Application Note: 511

Characterization of Spider Venom Peptides by High-Resolution LC-MS/MS Analysis

Claire Dauly[1], Pierre Escoubas[2,3], Glenn F. King[4], Graham M. Nicholson[5]
[1]Thermo Fisher Scientific, Courtaboeuf Cedex, France
[2]University of Nice Sophia Antipolis, Institute of Molecular and Cellular Pharmacology, Valbonne, France
[3]VenomeTech, Valbonne, France
[4]Institute for Molecular Bioscience, The University of Queensland, St. Lucia, Australia
[5]School of Medical & Molecular Biosciences, University of Technology, Sydney, Australia

Key Words

- Peptidomics workflow
- Spider venom peptides
- ETD
- HCD
- *De novo* sequencing
- LTQ Orbitrap XL

Introduction

Animal venoms are natural libraries of biologically active peptides. They encompass a wide variety of structures and pharmacological activities and represent an enormous resource of novel molecules to be used as insecticide, therapeutic and drug models. Australian funnel-web spiders contain many hundreds of peptides that follow a bimodal distribution, with about 75% of the peptides ranging from 3,000 to 5,000 Da[1]. Although the up-to-date, funnel-web spiders protein database contains about 100 sequences, it was shown recently that the number of peptides present in venom is in the high hundreds[1]. Venom profiling by high-resolution and accurate-mass (HR/AM) LC-MS/MS can therefore be used for species identification and to indicate the presence of potentially unknown toxins. Combining MS-based peptide sequence tags with cutting-edge transcriptomics studies and the ability to generate full peptide sequences using molecular biology techniques such as RACE (Rapid Amplification of Complementary Ends) will allow the generation of large peptide sequence libraries that can be mined for drug discovery purposes, in a global approach termed "venomics".

The Thermo Scientific LTQ Orbitrap XL ETD hybrid mass spectrometer is a versatile instrument that allows the fragmentation of peptides and proteins by several different dissociation techniques including collision-induced dissociation (CID), higher-energy collisional dissociation (HCD) and electron transfer dissociation (ETD) (Figure 1). Large peptides require a high resolving power to achieve isotopic resolution of intact parent ions and MS/MS product ions that can then be utilized for *de novo* sequencing. In this study a combination of HCD and ETD was used to generate MS/MS spectra with high fragment ion coverage to characterize a funnel-web spider venom peptidome and to demonstrate that *de novo* sequencing with HR/AM spectrometry is applicable to venomics research.

Goal

Use high-resolution accurate-mass LC-MS/MS with a combination of HCD and ETD fragmentation techniques to characterize endogenous peptides in spider venom.

Experimental

Sample Preparation

Australian funnel-web Hadronyche infensa: Orchid beach (Hi:OB) spiders were collected from the Fraser Island, Queensland area. Crude venom was 'milked' by direct aspiration from the chelicerae of aggravated spiders. Venom was collected into silanized glass pipettes, eluted by rinsing the pipettes with 0.1% (v/v) trifluoroacetic acid (TFA), and immediately freeze-dried. A 1 µL (200 µg) sample of crude venom was reduced with 10 mM DTT at 56 °C for 45 minutes and alkylated with 55 mM iodoacetamide in 25 mM ammonium bicarbonate for 30 minutes. Crude and reduced samples were diluted in 500 µL of 0.1% formic acid prior to analysis by reversed-phase nanoflow HPLC-MS/MS.

LC-MS Analysis

Samples were separated by online reversed-phase chromatography using a Thermo Scientific Accela MS pump and Accela autosampler equipped with a C18 Reprosil analytical column (100-µm ID x 1-cm packed-tip column, Nikkyo Technos Co. Ltd) and a C18 trapping cartridge (Captrap,

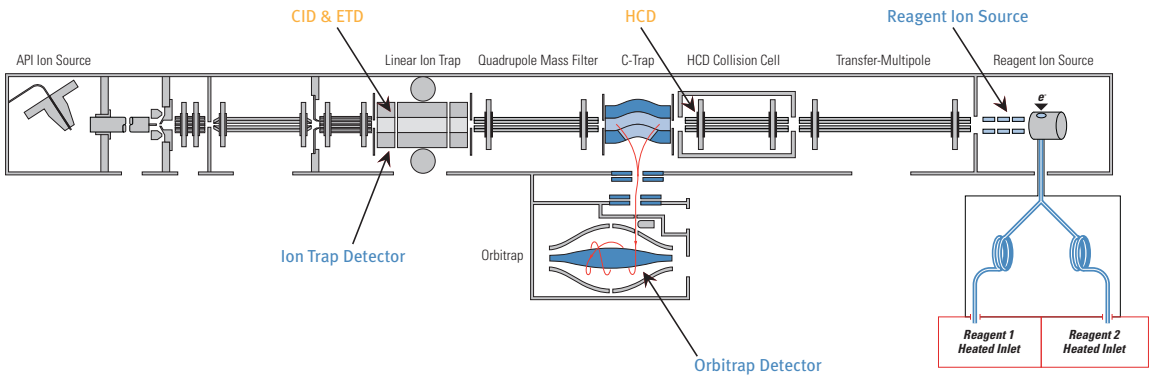


Figure 1: Schematic of the LTQ Orbitrap XL hybrid mass spectrometer with ETD.

Michrom Bioresources) at a flow rate of 300 nL/min post flow-split. The chromatography system was coupled with an LTQ Orbitrap XL™ mass spectrometer equipped with ETD. Details of chromatographic and mass spectrometric settings are listed in Tables 1 and 2. A total of about 1 µg of venom peptides was used for all experiments.

Crude venom peptides were analyzed by LC-MS with full-scan MS detection in the Orbitrap mass analyzer at 100,000 resolving power. Reduced and alkylated sample was then analyzed with data-dependent tandem MS acquisition methods in order to obtain sequence information. Peptides detected in full-scan MS were selected for HCD and ETD fragmentation in two separate injections. All experiments were performed with Orbitrap detection to resolve highly charged peptides and fragment ions. For data-dependent acquisitions, the method was set to analyze the three most intense ions.

Table 1: LC parameter settings.

LC Separation	
HPLC System	Thermo Scientific Accela MS Pump equipped with a flow splitter
Column	C18 Reprosil analytical column (100 µm id x 15 cm packed tip column, Nikkyo Technos Co. Ltd)
Pre-column	C18 trapping cartridge (Captrap, Michrom Bioresources)
Mobile Phases	0.1% formic acid in water (eluent A); 0.1% formic acid in acetonitrile (eluent B)
Gradient	5% – 40% B in 40 min
Flow	300 nL/min

Table 2: Mass spectrometer parameter settings.

Mass Spectrometer Settings		
Source		nano-ESI
Capillary temperature		200 °C
Tube lens voltage [V]		140
Source voltage [kV]		1.6
Full MS mass range *[m/z]*		800 - 2000
Resolution settings	Full MS (Full scan only method)	100000
	Full MS (Data dependent MS^n method)	60000
	MS^n (Orbitrap)	15000
Target value	Full MS	1.00E+06
	MS^n (Orbitrap)	3.00E+05
Max. injection times	Full MS	500 ms
	MS^n (Orbitrap)	1000 ms
Dynamic Exclusion	Repeat count	1
	Exclusion list size	500
	Exclusion duration	90 s
	Exclusion mass width low/high	10 ppm
MS^n parameters		
Isolation width [u]		5
Minimum signal required		10000
Collision energy (HCD)		35
Activation time [ms] (ETD)		30
Top n MS/MS		3
FT first mass value *(m/z)*		110
Charge state screening on: +1, +2 and unassigned charge states rejected		
Monoisotopic precursor selection enabled		
Non-peptide monoisotopic recognition enabled		

Data Processing

Full-scan spectra were deconvoluted with Thermo Scientific Xtract software to obtain the exact masses of native venom peptides.

MS/MS spectra were first processed with Thermo Scientific Proteome Discoverer software version 1.2. A signal/noise filter of 1.5 was applied and spectra with precursor mass within 10 ppm and 5 minutes retention time were grouped. Non-fragment peaks were also removed from ETD spectra (see Table 3 for details) and peak lists corresponding to grouped spectra were exported in .mgf format.

