associated to doi: 10.1016/j.talanta.2021.123188

Running title: *Searching for allergenic proteins in novel foods.*

A new paradigm to search for allergenic proteins in novel foods by integrating proteomics analysis and in silico sequence homology prediction: focus on spirulina and chlorella microalgae

Mariachiara Bianco[1], Giovanni Ventura[1]*, Cosima Damiana Calvano[1,2], Ilario Losito[1,2], Tommaso R.I. Cataldi1,2*

[1] *Department of Chemistry and* [2]*Interdepartmental Research Center SMART, University of Bari Aldo Moro, via Orabona 4, 70126, Bari, Italy*

Number of Figures: 4

Number of Tables: 1

Supplementary Material: yes

Keywords: food allergens, proteomics, spirulina, chlorella, microalgae, FTMS, tandem MS.

Author for correspondence, email: giovanni.ventura@uniba.it, tommaso.cataldi@uniba.it

Abstract

Since novel nutrient sources with high protein content, such as yeast, fungi, bacteria, algae, and insects, are increasingly introduced in the consumer market, safety evaluation studies on their potentially allergenic proteins are required. A pipeline for *in silico* establishing the sequence-based homology between proteins of spirulina (*Arthrospira platensis*) and chlorella (*Chlorella vulgaris*) micro-algae and those included in the AllergenOnline (AO) database (AllergenOnline.org) is described. The extracted proteins were first identified through tryptic peptides analysis by reversed-phase liquid chromatography and high resolution/accuracy Fourier-transform tandem mass spectrometry (RPLC-ESI-FTMS/MS), followed by a quest on the UniProt database. The AO database was subsequently interrogated to assess sequence similarity between identified microalgal proteins and known allergens, based on criteria established by the World Health Organization (WHO) and Food and Agriculture Organization (FAO). A direct search for microalgal proteins already included in allergen databases was also performed using the Allergome database. Six proteins exhibiting a significant homology with food allergens were identified in spirulina extracts. Five of them, *i.e.*, two thioredoxins (D4ZSU6, K1VP15), a superoxide dismutase (C3V3P3), a glyceraldehyde-3-phosphate dehydrogenase (K1W168), and a triosephosphate isomerase (D5A635), resulted from the search on AO. The sixth protein, C-phycocyanin beta subunit (P72508), was directly obtained after examining the Allergome database. Two proteins exhibiting significant sequence homology with food allergens were retrieved in chlorella extracts, *viz.* calmodulin (A0A2P6TFR8), which is related to troponin c (D7F1Q2), and fructose-bisphosphate aldolase (A0A2P6TDD0). Specific serum screenings based on immunochemical tests should be undertaken to confirm or rule out the allergenicity of the identified proteins.

1. INTRODUCTION

Food allergens are food components, chiefly proteins or protein epitopes, affecting susceptible individuals by immune-mediated reactions [1]. The initial list of allergenic foods of concern, also known as “the big 8”, was incremented by regulation 1169/2011 in the European Union and currently contains 13 foods, *i.e.* eggs, milk, fish, peanuts, crustaceans, soybeans, wheat, tree nut, lupin, shellfish, celery, mustard, and sesame, alongside sulfur dioxide [2]. Many studies have been focused on these major food allergens, establishing proteins responsible for allergenicity and identifying marker peptides for quantitation based on *bottom-up* proteomics [3–11].

Presently, microalgae and cyanobacteria are used as tablets/capsules or introduced in foodstuffs, such as snacks, pasta, cookies, bread, and so on [12–14]. Unfortunately, the question of risk assessment has not been systematically addressed for proteins extracted from microalgae and lacking attentions have been devoted to such novel foods' side effects, including allergenic reactions in sensitized populations [15–17]. Tiberg *et al.* [18, 19] studied the effects of chlorella (*Chlorella vulgaris*) extracts triggered in 6-17 years old children, with or without other allergies, using skin prick and radio-allergosorbent test (RAST). Microalgae and molds share similar growing conditions, and the possibility that mold-sensitive people develop equivalent adverse reactions to algae cannot be ruled out [18, 19]. Tiberg *et al.* [20] examined different extracts of chlorella, discovering that purified samples exhibited a reduced IgE binding activity in comparison with crude ones. Yim *et al.* [21] reported acute tubulointerstitial nephritis developed in a boy, after three months of ingestion of chlorella tabs as a food supplement [21]. The first case of anaphylaxis to spirulina (*Arthrospira platensis*) was described by Petrus *et al.* [22] in a 14 years old adolescent. Through Western blot analysis on spirulina protein extracts, and MALDI-ToF-MS analysis of IgE-reactive fractions digests, the β-chain of phycocyanin C was identified as an allergenic protein [22]. However, most food

allergens are members of large protein families with high sequence similarity that cannot be distinguished by antibody-based assays.

Current methods to investigate of food allergens and allergenic epitopes are based on biochemical and immunological tests, such as ELISA, protein/peptide microarray, immunofluorescence, radioimmunoassay, Western blotting, and immunohistochemistry [4, 23–29]. However, most of them are expensive, time-consuming, and not sufficiently selective. The allergenic potential of a protein can be predicted through the analysis of its sequence, structure, and B- or T-cell epitopes. This last classification is based on their respective receptors: B-cell epitopes are also known as IgE epitopes and can be continuous (or linear) and discontinuous (or conformational). Continuous epitopes are short linear peptide fragments that are usually sequences of contiguous amino acids, while conformational epitopes are comprised of amino acids that line up because of the tertiary structure of an allergenic protein [30–33].

The groundwork of allergen/epitope recognition was laid by Hopp and Woods[34] in 1981, when the first B-cell epitope prediction method was developed. Since then, many procedures have been developed or adapted from other computational tools. Based on the type of information being used, prediction methods can be categorized into *i)* sequence homology, *ii)* structure-based, and *iii)* hybrid ones.

In silico search for amino acid sequence homologies and/or structural similarities with known allergens is becoming an interesting alternative to food allergen discovery in novel foods. Based on a comparison between amino acid sequences, Food and Agriculture Organization (FAO) of the United Nations and World Health Organization (WHO) established guidelines for the determination of cross-reactivity between the expressed proteins and known allergens [35]. Cross-reactivity must be considered when (i) the percentage of identity (PID) of amino acid sequences is higher than 35% using a window of 80 amino acids (*vide infra*) or (ii) there is an identity of six contiguous amino acids.

Using these rules, Polikovsky *et al.* [36] evaluated *in silico* the potential allergenicity of proteins extracted by *Ulva* sp., a macroalga used as food supplement, leading to the recognition of superoxide dismutase as a possible food allergen [36]. Abdelmoteleb *et al.* [37] investigated the potential risks of chlorella using *in silico* genomic data and an online allergenic protein software, leading to an over-prediction of allergens. To greatly lower the number of false positives, the same authors suggested the use of stricter parameters [37] such as matching criteria identity higher than 35% and a statistical expectation score (E-score, *vide infra*) below 10^{-7}.

To address genomic data over-prediction as well as the existence of multiple redundant isoforms of allergenic proteins and their relatively low abundance, an integrated approach to food allergen discovery is proposed here. The strategy started with the analysis of tryptic digests of microalgal proteins by reversed-phase liquid chromatography and Fourier-transform tandem mass spectrometry with electrospray ionization (RPLC-ESI-FTMS/MS), followed by single protein quest on the UniProt database using dedicated tools, *e.g.* Thermo Protein Discoverer™. The identified proteins of chlorella (*C. vulgaris*) and spirulina (*A. platensis*) microalgae were then systematically compared with those listed in the AllergenOnline (AO) database [38], (https://www.allergenonline.org), containing the amino acid sequences of most allergenic food proteins. AO output data were filtered with the aim to identify proteins amenable to satisfy both the FAO/WHO criteria, separately and together. E-score, percentage of identity (PID), amino acid overlap (aao) and the possible existence of identical sub-sequences including at least 6 amino acids were calculated and used to filter out the output by a data processing pipeline written using MATLAB®. The *in silico* results were critically evaluated after inspecting Allergome (https://www.allergome.org) and Allergen (https://www.allergen.org) databases, being two well-known and very useful online resources of allergenic proteins.

2. MATERIALS AND METHODS

2.1 Chemicals. Water, hexane, acetonitrile, methanol, chloroform, and formic acid were purchased from Sigma-Aldrich (Milan, Italy). All solvents used were LC-MS grade except for hexane and chloroform (HPLC grade). Tris(hydroxymethyl) aminomethane hydrochloride (tris-HCl), DL-dithiothreitol (DTT), iodoacetamide (IAA), ammonium bicarbonate, and trypsin from porcine pancreas were obtained from Sigma-Aldrich (Milan, Italy). RapiGest surfactant was obtained from Waters Corporation (Milan, Italy). Solid-phase extraction (SPE) C18 tubes were from Supelco (Milan, Italy). Standard solutions for mass spectrometer calibration were purchased from Thermo Scientific (Waltham, Massachusetts, United States). Capsules as food supplements of spirulina and chlorella microalgae (Longlife nutritional supplements, Phoenix srl, Milan, Italy) were purchased from local stores.

2.2 Microalgae protein extraction, digestion, and purification. Protein extraction, digestion, and purification. Specifically, to 0.1 g of microalgae powder 9.5 mL of 50 mM Tris-HCl were added. After vortexing, solutions were incubated for 1 h at 55 °C, stirring samples every 20 minutes. Upon addition of 0.5 mL of methanol, samples were incubated in an ultrasound bath for 10 minutes to promote protein extraction. Thereafter, 3.5 mL of hexane were added, and the mixture was vigorously vortexed to facilitate fat transfer into the organic layer; samples were subsequently centrifuged for 15 minutes at 5000 g (*viz.,* 6000 rmp and 115 mm in radius). Finally, the organic phase was discharged, and 100 µL of the aqueous phase were collected and dried under a nitrogen flow.

