



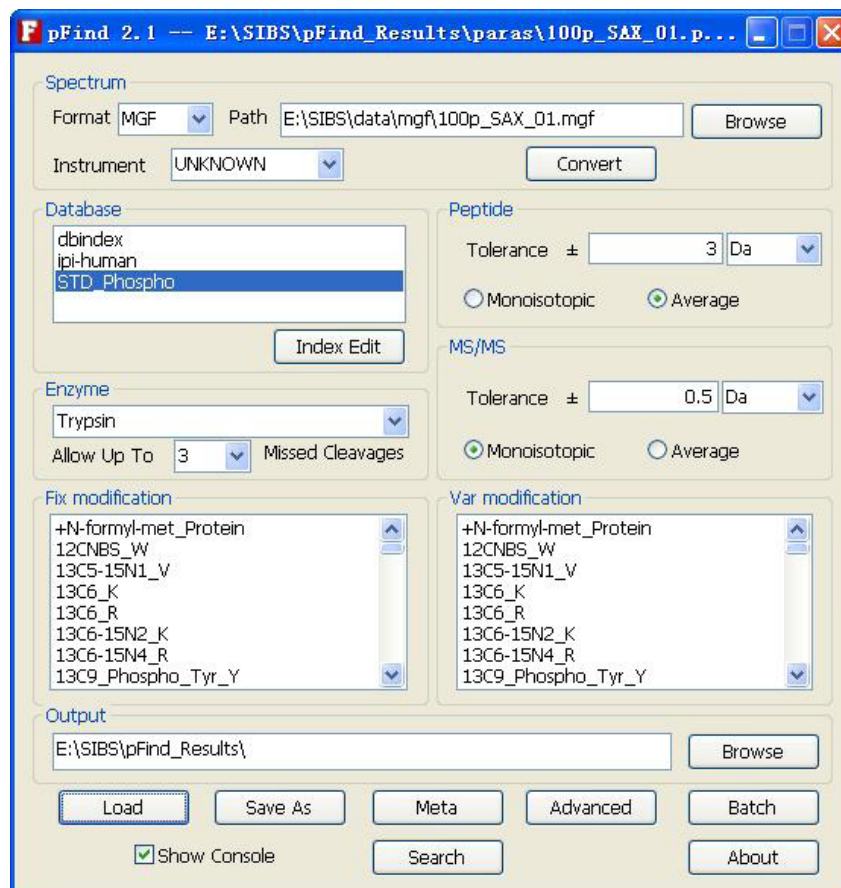
pFind Studio : Speedy and Precise Finding of Proteins

<http://pfind.ict.ac.cn>

Institute of Computing Technology, Chinese Academy of Sciences

pFind Studio is a software platform designed for high-throughput protein identification, including the database searching software pFind, together with other convenient tools for analyzing MS/MS data and protein sequence databases, such as pBuild, pLabel and pScan.

