Menschaert et al. 2010

DATASET · FEBRUARY 2013

READS
66

7 AUTHORS, INCLUDING:

Walter Luyten
University of Leuven
97 PUBLICATIONS **5,657** CITATIONS
SEE PROFILE

Gerben Menschaert
Ghent University
31 PUBLICATIONS **465** CITATIONS
SEE PROFILE

Liliane Schoofs
University of Leuven
378 PUBLICATIONS **11,321** CITATIONS
SEE PROFILE

Geert Baggerman
Flemish Institute for Technological Research
127 PUBLICATIONS **4,830** CITATIONS
SEE PROFILE

reviews Journal of proteome research

Peptidomics Coming of Age: A Review of Contributions from a Bioinformatics Angle

Gerben Menschaert,*,† Tom T. M. Vandekerckhove,† Geert Baggerman,‡ Liliane Schoofs,§ Walter Luyten,|| and Wim Van Criekinge†

Laboratory for Bioinformatics and Computational Genomics, Department of Molecular Biotechnology, Faculty of Bioscience Engineering, Ghent University, B-9000 Ghent, Belgium, ProMeta, Interfaculty Center for Proteomics and Metabolomics, Catholic University of Leuven, B-3000 Leuven, Belgium, Research Group of Functional Genomics and Proteomics, K.U. Leuven, B-3000 Leuven, Belgium, and Biomedical Sciences Group, Department of Woman & Child, K.U. Leuven, B-3000 Leuven, Belgium

Received October 16, 2009

The term peptidomics for a new promising "omics" field was not introduced until the beginning of 2000. The approach has been proven successful in several domains such as neuroendocrine research and biomarker or drug discovery. This review reports on bioinformatics tools and methodologies within the peptidomics field and the application thereof. Obviously, a plethora of proteomics data analysis tools lends themselves to direct use in peptidomics because the latter is a subfield of the former, at least to a certain extent. Nevertheless, peptidomics-specific tool extensions, inventions, and validation procedures have emerged, and certain tools are more suitable for this subfield than others due to small but important differences in peptidomics sample analysis. This paper focuses on these topics. Furthermore, it gives a comprehensive overview of available online tools tailored to the peptidomics field. To conclude, an ideal pipeline for bioactive peptide identification is presented.

Keywords: peptidomics • bioinformatics • mass spectrometry identification • MS/MS fragmentation spectra

1. Introduction

Although the neuropeptide concept was established at the end of the 1960s by David de Wied, it took three more decades to publish the first trend-setting peptidomics research papers on the identification of the whole of the biologically active peptide content of a cell, tissue or organism. The term peptidomics for a new promising "omics" field was not introduced until the beginning of 2000.[1−6] This delay was mainly caused by the fact that proteomics/peptidomics analysis was only made possible after several advances in mass spectrometry (MS) and related techniques on the one hand and genome projects that delivered comprehensive data pools for proteomics/peptidomics studies on the other hand.

Meanwhile, the peptidomics approach has been proven successful in several domains such as neuroendocrine research[6−9] and biomarker[10−16] or drug discovery[17−19] for medical areas such as oncology, neurodegeneration, osteoporosis, blood pressure regulation, fat metabolism, fertility and immunology.

Reviews covering diverse topics within the field are available. Recent manuscripts have been published on the integrated peptidomics approach and methods and technologies developed in the last years for the peptidome analysis upon selective extraction and multidimensional separation.[20,21] Others focus on biomarker or drug discovery within specific health-related areas[13,17,22] or also on quantitative peptidomics.[8] Hummon et al.[9] recently presented an overall picture of invertebrate neuropeptide discovery.

This review reports on bioinformatics tools and methodologies within the peptidomics field and the application thereof. Obviously, a plethora of proteomics data analysis tools lends themselves to direct use in peptidomics because the latter is a subfield of the former, at least to a certain extent. Nevertheless, peptidomics-specific tool extensions, inventions and validation procedures have emerged, and certain tools are more suitable for this subfield than others due to small but important differences in peptidomics sample analysis. This paper focuses on these topics.

2. Peptide Identification Based on MS/MS Fragmentation Spectra

Several matters complicate the identification process in peptidomics studies. First, endogenous peptides are most often cleaved from larger precursors by various releasing or processing enzymes,[23,24] some of which still have an unknown specificity today. In contrast to proteomics analysis, where proteolytic digestion with trypsin is routinely applied to obtain

* To whom correspondence should be addressed: Gerben Menschaert, Laboratory for Bioinformatics and Computational Genomics, Department of Molecular Biotechnology, Faculty of Bioscience Engineering, Ghent University, Coupure Links 653, B-9000 Ghent, Belgium. E-mail: Gerben.Menschaert@ugent.be.
† Ghent University.
‡ Catholic University of Leuven.
§ Research Group of Functional Genomics and Proteomics, K.U. Leuven.
|| Department of Woman & Child, K.U. Leuven.

10.1021/pr900929m

Published on Web 04/08/2010

Table 1. Auxiliary Validation Steps to Further Strengthen Database Search Identifications

	related publications	validation is applied
Biological characteristics		
Signal peptide in precursor	132	6,26,43,133−135
Peptide cleavage patterns	24,99,100,136	43,82,134,137
Blast homology searching		26,27,43,113,138−141
Database search inherent or MS/MS specific characteristics		
Difference experimental/theoretical ion mass		28,113,134,137,139,141
Inspection of first and second hit score (s1 − s2 > 1)		10,27
Inspection of error distribution scores (Da and ppm)		26
Inspection of % experimental spectra matching with predicted		31,43,113,141,142
Count of consecutive, accurate *b*/*y* ion matches		43,135,142
Presence of *b* and *a* ions		135
Characteristics based on other tools		
Peptide sequence tag overlap		86,129,143,144
Quality scoring to retain good quality spectra	44,49	
Member of identified spectral cluster	51−54	26

shorter, more workable protein fragments with highly regular patterns at both ends, no such regular cleavage enzyme is specified in peptidomics analysis, resulting in very large search spaces. Second, recent studies show that the extent of amino acid modification is underestimated in proteome samples;[25] from bioactive peptides it has long been known that they are sometimes even more heavily post-translationally modified before becoming active.[20,26] The former two phenomena hinder the identification process considerably.

Furthermore, the ever-growing search space hugely contrasts with the sometimes very small in vivo available peptide quantity, making it exceedingly difficult to identify bioactive peptides without strong experimental support.

In addition, protein-level precursors are no longer intact after biological processing; only the cleaved peptides are present in the sample. In other words, real protein-level inference is unneeded or even unwise, and a peptide identification cannot be supported by other high-confidence identifications from the same precursor.

A last difficulty stems from the less informative and inadequately understood fragmentation patterns of endogenous peptides compared with those of tryptic peptides.[27]

Although this review deals with bioinformatics tools for improved peptide identification upon MS fragmentation, the sample preparation step at the very beginning of the whole procedure is too crucial to omit. Dedicated sample preparation should warrant detection of bioactive peptides over the background from protein degradation. Although this appears to be less of a problem with lower organisms, the sometimes already minute in vivo peptide quantity next to their high modification rate requires adequate peptidome sample extraction protocols indeed. Several techniques such as microwave irradiation,[28,29] snap-freezing in liquid nitrogen optionally combined with rapid heat stabilization,[30] and use of an extraction buffer containing 90% methanol, 9% water, and 1% formic acid[6,31] have been successfully applied to reduce proteolysis to a minimum. A recent paper[32] compares microwave irradiation with snap-freezing, showing that the latter gives rise to lower titers of endogenous peptides and more degradation products. Contaminating protein fragments produced *post mortem* during conventional sample handling often compromise the study of endogenous peptides.

2.1. Spectral Identification by Database Searching. The most commonly applied strategy is the database search engine; hereby the ion spectra (whether parent ions—i.e. before fragmentation—or daughter ions) are compared with theoretical spectra predicted for each protein or peptide contained in the sequence database. The most routinely used engines are Mascot,[33] SeQuest,[34] X!Tandem,[35] OMSSA,[36] and MS-Fit,[37] the latter three being open-source software. More general information on database search engines, their algorithms and scoring schemes was provided by Nesvizhskii et al.[38] The database search approach has several drawbacks with regard to the peptidomics issues mentioned earlier: (1) the fact that the naturally correct cleavage enzymes (releasing enzymes or convertases) cannot be specified for precursor protein digestion weighs heavily on both the search performance and the quality of the search results; (2) the same is true for the abundance of post-translational modifications (PTMs); (3) the technique is always limited by its inherently incomplete set of known or predicted proteins offered by the database to be searched; (4) deficiencies in the scoring scheme used to quantify the degree of similarity between the experimental spectrum and those predicted may decrease or bias the identification rate. The latter two disadvantages are not restricted to peptidomics alone. In section 3 of this review, which deals with validation of peptide identifications, the statistical scoring principles of a few database search engines will serve as a basis for discussing the need and the development of better validation methods.

Several methods have been introduced to improve the database-driven approach. Preprocessing of the sequence database by using different cleavage patterns[39,40] helps to speed up the search. A recent publication describes an online tool, "Database on Demand",[41] allowing to create such search spaces using custom digestion algorithms. Others claim that mimicking the peptidome rather than the proteome is a strategy for specific and sensitive identification of endogenous peptides:[26,27,42,43] rather than using the proteomic search space, a custom database of bioactive peptides or precursors thereof is used.

Very often database-supported identifications with a below-threshold score are nevertheless retained thanks to extra validation methods to lower the false negative and/or false positive rate. Several of these can be found in table 1 with an indication of their application throughout different recent peptidomics studies. Some of the criteria are based on biological characteristics of the peptide precursor, others tackle shortcomings of database search engines or of the mass

spectrometer software itself (which produces the raw MS/MS spectra), and a fourth category is based on validation by complementary tools.

