

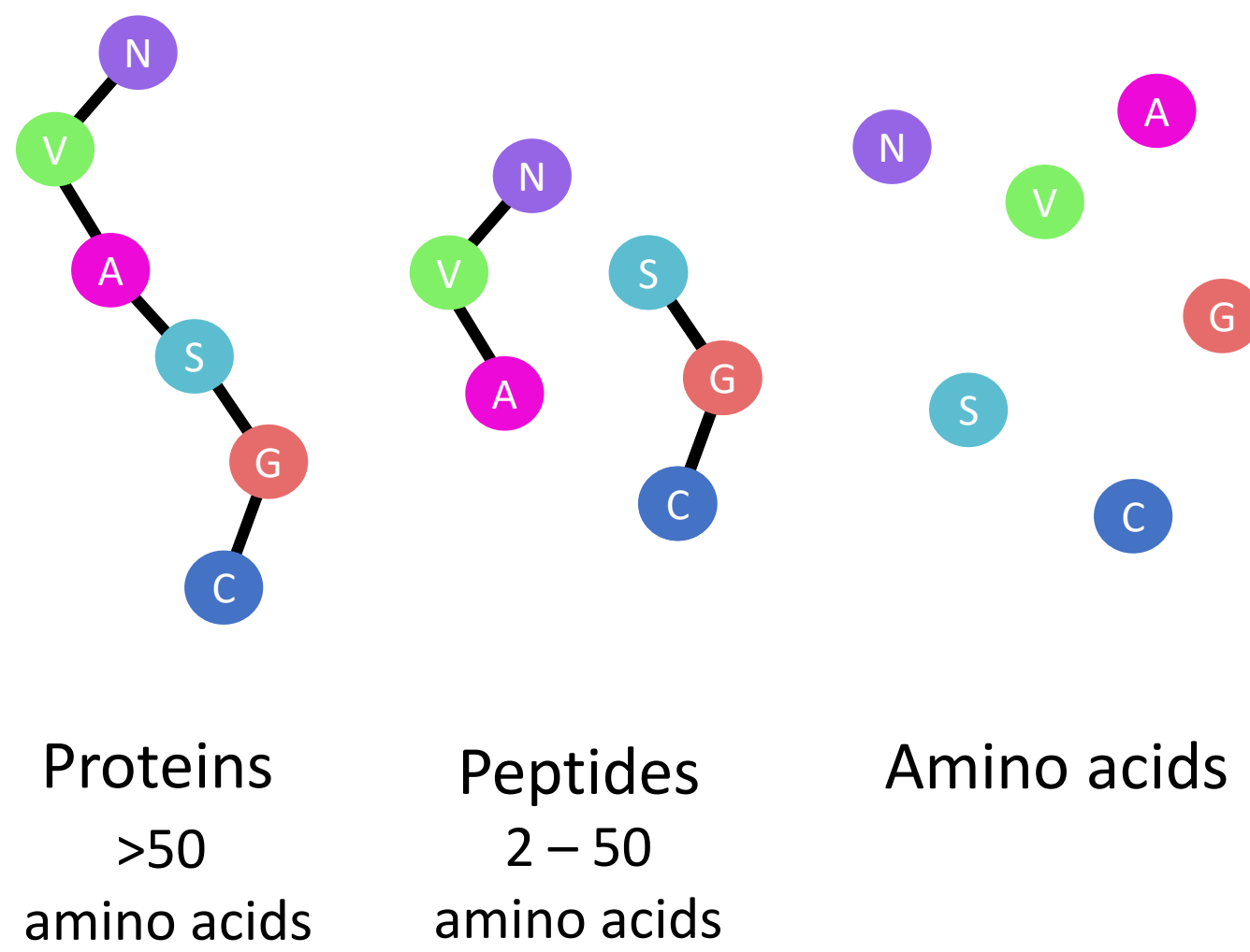
MutationDetector[®]. Software tool for detecting amino acid substitutions

Student: Brilliantov K., St. Petersburg Academic Lyceum Physical Technical High School, Russia
Supervisor: Vyatkina K., St. Petersburg Academic University, Russia



