# Rapid Protein Identification Using AP-MALDI TOF and the Spectrum Mill MS Proteomics Workbench

## Technical Overview

Donghui Yi and Jian Bai
*Agilent Technologies*

### Introduction

Peptide mass fingerprinting (PMF) has been widely embraced as a key methodology for protein identification. With this technique, proteins are digested, peptides are analyzed by mass spectrometry (MS), and the peptide masses are searched versus theoretical digest fragments from protein and DNA databases.

The explosion of research in proteomics has necessitated more rapid, sensitive, accurate analysis of peptides. Atmospheric pressure-matrix assisted laser desorption/ionization (AP-MALDI) time-of-flight (TOF) mass spectrometry addresses this need because AP-MALDI permits rapid, sensitive analysis and TOF provides mass accuracy that exceeds that of other commonly used mass analyzers. In particular, the Agilent AP-MALDI source in combination with the Agilent LC/MSD TOF allows low femtomole and even attomole-level analyses to be accomplished in only minutes with low-ppm mass accuracy.

Because of the large volume of samples that are typically analyzed with AP-MALDI TOF, it is also important to have high-throughput software to take full advantage of the available data. The software needs to accomplish the common data processing steps in a fast and reliable manner. These processing steps include extraction of mass data from the raw data file, PMF searching, and results summary. This overview describes how the Agilent Spectrum Mill MS Proteomics Workbench fulfills this need.

Agilent Technologies

## Experimental

### *Sample preparation*

Tryptic digests of apotransferrin, conalbumin, myoglobin, and peroxidase were purchased from Michrom BioResources, Auburn, CA. The digests (500 pmol) were reconstituted with 500 µL 15% acetonitrile/85% water with 0.1% formic acid. The stock solutions (1 pmol/µL) were split into aliquots (20 µL) and stored at –20°C prior to use.

Digests were analyzed individually and in mixtures. All samples were diluted with matrix solution, which consisted of 0.6 mg/mL α-cyano-4-hydroxy-cinnamic acid (CHCA) in 50% methanol/water with 0.5% acetic acid. For some samples, a mass calibrant of 20 fmol/µL of neurotensin (*m/z* 1672.9170) was added to the matrix solution.

### *AP-MALDI analyses*

All samples were analyzed with an Agilent LC/MSD TOF equipped with an AP-MALDI source. The initial mass axis calibration and tuning were performed with an interchangeable ESI source. With the AP-MALDI source, the protein digests were analyzed at the rate of one per minute. Spectra were recalibrated using *m/z* values of matrix ions and the neurotensin internal standard.

**Instrument Conditions**

| | |
|---|---|
| Instrument: | Agilent LC/MSD TOF with AP-MALDI source |
| Ionization mode: | Positive ion |
| Capillary voltage: | –3200 V |
| Drying gas: | 5 L/min at 325°C |
| Scan range: | 100–3200 |
| Transients/scan: | 40000 |
| Fragmentor: | 300 V |
| Skimmer: | 60 V |
| Octopole RF: | 300 V |
| Averages (scans): | 21 |
| Total acquisition time: | 1.0 min. |
| Laser: | Nitrogen laser, 10 Hz |

**Spectrum Mill Settings**

| | |
|---|---|
| Instrument: | MALDI MSD TOF |
| Search: | PMF Search |
| Database: | SwissProt |
| Enzyme: | Trypsin |
| Cys: | Carboxymethylation |
| Max missed cleavages: | 1 |
| N-terminus: | Hydrogen |
| C-terminus: | Free Acid |
| Peptide mass tolerance: | 10 ppm |

## Results

Before processing data with the Spectrum Mill MS proteomics workbench, the mass accuracy of the AP-MALDI TOF was verified by comparison of theoretical and experimental masses for a tryptic digest of 20 fmol serotransferrin. The spectrum was recalibrated as described above, using the Analyst QS Software for Agilent TOF. The results, shown in Table 1, indicate excellent mass accuracy, with a mean error of –0.98 ppm, and a standard deviation of 3 ppm.

The Spectrum Mill MS proteomics workbench was then used to convert the raw data files to identified proteins. The Spectrum Mill workbench data processing consisted of extraction of a mass list file from the raw data file, PMF searching, and results summary and review.

