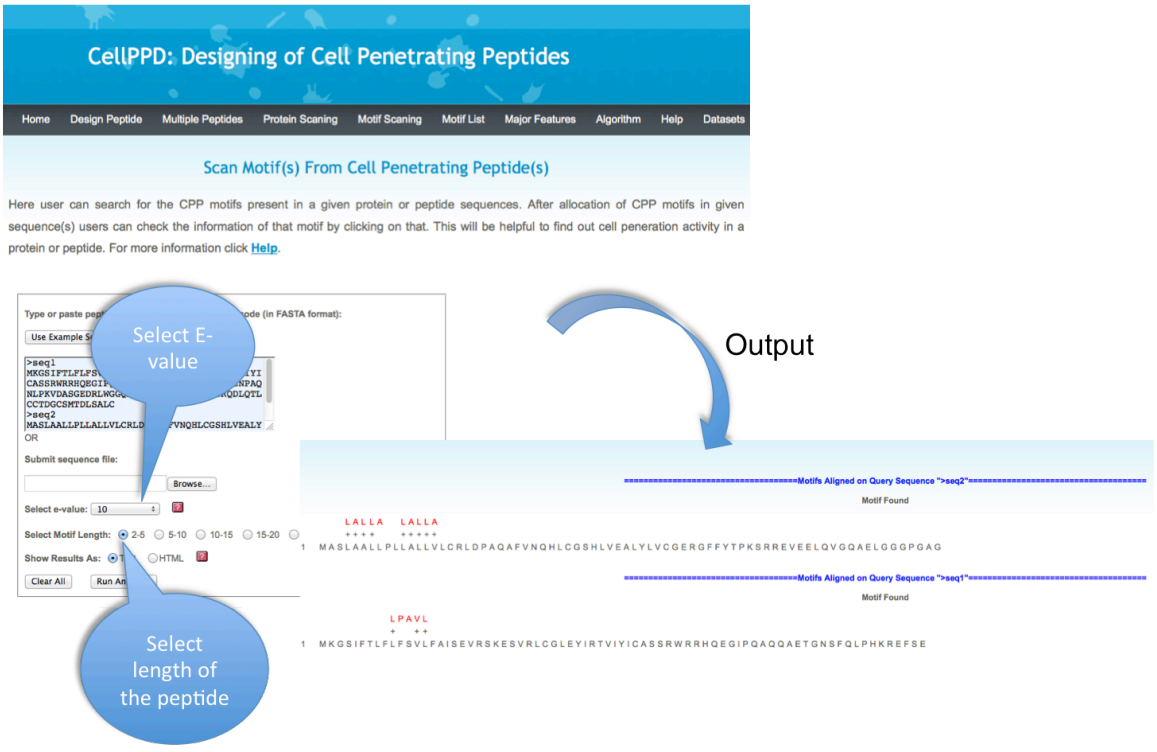


**Figure 6.** Motif scanning in a given protein sequence.

For this users have to choose the E-value cut-off, which will be required by the MAST program to find out the motif in given protein sequences. Users also have to choose the length of the motifs they want to search their protein sequences.

## 5. Motif List

Here we have provided all the CPP motifs present in the CPPs, which were obtained by the MEME suite (Figure 7).

