

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

Claire Dauly¹, Pierre Escoubas^{2,3}, Glenn F. King⁴, Graham M. Nicholson⁵

¹Thermo Fisher Scientific, Courtaboeuf Cedex, France

²University of Nice Sophia Antipolis, Institute of Molecular and Cellular Pharmacology, Valbonne, France

³VenomeTech, Valbonne, France

⁴Institute for Molecular Bioscience, The University of Queensland, St. Lucia, Australia

⁵School 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 Da¹. 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 hundreds¹. 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 gener-

ate 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 μ L (200 μ g) 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 μ L 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- μ m ID x 1-cm packed-tip column, Nikkyo Technos Co. Ltd) and a C18 trapping cartridge (Captrap,

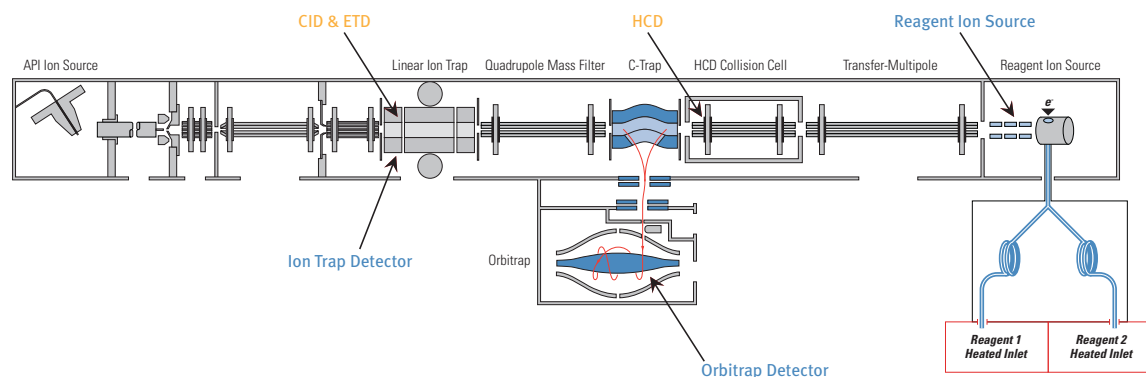


Figure 1: Schematic of the LTQ Orbitrap XL hybrid mass spectrometer with ETD.

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

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.

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		

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://pro-sightptm.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.

MGF Generation Parameters (in Proteome Discoverer)

A

Crude venom
RT: 0.00-64.98

Relative Abundance

824.6620 911.2701 961.3857 975.8767 1012.9126 847.1646 1402.1632 1074.0143 1055.6837 921.3653 1343.2825 1337.2786 1335.9357 803.5432

NL: 7.84E6
Base Peak F:
FTMS +p NSI Full
ms
[800.00-2000.00]
MS
Escoubas_HiOB

B

Reduced/alkylated venom

Relative Abundance

824.6641 895.7416 854.7484 854.9456 1173.2062 1092.7361 937.5959 1033.7128 1090.0098 1104.5114 1111.2622 1112.9891 1085.7665 921.5040 1604.0344

NL: 5.38E7
Base Peak F:
FTMS +p NSI Full
ms
[800.00-2000.00]
MS
escoubas_hiob_ra

Time (min)

