High-Efficiency Recognition and Identification of Disulfide Bonded Peptides in Rat Neuropeptidome Using Targeted Electron Transfer Dissociation Tandem Mass Spectrometry

Xi Yu,† Abbas Khani,‡ Xueting Ye,§ Filomena Petruzziello,‡ Huiyuan Gao,§ Xiaozhe Zhang,§,† and Gregor Rainer†

†Division of Biological Technology, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Zhongshan Road 457, Dalian, China
‡Visual Cognition Laboratory, Department of Medicine, University of Fribourg, Chemin de Musee 5, Fribourg, CH-1700, Switzerland
§Shenyang Pharmaceutical University, Wenhua Road 103, Shenyang, China

Supporting Information

ABSTRACT: The main goal of the present study is to develop a method to recognize and identify endogenous intrachain disulfide bonded peptide, which are rarely sequenced in current peptidomics studies. In order to achieve highly efficient detection of these peptides in a neuropeptidome analysis, we alkylated the peptides, mined the raw mass spectrometry data, and then recognized the candidates of untreated disulfide bonded peptides from unalkylated peptide extracts. After removing more than 90% features, targeted electron transfer dissociation fragmentation was performed for detecting and fragmenting disulfide bonded peptides, and even most of them were present in low abundance in the original sample. Diverse endogenous disulfide bonded peptides were then detected and sequenced, opening up new perspectives for comprehensively understanding the response of a neuropeptidome.

The vital role that disulfide bonded neuropeptides play in regulating brain function depends critically on their intact three-dimensional cyclic structure. These peptides have important functions in the mediation of physiological processes and behaviors such as food intake, memory, pain, stress, reward, and cognition. In spite of their critical importance, the identification of these peptides remains challenging due to their low abundance and also due to the high stability of their disulfide bonds, which inevitably results in low fragmentation efficiency when using mass spectrometric analysis. For biological applications, it is however highly in demand to simultaneously monitor different peptides, including disulfide bonded ones, in order to have a comprehensive understanding of the response of a neuropeptidome to stimulation, behavior, or drug application.

Diverse methods have been developed for the analysis of disulfide bonded peptides of different sources, including endogenous neuropeptides. Material scientists have developed different kinds of functionalized micrometer and submicrometer particles with selective affinity to a specific category of peptides. These materials can be employed as packing material for chromatography separation as well as adsorbents for selective solid phase extraction. For example, disulfide bonded peptides can be enriched using highly selective affinity chromatography or gold nanoparticles. For such methods, reduction of disulfide bridges during sample preparation using reagents such as dithiothreitol (DTT) is an inevitable procedure, which opens the cyclic disulfide bond and allows highly efficient fragmentation and identification of the peptides of interest. The peptides detected using such reduction-enrichment based methods are presented in reduced form, which means the crucial structural information is actually lost before identification. Additionally, as choosing extraction methods with specific selectivity, the nondisulfide bonded peptides are meant to be abandoned at the beginning for this batch identification. These facts consequently lead to difficulty in simultaneous monitoring of a global neuropeptidome.

The introduction of electron transfer dissociation (ETD)-mass spectrometry (MS) offers a promising alternative to traditional fragmentation techniques for disulfide bonded peptides. By preserving the post translational modifications (PTMs) to the greatest extent, it is considered to be an ideal tool for elucidation of sequences and identification of PTMs including disulfide bonding sites. Despite these advantages, ETD still suffers from a low fragmentation rate of precursor ions with weak intensity just like other fragmentation techniques. Therefore, only the top few precursor ions from primary mass (MS) fragmentation can be picked out for ETD
fragmentation in each running test. The identification of a large number of neuropeptides bearing the cysteine disulfide bond, in particular those of low abundance, thus remains a considerable technical challenge.

