

Identification of endogenous peptides in maize

Haoqiang Yin[†], Yuzheng Li[†], Jiaqi Zhao[†], Shanshan Liu, Anran Li, Liuqi Yang, Tao Liu, Mingchun Tian, Xinyang Li, Shunxi Wang^{*}, Lei Tian^{*}

College of Agronomy, Henan Agricultural University, Zhengzhou 450046, China

^{*} **Corresponding author:** Shunxi Wang (Email: angshunxi2008@163.com); Lei Tian (Email: tianney@126.com)

[†]These authors contributed equally to this work

Abstract: The extraction and identification of endogenous peptides from plants are relatively unexplored because of the degradation of peptides after extraction from plant tissues. In this study, we developed an optimized sample preparation protocol that combined constant water heating with plant protease inhibitors and TCA-acetone precipitation, resulting in the effective extraction of endogenous peptide from plants, such as maize, while diminishing unspecific protease activity. The results showed that a total of 2867 endogenous peptides were identified in three maize samples using this method, of which 2119 (73.9%) peptides were commonly identified in all three samples. The length and molecular weight of these peptides ranged from 8 to 25 amino acids and from 729.44 to 2980.62 Da, respectively, with 96.4% of the peptide scores greater than 40. These results indicated that our extraction method is highly reproducible, precise, and wide-ranging, providing optimized information for large-scale identification of plant endogenous peptides.

Keywords: Plant; Endogenous peptides; Maize; Heating; Protease inhibitors; TCA-acetone.

1. Introduction

Peptides, which are the smallest biological molecules in the protein group, typically consist of 2-100 amino acids (Tavormina et al., 2015). Endogenous peptides play important biological functions in various life activities of organisms, including growth and development (Bi et al., 2017), transcriptional regulation (Hinnebusch et al., 2016), stress adaptation (Jackson et al., 2018). Moreover, they hold significant value in development and application (Tavormina et al., 2015). With a wide variety of land plant species available, the cost-effective utilization of plant peptide resources to develop high-value peptide products can yield substantial economic and social benefits.

Peptidomics, a branch of proteomics that has gained prominence in recent years, primarily focused on the study of animal neuropeptides and proteases. However, in the field of plant science, peptidomics is a relatively new field that enables the large-scale identification of endogenous peptides from tissue samples (Schulz-Knappe et al., 2005). Plant endogenous peptides are a class of protein molecules that can be essentially divided into two types: bioactive peptides produced by the selective action of peptidases on larger precursor proteins, and degradation peptides generated by protein hydrolases in the protein turnover process (Rubinsztein et al., 2006). Given the diverse nature of peptides in biological systems and their critical regulatory roles, there is an increasing demand for the discovery of more endogenous peptides. In animals, peptide identification methods have been used to discover new bioactive peptides and elucidate their functions, particularly neuropeptides (Falth et al., 2007; Secher et al., 2016; Tharakan et al., 2019). However, the application of peptidomics is highly challenging due to the presence of non-specific protease digestion during sample preparation (Secher et al., 2016; Finoulst et al., 2011).

Therefore, new sample preparation methods are necessary to identify reliable endogenous peptides. In addition to the

commonly used acid extraction method, heat-stable methods or methods combining heat stability with protease inhibitors can effectively prevent the production of protein hydrolysis peptide fragments in animal tissues (Secher et al., 2016; Che et al., 2005). Although there has been rapid progress in the extraction of peptides from animal tissues (Finoulst et al., 2011), currently, only two methods are available for isolating peptides from plant tissues: acid extraction and acid extraction with a mixture of plant protease inhibitors (Fesenko et al., 2015; Ye et al., 2016; Chen et al., 2014). Therefore, optimizing the sample preparation method for extracting endogenous peptides from plant tissues is necessary.

China, being one of leading producers and consumers of corn, has the second highest corn yield in the world. Corn is not only an important economic crop, but also a model species in biological research. To address the need for large-scale identification of plant endogenous peptides, this study developed a method for extracting plant endogenous peptides using corn as a model plant. The method employed a combination of heating, plant protease inhibitors, and TCA-acetone, which identified a total of 2867 endogenous peptides in three corn samples. Among them, 2119 peptides (73.9%) were common to all three samples, and they varied widely in terms of lengths, molecular weights, and high peptide scores. These results indicate that the plant endogenous peptide extraction method we developed is highly reproducible, accurate, and provides broad coverage, thereby laying the foundation for large-scale identification of plant endogenous peptides.