As in proteomics studies, it is worthwhile to routinely include an extra quality score calculation in peptidomics studies. Several quality measures have been proposed over the years. Nesvizhskii et al.[44] report on the Spectrum Quality Score (SQS), which is dynamically calculated, meaning that the statistical classifier is based on the data itself, after prior database search to obtain high-confidence peptide assignments; thus, one becomes independent of training data sets created in different experiments. The classifier is automatically developed for each new data set, ensuring robustness of the method against variation in the MS/MS spectrum properties caused by differences in acquisition methods or instrument variability. This sets the method apart from other approaches,[45−48] and gives it a very important advantage, taking into account the specific nature of peptidomics MS/MS fragmentation spectra. Moreover, the method proved to perform well even for the typically small peptidomics data sets. The SQS is based on both general spectrum features (e.g., number of peaks, mean and standard deviation of peak intensities,.. .), sequence tag features (e.g., length of longest tag, number of peak pairs corresponding to an amino acid mass difference,.. .), and presence of complementary fragment ions (e.g., *b* and *y* ions) or neutral losses (e.g., loss of ammonia, water or carbon monoxide).

If multiple complementary fragmentation spectra (see section 5) are available from a Fourier transform MS instrument, the S-score described by Savitski et al.[49] seems a good alternative. This score is based on built-in discovery of reliable sequence tags (short, partial peptide sequences derived directly from the mass spectrum; see section 2.3).

Another very recently presented classifier[50] is based on 12 different MS/MS spectrum features—more than any other classifier described hitherto—and could be useful when no prior database search can be performed to make a distinction between good (assigned) and bad (unassigned) spectra. Unidentified spectra having a high quality score may thus result in identification of post-translational modifications or novel peptides.

PTMs are extremely important for bioactive peptides; they are often required for activity or to increase stability. Presence of PTMs should therefore be carefully analyzed in peptidomics studies. Spectral clustering[51−54] can facilitate identification of new and unexpected modifications; as a consequence, this small but clearly useful preliminary step ought not to be skipped, even for routine purposes. Aforementioned literature[51−54] describes different algorithms to cluster modified peptides with their unmodified counterparts based on peak pattern similarity, enabling inference of identifications of modified forms. This is of course only possible when both forms are present in the sample to be analyzed. The Bonanza clustering approach[51,52] was successfully applied to a peptidomics study of endocrine tissues (pituitary and islets of Langerhans) resulting in doubled identification rates.[26]

Such spectral clustering tries to define the entire modification profile within a MS/MS data set prior to database searching, which helps the identification process to a very significant extent. Although the peptide's amino acid sequence may be in the database, it may carry additional mass in vivo due to post-translational modifications, causing it to never be found by database search engines as they depend on small experimental errors on peptide masses (mostly less than 0.5 Da) for both performance and reliability reasons. Thus, they will only search (unmodified) peptide fragments with a close approximation of the experimental mass, often in vain in the case of peptidomics data sets. In theory, allowing the mass of one or more post-translational modifications to be facultatively added to the experimental mass, and then searching the protein fragment database with all possible resulting masses, can solve this problem. However, to avoid a combinatorial explosion causing identification speed and reliability of results to plummet in practice, one has to limit the number of post-translational modifications or combinations thereof, or both. As such, a preliminary clustering step can help to perform a more targeted database search with regard to the PTM profile.

Other methods are based on alignment of *de novo* derived sequence tags (see section 2.3) with known peptide sequences: OpenSea[55,56] (mass-based sequence alignment) and MS-Alignment[57] could further improve PTM identification, mostly because they apply dynamic parent mass thresholds to facilitate identification of modified peptides. MS-Alignment supports a "blind search" mode, allowing all possible mass shifts in the alignment process.

2.2. Peptide Identification by Spectral Matching. Certain bioactive peptide databases (SwePep,[58] Erop-Moscow,[59] PeptideDB,[60] Peptidome[61]) can be seen as the heralds of real peptidomics spectral matching engines. These databases are in fact knowledgebases and can be used as a validation tool for sequence database identification. They can be searched using different peptide characteristics: peptide monoisotopic mass with(out) PTMs, length, and amino acid sequence. In a later stage, real spectral matching became possible within the SwePep environment[62] by extending the existing version with tandem mass spectra from a locally curated version of the Global Proteome Machine database (GPMDB). Validation is obtained by pairwise comparison of the fragmentation patterns of two spectra using algorithms for calculating the correlation coefficient between them. More general, proteomics-based peptide MS libraries[63] and their matching methods are also available: SpectraST,[64] X!Hunter,[65] NIST MSPepSearch (http://peptide.nist.gov), and BiblioSpec.[66] On top of that, SpectraST allows the creation of spectral libraries; a focused peptidomics library can be constructed based on identified spectra, improving subsequent peptide identifications.

In peptidomics experiments, as in shotgun proteomics studies, the same peptides are repeatedly picked up by sequence database searching methods, which are often time-consuming and error-prone. Their runtime is especially long when searching with multiple variable PTMs. Over the years, as research in the field grows, the peptidome will reach a more and more comprehensive depth. The more complete peptidome map that this entails opens the possibility of inferring peptide sequences by matching fragmentation patterns against a library of spectra representing this peptidome map. The great advantage over the classical sequence database search is that it substantially outperforms it in speed, error rate and sensitivity characteristics of the results.[38] The disadvantage is that no (modified) peptides will be identified for which representative spectra were not submitted to the spectral library. Since the current peptidome map is far from complete, the spectral matching approach for the time being might be used most effectively as a rapid first pass in the incremental search strategy.

2.3. Peptide Identification by *de novo* Sequencing and Hybrid Approaches. Whenever peptidomic research is performed on species with unannotated genomes (nonmodel systems) or if high-quality tandem mass spectra cannot be identified, *de novo* sequencing comes into play. Popular

algorithms in proteomics are Peaks,[67] PepNovo,[68−71] Sherenga,[72] DirecTag,[73] and MS-Tag.[74] General information on *de novo* algorithms can be found in Nesvizhskii et al.[38] Frank et al.[68] made a comparative survey benchmarking performance and identification hit rate of the aforementioned *de novo* algorithms. Their own PepNovo algorithm scored best; it is basically an enhancement of the Sherenga algorithm using a probabilistic network (a maximum likelihood method) taking into account fragmentation rules governing the origin of daughter peptides in certain types of mass spectrometer. Recently they released a newer version, PepNovo+, outperforming the older version.[70] The major minus of this bottom-up approach is that usually only incomplete sequence tags can be extracted from the tandem mass spectra, mostly dependent on completeness and accuracy of their fragmentation. Moreover, even for partial peptide sequences the "best" *de novo* program hitherto still appears to err in up to 30% of the administered mass spectra. This may be a consequence of a number of factors, for example the probability score of the only true *de novo* sequence (the one present in the biological sample) may not be the highest, or based on the applied likelihood model for fragmentation equally good scores may be given to other theoretically possible but naturally erroneous sequences.

Several tools were designed to filter or search protein databases with the generated peptide sequence tags (PSTs): MS-Blast,[75,76] SPIDER,[77] Inspect.[78] Sequencing errors, mutations and homology searching can sometimes be taken into account. Looking into recent peptidomics projects it is notable that manual *de novo* sequencing is still frequently performed,[79−81] sometimes combined with automated *de novo* tools like Peaks,[81−84] MS-Tag,[85] or PepSeq (Micromass Co., Manchester, U.K.).[80]

Other tools, aiming for novel peptide discovery, scan the complete genome translated in its six reading frames.[86,87] MS-Dictionary[87] is extremely fast, it carries out pattern-mapping of the complete spectral dictionary to an indexed translated genome. IggyPep[86] on the other hand is an order of magnitude slower, but is capable of pinpointing the genomic location of a peptide with little more input than a few short sequence tags. IggyPep successfully passed the test identifying the sea urchin peptidome.

To allow further automation of *de novo* analysis, PepNovo+[70] outputs, next to its standard format, MS-Blast and Inspect readable input. IggyPep[86] allows batch processing of standard PepNovo+ output.

3. Validation of Peptide Identifications

Recent guidelines for protein and peptide identification[88,89] require statistical validation of the results by means of identification probabilities and false discovery rates (FDRs). A common method for FDR calculation among database search engines is a parallel search in a decoy database of scrambled or reversed proteins:[90,91] peptide assignments are then kept from both the target and the decoy database using a common score threshold. The FDR for that threshold is estimated as $2N_d/N$, N_d and N being the number of matches in the decoy and target databases, respectively.

Some search engines include their own statistical scoring mechanism. For example, Mascot uses the probability-based Mowse scoring algorithm,[92] which yields a score based on the probability that the top hit is a random event. Given an absolute probability of the top match being random, and knowing the size of the sequence database being searched, the engine calculates an objective measure of the significance of the hit. Comparably, X!Tandem calculates statistical confidence (expectation values) for all of the individual spectrum-to-sequence assignments.

A brand-new validation methodology for database identifications, MS-Generating Function (MS-GF),[93] could be a very good replacement for the former strategy. Kim et al.[93] use properties obtained from the computation of so-called "generating functions" (a mathematical technique used in combinatorics) to evaluate the error rates of peptide identifications. They conclude that error rates from existing database search tools do not provide accurate estimates of the statistical significance of *individual* peptide identifications whereas those evaluated by MS-GF do. Identification validation in peptidomics (i.e., for individual endogenous peptides) could benefit a lot from this approach. In addition, MS-GF was shown capable of completely replacing the decoy database strategy, saving half of the search time: a warmly welcomed feature, account taken of the vastly expanded peptidomics search space owing to aspecific cleavage and PTM abundance.

