Candidate Biomarker Discovery for Angiogenesis by Automatic Integration of Orbitrap MS1 Spectral- and X!Tandem MS2 Sequencing Information

Mark K. Titulaer *

Academic Medical Center, University of Amsterdam, 1100 DD Amsterdam, The Netherlands

Received 30 November 2012; revised 21 February 2013; accepted 28 February 2013
Available online 2 April 2013

Abstract  Candidate protein biomarker discovery by full automatic integration of Orbitrap full MS1 spectral peptide profiling and X!Tandem MS2 peptide sequencing is investigated by analyzing mass spectra from brain tumor samples using Peptrix. Potential protein candidate biomarkers found for angiogenesis are compared with those previously reported in the literature and obtained from previous Fourier transform ion cyclotron resonance (FT-ICR) peptide profiling. Lower mass accuracy of peptide masses measured by Orbitrap compared to those measured by FT-ICR is compensated by the larger number of detected masses separated by liquid chromatography (LC), which can be directly linked to protein identifications. The number of peptide sequences divided by the number of unique sequences is 9248/6911 = 1.3. Peptide sequences appear 1.3 times redundant per up-regulated protein on average in the peptide profile matrix, and do not seem always up-regulated due to tailing in LC retention time (40%), modifications (40%) and mass determination errors (20%). Significantly up-regulated proteins found by integration of X!Tandem are described in the literature as tumor markers and some are linked to angiogenesis. New potential biomarkers are found, but need to be validated independently. Eventually more proteins could be found by actively involving MS2 sequence information in the creation of the MS1 peptide profile matrix.

Introduction

Orbitrap mass spectrometry (MS) plays an increasingly important role in proteomics research. Orbitrap combines great mass accuracy (<10 ppm) with a high processing speed for peptide mixtures separated by nano-liquid chromatography (nano-LC). A comparable instrument, the Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer, has a slightly better mass accuracy of ≤1 ppm [1]. However, the performance of FT-ICR depends on the magnet strength chosen, e.g., 9.4 T, and thus it takes up a lot more space because of the sizeable magnet. Orbitrap has a shorter time scale of measuring than FT-ICR and can therefore be more conveniently linked to a nano-LC column to measure peptide fractions of a trypsin-digested sample.

Comparison of mass spectra from different samples is called peptide profiling. Peptide profiling creates a matrix of
Figure 1  Architecture of Peptrix

Peptrix processes raw Orbitrap files to a peptide profile matrix of all the intensities measured for every peptide mass with a certain retention time in the different samples with protein identifications. Peptrix implements Readw.exe (version 4.3.1), *.fasta file, and tandem.exe. Peptrix generates a Mascot generic file (MGF) for each Orbitrap file. The *.fasta file is the text database to correlate MS2 fragmentation masses to a protein.
## Table 1  Differentially expressed peptides of proteins between GV and NV groups

| Mass MH+ (Da) | Time (s) | Calculated MH+ (Da) | ppm | Mod | Sequence | Miss | P | Up | Ave. intensity (×100) |
|---------------|----------|---------------------|-----|-----|----------|------|---|---|----------------------|
| **GV/NV**     | **GV**   | **NV**              |     |     |          |      |   |   |                      |
| 1051.5951     | 11,281   | 1051.5935           | 1.5 | SFLAFTPLR | 0 | 0.0022 | + | High | 124 | 0                   |
| 1154.6514     | 6352     | 1154.6527           | 1.1 | LELGIPGAAATR | 0 | 0.1207 | + | 3.1 | 364 | 117                |
| 1188.6501     | 8036     | 1205.7000           | 19.4 | QLLTPHVL | 0 | 0.0448 | + | 2.2 | 552 | 255                |
| 1246.6788     | 5956     | 1246.6790           | 0.2 | GSGLILNYGQK | 0 | 0.9692 | - | 0.8 | 175 | 215                |
| 1293.6452     | 15,752   | 1295.6477           | 1.9 | SIESTLDDLFR | 0 | 0.0283 | + | 13.8 | 219 | 16                 |
| 1323.6219     | 5687     | 1323.6175           | 3.3 | SAGDVDTLAFDGR | 0 | 0.0072 | + | 11.7 | 494 | 422                |
| 1628.6876     | 9105     | 1628.8754           | 0.6 | ALEPGQLLLYGNAR | 0 | 0.0110 | + | 10.5 | 145 | 14                 |
| **sp|O75891|AL1L1_HUMAN aldehyde dehydrogenase family 1 member L1** |
| 1363.6835     | 10,783   | 1363.6852           | 1.2 | DLGEAALNEYLR | 0 | 0.0060 | + | High | 146 | 0                  |
| 1787.9414     | 5112     | 1787.9174           | 13.4 | LQQPEEAGTYEGIQK | 1 | 0.0187 | + | 39.1 | 312 | 8                  |
| **tr|E9PJL7|E9PJL7_HUMAN Uncharacterized protein** |
| 1165.6595     | 5443     | 1165.6575           | 1.7 | VLGVDVEVHGK | 0 | 0.1000 | + | 14.6 | 1054 | 72                 |
| **tr|E7ETO3|E7ETO3_HUMAN Uncharacterized protein (cell division control protein 42 homolog)** |
| 1244.6842     | 6133     | 1241.6736           | 0.9 | 1 | QKPHPTEAK | 1 | 0.7910 | + | 11.8 | 168 | 153               |
| 1472.8491     | 13,709   | 1472.8471           | 1.4 | TFPQGLVTQIDL | 0 | 0.0298 | + | 17.2 | 109 | 6                  |
| **sp|P10909|CLUS_HUMAN clusterin** |
| 1117.6073     | 7057     | 1117.6099           | 2.3 | TLLSNLEAK | 0 | 0.0125 | + | 25.5 | 876 | 34                  |
| 1245.6996     | 5854     | 1245.7049           | 4.3 | TLLSNLEAK | 0 | 0.0146 | + | 80.7 | 1019 | 13                |
| 1296.7549     | 7499     | 1296.7521           | 2.2 | 6 | PTVTPVEVSR | 0 | 0.0060 | + | High | 460 | 0                  |
| 1370.7441     | 5927     | 1360.7430           | 0.8 | TLEKNTNEERK | 2 | 0.0306 | + | 58.5 | 288 | 5                  |
| 1373.8000     | 5205     | 1373.7998           | 0.1 | KTLIATNLEAK | 2 | 0.0022 | + | High | 1404 | 0               |
| 1573.8515     | 5619     | 1575.8489           | 1.6 | YVVEKIQAGVNGK | 0 | 0.0022 | + | High | 417 | 0                  |
| 1576.8315     | 5200     | 1575.8499           | 0.9 | 4 | YVVEKIQAGVNGK | 0 | 0.0426 | + | 14.4 | 621 | 43                |
| 1837.9827     | 12,680   | 1837.9905           | 4.2 | LFSDPITVPVEVSR | 0 | 0.0187 | + | 141.2 | 1721 | 12              |
| 2028.8660     | 6065     | 2009.8716           | 2.8 | DQTVSDNELQEMSNGK | 0 | 0.0889 | + | 14 | 1469 | 105           |
| 2078.8252     | 6109     | 2009.8716           | 3.8 | 4 | DQTVSDNELQEMSNGK | 0 | 0.0666 | + | 18.2 | 246 | 14               |
| **tr|B4DN87|B4DN87_HUMAN Uncharacterized protein (collagen-binding protein 2 P50454)** |
| 1229.6574     | 5092     | 1224.6582           | 0.7 | GVVEVTHDLQK | 0 | 0.1613 | + | 25.5 | 876 | 34                  |
| 1337.7452     | 6553     | 1337.7423           | 2.2 | HLAGGLLTEAIDK | 0 | 0.0172 | + | 8.4 | 102 | 12                |
| 1659.8017     | 11,433   | 1659.8013           | 0.2 | LGYGPSVSFADDVF | 0 | 0.3056 | + | 16.2 | 409 | 25                |
| 1887.0880     | 9896     | 1887.0810           | 3.7 | DTQGSSLIFGLRVPK | 0 | 0.0779 | + | High | 65 | 0                  |
| **tr|B4DF14|B4DF14_HUMAN Uncharacterized protein (excitatory amino acid transporter 1)** |
| 1723.7509     | 6962     | 1723.7513           | 0.2 | DVMGNSVVEFENEMK | 0 | 0.0139 | + | 11.8 | 168 | 14                 |
| 1759.7289     | 6837     | 1723.7513           | 12.1 | 2 | DVMGNSVVEFENEMK | 0 | 0.0229 | + | 16.3 | 449 | 28               |
| 2040.0745     | 11,079   | 2040.0720           | 1.2 | TITTVSDNELQEMSNGK | 0 | 0.0060 | + | High | 211 | 0                  |
| **sp|Q16658|FSCN1_HUMAN fascin** |
| 1076.5313     | 5760     | 1076.5119           | 1.3 | YSVQTDLR | 0 | 0.2247 | + | 2.3 | 272 | 120               |
| 1076.5314     | 5880     | 1076.5119           | 18.2 | YSVQTDLR | 0 | 0.5149 | - | 0.2 | 1159 | 4652             |
| 1190.6392     | 5446     | 1190.6415           | 1.9 | YAPSGPSGTLLK | 0 | 0.0448 | + | 2.7 | 404 | 148                |
| 1200.6740     | 5289     | 1200.6735           | 0.4 | YLKGHDAGVLK | 1 | 0.0489 | + | 5.6 | 241 | 43                |
| 1240.7888     | 6908     | 1240.7888           | 0.4 | YLNPFFIVF | 1 | 0.0008 | + | High | 225 | 0                  |
| 1819.9725     | 7243     | 1819.9701           | 1.3 | LVARPEPATGTLFR | 1 | 0.1012 | + | 48.2 | 294 | 6                  |
Table 1 continued