De novo sequencing was performed with PEAKS® Studio 5.2 (Bioinformatics Solutions Inc.) after refinement of MS/MS spectra. Sequence tags generated by the *de novo* process were used for a homology search with a specific spider protein database (www.arachnoserver.org). PEAKS parameters are summarized in Table 3.

ProSight PTM software (web portal https://prosightptm.northwestern.edu) was used to match fragment ions to the identified protein sequence. Tandem mass spectra were deconvoluted with Xtract software and accurate fragment masses were matched with 5-ppm mass tolerance.

Table 3: Data analysis parameter settings.

MGF Generation Parameters (in Proteome Discoverer)	
Peak list generation conditions	
Total intensity threshold	100
Minimum peak count	8
S/N threshold	1.5
Spectral grouping	
Mass tolerance	10 ppm
RT window	5 min
Non-fragment filter (ETD)	
Mass window offset for precursor peak	4 Da
Mass window offset for charge reduced precursors	4 Da
Mass window offset for neutral losses from charge reduced precursors	4 Da

PEAKS Parameters	
Data Refine	
Filter quality	>0.65
Data preprocess	Yes
De Novo	
Parent mass error tolerance	10 ppm
Fragment mass error tolerance	0.01 Da
Enzyme	None
Fixed modification	Carbamidomethylation
Report # match	1
Spider search	
Database	Funnel-web from Arachnoserver
Query type	Homology match
Fixed modification	Carbamidomethylation
Mass tolerance	0.01 Da
Report # match	1

Results and Discussion

Hi:OB venom peptide masses were measured with Orbitrap detection at 100,000 resolving power. Crude and reduced samples were analyzed with the same chromatographic gradient. Figure 2 displays the base peak chromatogram of the two samples. A shift towards lower retention time was observed as well as improved chromatographic resolution for the reduced sample. This phenomenon could be explained by differences in protein folding and the reduction of intramolecular disulfide bridges. Spectra were deconvoluted with Xtract software to generate a mass list and estimate the number of non-redundant peptides. Figure 3A displays a zoom into the full-scan mass spectrum at a retention time of 33.85 minutes. The full MS scan contained many isotope patterns indicative of peptides that were resolved to isotopic resolution and could thus be readily deconvoluted to obtain monoisotopic peptide masses (Figure 3B). Peptide masses ranged from approximately 1,000 to 7,000 Da, and at least 8 monoisotopic peptide masses could be observed with high mass accuracy from this single MS scan. A second full-scan mass spectrum obtained at 29.11 minutes is displayed in Figure 3C. Several peptides with a charge state of +8 can be observed with good resolution (see zoomed frame). A total of 9 monoisotopic peptide masses were obtained after deconvolution with a mass range from 7991.3402 to 8502.3688 Da.

Mass lists were cleaned up by removing all masses below 1,000 Da and peptides with intensities below 1E4. A total of 479 monoisotopic peptide masses were detected from the crude venom sample. Peptides ranged from 1,000 to 15,000 Da with 85% of the peptides in the 3,000 to 9,000 Da range (Figure 4). Our results are in accordance with the expected bimodal distribution of venom peptides: one group ranging from 3,000–5,000 Da and another group ranging from 7,000–9,000 Da. However, in this study the distribution between the two groups is more equal with approximately 50% of peptides ranging from 3,000–5,000 Da versus 75% for the previously published MALDI-based analysis[1,2]. This difference can be likely explained by the use of a different ionization technique in this study (electrospray ionization) versus the original study (matrix assisted laser desorption ionization).

A combination of HCD and ETD experiments was used to further characterize and obtain peptide identities. Fragmentation with HCD was selected because this technique produces complete fragmentation spectra including low-mass sequence product ions and immonium ions that are useful for *de novo* sequenc-

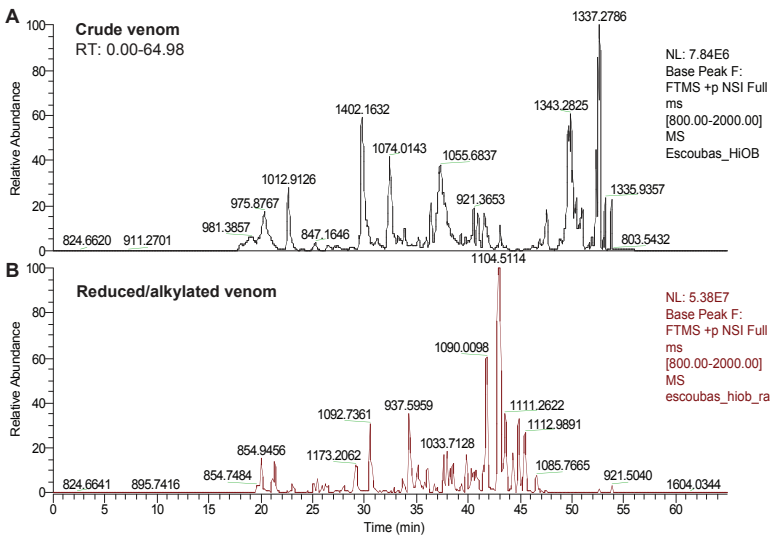


Figure 2: Base peak chromatogram of crude (A) and reduced/alkylated (B) venom analyzed by full-scan MS with Orbitrap detection at 100,000 resolving power.

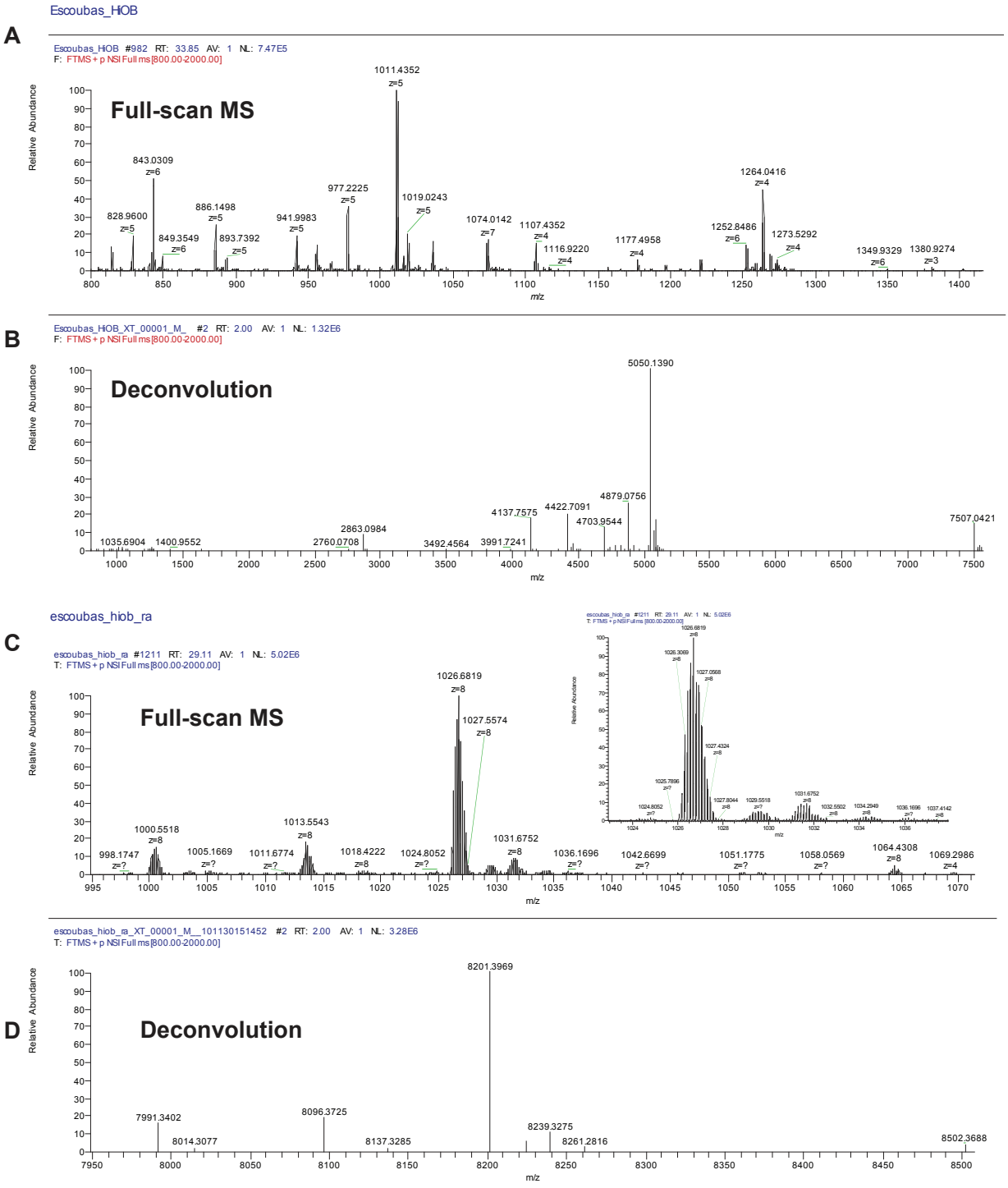


Figure 3: Zoom-in full-scan MS at 33.85 min (A) and 29.11 min (C). Mass spectra were deconvoluted with the Xtract algorithm to obtain accurate monoisotopic masses (B and D).