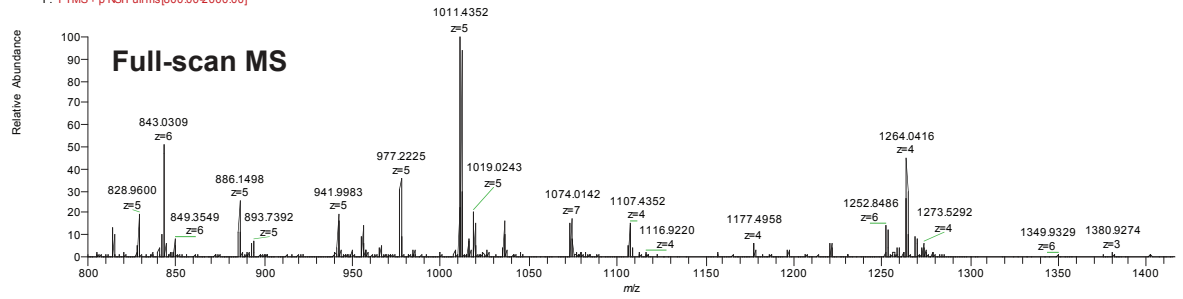
Results and Discussion

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

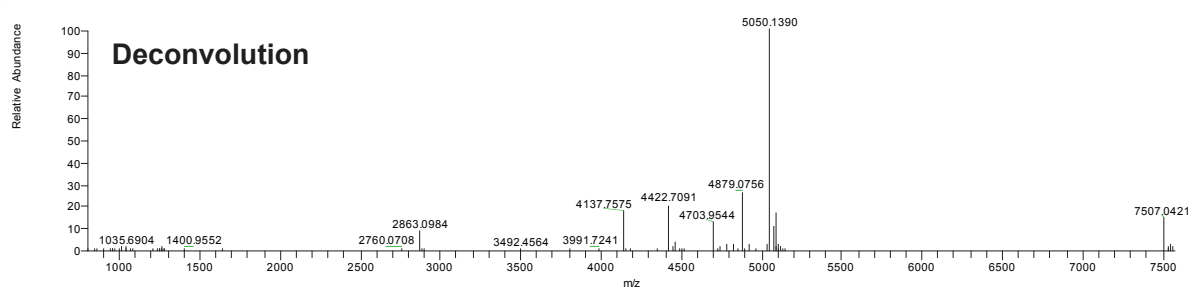
A combination of HCD and ETD experiments was used to further characterize and obtain peptide identities. Fragmentation with HCD was selected because this technique produces complete fragmentation spectra including low-mass sequence product ions and immonium ions that are useful for *de novo* sequenc-

A

Escoubas_HOB #982 RT: 33.85 AV: 1 NL: 7.47E5
 F: FTMS + p NSI Fullms[800.00-2000.00]

**B**

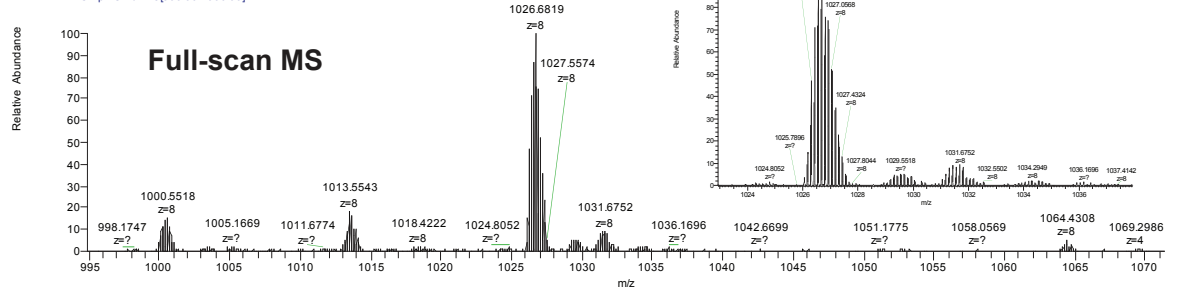
Escoubas_HOB_XT_00001_M_ #2 RT: 2.00 AV: 1 NL: 1.32E6
 F: FTMS + p NSI Fullms[800.00-2000.00]



escoubas_hibb_ra

C

escoubas_hibb_ra #1211 RT: 29.11 AV: 1 NL: 5.02E6
 T: FTMS + p NSI Fullms[800.00-2000.00]



escoubas_hibb_ra_XT_00001_M_101130151452 #2 RT: 2.00 AV: 1 NL: 3.28E6
 T: FTMS + p NSI Fullms[800.00-2000.00]

