Waters™
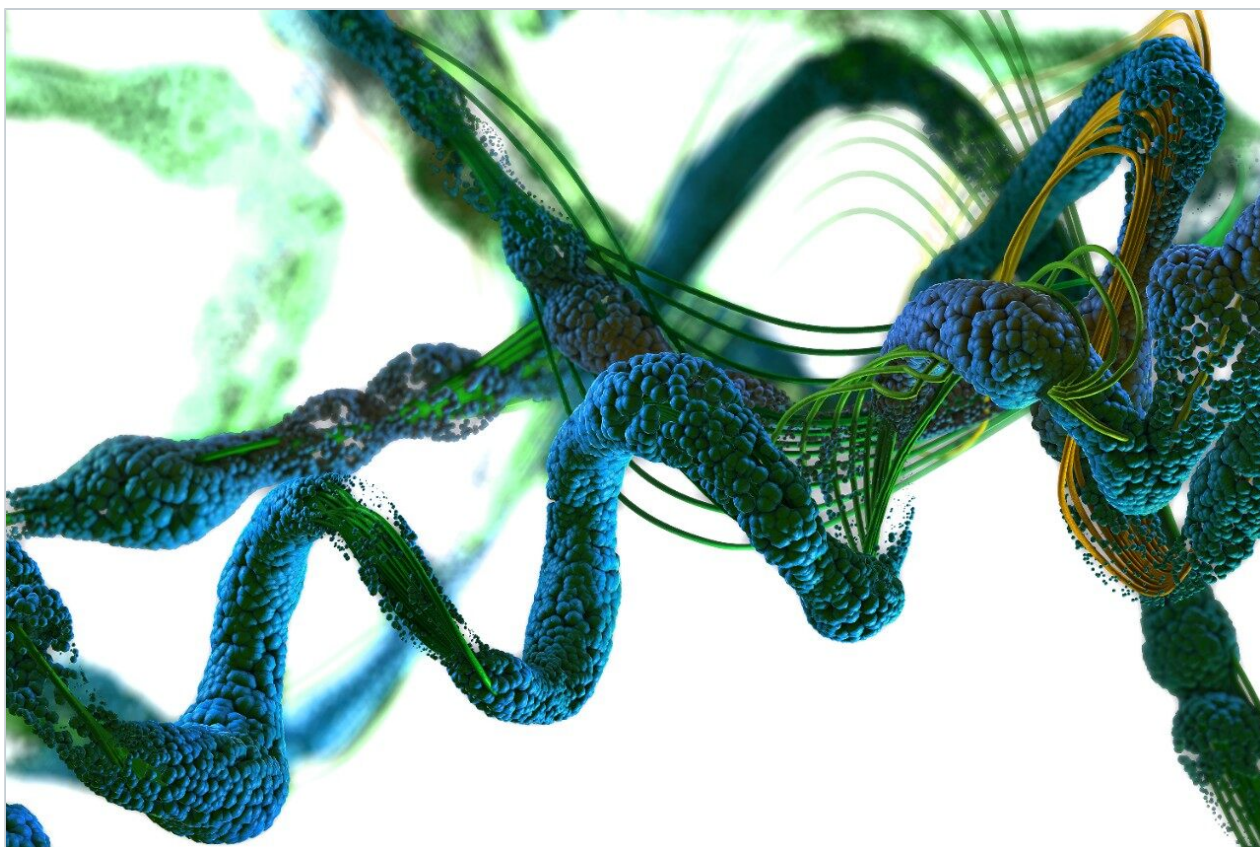
The Use of Exact Mass and High Resolution Precursor Ion Selection on a Q-Tof Ultima MALDI for *de novo* Peptide Sequencing

Marten F. Snel, James I. Langridge

Waters Corporation

Abstract

· The aim of the analysis was the identification by MALDI MS/MS of two unidentified peptides, that differ in mass by only 3 Daltons in the MALDI MS experiment