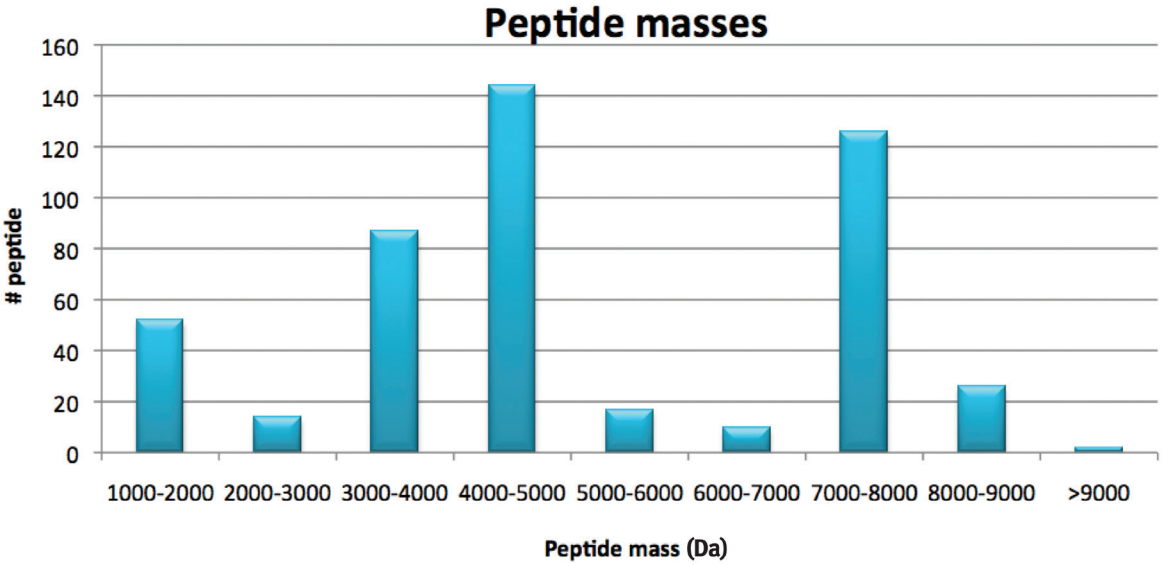


Figure 4: Venom peptide mass distribution.

ing. ETD was also expected to be very efficient for the fragmentation of large peptides and small proteins. Venom toxins are known to contain a high degree of cysteine disulfide bridges which confer peptide stability and activity. The Hi:OB venom sample was reduced and alkylated prior to analysis by tandem mass spectrometry in order to improve peptide fragmentation and facilitate MS/MS spectral interpretation. Tandem mass spectra resulting from either HCD or ETD fragmentation were transformed into peak lists (.mgf format) with Proteome Discoverer™ software after spectral grouping. In addition, ETD spectra were cleaned-up using the Non-Fragment Filter module of Proteome Discoverer software which removes precursors, charge-reduced species and neutral-loss ions from the peak list (Table 3). Resulting .mgf files were imported into PEAKS Studio software.

Spectra with a quality above 0.65 were further preprocessed (deconvolution and centroiding) by PEAKS Studio software and selected for *de novo* analysis. *De novo* sequencing was performed with 10-ppm mass accuracy for the precursor mass and 0.01-Da mass accuracy on fragment ions. All cysteines were considered as being carbamidomethylated for the analysis. A total of 362 *de novo* tags were generated with HCD fragmentation and 139 with ETD. Homology searches were performed using the identified *de novo* tags and a funnel-web spider protein database derived from the Arachnoserver spider toxin database (www.arachnoserver.org). A total of 46 homology matches were generated with SPIDER scores above 10. Results from this homology search are summarized in Table 4.

Figure 5 displays an example of an ETD mass spectrum that generated a *de novo* sequence tag that could be matched to a known peptide sequence. The data produced a SPIDER score of 48.11 and a delta mass of 50.0396 Da. This mass shift may be due to a sequence mutation that is very common in venom toxins or caused by post-translational modifications. In this case, fragment ions matched the predicted sequence for both the N- and C-terminal portions of the peptide (respectively c and z ions) indicating that the difference was located in the middle of the sequence and eliminating a possible truncation of the native protein.

Figure 6 shows another example where the *de novo* sequence tag could be matched to a protein sequence with the same exact mass. In this case, HCD fragmentation allowed the identification of a truncated form of omega-hexatoxin-Hi1a with a mass error of 0.0132 Da (5 ppm) (Table 4). It must be noted that a total of 22 out of 44 b or y ions matched to this peptide sequence and that the mass accuracy of all ions fell below 0.01 Da with a majority of fragments' mass deviation being below 0.005 Da. In the case of *de novo* sequencing, mass accuracy of fragment ions is crucial for the confirmation of amino acid sequences.

Table 4: Homology match results from PEAKS Studio.