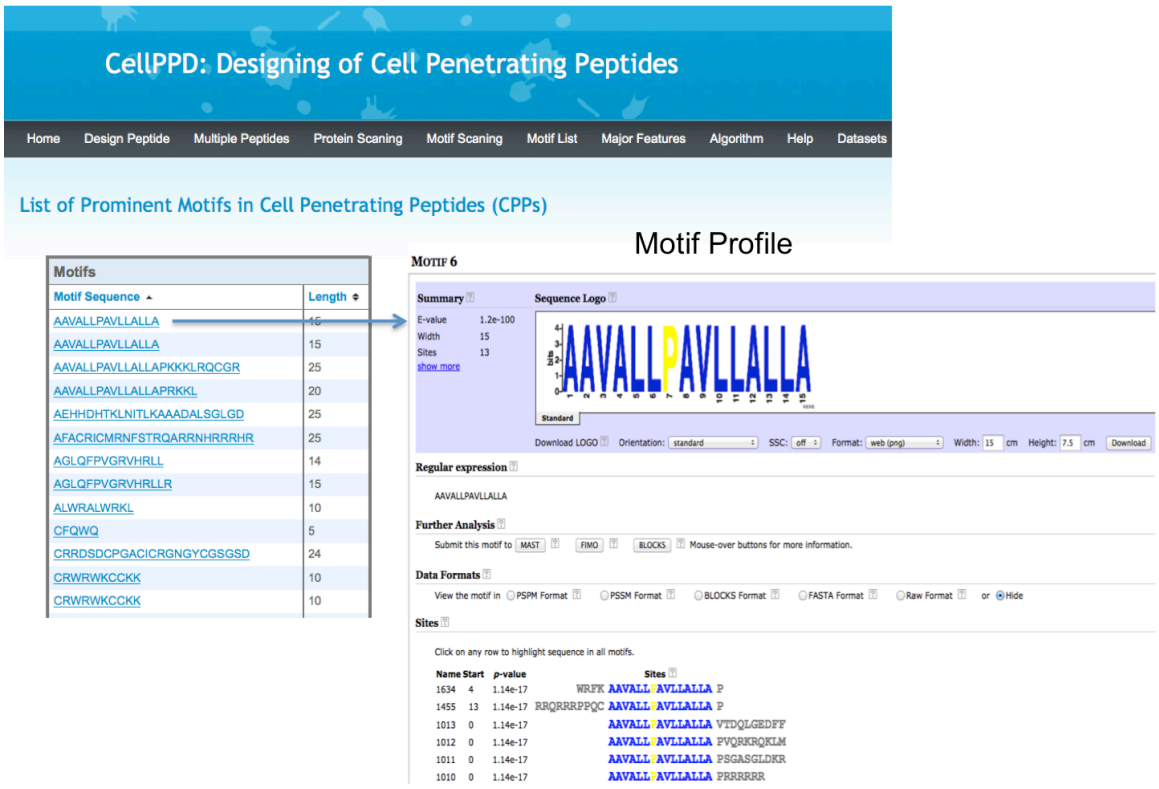



**Figure 7.** CPP Motif list and its description.

## 6. Case Study

CellPPD web server not only provides facility to predict peptides as CPP or non-CPP, but also it offers opportunity to design better CPP analogues. Besides prediction of a given peptide, CellPPD also generates all possible single substitution mutants of original peptide with their SVM scores and prediction status. Along with this, server also calculates important physicochemical properties (*e.g.* hydrophobicity, amphipathicity, charge pI, *etc*.) in an aesthetic tabular format with sorting option. This feature is helpful for user to select better CPP analogues having desired physicochemical properties, as many analogues may have better SVM score or better desired properties than the original peptide. In addition, user can further generates all possible mutants (2$^{nd}$ round) of their selected CPP if they wish to and may get the even better CPP analogues with more SVM score. This cycle (called CPP designing cycle, Figure 8) can be run until the desired CPP is obtained. One example is given below to explain designing of CPPs using CellPPD.