Despite all validation techniques put forward and guidelines published, manual validation is often still part of peptidomics efforts. This is especially true for endogenous peptides from organisms without sequenced genomes. Manual validation is mainly based on biological characteristics of the identified peptide: basic cleavage patterns, homology search results, presence of a secretion signal peptide at the beginning of the precursor sequence (see Table 1).

4. Publicly Available Tools/Databases and Listing of Relevant in Silico Peptide Discovery Research

Table 2 gives an overview of publicly available tools often encountered in the peptidomics field; it also lists relevant research based on *in silico* peptide precursor discovery. As mentioned in previous sections, numerous tools can identify peptides: database search engines, spectral matching algorithms, *de novo* sequencing or hybrid approaches. Most of these tools are available online (see Table 2) or can be downloaded and installed locally (e.g.: X!Tandem, X!Hunter, PepNovo, Inspect).

Several neuropeptide information databases are accessible online (see Table 2). SwePep[58] gathers information on many thousands of endogenous peptides, classified in three categories: biologically active, possibly biologically active, and uncharacterized. Likewise, EROP-Moscow[59] and Peptidome[61] offer information on endogenous regulatory oligopeptides but without classification. The PeptideDB[60] database encompasses all naturally occurring signaling peptides of animal origin, holding over 20 000 peptides derived by cleavage from prepropeptide precursor proteins. The collection is subdivided into the following families: cytokines and growth factors, peptide hormones, antimicrobial peptides, toxins and venom peptides, and antifreeze proteins. PepBank[94] is a database of peptides compiled from sequence text-mining and public peptide data sources. Only peptides of 20 amino acids or shorter are stored. SwePep is the only resource to also store MS/MS fragmentation spectra. All online databases allow basic querying by accession number, organism, and peptide name, or also by physicochemical features such as amino acid sequence and length, molecular mass, and isoelectric point (p*I*). Some[59,61,94] also accept related literature queries. PepBank differs from the other resources in that it can be queried with amino acid strings,

Table 2. Publicly Available Tools Often Used in the Peptidomics Field

	Web site	ref
Database search engines		
X!Tandem	http://www.thegpm.org	35
MS-Fit	http://prospector.ucsf.edu	37
OMSSA	http://pubchem.ncbi.nlm.nih.gov/omssa	36
Mascot	http://www.matrixscience.com	33
Sequest	http://www.thermo.com	34
Peptidome database and spectral matching		
SwePep	http://www.swepep.org	58
Erop-Moscow	http://erop.inbi.ras.ru	59
PeptideDB	http://www.peptides.be	60
Peptidome	http://www.peptidome.jp	61
PepBank	http://pepbank.mgh.harvard.edu	94
SpectraST	http://www.peptideatlas.org/spectrast	64
X!Hunter	http://www.thegpm.org	65
NIST MSPepSearch	http://peptide.nist.gov	
BiblioSpec	http://proteome.gs.washington.edu/software/bibliospec	66
De novo and hybrid tools		
PepNovo	http://proteomics.ucsd.edu	68−70
DirecTag		67
Peaks	http://www.bioinformaticssolutions.com	72
Sherenga		73
MS-Blast	http://genetics.bwh.harvard.edu/msblast	75,76
Spider	http://bif.csd.uwo.ca/spider	77
Inspect	http://proteomics.ucsd.edu	145
IggyPep	http://www.iggypep.org	86
MS-Dictionary	http://proteomics.ucsd.edu	87
Other peptidomics tools and in silicoprediction techniques		
In silico experiments for precursor or neuropeptide prediction		98,101−103
Conserved motif screening for precursor prediction		95−97
NeuroPred, cleavage site prediction	http://neuroproteomics.scs.uiuc.edu/neuropred.html	99,100
Bonanza spectral clustering for peptidomics	http://peppus.ugent.be:7777/pls/apex/f?p = 120:2	26
Database on Demand	http://www.ebi.ac.uk/pride/dod	41
General tools		
Unimod, peptide modification database	http://www.unimod.org	114
Collection of proteomics tools	http://www.proteomecommons.org	115

whereby the results can be further filtered (with the help of a machine-learning algorithm) using a heat map of predefined categories of interest, for example: related to cancer, binding data availability and so on. Furthermore, Blast and Smith-Waterman searches are also possible.

Many efforts have been made to predict peptide precursors by motif-searching,[95−97] hidden Markov model applications,[98] cleavage site prediction[99,100] and evolutionary sequence modeling.[101] Liu et al.[95−97] used the pattern-matching program Pratt to look for conserved patterns among peptides within families. They discovered 155 novel peptide patterns in addition to the 56 established ones in the PROSITE database. Using the newly identified peptide signatures as a search tool, they predicted 95 hypothetical proteins as putative peptide precursors. Mirabeau et al.[98] developed a hidden Markov tool that uses several peptide hormone sequence features to estimate the likelihood that a protein contains a processed and secreted peptide of the bioactive class. Analysis of the top-scoring hypothetical and poorly annotated human proteins identified two novel candidate peptide hormones, which they named spexin and augurin. Their findings were confirmed by expression data. Another peptide precursor predicting endeavor was based on cleavage pattern recognition. Southey et al.[99,100] trained logistic regression models on experimentally verified or published cleavage data from molluscs, mammals and insects, and on amino acid motifs reportedly associated with cleavage. Their tool, NeuroPred, also computes the mass of the predicted peptides, including user-selectable post-translational modifications. The resulting mass aids the discovery and confirmation of new neuropeptides by means of mass spectrometry techniques. Sonmez et al.[101] devised a computational method that models protein sequence patterns simultaneously with evolutionary differences across species in order to identify peptide hormones unknown as yet; they were able to identify a previously unknown putative prohormone that contains up to four potential neuropeptides. Shemesh et al.[102] used machine-learning algorithms (Random Forest) to predict potential peptide ligands. The most promising ones were tested in vitro on a panel of GPCRs by measuring intracellular calcium levels. Muggleton et al.[103] resorted to the Inductive Logic Programming (ILP) Bayesian approach to generate a grammar for recognizing neuropeptide precursors, learnt from positive examples. Their best predictor made the search for novel neuropeptide precursors more than 100 times more efficient than randomly selecting proteins for synthesis and testing them for biological activity.

As mentioned above (see section 1), an online tool "Database on Demand" is available to predigest sequence databases *in silico* with different cleavage patterns in order to accelerate the search. Furthermore, a web interface is freely accessible to perform preliminary Bonanza spectral clustering on peptidomics MS samples.[26]

5. Quantitative Peptidomics, Biomarker Discovery, and Spatial Imaging

Next to bioinformatics solutions for identifying peptides, boosting hit rates or elucidating modifications, software tools have also been developed in closely related fields such as quantitative peptidomics, biomarker discovery, and spatial imaging.

Mass spectrometry is increasingly used for relative or absolute quantification of proteins and peptides. Some previous peptidomics studies have yielded estimates of the relative concentrations of peptides in two or more samples by comparing the relative intensities measured during different LC−MS runs with the peptides present in a complex sample.[104−107] Furthermore, as in proteomics research, stable isotope labeling and label-free quantification can be envisaged. Since the bioactive peptide rather than the protein precursor is the subject of quantification, some adaptations are deemed necessary. The exponentially modified protein abundance index (emPAI[108]) is a simple measure based on the number of observed peptides, normalized to account for expected number of tryptic peptides. In peptidomics research an alternative peptide-centered measure could be obtained based on the actual spectral count of fragmentation spectra mapping to a certain bioactive peptide. A tighter integration between database search engines on the one hand and clustering algorithms[26,52,109] on the other hand is indispensable to perform this task, as is thorough validation. The cluster size could serve as an initial indicator of the endogenous peptide amount, but stable isotope labeling would of course be more accurate.

Most of the efforts toward the development of stable isotopic tags have been directed at proteins rather than peptides. The Fricker group performed extensive research on amine-labeling reagents, ideal for endogenous peptides, as most contain either a free N-terminal amine or at least one Lys residue.[110−113] They compared succinic anhydride with either four hydrogens or deuteriums and [3-(2,5-dioxopyrrolidin-1-yloxycarbonyl)-propyl]-trimethylammonium chloride with either nine hydrogens or deuteriums. These two labels react with amines and impart either a negative charge (succinyl) or a positive charge (4-trimethylammoniumbutyryl: TMAB). They prove to be stable during MS experiments, are rather inexpensive, give reproducible results, and are a modification option in common computer programs implementing the database search method (e.g., Mascot, X!Tandem) for identifying peptides based on MS/MS data. Two additional TMAB forms containing three and six deuteriums have been synthesized and tested later on, performing identically to the former tags.[114] A very recent paper reports on the development and application of a promising set of novel *N*,*N*-dimethyl leucine (DiLeu) 4-plex isobaric tandem mass tagging (TMT) reagents with high quantification efficacy, greatly lowering the cost of neuropeptide analysis.[115]

For differential peptidomics or biomarker studies, several software packages on the market can detect and quantify the abundance of single peptides for either label-free or isotope-labeled systems: e.g. DeCyder MS Differential Analysis Software (GE Healthcare), ProfileAnalyse (Bruker Daltonics), Quant,[116] OpenMS,[117] Mascot,[33].. .). A detailed survey can be consulted in Vaudel et al.[118]

Another promising proteomics technology readily translatable to the peptidome research field is spatial imaging using MALDI-MS Imaging or Secondary Ion Mass Spectrometry (SIMS). General reviews on these topics have been published,[119,120] more specific peptidomics applications are explained in Boonen et al.[20] Spatial imaging with MS boils down to cutting a slice from a biological sample and mounting it on a carrier for a mass spectrometer in such a way that it is subdivided by a grid overlay before the ionizer and mass analyzer do their work. By and large the spatial resolution of the peptide image analysis of extract slices delivered by the two aforementioned approaches is fairly low. A major advantage of this emerging technology is nonetheless that it allows true label-free molecular imaging of flat samples, for example, biological tissue sections. Analysis of the generated 3D data cube (set of serial tissue sections) is possible by means of specially designed bioinformatics tools like BioMap (Novartis) or an unsupervised spatial tissue exploration technique.[121]

6. Need for General-Purpose Peptidomics Data Sets for Comparing Bioinformatics Tools

It would be powerful if several sets of general-purpose peptidomics MS/MS data could be compiled so that comparison can be made for performances when using different strategies and algorithms discussed in this review. Two meritorious first attempts at making proteomics-specific mass spectrometry data sets publicly available came from the SwedCAD project[122] from which thousands of annotated high-accuracy MS/MS spectra of tryptic peptides emanated, and the PRIDE project[123] which strives to standardize proteomics MS/MS data exchange in general. Nonetheless, a lot of further collaborative effort is needed to reconcile all of this with the peptidomics issue.