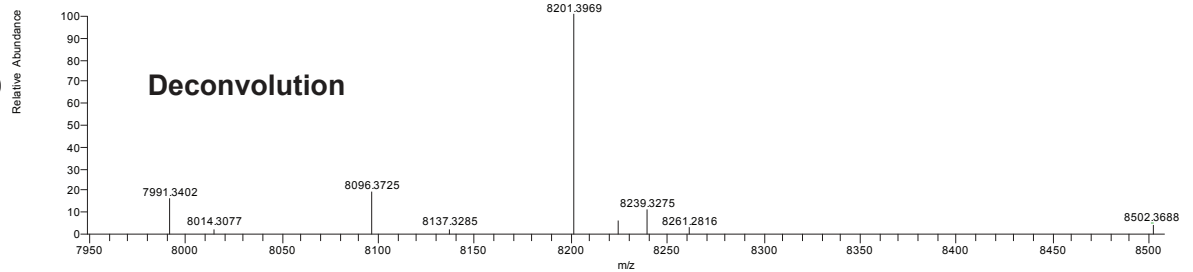
D

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).

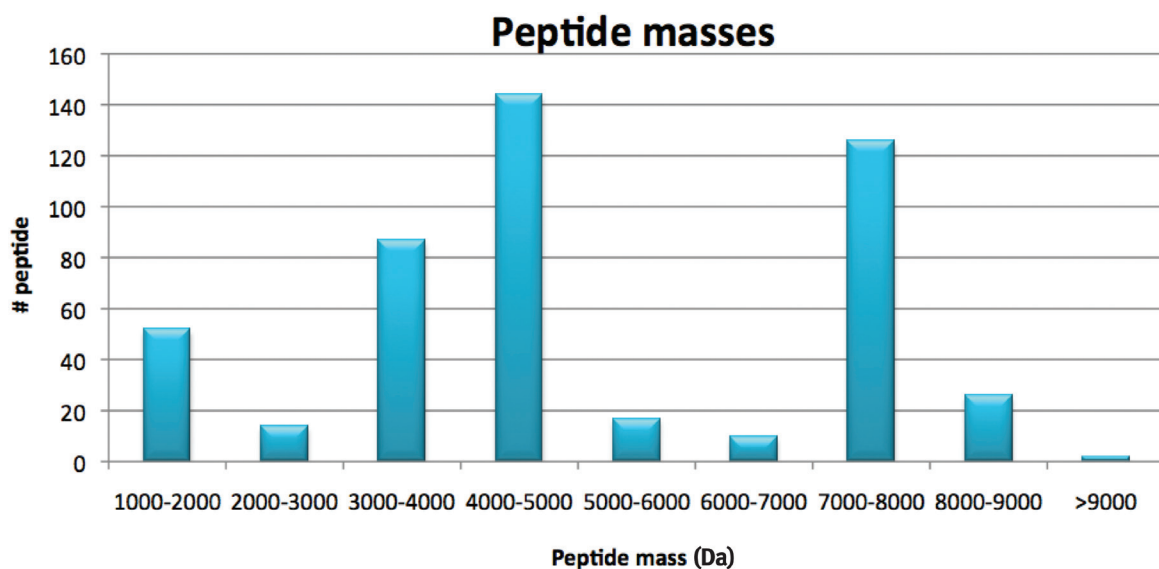


Figure 4: Venom peptide mass distribution.