· The experiment involved MS/MS analysis of tryptically digested peptides on a MALDI quadrupole orthogonal acceleration time-of-flight mass spectrometer

· High resolution precursor ion selection was achieved using a quadrupole mass filter, allowing MS/MS to be performed on peptides of similar mass

· Confident *de novo* sequencing of the peptides was performed using Waters MassSeq algorithm using exact mass measured MS/MS data

Introduction

MALDI MS/MS is a powerful technique for the unambiguous identification and characterization of proteins. MS/MS is often used in conjunction with peptide mass fingerprinting to confirm the protein identification and characterize peptides not matched in the MS spectrum. When peptides are separated by only a few Daltons and their isotopic peak envelopes overlap, it is important in an MS/MS experiment to be able to set the precursor window of MS1, to transmit only the ions of interest and not isotopes from other peptides. Peptides with a similar *m/z* can often be observed with large proteins, or mixtures of proteins that, after enzymatic digestion, produce large numbers of peptides—some of which are likely to be similar in mass. If the precursor ion transmission window is too large, more than one ion will be transmitted and the MS/MS spectra will contain fragment ions from these other peptide species. This makes interpretation extremely difficult.

Peptides that do not match to protein sequences identified by peptide mass fingerprinting, are often chemically modified, contain amino acid substitutions or were produced by non-specific cleavage during the digestion process. Alternatively, unmatched peptides may occur when a mixture of proteins is present, some of which are at low levels or produce only a few tryptic peptides. Peptide mass fingerprinting normally relies on at least five matching peptides to confidently identify a protein.

A different approach that can be applied is to fragment the unidentified peptide and compare the MS/MS fragment ion data with a sequence databank. One advantage of this approach is that proteins can often be

identified from a single peptide. This is possible due to the specificity of exact mass MS/MS sequence information.

If a protein hit is still not identified then *de novo* sequence information can be obtained and compared with protein sequences in a databank using homology-based BLAST searching.

In this application note we highlight the need for accurate precursor ion selection windows, using as an example, two peptides from a tryptic digest of glycogen phosphorylase B that are not matched by peptide mass fingerprinting. These peptides at mass 1530 Da are only separated by 3 Da. We also discuss the advantages of having exact mass measurement for *de novo* sequencing of tryptic peptides from singly charged MALDI MS/MS data.

Waters Micromass Q-Tof Ultima MALDI Mass Spectrometer.

Experimental

Sample Preparation

Equal volumes of aqueous 1 pmol/µL solution of glycogen phosphorylase B (rabbit) and 2 µg/µL alpha-cyano-4-hydroxycinnamic acid matrix solution (1:1 acetonitrile:0.1% trifluoroacetic acid) were mixed. A 1 µL aliquot of the resultant mixture was spotted onto a stainless steel MALDI target plate.

Mass Spectrometry

All data were acquired on a Waters Micromass Q-Tof Ultima MALDI Mass Spectrometer in positive ion mode. The instrument was calibrated with a mixture of PEG 200, 600, 1000 and 2000. Each spectrum was lock mass corrected, from a near point sample well, containing (Glu1)-fibrinopeptide B.

Data Processing

ProteinLynx Global SERVER v2.0 Bio-informatics Software was used for all database searching. Data were converted to XML format and searched against the Swiss-Prot Database v40. MS/MS spectra were initially deisotoped using MaxEnt 3. *De novo* sequence analysis was performed using MassSeq Software.

Results and Discussion

The tryptic digest was confidently identified as glycogen phosphorylase B by peptide mass fingerprinting using ProteinLynx Global SERVER v2.0 (Figure 1). However, not all peaks in the MS spectrum were matched to peptides from glycogen phosphorylase B of particular interest were the unmatched peaks at 1530.871 Da and 1533.786 Da, which were separated by 3 Da and had overlappinsg isotope envelopes (Figure 2).