### *Data Extraction*

Once the raw data files were copied to the Spectrum Mill server, the Spectrum Mill data extractor was used to mine the mass peaks from the data files. The form for extraction of AP-MALDI TOF data, shown in Figure 1, requires input of only a few simple parameters. Then the software extracts the mass peaks in a completely automated fashion. A single mouse click initiates extraction of an entire folder of data files (for example, all data from a MALDI sample plate). For AP-MALDI TOF data, the extractor averages and centroids the spectra from the raw data files. No further preparation is necessary since the PMF search de-isotopes, deconvolutes, filters the spectra by signal-to-noise, and accomplishes the background subtraction.

**Table 1. Mass errors for tryptic digest of 20 fmol serotransferrin**

| Theoretical mass | Measured mass | Delta (ppm) |
|---|---|---|
| 888.399251 | 888.3997 | 0.51 |
| 936.500951 | 936.5020 | 1.12 |
| 1016.494151 | 1016.4942 | 0.05 |
| 1064.595951 | 1064.5954 | –0.52 |
| 1074.540051 | 1074.5361 | –3.68 |
| 1097.505051 | 1097.5061 | 0.96 |
| 1122.578951 | 1122.5791 | 0.13 |
| 1157.515651 | 1157.5164 | 0.65 |
| 1216.603051 | 1216.5992 | –3.17 |
| 1305.679751 | 1305.6761 | –2.80 |
| 1311.653951 | 1311.6527 | –0.95 |
| 1336.681751 | 1336.6831 | 1.01 |
| 1347.599751 | 1347.6000 | 0.18 |
| 1363.692651 | 1363.6906 | –1.50 |
| 1389.675751 | 1389.6741 | –1.19 |
| 1397.643951 | 1397.6505 | 4.69 |
| 1448.643451 | 1448.6555 | 8.32 |
| 1464.776651 | 1464.7696 | –4.81 |
| 1466.643451 | 1466.6466 | 2.15 |
| 1483.684551 | 1483.6852 | 0.44 |
| 1511.727151 | 1511.7267 | –0.30 |
| 1594.738351 | 1594.7338 | –2.85 |
| 1604.806751 | 1604.8022 | –2.84 |
| 1640.766351 | 1640.7616 | –2.90 |
| 1645.695151 | 1645.6938 | –0.82 |
| 1732.901651 | 1732.8939 | –4.47 |
| 1757.860551 | 1757.8528 | –4.41 |
| 1768.861251 | 1768.8590 | –1.27 |
| 1940.924251 | 1940.9154 | –4.56 |
| 1996.784051 | 1996.7846 | 0.27 |
| 2017.920951 | 2017.9153 | –2.80 |
| 2263.963451 | 2263.9701 | 2.94 |
| 2275.974651 | 2275.9687 | –2.61 |
| 2411.100651 | 2411.0806 | –8.32 |
| | **Mean** | **–0.98** |
| | **Standard deviation** | **3.05** |

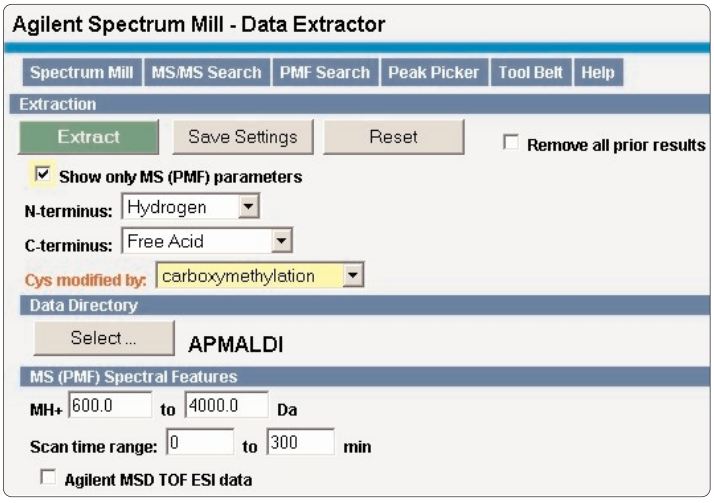


**Figure 1. Data extractor form set up for AP-MALDI TOF data**

### *PMF search*

The PMF search form, shown in Figure 2, illustrates typical settings. For the analyses shown below, extracted mass peaks were searched against the SwissProt protein database. A list of contaminant ions was automatically subtracted prior to searching. Mixture scoring was invoked so that both individual hits and mixtures could be scored and compared using a common set of probability scores. Again, all data files in the selected folder were processed with a single mouse click.