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)AKKRNC(+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)TYKANENGNOVKRC(+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
DAQRFRSSPTC(+57.02)IPTGQPC(+57.02)PYNENC(+57.02)C(+57.02)SQSC(+57.02)TYKANEN
EIRKPIDTEKADAERVLD(+57.02)VNLLGC(+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)SEFPVHC(+57.02)RELGLVDDL(+57.02)MSPGETTDSGRYLFFC(+57.02)PC(+57.02)ETGLRC(+57.02)DKNDWTC(+57.02)
GVLD(+57.02)VVNTLGC(+57.02)SSDKDC(+57.02)C(+57.02)GMTPSC(+57.02)TLG
GVLD(+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)APSVGGLVGG
IMASTFAQKC(+57.02)GDQVC(+57.02)GAGTC(+57.02)C(+57.02)AEYPELHC(+57.02)KRVGQLYDL(+57.02)VDSEATKDSGNHLFFC(+57.02)PC(+57.02)DEGMYC(+57.02)DMNSWSC(+57.02)Q
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)KRVGQLYDL(+57.02)VDSEATKDSGNHLFFC(+57.02)PC(+57.02)DEGMYC(+57.02)DMNSWSC(+57.02)QKKT
KC(+57.02)GDQVC(+57.02)GAGTC(+57.02)C(+57.02)AEYPELHC(+57.02)KRVGQLYDL(+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)KRVGQLYDL(+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)SDPNKTDGSHLFFC(+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)VNLLGC(+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)VPSVGGVGG
LLDVATVLGGLGAGESDMRKDVMGLFRRALC(+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)AVFSQNH(+57.02)RRLSQMYDL(+57.02)SDHSDASPSGNLFFC(+57.02)PC(+57.02)EPGLHC(+57.02)DRNTW
NTATGVLALLVATLGC(+57.02)LEVEETRADLQAFESYEGEAAEKI
PGSVC(+57.02)DGNESDC(+57.02)KC(+57.02)YGKWHKC(+57.02)JRC(+57.02)PWKWHFTGEGPC(+57.02)TC(+57.02)EKGMKHTC(+57.02)LTCLHC(+57.02)PNKAEWGLDWRSEES
PTC(+57.02)IPTGQPC(+57.02)PYNENC(+57.02)C(+57.02)NQSC(+57.02)TYKANENGNOVKRC(+57.02)D
QC(+57.02)GDETC(+57.02)GAGTC(+57.02)C(+57.02)AVFSQNH(+57.02)RRLSQMYDL(+57.02)SDHSDASPSGNLFFC(+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)TYKANENGNOVK
SAQQC(+57.02)GDETC(+57.02)GAGTC(+57.02)C(+57.02)AVFSQNH(+57.02)RRLSRMYDL(+57.02)SDHADASPSGNLFFC(+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)TYKANENGNOVKRC(+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)TYKANENGNOV
STC(+57.02)TPTDQPC(+57.02)PYDESC(+57.02)C(+57.02)SGSC(+57.02)TYKANENGNOVKRC(+57.02)
STC(+57.02)TPTDQPC(+57.02)PYDESC(+57.02)C(+57.02)SGSC(+57.02)TYKANENGNOVKRC(+57.02)D
STC(+57.02)TPTDQPC(+57.02)PYDESC(+57.02)C(+57.02)SGSC(+57.02)TYKANENGNOVKRC(+57.02)D
STC(+57.02)TPTDQPC(+57.02)PYHESC(+57.02)C(+57.02)SGSC(+57.02)TYKANENGNOVKRC(+57.02)
STC(+57.02)TPTDQPC(+57.02)PYHESC(+57.02)C(+57.02)SGSC(+57.02)TYKANENGNOVKRC(+57.02)D
TC(+57.02)ILPTGQPC(+57.02)PYNENC(+57.02)C(+57.02)SQSC(+57.02)TYKANENGNOVKRC(+57.02)D
TC(+57.02)IPTGQPC(+57.02)PYNENC(+57.02)C(+57.02)SQSC(+57.02)TYKANENGNOVKRC(+57.02)D
TCIRTDQPC(+57.02)PYNESC(+57.02)C(+57.02)SGSC(+57.02)TYKANENGNOVKRC(+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
0Q	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

