Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography B, 803 (2004) 3–16
JOURNAL OF CHROMATOGRAPHY B
www.elsevier.com/locate/chromb

Review

Peptidomics

Geert Baggerman*, Peter Verleyen, Elke Clynen, Jurgen Huybrechts, Arnold De Loof, Liliane Schoofs

Laboratory of Developmental Physiology, Genomics and Proteomics, Katholieke Universiteit Leuven, Naamsestraat 59, B-3000 Leuven, Belgium

Abstract

Peptides occur in the whole animal kingdom, from the least evolved phyla with a very simple nervous system (coelenterates) to the highest vertebrates and are involved in most, if not all, physiological processes in animals.

Knowing the amino acid sequence of peptide hormones or neurotransmitters is important since this allows for synthesis of large quantities of peptides to perform further functional analysis. Immunocytochemistry, radioimmunoassays (RIA), enzyme-linked immunosorbant assays (ELISA) and mass spectrometry can then provide information on the temporal and spatial distribution and quantification of the (neuro)peptide. Ever since the 1970s, a wealth of peptides has been discovered and investigated and this flow seems to be far from over. This is partially due to the use of new approaches mainly based on chromatographical purifications as well as molecular biological techniques.

Surprisingly, peptides have so far been neglected in most proteomic studies. The finalization of the genome projects has opened new opportunities for rapid identification and functional analysis of (neuro)peptides as well. In analogy with the proteomics technology, where all proteins expressed in a cell or tissue are analyzed, the peptidomic approach aims at the simultaneous visualization and identification of the whole peptidome of a cell or tissue, i.e. all expressed peptides with their post-translational modifications (PTMs). This technology provides us with a fast and efficient tool to analyze the peptides from any tissue. This paper reviews the approaches that have been used so far to achieve this.

Keywords: Reviews; Peptide profiling; Proteomics; Neuropeptides

Contents

1. Introduction 3
2. Peptidomics technology 4
   2.1. Mass spectrometry 4
   2.2. Peptide separation 5
   2.3. Peptide quantification 7
   2.4. Post-translational modifications 7
3. Peptidomics: data 8
4. Differential peptidomics: data 9
5. Peptidomics in the post-genome era 10
6. General conclusion 13
Acknowledgements 13
References 13

1. Introduction

The finalization of the different genome projects at the end of the 20th century has entered biology into a new era. At the moment the genome of eight metazoa has been fully sequenced; the yeasts *Saccharomyces cerevisiae* [1] and *Schizosaccharomyces pombe* [2], the plant, *Arabidopsis thaliana* [3], three invertebrates, the nematode, *Caenorhabditis elegans* [4], the fruit fly, *Drosophila melanogaster* [5] and the malaria mosquito, *Anopheles gambiae* [6]. In 2001, the human genome was nearly completely sequenced [7] and recently the genome of the mouse was completed [8].

* Corresponding author.
E-mail address: geert.baggerman@bio.kuleuven.ac.be (G. Baggerman).

1570-0232/$ – see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2003.07.019

Although the sequences of all the genes of these organisms are known at present, for most genes very little is known about their function and their interaction. This has led to the introduction of "post-genomic" techniques such as cDNA arrays and proteomics to discover the genes and/or proteins that are up or down regulated in certain physiological processes.

"Proteomics" attempts to identify all the proteins, and their post-translational modifications (PTMs), in a certain organism or tissue. The most common approach in proteomic studies is to separate and visualize all the proteins by two-dimensional electrophoresis and to subsequently identify the expressed proteins by mass spectrometric techniques. However, there are several limitations to the use of 2D-PAGE as a proteome visualization technique. One of the major constraints is the limitation of the size of proteins that can be trapped in a gel. Proteins of a molecular mass lower than 10 kDa are generally not retained and overlooked in most of the proteomic studies. Nevertheless, this mass region contains the very important groups of peptides and neuropeptides. The latter are by far the largest class of neuroactive messengers found in animals. They outnumber largely the more conventional neurotransmitters such as acetylcholine, amino acids and amines. Their high diversity in structure allows a high variety of different and specific messages that can be transmitted from one cell to another. Neuropeptides act as neurotransmitter, neurohormone or neuromodulator and are involved in the regulation of most, if not all, physiological processes in metazoa.

Despite their importance in animal physiology, only a very small number of scientific manuscripts deal with an integrated analysis of the peptide content of an organism, body fluid, tissue or cell. This type of analysis was named "peptidomics" in analogy to "proteomics". This is clearly demonstrated by the fact that Pubmed retrieves only 12 manuscripts when using the term peptidomic(s) as a query. The term peptidomics, however, is fairly recent [9,10] and several other papers deal with similar integrated peptidome analysis.

2. Peptidomics technology

A proteomic analysis of a sample usually consists of four steps. These are: extraction of the proteins from the sample, their separation, detection and finally identification/analysis of the individual, separated proteins. The separation and quantitative detection of proteins is routinely performed by two-dimensional gel electrophoresis (2DE) and staining of the gel. Protein identification is performed by the mass spectrometric analysis of a tryptic digest of the individual proteins through peptide mass fingerprinting or tandem mass spectrometry [11].

In the past, it has been demonstrated that peptides can be analyzed by mass spectrometry requiring very little or no manipulations of the samples, just by placing the tissues directly on the MALDI target plate and applying the matrix solution [9,10,12,13] or, in the case of electrospray mass spectrometry (ESI), by a brief extraction in the spraying solvent [9,14]. In many cases, however, the complexity of the peptide sample requires a separation prior to the mass spectrometric analysis of the sample.

As in proteomics, the core technology in peptide identification is mass spectrometry. Several mass spectrometric techniques have been recently used to detect and identify neuropeptides from tissue extracts using a very limited amount of starting material. Both approaches with matrix assisted laser desorption time of flight (MALDI-TOF) and electrospray mass spectrometry have proven to be successful [9,10,15].

2.1. Mass spectrometry

As with the identification of any other molecule, knowing the molecular mass of a peptide is a crucial step in its identification process [11,16–18].

The recent success of mass spectrometry is partially due to its ability to obtain structural information from molecules, this is in addition to the relative mass. The general idea behind tandem mass spectrometry or MS/MS is: first select a precursor ion with a first mass analyzer (MS 1), then, fragment this and, finally, analyze the obtained fragments in the second mass analyzer (MS 2).

The most common tandem mass spectrometers used in peptide research are ESI-tandem mass spectrometers such as triple quadrupole instruments, ion-traps or quadrupole time of flight instruments [19,20].

In the past, discovery of novel peptides was accomplished through preparative scale isolation of bioactive compounds, purification and finally Edman-degradation-based sequencing [21]. The high sensitivity and the fact that peptide ions can be selected for fragmentation from a complex mixture (which means that it is no longer compulsory to fully purify the peptides) has dramatically reduced the amount of material needed to obtain the amino acid sequence from peptides [9].

Structural information can also be obtained with MALDI-TOF instruments equipped with an ion gate or ion selector allowing fragmentation by post-source decay (PSD). Although direct sequencing of peptides by PSD is less straightforward than with ESI-tandem MS (due to the high variety of daughter ions generated by PSD) fragmentation spectra can be used for identification of the peptides in a database. The major drawback of PSD, however, is the time-consuming spectrum acquisition and the low resolution of the ion selectors, which can be problematic when working with complex mixtures, such as tissue extracts. Recent instrument innovations, such as MALDI-Q-TOF [22] and MALDI-TOF-TOF [23,24] provide solutions for this problem and will become more important tools in protein and peptide identification. These instruments allow the se-

lection of an ion produced from a MALDI-source and subject it to collision induced dissociation (CID), combining the high speed and throughput of MALDI with the advantages of tandem mass spectrometry [19,20]. In addition, these techniques are sensitive at the low attomole level. This section gives only a brief discussion on the MS technology itself. More elaborate information can be obtained from a number of excellent reviews on the matter [11,25].

2.2. Peptide separation

In the past, it has been demonstrated that peptides can be analyzed by mass spectrometry requiring very little or no manipulations of the samples, just by placing the tissues directly on the MALDI target plate and applying the matrix solution [9,10,12,13] or, in the case of ESI, by a brief extraction in the spraying solvent [9,14]. In many cases, however, different neuropeptides are present in very different concentrations in tissues or cells. In addition, some organs or body fluids, such as brain, cerebrospinal fluid, blood, etc. may contain many hundreds to a thousands of different peptides [26–28]. In these cases, the complexity of the peptide sample requires a separation prior to the mass spectrometric analysis of the sample. In addition, some tissues or body fluids contain only very low concentrations of peptide hormones requiring a preconcentration prior to analysis [29].

The range of proteins that can be separated on a typical 2D-gel have a molecular mass that ranges between 10 and 150–200 kDa, which makes this technique obsolete for (neuro)peptides. Nevertheless, peptides can easily be separated using reversed phase chromatography [30]. Since in MALDI-TOF MS samples need to be crystallized in a matrix, the chromatographic system cannot be coupled directly to the mass spectrometer. Fractions are collected and applied off-line to the MALDI target [27]. Fractions can be collected directly on the MALDI target straight from the column by collecting the drops on the target or trough electrospray deposition. In electrospray deposition, a voltage difference is applied between the (steel) outlet of the HPLC and the sample target. In this way, the column eluent is sprayed in a thin layer on the target. These methods were applied successfully to peptidomics [27] and allow very sensitive detection of peptides even at the level of a single neuron [13,31].

Because the electrospray process transfers ions from solution directly into the gas phase, from the start of electrospray mass spectrometry, electrospray sources have been coupled to liquid chromatography [32]. Coupling an HPLC directly to the electrospray allows the analysis of peptides as they elute from the column. Following the evolution of the electrospray source to the nanoscale level [33], nanoscale capillary LC systems are taking full advantage of low flow electrospray technology. As with electrospray itself, the chromatography has gone through an evolution of down sizing this, to gain sensitivity [29].

