The Effect of MS/MS Fragment Ion Mass Accuracy on Peptide Identification in Shotgun Proteomics

Authors
David M. Horn
Christine A. Miller
Bryan D. Miller

Agilent Technologies
Santa Clara, CA

Abstract

It is widely accepted that high precursor ion mass measurement accuracy significantly improves confidence in MS/MS database search results through the reduction of false positives. In contrast, the impact of high fragment ion mass accuracy has been largely ignored. The Agilent Accurate Mass Q-TOF, a recent advancement in MS instrumentation, now enables high-throughput accurate mass measurement of both precursor and fragment ion data. This study examines the effects of precursor and fragment ion mass accuracy on protein identification using the Accurate Mass Q-TOF. Results demonstrate that fragment ion mass accuracy is as important (if not more important) than precursor ion mass accuracy in reducing false positive identification rates in MS/MS database searches, and plays a critical role in improving the throughput of proteomics research projects.

Introduction

Perhaps the most fundamental goal of proteomics research is the identification and characterization of proteins, particularly in the context of cellular expression. The shotgun technology of tandem mass spectrometry paired with advanced liquid chromatography has emerged as the standard technique for high-throughput protein identification (Liu, J. *et al.*, 2007; VerBerkmoes, N.C. *et al.*, 2002). To perform a typical shotgun experiment, a sample is first fractionated and the resulting mixture of proteins is then digested into peptides by an enzyme such as trypsin. The peptide mixture is then separated by high performance liquid chromatography (HPLC) and subsequently analyzed by mass spectrometry to determine the mass/charge (*m/z*) ratio of each peptide. Peptides of interest are selected for further fragmentation in a collision cell to produce tandem (MS/MS) mass spectra. This resultant data is then searched against a protein database to identify the peptide sequences and further infer the protein content of the sample.

It is well known that high mass accuracy measurement of precursor ions significantly improves confidence in MS/MS protein database search results. However, the impact of high mass accuracy measurement of fragment ions has been largely ignored. Up until recently, MS instruments that possessed sufficient throughput capability to capture both precursor and fragment ion data with high mass accuracy were not readily available. The advent of the Agilent Accurate Mass Q-TOF LC/MS provides the sensitivity, speed, and accurate mass capabilities to enable the assessment of precursor and fragment ion mass tolerance as the maximum mass deviation, or MMD (Zubarev, R. and Mann, M., 2007). The Accurate Mass Q-TOF LC/MS quickly captures information from complex samples with attomole-level sensitivity, a wide in-scan dynamic range that covers 3.5 orders of magnitude, and a fast acquisition rate (20 spectra/sec) at up to 15,000 resolving power.

Here, we apply the Q-TOF LC/MS system and demonstrate that fragment ion mass accuracy can be at least as significant as precursor ion mass accuracy in reducing false positive identification rates in MS/MS protein database searches. In addition, we compare false positive rates generated by the mass tolerance parameters particular to other instrument types used for protein identification research.

Materials and Methods

The experimental workflow for this analysis used a 3100 OFFGEL Fractionator to fractionate the HeLa sample, a 1200 HPLC-Chip/6510 Q-TOF LC/MS to acquire the data, and Spectrum Mill Protein Identification software to perform analysis of the collected data and assess the MMD for HeLa cells **(Figure 1)**.

HeLa Cell Culture

HeLa S3 cells were grown in Ham's F12 medium (F12K) with 2 mM L-glutamine and 1.5 g/L sodium bicarbonate supplemented with 10% fetal bovine serum until 90% confluency. Cells were washed in PBS, lysed in hypotonic solution containing protease inhibitors cocktail, then homogenized with a Dounce homogenizer. Soluble HeLa lysate was prepared by centrifugation (1 hr at 16,000 rpm at 4°C). An aliquot (~300 µg total protein) of the supernatant was digested with trypsin using a 2,2,2-trifluoroethanol (TFE) based protocol for solubilization and denaturation. The digest was dried to remove the TFE and ammonium bicarbonate.

OFFGEL Fractionation

The trypsinized lysate was fractionated into 23 fractions in the pH range from 3 to 10. The current was limited to 50 µA and fractionation was stopped after 50 kVh (about 24 hr).

LC/MS Analysis

Approximately 5% of each fraction (5 µL) was analyzed using a microfluidics based HPLC-Chip connected to an Agilent 6510 Quadrupole Time-of-Flight (Q-TOF) mass spectrometer or an Agilent 6330 Ion Trap mass spectrometer.

