CHEMICAL ENGINEERING TRANSACTIONS
VOL. 111, 2024
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.cetjournal.it
Guest Editors: Valerio Cozzani, Bruno Fabiano, Genserik Reniers

ISBN 979-12-81206-11-3; **ISSN** 2283-9216
DOI: 10.3303/CET24111119

Characterization of Peptides and Proteins Associated with Bacterial Proliferation and Bird's Nest Sample Matrix

Cao Ngoc Minh Trang[a*], Nguyen Khac Manh[b], Trenh Kim Lan[b], Vo Nguyen Dang Khoa[c]

[a]Faculty of Applied Technology, School of Technology, Van Lang University, Ho Chi Minh City, Vietnam
[b]University of Sciences, VNU-HCM, Ho Chi Minh City, Vietnam
[c]Institute of Applied Materials Science, Vietnam Academy of Science and Technology
trang.cnm@vlu.edu.vn

Bird's nest sample matrix is renowned for its nutritional and therapeutic properties, primarily attributed to its protein content. Despite its significance, research on its proteins has been limited, primarily hindered by challenges in extraction, isolation, and identification. The effectiveness of Biotool software and the Mascot protein database in recognizing insulin from LC-MS analysis data has been confirmed. The software adeptly matched the MS data with the cataloged proteins in the database, highlighting its capability in protein identification. The analysis produced 26 peptide fragments, primarily with charges of +2, +3, and +4, and a mass range of 1000-2000 Da. This first study endeavors to fill this gap by offering a thorough analysis of its peptides, aiming to enhance the quality assessment and provide valuable insights for future investigations into proteins analysis.

1. Introduction

Proteins are essential molecules in living organisms that play vital roles in catalysis, structural support, and biological regulation. They are made up of amino acids and can fold into intricate three-dimensional structures. Recent advancements in analysis tools, such as ionization (Whitelegge J.,2009) techniques like MALDI and ESI, coupled with mass separators and ion traps, allow for better characterization. Protein identification involves breaking down proteins into smaller peptide fragments, analyzing their masses, searching for matching sequences in databases, and identifying proteins present in the sample.

- Proteins and peptides are broken down into smaller peptide fragments using specific enzymes. Liquid Chromatography (LC) system separates these small fragments, which are then analyzed using Mass Spectrometry (MS) or Tandem Mass Spectrometry (MS/MS) to produce data.
- The identification of proteins is based on matching molecular masses of peptides against protein databases. This method is fast but requires the proteins to be of high purity and have their full sequences in the database.
- In the MS/MS method, peptide ion fragmentation occurs via collisional dissociation in a CID impact chamber by inert gas molecules such as N_2 or He.

The analysis of unidentified proteins involves extracting sequence data directly from MS/MS spectra. This process involves sequencing the amino acids of peptide chains using fragment ion series yi or bi. In these series, consecutive ions break apart an amino acid with a mass equal to $[M\text{-}H_2O]+$ (Neagu et al., 2022). This results in the establishment of the amino acid's identity and position within the peptide chain. By analyzing ion masses in the yi and bi series, the amino acid sequence of the corresponding peptide can be revealed.

ESI is used with high-resolution mass separators like quadrupoles, ion traps, and Q-TOF instruments to determine amino acid sequences with high accuracy and reliability (Anastasia et al., 2018). However, this approach is time-consuming due to the need for peptide fragment separation and amino acid sequencing. It's important to note that this technique cannot distinguish between leucine and isoleucine due to their identical masses (Alfaro et al., 2017).