Each and every tool or methodology discussed here was carefully tested, validated and if necessary compared at the time of publication by the respective makers and authors referred to. While it is true that a universal, common mass spectrometry data set was not used across the many dozens of research groups, they did have their reasons to do so and have recourse to case-specific data sets instead: familiarity with the data and/or the organisms they were collected from, (in)compatibility with certain mass spectrometry instruments or biological sample origins, internal agreements on disclosure before publishing, etc.

For some of the tools and methodologies enumerated here it can be easily predicted that using one or more current universal data sets for comparative purposes may cause readers to make wrong assumptions and conclusions. For example, the NeuroPred[99] bioinformatics tool for *in silico* prediction of neuropeptides would perform less well than claimed by its makers with a universal bioactive peptide data set, and even worse with a universal proteomics one. A comparative review illustrated with thoughtlessly designed general-purpose data sets would then underestimate NeuroPred undeservedly, and conclusions would be valid in part only.

The only plausible solution to this kind of unwanted bias is the compilation of one or more general-purpose data sets for peptidomics that pre-emptively embody antibiasing measures in their composition. However, such a compilation requires careful study and likely even discussion in a public forum, for example during an international conference.

In short, the universal reference data set for bioinformatics and mass spectrometry in peptidomics research is a project *per se*, which falls beyond the scope of this review but for which the latter may serve as an ideal point of departure.

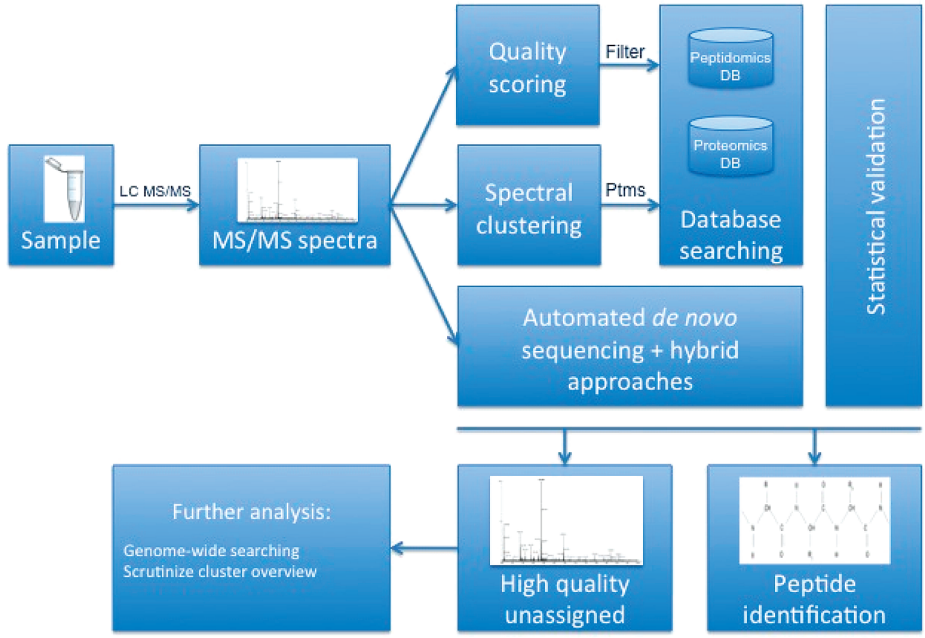

Figure 1. Overview of the fragmentation mass spectra identification analysis in a peptidomics study. Once past the sample preparation and tandem mass spectrometry stages, the bioinformatics section of the pipeline sets in. It contains the following elements: database searching preceded by quality control and spectral clustering, automated *de novo*/hybrid approach, and identification validation. The outcome consists of either peptide identifications or high-quality unassigned fragmentation spectra needing further detailed analysis.

7. Conclusion and Future Perspectives

Certain aspects typical of peptides (unknown cleavage specificity, presence of multiple PTMs, very small *in vivo* amounts) make their identification very difficult, laborious and even tedious. An elaborate peptidomics study ought to comprise several steps to try and boost the identification rate (see Figure 1). Clearly, some of the plethora of available tools proved to yield more satisfactory results and should, as a consequence, be included in a typical peptidomics identification pipeline.

First off, standard database search is best preceded by both a quality control[44,49] and a spectral clustering step.[26,51,52] Quality control values of tandem MS/MS spectra may serve as a first filtering criterion. Fewer spectra will consequently be introduced into the database engine, speeding up the search and reducing the rate of false positive identifications. Table 1 shows that quality scores like for example the SQS[44] are not routinely included as auxiliary validation in peptidomics studies. Nevertheless such scores represent a lot of spectral information, therefore they may well contribute to supplemental validation evidence in much the same way as peptide sequence tag features and peak list characteristics (see Table 1) do in *de novo* analyses. Quality scores could ultimately even incorporate the former features and serve as an alternative, standardized quality measure. The spectral clustering step is able to detect the complete modification profile of the sample.[26] This prior knowledge can be included further downstream in the analysis, for example in the database search stage. This undoubtedly conveys one of the most substantial advantages to a pipeline supposed to identify modified peptides. As for the database search *per se*, it is probably best to compare and/or combine at least two search engines. Most of the engines appear to agree on about 80% of all identifications, but for the remaining 20% different engines tend to assign high identification scores to different spectra.[124] Alternatively, Fälth et al.[27] demonstrated that mimicking a peptidomics search space yielded good results. Since spectral libraries become more comprehensive[62,63] this strategy may serve as a replacement of the former or may at least be used for cross-validation.

Second, an automated *de novo* routine[67,68,70] results in either complete *de novo* sequences or—much more often—partial sequence tags. The obtained sequence information may lead to discovery of new peptides not present in the scanned sequence databases due to genes overlooked by gene prediction algorithms, or due to splice isoforms and peptides from coding small open reading frames.[125,126] Provided that complete peptide sequences can be derived, they may be used to locate the corresponding gene through well-known blast algorithms.[127,128] Conversely, if only partial peptide sequences can be distilled from the experimental MS/MS spectrum, this task becomes rather difficult. With such fragmentary input other tools may then be applied to search or filter existing databases[76−78] or even the complete six-frame translation of the genomic sequence.[86,87]

Unfortunately, peptidomics analyses sometimes still result in a set of high-quality unassigned fragmentation spectra that need further analysis. Attempts may be undertaken at manual *de novo* sequencing—but that can be a time-consuming chore—followed by localizing the genomic position of the gene of interest.[86] The IggyPep web interface was erected for this purpose, among other things. Another and possibly even more profuse source of extra identifications of high-quality unassigned spectra is the spectral clustering output. Scrutinizing the cluster overview may help infer identifications by elucidating mass shifts between members within clusters holding at least one peptide bearing a high-confidence identification label.

Next to the identification process itself, sound statistical validation is equally necessary. The implementation of the MG-Generating Function,[93] account taken of its advantages over

the decoy database search tactics (see section 3), should be seen as a promising alternative to the latter, at least where peptidomics is concerned but likely also beyond.

Figure 1 depicts a thorough identification methodology wherein each component has already proved its usefulness. It should be noted that in neuroendocrine research or biomarker discovery the complete pipeline is likely to be the only sensible option, certainly if one has the ambition to discover new or modified forms of peptides. In other cases such as differential peptidomics or diagnostic biomarker tests, where the panel of differential peptides/biomarkers is already known, a slimmed version of the pipeline can be used encompassing quality scoring, clustering and regular database searching or spectral matching. Such analysis can be routine as it is less time-consuming as compared with the complete pipeline which also includes *de novo* elements.

Incomplete backbone fragmentation and frequent overlap of fragment masses are problems of conventional tandem MS/MS based sequencing, rendering the identification impossible. They may be alleviated by complementary fragmentation techniques.[129−131] Depending on the MS instrumentation, fragmentation based on electron capture dissociation (ECD) and conventional collisionally activated dissociation (CAD) could be combined, or alternatively, electron transfer dissociation (ETD) and collisionally induced dissociation (CID), or ECD and infrared multiphoton dissociation (IRMPD), All these combinations would result in enhanced analysis specificity and lowered false positive rates.