| Mass MH+ (Da) | Time (s) | Calculated MH+ (Da) | ppm | Mod | Sequence | Miss | P     | Up Ave. intensity (×100) |
|---------------|----------|---------------------|-----|-----|----------|------|-------|---------------------------------------------------|

**sp|O15540|FABP7_HUMAN** fatty acid-binding protein, brain

| sp|O15540|FABP7_HUMAN | fatty acid-binding protein, brain |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| 963.5163 | 5668 | 963.5258 | 9.9 | 8 | **T**FGDVAVR | 0 | 0.3994 | + 2.2 | 279 | 129 | **T**FGDVAVR | 0 | 0.3994 | + 2.2 | 279 | 129 |
| 1489.6637 | 6212 | 1489.6628 | 0.6 | | LTNNSQNFDEYMK | 0 | 0.0008 | + High | 676 | 0 | LTNNSQNFDEYMK | 0 | 0.1826 | - 0.4 | 58 | 157 |
| 1489.6672 | 5901 | 1489.6628 | 3.0 | | LTNNSQNFDEYMK | 0 | 0.0008 | + High | 676 | 0 | LTNNSQNFDEYMK | 0 | 0.0008 | + High | 676 | 0 |
| 1505.6539 | 6215 | 1489.6628 | 5.6 | 2 | LTNNSQNFDEYMK | 0 | 0.0022 | + High | 324 | 0 | LTNNSQNFDEYMK | 0 | 0.0022 | + High | 324 | 0 |
| 1558.8059 | 12,106 | 1538.8069 | 0.3 | 2 | **M**YMTLTFGDIVAVR | 0 | 0.8942 | + 1.9 | 26 | 14 | **M**YMTLTFGDIVAVR | 0 | 0.8942 | + 1.9 | 26 | 14 |
| 1637.9231 | 6219 | 1637.9220 | 0.7 | | SVSLSLGDKVLHVQK | 1 | 0.0147 | + 137.3 | 474 | 3 | SVSLSLGDKVLHVQK | 1 | 0.0147 | + 137.3 | 474 | 3 |
| 1896.0122 | 7155 | 1913.0338 | 2.6 | 1 | **Q**VGNVTKTPIVSSEQGDK | 1 | 0.0149 | + High | 53 | 0 | **Q**VGNVTKTPIVSSEQGDK | 1 | 0.0149 | + High | 53 | 0 |
| 2363.3206 | 9030 | 2380.3558 | 3.7 | 1 | **Q**VGNVTKTPIVSSEQGDKVIR | 2 | 0.0149 | + High | 113 | 0 | **Q**VGNVTKTPIVSSEQGDKVIR | 2 | 0.0149 | + High | 113 | 0 |
| 2380.3550 | 7027 | 2380.3558 | 0.3 | | **Q**VGNVTKTPIVSSEQGDKVIR | 2 | 0.0837 | + 215.1 | 379 | 2 | **Q**VGNVTKTPIVSSEQGDKVIR | 2 | 0.0837 | + 215.1 | 379 | 2 |

