Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Data processing for fennel protein characterization by Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS)

Maria Teresa Melfi[a], Basem Kanawati[b,*], Philippe Schmitt-Kopplin[b,c], Luigi Macchia[d], Diego Centonze[a], Donatella Nardiello[a,*]

[a] *Dipartimento di Scienze Agrarie, degli Alimenti e dell'Ambiente, Università degli Studi di Foggia, Via Napoli, 25, 71122 Foggia, Italy*
[b] *Research Unit Analytical BioGeoChemistry (BGC), Helmholtz Zentrum München, Ingolstaedter Landstrasse, 85764 Neuherberg, Germany*
[c] *Chair of Analytical Food Chemistry, Technical University of Munich, Alte Akademie 10, D-85354 Freising, Germany*
[d] *Dipartimento dell'Emergenza e dei Trapianti di Organi, Sezione di Allergologia ed Immunologia Clinica, Università degli Studi di Bari, Piazza G. Cesare, 11, 70124 Bari, Italy*

ARTICLE INFO

Article history:
Received 21 December 2020
Revised 3 March 2021
Accepted 10 March 2021
Available online 18 March 2021

Keywords:
Direct-infusion FT-ICR-MS
Fennel proteins
Peptide mass fingerprint
Matlab data processing
Custom-made protein database
In-silico enzymatic digestion

ABSTRACT

An untargeted shot-gun approach is described for the ultra-high-resolution analysis of fennel proteins by Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) combined with a home-made Matlab search algorithm. The first step of the proposed bioinformatic strategy was the development of a custom-made fennel protein database, starting from the well-known, on-line available, protein NCBI database, under *Foeniculum Vulgare* organism, consisting of 231 total proteins. Partial and redundant forms of proteins, repeatedly included in the official NCBI database under different codes, were removed. In the final custom-made database, in addition to the 92 fennel specific non-redundant proteins, 10 proteins belonging to recognized allergenic sources associated with spice-mugwort-allergy syndrome (celery, carrot, parsley, birch, and mugwort) were also

DOI of original article: 10.1016/j.foodres.2020.109919

* Corresponding authors.
E-mail addresses: basem.kanawati@helmholtz-muenchen.de (B. Kanawati), donatella.nardiello@unifg.it (D. Nardiello).

https://doi.org/10.1016/j.dib.2021.106960
2352-3409/© 2021 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

included. The second step was the in-silico enzymatic digestion, performed on all the 102 proteins, to obtain a theoretical list of *m/z* dataset of tryptic peptides. The Matlab processing data was the third and crucial step, necessary to search for in-silico mass calculated peptide sequences in the high resolution ICR mass spectra of the digested fennel extract. The final step was based on database searching in Peptide Mass Fingerprint (PMF) mode by using the matched *m/z* values as input data. The PMF search results confirmed the presence of 70 proteins (61 fennel specific and 9 allergenic proteins) inside the fennel extract.

© 2021 The Author(s). Published by Elsevier Inc.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Specifications Table

Subject	Proteomics, Analytical Chemistry, Food Chemistry
Specific subject area	Analytical Chemistry applied to protein analysis and food chemistry
Type of data	Table Figure
How data were acquired	Starting from the official NCBI database, a custom-made fennel protein database was derived by removing redundant and partial proteins and including recognized allergenic proteins associated with spice-mugwort-allergy syndrome. Tryptic peptide theoretical lists were obtained by in-silico enzymatic digestion of custom-made database proteins and compared to the experimental *m/z* values of the FT-ICR mass spectra of the digested fennel extract by a home-made Matlab search algorithm. A list of proteins inside the fennel extract was obtained by MASCOT Peptide Mass Fingerprint database search by using as input data that matched theoretical/experimental *m/z* dataset.
Data format	Raw Analyzed
Parameters for data collection	FT-ICR-MS analysis parameters, data processing and bioinformatics for protein characterization are briefly described in this paper and fully provided in the related research article.
Description of data collection	Data were collected by SolariX ion cyclotron resonance Fourier transform mass spectrometer (Bruker Daltonics GmbH, Bremen, DE) equipped with an Apollo II ESI source (Bruker Daltonics GmbH, Bremen, DE) and a 12 T superconducting magnet (Magnex Scientific, Yarnton, UK).
Data source location	Fennel extracts were prepared in the laboratories of the Section: Allergy and Clinical Immunology at the University of Bari (Bari, ITALY). The fennel tryptic peptide mixtures were prepared in the laboratories of the Research Unit Analytical BioGeoChemistry at Helmholtz Zentrum München (Neuherberg, Germany)
Data accessibility	Repository name: Mendeley Data: Data processing for fennel protein characterization by Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) Data identification number: http://dx.doi.org/10.17632/s354fkt6fp.1 Direct URL to data: http://dx.doi.org/10.17632/s354fkt6fp.1 Repository name: Mendeley Data: Supplementary Data for: Data processing for fennel protein characterization by Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) Data identification number: http://dx.doi.org/10.17632/9fxgd8wb2j.1 Direct URL to data: http://dx.doi.org/10.17632/9fxgd8wb2j.1
Related research article	M.T. Melfi, B. Kanawati, P. Schmitt-Kopplin, L. Macchia, D. Centonze, D. Nardiello, Investigation of fennel protein extracts by shot-gun Fourier transform ion cyclotron resonance mass spectrometry, Food Research International 139 (2021) 109,919 https://doi.org/10.1016/j.foodres.2020.109919