Paper Received: 15 May 2024; Revised: 15 June 2024; Accepted: 22 July 2024
Please cite this article as: Cao N.M.T., Nguyen K.M., Trenh K.L., Vo N.D.K., 2024, Characterization of Peptides and Proteins Associated with Bacterial Proliferation and Bird's Nest Sample Matrix, Chemical Engineering Transactions, 111, 709-714 DOI:10.3303/CET24111119

In this study, we employ advanced bioinformatics tools—specifically FASTA and BLAST—to align sequences against a reference database. The score value reflects alignment accuracy, while the expected value estimates the likelihood of chance matches. Our custom Biotool® software utilizes data lookup algorithms to analyse amino acid sequences within small and medium peptides, leveraging online protein libraries.
Our investigation focuses on three key areas: insulin, peptides derived from bacterial proliferation, and peptides from the bird's nest matrix. By employing state-of-the-art mass spectrometry, we aim to identify and characterize these peptides and associated proteins. Understanding the molecular intricacies of bacterial proliferation and their interactions across diverse environments represents a significant advancement in microbiological research.

2. Method

2.1. Instrumentation and chemicals

2.1.1. Instrumentation

The research utilized a MicroQTOF-QII high-resolution MS/MS mass spectrometry system from Bruker Daltonics (Germany) in conjunction with an Agilent 1200 high-pressure liquid chromatography system (USA). Chromatographic separation was executed on a Kromasil-C4 column measuring 2.1x150 mm. Subsequently, the obtained data underwent processing via the integrated Data Analysis software from Bruker (Germany) and Biotool software.

2.1.2. Reagents and Strains

The reagents used included formic acid (FA), trifluoroacetic acid (TFA), and triethylamine (TEA) of analytical grade (Merck®). Isopropanol (IPA) and acetonitrile (ACN) of HPLC grade were procured from Schaulour. Deionized water was obtained from a Pure Water System (WP710). An insulin standard was also used. The lactic acid bacteria strains that produce bacteriocin, namely *Lactococcus lactis spp lactis*, *Lactobacillus casei, Lactobacillus garvieae*, and *Lactobacillus acidophilus*, were provided by the Industrial Microbiology Center of the Institute of Food Industry. Strains LTD 003 and LTD 005 were obtained from the Ho Chi Minh City High-Tech Agricultural Research and Development Center, Vietnam. Indicator bacteria strains, including Escherichia coli, *Salmonella typhimurium*, *Listeria monocytogenes*, and *Lactobacillus plantarum* JCM 1147, were also provided by the Industrial Microbiology Center of the Institute of Food Industry. The bird's nest matrix utilized in the research was sourced from an authorized swiftlet establishment in Ho Chi Minh City, Vietnam. This bird's nest matrix underwent grinding into powder form and was subsequently preserved in an airtight container until subsequent analysis (Qi Hao et al.,2016).

2.2. Sample preparation and peptide separation procedure

Prepare the insulin sample, 1 mg of standard insulin, and dissolve it in deionized water (V=10mL). Centrifuge the solution at 5000 rpm for 10 minutes to separate any particulate matter. After centrifugation, carefully remove the supernatant and inject it into the LC-MS system for analysis. Prepare the bird's nest sample, 100 mg of the ground sample and mix it with 5 mL of 50 mM NH_4HCO_3 solution adjusted to pH 8. Add 100 µL of a 0.1 mg/mL trypsin solution to the mixture and shake it for 20 minutes. Subsequently, incubate the sample at 37°C for 10 hours. After incubation, centrifuge the sample at 5000 rpm for 10 minutes to separate the supernatant. Collect the clear solution and proceed to inject it into the LC-MS system for analysis (Hopfgartner et al, 2020). For preparing Nisin samples, grow bacterial strains in MRS broth (30°C, 18h), centrifuge them (4000 rpm for 20 minutes), adjust the NaOH 10N solution to pH 6.5, and centrifuge them again (10000 rpm for 10 minutes). Collect the supernatant and inject it into the LC-MS system (Ko K.Y. et al., 2016). The LC-MS system utilizes a Kromasil-C4 column and mobile phases A (deionized water + 0.1% formic acid) and B (acetonitrile + 0.1% formic acid). Maintain a flow rate of 0.25 mL/min and set the column temperature to 50°C.

Table 1: The mobile phase program that is used to separate the peptide in bird's nest extract.