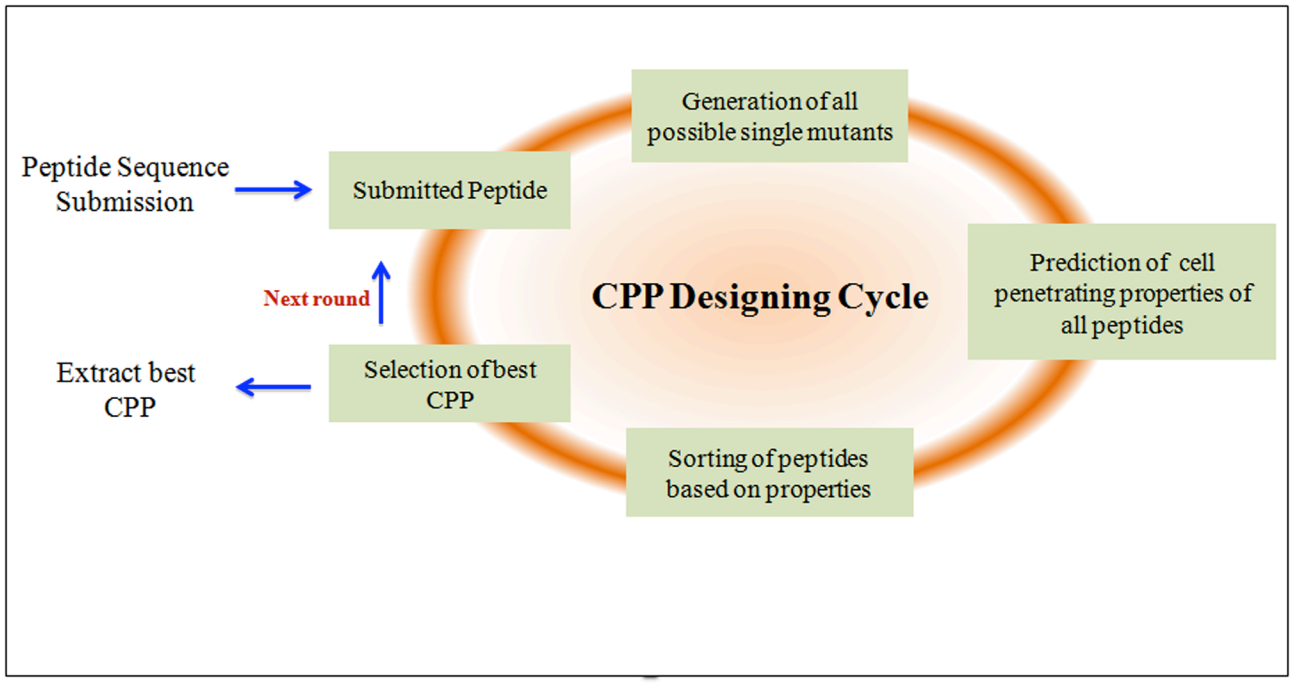



**Figure 8.** CPP Designing Cycle showing various steps for designing CPPs using CellPPD.

## Example: Designing of best CPP analogues based on a random sequence RRGIRLWSHLPRK

User can follow the below steps in order to design better CPP analogues.

### *Step 1. Submission of RRGIRLWSHLPRK*

Go to **"Design peptide"** tool and type the peptide sequence in single letter code as described in following Figure 9.

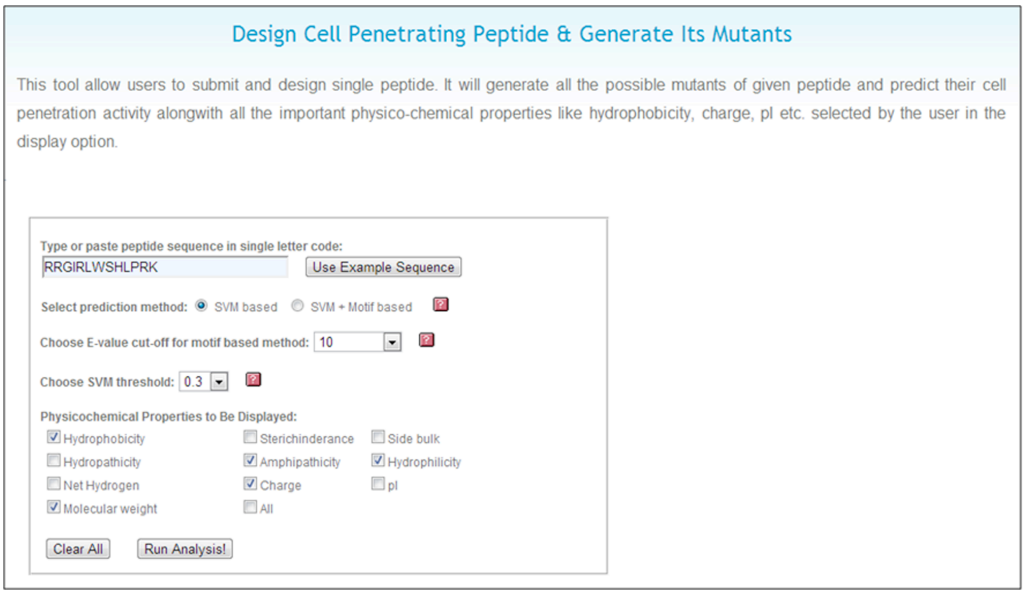



**Figure 9.** Submission page for design peptide tool.

There are two options for prediction, one is SVM based and other is SVM + motif based. User can select both options one by one as per the convenience. For the prediction, user has to select SVM threshold and E-value cut-off for SVM based and motif based method, respectively. As this server allows users to select a threshold, we suggest the users to select higher value, if they are interested in high specificity (high confidence). In addition, several physicochemical properties like hydrophobicity, amphipathicity, pI, charge, *etc*. can be selected to be displayed along with prediction status. For example, we choose SVM method for prediction with threshold 0.3.