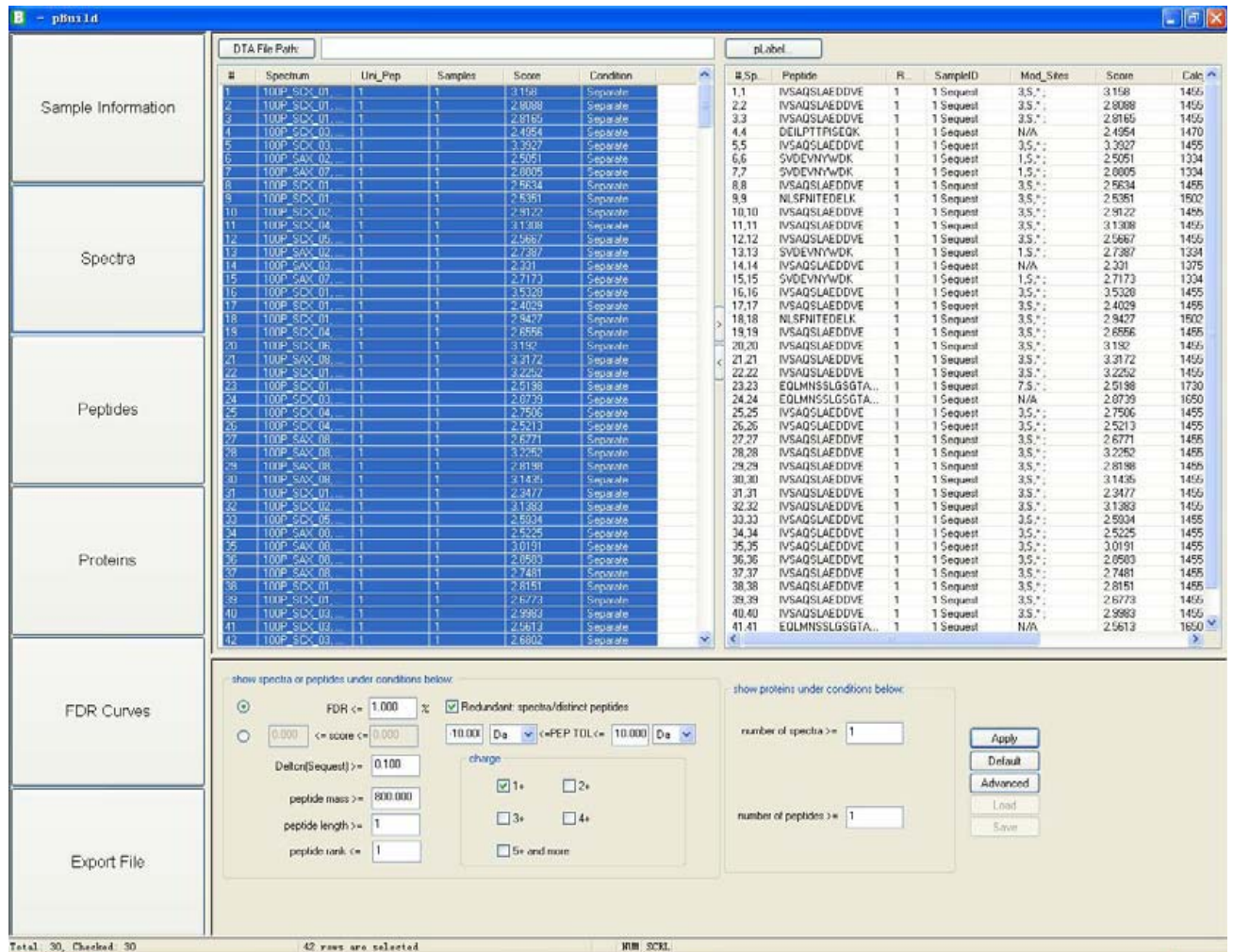
The newest version of pFind incorporates several newly developed or improved algorithms, modules and workflows: It incorporates the target-decoy database search strategy for automated FDR estimation. Users only need to specify a required FDR before searching. Then the system will calculate a threshold that achieves the FDR and filter search results automatically. We developed a toolbox to index protein databases for high-throughput application and designed all modules under a parallel-processing-oriented architecture for distributing the computational load efficiently among a lot of computers. These developments greatly improve the overall searching speed.





pBuild is a tool that can compare several search engines' results and combine them together. The latest version, pBuild v2.0, can process the search results of pFind, SEQUEST and Mascot. It is developed in the C++ programming language and you can run it in the Windows operating system. The main functions of pBuild are as follows:

1. Importing and parsing the engines' data, including TXT files of pFind, DAT files of Mascot and OUT file of SEQUEST.
 2. Comparing different engines' results and combining them together.
 3. Plotting FPR curves.
 4. Exporting files containing the list of spectra, peptides and proteins.
- It is a useful software system that can process large amount of data with a high speed.



The screenshot displays the pBuild v2.0 software interface. The main window is titled 'pBuild v2.0' and contains several panes:

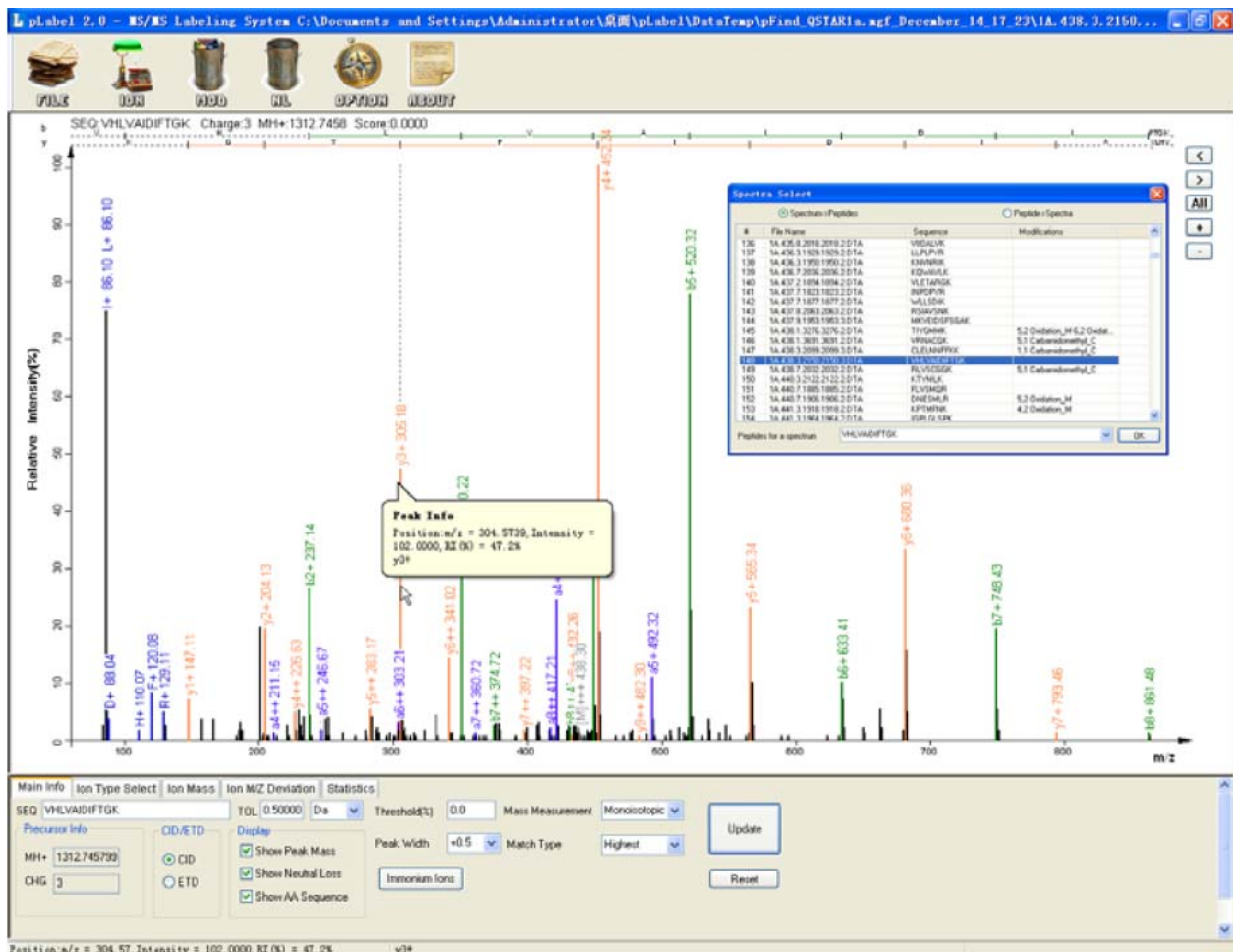
- Sample Information:** A table with columns: #, Spectrum, Uni_Pep, Samples, Score, Condition. It lists 42 samples with various IDs and scores.
- Spectra:** A table with columns: #, Sp, Peptide, R, SampleID, Mod_Sites, Score, Calc. It lists 42 spectra with peptide sequences and scores.
- Peptides:** A table with columns: #, Sp, Peptide, R, SampleID, Mod_Sites, Score, Calc. It lists 42 peptides with peptide sequences and scores.
- Proteins:** A table with columns: #, Sp, Peptide, R, SampleID, Mod_Sites, Score, Calc. It lists 42 proteins with peptide sequences and scores.
- FDR Curves:** A section for filtering results based on FDR, score, and other parameters. It includes checkboxes for 'Redundant: spectra/distinct peptides' and 'charge' (1+, 2+, 3+, 4+, 5+ and more).
- Export File:** A section for exporting results, including options for 'number of spectra' and 'number of peptides'.

At the bottom of the window, it shows 'Total: 30, Checked: 30' and '42 rows are selected'.