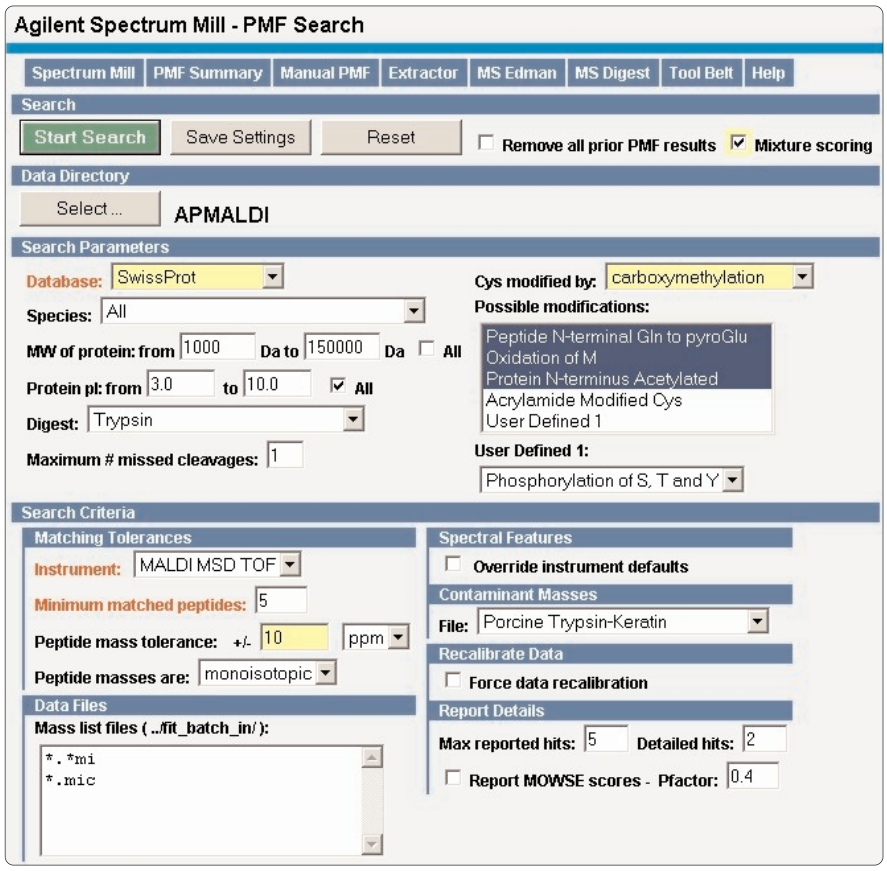


**Figure 2. PMF search form set up for AP-MALDI TOF data**

### *Results summary*

A number of mixtures were summarized using the default settings on the Spectrum Mill PMF summary page shown in Figure 3. Typical summary results are shown in Figures 4 and 5. Figure 4 shows PMF search results for a mixture of apotransferrin and myoglobin. The protein names (highlighted area 1) indicate that the correct proteins were identified. The mixture score (highlighted area 2) is quite good; the large negative