Value of the Data

- An innovative analytical platform by ultra-high-resolution FT-ICR-MS analysis combined with a home-made new Matlab search algorithm is proposed for the characterization of fennel proteins.
- The bioinformatics workflow allows for reliable protein characterizations by database searching in Peptide Mass Fingerprint mode, taking full advantages of the high measurement accuracy of the FT-ICR mass analyzer.
- The described strategy (direct-infusion FT-ICR-MS, peak list extraction, production of a sub-database of non-redundant specific protein entries, calculation of tryptic digestions, comparison between the theoretical and the experimental *m/z* datasets, and finally, the Mascot database searching in PMF mode) represents a very informative approach for a rapid and accurate protein characterization in food products, as well as in biological and clinical samples.
- The development of a custom-made protein database has allowed to overcome the intrinsic limits of NCBI protein database in terms of redundancy and incompleteness.

1. Data Description

All tables and Figures are reported in [1].

Table 1 shows the custom-made fennel protein database (protein name, code, amino acid sequence and molecular weight), generated starting from the official NCBI database associated to the organisms *Foeniculum Vulgaris*. From the original list of 231 total proteins, the redundant and partial proteins were removed and 92 fennel specific, non-redundant proteins were obtained.

In Table 2, the list of the most important allergenic proteins originating from other plants (celery, carrot, parsley, birch and mugwort), known as responsible, together with fennel, of the so-called birch-weed or fruit-spice syndrome [2–4] are listed. Therefore, in total, the custom-made protein database is composed of 102 different proteins: 92 fennel specific and 10 recognized allergenic proteins.

In Fig. 1, the primary structures of all the proteins found in the fennel extracts are shown, highlighting in bold red the amino acid sequences which correspond to the identified peptides. For each protein, peptide sequences and mass are reported in Table 3. For each identification, the number of matched peptides ranged from 3 (for smaller proteins, such as Cytochrome b6/ f complex subunit IV) and 99 (for larger proteins, such as Hypothetical chloroplast RF1). Further details are reported in the related scientific article [5].

Fig. 2A and B shows expanded views of the fennel MS spectrum in the mass segments 600–700 and 700–800 amu, respectively, also showing the amino acidic sequences matching the in silico peptide digests. Other mass segments: 800–1000 and 1000–1200 amu are shown in Fig. 3A and B, respectively. The MS spectra raw data files and the corresponding mass peak lists have been uploaded in a public repository [6].

Fig. 4 shows a comparison between the primary structure of the major pollen allergen Bet v 1-A, from *Betula pendula*, and the putative homologue protein found in the fennel extract; the bold gray strings correspond to the identified peptides.

2. Experimental Design, Materials and Methods

2.1. Protein extraction from fennel and enzymatic digestion

An amount of 100 g of the edible part of fresh *Foeniculum Vulgare* (purchased in local supermarkets) was washed properly, minced and homogenated (Heidolph DIAX 900 homogenizer with a Heidolph 10 F probe) for 15 min at 25,000 revolutions per minute (rpm), on ice, in

the presence of phosphate buffer solution (PBS) containing Ca^{2+}, Mg^{2+}, EDTA at a final concentration of 2 mM, and 700 µL of plant cell-specific protease inhibitor cocktail composed of 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF), 1,10-phenantroline, pepstatin A, bestatin and trans-epoxysuccinyl-Lleucylamido-(4-guanidino)butane (E64), at unknown concentration. The homogenate was then centrifuged at 12,000 *xg*, for 20 min at 4 °C and the supernatant was recovered. This step was twice repeated. Successively, the sample was ultracentrifuged at 100,000 *xg*, for 2 h, at 4 °C. From an initial amount of 100 g of fresh fennel, 40 mL of 100,000 *xg* supernatant were obtained. The extract was kept at −80 °C, until used. The protein content, determined according to the colorimetric Bradford method [7], was 3.5 mg L^{-1}.