Time (min(	% Phase A	% Phase B
0.00	95.00	5.00
5.00	95.00	5.00
50.00	5.00	95.00
70.00	5.00	95.00

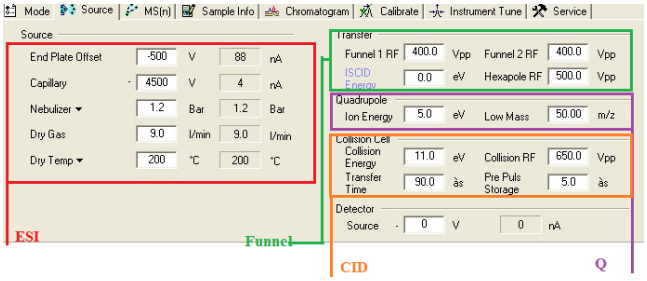


Figure 1: MS device parameters. Parameters of ESI ion generator, ion filter (Funnel), ion isolator (Q), internal ion generator (CID), mass separator (TOF).

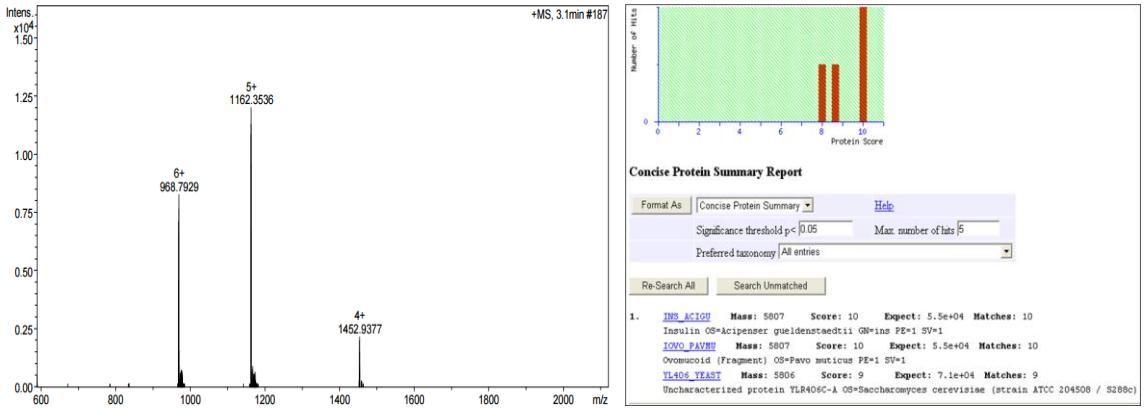
3. Result and discussion

3.1. Insulin identification via Biotool software following LC-MS analysis

The samples were separated by a Kromasil C4 column (2.1mm x 150 mm, 3.5 μm) The mobile phase was composed of 0.1% (v/v) formic acid in water (A) and acetonitrile (B). The flow rate was 0.25 mL/min, the column temperature was 50˚C. The mass spectrometry analysis detects insulin in a multi-ionized state, as outlined in Table 2. Retrieval of a list of m/z ions corresponding to insulin from the mass spectrometry results is conducted within the nasal region.

Table 2: Ion Charge States and Corresponding Deconvoluted Masses for Insulin

Ions m/z	Charge	Deconvolute (Da)
968.7929	6+	5807.7208
1162.3536	5+	5807.7391
1452.9377	4+	5808.7292