Workflow Results (1 proteins)

P00489 Glycogen phospho

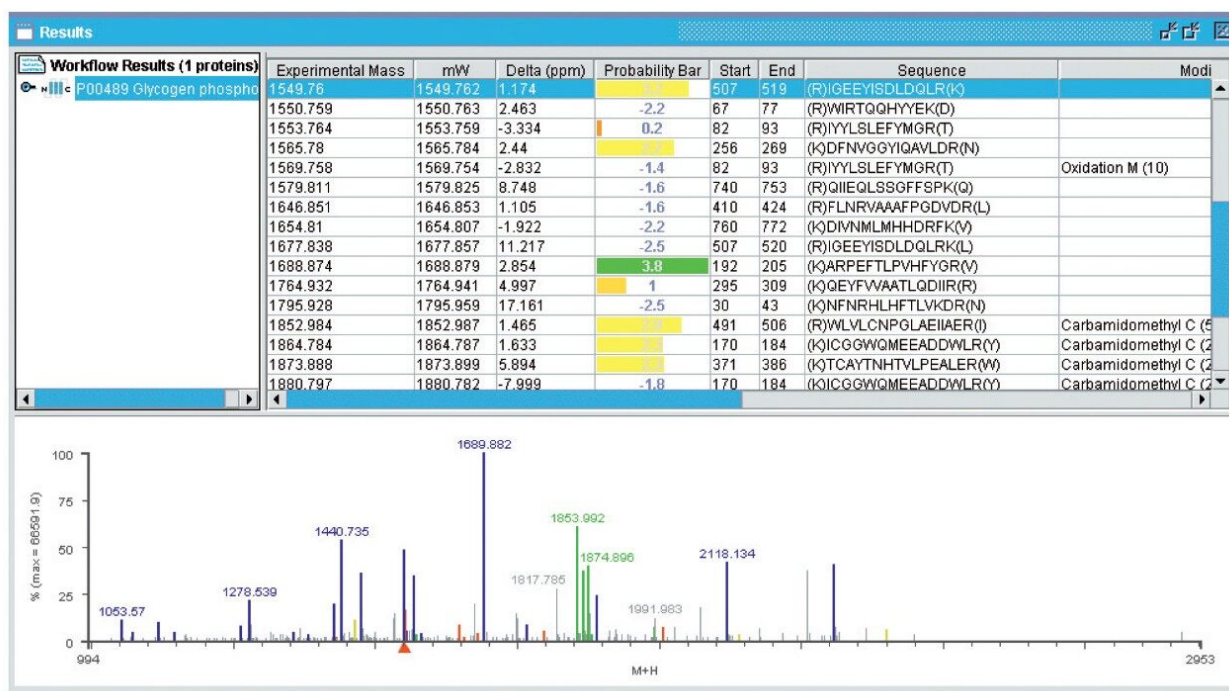| Experimental Mass	mW	Delta (ppm)	Probability Bar	Start	End	Sequence	Modi
1549.76	1549.762	1.174		507	519	(R)IGEEYISDLDQLR(K)	
1550.759	1550.763	2.463	-2.2	67	77	(R)WIRTQQHYYEK(D)	
1553.764	1553.759	-3.334	0.2	82	93	(R)IYYLSLEFYMGR(T)	
1565.78	1565.784	2.44		256	269	(K)DFNVGGYIQAVLDR(N)	
1569.758	1569.754	-2.832	-1.4	82	93	(R)IYYLSLEFYMGR(T)	Oxidation M (10)
1579.811	1579.825	8.748	-1.6	740	753	(R)QIIEQLSSGFFSPK(Q)	
1646.851	1646.853	1.105	-1.6	410	424	(R)FLNRVAAAFPGDVDR(L)	
1654.81	1654.807	-1.922	-2.2	760	772	(K)DIVNMLMHHDRFK(V)	
1677.838	1677.857	11.217	-2.5	507	520	(R)IGEEYISDLDQLRK(L)	
1688.874	1688.879	2.854	3.8	192	205	(K)ARPEFTLPVHFYGR(V)	
1764.932	1764.941	4.997	1	295	309	(K)QEYFVVAATLQDIIR(R)	
1795.928	1795.959	17.161	-2.5	30	43	(K)NFNRHLHFTLVKDR(N)	
1852.984	1852.987	1.465		491	506	(R)WLVLCNPGLAEIIAER(I)	Carbamidomethyl C (5
1864.784	1864.787	1.633		170	184	(K)ICGGWQMEEADDWLR(Y)	Carbamidomethyl C (2
1873.888	1873.899	5.894		371	386	(K)TCAYTNHTVLPEALER(W)	Carbamidomethyl C (2
1880.797	1880.782	-7.999	-1.8	170	184	(K)ICGGWQMEEADDWLR(Y)	Carbamidomethyl C (2