HPLC-Chip: Protein ID chip with 150 x 0.075 mm analytical column and 160 nL enrichment column. Sample load: 5 µL of fractions obtained from OFFGEL. Flow: 300 nL/min analytical pump, 4 µL/min loading pump. Mobile phases A: 0.1% formic acid (FA), B: 90% acetonitrile, 0.1% FA. Gradients: 3%B to 6%B at 2 min, then 10% B at 10 min, 30%B at 65 min, 40%B at 75 min, 80%B at 80 min until 85 min, then 3%B at 85.1 min. Stoptime: 90 min and post time 10 min.

Ion Trap MS Conditions: Drying gas: 4 L/min, 300°C; Skim 1: 30 V; Capillary exit: 75 V; Trap Drive: 85; Averages: 1; ICC: On; Max. Accumulation time: 50 ms; Smart Target: 500,000; MS Scan range: 300-1800; Ultra Scan; MS/MS Scan range: 100-2000; Ultra Scan.

Q-TOF MS Conditions: Drying gas: 4 L/min, 300°C; Skimmer: 65 V; Fragmentor: 175 V; Collision energy: slope 3.7 V, offset 2.5 V; MS Scan range and rate: 300-2000 at 3 Hz; MS/MS scan range and rate: 50-3000 at 3Hz. AutoMS/MS: 8 precursors, active exclusion on with 1 repeat and release after 0.17 min. Preferred charge state: 2, 3, >3, unknown.

Database Searches

Protein database searches were performed with Spectrum Mill Protein Identification software. All searches used the IPI Human Version 3.28 database with trypsin specificity, 2 missed cleavages, 50% minimum scored peak intensity, and dynamic peak thresholding. The "forward" database search of the ion trap data was performed using 2.5 Da precursor and 0.7 Da fragment ion tolerance while the Q-TOF search used 10 ppm precursor and 40 ppm fragment ion tolerance. Protein identifications were validated manually. A reversed version of

the IPI Human database was created using a Perl script. Searches against this reversed database were performed using a variety of different precursor and product ion mass tolerances. The false positive rates for a given Spectrum Mill score and MMD regime were estimated by comparison of the distribution of Spectrum Mill scores for the forward Q-TOF search and reversed database search results. All results were exported to Microsoft Excel and database search scores were divided into 0.5 unit bins and plotted.

Results and Discussion

Selection of MMD values for Q-TOF data

As a first step in this study, it was necessary to determine the optimal mass tolerances of the 6510 Q-TOF to use for the subsequent protein database search. Using various precursor and product ion mass tolerances, a series of forward database searches were performed. The search results that produced the largest number of validated protein identifications used 10 ppm precursor ion and 40 ppm fragment ion tolerance.

Figure 1. The experimental workflow consisting of a 3100 OFFGEL Fractionator, a 1200 HPLC-Chip/6510 Q-TOF LC/MS, and Spectrum Mill software for data analysis. As part of this study, the performance of the Q-TOF was compared with the 6330 Ion Trap mass spectrometer.

Figure 2 shows the distribution of precursor and fragment ion mass errors for the validated peptides. These results indicate that fragment ion mass tolerance displays a wider range than that of the precursor due to significantly reduced signal-to-noise ratios. However, the higher abundance fragments demonstrate similarly high mass accuracy and improvements to the MS/MS search algorithm could take this into account to improve the results.

Comparison of ion trap and Q-TOF results

Having established the optimal mass tolerance values for the Q-TOF, we next compared the performance of the Q-TOF with the 6330 Ion Trap mass spectrometer, an instrument that is commonly used for protein identification research. Both instruments analyzed aliquots of the same HeLa sample. **Table 1** displays the results generated from both instruments. In the ion trap, there were a number of very high scoring random protein hits. Further, the number of spectra with "random" peptide identifications was more than 10-fold for the ion trap compared to the Q-TOF system. This is presumably partly due to the higher precursor MMD of the ion trap, but also potentially due to the effect of high fragment ion tolerance as well. Also, the total number of MS/MS acquired on the Q-TOF was lower by 90,000 spectra, largely because the high resolution of the Q-TOF better distinguished real peptide signal from background noise. Finally, a greater number of validated peptide identifications were found in the Q-TOF data.

