pLabel: A Novel Tool for Tandem Mass Spectromic Data Labeling and Mining

Haipeng Wang[1, 2], Yan Fu[1, 2], Ruixiang Sun[1], Yonggang Wei[1], Rong Zeng[3], Xiaobiao Wang[1], Simin He[1], and Wen Gao[1, 2]

[1]Institute of Computing Technology, Chinese Academy of Sciences, Beijing, P. R. China
[2]Graduate School of Chinese Academy of Sciences, Beijing, P. R. China
[3]Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, P. R. China

Tandem mass spectrometry has become a mainstream technology for protein identification in routine proteomics experiments in many biological laboratories. There exist many weaknesses such as low accuracy and high false positive rate in most protein-identification systems based on tandem mass spectrometry. Although some systems incorporate empirical rules about peptide fragmentation into their peptide-scoring algorithms, automatic knowledge discovery from tandem mass spectromic data is still lacking. pLabel, developed by us, is a tool which allows automatic and efficient tandem mass spectromic data labeling and mining. Providing a tandem mass spectrum and the corresponding peptide sequence, pLabel can automatically label various ions (e.g., a, b, y series of ions and their variants due to neutral loss of water or ammonia, immonium ions, internal ions) appearing in the spectra, graphically display matching ion peaks in different colors between experimental and theoretical spectra, and record these information in an ASCII-format and predefined-structure file. With the label information, we can do data mining. For example, we can compute the probabilities of observing various ion series and analyze the corre-lations between different ions. These discovered rules are themselves of valuable biomedical signifi-cance. As an application of these rules, we aim to design a real probabilistic framework for peptide scoring. The tool is easy to use and has a friendly interface. Through the tool, we have found new frequent immonium ions such as Cys and Arg. pLabel can also be a useful aid for biologists to manually validate the peptide identification results. A free version of the tool can be downloaded at http://pfind.jdl.ac.cn/pLabel.

The System Framework

Other characteristics of the tool:
- Label peaks smartly
- Zoom spectrum by press & release
- All possible ion types
- Clear label results in ASCII-format

File menu

Batch labeling mode

Amino acid mass setting

Immonium ions mass setting

The main window of the labeling system.
(1. Ion type checkboxes 2. Parameters setting area 3. Matching table 4. Matching profile and zoom area)

Manually labeling dialog by double-clicking the lines