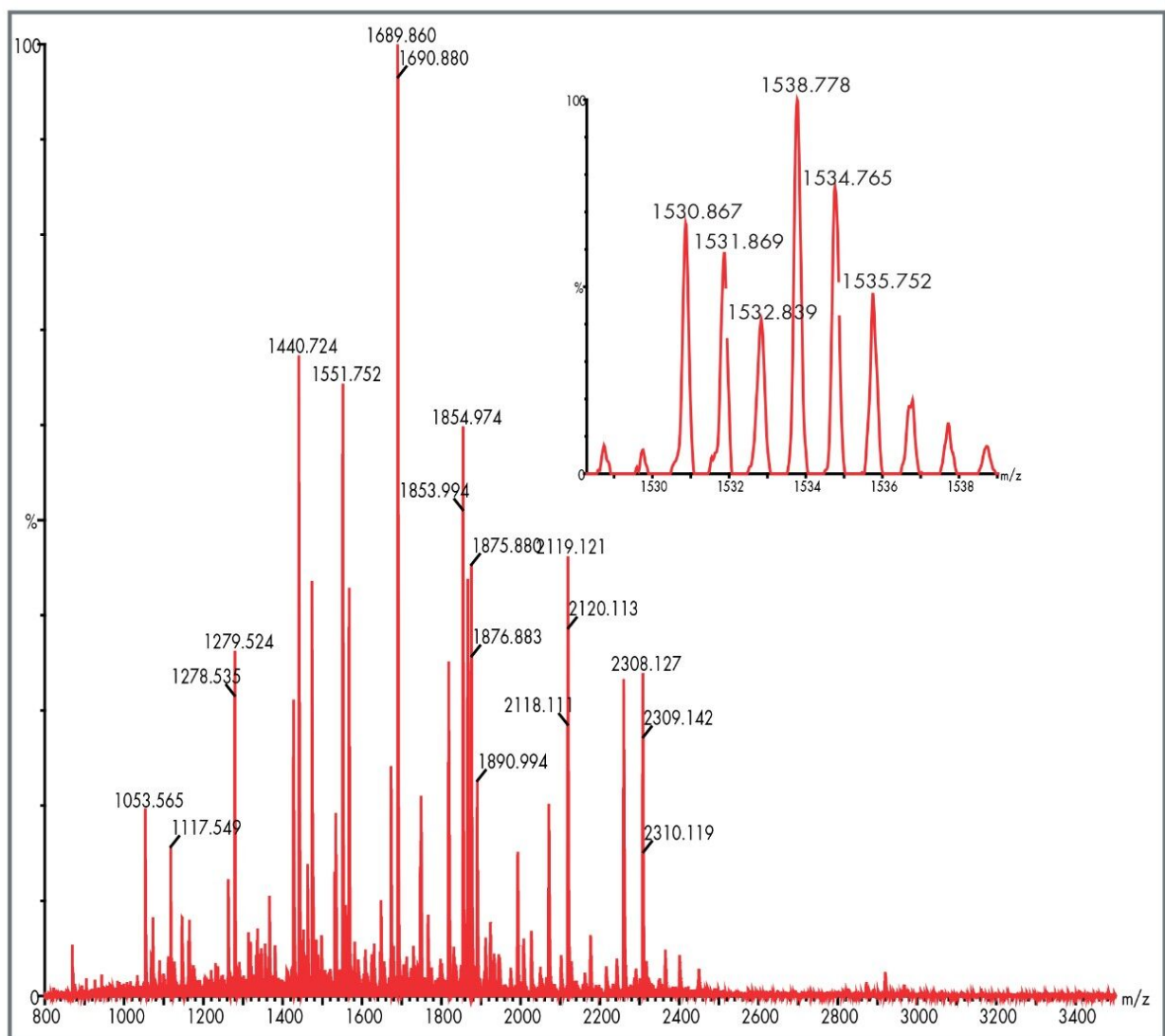Figure 1. ProteinLynx Global SERVER v2.0 database search result.

