An online engine for peptide mass fingerprinting from spectrum

Laetitia Jourdan, Jean-Charles Boisson, Pierre Laurence, El-Ghazali Talbi and Christian Rolando

[1] INRIA Futurs, Bat M3, Cité Scientifique, 59655 Villeneuve d'Ascq France.
`(jourdan,boisson,laurence,talbi)@lifl.fr`,
[2] Integrated Center of Bioinformatics, Bat M3, Cité Scientifique, 59655 Villeneuve d'Ascq, France.
[3] Plateforme de Protéomique / Centre Commun de Spectrométrie de masse, Bât C4-Cité Scientifique, 59655 Villeneuve d'Ascq, France.

1 Introduction.

For peptide mass fingerprinting, search engines like Sequest, Mascot, Protein-Prospector use as input data a list of monoisotopic mass values. This mass list is extracted from the mass spectrum compound of Gaussian peaks. The search engine relies entirely on the extraction of the monoisotopic mass by the extraction software which is proprietary software from the mass spectrometer manufacturer. If the extraction software miss-assigned the monoisotopic peak by one mass unit, the data base search is completely wrong. On the other hand in the case of noisy peaks which very often carry the information on the more interesting proteins and in the case of two overlapping isotopic clusters, the extraction software gave very often a poor result.

2 Materials and Methods.

In [1], we have presented a new search engine called ASCQ_ME based on an original algorithm for peptide mass fingerprinting directly from mass spectrum without mass list extraction. First, each protein present in the FASTA data base, is digested according to enzyme specificity. The isotopic cluster of each peptide is then calculated using Fast Fourier Transform using the algorithm introduced by A.L. Rockwood. The predicted spectrum is then matched against the experimental spectrum. The simplest scoring based on spectrum multiplication gave better results than the classical peptide mass fingerprinting engines. In this paper we present the online version of the software freely available.

3 Results

The online version of ASCQ_ME has been developed by the Center of Bioinformatic of the Genopole of Lille and is available on the website of the Genopole of Lille in France at the url `https://www.genopole-lille.fr/logiciel/`

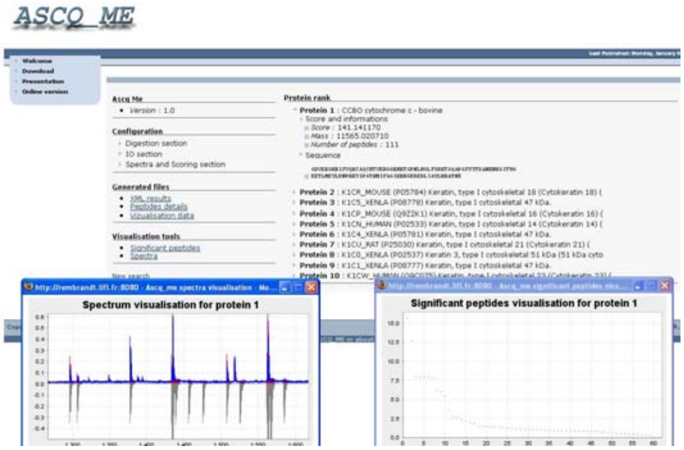
`ascq_me/`. This application is proposed to compare your results obtained by your favorite search engine. We propose a highly configurable environment developed in Java Server Page (jsp) but no so freely configurable as a local version of the engine ASCQ_ME that is available for download on the website. User can choose different parameters on the website:

- The digestion enzyme,
- The number of miscleavages (limited to 10),
- Amino acid for B, Amino acid for X, Amino acid for Z, Formula for peptides mass (e.g. "H"),
- A taxonomy,
- A fasta database (UniProtKB/SwissProt, your own database).

The online software provides:

- A ranked list of proteins with their sequence and their score,
- A graphical comparison of the experimental spectrum and the theoretical spectrum of the identified protein that allows to zoom on a part of the spectrum,
- A graphical view of the significant peptides.

The figure 1 shows an example of the results obtained for a search from a spectrum of a CytC protein. For confidentiality, user data are not keep on the server. This online version will be regularly updated to suit the evolution of the ASCQ_ME search engine.



Ascq Me
- Version : 1.0

Configuration
- Digestion section
- IO section
- Spectra and Scoring section

Generated files

Visualisation tools

Protein rank
- **Protein 1** : CCBO cytochrome c - bovine
  - Score and informations
    - Score : 141.141170
    - Mass : 11565.020710
    - Number of peptides : 111
  - Sequence
- **Protein 2** : K1CR_MOUSE (P05784) Keratin, type I cytoskeletal 18 (Cytokeratin 18) (
- **Protein 3** : K1C5_XENLA (P08778) Keratin, type I cytoskeletal 47 kDa.
- **Protein 4** : K1CP_MOUSE (Q9Z2K1) Keratin, type I cytoskeletal 16 (Cytokeratin 16) (
- **Protein 5** : K1CN_HUMAN (P02533) Keratin, type I cytoskeletal 14 (Cytokeratin 14) (
- **Protein 6** : K1C4_XENLA (P05781) Keratin, type I cytoskeletal 47 kDa.
- **Protein 7** : K1CU_RAT (P25030) Keratin, type I cytoskeletal 21 (Cytokeratin 21) (
- **Protein 8** : K1C0_XENLA (P02537) Keratin 3, type I cytoskeletal 51 kDa (51 kDa cyto
- **Protein 9** : K1C1_XENLA (P08777) Keratin, type I cytoskeletal 47 kDa.
- **Protein 10** : K1CW_HUMAN (Q9C075) Keratin, type I cytoskeletal 23 (Cytokeratin 23) (

Fig. 1. The webpages obtained for the identification of the CytC protein.

References

1. Boisson, J-C., Jourdan, L., Talbi, E-G., Cren-Olivé C. and Rolando, C: ASCQ_ME: A New Engine For Peptide Mass Fingerprinting Without Mass List Extraction. HUPO 2005. 2005. p. S18.