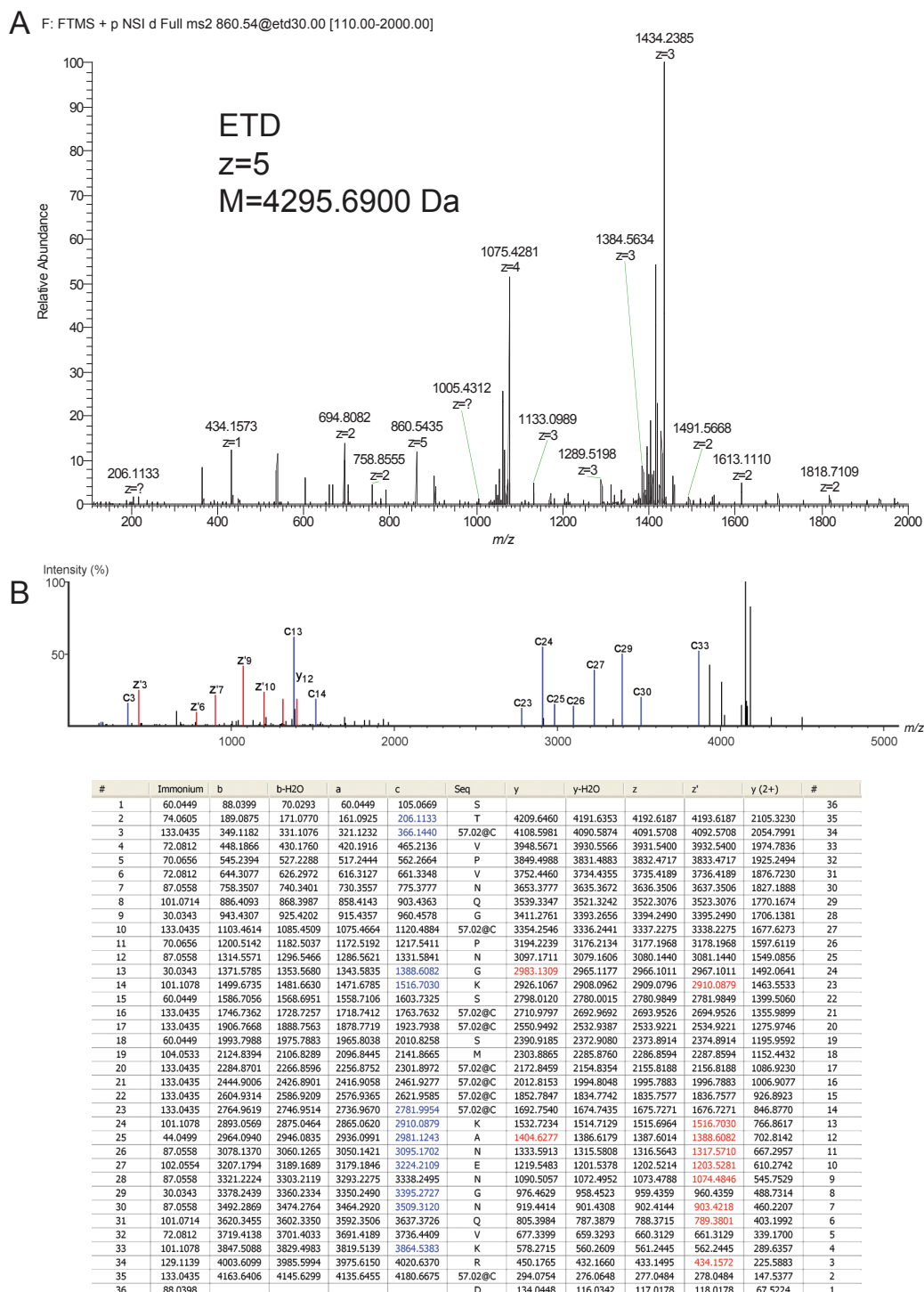


Figure 5: Example of homology-matched ETD spectrum. (A) Full-scan MS/MS spectrum. (B) Annotated MS/MS spectrum and fragment match table.