Thus, 100 µL of RapiGest solution (0.1% w/v in 50 mM NH_4HCO_3) were added to dried samples and were vigorously vortexed to dissolve them in the surfactant solution. Then, 10 µL of 50 mM DTT were added and the samples were incubated for 30 minutes at 60 °C. Subsequently, 10 µL of 150 mM IAA were added and the samples were kept in the dark for 30 minutes at room temperature.

Afterwards, 5 µL of trypsin from porcine pancreas (0.1 µg/µL in 25mM NH_4HCO_3) were added and the samples were incubated at 37 °C overnight. The enzymatic reaction was stopped by adding formic acid for a final pH≈2. The samples were dried again under nitrogen flow.

The purification of tryptic digests was performed using homemade C18 tips. In detail, 10 mg of C18 stationary phase obtained from disassembled Supelco C18 tubes and dissolved in 100 µL of acetonitrile were collected in 200 µL tips, appropriately latched. Then, 100 µL of 0.1% formic acid solution were added for conditioning the SPE C18 phase. The samples, resuspended in the same solution, were loaded into homemade tips. Tips were washed twice with 100 µL of 0.1 % of formic acid, and elution was performed using 50 µL of CH_3CN/H_2O (70/30 v/v with 0.1% of formic acid). The purified samples were dried under nitrogen and resuspended in 50 µL of H_2O/ CH_3CN (95/5 v/v with 0.1% of formic acid), a mixture which matches the initial composition of the gradient elution program used for RPLC run (see next section).

2.3 RPLC-ESI-FTMS instrumentation and operating conditions. RPLC-ESI-FTMS analyses were performed using an Ultimate 3000 UHPLC chromatographic station coupled to a quadrupole-Orbitrap spectrometer equipped with a higher-energy collision dissociation (HCD, Q-Exactive, Thermo Scientific, Waltham, MA, USA) through a HESI (heated electrospray ionization) source. The separation was performed at 40 °C using a Phenomenex Aeris WIDEPORE 200 Å C18 column (250 x 2.1 mm, 3.6 µm), equipped with a Phenomenex AJO 8783 WIDEPORE C18 (2 x 2.1 mm ID) security guard cartridge, using H_2O (solvent A) and CH_3CN (solvent B) both containing 0.1% of formic acid. In detail, the gradient used during each chromatographic run, at a flow rate of 0.200 mL/min, was the following: 0 − 2 min at 5% solvent B; 2 − 20 min linear from 5% to 60% (v/v) of B; 20 – 22 min linear from 60% to 100% B; 22−26 min isocratic at 100% of B; 26−30 min back to the initial composition, followed by 5 min equilibration time. Mass spectrometry analyses were carried out in data dependent mode in positive polarity. The ESI and ion optic parameters adopted were the following:

sheath gas flow rate, 10 (arbitrary units); auxiliary gas flow rate, 5 (arbitrary units); spray voltage, 3.5 kV in positive polarity; capillary temperature, 200 °C; S-lens radio frequency level, 100 arbitrary units. Positive MS *full-scan* spectra were acquired in the *m/z* range 500–2000 at a 70k resolution using an automatic gain control (AGC) target of $3\ 10^6$. Tandem MS Full-MS/ddMS^2 was performed with the five most dominant ions selected from the first MS scans by repetitively full scan MS with HCD in a Q-Exactive mass spectrometer using normalized collisional energy (NCE) fixed at 30, with a 17.5k resolution, AGC of $1e^5$, IT fill time of 50 ms, isolation window of 4 *m/z*, minimum AGC of $8.00e^3$, and dynamic exclusion of 10 s. After calibration with a solution containing caffeine, MRFA, and Ultramark® (Thermo Scientific), the mass accuracy ranged between 0.15 and 0.21 ppm in positive polarity. The LC-MS instrumental control and the first processing data were performed by the Xcalibur software 2.2 SP1.48 (Thermo Scientific).

2.4 Identification of microalgae proteins using Full-MS/ddMS2 analysis. The identification of extracted microalgal proteins was carried out by Proteome Discoverer™ (version 2.4, Thermo Fisher Scientific) starting from Full-MS/ddMS^2 raw files obtained after the analysis of tryptic digests. Specifically, tryptic peptide MS/MS data were used as input to quest against three sub-sets of the UniProt database (https://www.uniprot.org/uniprot/) downloaded on August 20, 2021. Two explorations were focused on proteins already listed in the database for the specific microalgae, *i.e.*, *A. platensis* (https://www.uniprot.org/uniprot/?query= Arthrospira + platensis) and *C. vulgaris* (https://www.uniprot.org/uniprot/?query=chlorella+vulgaris), respectively. The third search was extended to all proteins referred to the *Chlorophyta* phylum (https://www.uniprot.org/uniprot/?query=Chlorophyta#). Furthermore, the *Allergome* database embedded into UniProt (https://www.uniprot.org/uniprot/?query=Allergome#) was also exploited to identify allergenic proteins directly from Full-MS/ddMS^2 raw data, since it is currently the most

comprehensive database for allergenic proteins identified in all species. The Processing and Consensus workflows for Proteome Discoverer™ operation were PWF_QE_Basic_SequestHT.pdProcessingWF and CWF_Basic.pdConsensusWF, respectively (**see Figure S1**). Other adopted criteria, besides instrumental quest parameters reported in Supplementary Material, were the following: trypsin as the enzyme, 2 missed cleavages, minimum and maximum length of peptides equal to 6 and 144 amino acids respectively, 10 ppm and 0.02 Da as tolerance for precursor and fragment ions respectively, met-oxidation, acetyl, met-loss, and met-loss+acetyl as dynamic modifications and carbamidomethylation of cysteines as a static modification.

2.5 Search for sequence homology in the AllergenOnline database and data filtering using a homemade MATLAB® pipeline. Before data processing based on Proteome Discoverer™, downloaded DBs were imported in MATLAB® 2021 and processed to exclude redundant proteins by a homemade algorithm. All the identified microalgal proteins were subsequently evaluated for sequence homology with allergenic proteins listed in the AO database. As emphasized in **Figure 1**, the FASTA sequences of microalgal proteins were first retrieved, starting from the corresponding accession numbers using a MATLAB® pipeline and cumulatively introduced in the AO entry window, where *Full FASTA 36* was set as the search method. The corresponding output, often including several AO proteins for each microalgal protein, was carefully processed by MATLAB® (*vide infra*). The latter started with the selection of database proteins with an E-score not higher than $1*10^{-7}$, then those, among them, characterized by a PID higher than 35% and with at least 80 amino acids overlapping (aao) with the input microalgal protein were further selected. Finally, the ensuing presence of identical sequences of at least six amino acids with respect to the input microalgal protein was assessed for each of the proteins selected from the database and added to the information exported. The complete collection of the MATLAB® function will be shared upon

request.

We wish to emphasize that using FAO/WHO guideline of 35% identity in sliding windows of 80 aa is a computationally very time-consuming and slow process. For this reason, this criterion is not applicable for high throughput analysis. AO *Full FASTA 36* method uses a different approach whereby entire protein sequences are aligned, and then a similarity examination is performed. AO results were finally filtered considering only those having a PID> 35% in at least 80 aao, and contiguous sequences of at least six identical aa were searched for. As evidenced in the bottom-right part of **Figure 1**, an alternative, more direct, approach to the recognition of possible allergenic proteins among those extracted from microalgae entails the application of Proteome Discoverer to the *Allergome* database. A set of proteins arising from the AO and/or from the *Allergome* database could thus be obtained as the outcome.

3. RESULTS AND DISCUSSION

3.1 Proteomics analysis of microalgae extracts

To identify proteins of microalgae powders extracted in the buffer solution, a *bottom-up* proteomics approach based on tryptic digestion and LC-MS/MS analysis was adopted. Typical total ion current (TIC) chromatograms of both chlorella and spirulina samples were reported in **Figure S2** (Supplementary Material), plots A and B, respectively. Digested proteins were analyzed by RPLC-ESI-FTMS/MS in data-dependent mode (ddMS^2) and the resulting raw data were used as input of Proteome Discoverer^{TM} processing. Redundant proteins found in DBs, *i.e.*, identical proteins listed with different accession numbers, were purposely removed, thus leaving only single protein in each case. Therefore, the original raw DB of *A. platensis* was reduced to 23,354 protein sequences from 25,226, which means a redundancy reduction of 7.42%. Likewise, the DBs of *C. vulgaris* and

Chlorophyta were decreased to 499 proteins (redundancy reduction 10.01%) and 882,764 (redundancy reduction 0.83%), respectively.

Figure 2 shows, as an example, the tandem mass spectra of two peptides identified in the tryptic digests of chlorella protein extracts detected as doubly charged precursor ions at *m/z* 951.414 (plot A) and *m/z* 802.418 (plot B). The product ions in **Figure 2A** allowed us to identify the precursor ion at *m/z* 951.414 as the peptide EADQDGDGQVDYSEFVK, covering 15% of the protein known as calmodulin, which is involved in calcium-mediated signal transduction. The doubly charged ion at *m/z* 802.418 was recognized as AVNVTGPNGAPPEGAPR (**Figure 2B**) belonging to the sequence of the glycine-rich 2 protein, for which further peptides were recognized, leading to a final coverage of 31%. In both cases, the most intense ions detected in the spectra (**Figure 2**) were *y*- (to great extent) and *b*- types, together with the couple of a_2/b_2 ions, typically formed in the higher-energy collisional dissociation (HCD) cell of the Q-Exactive instrument [39, 40]. Plots A and B of **Figure 3** show tandem mass spectra of peptides detected at *m/z* 699.886^{2+} and 702.877^{2+}, respectively, upon analysis of spirulina protein digest. Whereas the precursor ion at *m/z* 699.886^{2+} (**Figure 3A**) was recognized as SLGTPIEAVAEGVR, which is a peptide of allophycocyanin alpha chain (13 peptides recognized, protein coverage 76%), the precursor ion at *m/z* 702.877^{2+} (**Figure 3B**) was identified as the peptide ITSNASTIVSNAAR, belonging to the C-phycocyanin beta subunit, an allergenic protein (13 peptides recognized and protein coverage 73%) (*vide infra*). Both the cited proteins from spirulina microalga correspond to light-harvesting photosynthetic bile pigment proteins from the phycobiliprotein complex. Further examples of MS/MS spectra assigned to tryptic peptides of spirulina and chlorella proteins are reported in the **Figures S3-S8** (Supplementary Material).