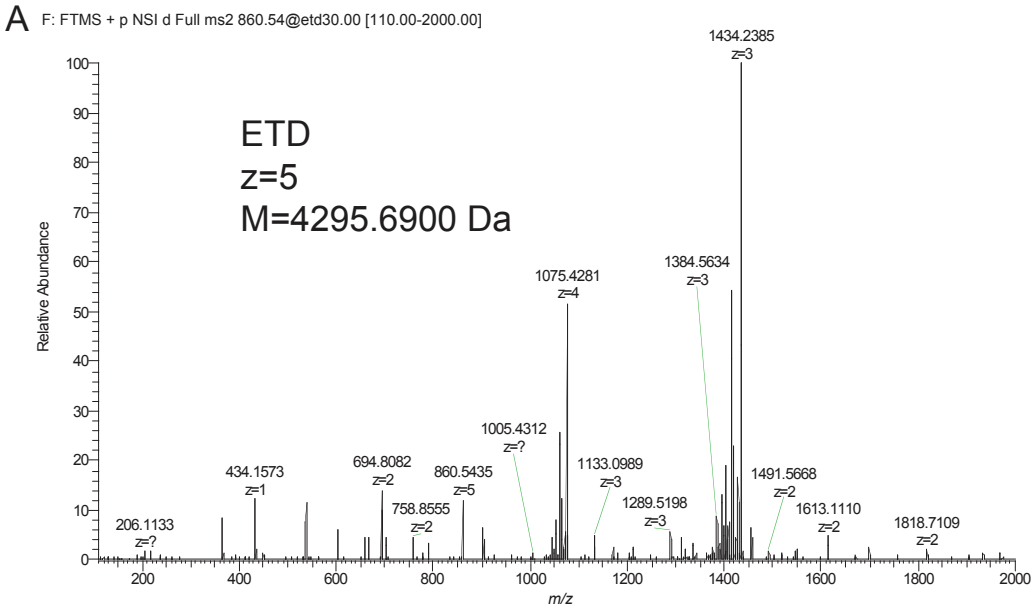
Homolog
C(+57.02)AKKRNWC(+57.02)GKTEDC(+57.02)C(+57.02)C(+57.02)PMKC(+57.02)VYAWYNEQGSC(+57.02)QSTLSALWK
C(+57.02)IPTGQPC(+57.02)PYNENC(+57.02)C(+57.02)SQSC(+57.02)TYKANENGNQVKRC(+57.02)D
C(+57.02)LAEAADC(+57.02)SPWSGDSC(+57.02)C(+57.02)KPYLC(+57.02)SC(+57.02)LFFYPC(+57.02)SC(+57.02)RPK
DAQRIFRRSPTC(+57.02)IPTGQPC(+57.02)PYNENC(+57.02)C(+57.02)SQSC(+57.02)TYKANEN
EIRKPIDTEKADAERVLDC(+57.02)VVNLLGC(+57.02)SSDKDC(+57.02)C(+57.02)GMTPSC(+57.02)TLGLC(+57.02)APS
EQC(+57.02)GDKVC(+57.02)GEGTC(+57.02)C(+57.02)SEFPVVHC(+57.02)RELGLVDDLC(+57.02)MSPGETTDSGRYLFFC(+57.02)PC(+57.02)ETGLRC(+57.02)DKNDWTC(+57.02)
GVLDC(+57.02)VVNTLGC(+57.02)SSDKDC(+57.02)C(+57.02)GMTPSC(+57.02)TLG
GVLDC(+57.02)VVNTLGC(+57.02)SSDKDC(+57.02)C(+57.02)GMTPSC(+57.02)TLGLC(+57.02)APSVG
GVVDC(+57.02)VLNTLGC(+57.02)SSDKDC(+57.02)C(+57.02)GMTPSC(+57.02)TLGLC(+57.02)APSVGGLVGGL
IMASTFAQKC(+57.02)GDQVC(+57.02)GAGTC(+57.02)C(+57.02)AEYPELHC(+57.02)KRVGQLYDLC(+57.02)VDSEATKDSGNHLFFC(+57.02)PC(+57.02)DEGMYC(+57.02)DMNSWSC(+57.02)C
KC(+57.02)GDQVC(+57.02)GAGTC(+57.02)C(+57.02)AEYPELHC(+57.02)K
KC(+57.02)GDQVC(+57.02)GAGTC(+57.02)C(+57.02)AEYPELHC(+57.02)KRVGQL
KC(+57.02)GDQVC(+57.02)GAGTC(+57.02)C(+57.02)AEYPELHC(+57.02)KRVGQLY
KC(+57.02)GDQVC(+57.02)GAGTC(+57.02)C(+57.02)AEYPELHC(+57.02)KRVGQLYDLC(+57.02)VDSEATKDSGNHLFFC(+57.02)PC(+57.02)DEGMYC(+57.02)DMNSWSC(+57.02)QKKTS
KC(+57.02)GDQVC(+57.02)GAGTC(+57.02)C(+57.02)AEYPELHC(+57.02)KRVGQLYDLC(+57.02)VDSEATKDSGNHLFFC(+57.02)PC(+57.02)DEGMYC(+57.02)DMNSWSC(+57.02)QKKTSG
KC(+57.02)GDQVC(+57.02)GAGTC(+57.02)C(+57.02)AEYPELHC(+57.02)KRVGQLYDLC(+57.02)VDSEATKDSGNHLFFC(+57.02)PC(+57.02)DEGMYC(+57.02)DMNSWSC(+57.02)QKKTSG
KC(+57.02)LAEAADC(+57.02)SPWSGDSC(+57.02)C(+57.02)KPYL
KC(+57.02)LAEAADC(+57.02)SPWSGDSC(+57.02)C(+57.02)KPYLC(+57.02)
KC(+57.02)LAEAADC(+57.02)SPWSGDSC(+57.02)C(+57.02)KPYLC(+57.02)SC(+57.02)LFFYPC(+57.02)SC(+57.02)RPKG
KC(+57.02)LAEAADC(+57.02)SPWSGDSC(+57.02)C(+57.02)KPYLC(+57.02)SC(+57.02)LFFYPC(+57.02)SC(+57.02)RPKG
KC(+57.02)LAEAADC(+57.02)SPWSGDSC(+57.02)C(+57.02)KPYLC(+57.02)SC(+57.02)LFFYPC(+57.02)SC(+57.02)RPKGW
KC(+57.02)LAEAADC(+57.02)SPWSGDSC(+57.02)C(+57.02)KPYLC(+57.02)SC(+57.02)LFFYPC(+57.02)SC(+57.02)RPKGW
KCGDQEC(+57.02)GEGTC(+57.02)C(+57.02)LDYSQQHC(+57.02)SRLGKLYDMC(+57.02)SDPNDKTDSGSHLFFC(+57.02)QC(+57.02)ETGLRC(+57.02)DKTSWSC(+57.02)
KDC(+57.02)C(+57.02)GMTPSC(+57.02)TMGLC(+57.02)VPSVGG
LDC(+57.02)VLNTLGC(+57.02)SSDKDC(+57.02)C(+57.02)GMTPSC(+57.02)TLGLC(+57.02)APNVGGL
LDC(+57.02)VVNLLGC(+57.02)SSDKDC(+57.02)C(+57.02)GMTPSC(+57.02)TLGLC(+57.02)APS
LDSRVC(+57.02)SSDKDC(+57.02)C(+57.02)GMTPSC(+57.02)TMGLC(+57.02)VPSVGGLVGG
LLDVATVLGGLEAGESDMRKDVMGLFRRALC(+57.02)TGADRPC(+57.02)AAC(+57.02)C(+57.02)PC(+57.02)C(+57.02)PG
LVDC(+57.02)VVNTLGC(+57.02)SSDKDC(+57.02)C(+57.02)GMTPSC(+57.02)TLGLC(+57.02)APSVGG
LVDC(+57.02)VVNTLGC(+57.02)SSDKDC(+57.02)C(+57.02)GMTPSC(+57.02)TLGLC(+57.02)APSVRGLVGGLLGR
LVIMASTSAQQC(+57.02)GDETC(+57.02)GAGTC(+57.02)C(+57.02)AVFSQNHC(+57.02)RRLSQMYDLC(+57.02)SDHSDASPSGNYLFFC(+57.02)PC(+57.02)EPGLHC(+57.02)DRNTW
NTATGVLALLVLATVLGC(+57.02)LEVEETRADLQGAFESYEGEAAEKI
PGSVC(+57.02)DGNESDC(+57.02)KC(+57.02)YGKWHKC(+57.02)RC(+57.02)PWKWHFTGEGPC(+57.02)TC(+57.02)EKGMKHTC(+57.02)LTKLHC(+57.02)PNKAEWGLDWRSEES
PTC(+57.02)IPTGQPC(+57.02)PYNENC(+57.02)C(+57.02)NQSC(+57.02)TYKANENGNQVKRC(+57.02)D
QC(+57.02)GDETC(+57.02)GAGTC(+57.02)C(+57.02)AVFSQNHC(+57.02)RRLSQMYDLC(+57.02)SDHSDASPSGNYLFFC(+57.02)PC(+57.02)EPGLHC(+57.02)DRNTWTC(+57.02)T
RRSTC(+57.02)TPTDQPC(+57.02)PYHESC(+57.02)C(+57.02)SGSC(+57.02)TYKANENGNQVK
SAQQC(+57.02)GDETC(+57.02)GAGTC(+57.02)C(+57.02)AVFSQNHC(+57.02)RRLSRMYDLC(+57.02)SDHADASPSGNYLFFC(+57.02)PC(+57.02)EPGLHC(+57.02)DRNTWTC(+57.02)TEGSS
SC(+57.02)C(+57.02)KPYLC(+57.02)SC(+57.02)LFFYPC(+57.02)SC(+57.02)RPKGW
SPTC(+57.02)IPTGQPC(+57.02)PYNENC(+57.02)C(+57.02)SQSC(+57.02)TYKANENGNQVKRC(+57.02)D
STC(+57.02)TPTDQPC(+57.02)PYDESC(+57.02)C(+57.02)SGSC(+57.02)TYK
STC(+57.02)TPTDQPC(+57.02)PYDESC(+57.02)C(+57.02)SGSC(+57.02)TYKANENGNQV
STC(+57.02)TPTDQPC(+57.02)PYDESC(+57.02)C(+57.02)SGSC(+57.02)TYKANENGNQVKRC(+57.02)
STC(+57.02)TPTDQPC(+57.02)PYDESC(+57.02)C(+57.02)SGSC(+57.02)TYKANENGNQVKRC(+57.02)D
STC(+57.02)TPTDQPC(+57.02)PYDESC(+57.02)C(+57.02)SGSC(+57.02)TYKANENGNQVKRC(+57.02)D
STC(+57.02)TPTDQPC(+57.02)PYHESC(+57.02)C(+57.02)SGSC(+57.02)TYKANENGNQVKRC(+57.02)
STC(+57.02)TPTDQPC(+57.02)PYHESC(+57.02)C(+57.02)SGSC(+57.02)TYKANENGNQVKRC(+57.02)D
STC(+57.02)TPTDQPC(+57.02)PYHESC(+57.02)C(+57.02)SGSC(+57.02)TYKANENGNQVKRC(+57.02)D
TC(+57.02)ILPTGQPC(+57.02)PYNENC(+57.02)C(+57.02)SQSC(+57.02)TYKANENGNQVKRC(+57.02)D
TC(+57.02)IPTGQPC(+57.02)PYNENC(+57.02)C(+57.02)SQSC(+57.02)TYKANENGNQVKRC(+57.02)D
TCIRTDQPC(+57.02)PYNESC(+57.02)C(+57.02)SGSC(+57.02)TYKANENGNQVKRC(+57.02)D
VDC(+57.02)VLNTLGC(+57.02)SSDKDC(+57.02)C(+57.02)GMTPSC(+57.02)TLGLC(+57.02)APSVGGLVGG
VLDC(+57.02)VVNTLGC(+57.02)SSDKDC(+57.02)C(+57.02)GMTPSC(+57.02)TLGLC(+57.02)APSVGGLVGG
VLNTLGC(+57.02)SSDKDC(+57.02)C(+57.02)GMTPSC(+57.02)TLGIC(+57.02)
VNLLGC(+57.02)SSDKDC(+57.02)C(+57.02)GMTPSC(+57.02)TLGLC(+57.02)APSVGGLVGGLLG