Figure 2. MS spectrum of glycogen phosphorylase B. The insert shows the separation of the unmatched peptides at 1530 and 1533 Da.

The MS/MS spectrum of the peak at 1530.871 Da was acquired with quadrupole resolution settings allowing selection of a 3 Da window, avoiding transmission of the peak at 1533.786 Da. The MS/MS spectrum acquired was subsequently *de novo* sequenced using MassSeq. The resulting sequence generated (Figure 3) was then searched against the Swiss-Prot Database using the BLAST algorithm contained as an integral part of ProteinLynx Global SERVER v2.0.
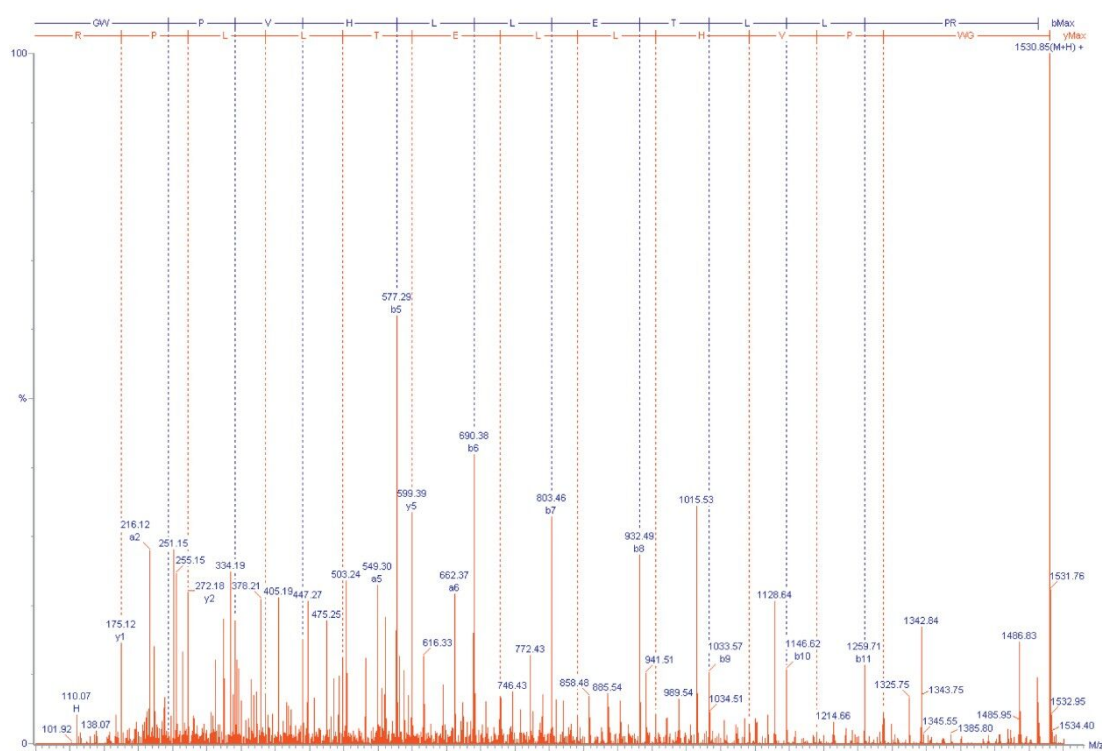
Figure 3. MassSeq de novo sequence of the MS/MS spectrum of the peptide at 1530.871 Da.