3.2 Lists of identified proteins in spirulina and chlorella microalgae extracts.

All proteins of *A. platensis* and *C. vulgaris* identified with a coverage greater than 10% are listed in

Tables S1 and **S2** (Supplementary Material), respectively. The use of this relatively low threshold for coverage was dictated by the fact that two proteins may share similar structures, and then exert comparable biological effects, even in the presence of several differences in their amino acidic sequences. As mentioned above, in the case of spirulina the assessment was carried out on protein extracts using the species-specific DB of *A. platensis* (**Table S1**), while both DBs referred to *C. vulgaris* and to the entire *Chlorophyta* phylum (**Table S2**) were used in the case of chlorella, since relatively few proteins (*i.e.*, 499 after elimination of redundances) were listed in the first one. Up to 103 and 26 proteins with coverage greater than 10% were recognized for spirulina and chlorella samples, respectively. Among them, protein with accession number P72508, a C-phycocyanin beta subunit of spirulina, is of great interest, being already known as a food allergen, as reported on the *Allergome* database (protein code: 8833), in which it was inserted after the case of anaphylaxis reported by Petrus *et al*. [22] in a 14 years old child. Further proteins related to as phycocyanin beta subunit/chain in the UniProt database, corresponding to accession numbers B1NJ40, A9UKJ0, Q6XAW9, Q6XAW5, Q208D1, and Q5SCD5 (see **Table S1**, Supplementary Material, and **Figure S9**), were recognized in spirulina extracts. Notably, the C-phycocyanin beta subunit protein is currently not reported as an allergen on the AO database. The overlap of these proteins, which share high identity percentages compared to P72508, is reported in **Figure S9**. According to the FAO/WHO guidelines, possible cross-reactions in subjects showing allergic reactions to P72508 cannot be ruled out and should be seriously considered.

3.3 In silico evaluation of sequence homology between spirulina proteins and known allergenic proteins.

Both WHO/FAO constrains, separately and together, were used to find sequence homologies between proteins included in the AO database and those proteins identified in spirulina and

chlorella extracts exhibiting an E-score lower than 10^{-7}. As a result, the information described, respectively in **Table S3** and **Table S4** (Supplementary Material), was retrieved. In these tables each microalgal protein, represented by its accession number and name, is associated with one or more proteins listed in the AO database, followed by the number of overlapping amino acids (aao), PID, number of identical sub-sequences including at least 6 amino acids and the E-score obtained through database search.

The flowchart that describes the MATLAB® pipeline used to processing data of proteins resulting from the AO database is illustrated in **Figure 4**. Starting from these outcomes, superfluous information of non-allergenic proteins were removed, and a table of data of potentially allergenic proteins was generated. All implemented MATLAB® functions will be freely shared under request with a complete user guide. In the flowchart of **Figure 4**, **i** and FP represent, respectively, the counter for those sequences and their total number. The E-score for each microalgal protein retrieved from the database was evaluated and only allergenic protein sequences with a value not higher than 10^{-7} were selected and grouped, thus obtaining the total number indicated as **AP** in the flowchart. E-score represents the probability for an alignment to be the result of chance and reflects amino acid identity or similarity, alignment length, and database size. The specific E-score cut-off was chosen since values lower than $1*10^{-7}$ are often obtained in the case of protein cross-reactivity. Each of the allergenic proteins selected, related to a specific value of the **j** counter, was subsequently checked for constraints PID $\geq$ 35% and aao $\geq$ 80. If just one of these was not satisfied, a false (F) flag was assigned, and the next candidate allergenic protein was checked. Conversely, if both were satisfied, an evaluation of identical sequences with at least six amino acids existing between the microalgal protein and the candidate allergenic protein was performed, and the ensuing results used as additional information, yet their occurrence was not considered as a constraint for the selection of the candidate protein. Finally, when i> FP, the MATLAB® pipeline ends, and results are stored in a

tabled format.

It is worth mentioning that the first protein in terms of coverage, retrieved directly from the *Allergome* database using Proteome Discoverer™, *i.e.*, C-phycocyanin beta subunit, was not found in the *AllergenOnline* database. It is thus reported in **Table 1** but without any information arising from the search for homology on the latter. Notably, the *Allergome* database contains 4446 proteins, with a level of a redundancy of 5.3%, whereas 2233 proteins are included in that of AO, which applies different criteria for allergen identification (see http://www.allergenonline.org/about.shtml). Proteins with accession numbers B1NJ40, A9UKJ0, Q6XAW9, Q6XAW5, Q208D1, and Q5SCD5 identified in spirulina extracts should be considered for potential allergenicity, based on the homology with the P72508 protein as emphasized in **Figure S9** (Supplementary Material).

The AO database returned six spirulina proteins for which sequence homology was found with several allergenic proteins (see **Table S3,** Supplementary Material), including two thioredoxins (D4ZSU6, K1VP15), glutathione-dependent peroxiredoxin (K1X048), superoxide dismutase (fragment, C3V6P3), glyceraldehyde-3-phosphate dehydrogenase (K1W168), and triosephosphate isomerase (D5A635). Thioredoxins are small proteins possessing two cysteine residues, capable of forming intramolecular disulphide bridges in a conserved motif. Their presence was assessed in both chlorella and spirulina extracted samples (*vide infra*), and two thioredoxins recognized in the latter showed a high percentage of identity (97.2%), as considered by simulated alignment based on the *Align* function, freely available on the UniProt website (https://www.uniprot.org/align/, **Figure S10**, Supplementary Material). Results obtained by AO database showed the occurrence of an additional protein, *i.e.*, a thioredoxin codified as M5E6V3, associated with spirulina thioredoxin K1VP15 but not with the other thioredoxin, D4ZSU6, being the overlapping with M5E6V3 lower than 80 amino acids. Weichel *et al*. [41] retrieved a novel family of cross-reactivity allergens named thioredoxins

that contribute to the symptoms of baker's asthma, a serious problem especially for workers in the food industry, most likely related to grass pollen allergy. Thioredoxin *h*1 is reported only as airway allergen in the *Allergen* database. According to Righetti and co-workers [42], thioredoxins *h* of corn seeds, reported as a food allergens in the allergome database, are analogous to spirulina thioredoxin proteins.

With a PID > 35%, aao > 130, E-score $<1*10^{-16}$ and at least one overlapping sequence with more than 6 identical aa, five allergenic proteins that were related to the spirulina glutathione-dependent peroxiredoxin (K1X048) are listed in **Table S3**. Apparently, these proteins are chiefly associated with the *fungi* kingdom and are not considered as food allergens, being their allergenicity defined by skin contact or inhalation and not by ingestion [43–46]. Protein C3V6P3 of spirulina, a superoxide dismutase fragment, was related to seven superoxide dismutases (see **Table 1**), which can be divided into three groups. The first one is represented by three isoforms (Q9FSJ2, P35017, Q9STB5) of superoxide dismutase found in *Hevea brasiliensis* (para rubber tree), classified as *Hev* b 10 allergen, whose allergenicity, according to the *Allergome* database, is related to inhalation and contact with skin [47–49]. Notably, the allergen.org site reports that only the contact with skin of this allergen poses a potential risk for susceptible subjects. The second group of superoxide dismutase reported in **Table S3** (Supplementary Material) is represented by a single protein from pistachio seeds (B2BDZ8), known as a food allergen, with an E-score lower than $2*10^{-24}$. Noorbakhsh *et al.* [50] demonstrated the allergenicity of pistachio superoxide dismutase in 40% of the explored cases through IgE test, ELISA and immunoblotting. Members of the third group of superoxide dismutase (Q92450, M5ECN9 and Q873M4) are allergens of fungi species with allergenicity due to inhalation or contact with skin [51–53].

As shown in **Table S3**, very high aao values (337-338), PID values greater than 45% and very low E-scores (lower than 10^{-63}) were found for two glyceraldehyde-3-phosphate dehydrogenases

(A0A5N5Q6M7, C7C4X1) of the AO database that were related to the same protein identified in spirulina extracts (K1W168). In both cases four identical amino acid sequences were also found, with lengths 13, 11, 10, 9 and 14, 9, 9, 6, respectively. Interestingly, one of the two proteins, C7C4X1, has been reported to be responsible of the allergenicity linked to inhalation of wheat and related to baker's asthma [54]. The other one, A0A5N5Q6M7, was recognized as a food allergen related to *Pangasianodon hypophthalmus* (striped catfish), reported in the *Allergen* database as *Pan h 13* [55]. The overlap between the sequences of proteins A0A5N5Q6M7 and K1W168 protein, leading to a very low E-score ($7.5*10^{-64}$), is reported in **Figure S11** (Supplementary Material). Eight triosephosphate isomerases, included in the AO database, were found to have relevant sequence homology with the last spirulina protein reported in **Table S3**, that is also a triosephosphate isomerase. The first two among them, L7UZA7 and A0A088SAX2, related to *Dermatophagoides farinae (i.e.,* the American house dust mite*)* [56] and protein Q9FS79, found in wheat [54, 57], are not considered food allergens. The remaining triosephosphate isomerases assessed on the AO database, related to fish or crustaceans, are indicated as food allergens in the Allergome database: A0A5N5Q6M9 as Pan h 8 from *P. hypophthalmus*), A0A1L5YRA2 as *Scy p 8* from *Scylla paramamosain* (Mud crab), D7F1Q0 as *Cra c 8* from *Crangon crangon* (Brown shrimp), F5A6E9 as *Pro c 8* from *Procambarus clarkia* (Red swamp crayfish) and B5DGL3 as *Sal s 8* from *Salmo salar* (Atlantic salmon) [55, 58].