	M homolog (Da)	SPIDER score	*m/z*	z	Delta mass (Da)	RT	Fragmentation type
	5020.1562	12.39	856.7114	6	-114.0684	38.4	ETD
	4108.665	19.87	836.3481	5	-68.0396	15.38	ETD
	4308.742	18.68	895.5808	5	-164.1255	49.88	HCD
	4478.931	20.18	919.5861	5	-113.9629	50.9	HCD
	5297.4272	20.88	900.8737	6	-101.7715	24.43	HCD
	7526.1133	41.19	1533.219	5	-134.9463	35.5	HCD
	3062.258	32.49	1040.772	3	-57.0364	42.98	HCD
	3746.5845	38.66	1268.885	3	-57.0471	48.18	HCD
	4242.8853	45.68	1434.312	3	-57.0298	53.68	HCD
QQ	8637.607	56.57	802.8148	11	-182.2754	31.08	HCD
	2599.0383	54.35	904.3699	3	-111.0496	20.46	HCD
	3152.3718	29.49	816.8616	4	-111.0454	24.26	HCD
	3430.462	14.08	886.3865	4	-111.0549	25.25	HCD
	7976.282	30.62	811.9532	10	-133.1777	31.73	HCD
	8033.3037	44.58	914.9373	9	-192.0664	30.48	ETD
	8033.3037	27.11	915.0503	9	-193.084	30.88	HCD
	2574.0645	25.04	877.3869	3	-55.0745	26.41	HCD
	2734.0952	59.89	912.3767	3	-0.0132	29.3	HCD
	4493.8584	26.33	913.791	5	-70.0605	48.83	ETD
	4493.8584	22.18	913.7837	5	-70.0239	48.61	HCD
	4679.9375	17.22	937.0002	5	-0.0273	43.63	HCD
	4679.9375	51.01	936.9985	5	-0.0186	45.11	ETD
	7491.011	19.63	1528.622	5	-147.0601	35.22	HCD
	2272.9077	19.6	802.9897	3	-133.0396	15.91	HCD
	3801.6265	15.18	1325.583	3	-172.1006	52.83	HCD
	3446.441	17.31	891.1364	4	-114.0757	48.73	HCD
	3618.5374	18.81	1215.208	3	-24.0652	38	HCD
	5182.3833	23.53	876.3791	6	-69.8477	39.06	ETD
	3746.5842	36.74	1278.212	3	-85.0303	50.58	HCD
	4611.15	22.58	931.0092	5	-38.8594	48.05	HCD
	8149.3823	36.03	821.0497	10	-51.0415	31.68	HCD
	4550.309	12.68	1167.001	4	-113.666	20.18	HCD
	8111.5127	29.51	915.6028	9	-119.8477	30.81	HCD
	4333.7764	48.99	1465.61	3	-60.0317	19.42	HCD
	7509.986	15.55	1528.825	5	-129.104	35.04	HCD
	4148.7256	43.95	1423.567	3	-118.9541	16.35	HCD
	8168.29	50.16	1183.491	7	-109.0977	29.01	HCD
	3032.2722	41.21	1011.775	3	-0.03	42.28	HCD
	4393.7974	33.96	1099.462	4	-0.0195	19.16	HCD
	2860.0388	35.97	961.7006	3	-22.041	16.68	HCD
	3686.3958	59.17	1237.15	3	-22.031	18.23	HCD
	4130.6226	38.98	1062.176	4	-114.0503	16.5	HCD
	4245.6494	48.11	860.145	5	-50.0396	16.54	ETD
	4245.6494	67.9	1062.428	4	-0.0342	18.75	HCD
	4152.655	39.75	854.3521	5	-114.0693	14.98	HCD
	4267.6816	56.58	857.7378	5	-15.9712	16.76	ETD
	4267.6816	45.59	854.5465	5	-0.0146	16.68	HCD
	4209.713	37.63	857.5316	5	-72.9087	16.36	ETD
	4209.713	39.74	1053.439	4	-0.0117	18.46	HCD
	4228.7188	44.7	854.5543	5	-39.0166	18.5	ETD
	3973.7112	28.99	815.1736	5	-97.1208	50.63	HCD
	4072.7795	37.07	1377.297	3	-56.0896	52.38	HCD
	2805.1572	26.27	950.3861	3	-42.9792	20.6	HCD
	3781.7273	22.26	1290.284	3	-86.1023	53.4	HCD

Overall, peptides from a wide mass range could be identified, with 8 peptides representing masses above 6,000 Da positively matched to known peptide sequences with HCD and/or ETD fragmentation spectra (Table 4).

Sequence homology match is a useful tool for matching sequence tags to known protein sequences. However, a lot of sequence tags generated by *de novo* analysis remain unmatched because spider venom databases are not complete. In the example displayed in Figure 7, a peptide with a mass of 8625.7891 Da generated a good *de novo* sequence tag covering a 15 amino acid stretch of the N-terminal part of the peptide with excellent mass accuracy. However this *de novo* sequence tag could not be matched to any known protein sequence.

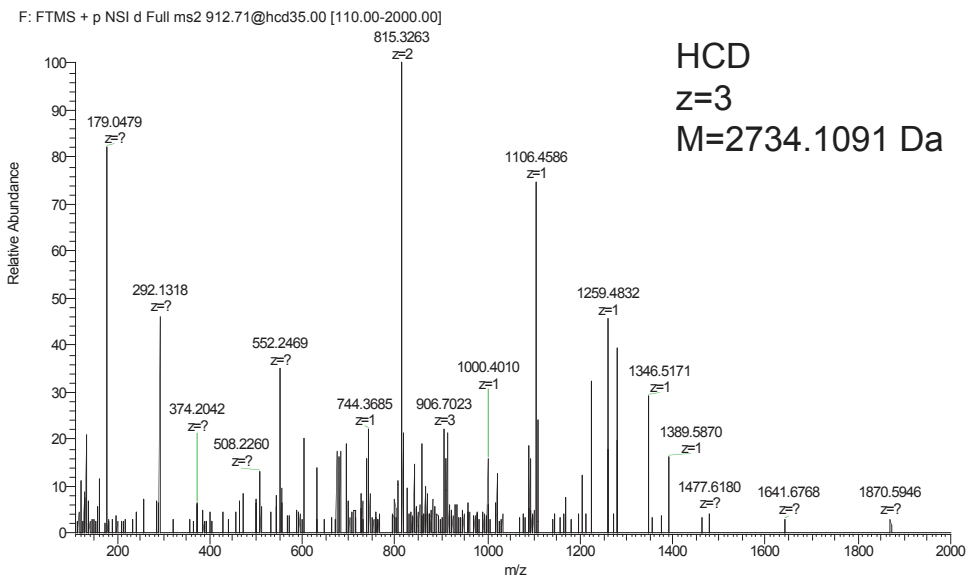
Several peptides could be positively matched to homologous sequences with both ETD and HCD. ETD