Introduction

Proteins and Peptides



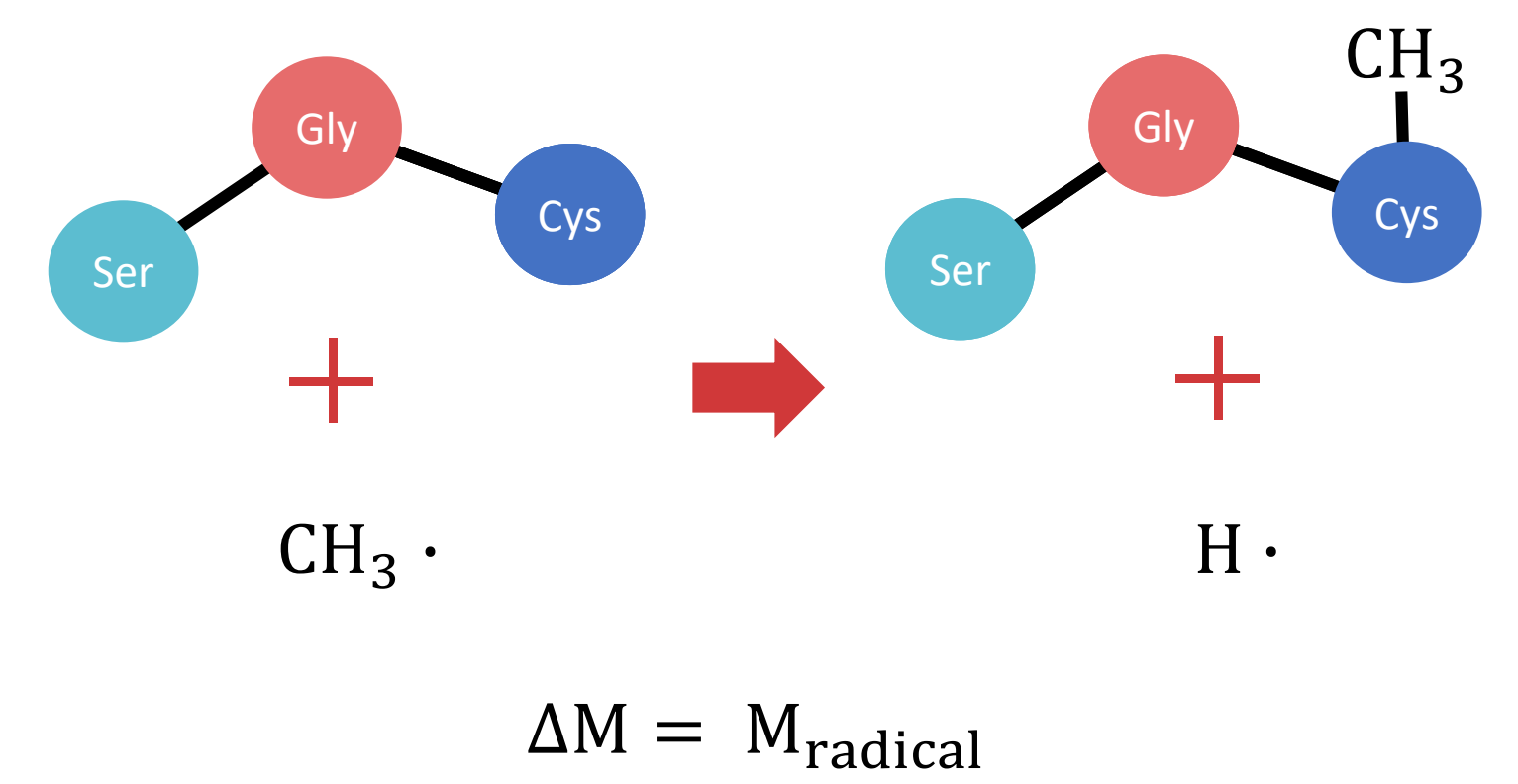
Single nucleotide polymorphism (SNP)

	Wild type	Variant type (A->T)
DNA strand	ACC AAA CCG AGT	ACC ATA CCG AGT
mRNA	UGG UUU GGC UCA	UGG UAU GGC UCA
Protein	-Trp-Phe-Gly-Ser-	-Trp-Tyr-Gly-Ser-