In Table 1 the sequences generated from the MALDI MS/MS data by MassSeq and from the Swiss-Prot entry for glycogen phosphorylase B are presented. The homologous part of the amino acid sequence from the experimental data and the glycogen phosphorylase B databank entry is highlighted in blue. All of the amino acid sequence information matched to the sequence from the database with the exception of the N-terminal residue. The sequence generated by MassSeq suggests a substitution at the N-terminus of a glycine for an arginine. Further confidence in the assignment of an amino acid substitution was obtained from the mass difference between the theoretical and the experimentally measured mass of the peptide, which was only 7 mDa (4.6 ppm).

Sequence	Mass (Da)	
GWPVHLLETLLPR	1530.8796	Sequence generated by MassSeq
R WPVHLLETLLPR	1629.9593	Sequence in Swiss database

Table 1. Comparison of MassSeq generated sequence with sequence from Swiss-Prot database.

It is interesting to note that increasing the minimum mass standard deviation parameter in the MassSeq Software from 0.005 to 0.025 results in the prediction of an alternative sequence KDPVHLLETLLPR which differs by 28 mDa (19.5 ppm) from the experimental mass (Table 2). The minimum mass standard deviation parameter in MassSeq is set to reflect the known mass measurement accuracy of the instrument used to acquire the spectrum. The MassSeq algorithm is specifically designed to work with exact mass data, and this example clearly illustrates the benefit of exact mass data in combination with MassSeq. This helps distinguish between different sequence possibilities in the case of *de novo* sequencing. The substitution (G for R) at the N-terminus of this peptide explains why it was not identified by peptide mass fingerprinting.

MassSeq mass standard deviation (mDa)	Sequence	Δ Mass (mDa)	Δ Mass (ppm)
5	GWPVHLLETLLPR	7	4.6
25	KDPVHLLETLLPR	28	19.5
25	GWPVHLLETLLPR	7	4.6

Table 2. Comparison of sequences generated by the MassSeq algorithm with different settings.