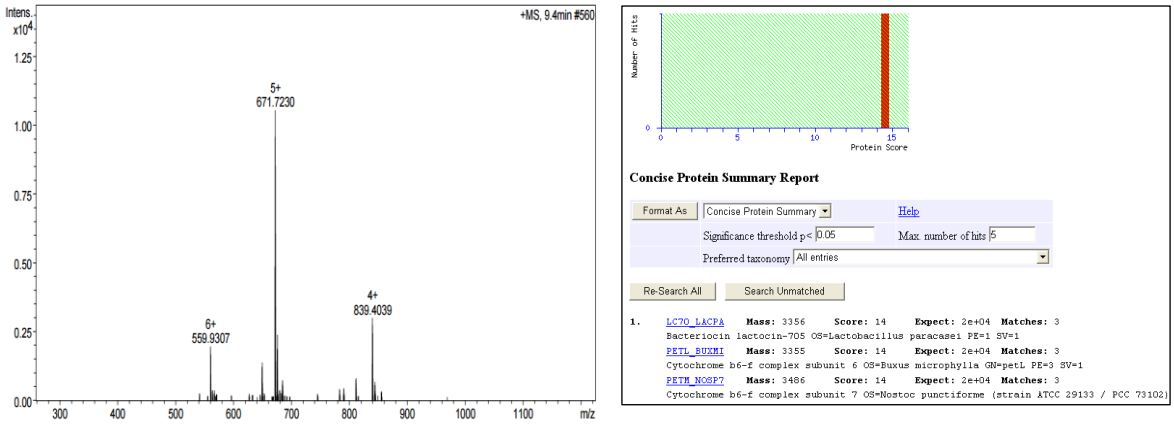
(A) *(B)*

Figure 2: (A) MS spectrum of insulin. (B) Identification of peptide on the Mascot's web interface.

Subsequently, Biotool software is utilized to import all MS spectral data alongside the list of *m/z* ions. This software interfaces with the internet-accessible Mascot protein database. Upon accessing the Mascot interface, essential information, illustrated in Figure 2B, is provided, facilitating the identification of proteins potentially aligning with the input MS data associated with insulin.

3.2. Peptide identification from bacteriocin-producing lactic acid bacteria growth process

The mass spectrometry analysis reveals that the growth sample of the lactic acid bacteria strain producing bacteriocin manifests in a multi-charged state, with the ions detailed in Table 3.



(A) *(B)*

Figure 3: (A) MS spectrum of bacteriocin-producing lactic acid bacteria. (B) Identification of peptide on the Mascot's web interface.

Consequently, positive outcomes were achieved from test samples of known origin and composition by transferring input MS data to the Mascot protein database through Biotool software. These results are consistent and align with the top position in the Mascot database matches (Grosse-Coosmann et al., 2005).

Table 3: Ion Charge States and Corresponding Deconvoluted Masses for bacteriocin-producing lactic acid bacteria

Ions m/z	Charge	Deconvolute (Da)
559.9307	6+	3354.5479
671.7230	5+	3354.5859
839.4039	4+	3354.5938

3.3. Peptide Identification from Bird's Nest Sample Extract

The intricate composition of bird's nest sample matrices offers a rich source of peptides and proteins that reflect the biological activities and adaptations of avian species. Building upon the acquired findings, our investigation progressed to identify peptides within a more complex sample, specifically, a base extract from bird's nests. The analysis unveiled 26 peptide fragments, accompanied by ion m/z and [M+H] values post-deconvolution (Figure 4). The identification of peptide fragments relied on the manifestation of ions m/z in a multi-charged state Notably, our study omitted the analysis of peptide fragments with a 1+ charge due to their unreliable detection outcomes. The results highlighted the presence of peptides within the mass range of 1000-2000 Da, primarily exhibiting charges of +2, +3, and +4. Remarkably, identical peptide fragments were detected in two extraction media: deionized water solution and Tris-HCl solution.

The peptide fragments P78, P70, P77, P22, P36, P67, P16, P32, P34, P8, P18, P3, P7, P68, P30, P20, P41, P80, P65, P60, P19, P27, P59, P1, P47, P55, P39, P76, P12, P9, P75, P4, P37, P44, P21, P25, P28 identified the *Gallus gallus* EST clone 16k18r1 protein from the *Gallus Gallus* family, exhibiting a sequence coverage of 82% and a corresponding score of 65 (Table 4). The recorded amino acid sequence of the protein is as follows: ^{1}CQLGDWRLTD RARHKPRFQG KVGLCCEV_A _MTRLHLAVL VCFLCHLVTF ^{51}PLCCNS_EYC LKFGPGEQPL IHRLD_E_LHD FNMLFQRGLQ SPAVN_FLAQ ^{101}PVSGQAATRV PASHCPQNMI RHWCRGSLSA QVAQSALY_V IIKNKNVRYL _^{151}KQSGSSRAF FFEC_KSIWK SKRTTATL_T PLRGVGVSMK GKVLIHLSAL ^{201}EDDP_SRGG.