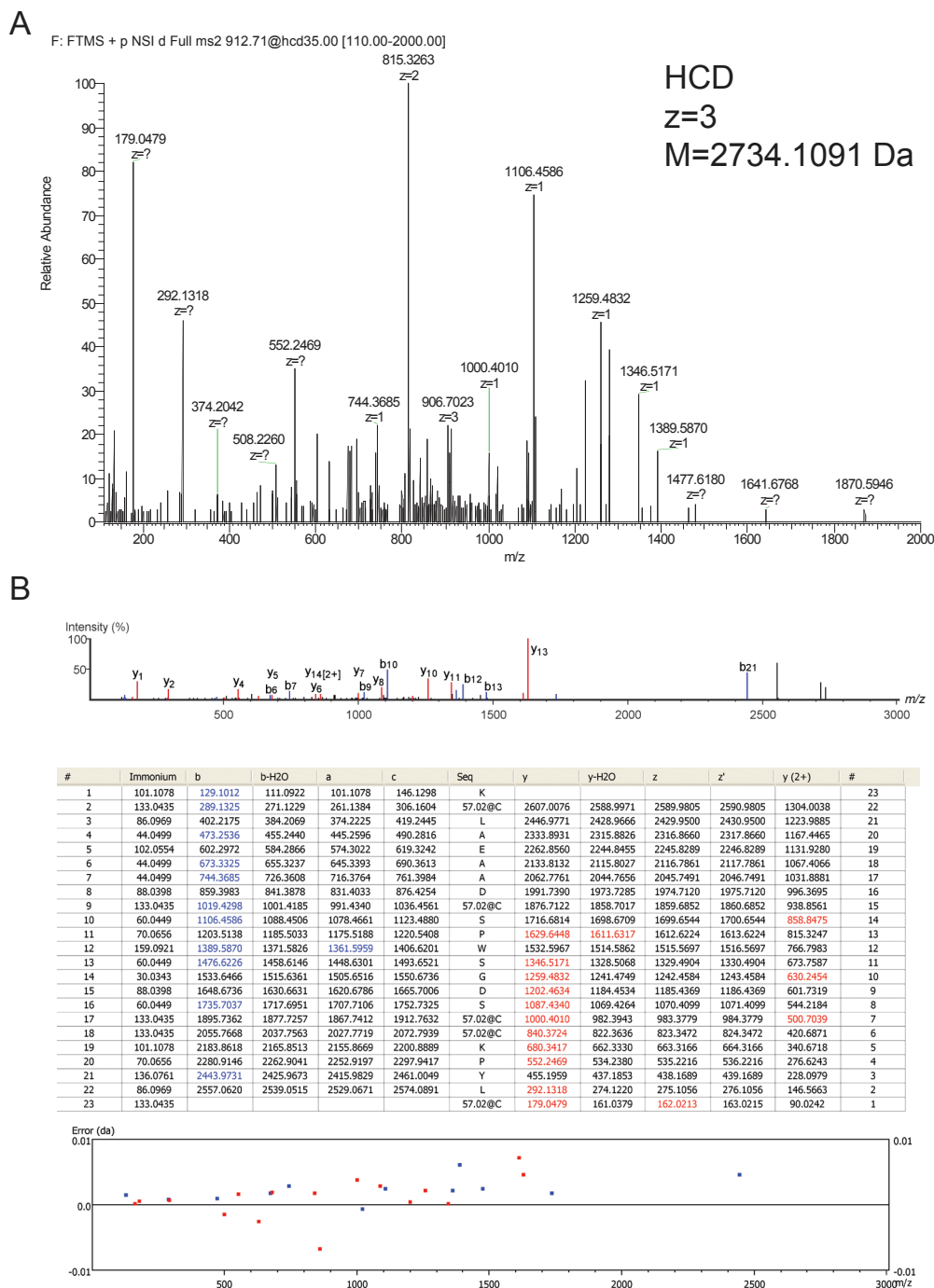


Figure 6: Example of homology-matched HCD mass spectrum. (A) Full-scan MS/MS spectrum. (B) Annotated MS/MS spectrum, fragment match table and fragment error distribution.

