Application Note: 30137

Discovery and *De Novo* Sequencing of Insect Neuropeptides Using the LTQ Orbitrap XL™

Markus Kellmann, Thomas Moehring, Thermo Fisher Scientific, Bremen, Germany,
Martijn Pinkse, Peter Verhaert, Delft University of Technology, Delft, The Netherlands

Key Words

- LTQ Orbitrap XL
- *De Novo* Sequencing
- HCD Higher Energy Collisional Dissociation
- Nano LC-MS/MS
- Neuropeptides

Introduction

Neuropeptides are important chemical messengers in animals. They are implicated in nearly all physiological processes, acting as neurotransmitters, neuromodulators or classical neurohormones. Recently, mass spectrometry (MS) has been used to localize neuropeptides in neuroendocrine tissue[1]. A sensitive and reliable identification of novel peptide sequences and post-translational modifications is of great interest, because bioactive (neuro)peptides may serve as drug leads for pharmaceutical development. The combination of on-line nano HLPC and high mass accuracy/high resolution mass spectrometry is ideally suited for *de novo* sequencing of peptides.

MS techniques such as MALDI-MS/MS give full structural elucidation only for a limited number of small peptides (< 20 amino acids). However, for larger peptides, the sequence coverage obtained is often incomplete. Analysis using QTOF instruments gives results which lack the necessary mass accuracy resulting in uncertainties or ambiguities for complete *de novo* sequence interpretation.

With the introduction of the LTQ Orbitrap XL (Figure 2) with its octapole collision cell based Higher Energy Collisional Dissociation (HCD), it is possible to carry out the most rigorous *de novo* sequencing, since high resolution, accurate mass data can be acquired over the entire mass range of *m/z* 50 to 2000. This is of considerable benefit especially because highly informative signals in the low mass range can be detected simultaneously with higher *m/z* b- and y-ions resulting in complete sequence coverage.

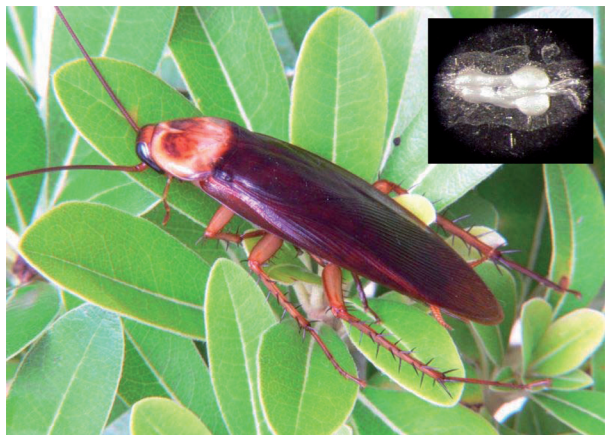

Figure 1: American cockroach (*Periplaneta americana*). The inset shows a microscope view of the corpora cardiaca/corpora allata complex (cc/ca).

For this study, the corpora cardiaca/corpora allata complex (cc/ca; the insect equivalent of the mammalian pituitary gland) of cockroaches (*Periplaneta americana* and *Periplaneta australasiae*; Figure 1) served as the source of sets of biological (neuro)peptides to be structurally elucidated by nano HPLC coupled to LTQ Orbitrap XL. The dataset combines information acquired in a single nano LC-MS/MS run using different types of fragmentation (both CID and HCD). The comprehensive information obtained led to the structural characterization of more than 20 known neuropeptides (including pyrokinins, sulfakinins, allatostatins, hypertrehalosemic hormones, corazonin) by *de novo* sequencing as well as by database searching. These identifications provide a clear demonstration for the robustness of our *de novo* sequencing approach.

