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

Motivation

Proteins play an essential role in our lives, because they are regulating sub-cellar processes[1]. If any disruption occurs, a protein will cease acting properly and may cause severe diseases. Many factors can result in such disruptions. We consider the most important of those – a single nucleotide polymorphism (SNP). Having learned to find positions in protein sequence where the substitution might occur, we would get a possibility to identify the so-called variant proteins, which can be biomarkers for a variety of hard diseases[2].

Single nucleotide polymorphism (SNP)

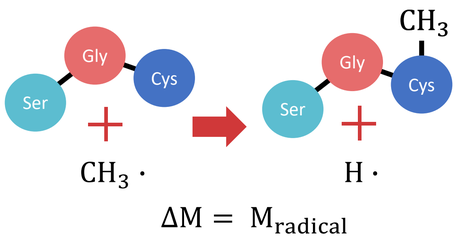
SNP is a substitution of nucleotide in a DNA strand. Since codons encode amino acid, SNP can lead to substitution of amino acid

	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-

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

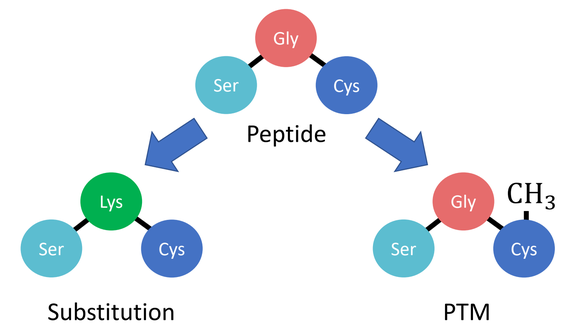
Post translational modifications (PTMs)

PTM is a change of amino acid's chemical composition through adherence of some chemical radicals.
(ex. the mass of methionine increases by approximately 14Da upon methylation)



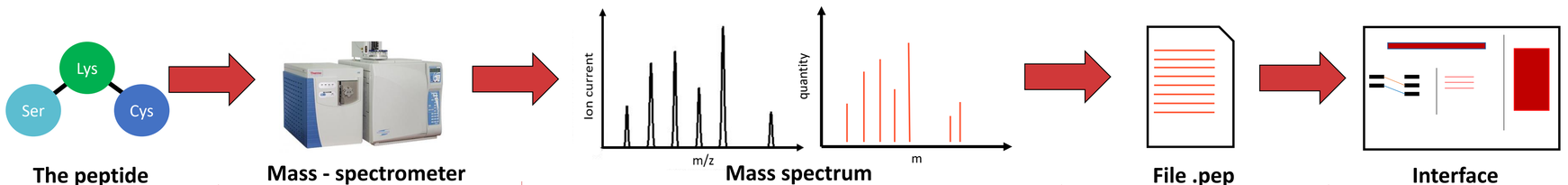
$$\Delta M = M_{\text{radical}}$$

Possible modifications



Problem statement

There are two different methods for analyzing such peptides. First is qualitive and second is quantitative. Our work was concentrated on qualitive analyze. Here is the scheme of such analyze



The peptide
The dissolved peptide is being bombarded with charged particles, so ideally each molecule divides in two parts.

Mass - spectrometer
After that, these fragments go through mass - spectrometer and this device gives us the mass spectrum.

Mass spectrum
Mass spectrum is a histogram, where X – axis is a mass of an ionized fragment, Y-axis is the ion current. Accordingly, this graph is a set of peaks. Based on this set we can establish amino acid's sequence of the investigated peptide. The with usage of algorithm Twister, we can generate a file.

File .pep
File has an information about the investigated peptide: amino acid sequence, fragment which was not modified and mass difference

Interface
To analyze it manually is uncomfortable. So, the idea about developing a software tool for analyzing such data arises.

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
- Learn how to generate a result file

Methodology

Programming tools

For developing this interface I used Java for a basic programming language. And an additional library for creating graphical interfaces, named, Swing

Input

Structure of .pep file

General	Example
<peptide>	…LTASMLAA…
<fragment which has not been modified>	SMLA
<Prefix/Suffix>	Prefix
<experimental mass of this peptide>	4653.40152 Da

Output

The developed program gives to user as an output: the peptide with highlighted positions and a .pepout file.

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

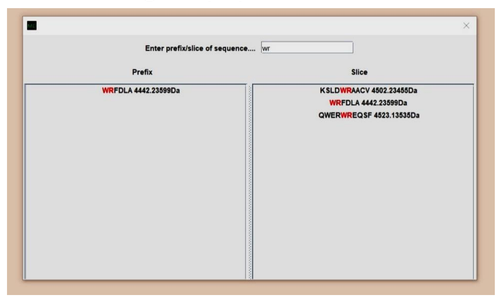
- PTM
- both
- substitution

Structure of .pepout file

General	Example
<peptide>	YASSV^{5}RSPHP^{10}AIQPL^{15}QAP…
<Prefix/Suffix> Selected	Prefix selected
at position(s) <numbers>: <sub/PTM>	at positions 11, 17: A>>V
…	at positions 13, 16: Q>>R