The need to detect lower concentrations of peptides from complex protein mixtures with a high degree of automation has driven the development of ultra low flow LC systems that can be coupled directly to nanospray sources. Columns with 50–200 μm internal diameter and HPLC systems with flow splitters that can handle flow rates of a few 100 nl/min are commercially available. Currently, chromatographical separations are even performed at flow rates as low as 20 nl/min using columns with internal diameters down to 15 μm [34].

However, with such low flow rates, loading a sample onto the column is time consuming. If, for instance, a 10 μl sample has to be loaded on a micro-column at a typical flow rate of 150 nl/min, this would take almost 1.5 h. This can be avoided by making use of column switching [29]. The purpose is to segregate the loading of the sample and the actual separation to reduce the loading time. This is obtained by first loading the sample on a short preconcentration-column that has a larger diameter than the capillary column (not in line with the capillary column) at a higher flow rate. Once the sample is loaded, the guard column is put in line with the capillary column to run the gradient at a low flow [29] (Fig. 1).

The advantages are clear; instead of analyzing the entire sample at once, the peptides are now separated and elute gradually, while each peptide is concentrated in a chromatographic peak, resulting in higher sensitivity. In addition, when coupled to a tandem mass spectrometer, such as a Q-TOF, peptides (or other compounds) can be analyzed and selected for fragmentation in the collision cell of the MS instrument as they elute from the column. For this purpose the software of most MS manufacturers has been adapted to allow automated charged state recognition of the ions (apart from the smallest neuropeptides (<6 AA) and depending on their amino acid composition most peptides will have more than one charge) and automated fragmentation.

Such a peptidomic analysis theoretically allows to identify all peptides present in a sample in a single chromatographic run [26,28,35]. However, in complex mixtures of peptides, the higher incidence of co-elution is the limiting factor for the number of peptides that can be identified. Since sensitivity in electrospray is concentration-dependent rather than flow rate-dependent, the ability to vary the flow in nanoscale liquid chromatography (nano-LC) separations is advantageous when analyzing more complex or more diluted mixtures of compounds. This technique, known as ‘peak parking’ has already shown its performance in the LC-MS analysis of enzymatic protein digests, resulting from a proteomic experiment [36]. By reducing the flow rate on the fly (during the separation) elution of peaks can be slowed down, thereby increasing the time for analysis of the mass spectra. This way, more peptides can be fragmented or a single low abundant peptide can be fragmented for a longer time.

Another method to increase the resolving power of a peptide separation is by making use of on-line coupled two-dimensional separations. In this approach, the second chromatography step enables full separation of all of the effluent material from the first step using multiple HPLC pumps connected through detailed valve con-

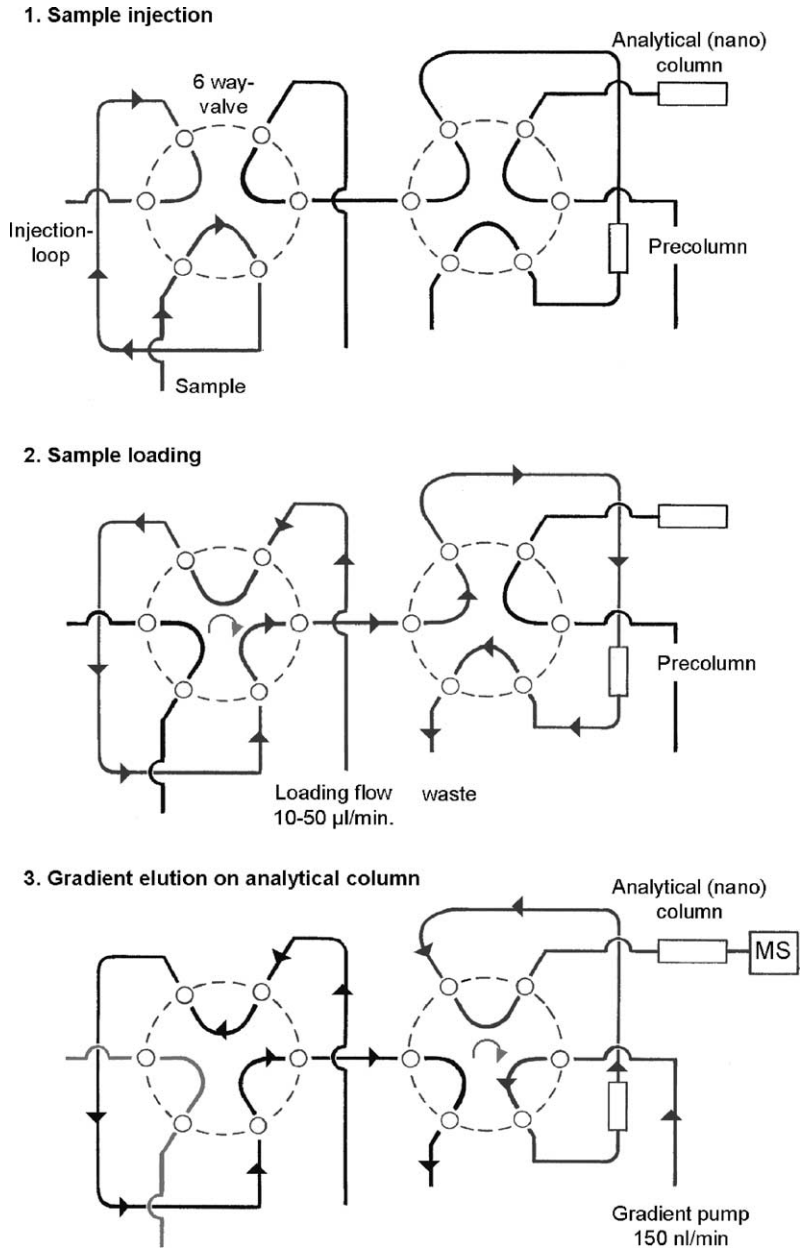

Fig. 1. Schematic diagram of a column switching setup for sample loading on a capillary nanoliter flow HPLC. In a first step (1) the sample is injected in the sample loop. (2) The sample is loaded on a small reversed phase cartridge at a high flow rate (typically 20 μl/min) and salts are rinsed of. (3) The cartridge is placed on-line with the capillary column and peptides are separated using a gradient with increasing hydrophobicity.

nection systems. Reverse-phase LC is often used as the second separation for its high speed and resolving power and its de-salting capability [37]. RPLC as the second dimension has been combined with several other chromatographic techniques in the first dimension such as size exclusion-reversed phase [29,38], reversed phase-reversed phase chromatography [29,39] and ion exchange-reversed phase [29,40]. These techniques are used increasingly frequent for protein identifications in complex protein digest samples so that time-consuming 2D-PAGE and in-gel-digestion prior to identification are avoided [29,41]. The ion exchange-reversed phase approach seems the most promising one. A high-throughput method for protein identification of very complex (whole tissue) mixtures was reported in a process called MudPIT [42]. MudPIT is short for multi-dimensional protein identification technology and uses a biphasic column (a column containing two different types of packing). This column uses a strong cation raisin (SCX) and a reversed phase C18 raisin combined in a single fused silica column. The column used in this setup had an internal diameter of only 100 μm, the flow rate was decreased and the eluate was sprayed directly from the column into the ESI mass spectrometer, increasing the sensitivity of the analysis dramatically. Using multiple steps

of chromatography. MudPIT allows effectively to perform a two-dimensional separation on a single column, thereby eliminating the use of valves and the problem of dead volume associated with it. We predict this technology will soon be introduced in peptide applications as well.

In protein research, the combination of capillary electrophoresis (CE) with mass spectrometry is becoming increasingly popular. A recent study illustrates that CE-MS/MS can be used to analyze peptide and neurotransmitter content of small brain regions [43]. It allows separation of compounds ranging from small molecules to large proteins in the same analysis. The advantages of CE are low sample consumption and short analysis time. However, the low volume required by CE means that the samples need to be highly concentrated. This can be achieved by preconcentrating the sample on hydrophobic media. Tong et al. [44], reported an LC-CE-MS/MS method to identify proteins from a complex mixture. In this setup the protein sample, a 75-protein complex from yeast ribosome was first digested. The peptides were preconcentrated on a solid-phase microextraction (SPE) C18 cartridge and de-salted. The SPE served as a first semi-separation dimension using an organic-phase stepped gradient. CE was used as the second dimension. While this technique is not fully automated yet, other authors have reported integrated LC-CE systems [45–47].

2.3. Peptide quantification

The relevance of proteomics as an analytical tool lies in the possibility to monitor changes in protein expression in different physiological conditions. In other words being able to determine the relative quantities of each protein in different conditions. The same holds true when looking at peptide profiles. When performing HPLC separations of peptides, such as in LC-MS, the simplest way to compare the relative abundances of peptides is through the use of a UV detector placed between the LC system and the mass spectometer. However, when dealing with complex samples this is not an option since one chromatographic peak may contain several compounds.

In general, the ion intensities of each compound observed in a mass spectrum do not correlate to the relative amount of these compounds. Nevertheless, for compounds with a similar mass and with similar functional groups, the relative ion intensities may correspond to their relative contents. ‘Mass labeling’ is a common strategy to quantify peptides by mass spectrometry. In short, a known concentration of a ‘mass labeled’ analogue of the peptide of interest is added to the sample and serves as internal reference. The most appealing way to mass label a compound is by incorporating stable isotopes. In this way, the two compounds are chemically identical and hence, have an equal ionization efficiency but are of a slightly different mass.