2. Materials and Methods

2.1. Plant materials

The B73 maize seeds used in this study were obtained from our laboratory. Seeds with uniform grains were selected from the same ear and were sown in the greenhouse under controlled conditions of 28°C and 15 hours of light followed

by 25°C and 9 hours of darkness. When the maize plants had reached the three-leaf stage, the third leaf was quickly cut off and immediately frozen in liquid nitrogen. This process was repeated three times, and all samples were stored at -80°C for subsequent experiments.

2.2. Peptide extraction

The maize leaves were first ground into a fine powder in liquid nitrogen. Approximately 2 g of powder was transferred to a 50 ml centrifuge tube and heated in a 95°C water bath for 5 minutes. Next, 10 ml of 10% TCA-acetone was added, and the mixture was precipitated at -20°C for 5 hours. Afterward, the mixture was centrifuged at 12000 rpm for 20 minutes at 4°C, and the supernatant was discarded. The precipitate was then washed with 10 ml of pre-cooled acetone at -20°C until it turned colorless. The precipitate was then dissolved in 10 ml of 1% TFA solution containing a plant proteinase inhibitor at a ratio of 250:1 and incubated at 4°C for 1 hour. The mixture was sonicated 5 times on ice (40w, 6 seconds each time with an 8-second interval) and centrifuged at 12000 rpm for 20 minutes at 4°C. The supernatant was transferred to a 30 kDa ultrafiltration tube and centrifuged at 12000 rpm for 20 minutes at 4°C. The peptides were desalted using C18 cartridges (Empore SPE Cartridges C18, 7 mm inner diameter, 3 mL volumes, Sigma), and then dried using a vacuum concentrator. The peptide segments were re-dissolved in 40 µl of 0.1% TFA for subsequent LCMS/MS mass spectrometry detection.

2.3. Liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis

2 µg of peptide samples were taken based on quantitative results and separated using a nanoflow HPLC liquid phase system EASY-nLC1000. The Thermo EASY column SC200 (150 µm × 100 mm RP-C18) was balanced with 100% of liquid phase A, which consisted of a 0.1% formic acid acetonitrile water solution (with 2% acetonitrile), while liquid phase B was a 0.1% formic acid acetonitrile water solution (with 84% acetonitrile). The sample was injected by an autosampler into Thermo EASY column SC001 traps (150 µm × 20 mm RP-C18) before separation by the chromatographic column at a flow rate of 300 nl min⁻¹. The relevant liquid phase gradient used was as follows: 0-115 min, linear gradient of B liquid from 0% to 45%; 115-117 min, linear gradient of B liquid from 45% to 100%; 117-120 min, B liquid was maintained at 100%. The enzymatic hydrolysate was separated by capillary high-performance liquid chromatography and analyzed by Q-Exactive mass spectrometer (Thermo Finnigan) in positive ion detection mode with a parent ion scan mass range of 300-1800 m/z-1. The resolution of MS1 was set to 70000 at M/Z 200, AGC target was 3e6, and the maximum IT of the first level was 10 ms with a single scan range. 20 fragment spectra were collected after full scan, with MS/MS activation type set to HCD, isolation window of 2 m/z, microscans of 1, maximum IT of the second level of 60 ms, normalized collision energy of 30 eV, and an underfill ratio of 0.1%. The LC-MS/MS analysis was conducted over a duration of 120 min.

2.4. Peptide identification and quantification by MaxQuant

The result MS files were processed using the MaxQuant software (version 1.5.5.1). The MGF files obtained were

searched against the UniProt Zea mays database (containing 132,354 sequences, downloaded on May 02, 2018) without specifying any enzyme cleavage rules. The search parameters were set as follows: peptide mass tolerance of ± 20 ppm, MS/MS tolerance of 0.1 Da, and a maximum of 2 missed cleavages allowed (with an allowance for 2 missed cleavages). The variable modifications included oxidation (M) and acetylation (Protein N-term). MaxQuant software was used for label-free peptide quantification based on extracted ion chromatograms and spectral counts, and validation. The cutoff value for global false discovery rate (FDR) for peptide identification was set at 0.01.