Research on fragmentation rules in general, and more specifically on fragmentation of bioactive peptides, improved sample extraction and preparation protocols, and ever more specific and accurate MS instrumentation are key components that will certainly advance the identification process in the near future, as long as the modern mass spectrometrist stays up to date with the bioinformatics revolution that accompanies the hardware evolution.

Acknowledgment. This work was supported by an SBO grant (IWT-50164) of the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT).

References

(1) Verhaert, P.; Uttenweiler-Joseph, S.; de Vries, M.; Loboda, A.; Ens, W.; Standing, K. G. Matrix-assisted laser desorption/ionization quadrupole time-of-flight mass spectrometry: an elegant tool for peptidomics. *Proteomics* **2001**, *1* (1), 118–31.

(2) Clynen, E.; Baggerman, G.; Veelaert, D.; Cerstiaens, A.; Van der Horst, D.; Harthoorn, L.; Derua, R.; Waelkens, E.; De Loof, A.; Schoofs, L. Peptidomics of the pars intercerebralis-corpus cardiacum complex of the migratory locust, Locusta migratoria. *Eur. J. Biochem.* **2001**, *268* (7), 1929–39.

(3) Schulz-Knappe, P.; Zucht, H. D.; Heine, G.; Jurgens, M.; Hess, R.; Schrader, M. Peptidomics: the comprehensive analysis of peptides in complex biological mixtures. *Comb. Chem. High Throughput Screen.* **2001**, *4* (2), 207–17.

(4) Bergquist, J.; Ekman, R. Future aspects of psychoneuroimmunology - lymphocyte peptides reflecting psychiatric disorders studied by mass spectrometry. *Arch Physiol Biochem* **2001**, *109* (4), 369–71.

(5) Sasaki, K.; Sato, K.; Akiyama, Y.; Yanagihara, K.; Oka, M.; Yamaguchi, K. Peptidomics-based approach reveals the secretion of the 29-residue COOH-terminal fragment of the putative tumor suppressor protein DMBT1 from pancreatic adenocarcinoma cell lines. *Cancer Res.* **2002**, *62* (17), 4894–8.

(6) Baggerman, G.; Cerstiaens, A.; De Loof, A.; Schoofs, L. Peptidomics of the larval Drosophila melanogaster central nervous system. *J. Biol. Chem.* **2002**, *277* (43), 40368–74.

(7) Clynen, E.; De Loof, A.; Schoofs, L. The use of peptidomics in endocrine research. *Gen. Comp. Endrocrinol.* **2003**, *132* (1), 1–9.

(8) Fricker, L. D.; Lim, J.; Pan, H.; Che, F. Y. Peptidomics: identification and quantification of endogenous peptides in neuroendocrine tissues. *Mass Spectrom. Rev.* **2006**, *25* (2), 327–44.

(9) Hummon, A. B.; Amare, A.; Sweedler, J. V. Discovering new invertebrate neuropeptides using mass spectrometry. *Mass Spectrom. Rev.* **2006**, *25* (1), 77–98.

(10) Yuan, X.; Desiderio, D. M. Human cerebrospinal fluid peptidomics. *J. Mass Spectrom.* **2005**, *40* (2), 176–81.

(11) Schrader, M.; Selle, H. The process chain for peptidomic biomarker discovery. *Dis. Markers* **2006**, *22* (1−2), 27–37.

(12) Menzel, C.; Kellmann, M.; Juergens, M.; Schulz-Knappe, P. Optimized and integrated identification strategies for Peptidomics (R) biomarker discovery. *Mol. Cell. Proteomics* **2005**, *4* (8), S158–S158.

(13) Westman-Brinkmalm, A.; Ruetschi, U.; Portelius, E.; Andreasson, U.; Brinkmalm, G.; Karlsson, G.; Hansson, S.; Zetterberg, H.; Blennow, K. Proteomics/peptidomics tools to find CSF biomarkers for neurodegenerative diseases. *Front. Biosci.* **2009**, *14*, 1793–1806.

(14) Selle, H.; Lamerz, J.; Buerger, K.; Dessauer, A.; Hager, K.; Hampel, H.; Karl, J.; Kellmann, M.; Lannfelt, L.; Louhija, J.; Riepe, M.; Rollinger, W.; Tumani, H.; Schrader, M.; Zucht, H. D. Identification of novel biomarker candidates by differential peptidomics analysis of cerebrospinal fluid in Alzheimer's disease. *Comb. Chem. High Throughput Screen.* **2005**, *8* (8), 801–6.

(15) Petricoin, E. F.; Belluco, C.; Araujo, R. P.; Liotta, L. A. The blood peptidome: a higher dimension of information content for cancer biomarker discovery. *Nat. Rev. Cancer* **2006**, *6* (12), 961–7.

(16) Svensson, M.; Sköld, K.; Nilsson, A.; Fälth, M.; Nydahl, K.; Svenningsson, P.; Andren, P. E. Neuropeptidomics: MS applied to the discovery of novel peptides from the brain. *Anal. Chem.* **2007**, *79* (1), 18–21, 15−6.

(17) Clynen, E.; Baggerman, G.; Husson, S. J.; Landuyt, B.; Schoofs, L. Peptidomics in drug research. *Expert Opin. Drug Discovery* **2008**, *3* (4), 425–440.

(18) Adermann, K.; John, H.; Standker, L.; Forssmann, W. G. Exploiting natural peptide diversity: novel research tools and drug leads. *Curr. Opin. Biotechnol.* **2004**, *15* (6), 599–606.

(19) Meltretter, J.; Armand, F.; Perret, F.; Menin, L.; Stocklin, R. Peptidomics Techinologies and Their Application for Biomarker Discovery and Drug Profiling. *Biopolymers* **2009**, *92* (4), 348–349.

(20) Boonen, K.; Landuyt, B.; Baggerman, G.; Husson, S. J.; Huybrechts, J.; Schoofs, L. Peptidomics: The integrated approach of MS, hyphenated techniques and bioinformatics for neuropeptide analysis. *J. Sep. Sci.* **2008**, *31* (3), 427–445.

(21) Hu, L.; Ye, M.; Zou, H. Recent advances in mass spectrometry-based peptidome analysis. *Expert Rev. Proteomics* **2009**, *6* (4), 433–47.

(22) Wei, X.; Li, L. Mass spectrometry-based proteomics and peptidomics for biomarker discovery in neurodegenerative diseases. *Int. J. Clin. Exp. Pathol.* **2009**, *2* (2), 132–48.

(23) Rholam, M.; Fahy, C. Processing of peptide and hormone precursors at the dibasic cleavage sites. *Cell. Mol. Life Sci.* **2009**, *66* (13), 2075–2091.

(24) Hook, V.; Funkelstein, L.; Lu, D.; Bark, S.; Wegrzyn, J.; Hwang, S. R. Proteases for processing proneuropeptides into peptide neurotransmitters and hormones. *Annu. Rev. Pharmacol. Toxicol.* **2008**, *48*, 393–423.

(25) Nielsen, M. L.; Savitski, M. M.; Zubarev, R. A. Extent of modifications in human proteome samples and their effect on dynamic range of analysis in shotgun proteomics. *Mol. Cell. Proteomics* **2006**, *5* (12), 2384–91.

(26) Menschaert, G.; Vandekerckhove, T. T.; Landuyt, B.; Hayakawa, E.; Schoofs, L.; Luyten, W.; Van Criekinge, W. Spectral clustering in peptidomics studies helps to unravel modification profile of biologically active peptides and enhances peptide identification rate. *Proteomics* **2009**, *9*, 4381–8.

(27) Fälth, M.; Sköld, K.; Svensson, M.; Nilsson, A.; Fenyo, D.; Andren, P. E. Neuropeptidomics strategies for specific and sensitive identification of endogenous peptides. *Mol. Cell. Proteomics* **2007**, *6* (7), 1188–97.

(28) Svensson, M.; Sköld, K.; Svenningsson, P.; Andren, P. E. Peptidomics-based discovery of novel neuropeptides. *J. Proteome Res.* **2003**, *2* (2), 213–9.

(29) O'Callaghan, J. P.; Sriram, K. Focused microwave irradiation of the brain preserves in vivo protein phosphorylation: comparison with other methods of sacrifice and analysis of multiple phosphoproteins. *J. Neurosci. Methods* **2004**, *135* (1−2), 159–68.

(30) Svensson, M.; Boren, M.; Sköld, K.; Fälth, M.; Sjogren, B.; Andersson, M.; Svenningsson, P.; Andren, P. E. Heat stabilization