**sp|P49327|FAS_HUMAN** fatty acid synthase

| sp|P49327|FAS_HUMAN | fatty acid synthase |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| 1038.6568 | 7667 | 1038.6557 | 1.1 | | GTPLISPLIK | 0 | 0.0403 | + 30.6 | 119 | 4 | GTPLISPLIK | 0 | 0.0403 | + 30.6 | 119 | 4 |
| 1115.6675 | 5470 | 1115.6670 | 0.4 | | SEGVVAVLTLK | 0 | 0.0350 | + High | 30 | 0 | SEGVVAVLTLK | 0 | 0.0350 | + High | 30 | 0 |
| 1263.7421 | 6778 | 1263.7419 | 0.2 | | LQVVDQPLPVR | 0 | 0.0299 | + 2.6 | 123 | 47 | LQVVDQPLPVR | 0 | 0.0299 | + 2.6 | 123 | 47 |
| 1298.6487 | 5094 | 1298.6499 | 16.3 | | YGDPQELNGITR | 0 | 0.2694 | + 5.9 | 382 | 84 | YGDPQELNGITR | 0 | 0.2694 | + 5.9 | 382 | 84 |
| 1308.6121 | 14,642 | 1250.6119 | 0 | 3,2 | **M**EEVIAIMSGK | 0 | 0.0041 | + 9.7 | 49 | 5 | **M**EEVIAIMSGK | 0 | 0.0041 | + 9.7 | 49 | 5 |
| 1386.7491 | 9654 | 1386.7474 | 1.3 | | GVDLVLNLSAAEKEK | 0 | 0.0030 | + 9 | 82 | 9 | GVDLVLNLSAAEKEK | 0 | 0.0030 | + 9 | 82 | 9 |
| 1426.7769 | 8686 | 1426.7732 | 3.3 | | SLLVNPQGPTLMR | 0 | 0.0366 | + 22 | 113 | 5 | SLLVNPQGPTLMR | 0 | 0.0366 | + 22 | 113 | 5 |
| 1469.6989 | 7733 | 1469.7019 | 2.0 | | FPQQLSTDTSFANSR | 0 | 0.0522 | + 4.9 | 97 | 20 | FPQQLSTDTSFANSR | 0 | 0.0522 | + 4.9 | 97 | 20 |
| 1622.9367 | 10,276 | 1622.9363 | 0.2 | | VVQVQVLAEPEAVLKK | 0 | 0.0366 | + 11.1 | 69 | 6 | VVQVQVLAEPEAVLKK | 0 | 0.0366 | + 11.1 | 69 | 6 |

**tr|E9P360|E9P77_HUMAN** Uncharacterized protein (**FINC_HUMAN.fibromectin**)
Table 1 continued

| Mass MH+ (Da) | Time (s) | Calculated MH+ (Da) | ppm | Mod | Sequence | Miss | P | Up Ave. intensity (×100) |
|--------------|---------|---------------------|-----|-----|----------|------|---|------------------------|

sp|Q13423|NNTM_HUMAN NAD(P) transhydrogenase, mitochondrial
953.5322 | 5046 | 953.5315 | 0.7 | FGHVPVAGR | 0 | 0.0366 | + 10.4 | 127 12
1072.6238 | 6122 | 1072.6248 | 0.9 | EVLASDLVVK | 0 | 0.0149 | + High | 123 0
1270.6883 | 7837 | 1270.6898 | 0.5 | SLGEAPELVDLK | 0 | 0.0049 | + 15.8 | 86 5
1429.7437 | 5824 | 1429.7434 | 0.2 | GITHYGTDPISR | 0 | 0.0277 | + 4.8 | 191 40
1524.9084 | 9253 | 1524.9107 | 1.5 | ILIVGGVYGLASGAAN | 0 | 0.0237 | + 7.6 | 151 20
1561.7473 | 8847 | 1578.7758 | 1.3 | QGFNVVVESGAGEASK | 0 | 0.0060 | + High | 53 0

sp|O43175|SERD_HUMAN D-3-phosphoglycerate dehydrogenase
1099.6110 | 6981 | 1099.6105 | 0.5 | GGIvDEGALLR | 0 | 0.0122 | + 93.5 | 7167 77
1298.7301 | 6156 | 1298.7314 | 0.5 | ILQDGGLQVVEK | 0 | 0.0149 | + High | 153 0
1345.7718 | 6477 | 1345.7685 | 2.5 | GTIQVITQITGKEDAANNYAR | 0 | 0.0187 | + 58.4 | 290 5

Note: MH+ indicates the protonated peptide mass. Amino acids with modifications are highlighted. Mod indicates modification: (1) – NH3, (2) + O, (3) + CO2H, (4) – NH2 + OH, (5) non-tryptic, TA, (6) DP, (7) AS, (8) LT, (9) FQ, (10) TS, (11) QS, (12) SL, (13) GDD, (14) EY. Miss indicates number of missed cleavages. In Up column, + indicates intensity GV > NV while – indicates intensity GV < NV.

all the intensities measured for every peptide mass with a certain retention time in different samples [2]. There are a number of open-source software applications for comparing Orbitrap mass spectra from large numbers of samples in different groups [3–5]. Open-source software applications running on Windows operating system (OS), which combine the full spectral MS1 peptide masses and protein identifications by fragmentation of peptide masses (MS2 or MS/MS), are scarce. A few such examples are MsInspect, MSight, MSQuant and MaxQuant [3–6]. The majority of open-source packages use commercial search engines such as Mascot and Sequest to correlate MS2 fragmentation spectra with proteins in *fasta* databases, such as MSQuant and the early version of MaxQuant [4,6,7]. Other open-source software packages that compare MS1 spectra from various samples with each other can only be installed on the Linux OS [5]. There are applications which use statistics to first determine differentially expressed MS1 peptide masses between the samples [8]. The masses that are differentially expressed with peak intensities in the groups are linked to a protein using MS2 fragmentation spectra. Some applications determine relative quantities of a protein between samples on the basis of the number of times that a protein’s peptide sequence is detected in an MS2 scan in a nano-LC Orbitrap measurement. For example, in MaxQuant [7], for label-free quantification, the maximum number of peptides between any two samples is compared, resulting in a matrix of protein ratios. There are a number of drawbacks associated with this spectral count technique. The sequence
Table 2 Differentially expressed proteins with number of peptides between GV and NV groups