2.5. Bioinformatics analysis

The heatmap plot was generated using an R package, while the Venn plot was created through the website (<https://bioinfogp.cnb.csic.es/tools/venny/index.html>). All other plots were produced using GraphPad Prism V.8.0.2.

3. Results

3.1. Development of plant peptidomics protocols

To thoroughly investigate endogenous peptides in maize, we developed a method for extracting plant endogenous peptides (Figure 1). Firstly, maize leaves were ground into a powder using liquid nitrogen, and the powdered samples were then placed in a centrifuge tube. The sample was heated at 95°C for 5 minutes in a water bath, and then precipitated with TCA-acetone for 5 hours. The resulting precipitate was washed three times with pre-chilled acetone solution until colorless, and then resuspended in a 1% TFA solution. To reduce non-specific protease degradation, a plant protease inhibitor was added to the solution. The mixture was then incubated at 4°C for 1 hour, followed by sonication on ice for 5 minutes. The resulting peptide mixture was enriched for plant endogenous peptides using a 30 kDa ultrafiltration tube, desalted and dried. The sample was resuspended in a 0.1% TFA solution and analyzed by high-throughput LC-MS/MS. The mass spectrometry data was searched against the Ensembl protein database, and plant endogenous peptides were identified following screening.

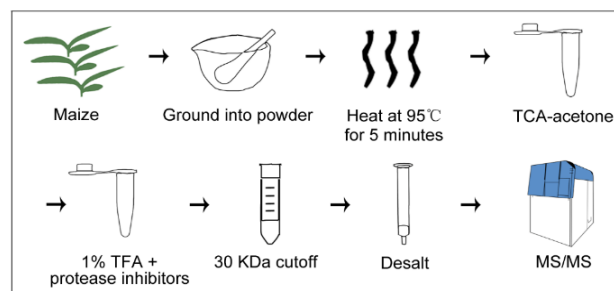


Fig. 1 The workflow for plant endogenous peptide

3.2. Large-scale identification of endogenous peptides

After performing statistical analysis on the identified endogenous peptides in the three groups of samples, a total of 2867 endogenous peptides were found, with 2536 identified in sample 1, 2539 in sample 2, and 2531 in sample 3 (Table 1). Upon further analysis using a Venn diagram, it was observed that 247 (8.6%) endogenous peptides were

exclusive to one sample, 501 (17.5%) were common to two samples, and 2119 (73.9%) were identified in all three samples (Figure 2).

Table 1. Number of peptides in three groups

	Sample 1	Sample 2	Sample 3
Number of repeated peptides	2536	2539	2531
The total number of peptides	2867		

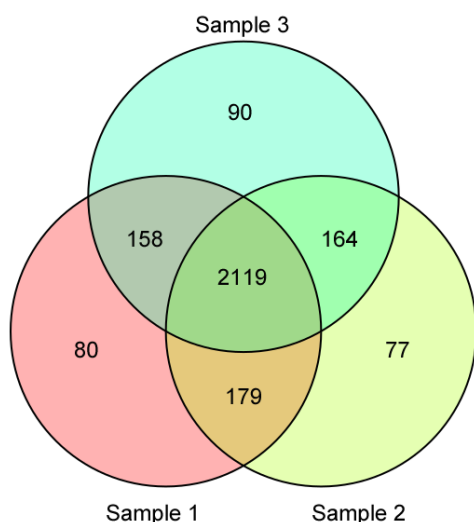


Fig. 2 Venn diagrams of peptide identified in three samples

In addition, we found the correlation between two samples ranged from 0.981 to 0.984 (Figure 3), indicating that the endogenous peptide extraction method employed in this study performed excellent reproducibility.