The MS/MS spectrum of the second peptide at 1533.786 Da (Figure 4) proved more complicated to sequence. The results from the MassSeq *de novo* sequencing, did not return a confident, complete sequence for the entire peptide, however it did have 100% confidence in a significant sub-sequence, PEFTLPV. BLAST searching of this sub-sequence resulted in a confident match to glycogen phosphorylase B, with the sequence being homologous in a variety of species. Further investigation suggested that the peptide could correspond to the sequence ARPEFTLPVHFYG(R). The theoretical monoisotopic mass of this peptide would be 1533.785 Da, which differs from the experimental mass by 1 mDa (0.7 ppm) giving a good indication that this sequence is correct. In addition, matching of the theoretical fragment ions to the MS/MS data shows that many of the low mass ions (100 Da–600 Da) match to internal acyl fragment ions (Figure 5), which are common in singly charged peptide MS/MS spectra. Internal acyl ions are not considered as primary fragmentation routes by the MassSeq algorithm, which accounts for them not being matched in the *de novo* sequence. However these ions are very useful in helping to corroborate the sequence. It can, therefore, be concluded that this second peptide is also from glycogen phosphorylase B but was not identified by peptide mass fingerprinting because the peptide is not the result of a specific tryptic cleavage, but has instead lost the C-terminal arginine residue.
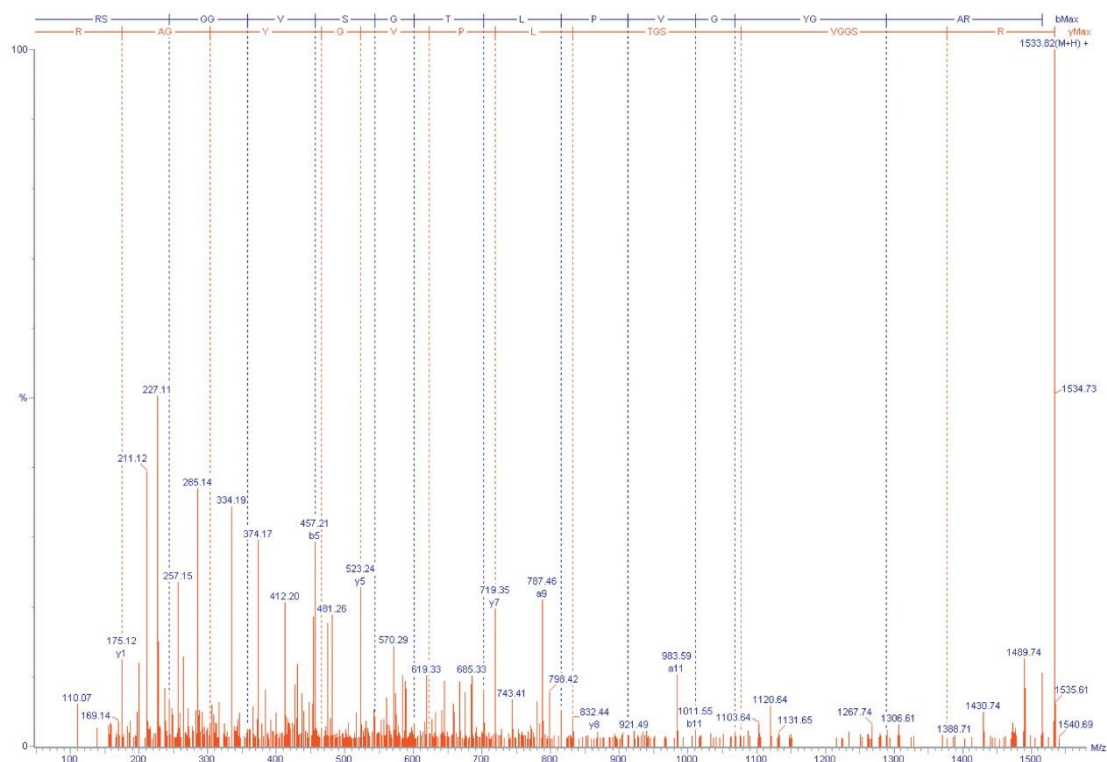
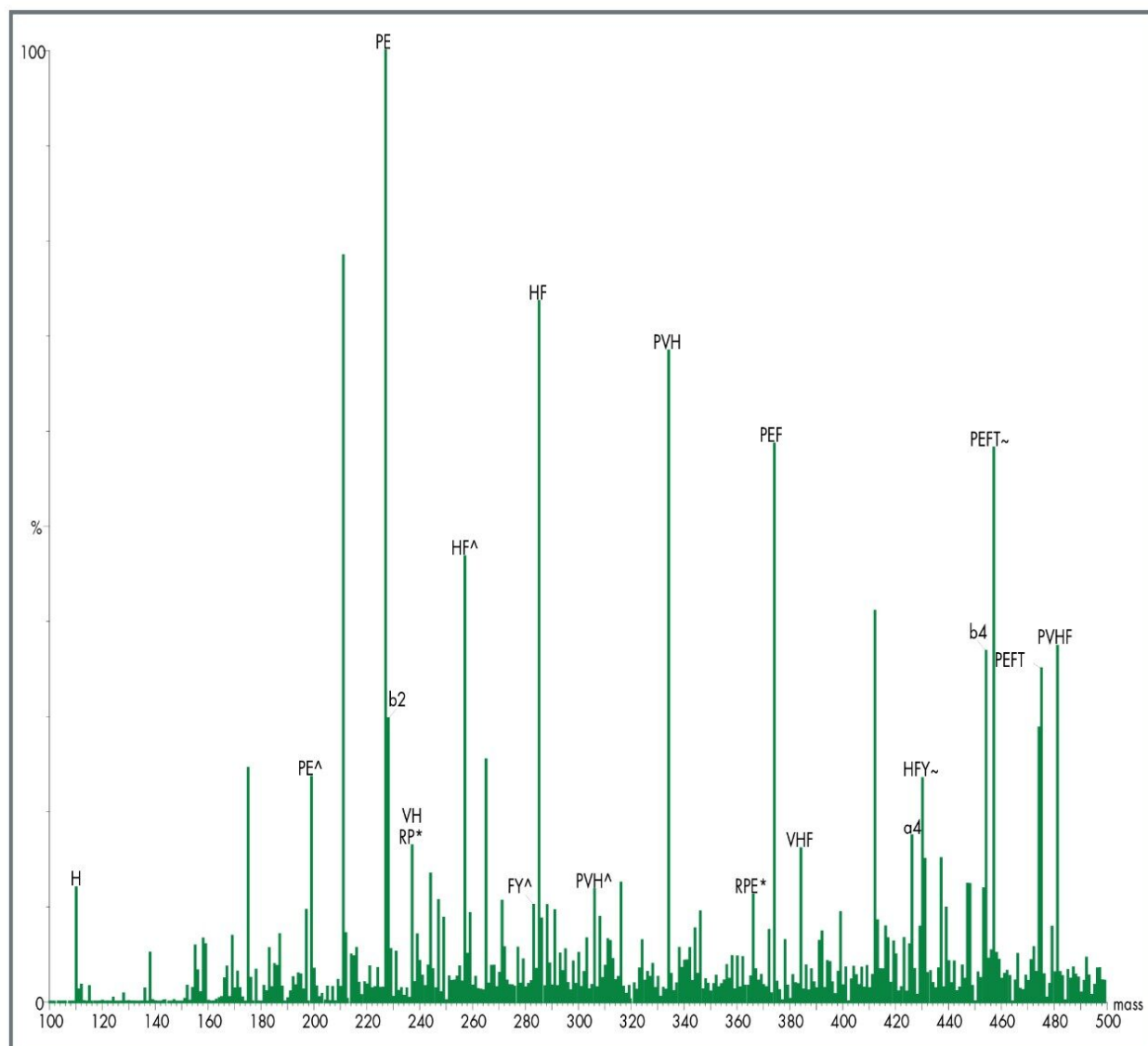Figure 4. MassSeq de novo sequence of the MS/MS spectrum of the peptide at 1533.786 Da.

