MS approaches to select peptides with post-translational modifications from amphibian defense secretions prior to full sequence elucidation

Martijn Pinkse, Geisa Evaristo, Mervin Pieterse, Yuanjie Yu, Peter Verhaert

A R T I C L E   I N F O

Article history:
Available online 24 November 2014

Keywords:
Amphibian skin secretion
Amidation
d-Amino acid
Tyrosine-sulfation
Pyroglutamic acid
Disulfide bridges

A B S T R A C T

Peptide families are characterized by structural motifs, which often comprise specific post-translational modifications (PTMs) required for biological activity. In conventional bioactivity-based peptidomics studies natural peptide mixtures are chromatographically separated and the bioactive fractions purified to homogeneity, prior to structural characterization. In this paper we illustrate the reverse methodology, in which the primary structures of peptides with presumed bioactivity are first determined before investigating functions/bioactivities. We exemplify mass spectrometry (MS)-based strategies (employing, in particular, high resolution MS) to specifically select peptides – from complex mixtures such as frog defensive secretions – by virtue of the occurrence of particular PTMs, including amidation, disulfide-bonding, L- to d-amino acid isomerization, tyrosine-sulfation, proline-hydroxylation, and aminoterminal pyroglutamate formation.

© 2014 The Authors. Published by Elsevier B.V. on behalf of European Proteomics Association (EuPA). This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

1. Introduction: conventional bioactivity-based peptide discovery

Skin secretions of anurans are a rich source of biologically active peptides with high pharmaceutical potential. Over the last decades various frog secretions have been subject of intense research and this has led to the discovery and characterization of several peptides with interesting biological activity (e.g., [1]). In the majority of cases, these bioactive peptides are first identified by specific bioassay screening and after this their structural sequence is further elucidated using Edman degradation and/or mass spectrometry [2]. Many of the peptides identified and characterized by this approach so far are post-translationally modified (e.g., [3,4]). These post-translational modifications (PTMs) contribute to the bioactivity of these peptides, either by enhancing receptor binding affinity or by extending the molecules’ half-life.
due to increased resistance against proteolytic degradation (e.g., [5]). In this traditional discovery scheme, a substantial amount of secretion/venom is required for individual peptide purification and bioassays screening. However, the amphibian secretions are complex cocktails of peptides with many of the highly active compounds present in low abundance [6]. The number of bioassays that can be performed on limited size samples is restricted, and many potentially interesting peptides risk to be missed by this narrow approach and thus to escape from the investigator’s attention.

2. Alternative peptide discovery strategies based on structural analyses

2.1. Peptide encoding mRNA based

In cases where full length messenger RNAs can be extracted, molecular cloning can be an excellent alternative technique to obtain (or confirm) the primary structure of frog peptides. Pioneering work by Chen, Shaw and co-workers has shown that amphibian skin secretions still contain clonable full length mRNAs which encode the typical defense peptides [6]. This thanks to the fact that frogs secrete peptides through a holocrine process which involves rupturing of the mature granular syncytium at the center of the poison gland, resulting in the release of all cytoplasmic components in the secretion, including intact poly-adenylated mRNAs. We emphasize that this is a very elegant and sustainable way of obtaining sequence information of amphibian secretory peptides, which is totally non-invasive and completely harmless to the sample donor. In this respect it is upsetting to read that seven years after Chen’s original publication, scores of specimens continue to be sacrificed for peptide transcript sequencing purposes [7]. Yet, it shall be clear that also this molecular cloning approach has its limitations. Not all peptide encoding mRNAs are equally abundant, and some may, therefore, be overlooked. More importantly, essential PTMs featuring on the mature secreted peptides cannot be promptly predicted from cDNA sequencing results. However, particularly when combined with high resolution tandem mass spectrometry (MS) data on the respective peptide gene products (as discussed below), nucleic acid derived sequence data are highly useful for full primary structure elucidation. We have personally used this method successfully to help elucidate the primary structure of several different post-translationally modified frog peptides [8,9].

2.2. Peptide based

It is evident that the best sample to analyze in order to get a full structure of a mature naturally occurring bioactive peptide, is the peptide itself, directly from its biological source.

To achieve this, the high resolution and dynamic range of today’s generation tandem MS offer elegant alternative as well as complementary approaches in the study of complex amphibian defense secretions. We here present an overview of different LC–MS/MS-based methodologies which we have used/elaborated thus far in our frog skin peptide research.

3. Concept: PTM directed MS analysis to screen for peptides with potential bioactivity

In view of the large number of different peptides which are detectable in the amphibian skin secretion, the overall aim to fully sequence all of them and perform bioassays to characterize their biological activity, as pioneered by Vittorio Erspamer [10], is a virtually impossible task. Whereas LC–MS/MS analyses on hybrid orbitrap systems yield Gbyte-size data sets, only in very rare cases can the full structure of a peptide be readily deduced from automated data acquisitions. Typically additional labor intensive manual (re-) analyses are required before the primary structure of a peptide ion peak of interest can be completely resolved.