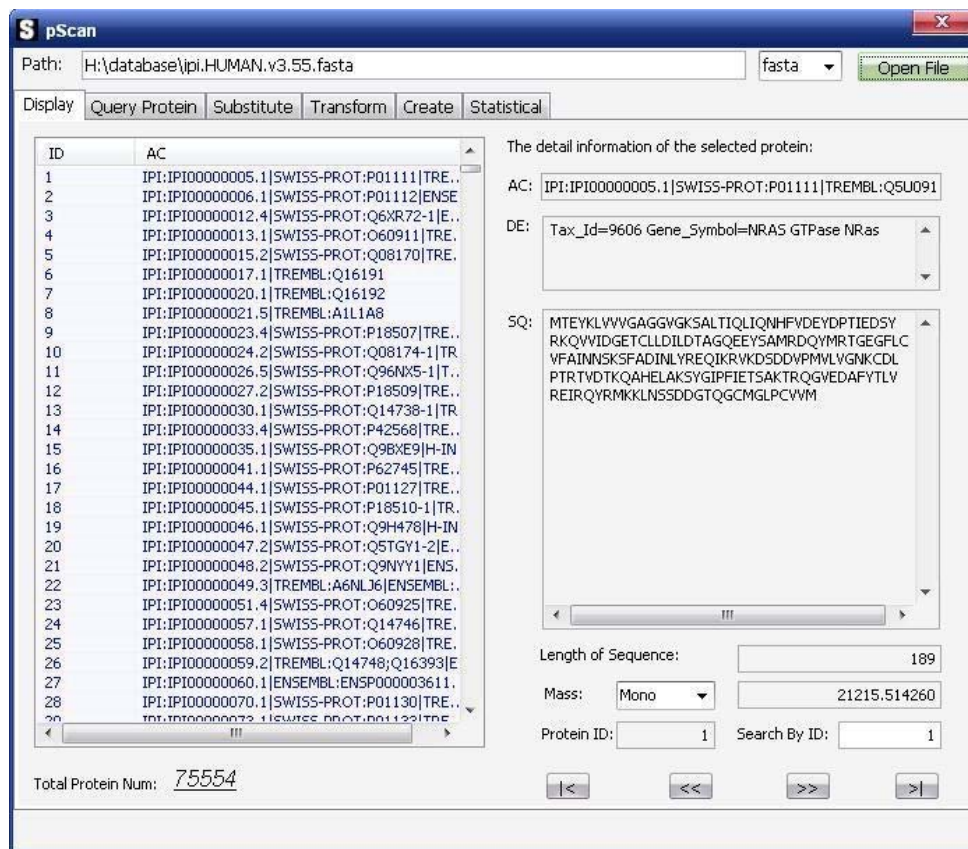
pLabel is a mass spectral peak labeling software developed for proteomics research. Follows are its main functions:

1. Supporting labeling a batch of spectra. Data and search result can be imported directly at the same time.
2. Supporting multiple formats of data type (dta, mgf, pkl) and results of multiple search engines (pFind, Mascot, SEQUEST)
3. Supporting ECD/ETD spectra labeling. [See an example of ETD labeling.](#)
4. Supporting multiple modification sites and types.
5. Supporting a flexible way of labeling the neutral losses. Users can self-define a neutral loss type with its molecular formula or mass.
6. Supporting statistical functions.
7. Supporting exporting spectra to different formats of picture (jpg, bmp, tiff, png) in single or batch mode.





pScan is a flexible tool that helps biologists to preprocess protein sequence databases in proteomics research. Besides the commonly used functions, such as sequence pattern-matching, building decoy databases, and converting protein sequence databases to peptide sequence databases, pScan also supports querying and substituting of protein entries based on the regular expression, creating customized databases, and conducting statistical characterization of the databases. pScan can greatly help biologists to improve the design of proteomics experiments and to facilitate the database search and analysis by making full use of the information content contained in the sequence databases.



We hope that pFind is helpful to your study in protein identification, post-translation modification (PTM) analysis and other important fields in proteomics. More details of pFind Studio could be found from our website: <http://pfind.ict.ac.cn>.

Reference:

1. pFind 2.0: a software package for peptide and protein identification via tandem mass spectrometry. *Rapid Communications in Mass Spectrometry*, 21(18), p2985-2991, 2007.
2. pFind: a novel database-searching software system for automated peptide and protein identification via tandem mass spectrometry. *Bioinformatics*, 21(13), 3049-3050, 2005.
3. Exploiting the kernel trick to correlate fragment ions for peptide identification via tandem mass spectrometry. *Bioinformatics*, 20, 1948-1954, 2004.

Address: No.6 Kexueyuan South Road Zhongguancun, Haidian District Beijing, China. Postcode: 100190

Tel: (8610)62601016 Website: <http://pfind.ict.ac.cn>