ASCQ-PROT : AN OPEN SOURCE SOFTWARE FOR PROTEOMICS ANALYSIS

Pierre Laurence, David Boens, Caroline Tokarski, Séverine Le Gac, Cécile Cren-Olivé et Christian Rolando
Université des Sciences et Technologies de Lille, UMR CNRS 8009 Chimie Organique et Macromoléculaire, 59655 Villeneuve d'Ascq, France

Database searching reaches an increasing place in proteomics and the overall quality depends heavily on the data mining step. In our hands, as already noticed by many research groups, results obtained for a same sample are very different according to the searching program used. It is very difficult to overcome this observation since the information given on the different program do not allow to know precisely the algorithm and even less the limitations set for speeding up the search. Contrary to other analytical fields and more particularly to NMR, very few free software exists in mass spectrometry. So we decide to develop an open source software for peptide fingerprint with four main goals: (i) to create an accessible software as part of the license Open Software Foundation (ii) to develop a software which does not run thought a web interface but rather uses as enter and exit parameters different text files so a to allow a larger flexibility, (iii) to find a fitting algorithm allowing no limitation on the depth of the search more particularly for the post-translational modification, and which can be used in recursively for the identification of minority proteins, (iv) to develop a software which optimizes the given answer in function of criteria corresponding to the ones used by experts rather than using discrimination based on statistics.

The weighting of the different parameters used (percentage of mass values matched, sequence coverage, localization of the matched peptides inside the protein sequence, average mass deviation, number of post-translational modifications) may be modified in function of different sample types: identification of post-translational modifications, samples very noisy in case of spot with a low intensity and identification of different protein in a mixture.

```
ASCQ-PROT [beta-test version 0.94]
Finds proteins containing specified masses fragments in a fasta format file.
options:
  -d <FASTA-FILE>    : fasta format file database containing sequences
  -w <TEXT-FILE>     : file containing masses fragments datas
  -o <TEXT-FILE>     : output file for result of search
                       (if no TEXT_FILE specified, result will be save in
                       "_results" directory of ASCQ-PROT with the name of
                       the masses fragments file)
  -f <FASTA-FILE>    : fasta format file for sequences result
  -p <CONFIG-FILE>   : file containing parameters of search
                       (examples in "_config" directory)
  -m <number>        : minimum masses hit (dft: 5)
                       % is also allowed example : "-m 50%"
  -c <number>        : maximum cleave misses (dft: 0)
  -t <number>        : tolerance (dft: 0.2)
  -l <number>        : low mass range (dft: 22000)
  -h <number>        : high mass range(dft: 40000)
                       "-h none" for no limit
  -s <number>        : minimum score for a selected protein (dft: 0.5)
  -covert <number>   : (dft: 4)
  -mispep <number>   : (dft: 1)
  -difavg <number>   : (dft: 1)
  -cleava <number>   : (dft: 1)
  -baryce <number>   : (dft: 1)
  -aat <AA-FILE>     : file contains amino-acids datas
  -gt <GENETIC-FILE> : file contains genetic datas
  -enzyme <string>   : enzyme to use for the digest
                       "-enzymeslist" option prints the enzymes list
  -modif             : with post-translationnal and chemical modifications
  -dna               : with DNA fasta sequences
  -filter <string>   : search filter (specie for example)
  -beep              : sound when work completed
  -silent            : silent mode
```

Command file

```
# fasta database file
-d .\_fasta\sprot.fas
# mass file
-w .\_mass_data\caro_apo137
# output file
-o
# Tolerance (amu)
-t 0.1
# Minimum mass required
-m 20%
# Missed cleavage allowed
-c 5
# Minimum score
-s 0.4
# Low mass range
-l 0
# High mass range
-h none
# Modifications
- modif
# Filter accelerates search
-filter HUMAN
# Silent mode
# -silent
# Beep when finished
-beep
```