The peptide fragments P30, P38, P73, P59, P5, P12, P40, P43, P50, P3, P56, P10, P6, P31, P61, P79, P11, P72, P15, P51, P54 identified the Gallus protein. The Gallus cDNA clone ChEST382g23 5', mRNA sequence belongs to the *Gallus Gallus* family, exhibiting a sequence coverage of 44% and a corresponding score of 55. The recorded amino acid sequence of this protein is as follows: ^{1}ADDEEEDVPFE EDSEEAGGGL DGGQGKRKRL FSKELRCMMY GFGDDQNPYT ^{51}ESVDILEDLV IEFITEMTHK AMSIGRQGRV QVEDIVFLIR KDPRKFARVK ^{101}DLLTMNEELK RARKAFDEAN YGS_FCVRAS RGHAVGSCCS RGRNVLCVAL ^{151}RCYLRWRSLL LAYLPCLSAF SGRHSSS_ML PDCTWYS_RV FIPFCEFEK_ ^{201}ALRIL_R_EV SLR_DC_IGH LEGP_QVLLY FGYCSVLC_Q ETWKTDLGYW ^{251}TNKVIKP

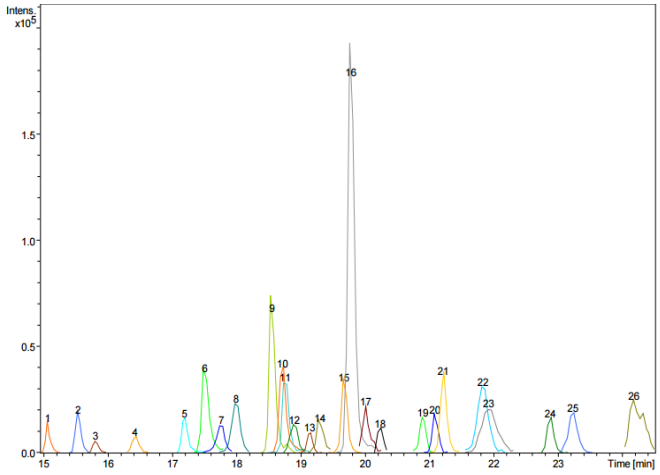


Figure 4: MS spectrum of peptide from bird's nest sample extract

Table 4: Identification of peptide derived from Bird's Nest Sample Extract on the Mascot's web interface

stt	Peptide	Amino acid sequence	Family	Protein	% coverage
1	P70, P22, P36, P16, P32, P34, P8, P18, P3, P7, P68, P30, P20, P41, P65, P60, P19, P27, P59, P1, P47, P55, P39, P76, P12, P9, P75, P4, P37, P44, P21, P25, P28	CQLGDWRLTD RARHKPRFQG KVGLCCEV_A _MTRLHLAVL VCFLCHLVTF PLCCNS_EYC LKFGPGEQPL IHRLD_E_LH DFNMLFQRGL QSPAVN_FLA QPVSGQAATR VPASHCPQNM IRHWCRGSLS AQVAQSALY_ VIIKNKNVRY L_KQSGSSRA FFFEC_KSIWK SKRTTATL_T PLRGVGVSMK GKVLIHLSAL EDDP_SRGG	Gallus Gallus	Gallus gallus EST clone 16k18r1	82%
2	P30, P38, P73, P59, P5, P12, P40, P43, P50, P3, P56, P10, P6, P31, P61, P79, P11, P72, P15, P51, P54	ADDEEEDVPFE EDSEEAGGGL DGGQGKRKRL FSKELRCMMY GFGDDQNPYT ESVDILEDLV IEFITEMTHK AMSIGRQGRV QVEDIVFLIR KDPRKFARVK DLLTMNEELK RARKAFDEAN YGS_FCVRAS RGHAVGSCCS RGRNVLCVAL RCYLRWRSLL LAYLPCLSAF SGRHSSS_ML PDCTWYS_RV FIPFCEFEK_ ALRIL_R_EV SLR_DC_IGH LEGP_QVLLY FGYCSVLC_Q ETWKTDLGYW TNKVIKP	Gallus Gallus	Gallus gallus cDNA clone ChEST382g23 5', mRNA sequence.	44%