The use of ‘mass labeling’ to obtain quantitative data has recently been introduced in the field of proteomics, with great success. With isotope-coded affinity tags (ICAT) the side chains of cysteinyl residues in a reduced protein sample of one particular physiological condition are derivatised with an isotopically light form of the ICAT reagent. The sample derived from a second physiological state is labeled with the isotopically heavy form of the ICAT reagent. After enzymatic digestion, the labeled peptides from the two combined samples are purified by affinity chromatography. In a subsequent LC-MS/MS analysis the peptides are identified. Next, the relative abundancy of these peptides in the two conditions can be compared by comparing their relative signal intensities [48]. The ICAT technology can so far only be used for cysteine containing peptides or proteins.

To circumvent this more general labeling methods (instead of the amino acid specific, such as ICAT) have been developed. Goodlett et al. [49] described a method in which peptides are methylated using either d0 (no deuterium in the molecule) or d3 (three deuterium atoms per molecule) methanol. The methyl esterification converts carboxylic acids, such as those on the side chains of aspartic and glutamic acid and the carboxyl-terminus of the peptide. Although not every peptide will be labeled using this method (e.g. peptides with no aspartic or glutamic acid and with a C-terminal amidation) it would be more useful in peptidomics than ICAT.

A similar approach has been used to quantify peptides in tissue extracts [50,51]. In these studies a peptide labeled with a stable isotope was introduced. In this way, one obtains a chemical identical analogue (with similar ionization) with a slightly higher mass. The concentration can be calculated by comparing the ratio of intensity of the labeled analogue and the naturally occurring peptide.

2.4. Post-translational modifications

Post-translational modifications are covalent processing events that change the properties of a protein, or peptide, by cleavage or by addition of a modifying group to one or more amino acids [52]. In general, a PTM of a protein can modify or regulate its function, state of activity, turnover and interactions with other proteins. In proteomics, several approaches were recently developed to study PTMs, the main focus of which is on phosphorylation analysis [53,54].

Phosphorylation of a neuropeptide is rare. Instead, other PTMs that serve to protect the peptide against proteolytic degradation are very common. At their N-terminus a large number of neuropeptides carry a pyroglutamic acid, formed by the cyclization of an N-terminal glutamine (−17 Da) [55]. Some peptides, such as melanotropines and β-endorphins, are acetylated at their N-terminus (+42 Da). This modification increases their stability and modulates their activity [56].

Another important protective PTM in peptides is C-terminal amide-formation. In about half of all known bioactive peptides the hydroxyl group of the C-terminal carboxyl group is replaced by an amide, resulting in a decrease in molecular mass of the peptide by 1 Da [55].

Other peptide modifications, such as tyrosine sulfation, methylation or glycosilation are important for interaction with the receptor and can therefore play a role in regulation of the activity of the peptide [57,58].

Determining these modifications by MS can be very challenging, especially in the case of de novo sequencing. Modification analysis is usually done by comparison of experimental data to a known amino acid sequence, which means that the peptide sequence has to be determined before the PTM can be identified. If the modification is labile (such as sulfation and glycosilation), it will be lost prior to fragmentation of the peptide itself. In this case, the peptide can be sequenced and identified and the mass increment between the obtained sequence and the measured parent mass will be that of the PTM. However, in this case the exact location of the PTM cannot be determined [59] (Fig. 2A).

In the case of a stable modification, such as acetylation, methylation, oxidation of methionine, N-terminal amidation or formation of a pyroglutamic acid, the fragmentation spectrum will be similar to that of the unmodified peptide, with the exception of the modified amino acid, which will have a mass increment or decrement corresponding to the PTM (Fig. 2B). In this case, the fragmentation pattern will allow the exact localization of the modification, although it will make “de novo” sequencing of the peptide more challenging. A short list of the most common (neuro)peptide modifications is given in Table 1.

Luckily, peptides can be identified in a database by their fragmentation spectrum using pattern searching with algorithms such as Sequest and Mascot that have been adapted to allow for the identification of modified peptides [60,61].

3. Peptidomics: data

To date, many neuropeptides have been purified from vertebrate and invertebrate sources. To determine its sequence in the pre-genomic era (before the realization of the genome projects), a peptide had to be fully purified from tissue extracts [30]. Only then one could study the expression in time and space of a single peptide through mRNA hybridisation, immunolocalisation or HPLC-analysis. This was a time-consuming task since only one peptide at a time could be studied in this way. However, many peptides belong to the same family and are structurally related. Often, cross-immunoreactivity impedes the identification of a single peptide within a tissue. The developments of mass spectrometry in the last decade allow very sensitive and unequivocal profiling of peptides within an organism, tissue or cell.

The need for confirmation of immunocytochemical results is clearly demonstrated in a number of manuscripts where the neurohaemal organs of insects were analysed. Insects have an open circulatory system and besides the corpora cardiaca, which are the functional equivalents of the pituitary in vertebrates, they have a number of additional neurohaemal release sites such as the perisympatic organs distributed in their abdomen. Based on immunocytochemical analysis it

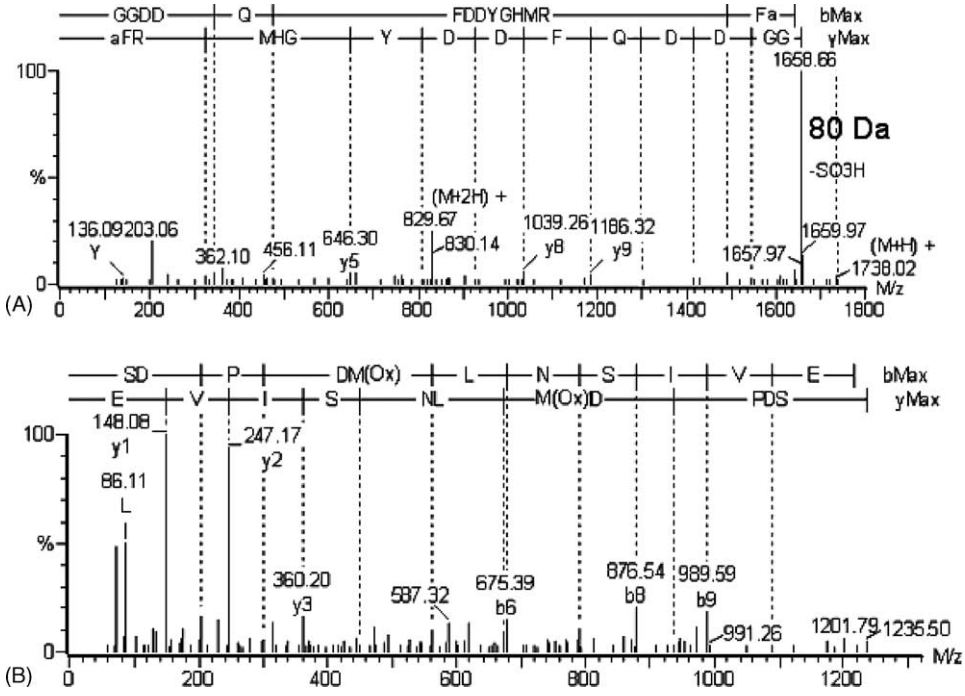

Fig. 2. (A) CID spectrum of *Drosophila* sulfakinin II (GGDDQFDDYGHMRFamide) b-type and y-type are indicated. The loss of 80 Da from the parent mass is indicative for the presence of a (labile) sulphate group, originating from the tyrosine residue. (B) CID spectrum of *Drosophila* sNPF-associated peptide (Drm-sNPF 5; SDPDMLNSIVE), which has an oxidation on its methionine residue as a (stable) PTM. b-type and y-type fragment ions are indicated. The localization of the PTM can be deduced from the fragmentation spectrum.

Table 1
List of post-translational modifications (PTMs), common in (neuro)peptides

PTM	Mass difference (Da)	Stability in MS/MS[a]	Function
Acetylation	+42	3	Blocked N-terminus, peptide stability, regulation of bio-activity
Amidation	−1	3	Peptide stability, blocked C-terminus
Disulfide bridge	−2	2	Intra and intermolecular crosslink
Glycosylation		1	Regulation of bio-activity, receptor–ligand interaction
N-linked	>800		
O-linked	203, >800		
Methylation	+14	3	?
Oxidation	+16	3	Methionine side chain
Pyroglutamic acid	−17	3	Peptide stability, blocked N-terminus
Sulfation	+80	1	Tyrosine side chain, receptor–ligand interaction

Source: http://www.abrf.org/index.cfm/dm.home and Mann and Jensen [52].
[a] Stability: 1: labile; 2: more stable; 3: stable.

has long time been assumed that these would contain neuropeptides identical to those found in the corpora cardiaca to compensate for their poorly developed circulatory system. It was only until sensitive mass spectrometric techniques became available that these small organs could be studied in detail. Through peptide profiling it became apparent that in both cockroaches and locusts these organs contain a number of previously unknown neuropeptides [15,62–64]. Similar studies have shown that peptide profiles can be obtained from a single insect organ making it possible to analyse the peptide profile of all endocrine organs from a single insect [9,10,35,65].

An interesting and early approach for peptide and protein profiling in tissues was shown by Caprioli et al. [66]. They produced MALDI ion images of tissue sections by coating the sections on a MALDI target with a thin layer of matrix and analysing the sections by mass spectrometry. Peptide tissue profiling was demonstrated with the mapping of insulin in an islet on a section of rat pancreas as well as with the mapping of hormones in a small area of rat pituitary.

In terms of sensitivity MALDI-TOF mass spectrometry is unsurpassed. It has shown to allow analysis of the peptide content of individual neurons [13,67] or even individual organelles, in this case secretory vesicles from a neuron originating from a mollusc [68]. In these studies the high sensitivity was achieved by fixing the cell or organelle directly to the MALDI target and covering it with matrix solution. However, single cell analysis was also achieved making use of Micro LC-MALDI-TOF MS and post-source decay [69]. Nowadays, laser capture microdissection (LCM) permits to isolate individual cells or cell populations from frozen sample tissues. Cells obtained with this technique can be directly analysed by MALDI-MS [70].