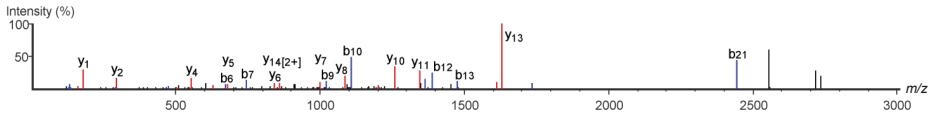
#	Immonium	b	b-H2O	a	c	Seq	y	y-H2O	z	z'	y (2+)	#
1	60.0449	88.0399	70.0293	60.0449	105.0669	S						36
2	74.0605	189.0875	171.0770	161.0925	206.1133	T	4209.6460	4191.6353	4192.6187	4193.6187	2105.3230	35
3	133.0435	349.1182	331.1076	321.1232	366.1440	57.02@C	4108.5981	4090.5874	4091.5708	4092.5708	2054.7991	34
4	72.0812	448.1866	430.1760	420.1916	465.2136	V	3948.5671	3930.5566	3931.5400	3932.5400	1974.7836	33
5	70.0656	545.2394	527.2288	517.2444	562.2664	P	3849.4988	3831.4883	3832.4717	3833.4717	1925.2494	32
6	72.0812	644.3077	626.2972	616.3127	661.3348	V	3752.4460	3734.4355	3735.4189	3736.4189	1876.7230	31
7	87.0558	758.3507	740.3401	730.3557	775.3777	N	3653.3777	3635.3672	3636.3506	3637.3506	1827.1888	30
8	101.0714	886.4093	868.3987	858.4143	903.4363	Q	3539.3347	3521.3242	3522.3076	3523.3076	1770.1674	29
9	30.0343	943.4307	925.4202	915.4357	960.4578	G	3411.2761	3393.2656	3394.2490	3395.2490	1706.1381	28
10	133.0435	1103.4614	1085.4509	1075.4664	1120.4884	57.02@C	3354.2546	3336.2441	3337.2275	3338.2275	1677.6273	27
11	70.0656	1200.5142	1182.5037	1172.5192	1217.5411	P	3194.2239	3176.2134	3177.1968	3178.1968	1597.6119	26
12	87.0558	1314.5571	1296.5466	1286.5621	1331.5841	N	3097.1711	3079.1606	3080.1440	3081.1440	1549.0856	25
13	30.0343	1371.5785	1353.5680	1343.5835	1388.6082	G	2983.1309	2965.1177	2966.1011	2967.1011	1492.0641	24
14	101.1078	1499.6735	1481.6630	1471.6785	1516.7030	K	2926.1067	2908.0962	2909.0796	2910.0879	1463.5533	23
15	60.0449	1586.7056	1568.6951	1558.7106	1603.7325	S	2798.0120	2780.0015	2780.9849	2781.9849	1399.5060	22
16	133.0435	1746.7362	1728.7257	1718.7412	1763.7632	57.02@C	2710.9797	2692.9692	2693.9526	2694.9526	1355.9899	21
17	133.0435	1906.7668	1888.7563	1878.7719	1923.7938	57.02@C	2550.9492	2532.9387	2533.9221	2534.9221	1275.9746	20
18	60.0449	1993.7988	1975.7883	1965.8038	2010.8258	S	2390.9185	2372.9080	2373.8914	2374.8914	1195.9592	19
19	104.0533	2124.8394	2106.8289	2096.8445	2141.8665	M	2303.8865	2285.8760	2286.8594	2287.8594	1152.4432	18
20	133.0435	2284.8701	2266.8596	2256.8752	2301.8972	57.02@C	2172.8459	2154.8354	2155.8188	2156.8188	1086.9230	17
21	133.0435	2444.9006	2426.8901	2416.9058	2461.9277	57.02@C	2012.8153	1994.8048	1995.7883	1996.7883	1006.9077	16
22	133.0435	2604.9314	2586.9209	2576.9365	2621.9585	57.02@C	1852.7847	1834.7742	1835.7577	1836.7577	926.8923	15
23	133.0435	2764.9619	2746.9514	2736.9670	2781.9954	57.02@C	1692.7540	1674.7435	1675.7271	1676.7271	846.8770	14
24	101.1078	2893.0569	2875.0464	2865.0620	2910.0879	K	1532.7234	1514.7129	1515.6964	1516.7030	766.8617	13
25	44.0499	2964.0940	2946.0835	2936.0991	2981.1243	A	1404.6277	1386.6179	1387.6014	1388.6082	702.8142	12
26	87.0558	3078.1370	3060.1265	3050.1421	3095.1702	N	1333.5913	1315.5808	1316.5643	1317.5710	667.2957	11
27	102.0554	3207.1794	3189.1689	3179.1846	3224.2109	E	1219.5483	1201.5378	1202.5214	1203.5281	610.2742	10
28	87.0558	3321.2224	3303.2119	3293.2275	3338.2495	N	1090.5057	1072.4952	1073.4788	1074.4846	545.7529	9
29	30.0343	3378.2439	3360.2334	3350.2490	3395.2727	G	976.4629	958.4523	959.4359	960.4359	488.7314	8
30	87.0558	3492.2869	3474.2764	3464.2920	3509.3120	N	919.4414	901.4308	902.4144	903.4218	460.2207	7
31	101.0714	3620.3455	3602.3350	3592.3506	3637.3726	Q	805.3984	787.3879	788.3715	789.3801	403.1992	6
32	72.0812	3719.4138	3701.4033	3691.4189	3736.4409	V	677.3399	659.3293	660.3129	661.3129	339.1700	5
33	101.1078	3847.5088	3829.4983	3819.5139	3864.5383	K	578.2715	560.2609	561.2445	562.2445	289.6357	4
34	129.1139	4003.6099	3985.5994	3975.6150	4020.6370	R	450.1765	432.1660	433.1495	434.1572	225.5883	3
35	133.0435	4163.6406	4145.6299	4135.6455	4180.6675	57.02@C	294.0754	276.0648	277.0484	278.0484	147.5377	2
36	88.0398					D	134.0448	116.0342	117.0178	118.0178	67.5224	1

Figure 5: Example of homology-matched ETD spectrum. (A) Full-scan MS/MS spectrum. (B) Annotated MS/MS spectrum and fragment match table.