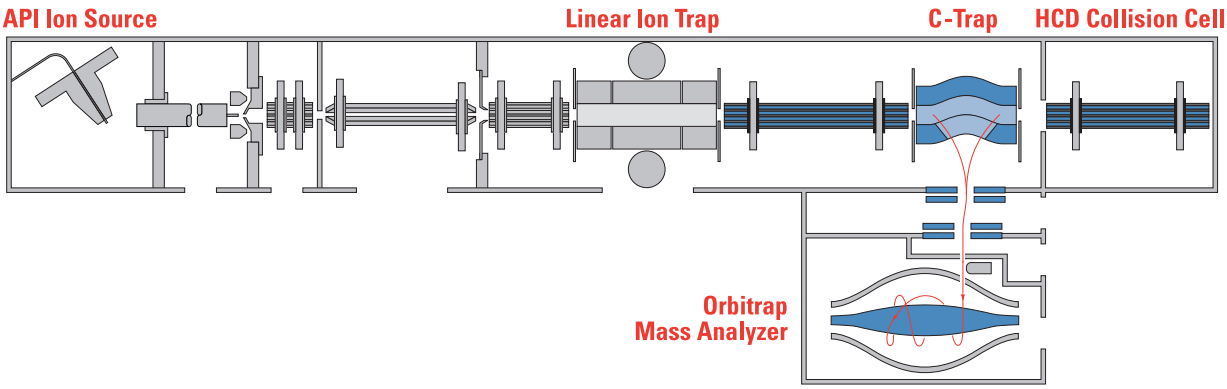

Figure 2: Schematic of LTQ Orbitrap XL. The additional octapole collision cell is placed behind the curved C-Trap. For Higher Energy Collisional Induced Dissociation (HCD) ions are transferred to the new octapole and, after fragmentation, they enter C-Trap again before being ejected into the Orbitrap mass analyzer.

Thermo SCIENTIFIC

Because of the high sample complexity (shown in Figure 3), *de novo* sequencing was setup in an automated fashion using the PEAKS Studio software, which incorporates also database and error tolerant search routines.

Methods

Cockroach species were obtained from local zoological gardens. Neuropeptides from isolated cc/ca were extracted in 1 min. using 50% MeOH, 0.1% formic acid (10 µL per animal tissue equivalent, and after a 5-fold dilution with water (0.1% FA), injected into a nano-HPLC system using a 75 µm ID (15 cm, C18, 5 µm) analytical column. The injection volume of 10 µL corresponded to the peptide equivalent of one fifth of a cc/ca-complex. Data were acquired with a hybrid linear ion trap - orbitrap mass spectrometer (LTQ Orbitrap XL, Thermo Fisher Scientific, Bremen) operating in a Data Dependent™ mode. Full scan MS spectra were acquired at 60,000 (FWHM) resolving power settings. The most abundant multiply charged ion was selected for collision-induced dissociation (CID) and HCD fragmentation of the same precursor both detected in the orbitrap mass analyzer.

Database search was performed by MASCOT™ 2.1.04 (Matrix Science, London, UK) on both SwissProt or nrNCBI databases. Tolerances were adjusted to 3 ppm (precursor) and 0.05 Da (fragment).

For *de novo* sequencing we used the PEAKS Studio 4.5 (BSI, Waterloo, Canada) software. PEAKS *de novo* sequencing as well as error tolerant searches and PEAKS protein identification were performed with tolerances of 3 ppm for the precursor and 0.05 Da for the fragments and with the same databases as used for MASCOT searches. For all data processing we used amidated *C*-terminus (X_{amide}), *N*-terminal pyroglutamic acid Q (q), and tyrosine sulfation (y) as variable modifications.

Instrument parameters:

Name	Value
Max. inject time (MS/MS²) [ms]	100/500
AGC target value	$2e^5$
Normalized collision energy [arb]	40
Isolation width [Da]	3