4. Differential peptidomics: data

Over the last few decades an ever-growing number of peptides have been isolated and characterized from different animal species. In quite a number of cases, not much information about these peptides is available apart from their structure and place of synthesis [71]. A number of research groups have understood the need for differential analysis of peptide expression in different physiological conditions to understand their function and interaction.

One of the earlier peptidomics papers describes the differential display of peptides induced in the haemolymph of individual fruit flies after immune challenge [12]. The authors described 24 immune induced peptides in the molecular mass range of 1.5–11 kDa, four of which were sequenced by Edman-degradation after their purification. A similar approach was used in our group to identify peptides involved in phase transition in locusts. Here, extracts from corpora cardiaca and haemolymph were compared in solitary and gregarious locusts [72]. Differential peptide display can even be used to detect post-translational alternative splicing of primary mRNA transcripts. In the central nervous system of the snail *Lymnea*, a common, multi-exon gene encodes FMRFamide-like peptides. Alternate mRNA splicing of the FMRFamide gene leads to the expression of two different mRNAs. MALDI-TOF MS analysis of individual neurons expressing the FMRFamide gene confirmed the mutually exclusive expression of the peptides encoded on the two transcripts in different neurons [31].

Another early paper by Jiménez et al. made use of an internal peptide standard to quantitatively analyse the difference in peptide expression in the neurointermediate lobe (NIL) of individual rats subjected to salt loading [73].

For a truly comparable and systematic analysis of a peptidome, however, an easily interpretable visualization of all peptides in a sample is needed. Schulz-Knappe et al. developed such a technology over the last few years. In short, the peptidome of a sample, in this case blood ultrafiltrate or cerebrospinal fluid is separated by reversed phase chromatography and all fractions are analysed by MALDI-TOF MS. From these analyses a virtual “2D” image is created in which one dimension represents the mass spectra and the other dimension the HPLC fractions. Signal intensities of the spectra are indicated by colour or intensity of the spot

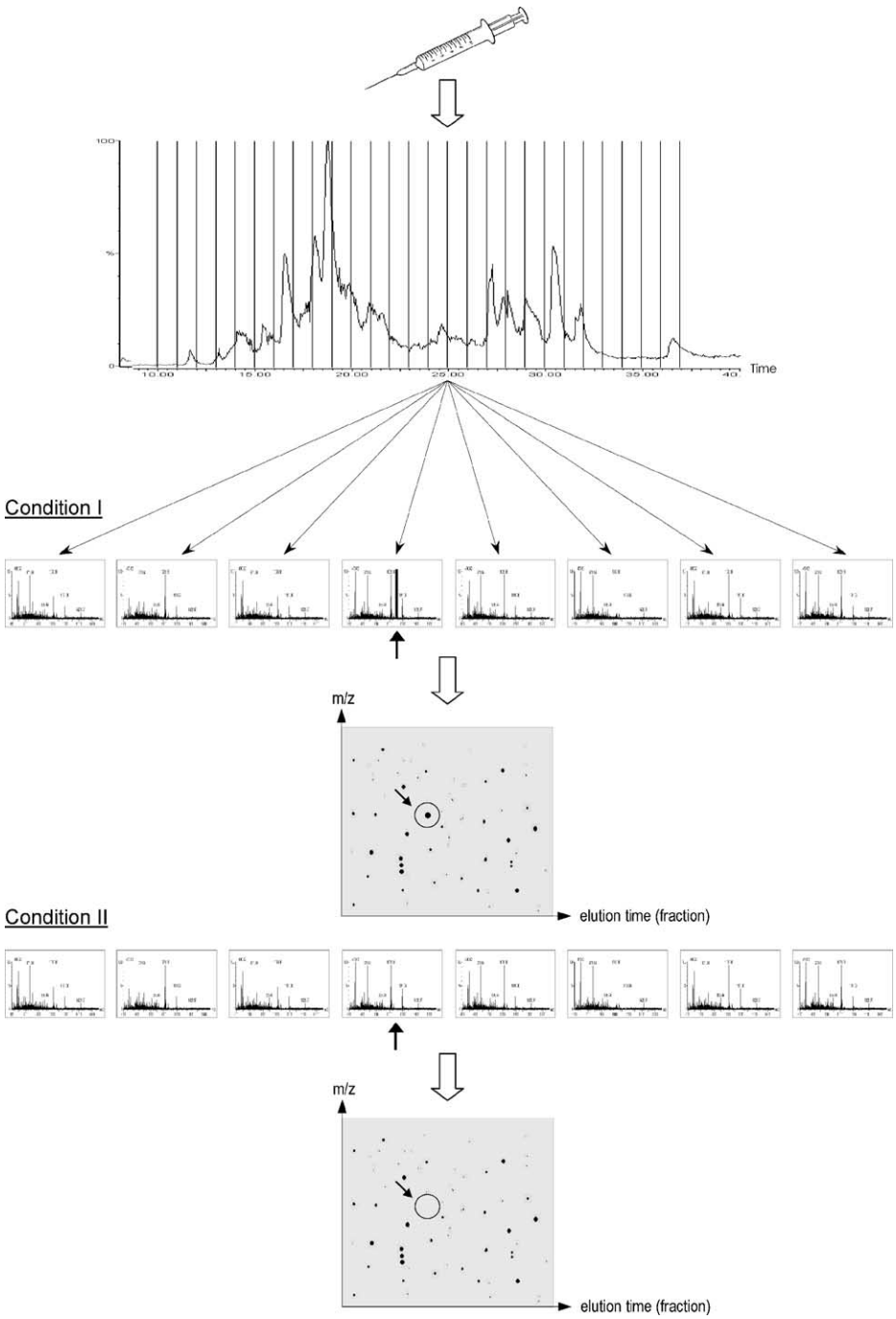

Fig. 3. Flow scheme of the process of generating a virtual 2D plot of a peptidome. The peptide extract is separated on a HPLC (capillary or not) and fractions are collected. Every single fraction is then analysed by MALDI-TOF MS. The separate spectra are combined to a virtual 2D-picture in which one dimension displays the elution-time and hence, hydrophobic nature, and the other one the mass-to-charge (m/z) ratio. In this way, two samples can be quantitatively and qualitatively compared (figure modified from Schulz-Knappe et al. [27]).

(Fig. 3). By comparing 2D plots from different physiological conditions, this system allows an easy detection of differences in peptide content [27,74]. Such profiling has been applied in the analysis of peptides involved in human diseases. A number of manuscripts describe the identification of biomarkers for diagnosis or classification of cancers using differential peptidomics [75,76].

5. Peptidomics in the post-genome era

Initially, mass spectrometry was used as a technique complementary to Edman-degradation-based sequencing of purified peptides. It was used to confirm whether the experimentally obtained peptide mass corresponds to the calculated mass derived from the amino acid sequence obtained

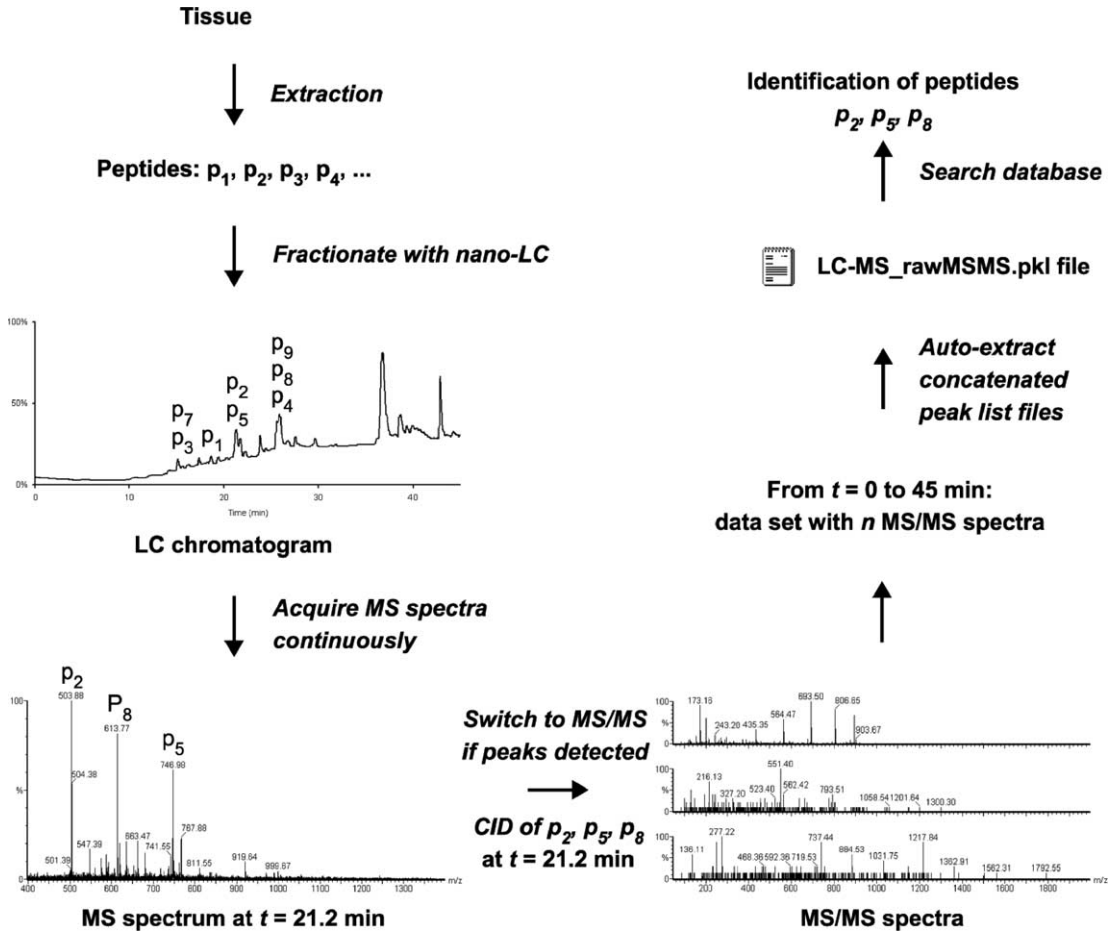

Fig. 4. Flowchart of the experimental setup used in the peptidomic analysis of parts of the rat brain (28) and the fruit fly central nervous system (30). Brain tissue was homogenized and the extract was separated on a nano-LC system coupled directly to a Q-TOF mass spectrometer. At each moment in the LC run, the mass spectrometer software uses the MS spectrum to decide on peak selection for fragmentation. Typically, ions with 2–4 positive charges were automatically selected for MS/MS analysis. The obtained fragmentation spectra are converted to peak lists which are then used for query of a database.

by Edman-degradation. If this was not the case, a mass difference between the theoretical and experimental mass could indicate the presence of a post-translational modifications. Post-translational modifications such as a C-terminal amidation, a tyrosine sulfation or glycosylations are not detected by Edman-degradation chemistry. Mass spectrometry often provides an essential tool to identify these modifications.