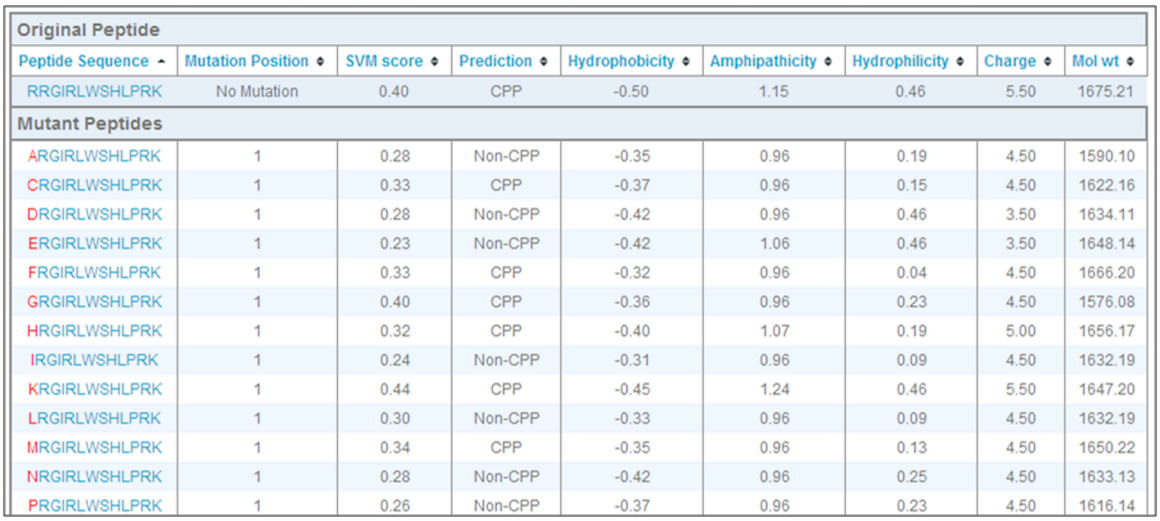
***Step 2. Prediction of submitted peptide and its all possible mutants.***

Server predicted the submitted peptide as CPP with SVM score 0.4 (Figure 10).

| Original Peptide | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Peptide Sequence | Mutation Position | SVM score | Prediction | Hydrophobicity | Amphipathicity | Hydrophilicity | Charge | Mol wt |
| RRGIRLWSHLPRK | No Mutation | 0.40 | CPP | -0.50 | 1.15 | 0.46 | 5.50 | 1675.21 |

**Figure 10.** Result of SVM based prediction.

In addition to prediction of original peptide, server also generates all possible single substitution mutants (depicted in red color) of the original peptide with their SVM score and prediction status (Figure 11). Various physicochemical properties have also displayed.



| Original Peptide | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Peptide Sequence | Mutation Position | SVM score | Prediction | Hydrophobicity | Amphipathicity | Hydrophilicity | Charge | Mol wt |
| RRGIRLWSHLPRK | No Mutation | 0.40 | CPP | -0.50 | 1.15 | 0.46 | 5.50 | 1675.21 |
| Mutant Peptides | | | | | | | | |
| ARGIRLWSHLPRK | 1 | 0.28 | Non-CPP | -0.35 | 0.96 | 0.19 | 4.50 | 1590.10 |
| CRGIRLWSHLPRK | 1 | 0.33 | CPP | -0.37 | 0.96 | 0.15 | 4.50 | 1622.16 |
| DRGIRLWSHLPRK | 1 | 0.28 | Non-CPP | -0.42 | 0.96 | 0.46 | 3.50 | 1634.11 |
| ERGIRLWSHLPRK | 1 | 0.23 | Non-CPP | -0.42 | 1.06 | 0.46 | 3.50 | 1648.14 |
| FRGIRLWSHLPRK | 1 | 0.33 | CPP | -0.32 | 0.96 | 0.04 | 4.50 | 1666.20 |
| GRGIRLWSHLPRK | 1 | 0.40 | CPP | -0.36 | 0.96 | 0.23 | 4.50 | 1576.08 |
| HRGIRLWSHLPRK | 1 | 0.32 | CPP | -0.40 | 1.07 | 0.19 | 5.00 | 1656.17 |
| IRGIRLWSHLPRK | 1 | 0.24 | Non-CPP | -0.31 | 0.96 | 0.09 | 4.50 | 1632.19 |
| KRGIRLWSHLPRK | 1 | 0.44 | CPP | -0.45 | 1.24 | 0.46 | 5.50 | 1647.20 |
| LRGIRLWSHLPRK | 1 | 0.30 | Non-CPP | -0.33 | 0.96 | 0.09 | 4.50 | 1632.19 |
| MRGIRLWSHLPRK | 1 | 0.34 | CPP | -0.35 | 0.96 | 0.13 | 4.50 | 1650.22 |
| NRGIRLWSHLPRK | 1 | 0.28 | Non-CPP | -0.42 | 0.96 | 0.25 | 4.50 | 1633.13 |
| PRGIRLWSHLPRK | 1 | 0.26 | Non-CPP | -0.37 | 0.96 | 0.23 | 4.50 | 1616.14 |

**Figure 11.** A screenshot of complete result of SVM-based prediction showing all possible mutants with their physicochemical properties, SVM score and prediction status.