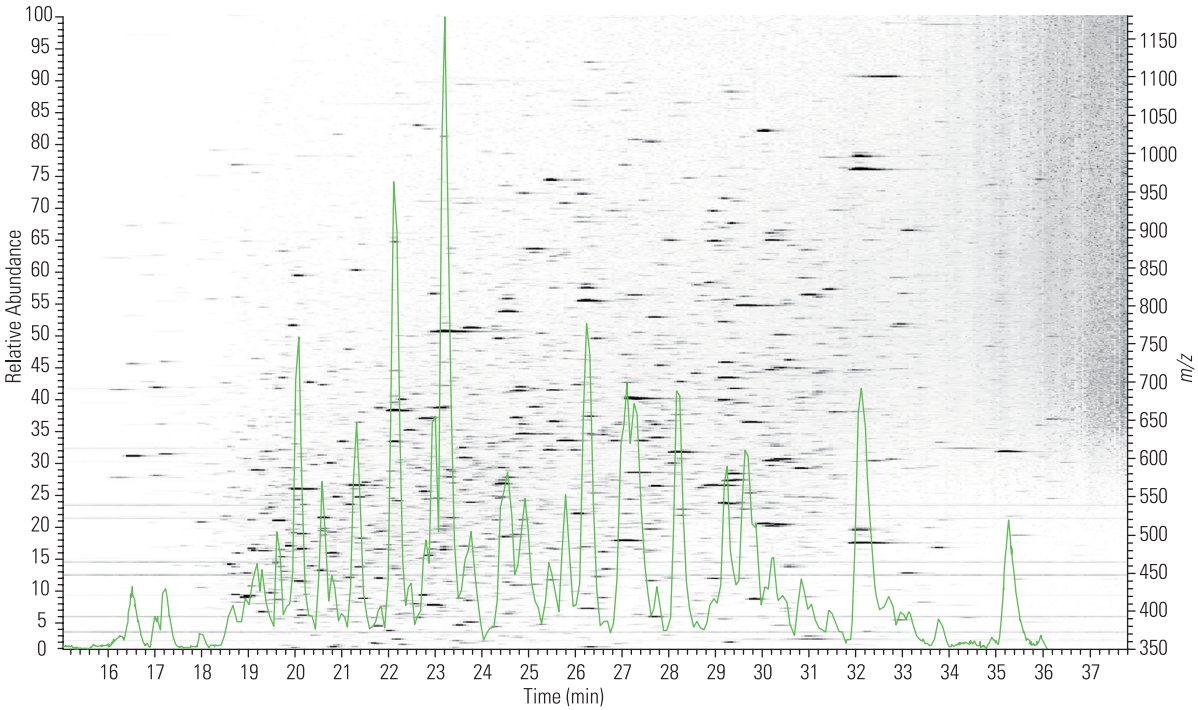

Figure 3: Peptide display with overlayed base peak chromatogram, showing the complexity of the sample (one fifth cc/ca equivalent of *Periplaneta americana*).