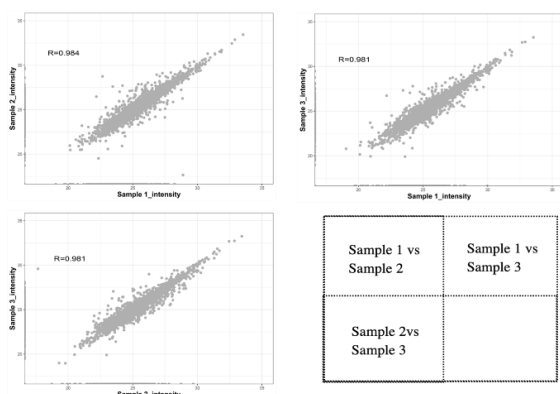


Fig. 3 Correlation among the three samples. R2 values are indicated.

To discover the distribution pattern of peptide length in plant endogenous peptides, we analyzed the peptide length of 2119 identified endogenous peptides. The results showed that the length of endogenous peptides ranged from 8 to 25 amino acids, with an average length of 16 amino acids. Among them, endogenous peptides with a length of 14 amino acids were the most abundant, and 65.5% (1387) of peptides fell within the range of 11 to 18 amino acids (Figure 4).

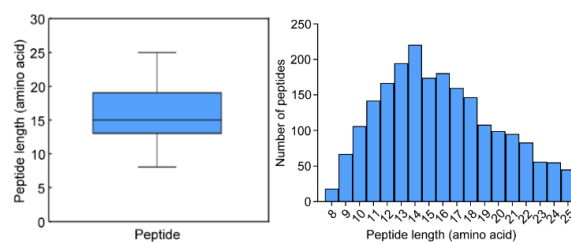


Fig. 4 Length distribution of peptides

The analysis of molecular weight of the 2119 identified endogenous peptides revealed that their range was from 729.44 to 2980.62 Da, with all peptides having a molecular weight less than 3000 Da. Furthermore, 86.6% of the peptide had a molecular weight between 1000-2300 Da, and the average molecular weight was 1698.25 Da (Figure 5).

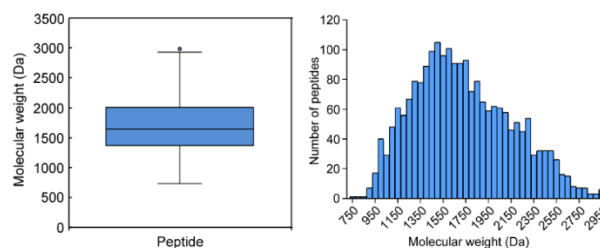


Fig. 5 Molecular weight distribution of endogenous peptides

The observed distributions of peptide length and molecular weight indicate that the identified endogenous peptides have a broad range, providing further evidence for the reliability of the extraction method.

Subsequently, we analyzed the distribution of peptide segment scores for the 2119 endogenous peptides identified in all three samples, and found that 2042 (96.4%) had scores >40, with 75% of the peptides having scores above 65. Additionally, 888 (41.9%) peptides had scores >100, and 95 (4.5%) peptides had scores >200 (Figure 6). These results demonstrate that the endogenous peptides extracted using this method possessed good and stable quality.

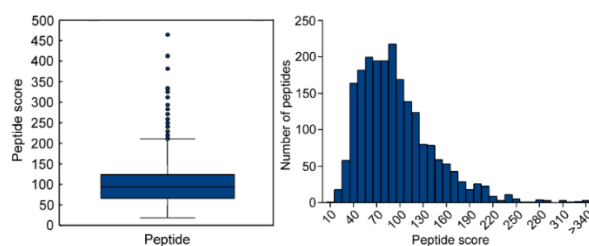


Fig. 6 Distribution of peptide score identified by the developed pipeline

4. Discussion

Endogenous peptides play crucial roles in regulating many physiological processes in multicellular organisms, including immune response, hormones, neurotransmitters, and signaling functions (Butenko et al., 2009; Fesenko et al., 2015; Hokfelt et al., 2003; Lightfoot et al., 2019; Montowska et al., 2015; Takahashi et al., 2018). Compared to the large number of peptides discovered and identified in animals, only a few peptides have been found and identified in plants. In this study, we implemented pre-treatment steps for the peptides prior to extraction, and the extraction and identification method used demonstrated high repeatability and resulted in good and stable peptide quality.