of the tissue proteome: a new technology for improved proteomics. *J. Proteome Res.* **2009**, *8* (2), 974–81.
(31) Boonen, K.; Baggerman, G.; Hertog, W. D.; Husson, S. J.; Overbergh, L.; Mathieu, C.; Schoofs, L. Neuropeptides of the islets of Langerhans: A peptidomics study. *Gen. Comp. Endrocrinol.* **2007**, *152* (2−3), 231–241.
(32) Scholz, B.; Sköld, K.; Kultima, K.; Fernandez, C.; Waldemarson, S.; Savitski, M. M.; Svensson, M.; Boren, M.; Stella, R.; Andren, P. E.; Zubarev, R.; James, P. Impact of temperature dependent sampling procedures in proteomics and peptidomics - A characterization of the liver and pancreas post mortem degradome. *Mol. Cell. Proteomics* 2010.
(33) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20* (18), 3551–67.
(34) Eng, J. K.; Mccormack, A. L.; Yates, J. R. An Approach to Correlate Tandem Mass-Spectral Data of Peptides with Amino-Acid-Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976–989.
(35) Craig, R.; Beavis, R. C. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **2004**, *20* (9), 1466–7.
(36) Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X. Y.; Shi, W. Y.; Bryant, S. H. Open mass spectrometry search algorithm. *J. Proteome Res.* **2004**, *3* (5), 958–964.
(37) Jimenez, C. R.; Huang, L.; Qiu, Y.; Burlingame, A. L. Searching sequence databases over the internet: protein identification using MS-Fit. *Curr. Protoc. Protein Sci.* **2001**, *Chapter 16*, Unit 16.5.
(38) Nesvizhskii, A. I.; Vitek, O.; Aebersold, R. Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat. Methods* **2007**, *4* (10), 787–97.
(39) Martens, L.; Vandekerckhove, J.; Gevaert, K. DBToolkit: processing protein databases for peptide-centric proteomics. *Bioinformatics* **2005**, *21* (17), 3584–3585.
(40) Vandekerckhove, T. T. M.; Menschaert, G.; Landuyt, B.; Boonen, K.; Van Criekinge, W. In *SOS MS: Enhancements for Mascot Search Engine and Other Mass Spectrometry Software*; ISMB-ECCB: Vienna, Austria, 2007.
(41) Reisinger, F.; Martens, L. Database on Demand - An online tool for the custom generation of FASTA-formatted sequence databases. *Proteomics* **2009**, *9*, 4421–4.
(42) Savitski, M. M.; Fälth, M. Peptide fragmentation and phosphosite detection. *Expert Rev. Proteomics* **2007**, *4* (4), 445–6.
(43) Zhang, X.; Che, F. Y.; Berezniuk, I.; Sonmez, K.; Toll, L.; Fricker, L. D. Peptidomics of Cpe(fat/fat) mouse brain regions: implications for neuropeptide processing. *J. Neurochem.* **2008**, *107* (6), 1596–613.
(44) Nesvizhskii, A. I.; Roos, F. F.; Grossmann, J.; Vogelzang, M.; Eddes, J. S.; Gruissem, W.; Baginsky, S.; Aebersold, R. Dynamic spectrum quality assessment and iterative computational analysis of shotgun proteomic data: toward more efficient identification of post-translational modifications, sequence polymorphisms, and novel peptides. *Mol. Cell. Proteomics* **2006**, *5* (4), 652–70.
(45) Bern, M.; Goldberg, D.; McDonald, W. H.; Yates, J. R. 3rd, Automatic quality assessment of peptide tandem mass spectra. *Bioinformatics* **2004**, *20* (Suppl 1), i49–54.
(46) Moore, R. E.; Young, M. K.; Lee, T. D. Method for screening peptide fragment ion mass spectra prior to database searching. *J. Am. Soc. Mass Spectrom.* **2000**, *11* (5), 422–6.
(47) Purvine, S.; Kolker, N.; Kolker, E. Spectral quality assessment for high-throughput tandem mass spectrometry proteomics. *OMICS* **2004**, *8* (3), 255–65.
(48) Xu, M.; Geer, L. Y.; Bryant, S. H.; Roth, J. S.; Kowalak, J. A.; Maynard, D. M.; Markey, S. P. Assessing data quality of peptide mass spectra obtained by quadrupole ion trap mass spectrometry. *J. Proteome Res.* **2005**, *4* (2), 300–5.
(49) Savitski, M. M.; Nielsen, M. L.; Zubarev, R. A. New data base-independent, sequence tag-based scoring of peptide MS/MS data validates Mowse scores, recovers below threshold data, singles out modified peptides, and assesses the quality of MS/MS techniques. *Mol. Cell. Proteomics* **2005**, *4* (8), 1180–1188.
(50) Wu, F. X.; Gagne, P.; Droit, A.; Poirier, G. G. Quality assessment of peptide tandem mass spectra. *BMC Bioinformatics* **2008**, *9* (Suppl 6), S13.
(51) Falkner, J.; Searle, B.; Andrews, P. Spectral clustering for increasing MSMS-based peptide identifications and identifying unknown modifications. *Mol. Cell. Proteomics* **2007**, *6* (8), 23–23.
(52) Falkner, J. A.; Falkner, J. W.; Yocum, A. K.; Andrews, P. C. A spectral clustering approach to MS/MS identification of post-translational modifications. *J. Proteome Res.* **2008**, *7* (11), 4614–22.
(53) Baumgartner, C.; Rejtar, T.; Kullolli, M.; Akella, L. M.; Karger, B. L. SeMoP: a new computational strategy for the unrestricted search for modified peptides using LC−MS/MS data. *J. Proteome Res.* **2008**, *7* (9), 4199–208.
(54) Savitski, M. M.; Nielsen, M. L.; Zubarev, R. A. ModifiComb, a new proteomic tool for mapping substoichiometric post-translational modifications, finding novel types of modifications, and fingerprinting complex protein mixtures. *Mol. Cell. Proteomics* **2006**, *5* (5), 935–948.
(55) Searle, B. C.; Dasari, S.; Turner, M.; Reddy, A. P.; Choi, D.; Wilmarth, P. A.; McCormack, A. L.; David, L. L.; Nagalla, S. R. High-throughput identification of proteins and unanticipated sequence modifications using a mass-based alignment algorithm for MS/MS de novo sequencing results. *Anal. Chem.* **2004**, *76* (8), 2220–30.
(56) Searle, B. C.; Dasari, S.; Wilmarth, P. A.; Turner, M.; Reddy, A. P.; David, L. L.; Nagalla, S. R. Identification of protein modifications using MS/MS de novo sequencing and the OpenSea alignment algorithm. *J. Proteome Re.s* **2005**, *4* (2), 546–54.
(57) Tsur, D.; Tanner, S.; Zandi, E.; Bafna, V.; Pevzner, P. A. Identification of post-translational modifications by blind search of mass spectra. *Nat. Biotechnol.* **2005**, *23* (12), 1562–7.
(58) Fälth, M.; Sköld, K.; Norrman, M.; Svensson, M.; Fenyo, D.; Andren, P. E. SwePep, a database designed for endogenous peptides and mass spectrometry. *Mol. Cell. Proteomics* **2006**, *5* (6), 998–1005.
(59) Zamyatnin, A. A.; Borchikov, A. S.; Vladimirov, M. G.; Voronina, O. L. The EROP-Moscow oligopeptide database. *Nucleic Acids Res.* **2006**, *34* (Database issue), D261–6.
(60) Liu, F.; Baggerman, G.; Schoofs, L.; Wets, G. The construction of a bioactive peptide database in Metazoa. *J. Proteome Res.* **2008**, *7* (9), 4119–31.
(61) Minamino, N. [Peptidome: the fact-database for endogenous peptides]. *Tanpakushitsu Kakusan Koso* **2001**, *46* (11 Suppl), 1510–7.
(62) Fälth, M.; Svensson, M.; Nilsson, A.; Sköld, K.; Fenyo, D.; Andren, P. E. Validation of endogenous peptide identifications using a database of tandem mass spectra. *J. Proteome Res.* **2008**, *7* (7), 3049–53.
(63) NIST/EPA/NIH (NIST 08) Mass Spectral Library.
(64) Lam, H.; Deutsch, E. W.; Eddes, J. S.; Eng, J. K.; Stein, S. E.; Aebersold, R. Building consensus spectral libraries for peptide identification in proteomics. *Nat. Methods* **2008**, *5* (10), 873–5.
(65) Yen, C. Y.; Meyer-Arendt, K.; Eichelberger, B.; Sun, S.; Houel, S.; Old, W. M.; Knight, R.; Ahn, N. G.; Hunter, L. E.; Resing, K. A. A simulated MS/MS library for spectrum-to-spectrum searching in large scale identification of proteins. *Mol. Cell. Proteomics* **2009**, *8* (4), 857–69.
(66) Frewen, B.; MacCoss, M. J. Using BiblioSpec for creating and searching tandem MS peptide libraries. *Curr. Protoc. Bioinf.* **2007**, *Chapter 13*, Unit 13.7.
(67) Ma, B.; Lajoie, G. De novo interpretation of tandem mass spectra. *Curr. Protoc. Bioinf.* **2009**, *Chapter 13*, Unit 13.10.
(68) Frank, A.; Pevzner, P. PepNovo: de novo peptide sequencing via probabilistic network modeling. *Anal. Chem.* **2005**, *77* (4), 964–73.
(69) Frank, A. M.; Savitski, M. M.; Nielsen, M. L.; Zubarev, R. A.; Pevzner, P. A. De novo peptide sequencing and identification with precision mass spectrometry. *J. Proteome Res.* **2007**, *6* (1), 114–23.
(70) Frank, A. M. A ranking-based scoring function for peptide-spectrum matches. *J. Proteome Res.* **2009**, *8* (5), 2241–52.
(71) Frank, A. M. Predicting intensity ranks of peptide fragment ions. *J. Proteome Res.* **2009**, *8* (5), 2226–40.
(72) Dancik, V.; Addona, T. A.; Clauser, K. R.; Vath, J. E.; Pevzner, P. A. De novo peptide sequencing via tandem mass spectrometry. *J. Comput. Biol.* **1999**, *6* (3−4), 327–42.
(73) Tabb, D. L.; Ma, Z. Q.; Martin, D. B.; Ham, A. J.; Chambers, M. C. DirecTag: accurate sequence tags from peptide MS/MS through statistical scoring. *J. Proteome Res.* **2008**, *7* (9), 3838–46.
(74) Jimenez, C. R.; Huang, L.; Qiu, Y.; Burlingame, A. L. Searching sequence databases over the internet: protein identification using MS-tag. *Curr. Protoc. Protein Sci.* **2001**, *Chapter 16*, Unit 16.6.
(75) Habermann, B.; Oegema, J.; Sunyaev, S.; Shevchenko, A. The power and the limitations of cross-species protein identification by mass spectrometry-driven sequence similarity searches. *Mol. Cell. Proteomics* **2004**, *3* (3), 238–49.