***Step 3. Selection of best CPP with desired properties.***

In CellPPD, sorting options for all the properties have been provided. User can sort peptide analogues having desired SVM score and physicochemical properties. For example, we have sorted analogues according to their SVM scores to select CPPs with highest SVM score (Figure 12). After sorting, analogue RRGRRLWSHLPRK displayed

the highest SVM score (0.63) amongst all analogues and original peptide. Similarly, user can sort other properties as well.

| Original Peptide | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Peptide Sequence | Mutation Position | SVM score | Prediction | Hydrophobicity | Amphipathicity | Hydrophilicity | Charge | Mol wt |
| RRGIRLWSHLPRK | No Mutation | 0.40 | CPP | -0.50 | 1.15 | 0.46 | 5.50 | 1675.21 |
| **Mutant Peptides** | | | | | | | | |
| RRGRRLWSHLPRK | 4 | 0.63 | CPP | -0.70 | 1.34 | 0.83 | 6.50 | 1718.23 |
| RRGIRLWRHLPRK | 8 | 0.62 | CPP | -0.62 | 1.34 | 0.67 | 6.50 | 1744.32 |
| RRGIRLWSHRPRK | 10 | 0.62 | CPP | -0.68 | 1.34 | 0.83 | 6.50 | 1718.23 |
| RRGKRLWSHLPRK | 4 | 0.59 | CPP | -0.64 | 1.43 | 0.83 | 6.50 | 1690.22 |
| RRGIRLRSHLPRK | 7 | 0.58 | CPP | -0.67 | 1.34 | 0.95 | 6.50 | 1645.18 |
| RRGIRLWSHKPRK | 10 | 0.58 | CPP | -0.63 | 1.43 | 0.83 | 6.50 | 1690.22 |
| RRGIRRWSHLPRK | 6 | 0.57 | CPP | -0.68 | 1.34 | 0.83 | 6.50 | 1718.23 |
| RRRIRLWSHLPRK | 3 | 0.56 | CPP | -0.65 | 1.34 | 0.69 | 6.50 | 1774.34 |
| RRGIRLWSRLPRK | 9 | 0.56 | CPP | -0.61 | 1.22 | 0.73 | 6.00 | 1694.25 |
| RRGIRLWKHLPRK | 8 | 0.55 | CPP | -0.57 | 1.43 | 0.67 | 6.50 | 1716.31 |
| RRGIRLWLHLPRK | 8 | 0.54 | CPP | -0.44 | 1.15 | 0.30 | 5.50 | 1701.30 |
| RRGIRLWSKLPRK | 9 | 0.53 | CPP | -0.56 | 1.32 | 0.73 | 6.00 | 1666.24 |
| RRGIRLWWHLPRK | 8 | 0.52 | CPP | -0.46 | 1.15 | 0.18 | 5.50 | 1774.35 |
| RRGWRLWSHLPRK | 4 | 0.51 | CPP | -0.53 | 1.15 | 0.34 | 5.50 | 1748.26 |

**Figure 12.** Sorting of results obtained in step 2.

***Step 4. Generation of further all possible mutants of RRGRRLWSHLPRK with prediction status.***

User can further generate all possible mutants of their desired analogues obtained in step 3 by clicking on the peptide. For example, we have selected and clicked on RRGRRLWSHLPRK analogue and server generated further all possible mutants of the RRGRRLWSHLPRK with SVM score, prediction status and all physicochemical properties (Figure 13). User can again sort the obtained results and select best CPP analogue. This cycle can be run until desired sequence is obtained.