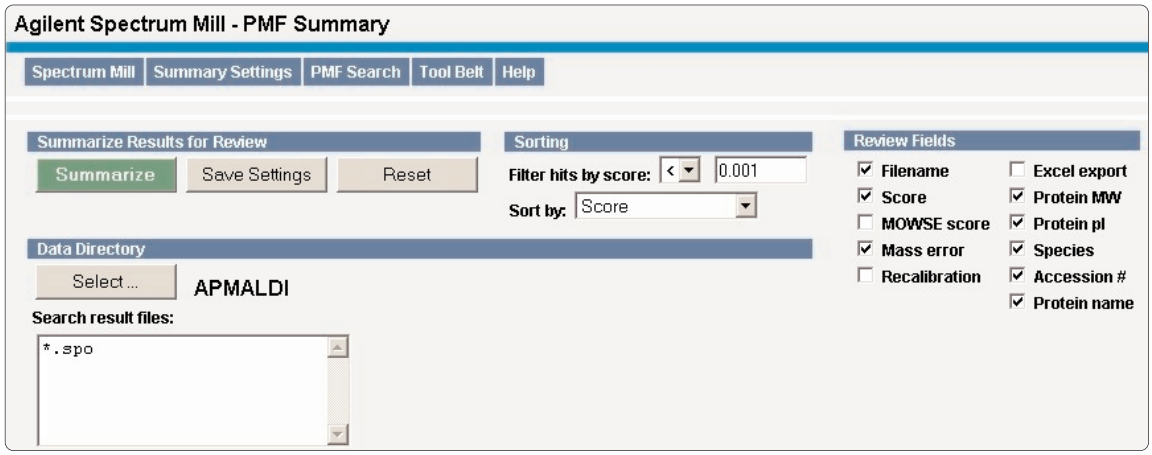


**Figure 3. PMF summary form**

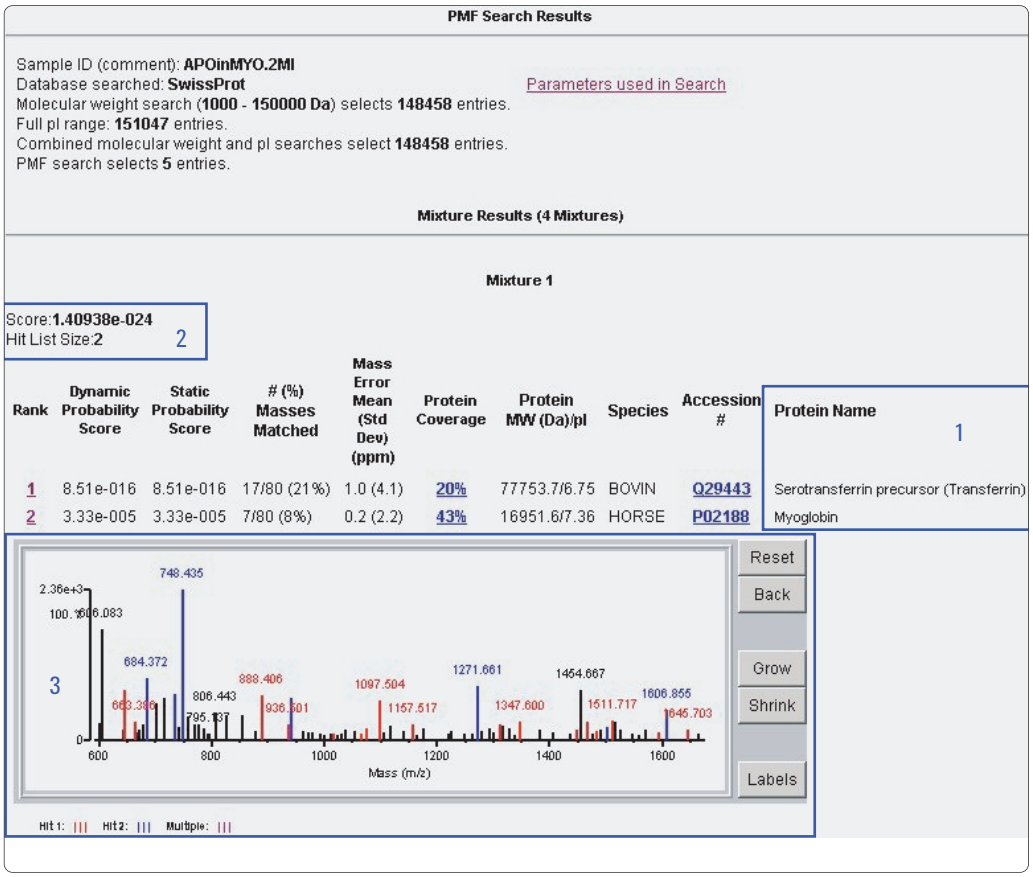


**Figure 4. PMF search results for AP-MALDI TOF analysis of 5 fmol apotransferrin digest in the presence of 100 fmol myoglobin digest, yielding correct identifications**

exponent indicates that this result has a very low probability of chance occurrence. Note that the mixture score is better than the scores of either of the individual components; the fact that these scores can be easily compared provides increased confidence that the sample is indeed a mixture. The spectrum (highlighted area 3) shows the transferrin peaks in red and the myoglobin peaks in blue, with unmatched peaks in black.

Figure 5 shows PMF search results for a mixture of four protein digests, three of which were readily identified. The fourth protein, myoglobin, was likely missed because with MALDI, smaller proteins such as myoglobin generate fewer peptides for detection.

For rapid results review, the Spectrum Mill workbench incorporates both high-level PMF summary results and more detailed results, with easy switch between the two. Figure 6 shows more detailed results for one of the components of the four-component digest mixture. The top shows the matched peptides, all with very small mass errors (see **Delta ppm** column). The bottom shows the sequence coverage in red.