Based on the results described in **Table S3**, six proteins identified in spirulina extracts exhibited a significant sequence homology with one or more proteins from other organisms that have already been classified as food allergens. They should thus be carefully considered to evaluate the risk of allergenicity posed by the ingestion of spirulina-based products.

3.4 In silico evaluation of sequence homology between chlorella proteins and known allergenic

proteins.

As summarized in **Table S4 (**Supplementary Material), four proteins identified in *C. vulgaris* extracts exhibited aao > 80 and PID > 35% and, eventually, one or more sequences of at least six identical amino acids with respect to allergenic proteins listed in the AO database. The first one is calmodulin (A0A2P6TFR8), which is related to seven allergenic proteins, most of which corresponding to Troponin C identified in different organisms. In detail, six of these proteins were found to exhibit airway-mediated allergenicity and were identified in mold mite (D2DGW3, *Allergome* code: Tyr p 24)[59], in the American cockroach (Q1M0Y3, *Per a 6*) [60], in the pollen of common olive (Q9M7R0, *Ole e 8*) [46, 61] and in the German cockroach (three isoforms of Troponin C with accession numbers Q1A7B1, Q1A7B2 and Q1A7B3, corresponding to allergen *Bla g 6*) [62, 63].

The last protein reported as homologous to chlorella calmodulin was reported as a food allergen identified in brown shrimp (D7F1Q2, Cra c 6) [58], whose allergenicity is due to ingestion. Notably, E-scores not much lower than 10^{-7} were found for all the cited proteins and no sequence with at least six identical amino acids was found upon comparison with chlorella calmodulin.

The second chlorella protein reported in **Table S4** is fructose-bisphosphate aldolase that showed a relatively high aao (> 350 aa) and PID values greater than 48.5% with five allergens corresponding to the same protein in other organisms. In this case E-scores values were very low (< 10^{-67}) and 5 or 6 sequences of identical amino acids (including up to 11 aa) were recognized, thus suggesting the occurrence of a relevant sequence homology. Specifically, the fructose-bisphosphate aldolase with accession number A0A068FCL9, related to *Penaeus chinensis* (Chinese white shrimp) [64], is reported as an allergen in the *AO* but not in the *Allergome* database. Protein XP_026771637 belongs to the already cited *Pangasianodon hypophthalmus* (striped catfish) [55], and has been reported as a food allergen (*Pan h 3*). The remaining fructose-bisphosphate aldolases indicated among allergenic proteins in **Table S4** were identified as allergens of yellowfin tuna (D4HTS6) and Atlantic salmon

(B5DGM7 and I0J1J3) [65, 66].

Ribosome biogenesis brx1 (A0A2P6U528) was the third protein identified in *C. vulgaris* extracts that displayed significant homologies with allergen proteins, namely with thioredoxin proteins from organism of the fungi kingdom, some of which matched with thioredoxins already found in the case of spirulina, whose allergenicity is related to inhalation or contact with skin [67–69]. The last protein reported in **Table S4** for chlorella is the cytochrome c domain-containing protein (A0A6V2GHH7), for which an aao of 102 and PID of 58.8% was found with cytochrome c of *Curvularia lunata* (Q96VP3), a filamentous fungus belonging to the family of *Pleosporaceae*, yet this protein is not considered a food allergen (*Allergome* code *Cur l 3*) [70, 71]. Based on data reported in **Table S4**, fructose-bisphosphate aldolase (A0A2P6TDD0) can be considered the chlorella protein with the highest potential for cross reaction upon ingestion, due to its significant sequence homology with the same proteins of edible fish or crustaceans.

The proposed proteomics analysis of microalgae and *in silico* sequence homology prediction developed in this work expand upon the utility of LC-MS/MS to evaluate potentially allergenic proteins from poorly characterized sources. A summary of proteins identified in spirulina and chlorella protein extracts that showed significant homology with proteins listed in the Allergome and/or Allergen databases as allergens upon ingestion, and that might thus represent potential allergens of the two microalgae, is given in **Table 1**. This is the first characterization step of the possible hazards posed by ingestion of allergen food proteins that permits consumers to be sufficiently informed regarding the risks they are undertaking.

CONCLUSIONS

Some suspected proteins occurring in extracts of chlorella and spirulina microalgae, showing significant sequence homology with known food allergens, most of which associated with fish and

crustaceans, were annotated. Moreover, the existence of a recognized allergen, viz. the C-phycocyanin beta subunit was ascertained in spirulina samples. The high percentage of identity along with contiguous amino acid sequences of some algal proteins, marked as potential allergens, might expose sensitized subjects' health to serious risks. Although adequate clinical studies will be needed to assess the tangible allergenicity of these microalgae, stakeholders are sufficiently informed to make well-reasoned decisions. Further improvements to the proposed approach should make the discovery of allergens in novel foods more reliable and less time-consuming.

ACKNOWLEDGMENTS

This work was supported by projects: (i) PONa3_00395/1 “BIOSCIENZE & SALUTE (B&H)” and (ii) Progetto di Ricerca di Interesse Nazionale—PRIN 2017YER72K—“Development of novel DNA-based analytical platforms for the rapid, point-of-use quantification of multiple hidden allergens in food samples”, financed by the Italian Ministero per l’Istruzione, l’Università e la Ricerca (MIUR).

Author contributions

M.B.: investigation, data curation, writing- original draft preparation; G.V.: conceptualization, methodology, writing- original draft preparation; C.D.C.: supervision, writing- original draft preparation; I.L: writing- reviewing and editing; T.R.I.C.: writing- reviewing and editing, funding acquisition, resources.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

This article contains supporting information.

458 The Supplementary Material is available free of charge on the publication website.

459 Supplemental methods, tables, and figures, including materials, sample preparation, methods for
460 MS, supplemental figures for the paper, and tables with found proteins in both spirulina and
461 chlorella samples.

462 **REFERENCES**

463 1. Sicherer, S.H., Sampson, H.A.: Food allergy. J. Allergy Clin. Immunol. 125, S116–S125 (2010).
464 https://doi.org/10.1016/j.jaci.2009.08.028
465 2. The European Parliament and the Council of the European Union: Regulation (EU) 1169/2011. Off. J.
466 Eur. Union. 17, 18–63 (2011)
467 3. Hoffmann, B., Münch, S., Schwägele, F., Neusüß, C., Jira, W.: A sensitive HPLC-MS/MS screening
468 method for the simultaneous detection of lupine, pea, and soy proteins in meat products. Food
469 Control. 71, 200–209 (2017). https://doi.org/10.1016/j.foodcont.2016.06.021
470 4. Monaci, L., De Angelis, E., Montemurro, N., Pilolli, R.: Comprehensive overview and recent advances
471 in proteomics MS based methods for food allergens analysis, (2018)
472 5. Montowska, M., Fornal, E.: Detection of peptide markers of soy, milk and egg white allergenic
473 proteins in poultry products by LC-Q-TOF-MS/MS. LWT - Food Sci. Technol. 87, 310–317 (2018).
474 https://doi.org/10.1016/j.lwt.2017.08.091
475 6. Pilolli, R., Nitride, C., Gillard, N., Huet, A.C., van Poucke, C., de Loose, M., Tranquet, O., Larré, C.,
476 Adel-Patient, K., Bernard, H., Mills, E.N.C., Monaci, L.: Critical review on proteotypic peptide marker
477 tracing for six allergenic ingredients in incurred foods by mass spectrometry, (2020)
478 7. Pilolli, R., De Angelis, E., Monaci, L.: In house validation of a high resolution mass spectrometry
479 Orbitrap-based method for multiple allergen detection in a processed model food. Anal. Bioanal.
480 Chem. 410, 5653–5662 (2018). https://doi.org/10.1007/s00216-018-0927-8
481 8. Parker, C.H., Khuda, S.E., Pereira, M., Ross, M.M., Fu, T.J., Fan, X., Wu, Y., Williams, K.M., DeVries, J.,
482 Pulvermacher, B., Bedford, B., Zhang, X., Jackson, L.S.: Multi-allergen Quantitation and the Impact of
483 Thermal Treatment in Industry-Processed Baked Goods by ELISA and Liquid Chromatography-
484 Tandem Mass Spectrometry. J. Agric. Food Chem. 63, 10669–10680 (2015).
485 https://doi.org/10.1021/acs.jafc.5b04287
486 9. Planque, M., Arnould, T., Dieu, M., Delahaut, P., Renard, P., Gillard, N.: Advances in ultra-high
487 performance liquid chromatography coupled to tandem mass spectrometry for sensitive detection
488 of several food allergens in complex and processed foodstuffs. J. Chromatogr. A. 1464, 115–123
489 (2016). https://doi.org/10.1016/j.chroma.2016.08.033
490 10. Planque, M., Arnould, T., Gillard, N.: Food Allergen Analysis: Detection, Quantification and Validation
491 by Mass Spectrometry. Chapter 2. In: Allergen. pp. 7–41. InTech (2017)
492 11. Bianco, M., Calvano, C.D., Ventura, G., Losito, I., Cataldi, T.R.I.: Determination of hidden milk
493 allergens in meat-based foodstuffs by liquid chromatography coupled to electrospray ionization and
494 high-resolution tandem mass spectrometry. Food Control. 108443 (2021).
495 https://doi.org/10.1016/j.foodcont.2021.108443
496 12. Niccolai, A., Venturi, M., Galli, V., Pini, N., Rodolfi, L., Biondi, N., D'Ottavio, M., Batista, A.P.,