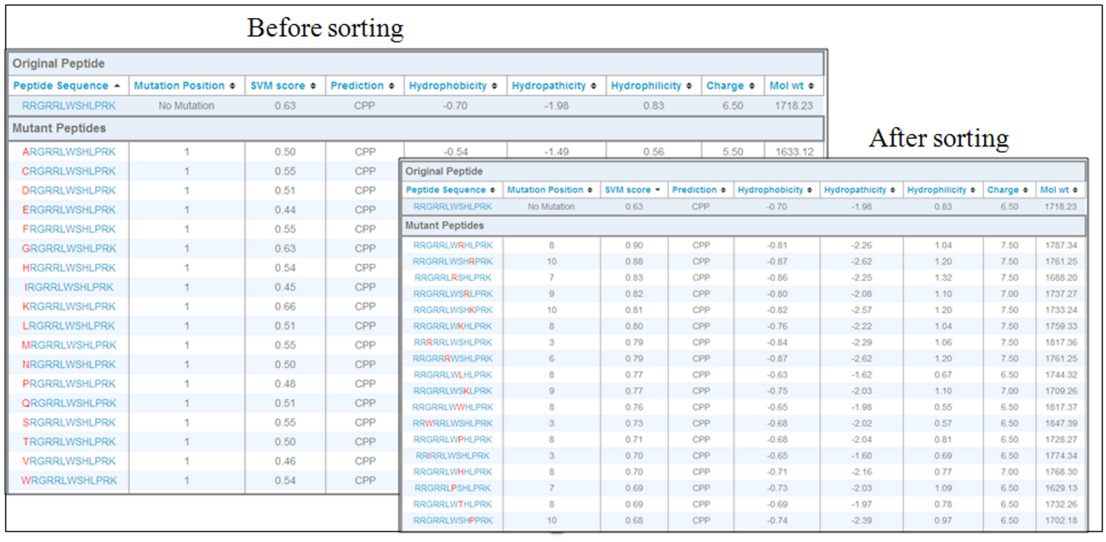


**Figure 13.** All possible mutants with their SVM score and prediction status of RRGRRLWSHLPRK peptide.

## Example: Designing of CPPs from a protein sequence.

Since most of the existing CPPs are derived from natural proteins. CellPPD provides facility to identify potential CPPs from a protein sequence. A tool **protein scanning** has been implemented to web server for the detection of putative CPPs in a protein sequence. In this tool, after submission of a query protein sequence, server first generates overlapping peptides of window length selected by the user, where all the peptides will be clickable. All peptides are then predicted by the server and presented in tabular format with their SVM score and prediction status. Next, CPP designing cycle can be used for further generating the mutants with prediction status as described earlier. The overall approach is demonstrated in Figure 14.

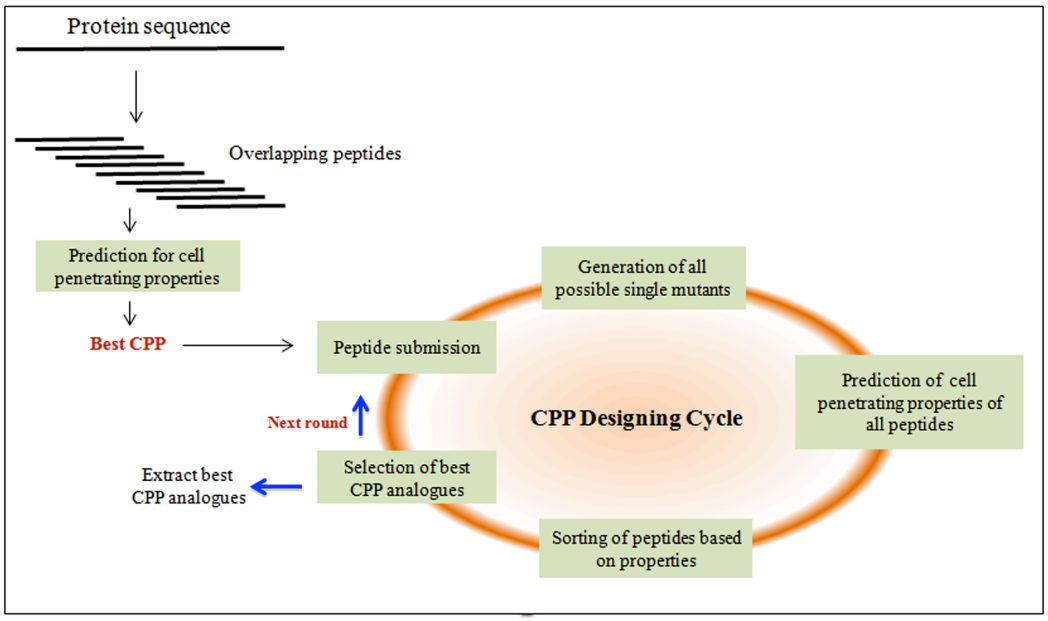



**Figure 14.** Designing of CPP by protein scanning tool.

For example, we wish to identify putative CPP sequences in the following protein sequence