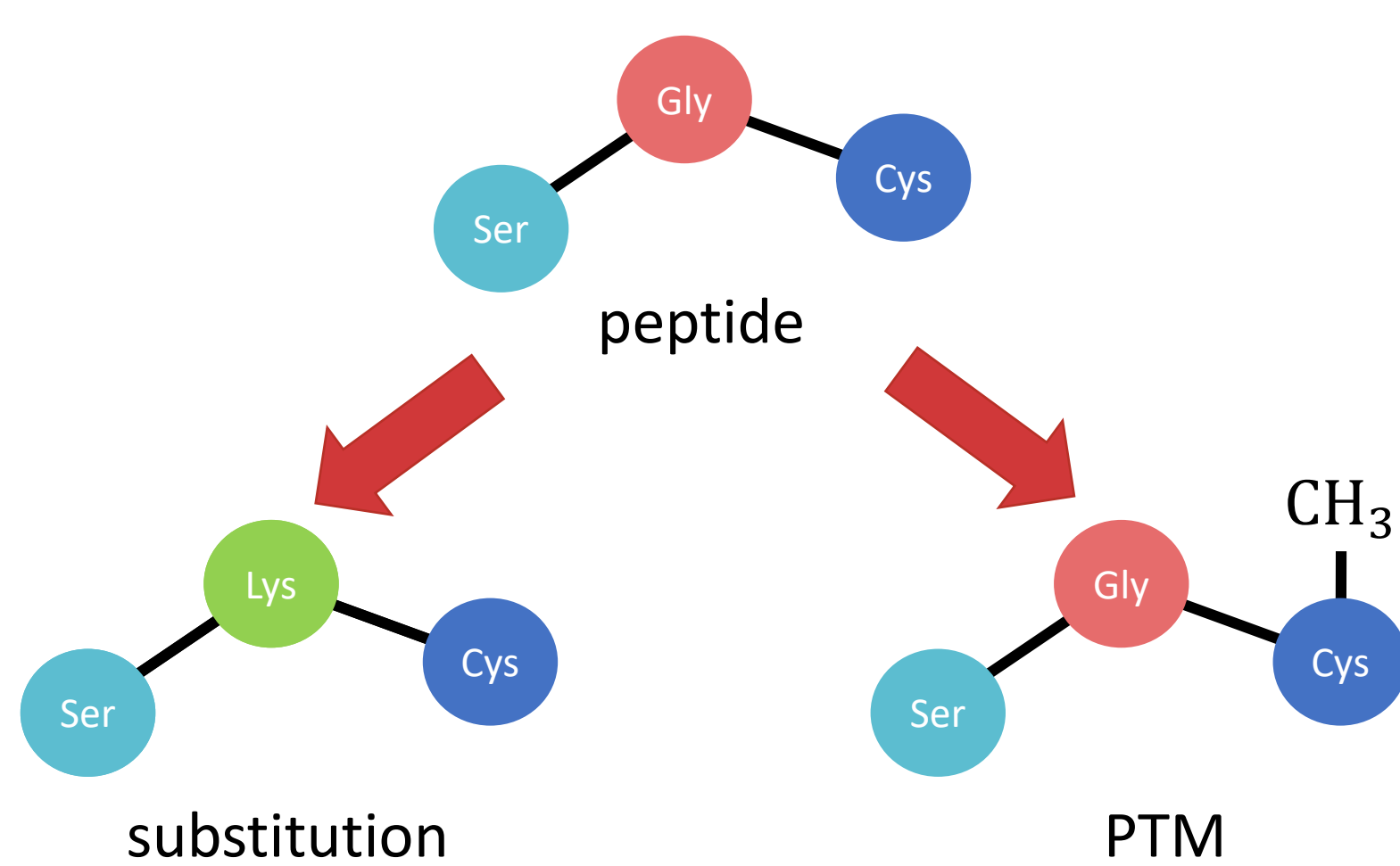
$$\Delta M = M_{\text{variant type}} - M_{\text{wild type}} = M_{\text{Tyr}} - M_{\text{Phe}}$$

Since three consequent nucleotides in a DNA strand, together forming a codon, encode an amino acid, SNP can lead to an amino acid substitution, thereby implying a change of the mass of the protein.