4. Conclusions

The effectiveness of the Biotool® software and the Mascot protein database in identifying insulin from LC-MS analysis data has been firmly established. The software adeptly aligned the mass spectrometry (MS) data with the proteins cataloged in the database, underscoring its potential for precise protein identification. Employing the same method, we scrutinized the growth sample of the bacteriocin-producing lactic acid bacteria strain, consistently achieving alignment with the Mascot database. This validation reaffirms the method's efficacy. Furthermore, our investigation successfully identified peptides within a complex sample—specifically, a bird's nest sample base. The analysis revealed 26 peptide fragments, predominantly exhibiting charges of +2, +3, and +4, falling within the mass range of 1000-2000 Da. This outcome underscores the method's capability to handle

intricate samples and its reliability in peptide identification. In summary, the amalgamation of LC-MS analysis, Biotool software, and the Mascot™ protein database constitutes a robust and dependable approach for protein and peptide identification across diverse sample types.

Acknowledgments

We extend our sincere thanks to all colleagues in the Faculty of Applied Technology, School of Technology, Van Lang University, Ho Chi Minh City, Viet Nam for contributing to this research paper

References

Andrianova, A.A., DiProspero, T., Geib, C. et al. Electrospray Ionization with High-Resolution Mass Spectrometry as a Tool for Lignomics: Lignin Mass Spectrum Deconvolution. J. Am. Soc. Mass Spectrom. 29, 1044–1059 (2018)

Aguilar, M.I. HPLC of Peptides and Proteins:Methods and Protocols. humana press. 251, 227-244 (2004)

Alfaro, J.A., Ignatchenko, A., Ignatchenko, V. et al. Detecting protein variants by mass spectrometry: a comprehensive study in cancer cell-lines. Genome Med 9, 62 (2017)

Bork, P. Analysis of amino acid sequences. Academic Press. 54, 7-17 (2000)

Whitelegge J. Comprehensive analytical chemistry protein mass spectrometry. Elsevier, 52, 14-16 (2009).

Neagu, A.-N.; Jayathirtha, M.; Baxter, E.; Donnelly, M.; Petre, B.A.; Darie, C.C. Applications of Tandem Mass Spectrometry (MS/MS) in Protein Analysis for Biomedical Research. Molecules, 27, 2411 (2022)

Hopfgartner, G. Wenkui Li, Wenying Jian, and Yunlin Fu (Eds.): Sample preparation in LC-MS bioanalysis. Anal Bioanal Chem 412, 803–804 (2020).

Ko, K.Y., Park, S.R., Lim, HS. et al. Improved Pretreatment Method for Determination of Nisins A and Z and Monitoring in Cheese Using Liquid Chromatrography-Tandem Mass Spectrometry. Food Anal. Methods 9, 122–130 (2016).

Grosse-Coosmann, F., Boehm, A.M. & Sickmann, A. Efficient analysis and extraction of MS/MS result data from Mascot™ result files. BMC Bioinformatics 6, 290 (2005).

Qi Hao, A.R. Swiftlets and Edible Bird’s Nest Industry in Asia. PJSRR, 2, 32-48 (2016)

Liu, X. et al., Proteomic profile of edible bird's nest proteins. Journal of Agricultural and Food Chemistry, 60(51), 77-81 (2012)