A



B



#	Immonium	b	b-H2O	a	c	Seq	y	y-H2O	z	z'	y (2+)	#
1	101.1078	129.1012	111.0922	101.1078	146.1298	K						23
2	133.0435	289.1325	271.1229	261.1384	306.1604	57.02@C	2607.0076	2588.9971	2589.9805	2590.9805	1304.0038	22
3	86.0969	402.2175	384.2069	374.2225	419.2445	L	2446.9771	2428.9666	2429.9500	2430.9500	1223.9885	21
4	44.0499	473.2536	455.2440	445.2596	490.2816	A	2333.8931	2315.8826	2316.8660	2317.8660	1167.4465	20
5	102.0554	602.2972	584.2866	574.3022	619.3242	E	2262.8560	2244.8455	2245.8289	2246.8289	1131.9280	19
6	44.0499	673.3325	655.3237	645.3393	690.3613	A	2133.8132	2115.8027	2116.7861	2117.7861	1067.4066	18
7	44.0499	744.3685	726.3608	716.3764	761.3984	A	2062.7761	2044.7656	2045.7491	2046.7491	1031.8881	17
8	88.0398	859.3983	841.3878	831.4033	876.4254	D	1991.7390	1973.7285	1974.7120	1975.7120	996.3695	16
9	133.0435	1019.4298	1001.4185	991.4340	1036.4561	57.02@C	1876.7122	1858.7017	1859.6852	1860.6852	938.8561	15
10	60.0449	1106.4586	1088.4506	1078.4661	1123.4880	S	1716.6814	1698.6709	1699.6544	1700.6544	858.8475	14
11	70.0656	1203.5138	1185.5033	1175.5188	1220.5408	P	1629.6448	1611.6317	1612.6224	1613.6224	815.3247	13
12	159.0921	1389.5870	1371.5826	1361.5959	1406.6201	W	1532.5967	1514.5862	1515.5697	1516.5697	766.7983	12
13	60.0449	1476.6226	1458.6146	1448.6301	1493.6521	S	1346.5171	1328.5068	1329.4904	1330.4904	673.7587	11
14	30.0343	1533.6466	1515.6361	1505.6516	1550.6736	G	1259.4832	1241.4749	1242.4584	1243.4584	630.2454	10
15	88.0398	1648.6736	1630.6631	1620.6786	1665.7006	D	1202.4634	1184.4534	1185.4369	1186.4369	601.7319	9
16	60.0449	1735.7037	1717.6951	1707.7106	1752.7325	S	1087.4340	1069.4264	1070.4099	1071.4099	544.2184	8
17	133.0435	1895.7362	1877.7257	1867.7412	1912.7632	57.02@C	1000.4010	982.3943	983.3779	984.3779	500.7039	7
18	133.0435	2055.7668	2037.7563	2027.7719	2072.7939	57.02@C	840.3724	822.3636	823.3472	824.3472	420.6871	6
19	101.1078	2183.8618	2165.8513	2155.8669	2200.8889	K	680.3417	662.3330	663.3166	664.3166	340.6718	5
20	70.0656	2280.9146	2262.9041	2252.9197	2297.9417	P	552.2469	534.2380	535.2216	536.2216	276.6243	4
21	136.0761	2443.9731	2425.9673	2415.9829	2461.0049	Y	455.1959	437.1853	438.1689	439.1689	228.0979	3
22	86.0969	2557.0620	2539.0515	2529.0671	2574.0891	L	292.1318	274.1220	275.1056	276.1056	146.5663	2
23	133.0435					57.02@C	179.0479	161.0379	162.0213	163.0215	90.0242	1

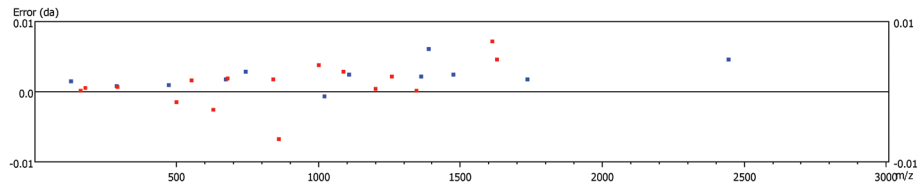
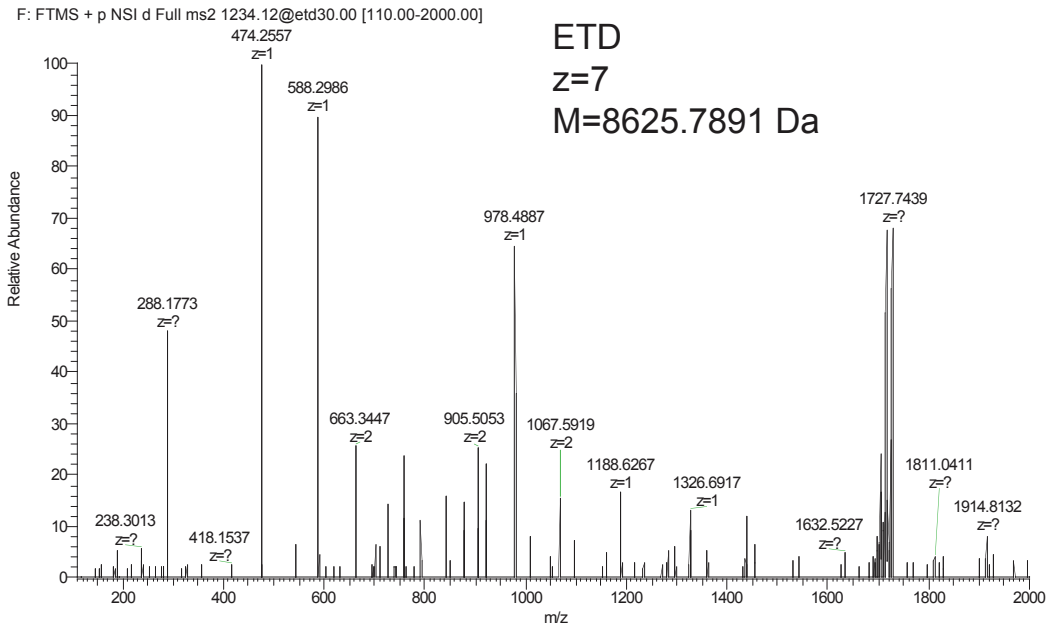


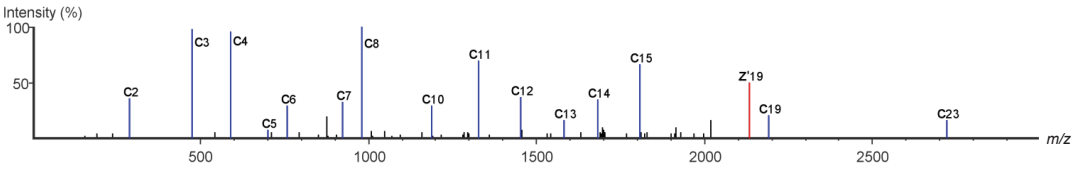
Figure 6: Example of homology-matched HCD mass spectrum. (A) Full-scan MS/MS spectrum. (B) Annotated MS/MS spectrum, fragment match table and fragment error distribution.

is known to give good fragmentation spectra for highly charged peptides. In the case of the peptide with a mass of 8033.3037 Da (z=9) the SPIDER score for the ETD mass spectrum was 44.58 versus 27.11 for the HCD mass spectrum. In cases such as the peptide STCTPTDQPCPY-DESCCSGSCTYKANENGNQVKRCD (4245.6494 Da), the HCD spectrum gave a better SPIDER score with 67.9 for HCD (z=4) versus 48.11 for ETD (z=5). Figure 8 displays an example of fragment-ion match based on protein sequence performed by ProSight PTM online software for the peptide with a mass of 4679.9375 Da. The comparison between HCD and ETD demonstrates the efficiency and complementary of the two fragmentation techniques. In this work, amino acid sequences that were not determined by HCD were in most cases identified by ETD. This is clearly important for novel protein identification and sequence validation.