Figure 5. Theoretical fragmentation of the peptide sequence ARPEFTLPVHFYG matched to the MS/MS spectrum of the peptide at 1533.786. The region 100–600 Da is displayed.

The high resolution, precursor ion selection capability of the quadrupole in the Q-Tof geometry allows confident and precise selection of peptides for MS/MS. This prevents peptides, similar in mass, from being fragmented at the same time, which results in an MS/MS spectrum, which is easier to interpret.

The exact mass capability of this instrument also reduces the number of false positives returned from database searching and *de novo* sequence algorithms, which provides greater confidence in the results obtained. These important capabilities are not possible on MALDI MS/MS instrumentation that uses ion gate technology for precursor ion selection and axial Tof analyzers for mass separation and measurement.

Conclusion

Using the Q-Tof Ultima MALDI we have been able to characterize two peaks in a tryptic digest that were not identified by peptide mass fingerprinting and were separated by only 3 Da in mass at *m/z* 1530.

A precursor selection window of 1 Da can be set using the Q-Tof Ultima MALDI. It is therefore possible to accurately select precursor ions for MS/MS experiments, even when ions have a similar *m/z*, thus avoiding the simultaneous fragmentation of two different precursor ions.

The exact mass capabilities of this instrument allow confident *de novo* sequence information to be obtained by the automated computer algorithm, MassSeq.

Sequence information obtained from *de novo* sequencing is useful for the identification of proteins by homology based BLAST searching and the characterization of modifications, e.g. phosphorylations.

Featured Products

ProteinLynx Global SERVER (PLGS) <https://www.waters.com/513821>

720000852, August 2004