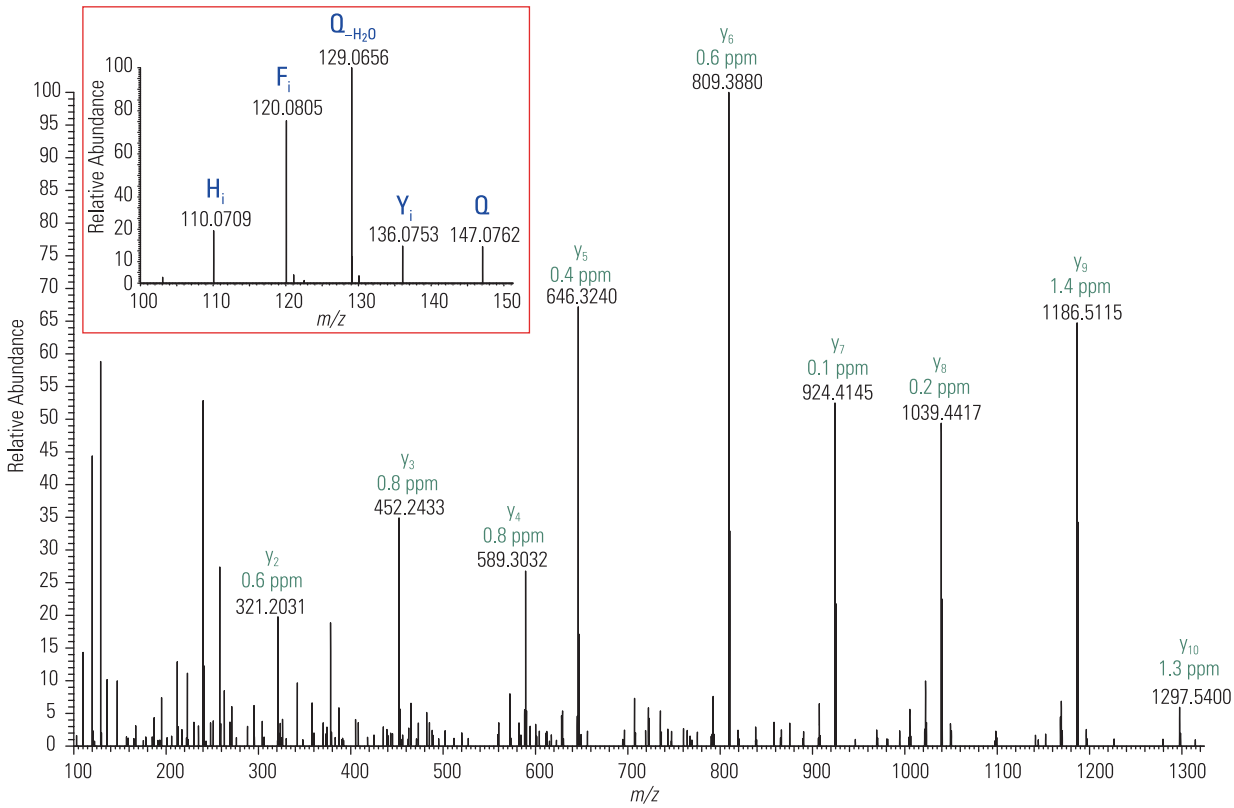

Figure 4: MS/MS Spectrum of perisulfakinin (Pea-SK-I) (*m/z*=722.310552$^+$). Mass accuracy is better than 3 ppm over the whole mass range. The insert shows the low *m/z* range with indicative immonium and other amino acid-specific ions (sequence: EQFDDYGHMRF$_{amide}$).

Results

In this work we have addressed the main requirements for *de novo* sequencing.
These are:

- High mass accuracy of both precursor and fragment ions (< 3 ppm)
- Sensitive and highly efficient generation of fragmentation spectra
- Generation of spectra with maximal interpretable information content
- High degree of automation

High mass accuracy is achieved by detecting MS/MS ions in the orbitrap mass analyser. As it is demonstrated in Figure 4, the mass accuracy for the parent and fragment ions is less than 3 ppm (with external calibration) for the complete mass range. In this example a virtually complete sequence of the peptide EQFDDYGHMRF$_{amide}$ is determined (nearly complete y-ion series). The C-terminus is incompletely assigned as there are two possible alternatives: An oxidized and amidated methionine or an amidated phenylalanine. This information is directly available from the low mass range, where immonium and amino acid specific signals can be found. In this example we detect immonium ions for histidine, tyrosine and phenylalanine, as well as two glutamine specific ions., indicating the presence of phenylalanine at the C-terminal position.

Furthermore, the finding is supported by the fact that a potential oxidized and amidated methionine would have a *m/z* of 147.05866, which is a mass deviation of approximately 120 ppm compared to the measurement. We conclude, thus, the correct amino acid sequence contains a phenylalanine at its C-terminus, which is in excellent agreement with the precursor *m/z* of 722.310552$^+$ (mass deviation 2 ppm).

Another interesting finding from the dataset is depicted in Figure 5. CID fragmentation in the linear ion trap results in very low energy fragmentation pathways. This has many advantages, but in some cases the fragment spectrum consists only of a neutral loss ion, in particular when post-translational modifications, such as phosphorylations and tyrosine sulfations are present. The latter is a common modification in neuropeptides.

Figure 5 (upper panel) shows a typical CID MS/MS spectrum from the tyrosine sulfated neuropeptide leucosulfakinin-2 (qSDDyGHMRF$_{amid}$) indicating the neutral loss of 79.9578 Da, which corresponds to a loss of SO_3 (mass deviation of 2.1 ppm). In case of CID, an additional scan event, i.e. performing an MS^3 experiment, is necessary to get sequence information. However, the lower panel in Figure 5 shows the same precursor, which has undergone HCD fragmentation. Due to the slightly higher collision energy, the sequence information can be directly deduced from the MS/MS experiment.