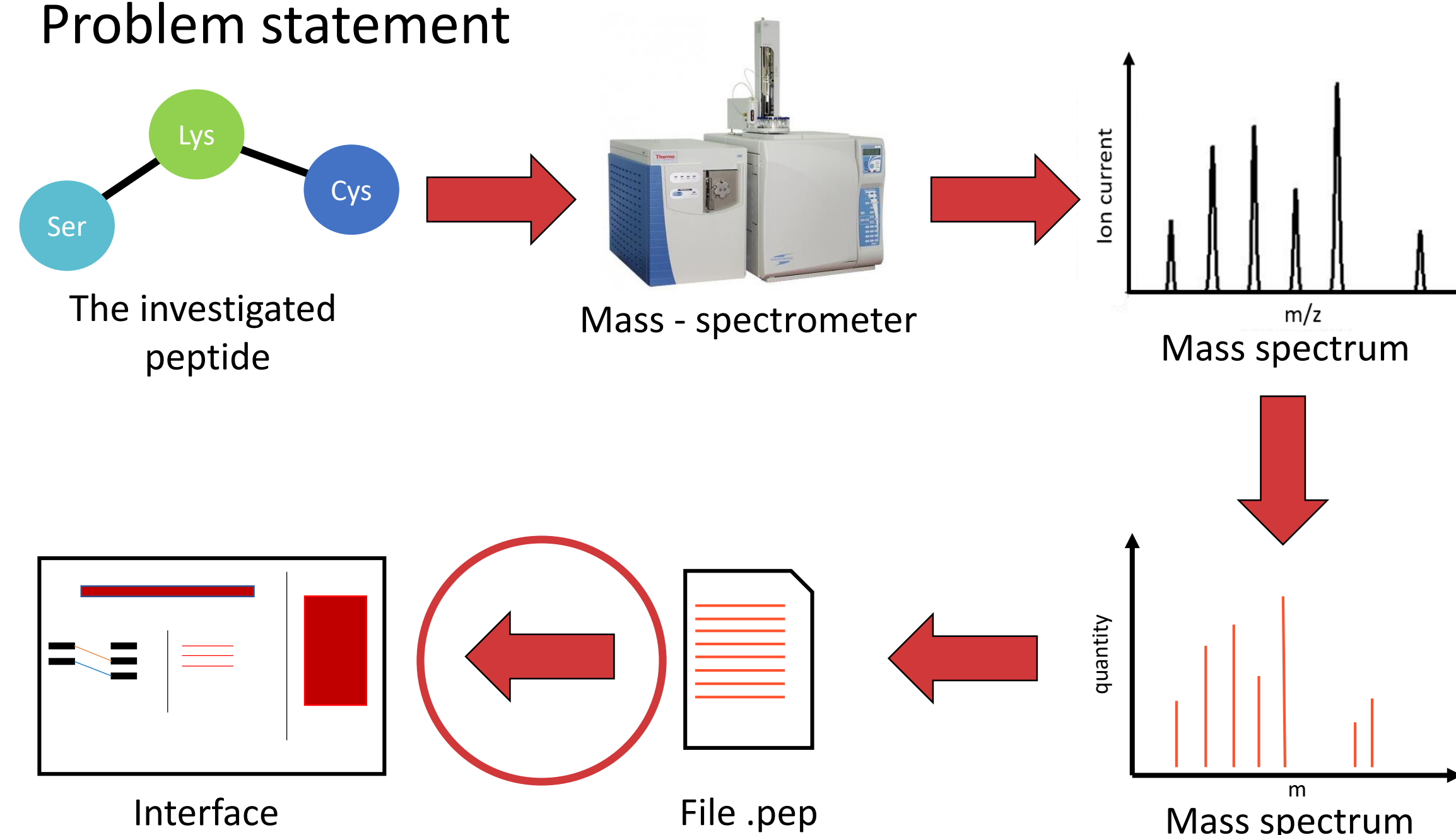
Post translational modifications (PTMs)



Possible modifications



Problem statement



The Goal

To develop a software tool for analysing data, obtained from modified peptide

The sub-tasks

- Learn how to read file .pep
- Learn how to find positions in the peptide where a substitution or a PTM could occur

Methodology

Programming tools



Input

The developed interface as an input takes a .pep file. You can see an example of such file.

YASSVRSHPHAIQLQAPQPA
VHVQGQEPLTASMLAAAPP
SMLA prefix
4653.40152Da

File .pep

Output

The developed program gives to user as an output: first, the peptide with a highlighted positions, secondly, a .pepout file. You can see an example of this file.