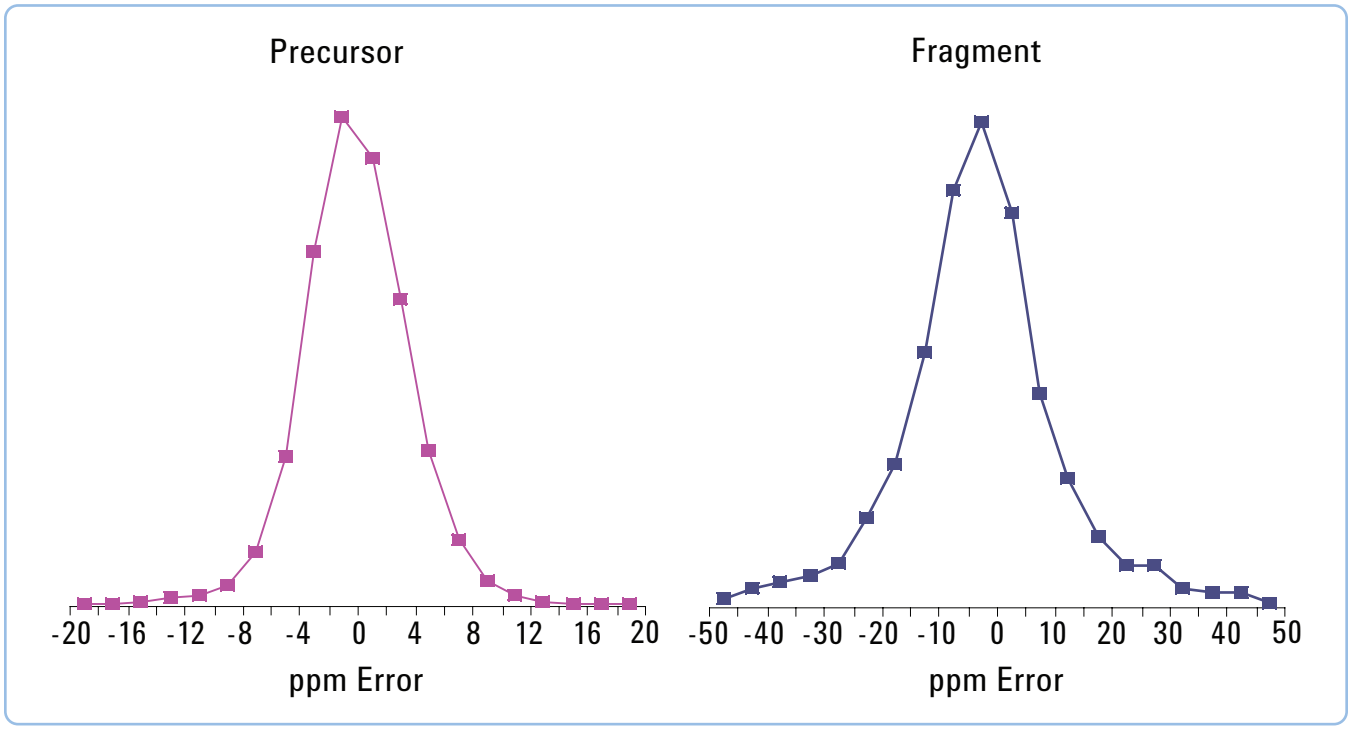

Figure 2. Mass measurement deviation distributions for precursor (3.9 rms ppm) and fragment ions (13.8 rms ppm) using the 6510 Q-TOF.

This combination of factors led to a significantly higher identification yield for the Q-TOF compared to the ion trap. Besides the higher mass accuracy, other factors that might have contributed to the high number of valid spectra generated by the Q-TOF include the ability to assign charge state for the fragment ions and the general cleanliness of the Q-TOF MS/MS data, especially when compared to the ion trap data for MS/MS acquired and the number of extracted files. Regardless, the reduction in matches to random proteins exhibited by the data obtained from the Q-TOF enables a vast improvement in confidence related to peptide identification.