| Classification | Protein name | No. of peptides | References | Classification | Protein name | No. of peptides | References |
|----------------|--------------|-----------------|------------|----------------|--------------|----------------|------------|
| Major blood proteins | Fibrinogen | 96 [11,12,14,16] | Cytoskeleton | Actin-related protein 2 | 4 [14] | | |
| | Hemoglobin subunit alpha | 22 [1,14] | | | Actin-related protein 2/3 | 2 [14,15] | complex subunit 3 |
| | Ig kappa chain C | 3 [11] | | | Actin-related protein 3 | 2 [14,15] | |
| | Serum albumin | 22 [11] | | | Alpha-internexin | 2 [11] | |
| Extracellular matrix/cell membrane | Agrin | 7 [10] | | Calponin-3 | 5 [13,16] | |
| | Alpha-1-antitrypsin | 18 [11,14] | | | Catenin beta-1 | 3 | |
| | Alpha-2-macroglobulin | 22 [14] | | | Cell division control protein 42 homolog | 2 [14] | |
| | Annexin A2 | 24 [12,14] | | Collapsin response mediator protein 4 long variant | 14 | |
| | Basement membrane-specific heparan sulfate proteoglycan core protein | 25 [10,15] | | | | | |
| | Basigin | 6 [14] | | Cytoplasmic dynein 1 heavy chain 1 | 15 | |
| | Brevican core protein | 2 | | Cytoplasmic dynein 1 light intermediate chain 2 | 2 | |
| | CD44 antigen | 5 [12] | | Cytoskeleton-associated protein 4 | 10 | |
| | CD99 antigen, isoform | 4 | | Differentiation-related gene 1 protein | 2 | |
| | Cell surface glycoprotein MUC18 | 3 | | Dihydropyrimidinase-related protein 2 | 11 [14] | |
| | Chondroitin sulfate proteoglycan 4 | 4 | | Ezrin | 5 [11,13,14] | |
| | Collagen | 43 [12] | | Fascin | 6 [13] | |
| | Complement component C9 | 8 | | Gamma-adducin | 2 [14] | |
| | Erythrocyte band 7 integral membrane protein | 4 [14] | | Glial fibrillary acidic protein | 135 | |
| | Fibronectin | 49 [1,11,12,14,15] | | Keratin, type II cytoskeletal 78 | 14 | |
| | Galectin-3 | 4 [12] | | Lamin-B1 | 11 | |
| | Glypican-1 | 3 [12] | | Microtubule-associated protein | 2 | |
| | Integrin alpha-V light chain | 4 [10,12] | | Microtubule-associated protein 1B | 38 | |
| | Inter-alpha-trypsin inhibitor heavy chain H1 | 8 | | Myosin regulatory light chain 12B | 3 [14,15] | |
| | Inter-alpha-trypsin inhibitor heavy chain H2 | 6 | | Nestin | 27 | [10] |
| | Laminin | 30 [10,12,14] | | Neurofilament light polypeptide Plectin | 80 [14] | |
| | Major histocompatibility complex, class I, C | 2 | | | | |
| | Neuronal membrane glycoprotein M6-a | 2 | | Profilin-2 | 2 | [11] |
| | Nidogen-1 | 10 [10,14] | | Septin-7 | 4 [14] | |
| | Nidogen-2 | 5 [10,12,14] | | Septin-8 | 2 | |
| | Periostin | 16 [10,12,15] | | Spectrin alpha chain, non-erythrocytic | 25 | |
| | Protein MAL2 | 3 | | Spectrin beta chain, brain 1 Synein | 21 | [14] |
| | Reticulon-4 | 6 [14] | | Talin-1 | 3 | |
| | Tenascin | 42 [10,12,15] | | Transgelin-2 | 7 [14,15] | |
| | Thrombospondin-1 | 4 [14,15] | | Tubulin | 106 | [11,14,15,22] | |
| | Thy-1 membrane glycoprotein | 2 | | Vimentin | 155 | [1,11] |
| Classification | Protein name | No. of peptides | References | Classification | Protein name | No. of peptides | References |
|----------------|--------------|-----------------|------------|----------------|--------------|----------------|------------|
| Lipid and fatty acid metabolic process and regulation | Vitronectin | 12 | [14] | Protein folding/chaperone/transport/channel function | 4F2 cell-surface antigen heavy chain | 3 | |
| | 3-Hydroxyacyl-CoA dehydrogenase type-2 | 6 | [14] | | 60 kDa heat shock protein, mitochondrial | 14 | [14] |
| | Acid ceramidase subunit beta 2 | 2 | | | 78 kDa glucose-regulated protein | 29 | [14,15] |
| | Enoyl-CoA hydratase 2 | 2 | [14] | | Annexin A6 | 23 | |
| | Fatty acid synthase | 9 | [14] | | Apolipoprotein E | 6 | |
| | Perilipin-3 | 7 | [14] | | Aquaporin-4 | 5 | [11] |
| | Peroxisomal multifunctional enzyme type 2 | 2 | [14] | | Band 3 anion transport protein | 9 | [10] |
| Mitochondrial transport | ADP/ATP translocase 2 | 10 | | | Clusterin | 15 | [14] |
| | ADP/ATP translocase 3 | 2 | [15] | | Coatomer subunit alpha | 3 | [14] |
| | Phosphate carrier protein, mitochondrial | 4 | [14] | | Collagen-binding protein | 2 | [1,15,16] |
| | Voltage-dependent anion-selective channel protein 1 | 11 | [12,14,15] | | Cytochrome b-245 heavy chain | 2 | |
| | Voltage-dependent anion-selective channel protein 2 | 4 | [14] | | Cytochrome c oxidase subunit 2 | 4 | |
| Metabolic enzymes | 2,3'-Cyclic-nucleotide 3'-phosphodiesterase | 2 | | | Electrogenic sodium bicarbonate cotransporter | 2 | |
| | 26S Protease regulatory subunit 4 | 2 | [14] | | Endoplasmic reticulum resident protein 29 | 3 | |
| | 26S Proteasome non-ATPase regulatory subunit 11 | 2 | | | Excitatory amino acid transporter 1 | 3 | |
| | 26S Proteasome non-ATPase regulatory subunit 3 | 3 | | | Fatty acid-binding protein, brain | 9 | [1,11,13] |
| | 3-Ketoacyl-CoA thiolase | 5 | | | Heat shock cognate 71 kDa protein | 17 | |
| | 6-Phosphogluconolactonase | 2 | [14] | | Heat shock protein HSP 90 | 57 | [12,14,15] |
| | Acetyl-CoA acetyltransferase, mitochondrial | 3 | | | Hsc70-interacting protein | 3 | [14] |
| | Adipocyte plasma membrane-associated protein | 3 | [14] | | Lactotransferrin | 8 | |
| | Alanyl-tRNA synthetase | 4 | [14] | | Prohibitin | 7 | [11,14] |
| | Aldehyde dehydrogenase family 1 member | 2 | [12] | | Regulator complex protein LAMTOR1 | 2 | |
| | Alpha-enolase | 7 | [12,14] | | Serotransferrin | 8 | [14] |
| | Amine oxidase (flavin-containing) B | 9 | [14] | | Sideroflexin 3 | 3 | |
| | ATP synthase | 38 | [11,12,14] | | Solute carrier family 2 (facilitated glucose transporter), member 1 | 3 | |
| | ATP-dependent RNA helicase DDX3Y | 3 | | | Sorting nexin 1, isoform 2 | 2 | |
| | Coagulation factor XIII A chain | 7 | [14] | | Sorting nexin-3 | 2 | [14] |
| | Cystosol aminopeptidase | 3 | | | T-complex protein 1 | 16 | [10,14] |
| | Cystosolic non-specific dipeptidase | 4 | [14] | | V-type proton ATPase subunit B, brain | 2 | |
| | D-3-phosphoglycerate dehydrogenase | 3 | [13] | | 40S ribosomal protein S23 | 6 | [10,14,15] |
| Classification | Protein name | No. of peptides | References | Classification | Protein name | No. of peptides | References |
|----------------|--------------|-----------------|------------|----------------|--------------|----------------|------------|
| Endonuclease domain-containing 1 protein | 4 | [14] | | 60S ribosomal protein | 5 | [14,15] |
| Extracellular signal-regulated kinase-2 splice variant | 4 | | | ADP-ribosylation factor 3 | 2 | |
| Farnesyl pyrophosphate synthase | 3 | | | Alpha-crystallin B chain | 9 | [11,13] |
| Fructose-bisphosphate aldolase A | 20 | [14,15] | | Calnexin | 11 | [14] |
| Glutamate dehydrogenase 1, mitochondrial | 5 | [10,11,14] | | Calreticulin | 5 | [14,15] |
| Glyceraldehyde-3-phosphate dehydrogenase | 15 | [12,14,15] | | Coatomer protein complex, subunit gamma | 5 | |
| Glycogen phosphorylase, brain | 2 | [14] | | Cullin-associated NEDD8-dissociated protein 1 | 4 | [14] |
| Haptoglobin | 3 | [14] | | Dolichyl-diphosphooligosaccharide–protein glycosyltransferase Elongation factor 1-gamma | 4 | [14,15] |
| Isocitrate dehydrogenase (NADP) | 3 | [14] | | Eukaryotic initiation factor 4A-I | 4 | [11,15] |
| L-lactate dehydrogenase | 7 | [14,15] | | Eukaryotic initiation factor 4A-III | 2 | |
| Malate dehydrogenase, mitochondrial | 3 | [14] | | Eukaryotic translation initiation factor 4H | 3 | [14] |
| Methyltransferase-like protein 7A | 2 | | | Heat shock 70 kDa protein 9 | 10 | |
| NAD(P) transhydrogenase, mitochondrial | 6 | [14] | | Isoform 3 of Heterogeneous nuclear ribonucleoprotein Q | 2 | [15] |
| Peptidylprolyl isomerase A (Cyclophilin A) | 3 | | | Protein disulfide-isomerase Ubiquitin-conjugating enzyme E2N | 58 | [11,14,15] |
| Peroxiredoxin-2 | 6 | [14] | | Vesicle-trafficking protein SEC22b | 3 | [14] |
| Phosphofructokinase, platelet | 2 | | | Clathrin heavy chain 1 | 14 | [14,15] |
| Phosphoglycerate kinase | 12 | [14] | | | | |
| Polyadenylate-binding protein 1 | 5 | | | Vesicular trafficking | | |
| Puromycin-sensitive aminopeptidase-like protein | 2 | [14] | | Rab GDP dissociation inhibitor alpha | 6 | [14,15] |
| Pyruvate kinase isozymes M1/M2 | 35 | [14,15] | | Ras-related protein Rab-10 | 3 | [14] |
| Pyruvate kinase | 6 | [12] | | Ras-related protein Rab-2A | 3 | [14] |
| Sarcoplasmic/endooplasmic reticulum calcium ATPase 2 Splicing factor 3A subunit 3 | 7 | [14] | | Secernin 1, isoform 3 | 3 | |
| Thioredoxin-dependent peroxide reductase, mitochondrial | 2 | [14] | | Transitional endoplasmic reticulum ATPase Transmembrane emp24 domain-containing protein 10 | 15 | [14,15] |
| Signal transduction | | | | | | |
| 14-3-3 Protein beta/alpha | 6 | [14] | Other | Vesicle-fusing ATPase | 4 | |
| 14-3-3 Protein epsilon | 26 | [14] | | FACT complex subunit SSRP1 | 3 | |
| 14-3-3 Protein zeta/delta | 4 | [14,15] | | Isoform 4 of Myelin basic protein | 2 | |
| Adenylyl cyclase-associated protein | 4 | [10,12] | | Receptor expression-enhancing protein 5 | 2 | |
and therefore protein identification is measured relatively less often for peptide masses with a low intensity in the MS1 spectrum [3]. The Xcalibur instrument software uses exclusion criteria in a nano-LC MS measurement to exclude selected MS1 parent masses for MS2 fragmentation in order to obtain as many protein identifications as possible (www.thermoscientific.com).