Part of the recent success of mass spectrometry in peptide and protein research is its ability to obtain, apart from the relative mass, structural information on molecules by fragmentation analysis. Fragment ions can be obtained due to decay of the ions as a result of the ionisation process, common in MALDI (post-source decay) and in fast atom bombardment (FAB)-MS. These techniques were used successfully for “de novo” sequencing of purified insect peptides that were modified post-translationally or for which only minute amounts were purified, ruling out Edman-degradation [77,78].

In tandem mass spectrometry or MS/MS, an ion is selected from a complex mixture and subjected to fragmentation. This has made it possible to “de novo” sequence peptides from tissue extracts without elaborate chromatographic separation [10,62,79]. However, in all these cases problems can arise with the “de novo sequencing” of as yet unidentified peptides present in the mixture. In the study of Clynen et al. on brain parts of the locust [9], many ion peaks remained unidentified, because sequence information obtained by mass spectrometry was not complete. Fragmentation spectra can be difficult to interpret or lead to only a partial sequence tag. In addition, some amino acids have the same (Leu/Ile) or very similar (Gln/Lys) masses making it difficult to differenciate between them.

The finalization of the genome projects has opened new opportunities for rapid identification and functional analysis of (new) (neuro)peptides. This was clearly demonstrated by two recent manuscripts using a similar approach (Fig. 4).

Skold et al. described a peptidomic approach that combines nanoscale liquid chromatography on-line with Q-TOF mass spectrometry to analyze the peptide content of different regions of the rat brain (motor cortex, striatum and thalamus) [28]. In this study they focused on the 300–5000 Da mass region (typical for peptides) and detected approximately 1500 peptide masses in a robust, specific and reproducible method. Ten percent of these ions were automatically selected for fragmentation by the mass spectrometer software during the

60 min nano-LC run. These fragmentation spectra were analyzed and peptide sequences were deduced after running specific mass spectrometric software programs. The peptides were identified by scanning the rat database (at the moment the rat genome project is near its finalisation) with the peptide sequence tags using the basic local alignment search tool (BLAST). A large number of peptides could be identified, including both known neuropeptides such as hemophorins as well as protein fragments. However, "typical" neuropeptides such as neurotensin, substance P, enkephalins, etc. were not detected in these experiments.

This is in contrast to the results we recently obtained in *Drosophila*. Using the same methodology (on-line nano-LC combined with Q-TOF mass spectrometry), we analysed the peptidome of 50 larval *Drosophila* brains. During the nano-LC run, a few 100 ions were selected for fragmentation. The obtained fragmentation spectra were used as a query to mine the fruit fly genome [5] to reveal the corresponding amino acid sequence and to identify the peptide precursor gene. Hardly any protein fragments were present in the sample. Instead, we sequenced and identified 28 neuropeptides (Table 2). Twenty-one of these peptides were never purified and sequenced from *Drosophila* before. The majority was predicted from the genome sequence but for some cases we found that the processing from the precursor was different from what had been predicted. In addition, eight entire novel peptides (encoded by five different genes) were identified. These peptides could not be predicted because sequence similarity-based searches failed to mine these peptides, as orthologs from other animal species were not isolated before. A BLAST search of the Swissprot database with these peptides did not show any homology with known peptides or peptide precursors.

Although a similar strategy was employed for both the rat and the fruit fly, different results were obtained. In the rat, almost all identified peptides are protein fragments and hardly any typical neuropeptides were found. In the fruit fly, the situation was exactly the opposite. This analysis yielded almost exclusively neuropeptides. Although a number of degradation products from proteins such as hemoglobin and cytochrome *c* are bioactive molecules [80,81], it is unlikely that all the identified protein fragments act as signaling molecules. Instead, as the author suggested, these might originate from post-mortem protease activity. Protease activity might also be responsible for the degradation of the classical neuropeptides. In order to avoid enzymatic activity, we used a different peptide extraction. Instead of homogenizing the tissue in water with 0.25% acetic acid [28], we used an extraction mixture containing 90% methanol, 9% water and 1% acetic acid, which will denature and inactivate any proteases immediately. A similar extraction method was used in the past for classical neuropeptide purifications [30] as well as in peptidomics [9,14,15,82].

Nevertheless, despite the sensitivity of the mass spectrometric methods, a large number of putative peptide ion peaks that were observed in both studies remained unidentified. In

Table 2
Peptides that were identified in the larval *D. melanogaster* central nervous system by nano-LC tandem MS

Peptide name (BDGP annotation)	Peptides characterised in larval *Drosophila* brain by LC-MS/MS (monoisotopic mass (Da))
FaRP family	
FMRFamide (CG2346)	
Drm-FMRFa-1	*DPKQDFMRFamide* (1111.51)
Drm-FMRFa-2	*TPAEDFMRFamide* (914.41)
Drm-FMRFa-3	*SDNFMRFamide* (924.43)
Drm-FMRFa-4	PDNFMRFamide (1181.56)
Dromyosupressin (CG6440)	
Drm-MS	*TDVDHVFLRFamide* (1246.64)
Drosulfakinin (CG18090)	
Drm-SK-2	GGDDQFDDY(SO3)GHMRFamide (1737.61)
Drm-sNPF (CG13968)	
Drm-sNPF 2^{1-10}	WFGDVNQKPI (1202.61)
Drm-sNPF-AP-1	SDPDMLNSIVE (1218.45)
Drm-sNPF-AP-1	SDPDM(ox)LNSIVE (1234.53)
Adipokinetic hormone family	
AKH (CG1171)	
Drm-AKH	*pQLTFSPDWamide* (974.45)
Corazonin (CG3302)	
Drm-COR	pQTFQYSRGWTNamide (1368.62)
Allatostatic peptides	
Allatostatin (CG13633)	
Drm-AST-2^{1-11}	AYMYTNGGPGM (1160.46)
Drm-AST-3	SRPYSFGLamide (924.48)
Drm-AST-4	TTRPQPFNFGLamide (1275.67)
MIP-like peptide (CG6456)	
Drm-MIP-2	AWKSMNVAWamide (1090.54)
Drm-MIP-5	DQWQKLHGGWamide (1252.61)
Diuretic peptides	
CAP2b peptide (CG15520)	
Drm-CAP-1	GANMGLYAFPRVamide (1293.67)
Drm-CAP-2	ASGLVAFPRVamide (1014.6)
DroLeucokinin (CG13480)	
DLK	*NSVVLGKKQRFHSWGamide* (1740.95)
Immune induced peptides	
Immune induced protein 2 (IM2)	
DIM2	*GNVVINGDCKYCVNGamide* (1689.77)
Immune induced protein 4 (CG15231)	
DIM4	*GTVLIQTDNTQYIRTamide* (1720.91)
Newly characterised peptides	
Neuropeptide-like precursor 1 (CG 3441)	
MTYamide peptide	YIGSLARAGGLMTYamide (1470.76)
IPNamide peptide	NVGTLARDFQLPIPNamide (1652.89)
NAP peptide	SVAALAAQGLLNAP (1422.81)
Hugin (CG6371)	
Drm-MT2	SVPFKPRLamide (841.58)
Neuropeptide-like precursor 2 (CG11051)	
NEF peptide	TKAQGDFNEF (1155.52)

Table 2 (*Continued*)

Peptide name (BDGP annotation)	Peptides characterised in larval *Drosophila* brain by LC-MS/MS (monoisotopic mass (Da))
Neuropeptide-like precursor 3 (CG13061)	
SHA peptide	VVSVVPGAISHA (1134.65)
VVIamide peptide	SVHGLGPVVIamide (975.58)
Neuropeptide-like precursor 4 (CG15361)	
YSY peptide	pQYYYGASPYAYSGGYYDSPYSY (2567.01)

Peptides that were identified earlier are in italic face. The other peptides were identified for the first time in the peptidomic analysis (30).

our study this was largely due to the low quality or signal intensity of some of the fragmentation spectra, rendering identification through a database search impossible. However, in some cases, even fairly good fragmentation spectra yielded a partial but inconclusive amino acid sequence (3–5 AA), too short to be used in an advanced BLAST search. The limited length of the sequence made also it impossible to identify these peptides through algorithms, such as Mascot, Sequest and the like, actually written to identify peptides obtained from an enzymatic protein digestion. The low quality of some of the fragmentation spectra is partially due to the limited time frame in which the spectra can be acquired (which, in an LC-MS setup, is dependent on the resolution of the separation). When using an off-line approach such as LC-MALDI-TOF-TOF this can be improved, since each collected HPLC fraction can be subjected to MS analysis as long as necessary to obtain good quality spectra.

Other causes for the absence of expected peptides in the peptidomic experiments are low concentration, timing of sample collection, ionization efficiency, tissue extraction efficiency, etc.