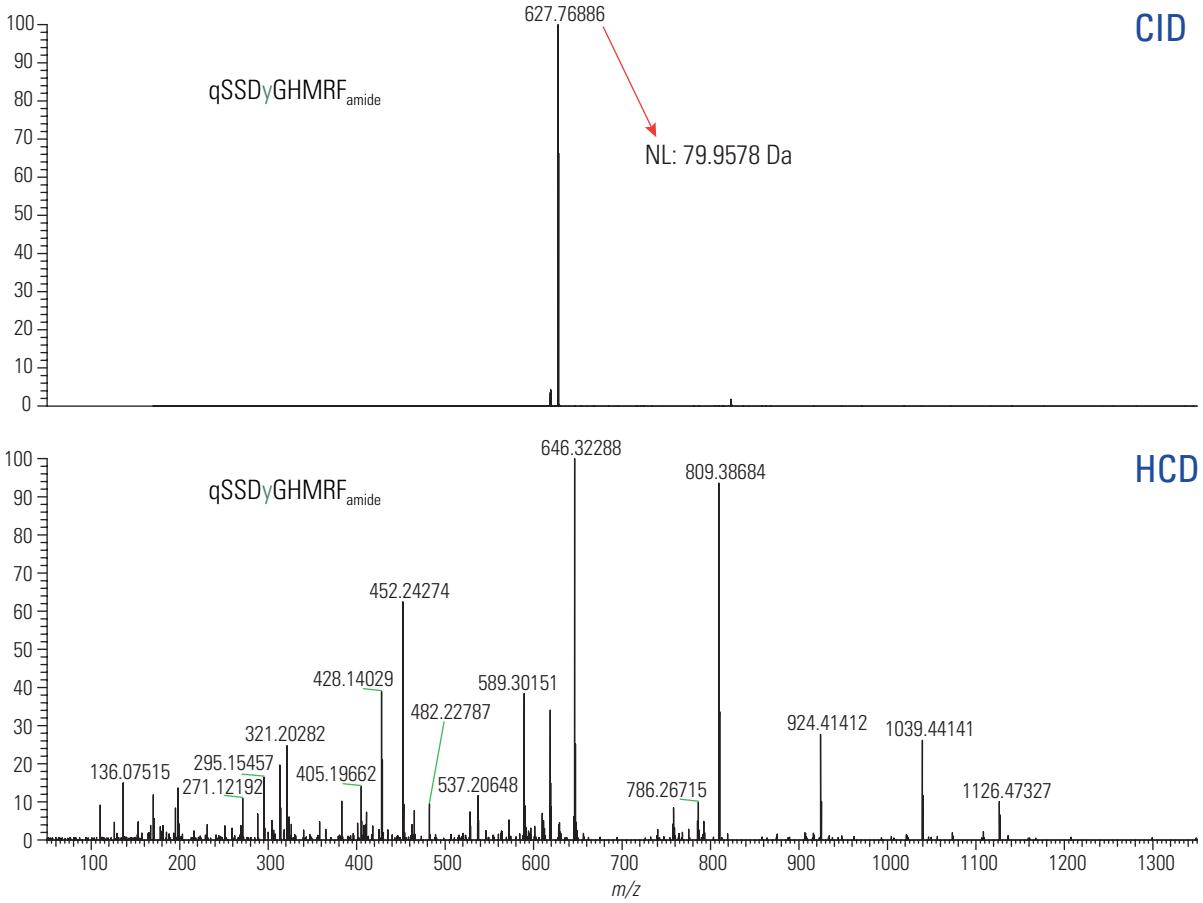

Figure 5: MS/MS fragmentation of a sulfated tyrosine neuropeptide from *P. americana*. The upper spectrum shows CID fragmentation with the neutral loss (NL) of 79.9578. Sequence information can not obtained from this spectrum. The lower spectrum shows HCD fragmentation of the same precursor ion. Here the sequence can be obtained directly from the MS/MS spectrum.

		F.-type		*De novo* search results [HCD]	
AA-sequence	**Accession**	**CID**	**HCD**	**score**	**rank**
SKYMYGFGL$_{amid}$	(P12764) Allatostatins precursor	✓	✓	69	1
ARPYSFGL$_{amid}$	(P12764) Allatostatins precursor	✓	✓	80	1
pQVNFSPNW$_{amid}$	(P84261) Hyper-trehalosaemicfactor1	✓	✓	66	1
pQLTFTPNW$_{amid}$	(P84256) Hyper-trehalosaemicfactor2	–	✓	49	2
DDYGHMRF$_{amid}$	(P36885) Perisulfakinin	✓	✓	78	1
pQSDDYGHMRF$_{amid}$	(P67802) Leucosulfakinin-2	–	✓	99	1
QSDDYGHMRF$_{amid}$	(P67802) Leucosulfakinin-2	✓	✓	99	1
pQSDDyGHMRF$_{amid}$	(P67802) Leucosulfakinin-2	–	✓	99	1
QSDDYGHMRF$_{amid}$	(P67802) Leucosulfakinin-2	–	✓	99	1
EQFDDYGHMRF$_{amid}$	(P36885) Perisulfakinin	✓	✓	75	1
pQDVDHVFLRF$_{amid}$	(P21144) Leucomyosuppressin	✓	✓	67	1
QDVDHVFLRF$_{amid}$	(P21144) Leucomyosuppressin	✓	✓	99	1
DHLPHDVVSPRL$_{amide}$	(P84410) Pyrokinin4	✓	✓	99	1
DHLPHDVVSPRL	(P84410) Pyrokinin4	✓	✓	99	1
NDPEVPGMWFGPRL$_{amide}$	(P84368) Pyrokinin6	✓	✓	40	2
PGMWFGPRL$_{amid}$	(P84368) Pyrokinin6	✓	✓	50	2
HLPHDVYSPRL$_{amid}$	(P84410) Pyrokinin4	✓	✓	75	1
HTAGFIPRL$_{amid}$	(P82691) Pyrokinin1	–	✓	99	1
HLPHDVYSPRL$_{amid}$	(P84410) Pyrokinin4	✓	✓	25	1
PEVPGMWFGPRL$_{amid}$	(P84368) Pyrokinin6	✓	✓	99	1
pQTFQYSRGWTN$_{amid}$	(Q26377) Corazonin	–	✓	50	2