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-DESCSGSCTYKANENGNGVKRCD (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.

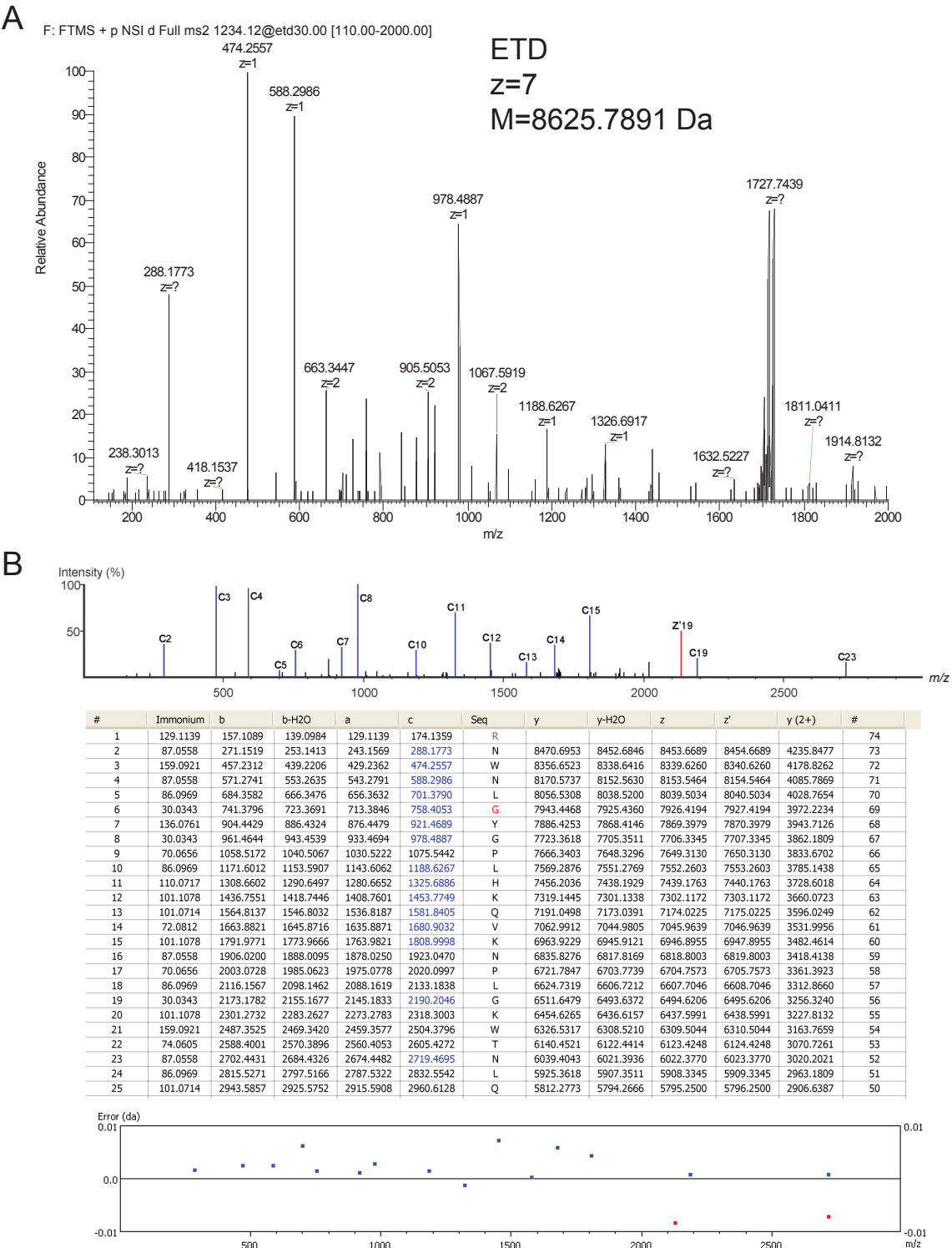


Figure 7: Example of unmatched ETD mass spectrum after homology search. (A) Full-scan MS/MS spectrum. (B) Annotated MS/MS spectrum, fragment match table and fragment error distribution.