However, both studies prove that peptidomics is a powerful and highly sensitive approach to study peptides, as the technique allows giving a snapshot of all the peptides that really interest the animal, rather than trying to find peptides by staring at genomes. In addition, it serves as a tool to identify peptide precursors that are missed by similarity-based searches. As already five novel peptide precursor genes were found to be expressed in the brain of the fruit fly alone, we believe that far more than the 31 predicted (neuro)peptide precursors are encoded by the *Drosophila* genome [83–86]. This is in accordance with the recent results obtained by bioinformatic analysis of the *C. elegans* genome [87]. This study was based on pattern-based searching and predicted the presence of 92 *C. elegans* peptide genes, corresponding to at least 200 peptides.

6. General conclusion

The finalization of several genome projects boosted research covering the analysis of the function of all the proteins encoded by all these genes. This is clearly demonstrated by the explosion of large and small proteome initiatives. Due to the limitations of the technology, most of these initiatives completely neglect the peptide portion of the proteome although most of the drugs in use today are small molecules, a number of which mimick the activity of peptides [88].

In analogy with the proteomics technology, where all proteins expressed in a cell or tissue are analyzed, the peptidomic approach aims at the simultaneous visualization and identification of the whole peptidome of a cell or tissue, i.e. all expressed peptides with their post-translational modifications. With nanoscale liquid chromatography combined with mass spectrometry one can analyze the peptidome of a tissue at the amino acid sequence level. Often, only partial amino acid sequences of peptides are obtained using tandem MS. This can pose a problem to get the entire de novo sequence of a peptide and often one cannot get the entire sequence. Nevertheless, certain bioinformatic programs, can be used to identify the peptides from a protein database or the corresponding gene from the genomic database, when available, with only this partial sequence or and the parent mass or with the uninterpreted fragmentation spectrum itself [89].

This technology provides us with a fast and efficient tool to analyze the peptidome of any tissue, not only from *Drosophila* and rat but also any metazoan of which the complete genome has been sequenced (*C. elegans*, *A. gambiae*, mouse, man, ...). It also solves one of the problems inherent to the use of mass spectrometry in de novo peptide sequencing. Since leucine and isoleucine residues have the same mass and lysine and glutamine residues only differ by 0.04 Da, they can only be distinguished from each other by W-ions. However, with the availability of the genome sequence the correct residue can be deduced. As such, the peptidomic tool described in this review will allow analysing the expression of the entire peptidome in different physiological and developmental conditions. In sum, historically, most endocrine factors have been defined based on phenotypic changes, monitored by diverse bioassays. Although the traditional biochemical and physiological characterization of endocrine factors will not be replaced completely, we expect that the peptidomic technology represents an opportunity to discover new peptide hormones and signalling molecules of the endocrine system. This will certainly be true for animals of which genome projects are completed or in progress.

Acknowledgements

This project was sponsored by the Flemish Science Foundation (FWO) (G.0356.98, G.0187.00 and G.0175.02). G. Baggerman is a postdoc of the FWO-Flanders. P. Verleyen, J. Hybrechts and E. Clynen benefit from a scholarship from the FWO.

References

[1] N. Williams, Science 272 (1996) 481.
[2] V. Wood, R. Gwilliam, M.A. Rajandream, M. Lyne, R. Lyne, A. Stewart, J. Sgouros, N. Peat, J. Hayles, S. Baker, D. Basham, S.

Bowman, K. Brooks, D. Brown, S. Brown, T. Chillingworth, C. Churcher, M. Collins, R. Connor, A. Cronin, P. Davis, T. Feltwell, A. Fraser, S. Gentles, A. Goble, N. Hamlin, D. Harris, J. Hidalgo, G. Hodgson, S. Holroyd, T. Hornsby, S. Howarth, E.J. Huckle, S. Hunt, K. Jagels, K. James, L. Jones, M. Jones, S. Leather, S. McDonald, J. McLean, P. Mooney, S. Moule, K. Mungall, L. Murphy, D. Niblett, C. Odell, K. Oliver, S. O'Neil, D. Pearson, M.A. Quail, E. Rabbinowitsch, K. Rutherford, S. Rutter, D. Saunders, K. Seeger, S. Sharp, J. Skelton, M. Simmonds, R. Squares, S. Squares, K. Stevens, K. Taylor, R.G. Taylor, A. Tivey, S. Walsh, T. Warren, S. Whitehead, J. Woodward, G. Volckaert, R. Aert, J. Robben, B. Grymonprez, I. Weltjens, E. Vanstreels, M. Rieger, M. Schafer, S. Muller-Auer, C. Gabel, M. Fuchs, C. Fritzc, E. Holzer, D. Moestl, H. Hilbert, K. Borzym, I. Langer, A. Beck, H. Lehrach, R. Reinhardt, T.M. Pohl, P. Eger, W. Zimmermann, H. Wedler, R. Wambutt, B. Purnelle, A. Goffeau, E. Cadieu, S. Dreano, S. Gloux, V. Lelaure, S. Mottier, F. Galibert, S.J. Aves, Z. Xiang, C. Hunt, K. Moore, S.M. Hurst, M. Lucas, M. Rochet, C. Gaillardin, V.A. Tallada, A. Garzon, G. Thode, R.R. Daga, L. Cruzado, J. Jimenez, M. Sanchez, F. del Rey, J. Benito, A. Dominguez, J.L. Revuelta, S. Moreno, J. Armstrong, S.L. Forsburg, L. Cerrutti, T. Lowe, W.R. McCombie, I. Paulsen, J. Potashkin, G.V. Shpakovski, D. Ussery, B.G. Barrell, P. Nurse, Nature 415 (2002) 871.

[3] The Arabidopsis genome initiative, Nature 408 (2002) 796.

[4] The *C. elegans* sequencing consortium, Science 282 (2002) 2012.

[5] M.D. Adams, S.E. Celniker, R.A. Holt, C.A. Evans, J.D. Gocayne, P.G. Amanatides, S.E. Scherer, P.W. Li, R.A. Hoskins, R.F. Galle, R.A. George, S.E. Lewis, S. Richards, M. Ashburner, S.N. Henderson, G.G. Sutton, J.R. Wortman, M.D. Yandell, Q. Zhang, L.X. Chen, R.C. Brandon, Y.H. Rogers, R.G. Blazej, M. Champe, B.D. Pfeiffer, K.H. Wan, C. Doyle, E.G. Baxter, G. Helt, C.R. Nelson, G.L. Gabor, J.F. Abril, A. Agbayani, H.J. An, C. Andrews-Pfannkoch, D. Baldwin, R.M. Ballew, A. Basu, J. Baxendale, L. Bayraktaroglu, E.M. Beasley, K.Y. Beeson, P.V. Benos, B.P. Berman, D. Bhandari, S. Bolshakov, D. Borkova, M.R. Botchan, J. Bouck, P. Brokstein, P. Brottier, K.C. Burtis, D.A. Busam, H. Butler, E. Cadieu, A. Center, I. Chandra, J.M. Cherry, S. Cawley, C. Dahlke, L.B. Davenport, P. Davies, B. de Pablos, A. Delcher, Z. Deng, A.D. Mays, I. Dew, S.M. Dietz, K. Dodson, L.E. Doup, M. Downes, S. Dugan-Rocha, B.C. Dunkov, P. Dunn, K.J. Durbin, C.C. Evangelista, C. Ferraz, S. Ferriera, W. Fleischmann, C. Fosler, A.E. Gabrielian, N.S. Garg, W.M. Gelbart, K. Glasser, A. Glodek, F. Gong, J.H. Gorrell, Z. Gu, P. Guan, M. Harris, N.L. Harris, D. Harvey, T.J. Heiman, J.R. Hernandez, J. Houck, D. Hostin, K.A. Houston, T.J. Howland, M.H. Wei, C. Ibegwam, M. Jalali, F. Kalush, G.H. Karpen, Z. Ke, J.A. Kennison, K.A. Ketchum, B.E. Kimmel, C.D. Kodira, C. Kraft, S. Kravitz, D. Kulp, Z. Lai, P. Lasko, Y. Lei, A.A. Levitsky, J. Li, Z. Li, Y. Liang, X. Lin, X. Liu, B. Mattei, T.C. McIntosh, M.P. McLeod, D. McPherson, G. Merkulov, N.V. Milshina, C. Mobarry, J. Morris, A. Moshrefi, S.M. Mount, M. Moy, B. Murphy, L. Murphy, D.M. Muzny, D.L. Nelson, D.R. Nelson, K.A. Nelson, K. Nixon, D.R. Nusskern, J.M. Pacleb, M. Palazzolo, G.S. Pittman, S. Pan, J. Pollard, V. Puri, M.G. Reese, K. Reinert, K. Remington, R.D. Saunders, F. Scheeler, H. Shen, B.C. Shue, I. Siden-Kiamos, M. Simpson, M.P. Skupski, T. Smith, E. Spier, A.C. Spradling, M. Stapleton, R. Strong, E. Sun, R. Svirskas, C. Tector, R. Turner, E. Venter, A.H. Wang, X. Wang, Z.Y. Wang, D.A. Wassarman, G.M. Weinstock, J. Weissenbach, S.M. Williams, T. Woodage, K.C. Worley, D. Wu, S. Yang, Q.A. Yao, J. Ye, R.F. Yeh, J.S. Zaveri, M. Zhan, G. Zhang, Q. Zhao, L. Zheng, X.H. Zheng, F.N. Zhong, W. Zhong, X. Zhou, S. Zhu, X. Zhu, H.O. Smith, R.A. Gibbs, E.W. Myers, G.M. Rubin, J.C. Venter, Science 287 (2000) 2185.