Here, we present a unique and highly efficient method which combines data mining with targeted ETD MS/MS detection for the highly efficient recognition and identification of intradisulfide bonded peptides in rat neuropeptidome. First, the classical reduction-derivatization reaction with DTT and iodoacetamide (IAM) was carried out to reduce and alkylate the disulfide bonded cysteines in the sample. Compared to the untreated sample, the disulfide bond containing peptides in the derivative sample gave a specific mass addition and increased retention time due to the alkylation. After nanoliquid chromatography (nanoLC)-MS analysis of the nonderivatized and the derivatized sample, a specific algorithm was developed to screen for primary mass fragment ion pairs with such characteristics. The potential candidates of intrachain disulfide bonded peptides were obtained for further targeted ETD MS/MS analysis of the untreated sample. By this way, disulfide bonded neuropeptides, including those low abundant ones, were recognized from untreated peptide samples and then fragmented for database-based sequencing. Through data mining using the raw data obtained from traditional CID fragmentation of alkylated sample, the present approach successfully developed a targeted technique which enhanced the strength of ETD in preferentially fragmenting and identifying the intrachain disulfide bond containing peptides in nonderivatized sample by offering a specific prescreened list of candidates with the highest possibility to possess intrachain disulfide bonds. The present approach greatly improved the identification rate of the intrachain disulfide bond containing peptides in the rat neuropeptidome. The identification of a large set of disulfide bonded peptides consequently extended the rat neuropeptidome for further biological applications.

**EXPERIMENTAL SECTION**

**Materials and Reagents.** LC–MS grade acetonitrile and formic acid were purchased from Sigma-Aldrich (China). Pure water was prepared using a Milli Q system. Siliconized microcentrifuge tubes (2 mL) were purchased from Eppendorf (Germany). Microcon centrifugal filter devices (Vivacon 500) were purchased from Sartorius (Germany). The self-packed C18 nanoLC column was packed with C18 (5 μm, 100 Å; Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) in capillary columns (75 μm i.d., New Objective). DL-DTT (99% purity) and IAM (98% purity) was obtained from Aladin (China). Melanin concentrating hormone was purchased from Botai Co. Ltd., Shanghai, China, and used directly after being dissolved in 20% acetonitrile for MS analysis.

**Animals.** Eight Sprague–Dawley rats (male, 8 weeks old) were used in the experiments. All animals were pair housed under constant temperature and humidity with free access to food and water.

**Extraction of Hypothalamic Peptides.** The rats were sacrificed by decapitation, and the brains were removed and the derivatized sample got a specific retention time due to the alkylation. After nanoliquid chromatography (nanoLC)-MS analysis of the nonderivatized and the derivatized sample, a specific algorithm was developed to screen for primary mass fragment ion pairs with such characteristics. The potential candidates of intrachain disulfide bonded peptides were obtained for further targeted ETD MS/MS analysis of the untreated sample. By this way, disulfide bonded neuropeptides, including those low abundant ones, were recognized from untreated peptide samples and then fragmented for database-based sequencing. Through data mining using the raw data obtained from traditional CID fragmentation of alkylated sample, the present approach successfully developed a targeted technique which enhanced the strength of ETD in preferentially fragmenting and identifying the intrachain disulfide bond containing peptides in nonderivatized sample by offering a specific prescreened list of candidates with the highest possibility to possess intrachain disulfide bonds. The present approach greatly improved the identification rate of the intrachain disulfide bond containing peptides in the rat neuropeptidome. The identification of a large set of disulfide bonded peptides consequently extended the rat neuropeptidome for further biological applications.

**NanoLC–MS Analysis of Nonderivatized and Derivatized Peptide Extracts.** Prior to analysis, 50 μL of 15% acetonitrile containing 0.2% formic acid were used for reconstituting the aqueous and organic extracts, respectively. LC–MS experiments were performed on a Linear Trap Quadrupole (LTQ)-orbitrap Elite system with an ETD mass spectrometer. An Ultimate 3000 nanoLC and a self-packed C18 nanoLC column were coupled online to the mass spectrometer through a nanospray ion source. Mobile phase A was 0.2% formic acid in water and mobile phase B was 0.2% formic acid in acetonitrile. The gradient consisted of the following: (i) 0 min at 0% B for sample loading, (ii) linear from 0% to 40% B over 40 min, (iii) linear from 40% to 80% B over 10 min, and finally (iv) isocratic at 80% B for 10 min. The flow rate of the column was maintained at 200 nL min⁻¹. Data acquisition on the LTQ-Orbitrap Elite system consisted of a full Fourier transform mass spectrometry (FTMS) scan event at the mass range of 350–2000 m/z. The lock mass (445.12002 z from polydimethylcyclosiloxane) was used for real time internal recalibration. The LC–FTMS analysis of each rat dorsal striatum (DS) sample was repeated twice.