Performance of Peptrix was compared with that of MSInspect [2] and the results were remarkably different, since the software applications differ greatly in their techniques for processing the mass spectra. Peptrix runs on an average computer system with the Windows OS. Peptrix is written in Java and uses around 1 GB of memory with a maximum memory heap size. – Xmx1024 M, setting of the Java executable. Peptrix does not use any statistics to make the peptide profile matrix [1]. Peptrix uses the freely available MS2 sequencing application X!Tandem (http://www.thegpm.org/tandem/) [9] for linking the protein identifications through MS2 peptide sequences to the MS1 peptide masses. The interesting points and results from this new link will be discussed.

Gliomas are among the most vascularized tumors. Therefore, identification of new angiogenesis-related proteins is important for the development of anti-angiogenic therapies [10]. A glioma type brain tumor dataset containing glioma and endometrium control samples is analyzed using Peptrix in this study. To discriminate between physiological and pathological angiogenesis, protein expression profiles of proliferating vessels in glioma are compared with those of endometrium tissue where physiological angiogenesis takes place. The potential protein biomarkers for glioma angiogenesis obtained are compared with proteins that have previously been reported in the literature [11–16]. The Orbitrap analysis results are also compared with FT-ICR MS analysis results from a comparable sample set [1].

Results

Peptrix performance

The processing of the 40 mass spectra in a peptide profile matrix shown in Figure 1 takes a total of 70.5 or 1.75 h per file on average. A peptide profile matrix for peptide masses with a sequence and protein label, and the average peak intensities of the masses in the spectra for the groups GV, GT, NV and NT are created. The peptide profile matrix consists of 24,249 mass-retention time bins, of which 9248 (38%) have a sequence and protein label. A large part of the low intensity peptide masses in the MS1 spectra is not selected for MS2 sequencing or sequencing is not successful. Among the 24,249 peptides, 52% are up-regulated with intensity ratio GV/NV > 1 (+), and 48% were down-regulated with intensity ratio GV/NV < 1 (−). The instrumental coefficient of variance (CV) of the Orbitrap mass spectrometer is 10% for measurements of high intensity peaks of technical replicates [2]. When working with peak lists, peak finding and matching low intensity peaks increase the CV. The average CV of measured peak intensities of technical replicates is 25% for Peptrix [2]. The biological variability of peak intensities is large. The CV of peak intensities when measured for all samples of a group, 7350 times in the 24,249 mass retention time bins of the peptide profile matrix, is about 100% of the mean intensity.

A small selection of proteins that have differentially expressed peptide peak intensities between GV and NV is shown in Table 1. The average spectrum peak intensities for the peptide masses for myosin-9 in the four groups examined are shown in Table S1. The average mass accuracy of all identified peptide masses was 4 ppm compared to the calculated value. The Orbitrap peptide masses are approximately less accurate by a factor of 4–7 than those measured with FT-ICR MS.

The number of unique sequences is 6911 in the Orbitrap peptide profile matrix, which means that sequences appear redundant 1.3 times on average in the matrix. The number of peptide sequences divided by the number of unique sequences is 9248/6911 ≈ 1.3. Majority of the sequences appear once (about 85%) or twice (about 10%), while some sequences appear three times (about 4%) or more (about 1%) (Tables 1 and S1).