497 Raymundo, A., Granchi, L., Tredici, M.R.: Development of new microalgae-based sourdough
498 "crostini": functional effects of Arthrospira platensis (spirulina) addition. Sci. Rep. 9, 19433 (2019).
499 https://doi.org/10.1038/s41598-019-55840-1
500 13. Hashemian, M., Ahmadzadeh, H., Hosseini, M., Lyon, S., Pourianfar, H.R.: Production of Microalgae-
501 Derived High-Protein Biomass to Enhance Food for Animal Feedstock and Human Consumption. In:
502 Advanced Bioprocessing for Alternative Fuels, Biobased Chemicals, and Bioproducts. pp. 393–405.
503 Elsevier Inc. (2019)
504 14. Bleakley, S., Hayes, M.: Algal proteins: Extraction, application, and challenges concerning
505 production. Foods. 6, 1–34 (2017). https://doi.org/10.3390/foods6050033
506 15. Fernandez, A., Mills, E.N.N.C., Koning, F., Moreno, F.J.: Allergenicity Assessment of Novel Food
507 Proteins: What Should Be Improved? Trends Biotechnol. 39, 4–8 (2021).
508 https://doi.org/10.1016/j.tibtech.2020.05.011
509 16. Pali-Schöll, I., Verhoeckx, K., Mafra, I., Bavaro, S.L., Clare Mills, E.N., Monaci, L.: Allergenic and novel
510 food proteins: State of the art and challenges in the allergenicity assessment. Trends Food Sci.
511 Technol. 84, 45–48 (2019). https://doi.org/10.1016/j.tifs.2018.03.007
512 17. Ververis, E., Ackerl, R., Azzollini, D., Colombo, P.A., de Sesmaisons, A., Dumas, C., Fernandez-
513 Dumont, A., Ferreira da Costa, L., Germini, A., Goumperis, T., Kouloura, E., Matijevic, L., Precup, G.,
514 Roldan-Torres, R., Rossi, A., Svejstil, R., Turla, E., Gelbmann, W.: Novel foods in the European Union:
515 Scientific requirements and challenges of the risk assessment process by the European Food Safety
516 Authority. Food Res. Int. 137, 109515 (2020). https://doi.org/10.1016/j.foodres.2020.109515
517 18. Tiberg, E., Dreborg, S., Björkstén, B.: Allergy to green algae (Chlorella) among children. J. Allergy Clin.
518 Immunol. 96, 257–259 (1995). https://doi.org/10.1016/S0091-6749(95)70016-1
519 19. Tiberg, E., Rolfsen, W., Einarsson, R., Dreborg, S.: Detection of Chlorella-specific IgE in mould-
520 sensitized children. Allergy. 45, 481–486 (1990). https://doi.org/10.1111/j.1398-
521 9995.1990.tb00523.x
522 20. Tiberg, E., Rolfsen, W., Einarsson, R.: Preparation of allergen extracts from the green alga Chlorella:
523 Studies of growth variation, batch variation, and partial purification. Int. Arch. Allergy Immunol. 92,
524 23–29 (1990). https://doi.org/10.1159/000235219
525 21. Yim, H.E., Yoo, K.H., Seo, W.H., Won, N.H., Hong, Y.S., Lee, J.W.: Acute tubulointerstitial nephritis
526 following ingestion of Chlorella tablets. Pediatr. Nephrol. 22, 887–888 (2007).
527 https://doi.org/10.1007/s00467-006-0420-z
528 22. Petrus, M., Culerrier, R., Campistron, M., Barre, A., Rougé, P.: First case report of anaphylaxis to
529 spirulin: Identification of phycocyanin as responsible allergen. Allergy Eur. J. Allergy Clin. Immunol.
530 65, 924–925 (2010). https://doi.org/10.1111/j.1398-9995.2009.02257.x
531 23. van Hengel, A.J.: Food allergen detection methods and the challenge to protect food-allergic

consumers. Anal. Bioanal. Chem. 389, 111–118 (2007). https://doi.org/10.1007/s00216-007-1353-5

24. Frémont, S., Kanny, G., Bieber, S., Nicolas, J.P., Moneret-Vautrin, D.A.: Identification of a masked allergen, ?-lactalbumin, in baby-food cereal flour guaranteed free of cow's milk protein. Allergy. 51, 749–754 (1996). https://doi.org/10.1111/j.1398-9995.1996.tb02121.x
25. Scheibe, B., Weiss, W., Ruëff, F., Przybilla, B., Görg, A.: Detection of trace amounts of hidden allergens: Hazelnut and almond proteins in chocolate. J. Chromatogr. B Biomed. Sci. Appl. 756, 229–237 (2001). https://doi.org/10.1016/S0378-4347(01)00083-4
26. Holzhauser, T., Vieths, S.: Indirect competitive ELISA for determination of traces of peanut (Arachis hypogaea L.) protein in complex food matrices. J. Agric. Food Chem. 47, 603–611 (1999). https://doi.org/10.1021/jf980775f
27. Stephan, O., Möller, N., Lehmann, S., Holzhauser, T., Vieths, S.: Development and validation of two dipstick type immunoassays for determination of trace amounts of peanut and hazelnut in processed foods. Eur. Food Res. Technol. 215, 431–436 (2002). https://doi.org/10.1007/s00217-002-0562-6
28. Hirao, T., Imai, S., Sawada, H., Shiomi, N., Hachimura, S., Kato, H.: PCR method for detecting trace amounts of buckwheat (Fagopyrum spp.) in food. Biosci. Biotechnol. Biochem. 69, 724–731 (2005). https://doi.org/10.1271/bbb.69.724
29. Lin, J., Alcocer, M. eds: Food Allergens. Springer New York, New York, NY (2017)
30. Liu, C., Sathe, S.K.: Food Allergen Epitope Mapping. J. Agric. Food Chem. 66, 7238–7248 (2018). https://doi.org/10.1021/acs.jafc.8b01967
31. Rupa, P., Mine, Y.: Oral immunotherapy with immunodominant T-cell epitope peptides alleviates allergic reactions in a Balb/c mouse model of egg allergy. Allergy Eur. J. Allergy Clin. Immunol. 67, 74–82 (2012). https://doi.org/10.1111/j.1398-9995.2011.02724.x
32. Wai, C.Y.Y., Leung, N.Y.H., Leung, P.S.C., Chu, K.H.: Immunotherapy of Food Allergy: a Comprehensive Review. Clin. Rev. Allergy Immunol. 2017 571. 57, 55–73 (2017). https://doi.org/10.1007/S12016-017-8647-Y
33. Zhou, F., He, S., Sun, H., Wang, Y., Zhang, Y.: Advances in epitope mapping technologies for food protein allergens: A review, (2021)
34. Hopp, T.P., Woods, K.R.: Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. U. S. A. 78, 3824–3828 (1981). https://doi.org/10.1073/pnas.78.6.3824
35. FAO/WHO: FAO/WHO (2001) Allergenicity of genetically modified foods. Report of Joint FAO/WHO Expert Consultation on Foods Derived from Biotechnology. (2001)
36. Polikovsky, M., Fernand, F., Sack, M., Frey, W., Müller, G., Golberg, A.: In silico food allergenic risk evaluation of proteins extracted from macroalgae Ulva sp. with pulsed electric fields. Food Chem. 276, 735–744 (2019). https://doi.org/10.1016/j.foodchem.2018.09.134

37. Abdelmoteleb, M., Zhang, C., Furey, B., Kozubal, M., Griffiths, H., Champeaud, M., Goodman, R.E.: Evaluating potential risks of food allergy of novel food sources based on comparison of proteins predicted from genomes and compared to www.AllergenOnline.org. Food Chem. Toxicol. 147, 111888 (2021). https://doi.org/10.1016/j.fct.2020.111888

38. Goodman, R.E., Ebisawa, M., Ferreira, F., Sampson, H.A., van Ree, R., Vieths, S., Baumert, J.L., Bohle, B., Lalithambika, S., Wise, J., Taylor, S.L.: AllergenOnline: A peer-reviewed, curated allergen database to assess novel food proteins for potential cross-reactivity. Mol. Nutr. Food Res. 60, 1183–1198 (2016). https://doi.org/10.1002/mnfr.201500769

39. Michalski, A., Neuhauser, N., Cox, J., Mann, M.: A Systematic Investigation into the Nature of Tryptic HCD Spectra. J. Proteome Res. 11, 5479–5491 (2012). https://doi.org/10.1021/pr3007045

40. Pejchinovski, M., Klein, J., Ramírez-Torres, A., Bitsika, V., Mermelekas, G., Vlahou, A., Mullen, W., Mischak, H., Jankowski, V.: Comparison of higher energy collisional dissociation and collision-induced dissociation MS/MS sequencing methods for identification of naturally occurring peptides in human urine. Proteomics - Clin. Appl. 9, 531–542 (2015). https://doi.org/10.1002/prca.201400163

41. Weichel, M., Glaser, A.G., Ballmer-Weber, B.K., Schmid-Grendelmeier, P., Crameri, R.: Wheat and maize thioredoxins: A novel cross-reactive cereal allergen family related to baker's asthma. J. Allergy Clin. Immunol. 117, 676–681 (2006). https://doi.org/10.1016/j.jaci.2005.11.040