We here discuss and illustrate various analytical workflows which can be employed to select peptides for full structure elucidation, on the basis of specific structural features detectable by MS, which they share with known regulatory peptides. These include the presence of certain PTMs, such as carboxyterminal amidation, disulphide-bonding, L- to D-amino acid isomerization, tyrosine-sulfation, proline-hydroxylation and aminoterminal pyroglutamate formation [11].

The idea behind this tactic is as follows: peptide ions with structural features in common with known bioactive peptides are analyzed to reveal specific signatures in overall secretome LC–MS profiles. These signatures can either be directly visible on a 2D image representation of the LC–MS run, or they may exhibit themselves after chemical treatment of the sample. By setting the MS system to focus on PTM-specific features, including neutral losses; and taking advantage of the high resolution of orbitrap-based instruments to distinguish between quasi-isobaric regular residues and modified residues, candidate bioactive peptides can be pinpointed with little effort. Subsequent de novo MS/MS sequencing of the selected peptides including the identification of their PTMs can then follow after the generation of sufficiently information rich fragmentation spectra. In our experience, this often requires a combined analysis of the fragmentation data of CID (collision induced dissociation)/HCD (higher energy collision induced dissociation) experiments on precursor ions of multiple charge states from the same peptide.

Fig. 1 shows the LC–MS profile (chromatogram) of a crude washed-off skin gland secretion – we typically use a ‘frog milk-ing’ procedure as originally described by Tyler et al. [12] – of Phyllomedusa burmeisteri as typical example. The total ion at each chromatographic time point (retention time) contains multiple peaks from several peptides in varying charge states (Fig. 2A). Some of these peptide ions yield good fragmentation spectra from which partial or full sequence information can be obtained straightforward (as demonstrated in Fig. 2B). A database search with LC–MS/MS data using search engines for proteomics data such as Mascot (www.matrixscience.com) or Sequest (http://fields.scripps.edu/sequest), therefore, usually yields a number of peptide identifications. However, this classically holds for only the minority of the fragmented precursors, and these obviously represent peptides that had been identified before and were submitted to public databases such as Uniprot or others. Not seldom it is necessary to combine the fragmentation spectra of multiple charge states of the
same precursor ion to get to a more complete sequence coverage. Manual interpretation of the MS/MS spectra that are not initially matched by these database (de novo sequencing) searches therefore often turns out to be very labor intensive and time consuming. Yet, to date this still remains the only way to solve a novel peptide sequence, although continually improved algorithms combining database mining with de novo sequencing (e.g., PEAKS v. 7; www.bioinfor.com), promise to be a step in the right direction.

To enhance the success-rate, various methods utilizing the performance of today’s generation high resolution MS can be employed to specifically target peptides that contain ‘predefined’ PTMs.

A quick view on the complexity of the sample is obtained by automated conversion of LC–MS data into a 2-dimensional (2-D) display, in which both retention time (RT) and mass-to-charge ratio (m/z) are plotted (Fig. 3). To achieve this, we typically employ the publicly available software tool MSight (SIB, Switzerland, http://web.expasy.org/MSight/; originally described by [13]) for this conversion. In addition, the thus generated LC–MS ‘images’ enable the experienced eye to either directly or indirectly (after chemical treatment) ‘spot’ peptides with one of the particular PTMs mentioned above.

Indeed, a chemical or enzymatic treatment affecting a specific PTM prior to a second LC–MS analysis results in specific shifts in RT and/or m/z. As a consequence, peptides carrying such PTM are readily detected by their altered location in the 2-D profile, when comparing treated with non-treated samples.

(Differential) 2-D profiling can thus be employed to create a list of those peptide ions with predicted PTMs that deserve subsequent characterization by tandem MS in more detail. To demonstrate the power of this approach we below list some illustrative examples.

3.1. C-terminally amidated peptides

Carboxyterminal amidation of peptides is one of the PTMs that is required for the bioactivity of many known signaling peptides. The mass difference between the free carboxyl and
the amide group in a peptide's carboxyterminal residue is 1 Da (more exactly −0.984 Da; the very same difference as between Glu and Gln or Asp and Asn). We observed that not seldom a small percentage of a C-terminally amidated peptide is found in the sample as free acid isoform. This is remarkable, as, to the best of our knowledge, it is not described that the typical peptidylglycine alpha-amidating monoxygenase (PAM) responsible for C-terminal amidation, can yield this side product. Anyhow, by virtue of this trait, amidated peptides are easily recognized in the 2-D display as two ions (isotope clusters) with a −0.984 Da shift in m/z and small shift in RT as illustrated in (Fig. 4A). Amidated and non-amidated/de-amidated peptides typically do not differ much in chromatographic elution behavior and in most cases they elute near each other. Once this feature is detected, the fragmentation spectra of both forms of this peptide need to be analyzed and compared to yield the full sequence (Fig. 4B). De novo sequencing of carboxyterminally amidated peptides is facilitated by comparative analysis with its free acid version. Since this PTM occurs at the final C-terminal

Fig. 3 – 2-Dimensional LC–MS peptide profile of *P. burmeisteri* skin secretion (image generated using MSight [http://web.expasy.org/MSight]), illustrating sample complexity.