**Data Mining for Recognition of Disulfide Bond-Containing Peptides.** The raw LC–FTMS spectra data were subjected to Sieve 2.1 (Thermo) for time alignment in order to reduce the effect of chromatographic variability between samples. Relative alignment of the spectra requires at least two set of data with one of them as the reference. The parameters used were as follows. Algorithm: frame; RT width = 6 min; intensity threshold = 100 000; frame = 50 000. In total, 50 000 frames were obtained after the alignment for further data screening of disulfide bond-containing peptides. The data screening was performed by eliminating invalid MS/MS information, deconvoluting isotope peaks, and retaining MS features with qualified signal intensity higher than 100 000. First, the data underwent a preliminary screening to get rid of...
invalid frames with zero charge state. After that, the peptides were screened out by searching for acetylation in both types of samples using a Matlab program of specific algorithm, followed by a deduction procedure. This self-developed program was first performed to execute a search between the raw data of the nonderivatized and the derivatized sample for ion pairs with specific mass difference of 2n × 58.0293 (n represents for the number of disulfide bonds in the peptide), which ended up with an output list of eligible ion pairs found in the two separate raw data file (L1). After that, the same search was carried out within the raw data of the nonderivatized sample itself to screen for fragments from naturally alkylated peptides, which ended up with another output list (L2). By deducting the duplicates in the two lists from L1, only the acetyl amination caused by the addition of DTT and IAM was listed.

**Targeted LC–MS/MS Analysis for Peptide Identification.** After data mining, the m/z and retention time of the peptide candidates were added to an inclusion list for further targeted nanoLC–ETD-MS/MS analysis, during which these peptides in the DS samples were selected for fragmentation. ETD in combination with Fourier transformation (FT) detection was used to fragment and analyze peptides. Additionally, both ETD in combination with ion trap (IT) detection and HCD were applied for the nonderivatized peptide extracts to increase the capability of detecting peptides of low charge states or low abundance. However, according to the data obtained, the identifications actually did not improve substantially by this additional run since peptides that are disulfide bonded in the sample are highly charged in the gas phase and consequently result in higher charged fragment ions. Therefore, it was merely used as a double check to guarantee the preciseness of our investigative study. Supplementary activation was used in ETD mode. ETD activation time was set at 75 ms, and maximum accumulation time was 300 ms. Dynamic exclusion was set as a repeat count of 2, an exclusion duration of 30 s, and a repeat duration of 30 s. A 25 ppm mass tolerance was set for precursor selection with disabled preview mode for FT mass scan and disabled monoisotopic precursor selection mode. Finally, all the peptides identified using targeted nanoLC-FT ETD-MS/MS in this step were loaded back to Sieve 2.1 for identification of the significantly changed peptides.

**Database Search and Spectral Interpretation.** All the raw nanoLC–FT–ETD-MS/MS data were subjected to Peaks Studio 7.0 (BSI, Canada) for spectral interpretation. The Data Refinement program allows correcting of the parent mass and charge states to provide accurate monoisotopic mass of a peptide. The scans of quality value >0.3 were kept for further sequence analysis. Data processing, including peak centroiding, charge deconvolution, and deisotoping, was conducted for data refinement.