Result file

```
.time: 0 min 25 sec
 .number of proteins checked: 122564
 .number of proteins found: 123
 .top 5:
                    1. [0.729]APA1_HUMAN (P02647) Apolipoprotein A-I precursor (Apo-AI).
                    2. [0.490]Z317_HUMAN (Q96PQ6) Zinc finger protein 317.
                    3. [0.489]Z141_HUMAN (Q15928) Zinc finger protein 141.
                    4. [0.488]CYL1_HUMAN (P35663) Cylicin I (Multiple-band polypeptide I) (Fragment)
                    5. [0.476]Z382_HUMAN (Q96SR6) Zinc finger protein 382.

DETAILS
-------
[score: 0.729]
[covert: 0.754][mispep: 0.420][difavg: 0.854]
[cleava: 0.865][baryce: 0.677][BC:   192-208  ]
>APA1_HUMAN (P02647) Apolipoprotein A-I precursor (Apo-AI).
                                  hits      : 34
                                  aa number : 267
                                  weight    : 30758.934
                                  frags     : 40
 2220.118 +0.075 +79.966   3     102-120    ETEGLRQEMSKDLEEVKAK
 1878.034 +0.020           2      35-51     VKDLATVYVDVLKDSGR
 1815.851 +0.029           1      48-64     DSGRDYVSQFEGSALGK
 1723.946 +0.012           2     141-155    QKVEPLRAELQEGAR
 1650.870 -0.011           1      37-51     DLATVYVDVLKDSGR
 1612.786 -0.006           0      70-83     LLDNWDSVTSTFSK
 1585.809 +0.012           1     185-197    THLAPYSDELRQR
 1514.811 -0.002           1     251-263    VSFLSALEEYTKK
 1467.792 +0.019           1     143-155    VEPLRAELQEGAR
 1462.852 -0.009           1      35-47     VKDLATVYVDVLK
 1451.754 -0.021           1     119-130    AKVQPYLDDFQK
 1411.668 +0.011 +15.995   1     131-140    KWQEEMELYR
 1400.670 -0.009           0      52-64    DYVSQFEGSALGK
 1386.716 -0.009           0     251-262    VSFLSALEEYTK
 1380.716 -0.004           1     121-131    VQPYLDDFQKK
 1302.648 -0.018 +15.995   1     165-175    LSPLGEEMRDR
 1235.689 -0.089 +79.966   0      37-47     DLATVYVDVLK
 1301.649 +0.001           0     185-195    THLAPYSDELR
 1283.573 -0.000 +15.995   0     132-140    WQEEMELYR
 1252.621 -0.012           0     121-130    VQPYLDDFQK
 1230.710 -0.014           0     240-250    QGLLPVLESFK
 1215.622 +0.014           0     220-230    ATEHLSTLSEK
 1157.628 +0.006           1     202-212    LEALKENGGAR
 1152.638 -0.011           1     156-164    QKLHELQEK
 1031.520 -0.009 +15.995   0     165-173    LSPLGEEMR
 1012.579 +0.004           1     231-239    AKPALEDLR
 1008.570 -0.011           1     176-184    ARAHVDALR
  931.510 -0.009           1     113-120    DLEEVKAK
  896.484 -0.013           0     158-164    LHELQEK
  873.443 -0.001           0     148-155    AELQEGAR
  869.521 -0.007           1     141-147    QKVEPLR
  831.437 -0.011           0     213-219    LAEYHAK
  781.432 -0.009           0     178-184    AHVDALR
```

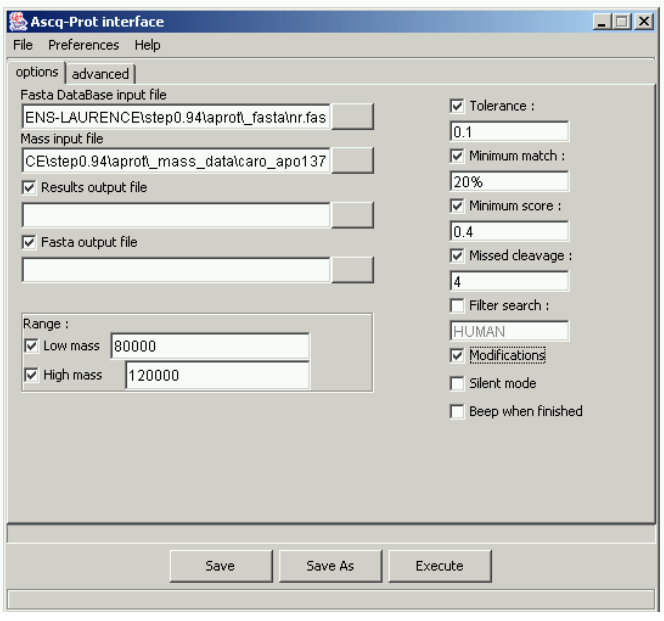
Result file (continued)

```
mkaavltlavlfltgsqarhfwqqdeppqspwdrVKDLATVYVDVLKDSGRDYVSQFEGS
ALGKqlnlkLLDNWDSVTSTFSKlreqlgpvtqefwdnlekETEGLRQEMSKDLEEVKAK
VQPYLDDFQKKWQEEMELYRQKVEPLRAELQEGARQKLHELQEKLSPLGEEMRDRARAHV
DALRTHLAPYSDELRQRlaarLEALKENGGARLAEYHAKATEHLSTLSEKAKPALEDLRQ
GLLPVLESFKVSFLSALEEYTKKlntq

missed mass:

 3259.568  2677.184  2300.160  2234.134  2211.103  2172.908
 2156.959  2140.947  2124.932  2108.942  2082.985  1878.054
 1815.881  1723.958  1706.941  1690.731  1650.859  1644.768
 1628.767  1616.786  1613.383  1612.780  1585.820  1514.809
 1504.796  1475.761  1467.811  1462.843  1459.680  1451.733
 1449.747  1427.674  1400.661  1386.707  1380.712  1331.563
 1318.625  1315.566  1302.210  1301.650  1299.568  1283.596
 1257.659  1252.609  1238.665  1230.696  1226.549  1215.636
 1213.678  1169.543  1159.150  1158.620  1157.634  1152.627
 1141.502  1135.605  1109.524  1066.593  1047.506  1045.574
 1031.520  1013.070  1012.583  1011.481  1008.560   973.497
  968.493   931.501   917.268   901.479   896.471   873.442
  870.343   869.514   851.440   842.510   831.425   807.396
  781.423   780.394   732.366
```

Window version (Java)



Download
http://ascqprot.free.fr/