42. Fasoli, E., Pastorello, E.A., Farioli, L., Scibilia, J., Aldini, G., Carini, M., Marocco, A., Boschetti, E., Righetti, P.G.: Searching for allergens in maize kernels via proteomic tools. J. Proteomics. 72, 501–510 (2009). https://doi.org/10.1016/j.jprot.2009.01.013

43. Kespohl, S., Maryska, S., Zahradnik, E., Sander, I., Brüning, T., Raulf-Heimsoth, M.: Biochemical and immunological analysis of mould skin prick test solution: current status of standardization. Clin. Exp. Allergy. 43, 1286–1296 (2013). https://doi.org/10.1111/cea.12186

44. Shen, H. Der, Wang, C.W., Chou, H., Lin, W.L., Tam, M.F., Huang, M.H., Kuo, M.L., Wang, S.R., Han, S.H.: Complementary DNA cloning and immunologic characterization of a new Penicillium citrinum allergen (Pen c 3). J. Allergy Clin. Immunol. 105, 827–833 (2000). https://doi.org/10.1067/mai.2000.105220

45. Andersson, A., Scheynius, A., Rasool, O.: Detection of Mala f and Mala s allergen sequences within the genus Malassezia. Med. Mycol. 41, 479–485 (2003). https://doi.org/10.1080/13693780310001615367

46. Ivanciuc, O., Midoro-Horiuti, T., Schein, C.H., Xie, L., Hillman, G.R., Goldblum, R.M., Braun, W.: The property distance index PD predicts peptides that cross-react with IgE antibodies. Mol. Immunol. 46, 873–883 (2009). https://doi.org/10.1016/j.molimm.2008.09.004

47. Wagner, S., Sowka, S., Mayer, C., Crameri, R., Focke, M., Kurup, V.P., Scheiner, O., Breiteneder, H.: Identification of a *Hevea brasiliensis* Latex Manganese Superoxide Dismutase (Hev b 10) as a Cross-

Reactive Allergen. Int. Arch. Allergy Immunol. 125, 120–127 (2001). https://doi.org/10.1159/000053805

48. Chardin, H., Raulf-Heimsoth, M., Chen, Z., Rihs, H.P., Mayer, C., Desvaux, F.X., Sénéchal, H., Peltre, G.: Interest of Two-Dimensional Electrophoretic Analysis for the Characterization of the Individual Sensitization to Latex Allergens. Int. Arch. Allergy Immunol. 128, 195–203 (2002). https://doi.org/10.1159/000064252

49. D'Amato, A., Bachi, A., Fasoli, E., Boschetti, E., Peltre, G., Sénéchal, H., Sutra, J.P., Citterio, A., Righetti, P.G.: In-depth exploration of Hevea brasiliensis latex proteome and "hidden allergens" via combinatorial peptide ligand libraries. J. Proteomics. 73, 1368–1380 (2010). https://doi.org/10.1016/j.jprot.2010.03.002

50. Noorbakhsh, R., Mortazavi, S.A., Sankian, M., Shahidi, F., Assarehzadegan, M.A., Varasteh, A.R.: Cloning, expression, characterization, and computational approach for cross-reactivity prediction of manganese superoxide dismutase allergen from pistachio nut. Allergol. Int. 59, 295–304 (2010). https://doi.org/10.2332/allergolint.10-OA-0174

51. Vilhelmsson, M., Glaser, A.G., Martinez, D.B., Schmidt, M., Johansson, C., Rhyner, C., Berndt, K.D., Scheynius, A., Crameri, R., Achour, A., Zargari, A.: Mutational analysis of amino acid residues involved in IgE-binding to the Malassezia sympodialis allergen Mala s 11. Mol. Immunol. 46, 294–303 (2008). https://doi.org/10.1016/j.molimm.2008.07.036

52. Schwienbacher, M., Israel, L., Heesemann, J., Ebel, F.: Asp f6, an Aspergillus allergen specifically recognized by IgE from patients with allergic bronchopulmonary aspergillosis, is differentially expressed during germination. Allergy Eur. J. Allergy Clin. Immunol. 60, 1430–1435 (2005). https://doi.org/10.1111/j.1398-9995.2005.00904.x

53. Marti, P., Truffer, R., Stadler, M.B., Keller-Gautschi, E., Crameri, R., Mari, A., Schmid-Grendelmeier, P., Miescher, S.M., Stadler, B.M., Vogel, M.: Allergen motifs and the prediction of allergenicity. Immunol. Lett. 109, 47–55 (2007). https://doi.org/10.1016/j.imlet.2007.01.002

54. Sander, I., Rozynek, P., Rihs, H.P., Van Kampen, V., Chew, F.T., Lee, W.S., Kotschy-Lang, N., Merget, R., Brüning, T., Raulf-Heimsoth, M.: Multiple wheat flour allergens and cross-reactive carbohydrate determinants bind IgE in baker's asthma. Allergy Eur. J. Allergy Clin. Immunol. 66, 1208–1215 (2011). https://doi.org/10.1111/j.1398-9995.2011.02636.x

55. Ruethers, T., Taki, A.C., Karnaneedi, S., Nie, S., Kalic, T., Dai, D., Daduang, S., Leeming, M., Williamson, N.A., Breiteneder, H., Mehr, S.S., Kamath, S.D., Campbell, D.E., Lopata, A.L.: Expanding the allergen repertoire of salmon and catfish. Allergy Eur. J. Allergy Clin. Immunol. 76, 1443–1453 (2021). https://doi.org/10.1111/all.14574

56. An, S., Chen, L., Long, C., Liu, X., Xu, X., Lu, X., Rong, M., Liu, Z., Lai, R.: Dermatophagoides farinae allergens diversity identification by proteomics. Mol. Cell. Proteomics. 12, 1818–1828 (2013).

https://doi.org/10.1074/mcp.M112.027136

57. Lupi, R., Denery-Papini, S., Rogniaux, H., Lafiandra, D., Rizzi, C., De Carli, M., Moneret-Vautrin, D.A., Masci, S., Larré, C.: How much does transgenesis affect wheat allergenicity? Assessment in two GM lines over-expressing endogenous genes. J. Proteomics. 80, 281–291 (2013). https://doi.org/10.1016/j.jprot.2013.01.028

58. Bauermeister, K., Wangorsch, A., Garoffo, L.P., Reuter, A., Conti, A., Taylor, S.L., Lidholm, J., DeWitt, Å.M., Enrique, E., Vieths, S., Holzhauser, T., Ballmer-Weber, B., Reese, G.: Generation of a comprehensive panel of crustacean allergens from the North Sea Shrimp Crangon crangon. Mol. Immunol. 48, 1983–1992 (2011). https://doi.org/10.1016/j.molimm.2011.06.216

59. Jeong, K.Y., Kim, C.R., Un, S., Yi, M.H., Lee, I.Y., Park, J.W., Hong, C.S., Yong, T.S.: Allergenicity of recombinant troponin C from Tyrophagus putrescentiae. Int. Arch. Allergy Immunol. 151, 207–213 (2010). https://doi.org/10.1159/000242358

60. Lee, M.-F.F., Song, P.-P.P., Hwang, G.-Y.Y., Lin, S.-J.J., Chen, Y.-H.H.: Sensitization to per a 2 of the American cockroach correlates with more clinical severity among airway allergic patients in Taiwan. Ann. Allergy, Asthma Immunol. 108, 243–248 (2012). https://doi.org/10.1016/j.anai.2012.01.014

61. Ledesma, A., Villalba, M., Vivanco, F., Rodríguez, R.: Olive pollen allergen Ole e 8: identification in mature pollen and presence of Ole e 8-like proteins in different pollens. Allergy. 57, 40–43 (2002). https://doi.org/10.1046/j.0105-4538.2001.00001.x-i5

62. HINDLEY, J., WUNSCHMANN, S., SATINOVER, S., WOODFOLK, J., CHEW, F., CHAPMAN, M., POMES, A.: Bla g 6: A troponin C allergen from Blattella germanica with IgE binding calcium dependence. J. Allergy Clin. Immunol. 117, 1389–1395 (2006). https://doi.org/10.1016/j.jaci.2006.02.017

63. Oseroff, C., Sidney, J., Tripple, V., Grey, H., Wood, R., Broide, D.H., Greenbaum, J., Kolla, R., Peters, B., Pomés, A., Sette, A.: Analysis of T Cell Responses to the Major Allergens from German Cockroach: Epitope Specificity and Relationship to IgE Production. J. Immunol. 189, 679–688 (2012). https://doi.org/10.4049/jimmunol.1200694

64. Fu, L., Wang, J., Ni, S., Wang, C., Wang, Y.: Identification of Allergenic Epitopes and Critical Amino Acids of Major Allergens in Chinese Shrimp (Penaeus chinensis) by Immunoinformatics Coupled with Competitive-Binding Strategy. J. Agric. Food Chem. 66, 2944–2953 (2018). https://doi.org/10.1021/acs.jafc.7b06042

65. Kuehn, A., Hilger, C., Lehners-Weber, C., Codreanu-Morel, F., Morisset, M., Metz-Favre, C., Pauli, G., de Blay, F., Revets, D., Muller, C.P., Vogel, L., Vieths, S., Hentges, F.: Identification of enolases and aldolases as important fish allergens in cod, salmon and tuna: Component resolved diagnosis using parvalbumin and the new allergens. Clin. Exp. Allergy. 43, 811–822 (2013). https://doi.org/10.1111/cea.12117