[6] R.A. Holt, G.M. Subramanian, A. Halpern, G.G. Sutton, R. Charlab, D.R. Nusskern, P. Wincker, A.G. Clark, J.M. Ribeiro, R. Wides, S.L. Salzberg, B. Loftus, M. Yandell, W.H. Majoros, D.B. Rusch, Z. Lai, C.L. Kraft, J.F. Abril, V. Anthouard, P. Arensburger, P.W. Atkinson, H. Baden, B. de V, D. Baldwin, V. Benes, J. Biedler, C. Blass, R. Bolanos, D. Boscus, M. Barnstead, S. Cai, A. Center, K. Chatuverdi, G.K. Christophides, M.A. Chrystal, M. Clamp, A. Cravchik, V. Curwen, A. Dana, A. Delcher, I. Dew, C.A. Evans, M. Flanigan, A. Grundschober-Freimoser, L. Friedli, Z. Gu, P. Guan, R. Guigo, M.E. Hillenmeyer, S.L. Hladun, J.R. Hogan, Y.S. Hong, J. Hoover, O. Jaillon, Z. Ke, C. Kodira, E. Kokoza, A. Koutsos, I. Letunic, A. Levitsky, Y. Liang, J.J. Lin, N.F. Lobo, J.R. Lopez, J.A. Malek, T.C. McIntosh, S. Meister, J. Miller, C. Mobarry, E. Mongin, S.D. Murphy, D.A. O'Brochta, C. Pfannkoch, R. Qi, M.A. Regier, K. Remington, H. Shao, M.V. Sharakhova, C.D. Sitter, J. Shetty, T.J. Smith, R. Strong, J. Sun, D. Thomasova, L.Q. Ton, P. Topalis, Z. Tu, M.F. Unger, B. Walenz, A. Wang, J. Wang, M. Wang, X. Wang, K.J. Woodford, J.R. Wortman, M. Wu, A. Yao, E.M. Zdobnov, H. Zhang, Q. Zhao, S. Zhao, S.C. Zhu, I. Zhimulev, M. Coluzzi, T.A. della, C.W. Roth, C. Louis, F. Kalush, R.J. Mural, E.W. Myers, M.D. Adams, H.O. Smith, S. Broder, M.J. Gardner, C.M. Fraser, E. Birney, P. Bork, P.T. Brey, J.C. Venter, J. Weissenbach, F.C. Kafatos, F.H. Collins, S.L. Hoffman, Science 298 (2002) 129.

[7] The human genome, Nature 409 (2001).

[8] R.H. Waterston, K. Lindblad-Toh, E. Birney, J. Rogers, J.F. Abril, P. Agarwal, R. Agarwala, R. Ainscough, M. Alexandersson, P. An, S.E. Antonarakis, J. Attwood, R. Baertsch, J. Bailey, K. Barlow, S. Beck, E. Berry, B. Birren, T. Bloom, P. Bork, M. Botcherby, N. Bray, M.R. Brent, D.G. Brown, S.D. Brown, C. Bult, J. Burton, J. Butler, R.D. Campbell, P. Carninci, S. Cawley, F. Chiaromonte, A.T. Chinwalla, D.M. Church, M. Clamp, C. Clee, F.S. Collins, L.L. Cook, R.R. Copley, A. Coulson, O. Couronne, J. Cuff, V. Curwen, T. Cutts, M. Daly, R. David, J. Davies, K.D. Delehaunty, J. Deri, E.T. Dermitzakis, C. Dewey, N.J. Dickens, M. Diekhans, S. Dodge, I. Dubchak, D.M. Dunn, S.R. Eddy, L. Elnitski, R.D. Emes, P. Eswara, E. Eyras, A. Felsenfeld, G.A. Fewell, P. Flicek, K. Foley, W.N. Frankel, L.A. Fulton, R.S. Fulton, T.S. Furey, D. Gage, R.A. Gibbs, G. Glusman, S. Gnerre, N. Goldman, L. Goodstadt, D. Grafham, T.A. Graves, E.D. Green, S. Gregory, R. Guigo, M. Guyer, R.C. Hardison, D. Haussler, Y. Hayashizaki, L.W. Hillier, A. Hinrichs, W. Hlavina, T. Holzer, F. Hsu, A. Hua, T. Hubbard, A. Hunt, I. Jackson, D.B. Jaffe, L.S. Johnson, M. Jones, T.A. Jones, A. Joy, M. Kamal, E.K. Karlsson, D. Karolchik, A. Kasprzyk, J. Kawai, E. Keibler, C. Kells, W.J. Kent, A. Kirby, D.L. Kolbe, I. Korf, R.S. Kucherlapati, E.J. Kulbokas, D. Kulp, T. Landers, J.P. Leger, S. Leonard, I. Letunic, R. LeVine, J. Li, M. Li, C. Lloyd, S. Lucas, B. Ma, D.R. Maglott, E.R. Mardis, L. Matthews, E. Mauceli, J.H. Mayer, M. McCarthy, W.R. McCombie, S. McLaren, K. McLay, J.D. McPherson, J. Meldrim, B. Meredith, J.P. Mesirov, W. Miller, T.L. Miner, E. Mongin, K.T. Montgomery, M. Morgan, R. Mott, J.C. Mullikin, D.M. Muzny, W.E. Nash, J.O. Nelson, M.N. Nhan, R. Nicol, Z. Ning, C. Nusbaum, M.J. O'Connor, Y. Okazaki, K. Oliver, E. Overton-Larty, L. Pachter, G. Parra, K.H. Pepin, J. Peterson, P. Pevzner, R. Plumb, C.S. Pohl, A. Poliakov, T.C. Ponce, C.P. Ponting, S. Potter, M. Quail, A. Reymond, B.A. Roe, K.M. Roskin, E.M. Rubin, A.G. Rust, R. Santos, V. Sapojnikov, B. Schultz, J. Schultz, M.S. Schwartz, S. Schwartz, C. Scott, S. Seaman, S. Searle, T. Sharpe, A. Sheridan, R. Shownkeen, S. Sims, J.B. Singer, G. Slater, A. Smit, D.R. Smith, B. Spencer, A. Stabenau, N. Stange-Thomann, C. Sugnet, M. Suyama, G. Tesler, J. Thompson, D. Torrents, E. Trevaskis, J. Tromp, C. Ucla, A. Ureta-Vidal, J.P. Vinson, A.C. Von Niederhausern, C.M. Wade, M. Wall, R.J. Weber, R.B. Weiss, M.C. Wendl, A.P. West, K. Wetterstrand, R. Wheeler, S. Whelan, J. Wierzbowski, D. Willey, S. Williams, R.K. Wilson, E. Winter, K.C. Worley, D. Wyman, S. Yang, S.P. Yang, E.M. Zdobnov, M.C. Zody, E.S. Lander, Nature 420 (2002) 520–562.

[9] E. Clynen, G. Baggerman, D. Veelaert, A. Cerstiaens, D.J. van der Horst, L. Harthoorn, R. Derua, E. Waelkens, A. De Loof, L. Schoofs, Eur. J. Biochem. 268 (2001) 1929.

[10] P. Verhaert, S. Uttenweiler-Joseph, M. de Vries, A. Loboda, W. Ens, K.G. Standing, Proteomics 1 (2001) 118.

[11] A.P. Jonsson, Cell Mol. Life Sci. 58 (2001) 868.

