



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

D.Q. Li¹, R.X. Sun¹, Y.G. We¹, H. Zhou², R. Zeng², S.M. He¹, W. Gao¹

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

²Research 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>.

pFind

- A new software to find peptide & protein by searching database via MS/MS
- A novel algorithm to compare experimental and theoretical MS/MS
- Kernel-based Correlations between fragments can be incorporated efficiently to reduce stochastic matches effectively!

FEATURES OF pFind

- A simple and user-friendly interface

pFind MS/MS SEARCH IDL Bio Group

DataBase	SwissProt		
Enzyme	Trypsin	Allow up to	0 Missed Clvs
Fixed Modification	Amide	Variable Modification	Amide
	Acetyl K		Acetyl K
Modification	Biotin K		Biotin K
	Carbamidomethyl CKH		Carbamidomethyl CKH
	Carbamyl KRC		Carbamyl KRC
	Carboxymethyl C		Carboxymethyl C
	Glutathione C		Glutathione C
	Methyl CHKNOR		Methyl CHKNOR
	Oxidation HMW		Oxidation HMW
Peptide Tol	2.0 Da	MS/MS Tol	0.8 Da
Data File	浏览...		
Data format	.DTA	<input checked="" type="radio"/> Monoisotopic	<input type="radio"/> Average
Report Top	10	Score Method	pFind KSDP
E-mail			
OK			

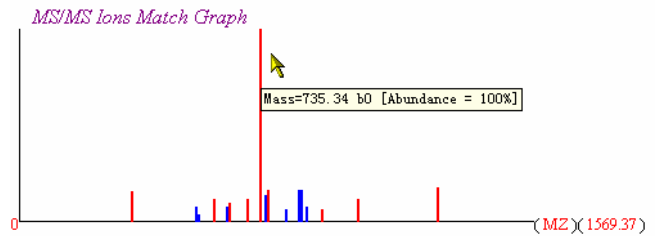
Select ions

a	<input type="checkbox"/>	a++	<input type="checkbox"/>	a+++	<input type="checkbox"/>
a0	<input type="checkbox"/>	a0++	<input type="checkbox"/>	a0+++	<input type="checkbox"/>
a*	<input type="checkbox"/>	a***	<input type="checkbox"/>	a****	<input type="checkbox"/>
b	<input checked="" type="checkbox"/>	b++	<input checked="" type="checkbox"/>	b+++	<input checked="" type="checkbox"/>
b0	<input checked="" type="checkbox"/>	b0++	<input checked="" type="checkbox"/>	b0+++	<input checked="" type="checkbox"/>

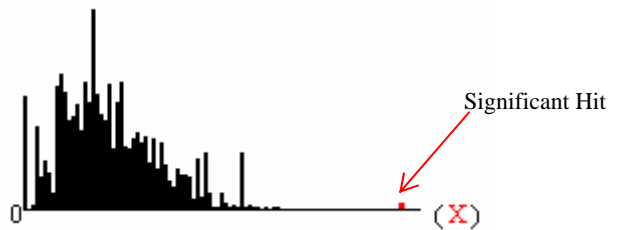
- Upload the compressed Winzip or WinRar archive files of many spectral data in the DTA format and to receive search results by an email - Web Batch Mode
- Create a querying output file in the XML format containing :
 - the best scoring peptide sequences,

#	Score	Expectation	Obsvrs	M(exp:0)	M(calc)	Delta	Mass	Rank						
1	1.70000e+001	1.40000e+001	1569.37	1569.37	1567.77	1.60	0	1						
Peptide	APGGPTQSGWLNQVR													
Protein	g 19915025 sp 046689 STAR_HORSE .. STREPTOCOCCUS ACUTE REGULATORY PROTEIN P8BCURSDR.g 921288 (AP031696) Streptococcus													
A	P	G	G	P	T	Q	S	G	W	I	N	Q	V	R
b	72.04	169.09	226.11	283.13	380.19	481.23	609.29	696.31	753.34	939.42	1052.50	1166.54	1294.60	1393.67
b0	54.04	151.09	208.11	265.13	362.19	463.23	591.29	678.31	735.34	921.42	1034.50	1148.54	1276.60	1375.67
b*	55.04	152.09	209.11	266.13	363.19	464.23	592.29	679.31	736.34	922.42	1035.50	1149.54	1277.60	1376.67
b++	36.52	85.04	113.56	142.07	190.59	241.12	305.15	348.66	377.17	470.21	526.75	583.77	647.80	697.33
b+++	27.52	76.04	104.56	133.07	181.59	232.12	296.15	339.66	368.17	461.21	517.75	574.77	638.80	688.33
b0++	28.02	76.54	105.06	133.57	182.09	232.62	296.65	340.16	368.67	461.71	518.25	575.27	639.30	688.83
y	1496.73	1399.68	1342.66	1285.64	1188.59	1087.54	959.48	872.46	815.43	629.36	516.27	402.23	274.17	175.10
y0	1478.73	1381.68	1324.66	1267.64	1170.59	1069.54	941.48	854.46	797.43	611.36	498.27	384.23	256.17	157.10
y*	1479.73	1382.68	1325.66	1268.64	1171.59	1070.54	942.48	855.46	798.43	612.36	499.27	385.23	257.17	158.10
y**	748.87	700.34	671.83	643.32	594.79	544.27	480.34	436.73	408.23	315.18	258.64	201.61	137.58	88.05
y0**	739.87	691.34	662.83	634.32	585.79	535.27	471.24	427.73	399.22	306.18	249.64	192.61	128.58	79.05
y***	740.37	691.84	663.33	634.82	586.29	535.77	471.74	428.23	399.72	306.68	250.14	193.11	129.08	79.55
A	P	G	G	P	T	Q	S	G	W	I	N	Q	V	R

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., *Omic*s, 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.