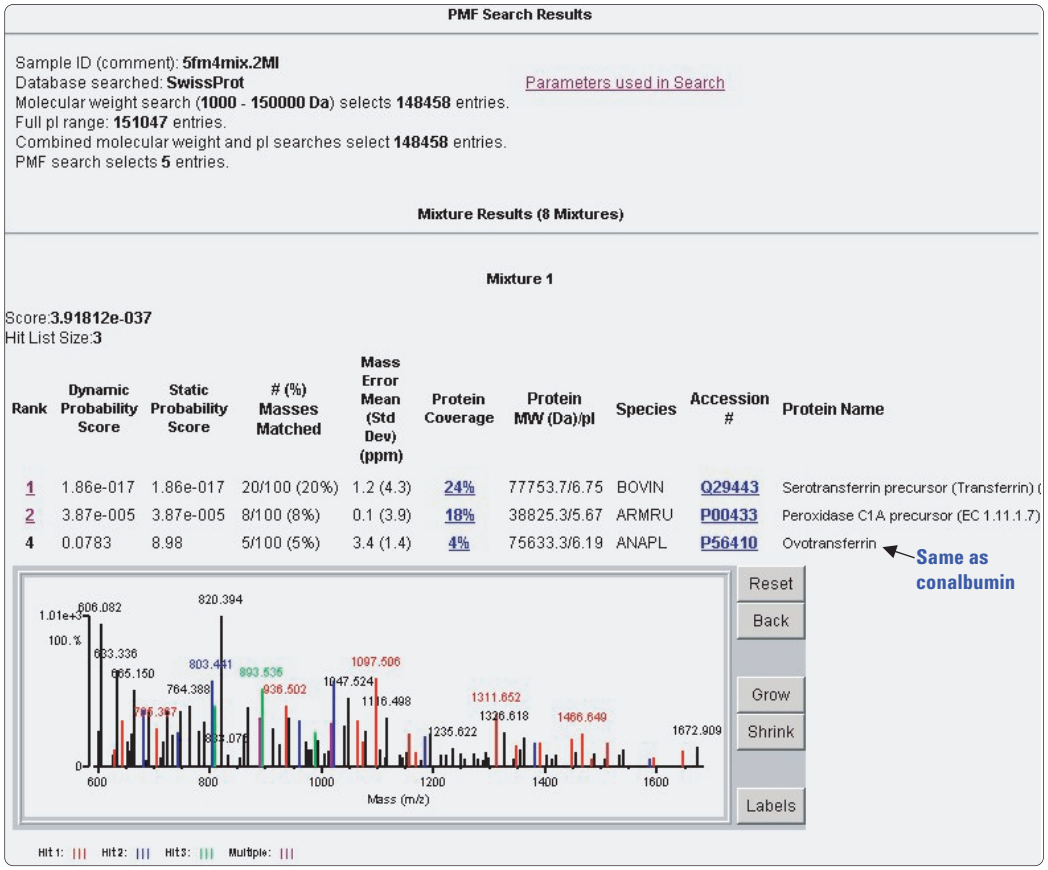


**Figure 5. PMF search results for a mixture of four protein digests (apotransferrin, conalbumin, myoglobin, and peroxidase), 5 fmol each. Three of the four proteins were identified**

**Detailed Results**

**1.** 20/100 matches (20%). BOVIN. Serotransferrin precursor (Transferrin) (Siderophilin) (Beta-1-metal binding globulin) ( 77753.7 Da) pI = 6.75