Table 1. Neuropeptides that are scored significantly high by database searching (MASCOT) are marked with ✓. *De novo* scores and ranking of spectra fragmented by means of HCD are shown in the last two columns. Database search and *de novo* sequencing was performed with tolerance windows of 3 ppm (precursor) and 0.05 Da (fragments).

A critical step towards successful *de novo* sequencing on a reasonable time scale – is an automation of the whole process. In order to properly set up automated sequencing runs, one has to carefully select the correct parameters. This includes the settings necessary to avoid false positives, while at the same time being sufficiently sensitive to deliver the maximum number of correct identifications. To optimize these parameters within PEAKS software, we (*de novo*) sequenced a whole data set of *Periplaneta australasiae* and performed a database search with MASCOT. We chose this species as there are very few entries in databases, nevertheless, it has close phylogenetic relationship to the better characterized *P. americana*. As a proof-of-concept we evaluated all unambiguous database hits and compared them to the *de novo* result obtained with PEAKS using different settings. The criterion for accepting the *de novo* sequencing settings was, that all of the database hits scored above the "identity threshold" in MASCOT, had to deliver the same sequence within the first five suggestions of PEAKS and having a *de novo* score above 20%. The results are summarized in Table 1. All 21 database hits with scores above the identity threshold are listed. The check mark (✓) indicates, if a peptide was found by database search. Notable is, that some peptides were not found by CID fragmentation, which results from either a post-translational modification (Y-sulfation), which results in only a neutral loss by CID fragmentation (Figure 5), or from different fragmentation pathways in CID, resulting in incomplete fragmentation and consequently lower scores.

Only HCD spectra were processed by PEAKS. The peaks scores are displayed in the last two columns in Table 1, indicating that *de novo* scores are almost above 50% and the correct sequence is always within the top five candidates.

The identified neuropeptides from database search can be sorted into different (functional) families:

– allatostatins (juvenile hormone synthesis)

– sulfakinins, myosuppressins, pyrokinins ([visceral] muscle contraction regulation)

– hypertrehalosaemic peptides (carbohydrate regulation)

– corazonin (heart beat regulation)

Based on these promising results, automated homology searches were performed by SPIDER (an equivalent to BLAST) to find undescribed neuropeptides. In this type of error tolerant search the top five candidates of the *de novo* search were searched against protein databases to look for homologies between the candidate and database entries from other species. The search can yield new PTMs of known peptides, amino acid exchanges and homologies to neuropeptides from other species. In this work the combination of *de novo* sequencing and subsequent homology searching the SwissProt and nrNCBI databases yielded 21 additional potential neuropeptides. Some of them need further inspection by a targeted approach to get the most informative fragment spectra for *de novo* sequencing (e.g. MS^n), since they indeed show partial sequences, which allow categorizing them in one of the above mentioned families, but have ambiguities in other parts of the sequence.

The example in Figure 6, shows a neuropeptide which had not been previously reported in cockroaches. The homology search suggests similarity to a putative allatostatin peptide of mosquitoes with the sequence $EPGWNNLKGLW_{amide}$ together with a mass shift of – 14.01565 Da, which is consistent with substitution of glutamic acid for aspartic acid at the *N*-terminus.