Precursor and fragment ion mass tolerance

Based on the results of these preliminary studies indicating the significance of mass accuracy on the ability to identify peptides, we next proceeded to determine the contributing effects of precursor and fragment ion mass tolerances on confident protein identification. Past research has indicated that high stringency in the precursor ion tolerance can dramatically reduce the rate of false positive identifications for a given search (Liu, T. *et al.*, 2007). To demonstrate this effect, a series of reversed database searches were performed holding the fragment ion mass tolerance constant at 40 ppm (the established threshold for the Q-TOF data obtained in this

	Ion Trap	*Q-TOF*
Total number spectra acquired	355,136	363,692
MS/MS acquired	271,507	186,049
Number of extracted files (.pkl)	92,081	129,752
Number of search result files (.spo)	72,028	20,871
Number of validated spectra	10,618	16,072
Number of validated unique peptides	4151	6253
Number of protein IDs	873	994
Number of unvalidated spectra	61,410	4799

Table 1. Comparison between 6330 Ion Trap and 6510 Q-TOF search results and hits.

study) while varying the precursor ion tolerance. It is known that reversed database searches are a good method for determining the false positive rate, because searching a reversed database would generate a number of matches that would be known to be false (Elias, J.E. and Gygi, S.P., 2007). As precursor ion tolerance decreased below the fragment ion tolerance of 40 ppm, the false positive rate decreased as expected **(Figure 3)**. Interestingly, as the precursor ion tolerance was increased above the fragment ion tolerance, there was little change from 50 to 200 ppm and only a slight increase with 500 and 1000 ppm. The increase in the number of false positives for 500 and 1000 ppm may be attributed to matches of peptides with 1 Da higher and lower nominal mass. These results suggested that while precursor ion mass tolerance most certainly does affect the false positive rate in database searching, the effect of a low fragment ion mass tolerance is also an important factor in limiting the false positive rate.

To determine the significance of low fragment ion mass tolerance on false positive rates, a series of reversed database searches were performed holding the precursor mass

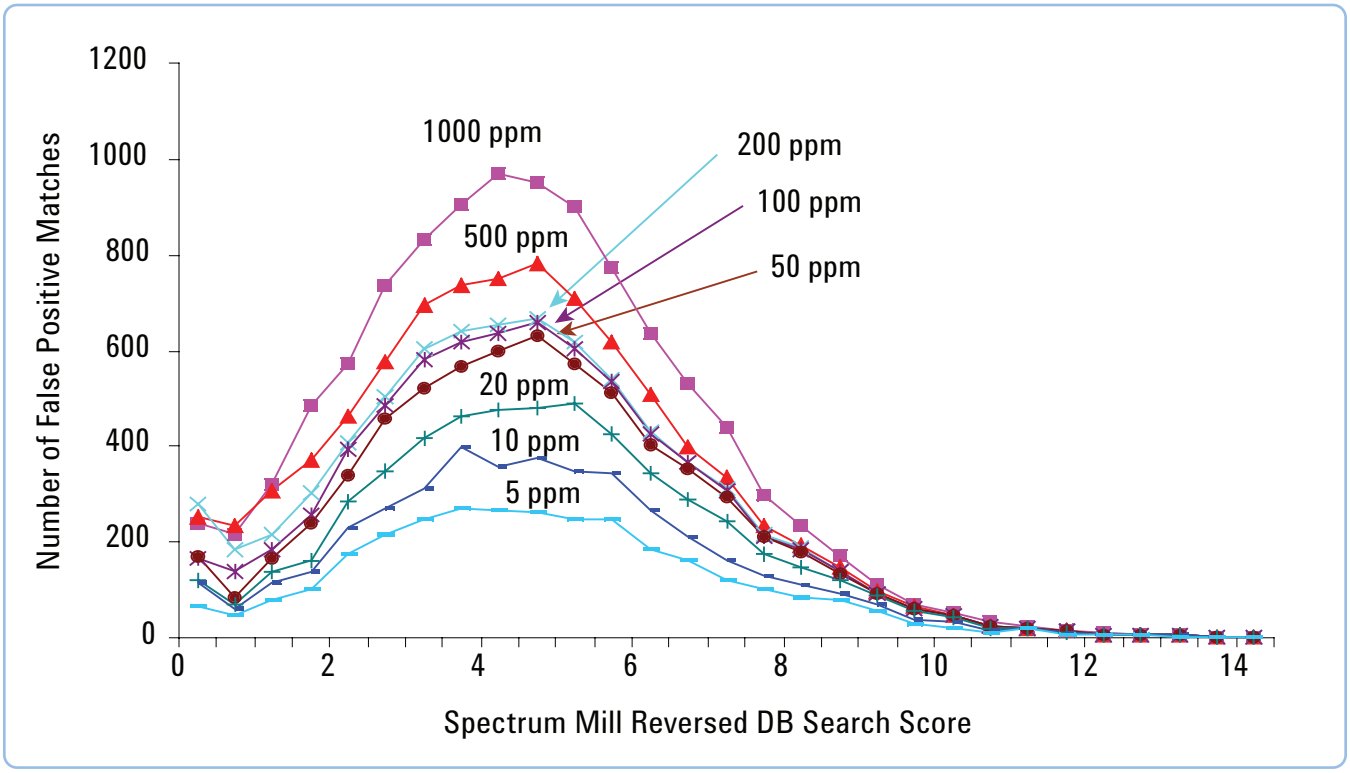

Figure 3. Spectrum Mill score distributions of reversed database hits for variable precursor tolerance at constant 40 ppm fragment tolerance using the 6510 Q-TOF.