![Image](image_url)

*Fig. 4 – Illustration of peptide with and without C-terminal amidation. (A) Zoom in area of 2-D peptide display depicting peptide ion isotopes. Characteristic is close retention time and 0.98 amu mass difference. (B) Orbitrap MS/MS spectra of both peptides showing similar fragmentation pattern. M/z values of b-ions are identical, whereas m/z values of y-ions are shifted by 0.98 amu in non-amidated peptide compared to amidated one. 2D display shows additional peptide ion cluster 4 Da heavier than carboxyterminally amidated peptide, illustrating sample complexity, revealed by analytical resolution.*
residue, both amidated and non-amidated peptide tandem mass spectra yield identical b-ions series, whereas all y-ions shift 0.984 Da (change of −OH to −NH$_2$ from free acid to amide). The easy annotation of y- and b-ion series greatly facilitates de novo sequencing (Fig. 4B). Tandem MS also allows to distinguish C-terminal amidation from other PTMs with identical mass shifts, such as de-amidations of Gln or Asn residues.

3.2. d-Amino acid containing peptides

Using the same 2-D display, other PTMs encountered on frog peptides can be observed as well. Several frog peptides have been characterized to contain a single d-amino acid at the second N-terminal position, for example deltorphin [14,15]. Although all peptides are initially fully synthesized in their l-amino acid form, peptidyl-aminoacyl-l/d isomerases, which have been identified in frog skin before [16] may specifically alter the stereochemistry of – typically – one amino acid within a peptide. This conversion is, however, seldom complete and, therefore, both versions of the peptide (the original full l-isoform and the one with a d-residue) can be ‘spotted’ in a sample 2D display. Both obviously have an identical mass, however their chromatographic behavior is slightly different. For example, deltorphin is detected at two different RTs in $P$. burmeisteri secretion with both differently eluting peptides displaying identical peptide fragmentation. Also in Bombina variegata skin secretion several peptides are found to show this behavior, suggesting the presence of a d-amino acid. As example, Fig. 5A highlights five l- and d-peptide pairs in the

Fig. 5 – (A) Zoomed-in part of 2D LC/MS analysis of skin secretion of $B$. variegata, highlighting several peptides pairs that have identical mass, but different retention times. This feature is indicative of presence of d-amino acid in one of two variants. (B) MS/MS spectra of two such peptides are virtually identical. Evidently higher resolution MS systems (here orbitrap) allow more accurate comparison of fragmentation spectra.

Fig. 6 – Schematic overview of neutral loss driven data-dependent MS$^2$/MS$^3$ scans-settings used for detection of tyrosine-sulfated peptides in hybrid ion trap-orbitrap system (LTQ-Orbitrap Velos; ThermoFischer).
Fig. 7 – Chromatograms of data dependent analyses using sulfate neutral loss driven MS3 analysis of sulfated peptides from skin secretion of *P. burmeisteri*. Mass range of this measurement was between 600 and 800 m/z. Shown are ion intensity chromatograms of (A) MS spectra, (B) MS2 spectra and (C) MS3 spectra.

2-D display of *B. variegata* crude secretion. The mass spectrometer cannot distinguish between the L- or D-residue as in both cases the MS/MS spectra are identical (Fig. 5B). Additional amino acid analysis is needed to solve this, however, the 2-D display allows one to quickly detect those peptides that potentially contain a D-amino acid. Also the use of ion mobility MS can help to determine the exact position of the D-amino acid [17,18].

Obviously also other phenomena may result in different retention times for peptides with (nearly) identical MS/MS profiles. These include Ile/Leu peptide isoforms or peptides with other isobaric sequence variations. Most of the latter, however, are not PTMs but rather sequence mutations/polymorphisms that are much less plausible to occur together in the same 2D display.

### 3.3. Disulfide bridge containing peptides

The cyclic nature of native peptides with Cys residues forming disulfide bonds (cystines) hampers efficient backbone cleavage in tandem MS. The resulting MS/MS spectra are relatively poor in fragment ions, yielding very ambiguous partial primary structures. After reduction of such peptides by dithiothreitol (DTT) or Tris-2-carboxyethyl phosphine (TCEP), the disulfide containing peptides appear much easier to fragment, yielding less ambiguous sequences. They are readily detected in the 2D differential display by a shift in +2 Da, per reduced cystine, as described in more detail before [19].