Protease inhibitors are commonly used during peptide

extraction to inhibit the activity of peptide-degrading enzymes (Finoulst et al., 2011). However, natural hydrolysis of proteins can increase the complexity of data and is unfavorable for the study of endogenous peptides. Studies on humans and animals have shown that protease inhibitors alone are insufficient to prevent the degradation of endogenous peptides (Beaudry et al., 2010). Heating, such as through the use of a constant-temperature water bath, can effectively deactivate protein hydrolysis activity in samples by denaturing animal protein hydrolysis enzymes. While microwave heating has been used in animal peptidomics, it has poor controllability (Finoulst et al., 2011).

The low protein content in plant tissue cells, as well as the presence of proteases and interfering compounds such as phenols, pigments, lipids, and nucleic acids (Zhang, 2018; Damerval et al., 1988), makes isolating endogenous peptides from plants more challenging than from animal cells. To precipitate and denature proteins, the TCA-mixed cold acetone solution can be used. Wu et al. (1984) found that TCA-acetone precipitation can effectively inhibit protease activity in plant tissues and remove interfering compounds. In this study, we used a constant-temperature water bath, plant protease inhibitors, and the TCA-acetone precipitation method to minimize non-specific proteolysis during peptide extraction. Therefore, this step helps to reduce interference from non-protein or non-peptide compounds during the extraction of endogenous peptides.

Endogenous peptides are considered as significant regulators in physiological and pathological processes. However, non-specific degradation related to proteases during peptide extraction remains a long-standing issue, and developing more effective peptide extraction protocols that maintain the endogenous peptides in the same state as in vivo for proteomic research is crucial. The irregular cleavage sites of endogenous peptides considerably increase the search space for peptide identification, making database searching time-consuming and accuracy reduced. Retrieval databases and algorithms suitable for endogenous peptides are urgently needed to quickly and accurately identify them. Exploring functional endogenous peptides is of great biological significance, as they play a critical role in regulating many physiological functions. It is anticipated that the development of methodology and technology will provide multi-peptidomics with sufficient qualitative and quantitative information, enabling the study of the physiological status, immune defense, and evolutionary processes of organisms with strong evidence.

In summary, our novel method for extracting endogenous peptides from plants involves a combination of constant temperature water bath heating, plant protease inhibitor, and TCA-acetone precipitation to pretreat samples, which enabled the identification of a large number of endogenous peptides. Compared to previous methods, this approach has provided a solid foundation for discovering new functional molecules in plant endogenous peptides.