tolerance at 10 ppm and varying the fragment mass tolerance **(Figure 4)**. In contrast to results obtained when the fragment mass tolerance was held constant, the results of this experiment showed that the number of false positives continued to increase as the fragment ion mass tolerance increased. Importantly, the number of false positive matches decreased quite dramatically as the fragment ion mass tolerance was decreased, indicating that high precursor ion mass accuracy is not sufficient to mitigate the effect of large fragment mass measurement errors.

We next examined the effects of mass accuracy on database search efficacy more closely by performing reversed database searches with 6510 Q-TOF data using different sets of precursor and fragment mass accuracy settings that represent three different MS instruments—an ion trap, a high-resolution

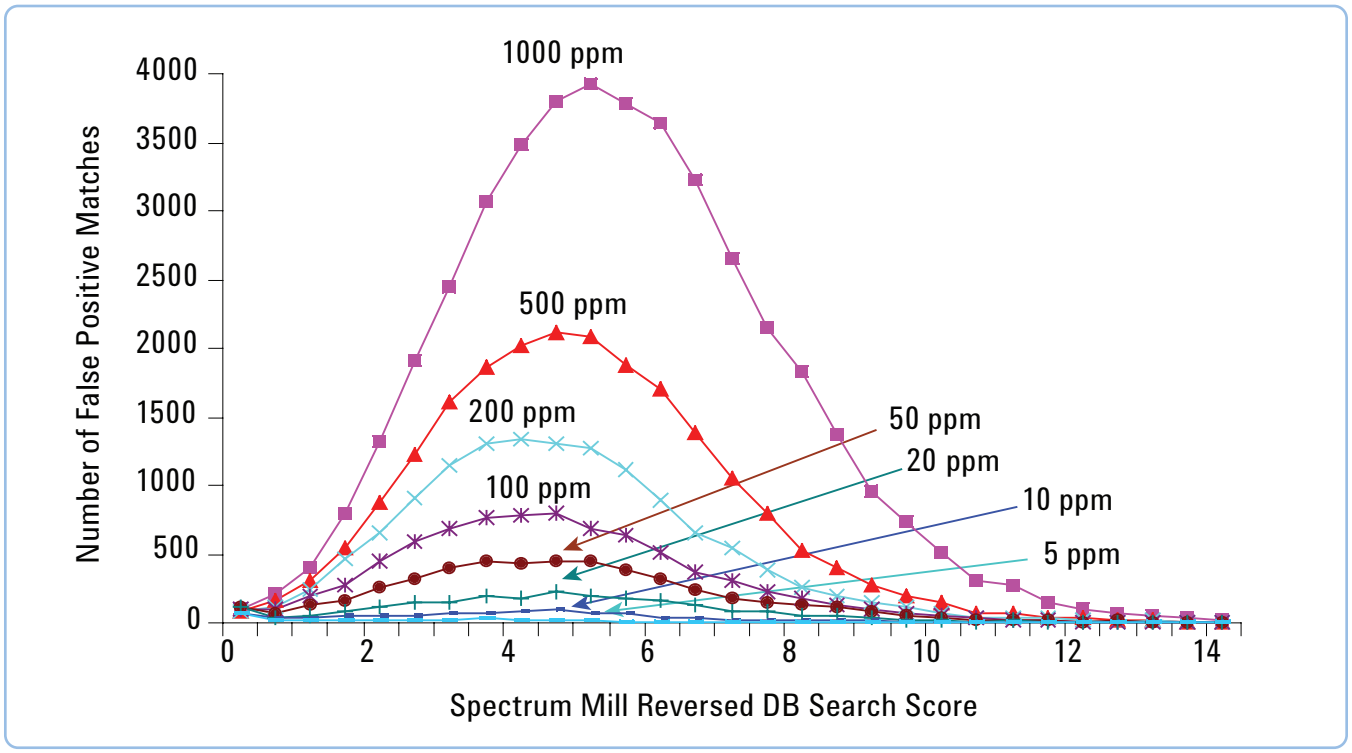

Figure 4. Spectrum Mill score distributions of reversed database hits for variable fragment tolerance at constant 10 ppm precursor tolerance using the 6510 Q-TOF.