66. Kuehn, A., Fischer, J., Hilger, C., Sparla, C., Biedermann, T., Hentges, F.: Correlation of clinical

672 monosensitivity to cod with specific IgE to enolase and aldolase. Ann. Allergy, Asthma Immunol. 113,
673 670-671.e2 (2014). https://doi.org/10.1016/j.anai.2014.09.005
674 67. Hoff, M., Ballmer-Weber, B.K., Niggemann, B., Cistero-Bahima, A., San Miguel-Moncín, M., Conti, A.,
675 Haustein, D., Vieths, S.: Molecular cloning and immunological characterisation of potential allergens
676 from the mould Fusarium culmorum. Mol. Immunol. 39, 965–975 (2003).
677 https://doi.org/10.1016/S0161-5890(03)00026-9
678 68. Glaser, A.G., Menz, G., Kirsch, A.I., Zeller, S., Crameri, R., Rhyner, C.: Auto- and cross-reactivity to
679 thioredoxin allergens in allergic bronchopulmonary aspergillosis. Allergy. 63, 1617–1623 (2008).
680 https://doi.org/10.1111/j.1398-9995.2008.01777.x
681 69. Limacher, A., Glaser, A.G., Meier, C., Schmid-Grendelmeier, P., Zeller, S., Scapozza, L., Crameri, R.:
682 Cross-Reactivity and 1.4-Å Crystal Structure of Malassezia sympodialis Thioredoxin (Mala s 13), a
683 Member of a New Pan-Allergen Family. J. Immunol. 178, 389–396 (2007).
684 https://doi.org/10.4049/jimmunol.178.1.389
685 70. Sharma, V., Singh, B.P., Gaur, S.N., Pasha, S., Arora, N.: Bioinformatics and Immunologic
686 Investigation on B and T Cell Epitopes of Cur l 3, a Major Allergen of Curvularia lunata. J. Proteome
687 Res. 8, 2650–2655 (2009). https://doi.org/10.1021/pr800784q
688 71. Sharma, V., Singh, B.P., Gaur, S.N., Arora, N.: Molecular and immunological characterization of
689 cytochrome c: A potential cross-reactive allergen in fungi and grasses. Allergy Eur. J. Allergy Clin.
690 Immunol. 63, 189–197 (2008). https://doi.org/10.1111/j.1398-9995.2007.01528.x
691 72. Yang, Y., Zhang, Y.-X.X., Liu, M., Maleki, S.J., Zhang, M.-L.L., Liu, Q.-M.M., Cao, M.-J.J., Su, W.-J.J., Liu,
692 G.-M.M.: Triosephosphate isomerase and filamin C share common epitopes as novel allergens of
693 Procambarus clarkii. J. Agric. Food Chem. 65, 950–963 (2017).
694 https://doi.org/10.1021/acs.jafc.6b04587
695
696

697 **Table 1.** Correspondence between proteins identified in spirulina and chlorella extracts and proteins
698 exerting allergenicity by ingestion reported in two major allergen databases.

Microalga	Protein Accession No.	Allergen Accession No.	Allergome code	Source of allergenicity (Allergome)	Source of allergenicity (Allergen.org)	Reference
Spirulina	P72508	P72508	Art pl beta_Phycocyanin	Ingestion	-	[22]
	D4ZSU6/K1VP15	Q4W1F7	Zea m 25 (Zea m 25.0101)	Ingestion	Airway	[41, 42]
	C3V6P3	B2BDZ8	Pis v 4 (Pis v 4.0101)	Ingestion	Food	[50]
	K1W168	A0A5N5Q6M7	Pan h 13 (Pan h 13.0101)	Ingestion	Food	[55]
	D5A635	A0A5N5Q6M9	Pan h 8	Ingestion	Food	[55]
		A0A1L5YRA2	Scy p 8 (Scy p 8.0101)	Ingestion	Food	[72]
		D7F1Q0	Cra c 8 (Cra c 8.0101)	Ingestion	Food	[58]
		F5A6E9	Pro c 8 (Pro c 8.0101)	Ingestion	Food	[72]
		B5DGL3	Sal s 8 (Sal s 8.0101)	Ingestion	Food	[55]
Chlorella	A0A2P6TFR8	D7F1Q2	Cra c 6 (Cra c 6.0101)	Ingestion	Food	[58]
	A0A2P6TDD0	A0A068FCL9[a]	-	-	-	[64]
		B5DGM7	Sal s 3 (Sal s 3.0101)	Ingestion	Food	[65, 66]
		XP_026771637	Pan h 3	Ingestion	Food	[55]
		I0J1J3	Sal s 3	Ingestion	Food	[65, 66]
		D4HTS6	Thu a 3	Ingestion	Food	[65, 66]

699 [a] Listed only in the AllergenOnline database.
700

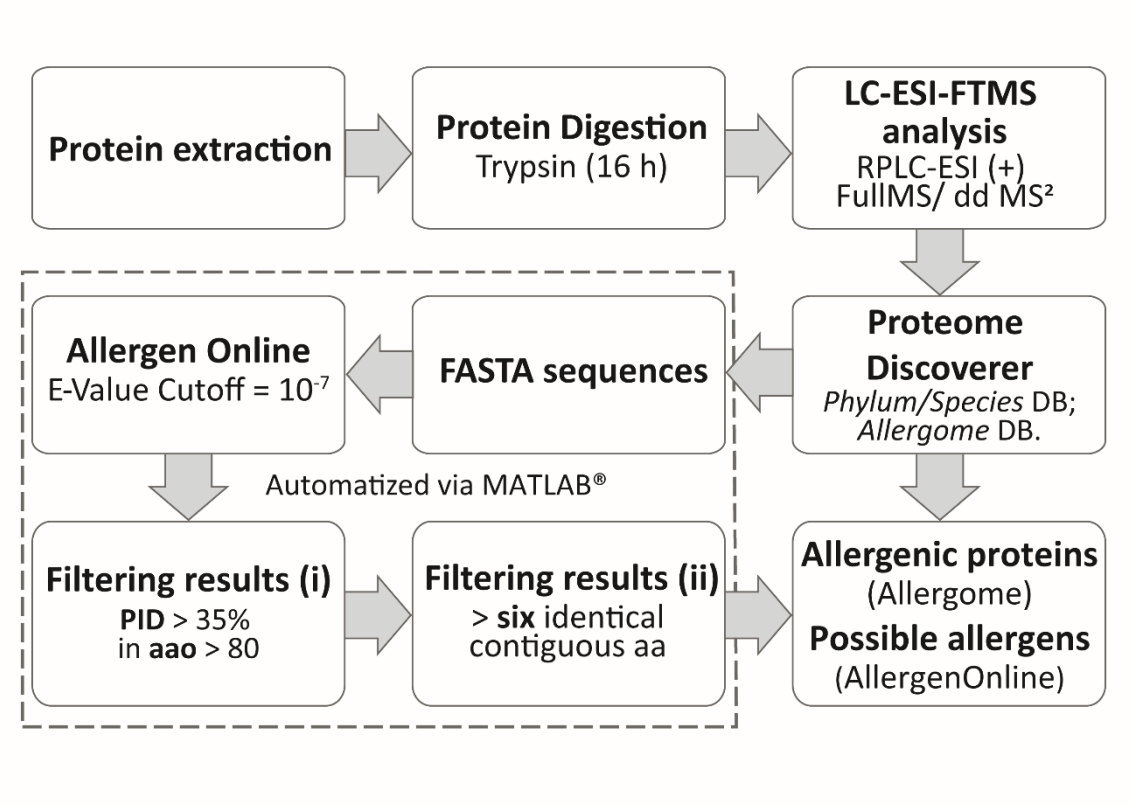


Figure 1. Workflow adopted to discover potentially allergenic proteins in the protein extracts of spirulina and chlorella microalgae.

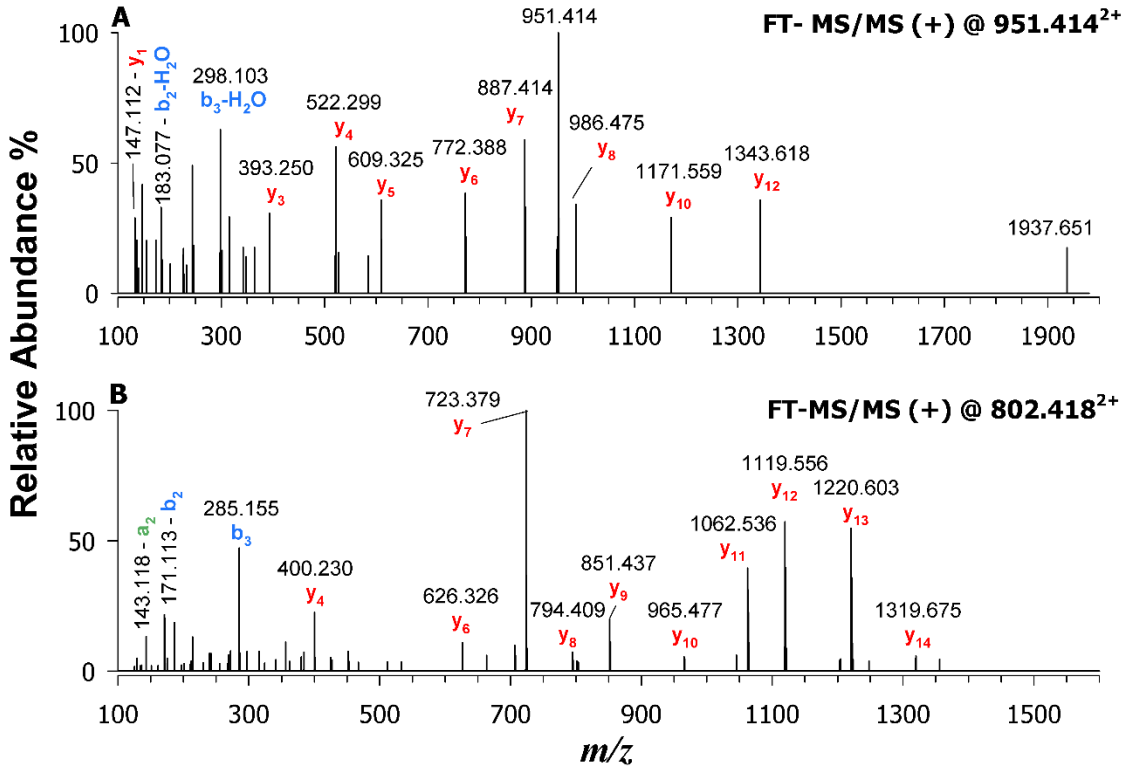


Figure 2. Tandem MS spectra by RPLC-ESI(+)-FTMS/MS of doubly charged peptide ions detected in the tryptic digest of chlorella microalga extract: (A) EADQDGDGQVDYSEFVK peptide of calmodulin; (B) AVNVTGPNGAPPEGAPR peptide of glycine-rich 2 protein.