V Q S S S A R C V N G N

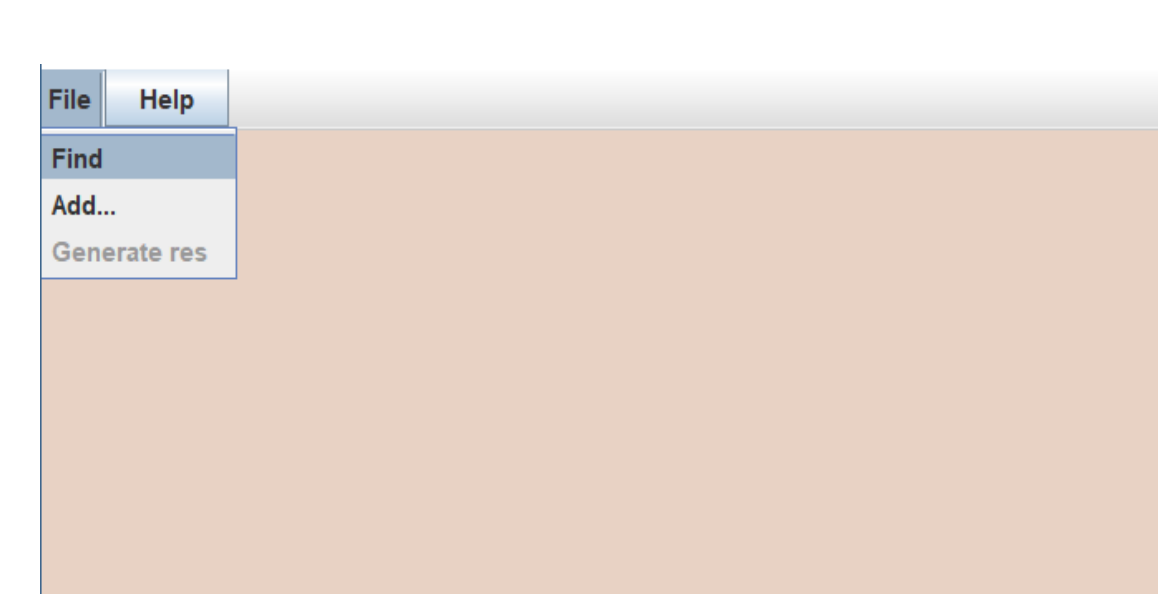
- PTM
- both
- substitution

YASSV⁵RSPHP¹⁰AIQ
PL¹⁵QAPQ P²⁰AV
Prefix selected
Possible modifications:
at positions 11, 17, 21 A->V
at positions 13, 16, 19: Q->R

File .pepout

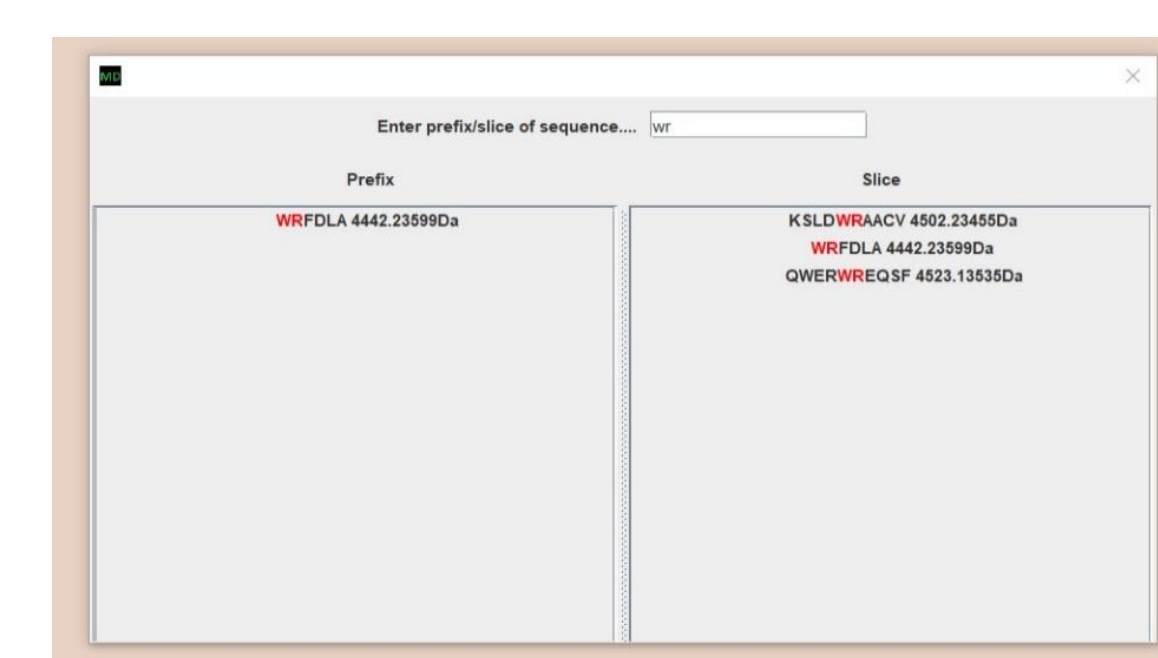
Interface

1. "File" tab



In the top right corner there is a pull-down menu. There are three items: Find, Add, Generate res. Item "Find" allows user to search for peptides. Add allows user to add .pep files. "Generate res" allows user to generate .pepout files