A



B



#	Immonium	b	b-H2O	a	c	Seq	y	y-H2O	z	z'	y (2+)	#
1	129.1139	157.1089	139.0984	129.1139	174.1359	R						74
2	87.0558	271.1519	253.1413	243.1569	288.1773	N	8470.6953	8452.6846	8453.6689	8454.6689	4235.8477	73
3	159.0921	457.2312	439.2206	429.2362	474.2557	W	8356.6523	8338.6416	8339.6260	8340.6260	4178.8262	72
4	87.0558	571.2741	553.2635	543.2791	588.2986	N	8170.5737	8152.5630	8153.5464	8154.5464	4085.7869	71
5	86.0969	684.3582	666.3476	656.3632	701.3790	L	8056.5308	8038.5200	8039.5034	8040.5034	4028.7654	70
6	30.0343	741.3796	723.3691	713.3846	758.4053	G	7943.4468	7925.4360	7926.4194	7927.4194	3972.2234	69
7	136.0761	904.4429	886.4324	876.4479	921.4689	Y	7886.4253	7868.4146	7869.3979	7870.3979	3943.7126	68
8	30.0343	961.4644	943.4539	933.4694	978.4887	G	7723.3618	7705.3511	7706.3345	7707.3345	3862.1809	67
9	70.0656	1058.5172	1040.5067	1030.5222	1075.5442	P	7666.3403	7648.3296	7649.3130	7650.3130	3833.6702	66
10	86.0969	1171.6012	1153.5907	1143.6062	1188.6267	L	7569.2876	7551.2769	7552.2603	7553.2603	3785.1438	65
11	110.0717	1308.6602	1290.6497	1280.6652	1325.6886	H	7456.2036	7438.1929	7439.1763	7440.1763	3728.6018	64
12	101.1078	1436.7551	1418.7446	1408.7601	1453.7749	K	7319.1445	7301.1338	7302.1172	7303.1172	3660.0723	63
13	101.0714	1564.8137	1546.8032	1536.8187	1581.8405	Q	7191.0498	7173.0391	7174.0225	7175.0225	3596.0249	62
14	72.0812	1663.8821	1645.8716	1635.8871	1680.9032	V	7062.9912	7044.9805	7045.9639	7046.9639	3531.9956	61
15	101.1078	1791.9771	1773.9666	1763.9821	1808.9998	K	6963.9229	6945.9121	6946.8955	6947.8955	3482.4614	60
16	87.0558	1906.0200	1888.0095	1878.0250	1923.0470	N	6835.8276	6817.8169	6818.8003	6819.8003	3418.4138	59
17	70.0656	2003.0728	1985.0623	1975.0778	2020.0997	P	6721.7847	6703.7739	6704.7573	6705.7573	3361.3923	58
18	86.0969	2116.1567	2098.1462	2088.1619	2133.1838	L	6624.7319	6606.7212	6607.7046	6608.7046	3312.8660	57
19	30.0343	2173.1782	2155.1677	2145.1833	2190.2046	G	6511.6479	6493.6372	6494.6206	6495.6206	3256.3240	56
20	101.1078	2301.2732	2283.2627	2273.2783	2318.3003	K	6454.6265	6436.6157	6437.5991	6438.5991	3227.8132	55
21	159.0921	2487.3525	2469.3420	2459.3577	2504.3796	W	6326.5317	6308.5210	6309.5044	6310.5044	3163.7659	54
22	74.0605	2588.4001	2570.3896	2560.4053	2605.4272	T	6140.4521	6122.4414	6123.4248	6124.4248	3070.7261	53
23	87.0558	2702.4431	2684.4326	2674.4482	2719.4695	N	6039.4043	6021.3936	6022.3770	6023.3770	3020.2021	52
24	86.0969	2815.5271	2797.5166	2787.5322	2832.5542	L	5925.3618	5907.3511	5908.3345	5909.3345	2963.1809	51
25	101.0714	2943.5857	2925.5752	2915.5908	2960.6128	Q	5812.2773	5794.2666	5795.2500	5796.2500	2906.6387	50

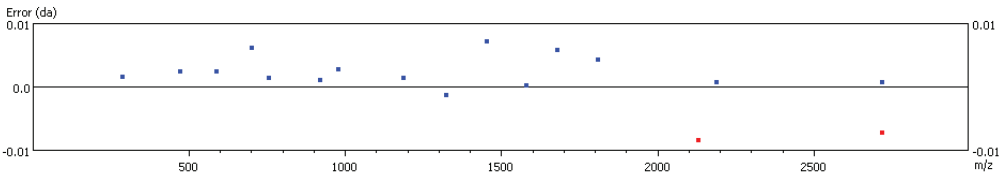


Figure 7: Example of unmatched ETD mass spectrum after homology search. (A) Full-scan MS/MS spectrum. (B) Annotated MS/MS spectrum, fragment match table and fragment error distribution.

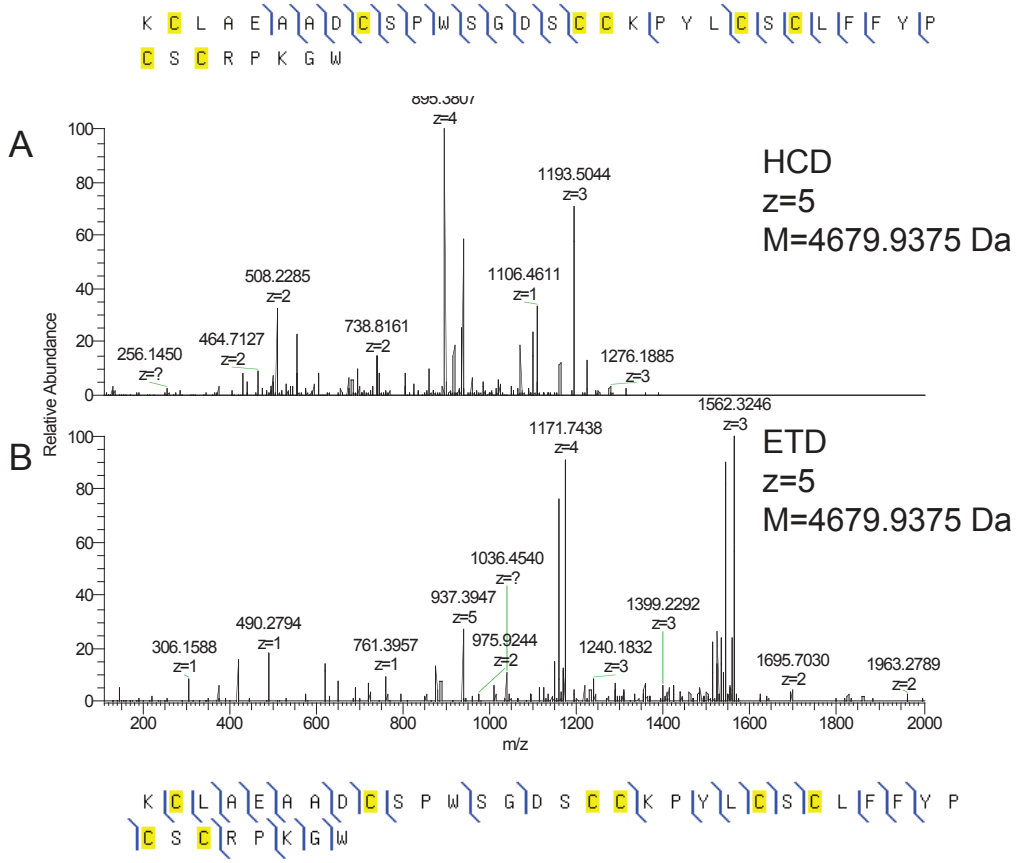


Figure 8: Sequence match with ProSight PTM from HCD (A) and ETD (B) MS/MS spectra of peptide sequence KCLAEAADCSPWSGDSCCKPYLCSCLFFYPCSCRPKGW.

Conclusion

- HR/AM nano-LC-MS/MS can be used to characterize crude spider venom peptides from small amounts of spider venom utilizing the high mass accuracy of the Orbitrap analyzer.
- A combination of HCD and ETD fragmentation allowed a large number of peptides to be identified by sequence homology matching after de novo sequencing.
- Spider venomics will be further developed to allow in depth mining of the enormous biological resource represented by small venomous animals.
- Peptide libraries based on venoms will be of major importance in the discovery of drugs and insecticides of the future.

References

1. Escoubas P, Sollod B, King GF. Venom landscapes: mining the complexity of spider venoms via a combined cDNA and mass spectrometric approach. Toxicon **2006**; 47: 650.
2. Escoubas P, King GF. Venomics as a drug discovery platform. Expert Rev. Proteomics **2009** Jun; 6(3): 221-4.

In addition to these offices, Thermo Fisher Scientific maintains a network of representative organizations throughout the world.

Africa-Other
+27 11 570 1840
Australia
+61 3 9757 4300
Austria
+43 1 333 50 34 0
Belgium
+32 53 73 42 41
Canada
+1 800 530 8447
China
+86 10 8419 3588
Denmark
+45 70 23 62 60
Europe-Other
+43 1 333 50 34 0
Finland/Norway/ Sweden
+46 8 556 468 00
France
+33 1 60 92 48 00
Germany
+49 6103 408 1014
India
+91 22 6742 9434
Italy
+39 02 950 591
Japan
+81 45 453 9100
Latin America
+1 561 688 8700
Middle East
+43 1 333 50 34 0
Netherlands
+31 76 579 55 55
New Zealand
+64 9 980 6700
Russia/CIS
+43 1 333 50 34 0
South Africa
+27 11 570 1840
Spain
+34 914 845 965
Switzerland
+41 61 716 77 00
UK
+44 1442 233555
USA
+1 800 532 4752

www.thermoscientific.com

AN63387_E 08/16S