| m/z submitted | MH+ matched | Delta ppm | Score Counted | Tolerance Bin Index (0-1) | start | end | Peptide Sequence (counted in score) | Modifications |
|---|---|---|---|---|---|---|---|---|
| 630.3570 | 630.3575 | -0.8 | 1 | **0** | 43 | 47 | **(R)ENVLR(I)** | |
| 644.3726 | 644.3731 | -0.9 | 1 | **0** | 127 | 131 | **(K)LNELR(G)** | |
| 705.3674 | 705.3718 | -6.2 | 1 | **0** | 683 | 688 | **(R)AMTNLR(Q)** | |
| 888.4046 | 888.3998 | 5.4 | 1 | **1** | 135 | 142 | **(K)SCHTGLGR(S)** | |
| 936.5024 | 936.5015 | 0.9 | 1 | **1** | 608 | 616 | **(R)GPNHAVVSR(K)** | |
| 1016.4985 | 1016.4947 | 3.7 | 1 | **1** | 134 | 142 | **(K)KSCHTGLGR(S)** | |
| 1064.5981 | 1064.5965 | 1.5 | 1 | **1** | 608 | 617 | **(R)GPNHAVVSRK(D)** | |
| 1074.5096 | 1074.5002 | 8.7 | 1 | **1** | 471 | 479 | **(K)KSCHTAVDR(T)** | |
| 1097.5063 | 1097.5056 | 0.6 | 1 | **1** | 540 | 548 | **(R)YYGYTGAFR(C)** | |
| 1157.5197 | 1157.5162 | 3.0 | 1 | **1** | 366 | 374 | **(K)WCAIGHQER(T)** | |
| 1167.5798 | 1167.5720 | 6.7 | 1 | **1** | 582 | 590 | **(K)KENFEVLCK(D)** | |
| 1311.6518 | 1311.6545 | -2.0 | 1 | **0** | 156 | 166 | **(K)ELPDPQESIQR(A)** | |
| 1347.6050 | 1347.6003 | 3.5 | 1 | **1** | 27 | 37 | **(R)WCTISTHEANK(C)** | |
| 1389.6650 | 1389.6763 | -8.1 | 1 | **0** | 457 | 468 | **(K)TSDANINWNNLK(D)** | |
| 1448.6429 | 1448.6440 | -0.7 | 1 | **0** | 527 | 539 | **(K)GTGKECVPNSNER(Y)** | |
| 1466.6488 | 1466.6440 | 3.3 | 1 | **1** | 671 | 682 | **(K)TYDSYLGDDYVR(A)** | |
| 1483.6755 | 1483.6851 | -6.5 | 1 | **0** | 244 | 255 | **(R)KNYELLCGDNTR(K)** | |
| 1511.7356 | 1511.7277 | 5.3 | 1 | **1** | 595 | 607 | **(R)KPVTDAENCHLAR(G)** | |
| 1594.7452 | 1594.7389 | 3.9 | 1 | **1** | 670 | 682 | **(K)KTYDSYLGDDYVR(A)** | |
| 1645.6994 | 1645.6957 | 2.3 | 1 | **1** | 499 | 513 | **(K)FDEFFSAGCAPGSPR(N)** | |
| | Mean: | 1.2 | | | | | | |
| | Std dev: | 4.3 | | | | | | |

```
1    MRPAVRALLA CAVLGLCLAD PERTVPWCTI STHEANKCAS FRENVLRILE SGPFVSCVKK TSHMDCIKAI SNNEADAVTL 80
81   DGGLVYEAGL KPNNLKPVVA EFHGTKDNPQ THYYAVAVVK KDTDFKLNEL RGKKSCHTGL GRSAGWNIPM AKLYKELPDP 160
161  QESIQRAAAN FFSASCVPCA DQSSFPKLCQ LCAGKGTDKC ACSNHEPYFG YSGAFKCLME CAGDVAFVKH STVFDNLPNP 240
241  EDRKNYELLC GDNTRKSVDD YQECYLAMVP SHAVVARTVG GKEDVIWELL NHAQEHFGKD KPDNFQLFQS PHGKDLLFKD 320
321  SADGFLKIPS KMDFELYLGY EYVTALQNLR ESKPPDSSKD ECMVKWCAIG HQERTKCDRW SGFSGGAIEC ETAENTEECI 400
401  AKIMKGEADA MSLDGGYLYI AGKCGLVPVL AENYKTEGES CKNTPEKGYL AVAVVKTSDA NINWNNLKDK KSCHTAVDRT 480
481  AGWNIPMGLL YSKINHCKFD EFFSAGCAPG SPRNSSLCAL CIGSEKGTGK ECVPNSNERY YGYTGAFRCL VEKGDVAFVK 560
561  DQTVIQNTDG NNNEAWAKNL KKENFEVLCK DGTRKPVTDA ENCHLARGPN HAVVSRKDKA TCVEKILNKQ QDDFGKSVTD 640
641  CTSNFCLFQS NSKDLLFRDD TKCLASIAKK TYDSYLGDDY VRAMTNLRQC STSKLLEACT FHKP                  704
```

**Figure 6. Detailed PMF search results for 5 fmol apotransferrin analyzed as part of a four-component digest mixture**

## Conclusions

The Spectrum Mill workbench enables fast, reliable processing for AP-MALDI TOF data. The software automates the preparation of multiple data files for PMF searching, as well as the PMF search itself. For rapid results review, the software features a results summary that provides both overview and detailed results. The Spectrum Mill workbench delivers reliable results for fast, high-throughput protein analysis with AP-MALDI TOF.

## Authors

***Donghui Yi*** and ***Jian Bai*** are scientists at Agilent Technologies in Santa Clara, California U.S.A.

**www.agilent.com/chem**



Printed in the U.S.A. October 22, 2004
5989-1778EN

**Agilent Technologies**