The refined data were subjected to Auto De Novo program for sequencing with the mass tolerance of parent ions and product ions set at 10 ppm and 1 Da, respectively. No enzyme was specified for cleavage. Variable PTMs, including, amidation (C-terminal), acetylation (N-terminal), pyroglutamalytion from glutamic acid, and glutamine (N-terminal) were selected in database search. Database search was conducted on a customized neuropeptide database Neuropeptide_Enargile and Sweepep. Estimation of false positives was conducted by searching all spectra against decoy databases. The cutoff of −10lgP for peptide identification in Peaks search was >15. To confirm sequences identified from the database search, all the search results were further subjected to manual inspection.

## RESULTS AND DISCUSSION

**Inspection of Disulfide Bonded Peptides Using Untargeted LC–FT-MS/MS Analysis.** Our initial analysis of HT extracts using FT ETD-MS/MS was in agreement with the previous reports,

which failed to find a sufficient number of disulfide bonded peptides. An analysis on the precursor ions of rat HT samples suggested that the failure might be caused by the presence of a vast number of MS features (~40 000 features/sample) alongside with those related to the disulfide bonded peptides. In a typical data set, the number of MS features can reach as many as 100 features s⁻¹, thus challenging the fragmentation speed of a mass spectrometer, such as the LTQ-Orbitrap. Previous studies have shown that independent of fragmentation technique, the peptide spectral matches (PSMs) are largely determined by the speed of fragmentation and abundance of peptides,

suggests that most of the low-abundance peptides are subject to a low chance of fragmentation,

consequently leading to tandem MS spectra with poor quality and an unsatisfactory identification rate.

**Workflow for Mining the MS Data.** To efficiently detect disulfide bonded peptides, we developed this data mining strategy to recognize disulfide bonded peptide candidates from brain extracts (Scheme 1). We first extracted all peptides from a brain sample and then loaded them using a MOC method to increase the peptide extraction coverage.

One group of peptide extracts were then submitted to reduction and alkylation using DTT and IAM, respectively, as is commonly done in bottom-up proteomics applications.

As verified using peptide Melanin-concentrating hormone (MCH, DFDMLFCLGLGRVYRRPCWQV), which contains an intrapeptide disulfide bond, the alkylation rendered disulfide bonded peptides with increasing molecular mass and hydrophobicity but without changing their charge states (Figure 1). This fact was also confirmed when checking other peptides such as somatostatin 14 and somatostatin 28. Our strategy thus fully
exploits the three criteria including the mass increase due to alkylation, retention time increase due to polarity change, and stable charge state to mine the vast number of MS1 features from the raw data to reliably recognize candidates of untreated disulfide bonded peptides.

Recognition of Disulfide Bonded Peptides Following Data Mining. The data mining was initialized with targeting the MS1 features corresponding to the newly formed peptides in the alkylated samples. Compared to the nonderivatized peptide extracts, a number of new MS1 features appear in the derivatized peptide extracts because of the addition of reagents, the generation of alkylated peptides, and the unavoidable side products. This allowed us to compare the MS1 features between unreduced and reduced samples and then extract only the features that were present exclusively in the alkylated samples. The mass error window was set at 10 ppm in the mining procedure. The MS1 features with charge state 1 were removed to reduce the interference from reaction reagents, since peptides rarely exhibit charge state 1. This step allowed reducing the number of disulfide bonded peptide candidates from 47,517 to 29,274 for a set of peptide extracts (bars 2 and 4 in Figure 2, respectively) and thus reducing the size of data to be processed significantly. These peptides candidates were however observed in the alkylated state so that they were unsuited for examination of endogenous peptides.