There are three reasons why peptide sequences appear in the peptide profile matrix more than once. Firstly, in approximately 40% of cases, the repetition of the sequences is caused by tailing possibly combined with mismatching of the peptide mass with other masses in the elution profile for the nano-LC. For example, the sequence FSVNLDVK for protein CRY-AB_HUMAN (alpha-crystallin B chain) is shown twice in Table S1. The difference in retention time between two bins is 1357 s (7157 – 5842 s), which is > 300 s (5 min) (Table 1). Another example is the peptide mass with the sequence IAQLEEQLD-NETK for myosin-9 in Table S1. The difference in retention time between two bins is 1357 s (7199 – 5842 s), which is > 300 s (5 min) too. Secondly, in approximately 40% of cases, the sequence is present more often in the peptide profile matrix

Table 2

| Classification | Protein name | No. of peptides | References | Classification | Protein name | No. of peptides | References |
|----------------|--------------|-----------------|------------|----------------|--------------|----------------|------------|
| Guanine nucleotide binding protein (G protein), beta polypeptide 2 | STAT1-alpha/beta | 3 | [15] | Vesicle-associated membrane protein-associated protein A | Serine/arginine-rich splicing factor 1 | 5 |

Note: Proteins in bold indicate proteins that have previously been named in the literature as a tumor biomarker, or linked with angiogenesis. The underlined proteins are also up-regulated in GT compared to NT.
through mass modifications of the peptide. The sequence QEEEMMAKEEELVK for myosin-9 is present in three mass retention time bins in Table S1, i.e., with mass 1705.7642, 1721.7544 and 1722.7951 Da. The mass of the peptide without modification is 1722.7951 Da. The first modification of peptide with mass 1705.7642 Da is an N-terminal loss of NH\textsubscript{3} with −17.0265 Da and cyclization of glutamine (Q) [17], while the second modification of peptide with mass 1721.7544 Da is an extra oxidation of methionine (M), which leads to an increase of the mass by 15.999 Da (net change 15.999−17.0265 ≈ −1 Da). Finally, in approximately 20% of cases, the sequence is measured more often because of faults in the determinations or mismatching of the masses of the peptides. The mass difference is slightly greater than 10 ppm and the peptide appears in a different mass retention time bin in the matrix, while it should be present in one bin. The peptide with sequence YSVQTADHR for fascin (Table 1) is such an example. The difference in mass is 17 ppm (10 ppm binning), while the retention time binning is not more than 5 min different, i.e., the bins differ by 5880 − 5760 = 120 s.

The number of unique proteins linked to the peptide profile matrix is 1873. Each protein is identified with approximately 4 (6911/1873) unique peptide sequences. The MS2 protein labels from the peak lists from the individual samples are currently passively matched with the MSI peptide profile matrix in the last step of the process shown in Figure 1. The number of unique peptide sequences in the peak lists from the individual samples is 10,259 and the number of protein labels is 2569. The peptide profile matrix has approximately 67% (6911/10,259) of the peptide sequence and 73% (1873/2569) of the protein information from the peak lists from the individual samples.

Modifications detected by X!Tandem

Compared to the large commercial search engine Mascot, the search engine X!Tandem detects a large number of non-tryptic peptide fragments due to a different search algorithm employed, approximately 11% of the total (Tables 1 and S1). Two of the sequences are the sequence DYEEVGDSVEEGEEEEGEE from tubulin alpha-1A chain split at EY (Table 1) and the sequence KTELEDTLDSAAQQELR from myosin-9 split at LK in Table S1.

Approximately 20% of the sequences have a modification (Tables 1 and S1). In approximately 1/3 of the cases, this involves an N-terminal loss of −NH\textsubscript{3} and cyclization of Q for −17.0265 Da, while in approximately 1/3 of the cases, oxidation of M (+ O) adds +15.999 Da. In addition, in approximately 1/4 of the cases there occurs deamidation (−NH\textsubscript{3} + OH) of asparagine (N) or Q to increase +0.984 Da, and in some cases (the remaining approximately 1/10), acetylation (+COCH\textsubscript{3}) of M or alanine (A) confers an augmentation of +42,0106 Da.

An N-terminal loss of −NH\textsubscript{3} and Q cyclization increases the hydrophobicity of the peptide. Consequently, the retention time of the peptide masses clearly increases by approximately 2200 s (36 min) through this N-terminal loss [17]. For peptide with sequence QAQQERDELAIENSSGK from myosin-9, the increased retention time is 7610 − 5496 = 2114 s and 7610 − 5363 = 2247 s (Table S1). There are two different values due to binning (errors), since the retention times 5496 and 5363 should have been in one mass-retention time bin.

From all observed modifications, only M oxidation is given as an input variable in the graphical user interface (GUI) of Peptrix. Peptrix stores the modifications in the file default_input.xml used by tandem.exe (Figure 1). The other modifications are detected by X!Tandem as standard, when the template default_input.xml is used, which can be downloaded together with tandem.exe in the distribution of X!Tandem. The default_input.xml can be changed according to personal needs. The modifications including N-terminal loss of ammonia and Q cyclization, oxidation of M, acetylation of M or A, as well as which amino acid this concerns, are reported in the file output.xml (Figure 1). The deamidation of N or Q is not detected as such, but is reported as an increase of mass of approximately 1 Da, compared to the theoretical mass.

Selection of candidate biomarkers for glioma angiogenesis

Candidate biomarkers for glioma angiogenesis (Table S2) are selected from the peptide profile matrix based on the following criteria: (1) at least two unique peptide sequences are up-regulated in the GV versus NV group with intensity ratio GV/NV > 1 (+). This results in 597 protein labels, which are about 32% of the total number of 1873 protein labels; (2) no more than 1 of 6 sequences exclusively down regulated for each protein (−), this is 17% of the peptide sequences for each protein. This results in 328 protein labels, which are about 18% of the total number of protein labels; (3) at least one or preferably more peptides with a Wilcoxon–Mann–Whitney P value < 0.1. This results in 235 protein labels (Table S2), which are about 13% of the total number of protein labels.

Apparent down-regulated peptide masses with intensity GV < NV (−) due to tailing in retention time are not considered. In most cases, the peptides of proteins that are up-regulated in GV are also up-regulated with peak intensities GV > NV (+). For some sequences of peptide masses in Table 1, a different pattern can be observed with peptide peak intensity GV < NV (−). This usually concerns sequences that appear more than once in the peptide profile matrix through tailing of the peptide mass in the elution profile from LC, modifications of the peptide or errors in determining the mass. Such examples include the sequences YSVQTADHR from fascin, LTNSQNFDEYMK from fatty acid-binding protein, brain, LGVRPSQGGEAPR and SYTITGLQGTDYK from fibronectin and LDHKFDLMYAKR from tubulin alpha-1A chain.

