



Sample: Bioinformatics - Protein Identification Using MASCOT Database Search

Protein was identified with the help of MASCOT server. Provided PMF file contains information about mass and charge of produced peptides. MASCOT server produced the following results using SwissProt database:

Database	: SwissProt 2014_04 (544996 sequences; 193815432 residues)
Timestamp	: 22 Apr 2014 at 16:03:40 GMT
Top Score	: 142 for FABPL_HUMAN, Fatty acid-binding protein, liver OS=Homo sapiens GN=FABP1 PE=1 SV=1

Similar results were obtained by MASCOT using NCBI database. The top score protein is FABPL. The single difference between SwissProt and NCBI databases is number of matching proteins. NCBI database produced several protein accession numbers; however all of them are relevant to FABPL:

Database	: NCBI nr 20140323 (38032689 sequences; 13525028931 residues)
Timestamp	: 22 Apr 2014 at 16:12:29 GMT
Top Score	: 142 for gi 4557577, fatty acid-binding protein, liver [Homo sapiens]

MASCOT service uses obtained MS spectrum and performs searches against certain database. A list of the most matched proteins is obtained and sorted according to their score. The top protein, FABL, has the highest score.

1::FABPL_HUMAN	Mass: 14199	Score: 142	Expect: 3.4e-09	Matches: 9	
Fatty acid-binding protein, liver OS=Homo sapiens GN=FABP1 PE=1 SV=1					
1::EOGT_HUMAN	Mass: 61971	Score: 42	Expect: 32	Matches: 6	
EGF domain-specific O-linked N-acetylglucosamine transferase OS=Homo sapiens GN=EOGT PE=1 SV=1					
1::COA3_HUMAN	Mass: 11724	Score: 37	Expect: 1.1e+02	Matches: 4	
Cytochrome c oxidase assembly factor 3 homolog, mitochondrial OS=Homo sapiens GN=COA3 PE=1 SV=1					
1::EFG_GRABC	Mass: 76486	Score: 54	Expect: 2.3	Matches: 7	
Elongation factor G OS=Granulibacter betshdensis (strain ATCC BAA-1260 / CGDNIH1) GN=fusA PE=3 SV=1					

Protein sequence coverage for FABL is 85%. Matched peptides shown in bold red.

```

1  MSFSGKYQLQ SQENFEAFMK AIGLPEELIQ KGKDIKGVSE IVQNGKHFKE
51  TITAGSKVIQ NEFTVGEECE LETMTGEKVK TVVQLEGDNK LVTTFKNIKS
101 VTELNGDIIT NTMTLGDIVF KRISKRI
    
```



The amino acid sequences of each protein fragment of FABL are listed below.

Start	End	Observed	Mr(expt)	Mr(calc)	Delta	M	Peptide
1	– 6	656.3072	655.2999	655.2999	-0.0000	0	-.MSFSGK.Y
7	– 20	1762.8110	1761.8037	1761.8032	0.0005	0	K.YQLQSQENFEAFMK.A
21	– 31	1210.7040	1209.6967	1209.6969	-0.0001	0	K.AIGLPEELIQK.G
37	– 46	1030.5530	1029.5457	1029.5455	0.0003	0	K.GVSEIVQNGK.H
50	– 57	824.4512	823.4439	823.4440	-0.0000	0	K.FTITAGSK.V
58	– 78	2386.0790	2385.0717	2385.0716	0.0001	0	K.VIQNEFTVGEECELETMTGEK.V
81	– 90	1102.5740	1101.5667	1101.5666	0.0001	0	K.TVVQLEGDNK.L
91	– 96	708.4290	707.4217	707.4218	-0.0001	0	K.LVTTFK.N
100	– 121	2381.2270	2380.2197	2380.2196	0.0001	0	K.SVTELNQDIITNTMTLGDIVFK.R

It is also possible to verify whether highly similar proteins exist in human proteome. This could be easily done with protein BLAST through human database. This kind of analysis revealed a single hit with high score corresponding to FABPL.