(76) Shevchenko, A.; Sunyaev, S.; Loboda, A.; Bork, P.; Ens, W.; Standing, K. G. Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching. *Anal. Chem.* **2001**, *73* (9), 1917–26.
(77) Han, Y.; Ma, B.; Zhang, K. SPIDER: software for protein identification from sequence tags with de novo sequencing error. *J. Bioinform. Comput. Biol.* **2005**, *3* (3), 697–716.
(78) Tanner, S.; Shu, H.; Frank, A.; Wang, L. C.; Zandi, E.; Mumby, M.; Pevzner, P. A.; Bafna, V. InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. *Anal. Chem.* **2005**, *77* (14), 4626–39.
(79) Roth, S.; Fromm, B.; Gade, G.; Predel, R. A proteomic approach for studying insect phylogeny: CAPA peptides of ancient insect taxa (Dictyoptera, Blattoptera) as a test case. *BMC Evol. Biol.* **2009**, *9*, 50.
(80) Ma, M.; Chen, R.; Sousa, G. L.; Bors, E. K.; Kwiatkowski, M. A.; Goiney, C. C.; Goy, M. F.; Christie, A. E.; Li, L. Mass spectral characterization of peptide transmitters/hormones in the nervous system and neuroendocrine organs of the American lobster Homarus americanus. *Gen. Comp. Endrocrinol.* **2008**, *156* (2), 395–409.
(81) Clynen, E.; Schoofs, L. Peptidomic survey of the locust neuroendocrine system. *Insect Biochem. Mol. Biol.* **2009**, *39* (8), 491–507.
(82) Bora, A.; Annangudi, S. P.; Millet, L. J.; Rubakhin, S. S.; Forbes, A. J.; Kelleher, N. L.; Gillette, M. U.; Sweedler, J. V. Neuropeptidomics of the supraoptic rat nucleus. *J. Proteome Res.* **2008**, *7* (11), 4992–5003.
(83) Bertile, F.; Robert, F.; Delval-Dubois, V.; Sanglier, S.; Schaeffer, C.; Van Dorsselaer, A. Endogenous plasma Peptide detection and identification in the rat by a combination of fractionation methods and mass spectrometry. *Biomark Insights* **2007**, *2*, 385–401.
(84) Ons, S.; Richter, F.; Urlaub, H.; Pomar, R. R. The neuropeptidome of Rhodnius prolixus brain. *Proteomics* **2009**, *9* (3), 788–92.
(85) Yasothornsrikul, S.; Greenbaum, D.; Medzihradszky, K. F.; Toneff, T.; Bundey, R.; Miller, R.; Schilling, B.; Petermann, I.; Dehnert, J.; Logvinova, A.; Goldsmith, P.; Neveu, J. M.; Lane, W. S.; Gibson, B.; Reinheckel, T.; Peters, C.; Bogyo, M.; Hook, V. Cathepsin L in secretory vesicles functions as a prohormone-processing enzyme for production of the enkephalin peptide neurotransmitter. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (16), 9590–5.
(86) Menschaert, G.; Vandekerckhove, T. T.; Baggerman, G.; Landuyt, B.; Sweedler, J. V.; Schoofs, L.; Luyten, W.; Van Criekinge, W. A hybrid, de novo based, genome-wide database search approach applied to the sea urchin neuropeptidome. *J. Proteome Res.* **2010**, *9* (2), 990–6.
(87) Kim, S.; Gupta, N.; Bandeira, N.; Pevzner, P. A. Spectral dictionaries: Integrating de novo peptide sequencing with database search of tandem mass spectra. *Mol. Cell. Proteomics* **2009**, *8* (1), 53–69.
(88) Bradshaw, R. A.; Burlingame, A. L.; Carr, S.; Aebersold, R. Reporting protein identification data: the next generation of guidelines. *Mol. Cell. Proteomics* **2006**, *5* (5), 787–8.
(89) Carr, S.; Aebersold, R.; Baldwin, M.; Burlingame, A.; Clauser, K.; Nesvizhskii, A. The need for guidelines in publication of peptide and protein identification data: Working Group on Publication Guidelines for Peptide and Protein Identification Data. *Mol. Cell. Proteomics* **2004**, *3* (6), 531–3.
(90) Elias, J. E.; Haas, W.; Faherty, B. K.; Gygi, S. P. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat. Methods* **2005**, *2* (9), 667–75.
(91) Wang, G.; Wu, W. W.; Zhang, Z.; Masilamani, S.; Shen, R. F. Decoy methods for assessing false positives and false discovery rates in shotgun proteomics. *Anal. Chem.* **2009**, *81* (1), 146–59.
(92) Pappin, D. J.; Hojrup, P.; Bleasby, A. J. Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.* **1993**, *3* (6), 327–32.
(93) Kim, S.; Gupta, N.; Pevzner, P. A. Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases. *J. Proteome Res.* **2008**, *7* (8), 3354–63.
(94) Shtatland, T.; Guettler, D.; Kossodo, M.; Pivovarov, M.; Weissleder, R. PepBank--a database of peptides based on sequence text mining and public peptide data sources. *BMC Bioinf.* **2007**, *8*, 280.
(95) Liu, F.; Baggerman, G.; D'Hertog, W.; Verleyen, P.; Schoofs, L.; Wets, G. In silico identification of new secretory peptide genes in Drosophila melanogaster. *Mol. Cell. Proteomics* **2006**, *5* (3), 510–22.
(96) Liu, F.; Baggerman, G.; Schoofs, L.; Wets, G. Uncovering conserved patterns in bioactive peptides in Metazoa. *Peptides* **2006**, *27* (12), 3137–53.
(97) Baggerman, G.; Liu, F.; Wets, G.; Schoofs, L. Bioinformatic analysis of peptide precursor proteins. *Ann. N.Y. Acad. Sci.* **2005**, *1040*, 59–65.
(98) Mirabeau, O.; Perlas, E.; Severini, C.; Audero, E.; Gascuel, O.; Possenti, R.; Birney, E.; Rosenthal, N.; Gross, C. Identification of novel peptide hormones in the human proteome by hidden Markov model screening. *Genome Res.* **2007**, *17* (3), 320–7.
(99) Southey, B. R.; Amare, A.; Zimmerman, T. A.; Rodriguez-Zas, S. L.; Sweedler, J. V. NeuroPred: a tool to predict cleavage sites in neuropeptide precursors and provide the masses of the resulting peptides. *Nucleic Acids Res.* **2006**, *34* (Web Server issue), W267–W272.
(100) Southey, B. R.; Sweedler, J. V.; Rodriguez-Zas, S. L. A python analytical pipeline to identify prohormone precursors and predict prohormone cleavage sites. *Front. Neuroinf.* **2008**, *2*, 7.
(101) Sonmez, K.; Zaveri, N. T.; Kerman, I. A.; Burke, S.; Neal, C. R.; Xie, X.; Watson, S. J.; Toll, L. Evolutionary sequence modeling for discovery of peptide hormones. *PLoS Comput Biol* **2009**, *5* (1), e1000258.
(102) Shemesh, R.; Toporik, A.; Levine, Z.; Hecht, I.; Rotman, G.; Wool, A.; Dahary, D.; Gofer, E.; Kliger, Y.; Soffer, M. A.; Rosenberg, A.; Eshel, D.; Cohen, Y. Discovery and validation of novel peptide agonists for G-protein-coupled receptors. *J. Biol. Chem.* **2008**, *283* (50), 34643–9.
(103) Muggleton, S. H.; Bryant, C. H.; Srinivasan, A.; Whittaker, A.; Topp, S.; Rawlings, C. Are grammatical representations useful for learning from biological sequence data?--a case study. *J. Comput. Biol.* **2001**, *8* (5), 493–521.
(104) Husson, S. J.; Clynen, E.; Baggerman, G.; Janssen, T.; Schoofs, L. Defective processing of neuropeptide precursors in Caenorhabditis elegans lacking proprotein convertase 2 (KPC-2/EGL-3): mutant analysis by mass spectrometry. *J. Neurochem.* **2006**, *98* (6), 1999–2012.
(105) Husson, S. J.; Janssen, T.; Baggerman, G.; Bogert, B.; Kahn-Kirby, A. H.; Ashrafi, K.; Schoofs, L. Impaired processing of FLP and NLP peptides in carboxypeptidase E (EGL-21)-deficient Caenorhabditis elegans as analyzed by mass spectrometry. *J. Neurochem.* **2007**, *102* (1), 246–60.
(106) Che, F. Y.; Yan, L.; Li, H.; Mzhavia, N.; Devi, L. A.; Fricker, L. D. Identification of peptides from brain and pituitary of Cpe(fat)/Cpe(fat) mice. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98* (17), 9971–6.
(107) Sköld, K.; Svensson, M.; Norrman, M.; Sjogren, B.; Svenningsson, P.; Andren, P. E. The significance of biochemical and molecular sample integrity in brain proteomics and peptidomics: stathmin 2−20 and peptides as sample quality indicators. *Proteomics* **2007**, *7* (24), 4445–56.
(108) Ishihama, Y.; Oda, Y.; Tabata, T.; Sato, T.; Nagasu, T.; Rappsilber, J.; Mann, M. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell. Proteomics* **2005**, *4* (9), 1265–72.
(109) Flikka, K.; Martens, L.; Vandekerckhove, J.; Gevaert, K.; Eidhammer, I. Improving the reliability and throughput of mass spectrometry-based proteomics by spectrum quality filtering. *Proteomics* **2006**, *6* (7), 2086–2094.
(110) Che, F. Y.; Biswas, R.; Fricker, L. D. Relative quantitation of peptides in wild-type and Cpe(fat/fat) mouse pituitary using stable isotopic tags and mass spectrometry. *J. Mass Spectrom.* **2005**, *40* (2), 227–37.
(111) Che, F. Y.; Eipper, B. A.; Mains, R. E.; Fricker, L. D. Quantitative peptidomics of pituitary glands from mice deficient in copper transport. *Cell. Mol. Biol. (Noisy-le-grand)* **2003**, *49* (5), 713–22.
(112) Che, F. Y.; Fricker, L. D. Quantitative peptidomics of mouse pituitary: comparison of different stable isotopic tags. *J. Mass Spectrom.* **2005**, *40* (2), 238–49.
(113) Che, F. Y.; Lim, J.; Pan, H.; Biswas, R.; Fricker, L. D. Quantitative neuropeptidomics of microwave-irradiated mouse brain and pituitary. *Mol. Cell. Proteomics* **2005**, *4* (9), 1391–405.
(114) Morano, C.; Zhang, X.; Fricker, L. D. Multiple isotopic labels for quantitative mass spectrometry. *Anal. Chem.* **2008**, *80* (23), 9298–309.
(115) Xiang, F.; Ye, H.; Chen, R.; Fu, Q.; Li, L. N,N-Dimethyl Leucines as Novel Isobaric Tandem Mass Tags for Quantitative Proteomics and Peptidomics. *Anal. Chem.* **2010**, *82*, 2817–25.
(116) Boehm, A. M.; Putz, S.; Altenhofer, D.; Sickmann, A.; Falk, M. Precise protein quantification based on peptide quantification using iTRAQ. *BMC Bioinf.* **2007**, *8*, 214.