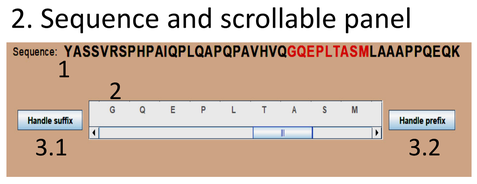
Interface

1. Searching for peptides



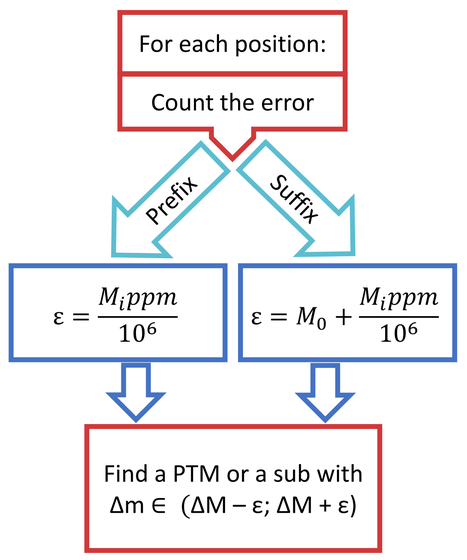
A user can search through the program's database with help of this dialog window to find peptides which are interest him

2. Sequence and scrollable panel



When a user clicks on some of the peptides, the main frame appears.
1 – the chosen amino acid sequence
2 – scrollable panel with zoomed amino acids
3.1 – button "handle suffix"
3.2 – button "handle prefix"
These buttons allows to choose the prefix or suffix, where the algorithm will search for PTMs and substitutions.

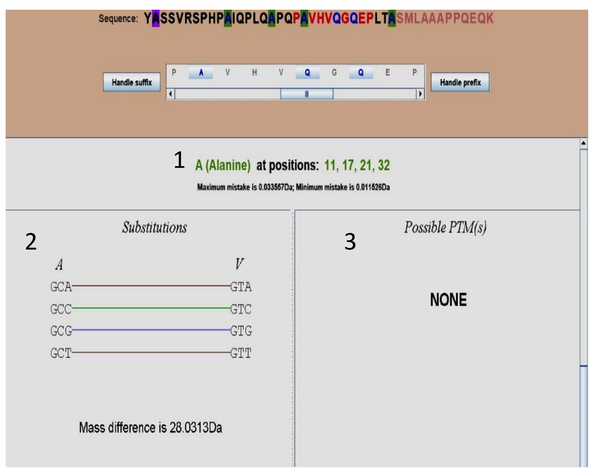
3. The algorithm



4. Highlighted positions

After this operation some of the positions appears highlighted. But not all of "A"s (and "Q"s) appears highlighted with blue color, because closer is position to the beginning of sequence smaller the error. So, in the beginning the error is not big enough.

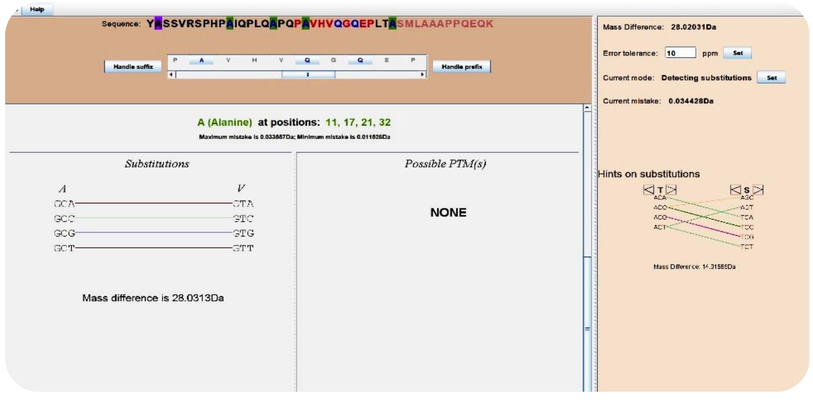
5. Information about substitutions



When a user clicks on some of the enabled buttons, the information about subs and PTMs appears.
1 – name of the amino acid under observation
2 – information about substitutions
3 – information about PTMs
Some of the codons are connected with colorful lines because they can to get into each other by SNP.

Results and conclusions

General view of an application



Additional functionality

1. Tab help
2. Hot keys
3. Hints on substitutions

Testing

To test our software tool we took some peptides, which were analyzed earlier.
.pep file:
YASSVRSPHPAIQPLQAPQPAVHVQGQEPLTASMLAA
APPQEQK SMLA prefix 4653.40152
Right result is
At positions 11, 17, 24: A>>V
At positions 13, 16, 22, 28, 30: Q>>R

Conclusion

The developed interface deals with high-resolution data, the algorithm which gives such data is new [3], so there is no a lot of applications that using such data.
Also, there is no a lot of *graphical* applications concerned with this subject

References

1. B. Lewin. *Cells*. BINOM Russia, 2011. 951 c.
2. 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.
3. K. Vyatkina, S. Wu, L. J. M. Dekker, M. M. VanDuijn, X. Liu, N. Tolic, M. Dvorkin, S. Alexandrova, T. M. Luider, 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.
4. Qisheng Peng, Zijian Wang, Donglin Wu, Xiaoou 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