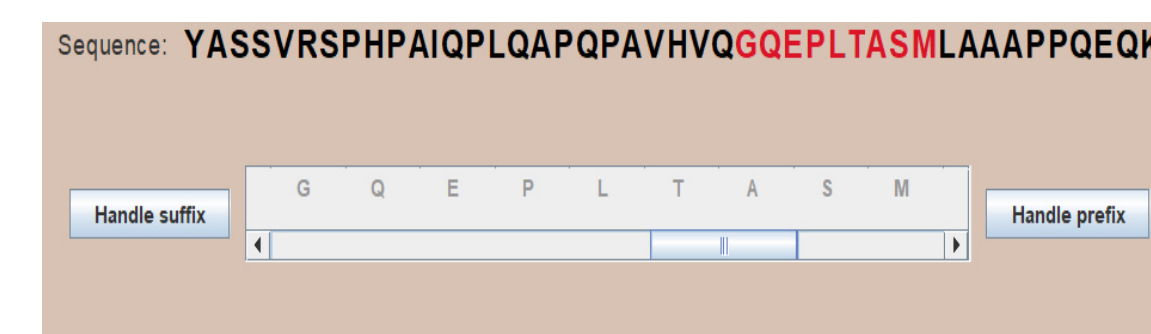
2. Searching for peptides



If a user clicks a "find" item, the dialog window appears. Then when user inputs Some sequence of the amino acids, the peptides where this fragment is a suffix appears on the left part of this window and,

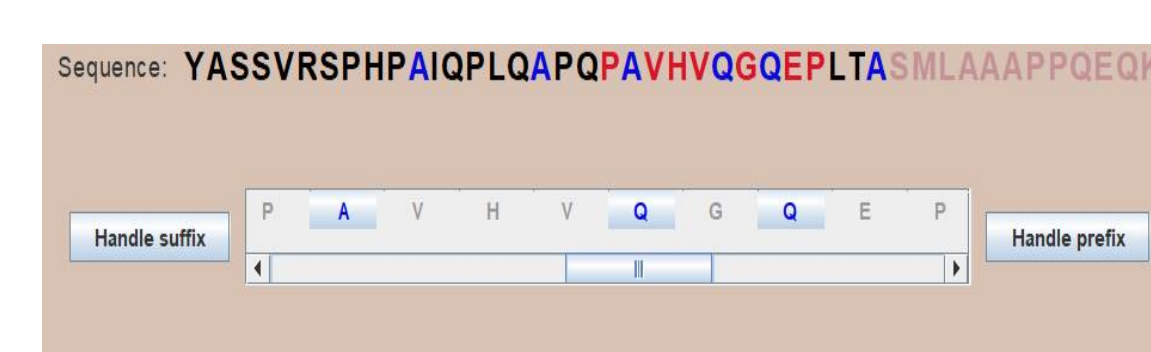
the peptides which just contain this fragment appears on the right part.

3. Sequence and scrollable panel



When user clicks on some of the peptides which were in that window, the main frame appears. In the top there is a chosen sequence of amino acids. Just below there is a scrollable panel. There are buttons symbolizing amino acids.

4. Highlighted positions



Also there are buttons "Handle prefix" and "Handle suffix". When a user clicks on some of these buttons, user can then click on one of the amino acids (buttons in the scrollable panel). After that the algorithm begins.