### 3.4. Other PTM containing peptides

Additional PTMs which are occasionally encountered on frog peptides are *sulfo-Tyrosine*, *hydroxyproline* (*Hyp* or *HyPro*), and an aminoterminal *pyroglutamate* (*pGlu*). We here illustrate this with a few examples from our lab.

#### 3.4.1. Tyrosine sulfation

Low energy CID spectra of sulfated peptides are often dominated by neutral loss of the SO3 group (−79.9595 amu), leading to a relatively poor sequence information for identification. However, the specific neutral loss scan can be used to trigger an MS3 experiment (Fig. 6). In a data dependent neutral loss driven analysis of *P. burmeisteri* crude secretion, all doubly charged peptides that showed a loss of 39.9798 in terms of m/z in their MS2 spectrum, were used as a trigger for a second stage of fragmentation analysis on the daughter ion (MS3 spectra, Fig. 7). The chance of selecting sulfated peptides is increased by analyzing the same sample 4 times, each at a different mass range: from m/z 400 to 600, 600 to 800, 800 to 1000 and 1000 to 1200. This way more MS2/MS3 events are triggered when compared to the same analysis performed with 1 broad mass band (m/z 400–1500). An example of a sulfated peptide discovered using this MS-based method is a phyllokinin, which...
shows no backbone fragmentation in MS\(^2\), only a neutral loss of sulfate. From the MS\(^3\) it is possible to \textit{de novo} sequence this tyrosine sulfated peptide (Fig. 8). Due to the high resolving power and accuracy of the orbitrap analyzer, the system can be programmed so that only peptides losing a sulfate group \((m/z = 39.9798)\) trigger MS\(^3\) fragmentation (and for instance not potential losses of the quasi isobaric phosphate group), making it a highly specific screening strategy for this PTM.

3.4.2. Proline hydroxylation
As an example, two phyllokinins from \textit{P. burmeisteri} are sequenced with identical structure to the original peptide identified previously in \textit{Phyllomedusa hypochondrialis} [20] – RPPGFTFRRF and RPPGFSFRRF. Both peptides were also found with the third proline residue modified to Hyp, in addition to the C-terminal residue being modified to sulfo-tyrosine. The Hyp residue in phyllokinins could be distinguished from Leu/Ile using high resolution (7500) orbitrap analysis. The difference of \(y_8\) and \(y_9\)-ions is 113.0481 Da (Fig. 9), which matches the mass difference of Hyp and not of Ile/Leu (113.0841 Da). We remark here that the presence or absence of immonium ions (including secondary ones) in the tandem MS spectra can be of additional value to confirm a proposed/deduced novel sequence: Hyp is prone to yield a 70 Da immonium ion which is not evident from Ile/Leu fragmentation, etc.

3.4.3. N-terminal cyclization of Gln to pGlu
A last example is a peptide identified from \textit{P. burmeisteri} with a primary structure originally described from \textit{P. hypochondrialis} as “phypo Xa” [21], pEFRPSYQIPP. This peptide was found to contain an aminoterminal pyroglutamic acid residue.

4. Perspectives and conclusion
Although none of the latter PTMs (sulfo-Tyr, Hyp or pGlu) can be directly “spotted” on a 2-D peptide display, one could, in principle, envisage a 2-D differential display method to assist in the specific targeting of peptides carrying these PTMs. In line with our strategy to detect cystine-containing peptides we suggest to analyze the peptide containing sample \(i.e.,\) a crude amphibian secretion) twice, once without any treatment and once after either chemical, or enzymatic treatment to alter the PTM of interest. From the comparison of the 2-D displays, the peptides with altered RT and or \(m/z\) ratio are then detected as new “spots”. For pGlu-containing peptides, pyroglutamate aminopeptidase \(i.e.,\) from calf liver or from \textit{Pyrococcus furiosus}) could be employed to treat the crude venom. This enzyme is frequently used to deblock pGlu-peptides prior to Edman degradation as it removes pyroglutamic acid from a peptide/protein aminoterminus [22].
Similarly, sulfo-tyrosine containing peptides could be detected after enzymatic removal of the sulfate group [23]. In these examples, the treatment would result in a mass decrease of 111 or 80 Da for pGlu and sulfoTyr, respectively.

Similar approaches can also be developed to look for, in principle, any other PTM that can be specifically chemically or enzymatically altered or removed.

Finally, it shall be clear that the approaches detailed above are not restricted to amphibian secretions, but could equally be employed to screen venoms of other organisms for interesting biological peptides.

Conflict of interest
None declared.

Transparency document
The Transparency document associated with this article can be found in the online version.

References