References

- [1] Beaudry, F. Stability comparison between sample preparation procedures for mass spectrometry-based targeted or shotgun peptidomic analysis. *Analytical biochemistry* 407, 290-292, doi:10.1016/j.ab.2010.08.017 (2010).
- [2] Bi P, Ramirez-Martinez A, Li H, et al. Control of muscle formation by the fusogenic micropeptide myomixer. *Science* 356(6335): 323-327, doi: 10.1126/science.aam9361 (2017).
- [3] Butenko, M. A., Vie, A. K., Brembu, T., Aalen, R. B. & Bones, A. M. Plant peptides in signalling: looking for new partners. *Trends in plant science* 14, 255-263, doi:10.1016 / j.tplants.2009.02.002 (2009).
- [4] Che, F. Y., Lim, J., Pan, H., Biswas, R. & Fricker, L. D. Quantitative neuropeptidomics of microwave-irradiated mouse brain and pituitary. *Molecular & cellular proteomics : MCP* 4, 1391-1405, doi:10.1074 / mcp.T500010-MCP200 (2005).
- [5] Chen, Y. L. et al. Quantitative peptidomics study reveals that a wound-induced peptide from PR-1 regulates immune signaling in tomato. *The Plant cell* 26, 4135-4148, doi:10.1105 / tpc.114.131185 (2014).
- [6] Damerval, C., Zivy, M., Granier, F., and de Vienne, D. (1988) Two-dimensional electrophoresis in plant biology, in *Advances in electrophoresis* (Chrambach, A., Dunn, M., and Radola, B., eds.), VCH, Weinheim, New York, pp. 265–340.
- [7] Falth, M. et al. Neuropeptidomics strategies for specific and sensitive identification of endogenous peptides. *Molecular & cellular proteomics: MCP* 6, 1188-1197, doi:10.1074 / mcp.M700016-MCP200 (2007).
- [8] Fesenko, I. A. et al. Specific pools of endogenous peptides are present in gametophore, protonema, and protoplast cells of the moss *Physcomitrella patens*. *BMC plant biology* 15, 87, doi:10.1186 / s12870-015-0468-7 (2015).
- [9] Finoulst, I., Pinkse, M., Van Dongen, W. & Verhaert, P. Sample preparation techniques for the untargeted LC-MS-based discovery of peptides in complex biological matrices. *Journal of biomedicine & biotechnology*, 245291, doi:10.1155 / 2011/245291 (2011).
- [10] Hinnebusch AG, Ivanov IP and Sonenberg N. Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science* 352(6292): 1413-1416, doi: 10.1126/science.aad9868 (2016).
- [11] Hokfelt, T., Bartfai, T. & Bloom, F. Neuropeptides: opportunities for drug discovery. *The Lancet. Neurology* 2, 463-472, doi: 10.1016/s1474-4422(03)00482-4 (2003).
- [12] Jackson R, Kroehling L, Khitun A, et al. The translation of endogenous open reading frames controls mucosal immunity. *Nature* 564(7736): 434-438, doi: 10.1038/s41586-018-0794-7 (2018).
- [13] Lightfoot, J. W. et al. Small peptide-mediated self-recognition prevents cannibalism in predatory nematodes. *Science* 364, 86-89, doi:10.1126 / science.aav9856 (2019).
- [14] Montowska, M., Alexander, M. R., Tucker, G. A. & Barrett, D. A. Authentication of processed meat products by peptidomic analysis using rapid ambient mass spectrometry. *Food chemistry* 187, 297-304, doi:10.1016/j.foodchem.2015.04.078 (2015).
- [15] Peng J, Zhang H, Niu H, et al. Peptidomic analyses: The progress in enrichment and identification of endogenous peptides. *TrAC trends in analytical chemistry*, 125:115835, doi: 10.1016/j.trac.2020.115835 (2020).
- [16] Rubinsztein D C. The roles of intracellular protein-degradation pathways in neurodegeneration. *Nature*, 443:780-786, doi: 10.1038/nature05291 (2006).
- [17] Schulz-Knappe, P., Schrader, M. & Zucht, H. D. The peptidomics concept. *Combinatorial chemistry & high throughput screening* 8, 697-704, doi: 10.2174 / 138620705774962418 (2005).
- [18] Secher, A. et al. Analytic framework for peptidomics applied to large-scale neuropeptide identification. *Nature communications* 7, 11436, doi:10.1038 / ncomms11436 (2016).

- [19] Svensson, M., Skold, K., Svenningsson, P. & Andren, P. E. Peptidomics-based discovery of novel neuropeptides. *Journal of proteome research* 2, 213-219 (2003).
- [20] Takahashi, F. et al. A small peptide modulates stomatal control via abscisic acid in long-distance signalling. *Nature* 556, 235-238, doi:10.1038/s41586-018-0009-2 (2018).
- [21] Tavormina P, Coninck B D, Nikonorova N, et al. The plant peptidome: an expanding repertoire of structural features and biological functions. *The Plant cell*, 27(8), doi: 10.1105 / tpc.15.00440 (2015).
- [22] Tharakan, R. et al. A methodology for discovering novel brain-relevant peptides: Combination of ribosome profiling and peptidomics. *Neuroscience research*, doi: 10.1016 / j.neures.2019.02.006 (2019).
- [23] Wu, F. and Wang, M. Extraction of proteins for sodium dodecyl sulfate polyacrylamide gel electrophoresis from protease-rich plant tissues. *Analytical biochemistry* 139, 100–103. doi: 10.1016/0003-2697(84)90394-4 (1984).
- [24] Ye, X. et al. Extensive characterization of peptides from *Panax ginseng* C. A. Meyer using mass spectrometric approach. *Proteomics* 16, 2788-2791, doi: 10.1002 / pmic.201600183 (2016).