[12] S. Uttenweiler-Joseph, M. Moniatte, M. Lagueux, A. Van Dorsselaer, J.A. Hoffmann, P. Bulet, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 11342.
[13] J.V. Sweedler, L. Li, P. Floyd, W. Gilly, J. Exp. Biol. 203 (2000) 3565.
[14] G.V. Kollisch, M.W. Lorenz, R. Kellner, P.D. Verhaert, K.H. Hoffmann, Eur. J. Biochem. 267 (2000) 5502.
[15] R. Predel, R. Kellner, G. Baggerman, T. Steinmetzer, L. Schoofs, Eur. J. Biochem. 267 (2000) 3869.
[16] M. Karas, F. Hillenkamp, Anal. Chem. 60 (1988) 1299.
[17] J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong, C.M. Whitehouse, Science 246 (1989) 64.
[18] H.R. Morris, M. Panico, M. Barber, R.S. Bordoli, R.D. Sedgwick, A. Tyler, Biochem. Biophys. Res. Commun. 101 (1981) 623.
[19] D.F. Hunt, A.M. Buko, J.M. Ballard, Biomed. Mass Spectrom. 8 (1981) 397.
[20] H.R. Morris, T. Paxton, M. Panico, R. McDowell, A. Dell, J. Protein Chem. 16 (1997) 469.
[21] L. Schoofs, G.M. Holman, P. Proost, J. Van Damme, T.K. Hayes, A. De Loof, Regul. Pept. 37 (1992) 49.
[22] A.A. Shevchenko, A. Loboda, W. Ens, K.G. Standing, Anal. Chem. 9 (2000) 2132.
[23] M. Vestal, P. Jurhasz, W. Hines, S. Martin, in: S. Carr, A. Burlingame, M. Baldwin (Eds.), Mass Spectrometry in Biology and Medicine, Humana Press, Totowa, NJ, 2000, pp. 1–16.
[24] A.L. Yergey, J.R. Coorssen, P.S. Backlund Jr., P.S. Blank, G.A. Humphrey, J. Zimmerberg, J.M. Campbell, M.L. Vestal, J. Am. Soc. Mass Spectrom. 13 (2002) 784.
[25] M. Mann, R.C. Hendrickson, A. Pandey, Ann. Rev. Biochem. 70 (2001) 437.
[26] G. Baggerman, A. Cerstiaens, A. De Loof, L. Schoofs, J. Biol. Chem. 277 (2002) 40368.
[27] P. Schulz-Knappe, H.D. Zucht, G. Heine, M. Jurgens, R. Hess, M. Schrader, Comb. Chem. High Throughput Screen 4 (2001) 207.
[28] K. Skold, M. Svensson, A. Kaplan, L. Bjorkesten, J. Astrom, P.E. Andren, Proteomics 2 (2002) 447.
[29] A.J. Oosterkamp, E. Gelpi, J. Abian, J. Mass Spectrom. 33 (1998) 976.
[30] L. Schoofs, G.M. Holman, L. Paemen, D. Veelaert, M. Amelinckx, A. De Loof, Peptides 14 (1993) 409.
[31] B.M. Worster, M.S. Yeoman, P.R. Benjamin, Eur. J. Neurosci. 10 (1998) 3498.
[32] M.L. Alexandrov, L.N. Gall, N.V. Krasnov, V.I. Nikolaev, V.A. Pavlenko, V.A. Shkurov, Bioorg. Khim. 10 (1984) 710.
[33] M. Wilm, M. Mann, Anal. Chem. 68 (1996) 1.
[34] Y. Shen, R. Zhao, S.J. Berger, G.A. Anderson, N. Rodriguez, R.D. Smith, Anal. Chem. 74 (2002) 4235.
[35] G. Baggerman, E. Clynen, A. Cerstiaens, J. Huybrechts, A. De Loof, L. Schoofs, in: D. Konopinska, G. Goldsworthy, J. Nachman, J. Nawrot, I. Orchard, G. Rosinsky (Eds.), Insects: chemical, physiological and environmental aspects, University of Wroclaw, Wroclaw, Poland, 2001.
[36] M.A. Moseley, K. Blackburn, W. Burkhart, H. Vissers, R. Bordoli, In: Presented at the 48th American Society for Mass Spectrometry Conference, Long Beach, CA, June 2000.
[37] H. Liu, D. Lin, J.R. Yates III, Biotechniques 32 (2002) 898, 900, 902.
[38] G.J. Opiteck, J.W. Jorgenson, R.J. Anderegg, Anal. Chem. 63 (1997) 2283.
[39] D.B. Wall, D.M. Lubman, S.J. Flynn, Anal. Chem. 71 (1999) 3894.
[40] M.T. Davis, E.T. Beierle, M.D. Bures, M.D. McGinley, J. Mort, J.H. Robinson, C.S. Spahr, W. Yu, J. Chromatogr. B 752 (2001) 281.
[41] C.L. Nilsson, P. Davidsson, Mass Spectrom. Rev. 19 (2000) 390.
[42] M.P. Washburn, D. Wolters, J.R. Yates III, Nat. Biotechnol. 19 (2001) 242.
[43] E.M. Javerfalk-Hoyes, U. Bondesson, D. Westerlund, P.E. Andren, Electrophoresis 20 (1999) 1527.
[44] W. Tong, A. Link, J.K. Eng, J.R. Yates III, Anal. Chem. 71 (1999) 2270.
[45] M.M. Bushey, J.W. Jorgenson, Anal. Chem. 62 (1990) 978.
[46] J.P. Larmann Jr., A.V. Lemmo, A.W. Moore Jr., J.W. Jorgenson, Electrophoresis 14 (1993) 439.
[47] A.W. Moore Jr., J.W. Jorgenson, Anal. Chem. 67 (1995) 3448.
[48] S.P. Gygi, B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, R. Aebersold, Nat. Biotechnol. 17 (1999) 994.
[49] D.R. Goodlett, A. Keller, J.D. Watts, R. Newitt, E.C. Yi, S. Purvine, J.K. Eng, P. von Haller, R. Aebersold, E. Kolker, Rapid Commun. Mass Spectrom. 15 (2001) 1214.
[50] D.M. Desiderio, U. Wirth, J.L. Lovelace, G. Fridland, E.S. Umstot, T.M. Nguyen, P.W. Schiller, H.S. Szeto, J.F. Clapp, J. Mass Spectrom. 35 (2000) 725.
[51] C. Fierens, L.M. Thienpont, D. Stockl, E. Willekens, A.P. De Leenheer, J. Chromatogr. A 896 (2000) 275.
[52] M. Mann, O.N. Jensen, Nat. Biotechnol. 21 (2003) 255.
[53] M. Mann, S.E. Ong, M. Gronborg, H. Steen, O.N. Jensen, A. Pandey, Trends Biotechnol. 20 (2002) 261.
[54] C.S. Loughrey, M.J. Huddleston, W. Shou, R.J. Deshaies, R.S. Annan, S.A. Carr, Mol. Cell Proteomics 1 (2002) 186.
[55] L. Schoofs, D. Veelaert, J. Vanden Broeck, A. De Loof, Peptides 18 (1997) 145.
[56] T.L. O'Donohye, G.E. Handelmann, R.L. Miller, D.M. Jacobowitz, Science 215 (1982) 1125.
[57] R.J. Nachman, G.M. Holman, W.F. Haddon, Peptides 9 (Suppl. 1) (1988) 137.
[58] J.A. Dias, Hum. Reprod. 16 (2001) 825.
[59] P. Torfs, G. Baggerman, T. Meeusen, J. Nieto, R.J. Nachman, J. Calderon, A. De Loof, L. Schoofs, Biochem. Biophys. Res. Commun. 299 (2002) 312.
[60] M.J. MacCoss, C.C. Wu, J.R. Yates III, Anal. Chem. 74 (2002) 5593.
[61] D.M. Creasy, J.S. Cottrell, Proteomics 2 (2002) 1426.
[62] R. Predel, G. Gäde, Peptides 23 (2002) 621.
[63] R. Predel, M. Eckert, G.M. Holman, Ann. N. Y. Acad. Sci. 897 (1999) 282.
[64] E. Clynen, J. Huybrechts, A. De Loof, L. Schoofs, Biochem. Biophys. Res. Commun. 300 (2003) 422.
[65] R. Predel, J. Comp. Neurol. 436 (2001) 363.
[66] R.M. Caprioli, T.B. Farmer, J. Gile, Anal. Chem. 69 (1997) 4751.
[67] L. Li, R.W. Garden, E.V. Romanova, J.V. Sweedler, Anal. Chem. 71 (1999) 5451.
[68] S.S. Rubakhin, R.W. Garden, R.R. Fuller, J.V. Sweedler, Nat. Biotechnol. 18 (2000) 172.
[69] S. Hsieh, K. Dreisewerd, R.C. van der Schors, C.R. Jimenez, J. Stahl-Zeng, F. Hillenkamp, J.W. Jorgenson, W.P. Geraerts, K.W. Li, Anal. Chem. 70 (1998) 1847.
[70] B.J. Xu, R.M. Caprioli, M.E. Sanders, R.A. Jensen, J. Am. Soc. Mass Spectrom. 13 (2002) 1292.
[71] L. Schoofs, E. Clynen, A. Cerstiaens, G. Baggerman, Z. Wei, T. Vercammen, R. Nachman, A. De Loof, S. Tanaka, Peptides 22 (2001) 219.
[72] E. Clynen, D. Stubbe, A. De Loof, L. Schoofs, Comp. Biochem. Physiol. B; Biochem. Mol. Biol. 132 (2002) 107.
[73] C.R. Jimenez, K.W. Li, K. Dreisewerd, H.D. Mansvelder, A.B. Brussaard, B.B. Reinhold, R.C. van der Schors, M. Karas, F. Hillenkamp, J.P. Burbach, C.E. Costello, W.P. Geraerts, Proc. Natl. Acad. Sci. U.S.A. 94 (1997) 9481.
[74] P. Schulz-Knappe, M. Raida, M. Meyer, E.A. Quellhorst, W.G. Forssmann, Eur. J. Med. Res. 1 (1996) 223.
[75] K. Sasaki, K. Sato, Y. Akiyama, K. Yanagihara, M. Oka, K. Yamaguchi, Cancer Res. 62 (2002) 4894.
[76] D. Basso, A. Valerio, R. Seraglia, S. Mazza, M.G. Piva, E. Greco, P. Fogar, N. Gallo, S. Pedrazzoli, A. Tiengo, M. Plebani, Pancreas 24 (2002) 8.

[77] G. Gäde, K.L. Rinehart, Biochem. Biophys. Res. Commun. 149 (1987) 908.
[78] G. Gäde, G.J. Goldsworthy, M.H. Schaffer, J.C. Cook, K.L. Rinehart Jr., Biochem. Biophys. Res. Commun. 134 (1986) 723.
[79] G. Baggerman, J. Huybrechts, E. Clynen, K. Hens, L. Harthoorn, H.D. van der, C. Poulos, A. De Loof, L. Schoofs, Peptides (2002) 635.
[80] V. Brantl, C. Gramsch, F. Lottspeich, A. Henschen, K.H. Jaeger, A. Herz, Eur. J. Pharmacol. 111 (1985) 293.
[81] V. Brantl, C. Gramsch, F. Lottspeich, R. Mertz, K.H. Jaeger, A. Herz, Eur. J. Pharmacol. 125 (1986) 309.
[82] J.M. Conlon, J.B. Kim, A. Johansson, S. Kikuyama, J. Endocrinol. 175 (2002) 769.
[83] J. Vanden Broeck, Peptides 22 (2001) 241.
[84] R.S. Hewes, P.H. Taghert, Genome Res. 11 (2001) 1126.
[85] A. De Loof, G. Baggerman, M. Breuer, I. Claeys, A. Cerstiaens, E. Clynen, T. Janssen, L. Schoofs, J. Vanden Broeck, Arch. Insect Biochem. Physiol. 47 (2001) 129.
[86] W. Brogiolo, H. Stocker, T. Ikeya, F. Rintelen, R. Fernandez, E. Hafen, Curr. Biol. 11 (2001) 213.
[87] A.N. Nathoo, R.A. Moeller, B.A. Westlund, A.C. Hart, Proc. Natl. Acad. Sci. U.S.A. 98 (2001) 14000.
[88] T. Kieber-Emmons, R. Murali, M.I. Greene, Curr. Opin. Biotechnol. 8 (1997) 435.
[89] J.S. Choudhary, W.P. Blackstock, D.M. Creasy, J.S. Cotrell, Trends Biotechechnol. 19 (2001) S17.