hybrid linear ion trap, and the Q-TOF itself **(Figure 5)**. As expected, when the precursor ion mass tolerance was improved from 1000 ppm (ion trap) to 5 ppm (high-resolution hybrid linear ion trap), the distribution of false positive matches decreased proportionally as well. However, when the fragment tolerance was lowered from 1000 ppm (ion trap & high-resolution hybrid linear ion trap) to 40 ppm (Q-TOF), this led to a substantial decrease in the score distribution of random matches. These results indicate that the decrease in false positive identification was attributable to not just the precursor ion mass tolerance but also just as strongly to the product ion mass tolerance.

In the final part of this study, we examined how both the precursor and fragment ion mass tolerances could ultimately affect confidence in protein identification. Using the Q-TOF data generated at 10 ppm precursor MMD and 40 ppm fragment MMD, the reverse score distributions and the

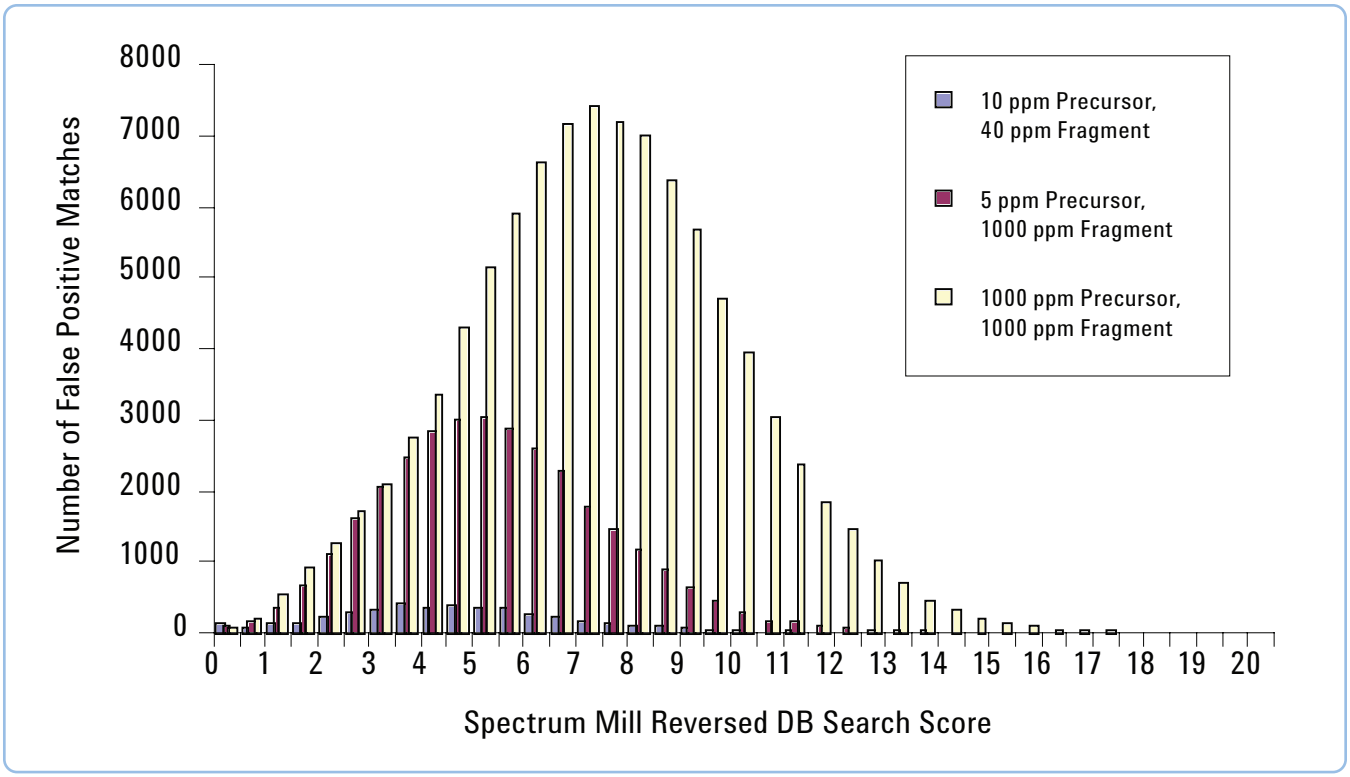

Figure 5. Spectrum Mill score distributions of reversed database hits using common precursor and fragment mass accuracy parameters on the 6510 Q-TOF.