In order to screen for the endogenous disulfide bonded peptide candidates, we computed the mass differences between the new MS1 features in the alkylated peptide extracts and the
corresponding MS^1 features in the untreated peptide extracts. According to the formula presented in Scheme 1, alkylation leads to an increase of 58.0293 amu \times 2n in the mass of peptides bearing n double disulfide bonds (n was set to 1, 2, and 3). The application of mass difference reduces the number of peptide candidates from 29 274 to 19 117 (bars 4 and 5 in Figure 2, respectively). At that, additional ways were used to restrict the number of disulfide bonded peptide candidates including the exclusion method using charge state and retention time (bars 6 and 7 in Figure 2, respectively). As the alkylation occurred on cysteines, which are not residues favored for addition of protons, it is expected to have negligible influence on the charge behavior of disulfide bonded peptides. Therefore, derivatized and corresponding nonderivatized peptides with different charge states were considered spurious and removed from the candidate list. Finally, the increase in the hydrophobicity of alkylated peptides was considered in the further data filtering to remove the peptide candidates that had larger retention time than its alkylated peptides. After the application of these data mining methods, the number of nonderivatized disulfide bonded peptide candidates for peptides bearing 1 disulfide bond was as few as 2998. The same data mining process was carried out to screen for peptides candidates bearing double and triple disulfide bonds as well, with final numbers of candidates obtained were 2572 and 2411, respectively. After significant reduction in the number of peptides, it would allow increased efficiency in the fragmentation of low abundance ions for a LTQ-Orbitrap Elite system.  

Identification of Disulfide Bonded Peptides Using Targeted LC–MS/MS Analysis. Having recognized the disulfide bonded peptide candidates, inclusion lists of precursor ions were loaded to perform targeted nanoLC–FT ETD-MS/MS fragmentation separately for peptides bearing 1, 2, and 3 disulfide bonds simultaneously in a single run. As only selected ions are for fragmentation, MS/MS parameters, including signal-to-noise ratio (n > 10) and maximum accumulation time of fragmentation (300 ms) were optimized to enhance the capability of generation of high-quality tandem MS spectra. Taken together, we unambiguously identified 43 disulfide bonded peptides (Table S1, Supporting Information), most of which are never reported in previous neuropeptidomics studies. These peptides include not only peptides bearing a single disulfide bond but also peptides bearing 2 or 3 disulfide bonds, such as orexin-A and CART (62-102). In addition to the known neuropeptides, as many as 35 novel disulfide bonded peptides were identified, which might be potential bioactive peptides in the brain. Figure 3 shows the identification of a representative peptide IPIYEKKGQVPMCDAGEQCVRKGAIRGKLCDCPRTSCNSFLLKCL, of which the tandem MS spectrum was dominated by c and z ions for high-confidence identification of disulfide bonded peptides.

Despite the efficiency in the data mining for finding the candidates of disulfide bonded peptides, the gap between the number of identified and suspected cysteine peptides is still presented. While the candidates can be any kind of cysteine containing peptides, neuropeptides are the sole targets of our study, in which the database search was performed singly for neuropeptides. This specification would in turn reduce the number of peptides finally identified. Additionally, the number of candidate peptides is heavily dependent on the accuracy of the MS instrument. We found there is almost one order magnitude of reduction in the number of candidate peptides if narrowing the mass tolerance window from 10 to 1 ppm. Although the Orbitrap Elite was well calibrated, the mass error of some peptides can be \pm 5 ppm. The mass tolerance window used in data mining in turn has to set as \pm 10 ppm to cover these peptides. Theoretically, the data mining efficiency can be remarkably improved if it is performed on a mass spectrometer of higher accuracy. This fact reflects the space of improvement of data mining strategy in combination with the use of a mass spectrometer of higher performance.

CONCLUSIONS

Overall, our results demonstrate that data mining, in conjunction with a specified chemical reaction, can allow the detection of peptides that are undetectable using conventional LC–MS methodology. A specified reaction can modify mass, hydrophobicity, charge state, or other properties of peptides and these modifications can be used to establish an inclusion list for targeted nanoLC–FT ETD-MS/MS analysis on samples that have not themselves undergone the chemical reaction. On the basis of the inclusion list, mass spectrometer parameters can be optimized, thus greatly improving the identification of the targeted peptides, particularly when they are present at low abundance. Even though current instruments are increasingly powerful for facilitating peptidomics and proteomics research,
the high complexity and large dynamic range of biological samples still stands in the way. Data mining for peptide recognition might be the road less traveled in solving the problem of monitoring additional classes of bioactive molecules that have hitherto been undetectable using conventional approaches and thus offer us a bigger chance to find out more.

**REFERENCES**