AGLQFPVGRVHRLLRDSDCPGACICFACRICMRNFSTRQARRNHRRRHGVCPKIL
KKCRCSIRICMRRDSDCPGACICRGNGYCGSGWTLNSAGYLLGKINLKALAALA
KKRQIKIWFQNRRMKWKK

An option for selection of peptide length has provided. We have chosen the peptide length 10 and SVM based method with threshold 0.7 for the prediction of CPPs (Figure 15). User can select either options (SVM based or Motif based) for the prediction.

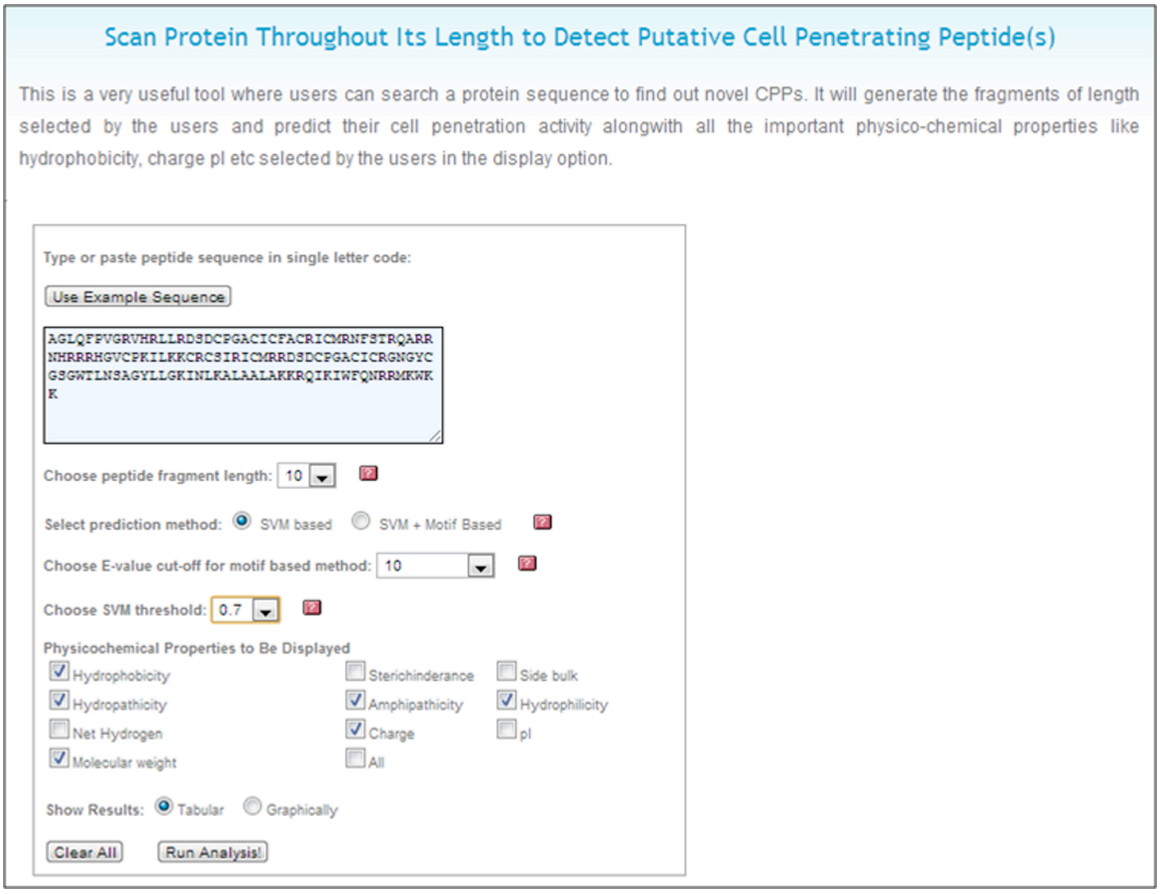



**Figure 15.** Submission page of protein scanning page.

Server generated overlapping peptides of window length 10 with their SVM score and prediction status, where all the peptides are clickable (Figure 16).