validated protein identifications from the forward database search were compared **(Figure 6)**. A false positive rate was estimated for each scoring bin by comparing the number of random matches in the reversed database search to the number of confidently identified peptide matches. The results indicated that high mass accuracy led to a downward shift in the score distribution of the reversed database matches relative to the forward search. As a result, the Spectrum Mill score required for a particular confidence level was lowered, leading to an increase in the number of confident hits. In fact, the number of confident matches increased up to 6.5-fold with the use of both highly accurate precursor and fragment ion mass measurements on the 6510 Q-TOF **(Table 2)**.

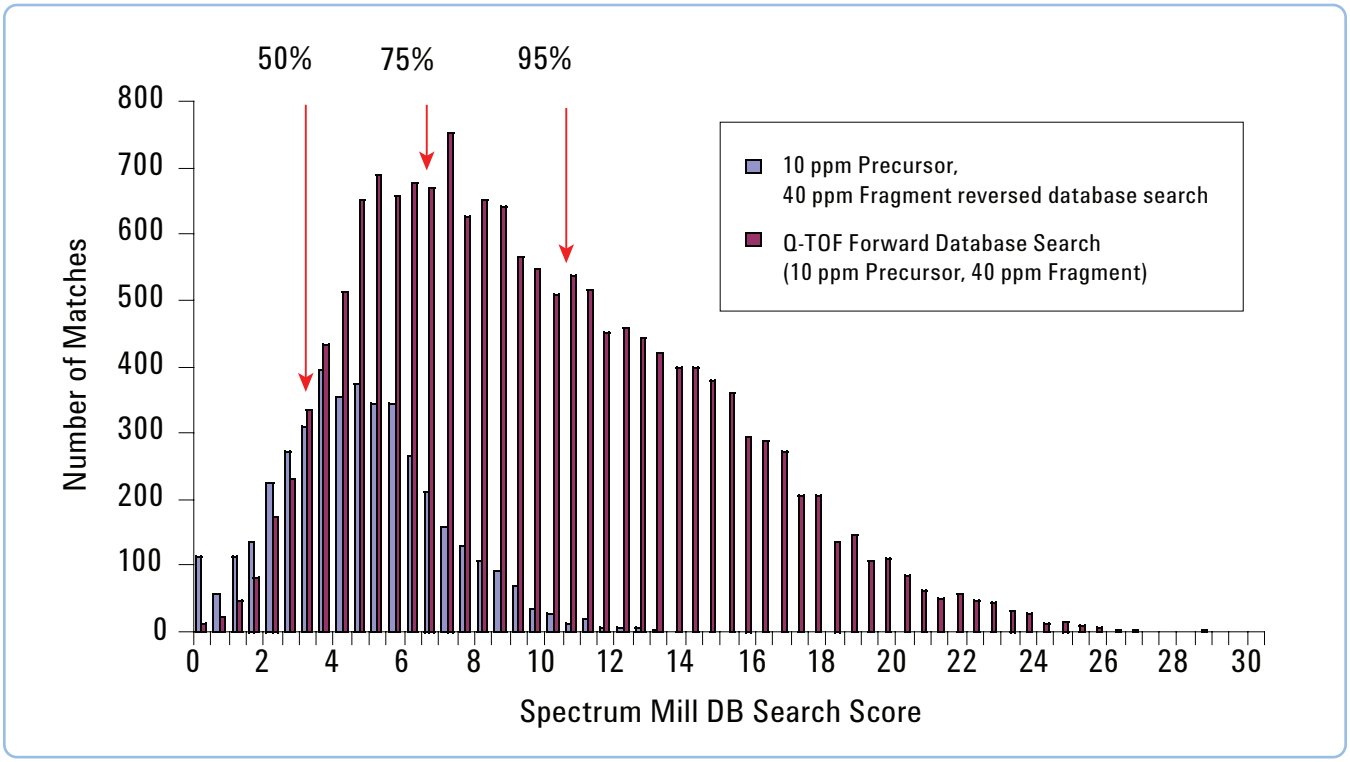

Figure 6. Calculation of peptide identification confidence by comparison of forward and reversed search results of 6510 Q-TOF data. The Spectrum Mill score required for a particular confidence level (denoted at the top of the chart) was lowered, leading to an increase in the number of confident hits.