Proteins can be selected from the list of 235 up-regulated protein labels (Table S2) and proteins that were previously named in the literature as a tumor biomarker or linked with angiogenesis [1,10–16] are displayed in bold in Table 1. The up-regulated proteins in GV were classified according to a scheme set-up [15] and information provided at www.uniprot.org. As presented in Table 2, a large number of proteins are cytoskeleton proteins, involved in cell migration and cell shape or cross linking of actin [1] in bundles, the filopodia [18], such as fascin, cell division control protein 42 homolog and spectrin beta chain, brain 1. Fibronectin (Table 2) is an example of the KEGG hsa04512 ECM–receptor interaction pathway, connecting with cell surface protein integrins, regulating cell-to-ECM and cell-to-cell adhesion. Other members of this pathway including agrin, integrin, laminin, tenascin [10,12,14,15] are also listed in Table 2.
New potential biomarkers for glioma angiogenesis can be selected from the list of up-regulated proteins in Table 2, but need to be validated independently. For example, clusterin was previously reported to be related to multiple sclerosis [19], while excitatory amino acid transporter 1 is perhaps associated with the glutamate dehydrogenase 1, mitochondrial (Accession No. P00367, Table S2) [10,11], since both glutamate transporters and glutamate dehydrogenase play roles in the developing brain [20]. In addition, 3-hydroxyacyl CoA dehydrogenase is an example of a candidate involved in the lipid and fatty acid metabolic process and regulation. It is interesting to note that angiogenesis and metastasis are reduced by the inhibition of fatty acid synthesis with the anti obesity drug Orlistat [21].

In the previous study, the comparison of GT and NT also shows a sharp skewed distribution toward low Wilcoxon–Mann–Whitney P values [2]. A large number of peptides are differentially expressed between GT and NT, these proteins are underlined in Table 2, 69 of the selected 235 protein labels (29%) appear also up-regulated in GT, compared to NT. Differences observed between GV and GT are not presented in Table 2, while no significant differences appear between NV and NT at all [2].

The candidate biomarker Cdc42 effector protein 3 [1] is not found in the peptide profile matrix. Peptide sequences from Cdc42 effector protein 4 (B3KUS7) are however defined in the peak lists of GV. The peak intensity of a peptide mass with sequence TPFLLVGTQIDLR from a related protein Cdc42 homolog (E7ETU3) is up-regulated in GV compared to NV with a factor of 17.2 (Table 1). Myosin-9 [15] is also not present in the list of selected candidates. The considerably lower Wilcoxon–Mann–Whitney P values and greater intensity ratios GV/NV measured by FT-ICR [1] are not observed in the present analysis. The Wilcoxon–Mann–Whitney P values between GV and NV of peptide masses (marked with * in Table S1) remain constantly high at about 0.6 from position 712 up to position 1755 of the primary structure of myosin-9. In addition, annexin A5 [1,12,14] is up-regulated in GV compared to NV, but absent from the list of selected candidates in Table 2, due to lack of peptide mass with Wilcoxon–Mann–Whitney P value < 0.1. Different from an earlier finding [1], desmin (Swiss-Prot Accession code P17661) does not appear to be up-regulated in GV compared to NV, but is instead down-regulated (−) as reported in another study [15]. The down regulation of desmin could be attributed to the use of a different control sample set, proliferating endometrium (representing physiological angiogenesis) instead of the normal control hemispheric brain used in the FT-ICR study. Even so, desmin is related to angiogenic micro vessels and is localized together with vimentin, a marker for pericytes [15,22].

Estimation of the FDR

The false discovery rate (FDR) of a protein is estimated based on how many single peptides of the protein are up- or down regulated.

All proteins mentioned in Table S2 are taken up-regulated and the majority of the proteins have been reported to be associated with tumor growth in the literature (Table 2). The 235 protein labels represent 2133 mass intensity bins in the peptide profile matrix (Table S2), from which 312 appear down-regulated (−) in GV versus NV group. The chance of having a single peptide measured as down-regulated by mistake is estimated as 0.15 (312/2133) for an assumed up-regulated protein.

Since approximately half of the peptides are either up-regulated or down-regulated in the peptide profile matrix, the FDR is therefore set as 0.15. The FDR of protein with two positive (+) peptides is 0.02. The FDR of fascin (Table 1) found with 5 positive (+) and one negative (−) peptide for example is 0.0004 [(0.15)^5 × (0.85) × 6!/(5! × 1!)].

Discussion

Peptrix implements the MS2 sequencing application X!Tandem and detects label-free differentially expressed candidate biomarkers for angiogenesis in a small dataset, by comparing the average peak intensities in combination with the non-parametric Wilcoxon–Mann–Whitney test. In this way, Peptrix is capable of detecting meaningful biomarkers, despite the large biological variability of peak intensities and zero values (peptide peak intensity is below detection limit or peak detection fails). As a result, biomarkers that have been reported previously in the literature are identified.

Peptide profiling from Orbitrap™ MS files results in less pronounced intensity ratios between GV and NV than with previous FT-ICR measurements. There is therefore no sign (Table S1) of the supposed up-regulation of peptide masses on the C-terminus of myosin-9 [1]. The level of up-regulation can now be determined with peptide masses at more positions in the protein, because more peptide masses are measured from myosin-9 through LC separation than those measured by FT-ICR.

The lower mass accuracy of Orbitrap™ MS compared to FTICR MS is compensated by greater number of masses and protein identifications, which can be directly linked to the peptide profile matrix. Through LC separation, the Orbitrap™ peptide profile matrix contains approximately 10 times more bins (24,249/2275) than the FT-ICR peptide profile matrix obtained from a comparable dataset. The signal is averaged over more MS1 measurements than with FT-ICR MS as well. A distorted image of up- or down-regulated peptide masses from a protein is however sometimes created through tailing, modifications, incorrectly-determined masses and mismatching. Peptide masses from up-regulated protein intensity GV > NV (+) in some mass retention time bins can appear down-regulated with intensity GV < NV (−). The peptide matrix contains approximately 70% of the MS2 labels of the individual peak lists together. Loss of MS2 sequencing information occurs when matching the individual peak lists in the last step of the creation of the peptide profile matrix (Figure 1). It is therefore important not to match the MS2 sequencing information passively, but to actively involve it in the creation of the matrix, so that all MS2 sequencing information is retained. This active matching should be implemented while retaining maximal speed of Peptrix and the minimal memory usage of the work station.

Materials and methods

Dataset

A glioma type brain tumor dataset containing glioma and endometrium control samples is analyzed in this study. The dataset consists of 10 micro-dissected tissue samples from
glioma blood vessels (GV), 10 from tissue around these blood vessels (GT), 10 from normal endometrium blood vessels (NV), and 10 from endometrium tissue around these blood vessels (NT). The origins and preparation of the micro-dissected tissue samples are described in [10]. In the present analysis, NV and NT of proliferating endometrium (representing physiological angiogenesis) are used as control samples [10], whereas normal control hemispheric brain samples were used in the FT-ICR study [1,16].

Peptrix label-free peptide profiling software

The Orbitrap™ MS measurements are described previously in [2,10]. Forty *.RAW files exported by the Xcalibur instrument software with an average size of 523 MB are imported in Peptrix for analysis. The Peptrix architecture is described in [23]. The imported *.RAW files are saved on an FTP server. Records with the file names of these *.RAW are created in the Table Sample in a MySQL database (Supplementary File 1). The files are assigned to GV, GT, NV or NT group, respectively, in the Java Swing graphic user interface (GUI) of Peptrix. The links between file name and group code are saved as records in the table Results of the MySQL database (Supplementary File 1). The mass and retention time window for binning the peptide masses in the peptide profile matrix, which is set as 10 ppm and 5 min, respectively, are entered in the GUI. The expected modifications of the peptide masses can also be entered. Only the (fixed) modification carbamidomethyl cysteine (C), mass + C4H9NO 57.022 Da, and (variable) oxidation of M, mass + O 15.999 Da, are currently implemented. The precursor mass tolerance of the parent peptide and MS2 fragment mass tolerance, which are set as 10 ppm and 0.6 Da, respectively, are entered in the GUI.