(117) Sturm, M.; Bertsch, A.; Gropl, C.; Hildebrandt, A.; Hussong, R.; Lange, E.; Pfeifer, N.; Schulz-Trieglaff, O.; Zerck, A.; Reinert, K.; Kohlbacher, O. OpenMS - an open-source software framework for mass spectrometry. *BMC Bioinf.* **2008**, *9*, 163.
(118) Vaudel, M.; Sickmann, A.; Martens, L. Peptide and protein quantification: a map of the minefield. *Proteomics* **2010**, *10* (4), 650–70.
(119) Amstalden van Hove, E. R.; Smith, D. F.; Heeren, R. M. A concise review of mass spectrometry imaging. *J. Chromatogr., A* **2010**; DOI: 10.1016/j.chroma.2010.01.033.
(120) Solon, E. G.; Schweitzer, A.; Stoeckli, M.; Prideaux, B. Autoradiography, MALDI-MS, and SIMS-MS imaging in pharmaceutical discovery and development. *AAPS J.* **2010**, *12* (1), 11–26.
(121) Van de Plas, R.; Ojeda, F.; Dewil, M.; Van Den Bosch, L.; De Moor, B.; Waelkens, E. Prospective exploration of biochemical tissue composition via imaging mass spectrometry guided by principal component analysis. *Pac. Symp. Biocomput.* **2007**, 458–69.
(122) Fälth, M.; Savitski, M. M.; Nielsen, M. L.; Kjeldsen, F.; Andren, P. E.; Zubarev, R. A. SwedCAD, a database of annotated high-mass accuracy MS/MS spectra of tryptic peptides. *J. Proteome Res.* **2007**, *6* (10), 4063–7.
(123) Martens, L.; Hermjakob, H.; Jones, P.; Adamski, M.; Taylor, C.; States, D.; Gevaert, K.; Vandekerckhove, J.; Apweiler, R. PRIDE: the proteomics identifications database. *Proteomics* **2005**, *5* (13), 3537–45.
(124) Deutsch, E. W.; Lam, H.; Aebersold, R. Data analysis and bioinformatics tools for tandem mass spectrometry in proteomics. *Physio.l Genomics* **2008**, *33* (1), 18–25.
(125) Galindo, M. I.; Pueyo, J. I.; Fouix, S.; Bishop, S. A.; Couso, J. P. Peptides encoded by short ORFs control development and define a new eukaryotic gene family. *PLoS Biol.* **2007**, *5* (5), e106.
(126) Savard, J.; Marques-Souza, H.; Aranda, M.; Tautz, D. A segmentation gene in tribolium produces a polycistronic mRNA that codes for multiple conserved peptides. *Cell* **2006**, *126* (3), 559–69.
(127) Altschul, S. F. Amino acid substitution matrices from an information theoretic perspective. *J. Mol. Biol.* **1991**, *219* (3), 555–65.
(128) Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25* (17), 3389–402.
(129) Savitski, M. M.; Nielsen, M. L.; Kjeldsen, F.; Zubarev, R. A. Proteomics-grade de novo sequencing approach. *J. Proteome Res.* **2005**, *4* (6), 2348–2354.
(130) Nielsen, M. L.; Savitski, M. M.; Zubarev, R. A. Improving protein identification using complementary fragmentation techniques in Fourier transform mass spectrometry. *Mol. Cell. Proteomics* **2005**, *4* (6), 835–845.
(131) Zubarev, R. A.; Zubarev, A. R.; Savitski, M. M. Electron capture/transfer versus collisionally activated/induced dissociations: Solo or duet. *J. Am. Soc. Mass Spectrom.* **2008**, *19* (6), 753–761.
(132) Bendtsen, J. D.; Nielsen, H.; von Heijne, G.; Brunak, S. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **2004**, *340* (4), 783–95.
(133) Li, B.; Predel, R.; Neupert, S.; Hauser, F.; Tanaka, Y.; Cazzamali, G.; Williamson, M.; Arakane, Y.; Verleyen, P.; Schoofs, L.; Schachtner, J.; Grimmelikhuijzen, C. J.; Park, Y. Genomics, transcriptomics, and peptidomics of neuropeptides and protein hormones in the red flour beetle Tribolium castaneum. *Genome Res.* **2008**, *18* (1), 113–22.
(134) Husson, S. J.; Landuyt, B.; Nys, T.; Baggerman, G.; Boonen, K.; Clynen, E.; Lindemans, M.; Janssen, T.; Schoofs, L. Comparative peptidomics of Caenorhabditis elegans versus C. briggsae by LC-MALDI-TOF MS. *Peptides* **2009**, *30* (3), 449–57.
(135) Zougman, A.; Pilch, B.; Podtelejnikov, A.; Kiehntopf, M.; Schnabel, C.; Kumar, C.; Mann, M. Integrated analysis of the cerebrospinal fluid peptidome and proteome. *J. Proteome Res.* **2008**, *7* (1), 386–99.
(136) Fricker, L. D. Neuropeptide-processing enzymes: applications for drug discovery. *AAPS J.* **2005**, *7* (2), E449–55.
(137) Southey, B. R.; Rodriguez-Zas, S. L.; Sweedler, J. V. Characterization of the prohormone complement in cattle using genomic libraries and cleavage prediction approaches. *BMC Genomics* **2009**, *10*, 228.
(138) Wang, J.; Ma, M.; Chen, R.; Li, L. Enhanced neuropeptide profiling via capillary electrophoresis off-line coupled with MALDI FTMS. *Anal. Chem.* **2008**, *80* (16), 6168–77.
(139) Yew, J. Y.; Wang, Y.; Barteneva, N.; Dikler, S.; Kutz-Naber, K. K.; Li, L.; Kravitz, E. A. Analysis of neuropeptide expression and localization in adult drosophila melanogaster central nervous system by affinity cell-capture mass spectrometry. *J. Proteome Res.* **2009**, *8* (3), 1271–84.
(140) Boerjan, B.; Cardoen, D.; Bogaerts, A.; Landuyt, B.; Schoofs, L.; Verleyen, P. Mass spectrometric profiling of (neuro)-peptides in the worker honeybee, *Apis mellifera*. *Neuropharmacology* **2010**, *58*, 248–58.
(141) Gomes, I.; Grushko, J. S.; Golebiewska, U.; Hoogendoorn, S.; Gupta, A.; Heimann, A. S.; Ferro, E. S.; Scarlata, S.; Fricker, L. D.; Devi, L. A. Novel endogenous peptide agonists of cannabinoid receptors. *FASEB J.* **2009**, *23* (9), 3020–9.
(142) Hummon, A. B.; Richmond, T. A.; Verleyen, P.; Baggerman, G.; Huybrechts, J.; Ewing, M. A.; Vierstraete, E.; Rodriguez-Zas, S. L.; Schoofs, L.; Robinson, G. E.; Sweedler, J. V. From the genome to the proteome: uncovering peptides in the Apis brain. *Science* **2006**, *314* (5799), 647–9.
(143) Antwi, K.; Hanavan, P. D.; Myers, C. E.; Ruiz, Y. W.; Thompson, E. J.; Lake, D. F. Proteomic identification of an MHC-binding peptidome from pancreas and breast cancer cell lines. *Mol. Immunol.* **2009**, *46* (15), 2931–7.
(144) Baggerman, G.; Boonen, K.; Verleyen, P.; De Loof, A.; Schoofs, L. Peptidomic analysis of the larval Drosophila melanogaster central nervous system by two-dimensional capillary liquid chromatography quadrupole time-of-flight mass spectrometry. *J. Mass Spectrom.* **2005**, *40* (2), 250–60.
(145) Tanner, S.; Shu, H.; Frank, A.; Wang, L. C.; Zandi, E.; Mumby, M.; Pevzner, P. A.; Bafna, V. InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. *Anal. Chem.* **2005**, *77* (14), 4626–39.

PR900929M