pFind: a Novel Web-based Application for Automated Identification of Peptide and Protein by MS/MS Data

D.Q. Li[1] , R.X. Sun[1], Y.G. We[i1], H. Zhou[2], R. Zeng[2], S.M. He[1], W. Gao[1]

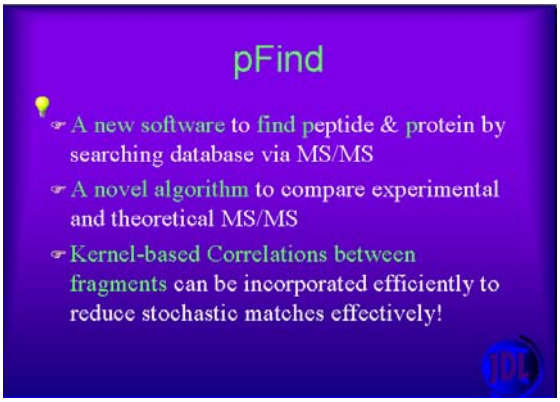
[1]Institute of Computing Technology, Chinese Academy of Sciences, Beijing, P. R. China

[2]Research Center for Proteome Analysis, Key Lab of Proteomics, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, P. R. China

Contact : dqli@jdl.ac.cn

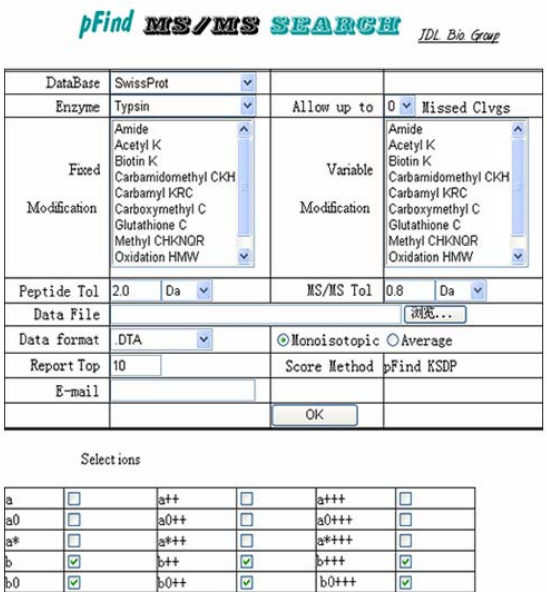
INTRODUCTION

To identify the unknown peptide sequences from the observed experimental MS/MS spectra, database searching is currently the most popular and reliable approach in high-throughput proteomics studies. We develop a new web-based database searching tool called pFind (peptide/protein Finder). The pFind is now freely available for public at http://pfind.jdl.ac.cn.

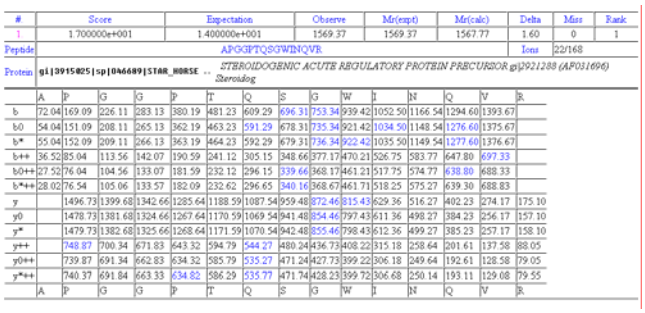


FEATURES OF pFind

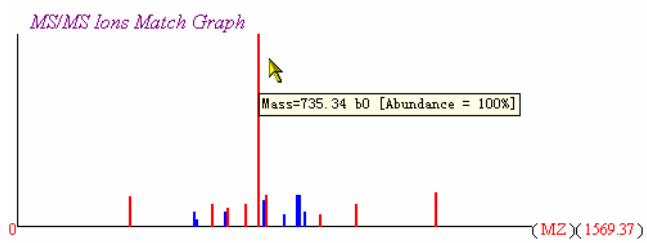
- A simple and user-friendly interface



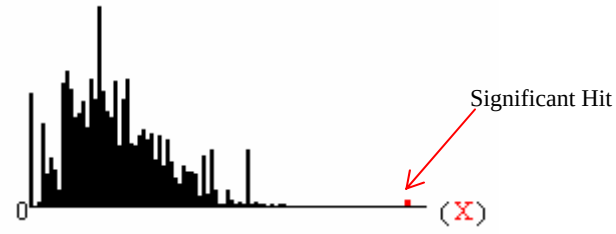
- Upload the compressed Winzip or WinRar archive files of many spectral data in the DTA format and to receive search results by an email - Web Batch Mode
- Create a querying output file in the XML format containing :
  - ✧ the best scoring peptide sequences,



  - ✧ the fragment ions match graph,



  - ✧ a statistical distribution histogram of score frequencies.



- The KSDP algorithm was coded in C++ using the Standard Template Library (STL).
- Provide the Expert Advanced mode to further improve the identification accuracy by specifying the fragment ions type and charge states.

EXPERIMENT RESULTS

Table 1. Comparison on the P18 dataset (Keller et al., *Omics*, 2002) against SWISS-PROT database

Dataset	Total	Peptides correctly identified			
		pFind		SEQUEST	
		Top 10	Top 1	Top 10	Top 1
P18A	1934	1318	1247	1320	1203
P18B	823	733	703	725	690

- Indicates that *top-1* identifications of pFind are more reliable.

ACKNOWLEDGEMENTS

This work was supported by the National Key Basic Re-search & Development Program (973) of China under Grant No. 2002CB713807.