The in-solution enzymatic digestion of the raw fennel extracts, diluted 1:10 dilution with water, was performed in duplicate, following the procedure of Khodadadi et al. with slight modifications [8]. Briefly, 400 µL of methanol, 100 µL of chloroform and 300 µL of water were added to 100 µl of fennel protein extract and mixed thoroughly. After centrifugation at 15,000 rpm for 15 min, the upper aqueous phase was discarded, whereas 300 µL of methanol was added slowly to the lower phase. Then, the extract was further centrifuged at 15,000 rpm for 15 min. After drying, the resulting pellet was resuspended in 50 mM NH_4HCO_3 to reach a pH of 8.5. After reduction with 50 mM dithiothreitol for 60 min at 56 °C, and alkylation with 50 mM iodoacetamide for 60 min at 37 °C in the dark, the enzymatic digestion was performed with trypsin at a 1:100 enzyme/protein concentration for 18 h of incubation at 37 °C. The resulting peptides mixtures were acidified with 5µL of 5% formic acid (FA) and centrifuged at 15,000 rpm for 15 min. Before ESI-FT-ICR mass spectrometry analyses, the peptide mixtures were diluted 1:10 in a mixture of acetonitrile/water (70:30, v/v) containing 0.2% FA.

2.2. ESI-FT-ICR mass spectrometry analyses

Ultra-high resolution ESI(+) mass spectra were acquired on a SolariX ion cyclotron resonance Fourier transform mass spectrometer (Bruker Daltonics GmbH, Bremen, DE) equipped with an Apollo II ESI source (Bruker Daltonics GmbH, Bremen, DE) and a 12 T superconducting magnet (Magnex Scientific, Yarnton, UK). The Bruker Solarix contains a linear ion beam guide, which consists of a splitted quadrupole (directly connected to an ion funnel), a quadrupole, which serves as a mass selection filter and a hexapole, which represents a relatively high pressure collision cell, for MS/MS experiments, which are assisted with argon atoms (collision gas). Samples were injected with a flow rate of 2 µL min^{-1}. The MS was calibrated with a 5 mg L^{-1} arginine solution reaching a mass error below 100 ppb and was tuned in order to obtain the highest sensitivity for peptides in the mass/charge (*m/z*) range of 122–3000 in broadband detection mode. The resolution was on average of $R = 400{,}000$ at *m/z* 400, enabling an excellent signal differentiation on a molecular level. Tryptic peptide mixtures of fennel proteins were analyzed by direct flow injections by the use of Electrospray Ionization double Quadrupole Fourier Transform-Ion Cyclotron Resonance mass spectrometry (ESI qQ-FT-ICR-MS). The used electrospray voltage was 3600 V in the positive ionization electrospray mode. No in-source ion fragmentation was done. This allowed for detection of the intact peptides resulting from the digested fennel proteins. A sine mathematic apodization function was applied on the time domain transients and is used to significantly reduce the FT-artifacts side lobes (wiggles), which do not count for the real (chemically relevant) signals in the mass spectra [9]. Technical replicates were performed by multiple analysis (at least 3) of the same sample fennel peptide mixture.

Extraction of mass spectra peak-lists, mass annotation and deconvolution were performed by using Data Analysis 4.4 (Bruker Daltonics). The mass spectra were calibrated by the use of the cluster ions of arginine in positive Electrospray ion mode, which range from *m/z* 175 from the [M + H]+ monomer ion until reaching *m/z* 1220 heptamer [7M+H]+ cluster ion.