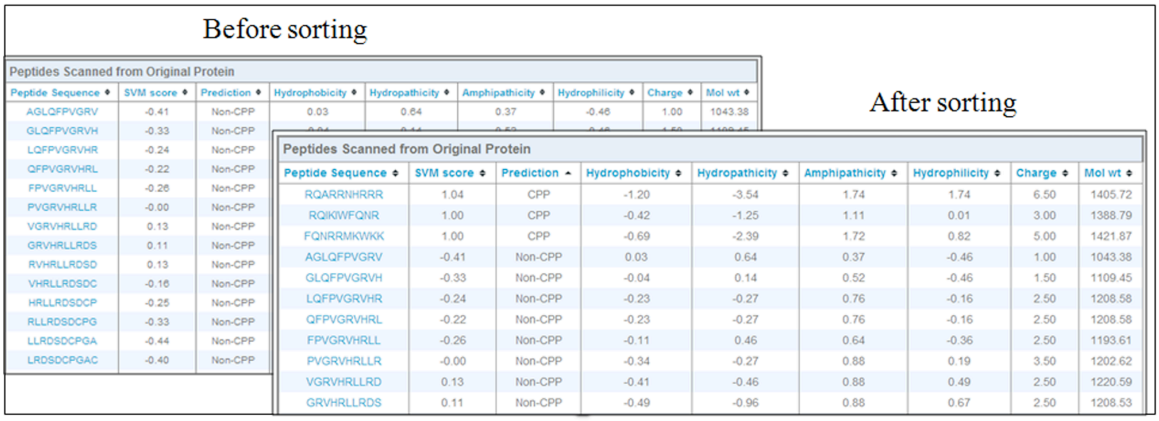



**Figure 16.** Screenshot of SVM based result of submitted protein sequence showing all possible peptides (length10) with their physicochemical properties, SVM scores and prediction status.

In protein scanning, user can also obtain their results in graphical format (Figure 17), where all the values *e.g.* SVM and other physicochemical properties can be plotted using this module.

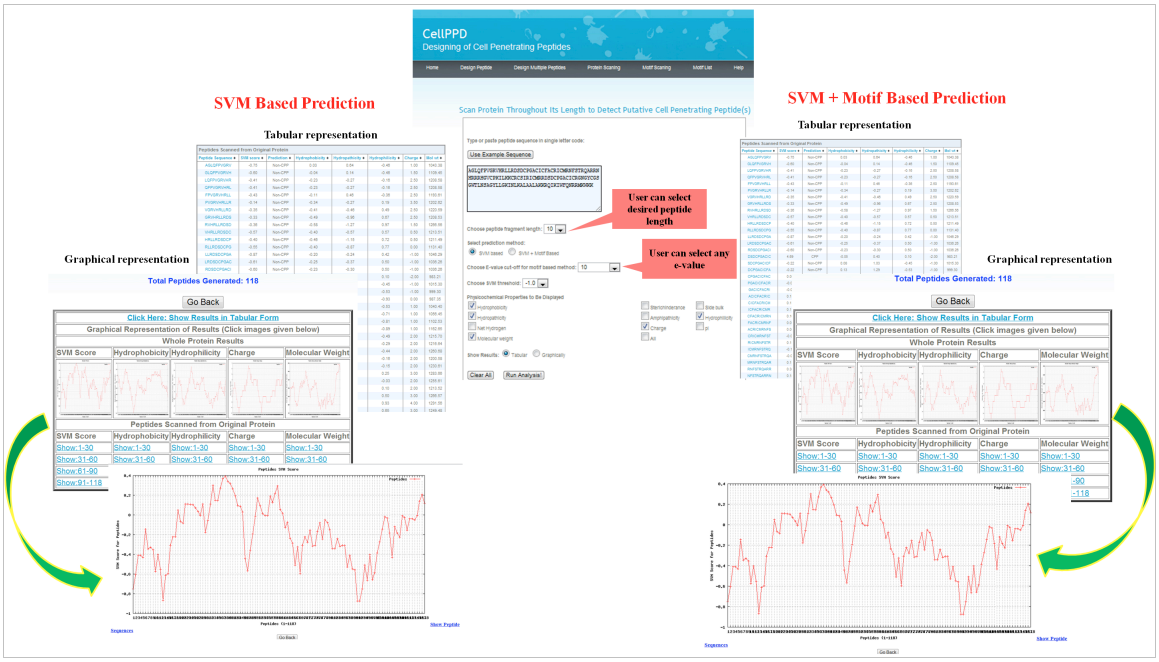



**Figure 17**. Schematic representation of SVM results in graphical format.