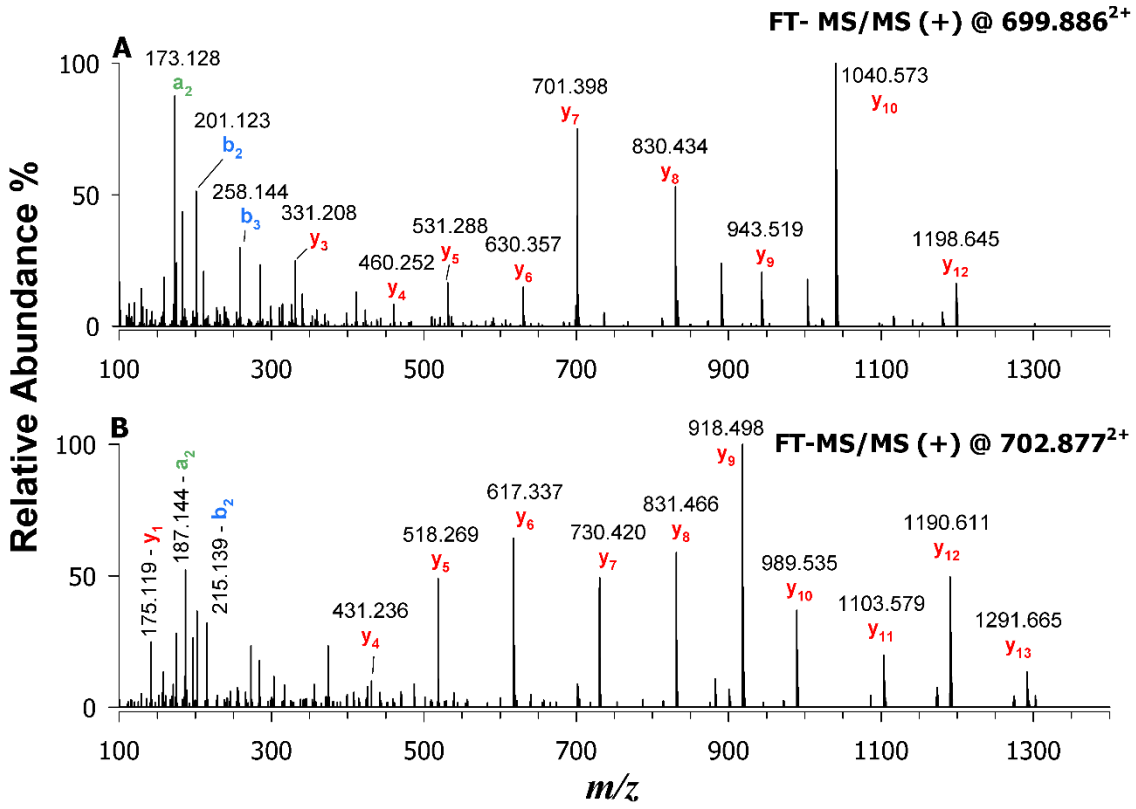


Figure 3 Tandem MS spectra by RPLC-ESI(+)-FTMS/MS of doubly charged peptide ions detected in the tryptic digest of spirulina microalga: (A) SLGTPIEAVAEGVR peptide of allophycocyanin alpha chain; (B) ITSNASTIVSNAAR peptide of C-phycocyanin beta subunit.

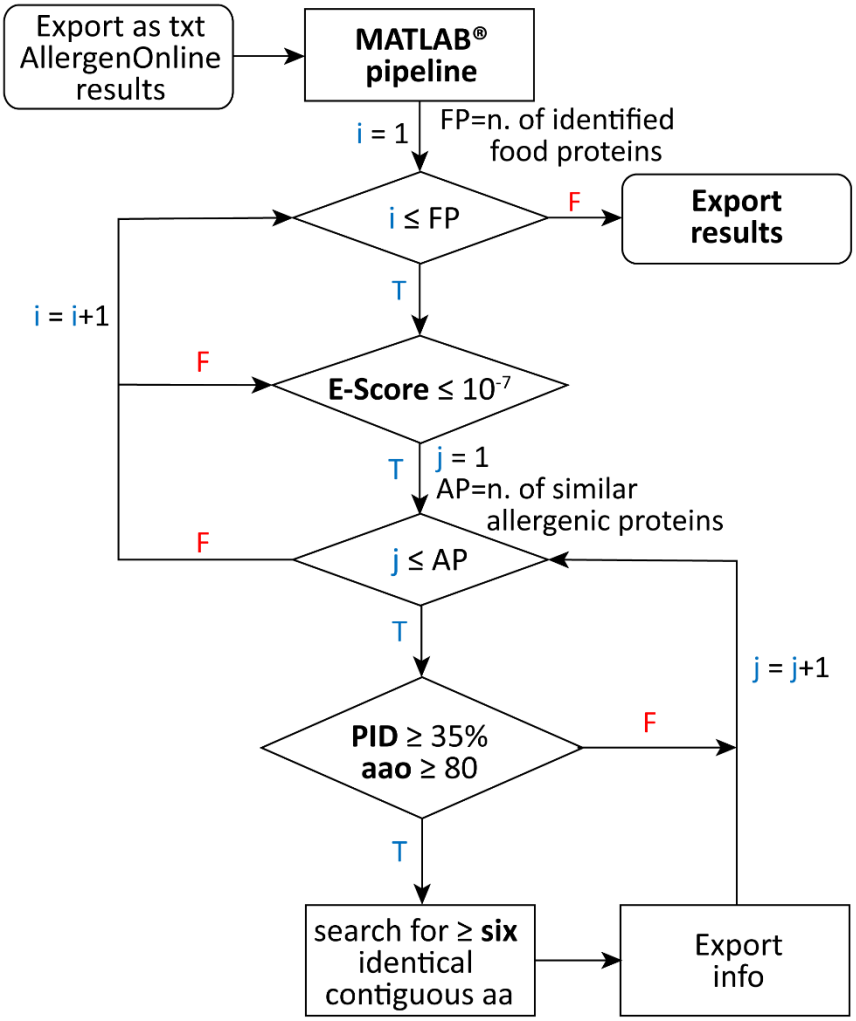
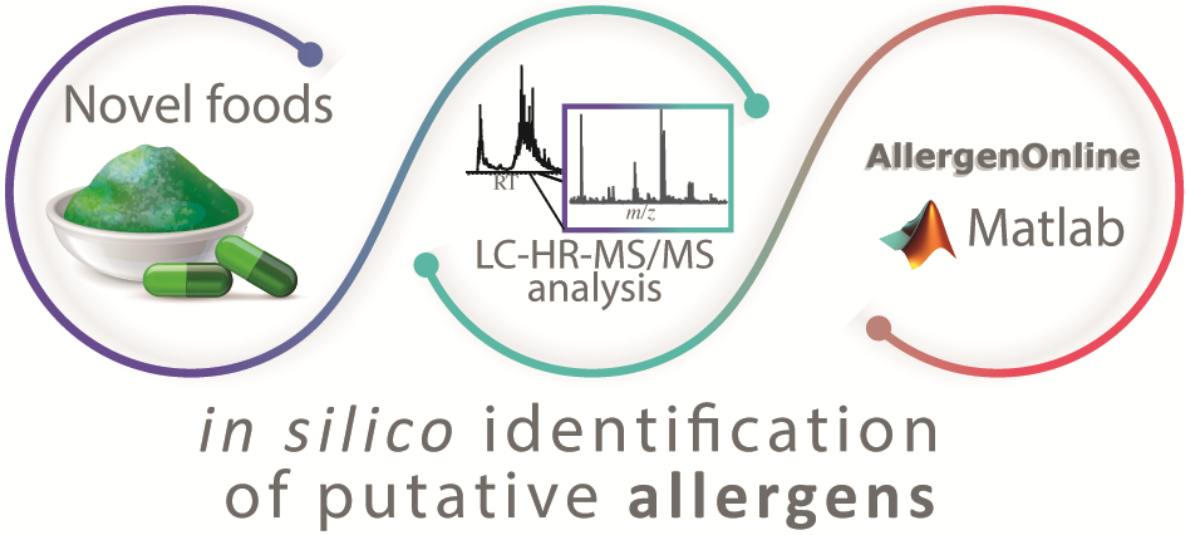


Figure 4. Flowchart describing the MATLAB® implemented workflow to select protein allergens from the AllergenOnline database exhibiting sequence homology with spirulina and chlorella proteins in accordance with WHO/FAO criteria. See text for details. T= True, F= false.

Graphical Abstract



Schematized workflow for in silico identification of potential Spirulina and Chlorella micro-algae allergens.