The processing of the MS files is displayed in Figure 1, which is done by pressing the button once without any further user interaction. Peptrix uses the following applications and files invisibly: (1) R.exe (http://www.r-project.org/) to trace differentially expressed peptides in the groups with the Wilcoxon–Mann–Whitney module; (2) Readw.exe (version 4.3.1, http://sourceforge.net/projects/sashimi/files/) for converting the *.RAW files into *.mzXML files; (3) the *.fasta file and HUMAN.fasta in this study, which is the text database to correlate MS2 fragmentation masses to a protein and (4) tandem.exe to search for the most likely protein. Tandem.exe reads both Mascot generic files, *.MGF, with the peptide parent mass and the MS2 fragmentation masses, and the 38.1 MB HUMAN.fasta file. The HUMAN.fasta file can be downloaded as HUMAN.fasta.gz archive (ftp://ftp.uniprot.org/pub/databases/uniprot/) in the directory current_release/knowledgebase/proteomes/. Peptrix generates a MGF file for each Orbitrap™ file.

At first, a pop-up window is displayed by Peptrix to search for the files and programs in the computer file system: (1) R.exe; (2) Readw.exe (version 4.3.1); (3) *fasta file and (4) tandem.exe. The paths to the files in the file system and file names are saved as records in the Table Itemvalue of the MySQL database (Supplementary File 1). Peptrix once again prompts for the file or the program when the files or programs are not found in a subsequent analysis, for example, because they have been deleted from the computer file system.

Availability and requirements

Peptrix is freely available. It requires Microsoft Windows 2000 OS or higher, R (R-2.15.1-win.exe or higher), Quick ‘n Easy FTP Server Lite version 3.2 or higher, MySQL 5.5.27 (mysql-5.5.27-win32.msi) or higher database, Java Runtime Environment (JRE) 7 Update 7 or higher (jre-7u7-windows-i586.exe), Eclipse Classic (Eclipse Juno 4.2.0) – Windows, edftftp-2.3.0 or higher (edftftp.jar), Connector/J (mysql-connector-java-5.1.22-bin.jar or higher), X!Tandem (tandem-win-11-12-01-1.zip) (tandem.exe), HUMAN.fasta (HUMAN.fasta.gz) and Readw.exe (version 4.3.1) (ReAdW-4.3.1.zip) for running. The source code of Peptrix is available as a zip file (http://sourceforge.net/projects/peptrix/files/), as well as the database script (Supplementary File 1), with detailed installation and running instructions and URIs of the software providers. The raw Orbitrap™ files conversion to mzXML formatted files was tested with Readw.exe version 4.3.1. Because the Readw.exe program depends on Windows-only vendor libraries from Thermo, the code for Orbitrap™ data handling will only work under Windows with Thermo Fischer Scientific’s Xcalibur software installed. If the Readw.exe program does not work properly, zlib1.dll should be downloaded (http://sourceforge.net/projects/peptrix/files/). The library zlib1.dll can be placed in the c:/windows/system32/directory and “regsvr32 c:/windows/system32/zlib1.dll” or “regsvr32 zlib1.dll” can be executed in the Windows command prompt (MSDOS box). If a 64-bit version of Windows is used, zlib1.dll should be copied in C:/Windows/SysWOW64/.

Competing interests

The author has declared that he has no competing interests.

Acknowledgements

John Shippey, BA and Drs. Els Spin from Univerital™ are gratefully thanked for reviewing the text, Dr. Dave Speijer from the Department of Medical Biochemistry, Academic Medical Center, University of Amsterdam, for advice and Prof. Dr. Johan M Kros from the Department of Pathology, Erasmus Medical Center, Rotterdam for providing the sample set. This study was initially financially supported at the Erasmus Medical Center in Rotterdam by the Virgo Consortium (www.virgo.nl) and the EU P-mark project, and finished at the Academic Medical Center, University of Amsterdam.

Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.gpb.2013.02.002.

References

[1] Titulaer MK, Mustafa DA, Siccama I, Konijnenburg M, Burgers PC, Andeweg AC, et al. A software application for comparing large numbers of high resolution MALDI-FTICR MS spectra demonstrated by searching candidate biomarkers for glioma blood vessel formation. BMC Bioinformatics 2008;9:133.
[2] Titulaer MK, de Costa D, Stingl C, Dekker LJ, Sillevis Smitt PA, Luider TM. Label-free peptide profiling of Orbitrap full mass spectra. BMC Res Notes 2011;4:21.

[3] America AH, Cordewener JH. Comparative LC–MS: a landscape of peaks and valleys. Proteomics 2008;8:731–49.

[4] Mortensen P, Gouw JW, Olsen JV, Ong SE, Rigbolt KT, Bunkenborg J, et al. MSQuant, an open source platform for mass spectrometry-based quantitative proteomics. J Proteome Res 2010;9:393–403.

[5] Mueller LN, Brusniaik MY, Mani DR, Aebersold R. An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. J Proteome Res 2008;7:51–61.

[6] Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 2008;26:1367–72.

[7] Luber CA, Cox J, Lauterbach H, Fancke B, Selbach M, Tschopp J, et al. Quantitative proteomics reveals subset-specific viral recognition in dendritic cells. Immunity 2010;32:279–89.

[8] Finney GL, Blackler AR, Hoopmann MR, Canterbury JD, Wu CC, MacCoss MJ. Label-free comparative analysis of proteomics mixtures using chromatographic alignment of high-resolution muLC–MS data. Anal Chem 2008;80:961–71.

[9] Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. Bioinformatics 2004;20:1466–7.

[10] Mustafa DA, Dekker LJ, Stingl C, Kremer A, Stoop M, Sillevis Smitt PA, et al. A proteome comparison between physiological angiogenesis and angiogenesis in glioblastoma. Mol Cell Proteomics 2012;11, M111.008466.

[11] Deighton RF, McGregor R, Kemp J, McCulloch J, Whittle IR. Glioma pathophysiology: insights emerging from proteomics. Brain Pathol 2010;20:691–703.

[12] Li C, Sasaroli D, Chen X, Hu J, Sandaltzopoulos R, Omidi Y, et al. Tumor vascular biomarkers: new opportunities for cancer diagnostics. Cancer Biomark 2010;2011;8:253–71.

[13] Zhang R, Tremblay TL, McDermid A, Thibault P, Stanimirovic D. Identification of differentially expressed proteins in human glioblastoma cell lines and tumors. Glia 2003;42:194–208.

[14] Qureshi AH, Chaoji V, Maiguel D, Faridi MH, Barth CJ, Salem SM, et al. Proteomic and phospho-proteomic profile of human platelets in basal, resting state: insights into integrin signaling. PLoS One 2009;4:e7627.