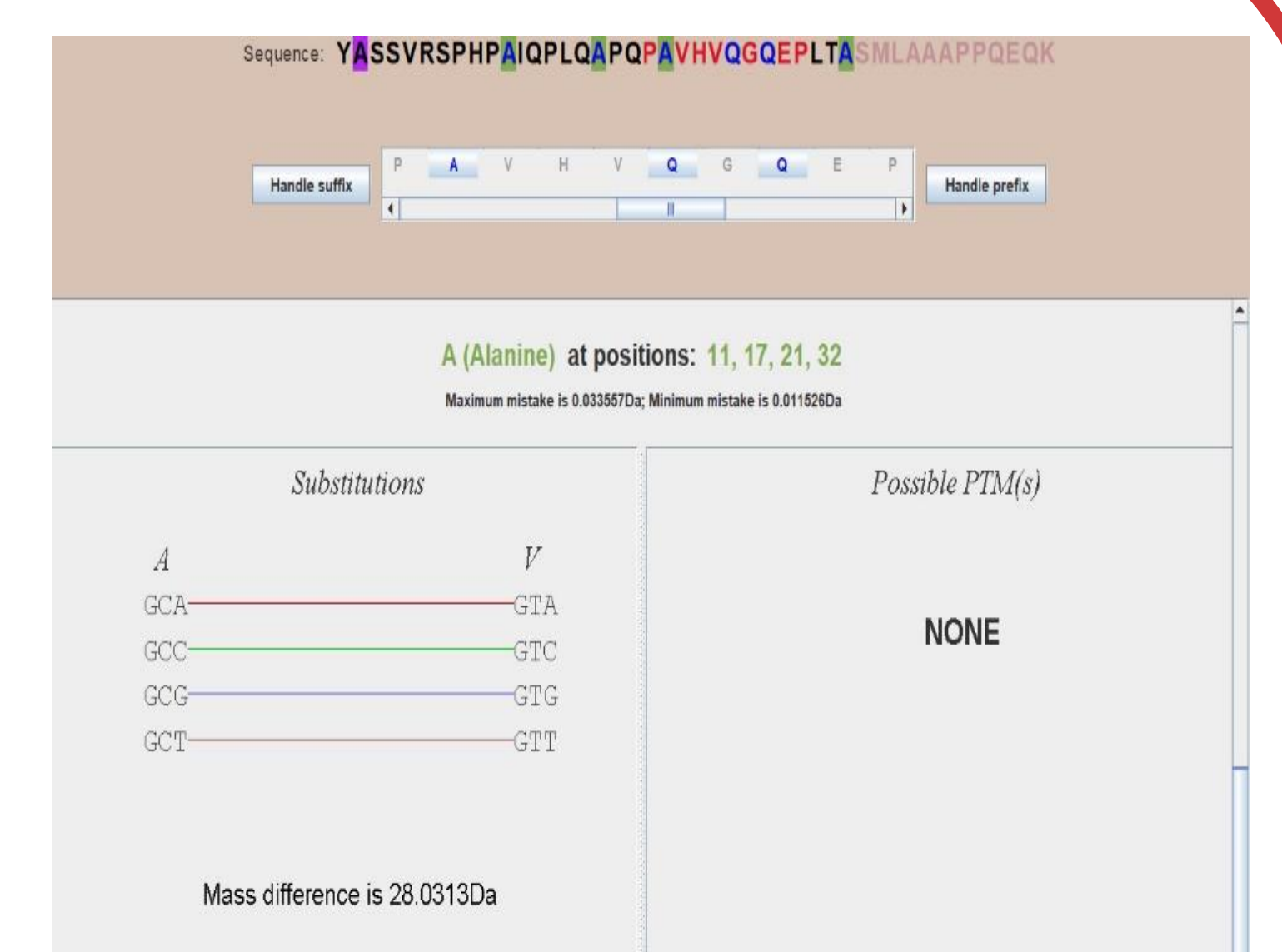
5. The algorithm

For each position in the suffix or in the prefix

$$\text{Count the error}$$
$$\text{If prefix: } \varepsilon = \frac{M_{\text{appm}}}{10^6}$$
$$\text{if suffix } \varepsilon = \frac{M_0 + M_{\text{appm}}}{10^6}$$

$$\text{Check if there is a sub or a PTM, with } \Delta m \in (\Delta M - \varepsilon; \Delta M + \varepsilon)$$