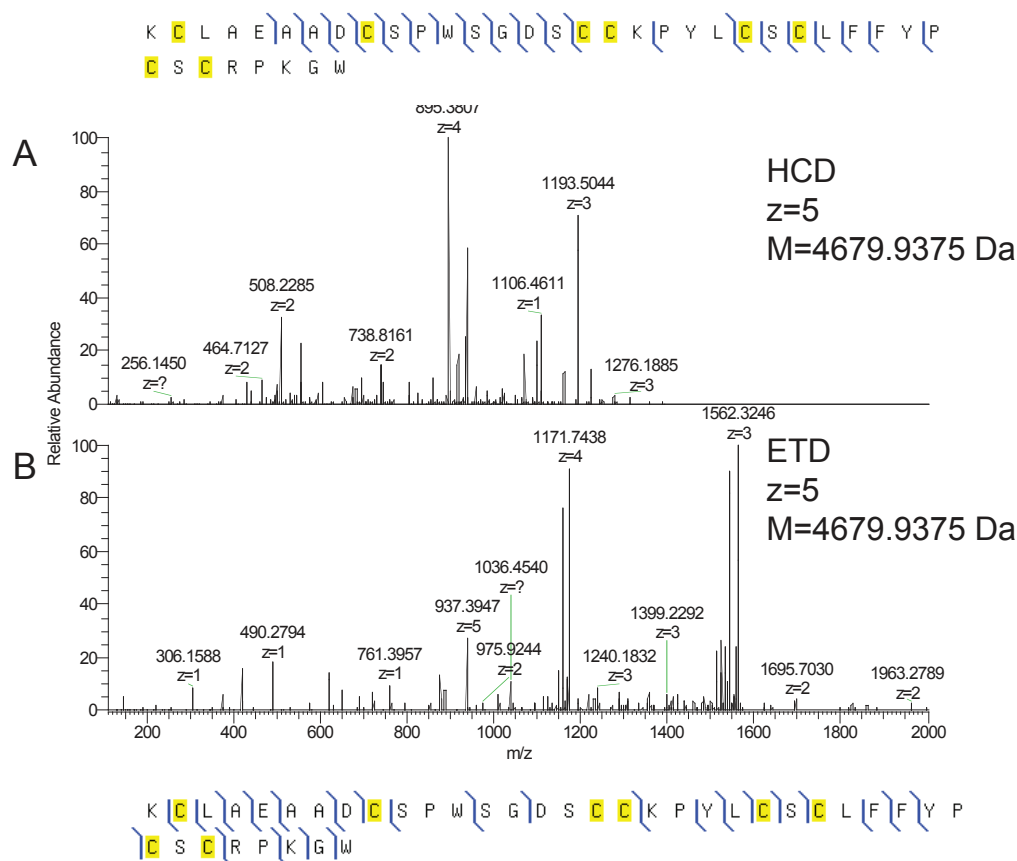


Figure 8: Sequence match with ProSight PTM from HCD (A) and ETD (B) MS/MS spectra of peptide sequence KCLAEADCSPWSDSCCKPYLCSCLFFYP CSCRPGW.

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 venomomics 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. Venomomics 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

Legal Notices: ©2016 Thermo Fisher Scientific Inc. All rights reserved. PEAKS is a registered trademark of Bioinformatics Solutions Inc. All other trademarks are the property of Thermo Fisher Scientific Inc. and its subsidiaries. This information is presented as an example of the capabilities of Thermo Fisher Scientific Inc. products. It is not intended to encourage use of these products in any manners that might infringe the intellectual property rights of others. Specifications, terms and pricing are subject to change. Not all products are available in all countries. Please consult your local sales representative for details.

AN63387_E 08/16S