Table 2 also shows the number of protein matches with 2 or more unique peptide identifications with at least 95%, 75%, or 50% confidence. Using both high-accuracy precursor and fragment ion mass measurements, the 6510 Q-TOF demonstrated roughly a 5-fold improvement in the number of proteins identified with confidence (95% confidence interval). As a result, a much larger proportion of sample identifications can be automatically validated, greatly increasing the automated throughput of proteomics experiments.

Conclusions

The results of this study indicate that fragment ion mass accuracy is an important component to accurate protein identification. Overall, fragment ion mass accuracy was found to render a significant impact on false positive rates. When the fragment tolerance was lowered from 1000 ppm to 40 ppm, a substantial decrease in the score distribution of random matches in database searches was realized. The positive effect of low fragment mass tolerance has at least two explanations. First, large fragments generated by MS/MS of a peptide will exhibit a similar distribution of masses for a given nominal mass as the precursor. Thus, lowering the fragment tolerance will have a similar effect as lowering the precursor tolerance for these fragments, with the result of limiting the distribution of potential database matches. Second, a wider fragment mass tolerance will greatly increase the chance that random peaks in the MS/MS spectrum will be matched against random entries in the database. Decreasing the fragment tolerance will reduce the chance of such random matches.

These data demonstrate that high fragment ion mass accuracy can significantly increase valid protein identification by reducing the number of false positives when searching large databases for proteins, and consequently decreasing the score required for a high-confidence match. Other protein database search engines, such as Mascot, that can accommodate accurate mass data would be expected to yield similar results. In addition, these data demonstrate that the Agilent Accurate Mass Q-TOF LC/MS provides high mass accuracy for both the precursor and product ions, while generating MS/MS spectra at a rate that is well matched to complex sample LC/MS, unlike other high-resolution hybrid linear ion trap instruments.

ppm Tolerance		*Number of Valid Peptide Matches with % Confidence*			*Number of Valid Protein Matches with % Confidence* (minimum of 2 or more unique peptides/protein)		
Precursor	**Fragment**	**>95%**	**>75%**	**>50%**	**>95%**	**>75%**	**>50%**
1000	1000	954	2220	3361	102	213	282
5	1000	4181	6590	7645	338	478	524
10	40	6590	11552	15509	478	728	829

Table 2. The number of confident hits for three different combinations of precursor and fragment mass accuracy settings on the 6510 Q-TOF.

References

1. Elias, J.E., S.P. Gygi (2007). **Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry**. *Nat Methods*, 4(3):207-14.
2. Liu, J., A.W. Bell, J.J. Bergeron, C.M. Yanofsky, B. Carrillo, C.E. Beaudrie, R.E. Kearney (2007) **Methods for peptide identification by spectral comparison**. *Proteome Sci,* 5:3.
3. Liu, T., M.E. Belov, N. Jaitly, W.J. Qian, R.D. Smith (2007) **Accurate Mass Measurements in Proteomics**. *Chem. Rev.*, 107:3621-3653.
4. VerBerkmoes, N.C., J.L. Bundy, J.L. Hause, K.G. Asano, J. Razumovskaya, F. Larimer, R.L. Hettich and J.L. Stephenson (2002) **Integrating top-down and bottom-up mass spectrometric approaches for proteomic analysis of Shewanella oneidensis**. *J of Proteom Res*, 1(3):239-252.
6. Zubarev, R., and M. Mann (2007) **On the proper use of mass accuracy in proteomics**. *Mol Cell Proteomics*, 6(3):377-81.

About Agilent Technologies

Agilent Technologies is a leading supplier of life science research systems that enable scientists to understand complex biological processes, determine disease mechanisms, and speed drug discovery. Engineered for sensitivity, reproducibility, and workflow productivity, Agilent's life science solutions include instrumentation, microfluidics, software, microarrays, consumables, and services for genomics, proteomics, and metabolomics applications.

Buy online:
www.agilent.com/chem/store

Find an Agilent customer center in your country:
www.agilent.com/chem/contactus

U.S. and Canada
1-800-227-9770
agilent_inquiries@agilent.com

Asia Pacific
adinquiry_aplsca@agilent.com

Europe
marcom_center@agilent.com

Printed in the U.S.A. January 10, 2008

5989-7820EN