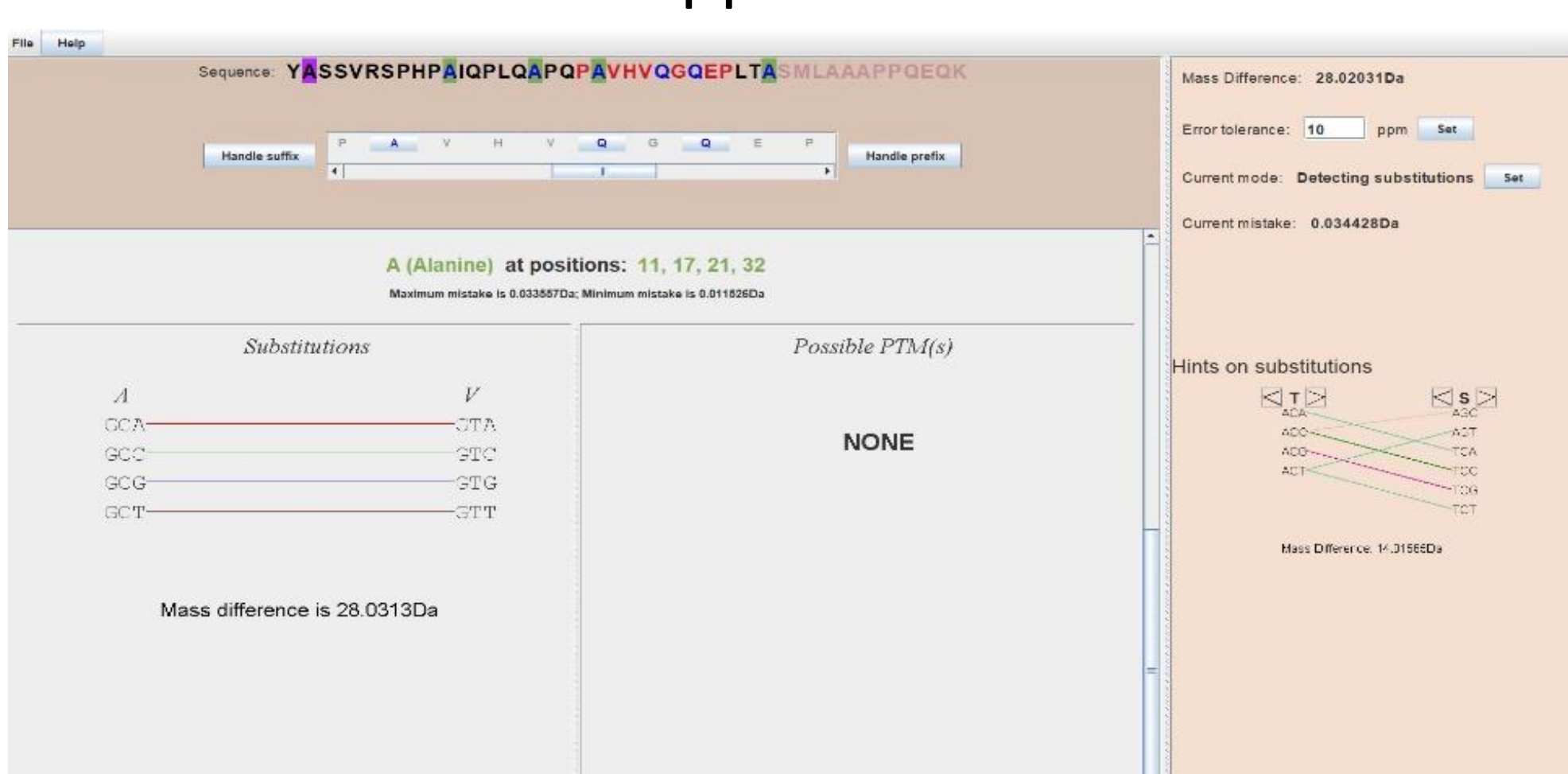
6. Information about substitutions



When a user clicks on one of the "A"s in the top sequence, first of these "A"s appears on the purple background since the remaining appear on pale green background, because first "A" does not fit the conditions since the mistake is not enough big (mass of prefix is not big enough). The substitution which could occur in those positions is A >> V some of the codons (encoding these amino acids) are connected with colorful lines. They are connected because the only difference between them is one nucleotide. For example, codon GCA encode A, GTA encode V, the difference between them is that on the second position there are different nucleotides. If SNP in which nucleotide C substitutes with nucleotide T occurs, a substitution A>>V occurs.

Results and conclusions

General view of an application



Additional functionality

- Tab help
- Hot keys
- Hints on substitutions

Testing

First peptide
YASSVRSHPHAIQLQAPQPAVHVQGQEPLTASMLAA
APPQEQK
In this peptide A>>V and Q>>R occurred (in the prefix).
Second peptide
EAATQEDPEQVPELAAHEVSASEAERPVAAEEILL
In this peptide A>>V occurred (in the suffix)

Conclusion

During this work, the software tool for analyzing data obtained from modified peptide has been developed. It has been tested, it works correctly. In the future, we intend to extend the functionality of MutationDetector in various ways thereby adapting it to solving special problems.

References

- B. Lewin. *Cells*. БИНОМ Russia, 2011. 951 c.
- S. Nie, H. Yin, Z. Tan, M. A. Anderson, M. T. Ruffin, D. M. Simeone, D. M. Lubman. *Quantitative Analysis of Single Amino Acid Variant Peptides Associated with Pancreatic Cancer in Serum by an Isobaric Labeling Quantitative Method*. J Proteome Res. 2014, 13(12):6058–6066.
- K. Vyatkina, S. Wu, L. J. M. Dekker, M. M. VanDuijn, X. Liu, N. Tolic, M. Dvorkin, S. Alexandrova, T. M. Luiders, L. Pasa-Tolic, P. A. Pevzner. *De Novo Sequencing of Peptides from Top-Down Tandem Mass Spectra*. J Proteome Res. 2015, 14(11):4450–4462.
- Qisheng Peng, Zijian Wang, Donglin Wu, Xiaou Li, Xiaofeng Liu, Wanchun Sun, Ning Liu. *Identification of single amino acid substitutions (SAAS) in neuraminidase from influenza A virus (H1N1) via mass spectrometry analysis coupled with de novo peptide sequencing*. Rapid Commun. Mass Spectrom. 2016