2.3. Data processing and database searching

Starting from the official protein NCBI database under *Foeniculum Vulgare* organism (consisting in 231 total proteins), a custom-made fennel database of 92 different proteins was derived by removing partial and redundant forms of proteins, repeatedly included in the original NCBI database under different codes. Then, in the final version of the custom-made protein database, 10 additional proteins belonging to recognized allergenic sources associated with spice-mugwort-allergy syndrome (celery, carrot, parsley, birch, and mugwort) were also included. Each protein of the above described NCBI subset was subjected to simulated tryptic digestion by the use of the ExPaSy peptide mass calculator tool (ExPaSy Bioinformatics Resource Portal, https://web.expasy.org/peptide_mass) by setting the following parameters: enzyme trypsin, two allowed missed cleavages; mass range from 0 to unlimited Dalton; cysteines treated with iodoacetamide. For each protein, the in silico enzymatic digestion was performed by selecting consecutively the option to save the theoretical peptide masses in form of $[M + H]^{+}$, $[M + 2H]^{2+}$ and $[M + 3H]^{3+}$ in order to have a complete and realistic mass list of all the putative *m/z* ions, as a result of the use of an ESI source. The obtained tryptic peptide theoretical *m/z* lists were compared to the experimental *m/z* dataset of the fennel FT-ICR mass spectrum by a home-made Matlab search algorithm (as reported in the Supplementary Material), with a mass tolerance of 5 ppm for this step. The matched *m/z* values (z=+1; +2; and +3) were converted in the query format required for database Peptide Mass Fingerprint (PMF) searching (MASCOT search engine, Matrix Science, London, UK) as $[M + H]^{+}$ ions according to Eq.: $[M + H]^{+}=z^{*}m/z$ +(1-z)proton mass, where z is the charge state). By using as input data the matched $[M + H]^{+}$ lists, the PMF searching was performed under the *Viridiplantae* category of the official full redundant NCBI database. A maximum number of 2 missed cleavages were allowed and the option carbamidomethylation at cysteine residues was checked. A peptide mass tolerance of 2 ppm was set in the error window.

CRediT Author Statement

Study concept and design: **B. Kanawati** and **D. Nardiello:** Acquisition of data; **M.T. Melfi** and **B. Kanawati:** Analysis and interpretation of data; **B. Kanawati, D. Nardiello** and **M.T. Melfi:** Drafting of the manuscript; **D. Nardiello, B. Kanawati** and **M.T. Melfi:** Clinical studies on hypersensitive patients with fennel allergy; **L. Macchia:** Critical revision of the manuscript for important intellectual content; **P. Schmitt-Kopplin** and **D. Centonze**.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

Acknowledgments

This work is part of the Ph.D. research project of Maria Teresa Melfi titled “The Challenge of Proteomics in the field of Food Safety: Allergen detection”, which is (partly) financed by the Italian Ministry of Education, University and Research (MIUR).

Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.dib.2021.106960.

References

[1] D. Nardiello, B. Kanawati, M.T. Melfi, Data processing for fennel protein characterization by Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS), 1 (2020). doi:10.17632/s354fkt6fp.1.
[2] R. Asero, Relevance of pollen-specific IgE levels to the development of Apiaceae hypersensitivity in patients with birch pollen allergy, Allergy Eur. J. Allergy Clin. Immunol. 52 (1997) 560–564, doi:10.1111/j.1398-9995.1997.tb02600.x.
[3] E. Jensen-Jarolim, A. Leitner, R. Hirschwehr, D. Kraft, B. Wüthrich, O. Scheiner, J. Graf, C. Ebner, Characterization of allergens in Apiaceae spices: anise, fennel, coriander and cumin, Clin. Exp. Allergy 27 (1997) 1299–1306, doi:10.1111/j.1365-2222.1997.tb01175.x.
[4] E.A. Pastorello, L. Farioli, C. Stafylaraki, J. Scibilia, M.G. Giuffrida, A. Mascheri, M. Piantanida, C. Baro, L. Primavesi, M. Nichelatti, J.W. Schroeder, V. Pravettoni, Fennel allergy is a lipid-transfer protein (LTP)-related food hypersensitivity associated with peach allergy, J. Agric. Food Chem. 61 (2013) 740–746, doi:10.1021/jf303291k.
[5] M.T. Melfi, B. Kanawati, P. Schmitt-Kopplin, L. Macchia, D. Centonze, D. Nardiello, Investigation of fennel protein extracts by shot-gun Fourier transform ion cyclotron resonance mass spectrometry, Food Res. Int. 139 (2021) 109919, doi:10.1016/j.foodres.2020.109919.
[6] D. Nardiello, B. Kanawati, M.T. Melfi, Supplementary data for: data processing for fennel protein characterization by Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS), 1 (2021). doi:10.17632/9fxgd8wb2j.1.
[7] M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, Anal. Biochem. 72 (1976) 248–254, doi:10.1016/0003-2697(76)90527-3.
[8] E. Khodadadi, B.A. Fakheri, S. Aharizad, A. Emamjomeh, M. Norouzi, S. Komatsu, Leaf proteomics of drought-sensitive and -tolerant genotypes of fennel, Biochim. Biophys. Acta Proteins Proteom. 1865 (2017) 1433–1444, doi:10.1016/j.bbapap.2017.08.012.
[9] B. Kanawati, T.M. Bader, K.-.P. Wanczek, Y. Li, P. Schmitt-Kopplin, Fourier transform (FT)-artifacts and power-function resolution filter in Fourier transform mass spectrometry, Rapid Commun. Mass Spectrom. 31 (2017) 1607–1615, doi:10.1002/rcm.7940.