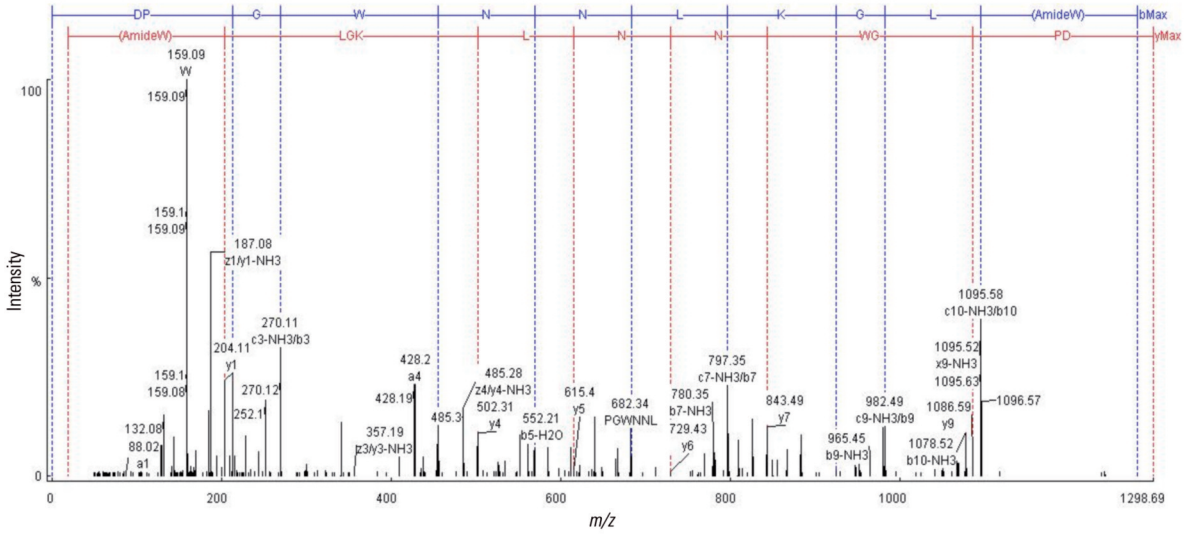

Figure 6: The automated *de novo* sequencing result and manual inspection agree on the sequence of $DPGWNNLKGLW_{amide}$.

Conclusions

We have carried out work specifically focused on automating the *de novo* sequencing. This included optimization of the settings for biologically relevant samples. In the course of this study we evaluated robustness and reliability by comparison to database searching. In addition we used *de novo* sequencing and error tolerant homology searching to find many new members of known neuropeptide families in cockroaches as well as two peptides previously not reported in the list of neuro-peptides from cockroaches.

Using HCD fragmentation in a new octapole collision cell improved considerably the *de novo* sequencing results compared to CID in linear ion trap or C-Trap.

We conclude that the method can be further improved by applying collision energy profiles to further improve fragmentation performance, or use base peak ejection to remove high intensity peaks prior to the detection to detect lower abundant peptide species.

Future work will concentrate on the new peptides, in particular on the characterization of their biological functions in cockroaches.

References

[1]Verhaert, P.D., Prieto Conaway, M.C., Pekar, T.M., Miller, K., Neuropeptide imaging on an LTQ with vMALDI source, The complete 'all-in-one' peptidome analysis. *Int. J. Mass Spec.* **2007**, 260, 177–184.

Acknowledgements

We thank Louwerens-Jan Nederlof from the Rotterdam Zoo and Ben Van Dijck from Antwerp Zoo for providing us with cockroachesused in this study.

View additional Thermo Scientific LC/MS application notes at: **www.thermo.com/appnotes**

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

Africa
+43 1 333 5034 127
Australia
+61 2 8844 9500
Austria
+43 1 333 50340
Belgium
+32 2 482 30 30
Canada
+1 800 530 8447
China
+86 10 8419 3588
Denmark
+45 70 23 62 60
Europe-Other
+43 1 333 5034 127
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
+1 608 276 5659
Middle East
+43 1 333 5034 127
Netherlands
+31 76 579 55 55
South Africa
+27 11 570 1840
Spain
+34 914 845 965
Sweden/Norway/ Finland
+46 8 556 468 00
Switzerland
+41 61 48784 00
UK
+44 1442 233555
USA
+1 800 532 4752

www.thermo.com

AN30137_E 08/16C