Journal of Proteomics 269 (2022) 104724

Contents lists available at ScienceDirect

Journal of Proteomics

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

Comparison of protein extraction protocols and allergen mapping from black soldier fly *Hermetia illucens*

Utpal Bose [a,b,1], James A. Broadbent [a,1], Angéla Juhász [b,1], Shaymaviswanathan Karnaneedi [c], Elecia B. Johnston [c], Sally Stockwell [a], Keren Byrne [a], Vachiranee Limviphuvadh [d,e], Sebastian Maurer-Stroh [d,e,f], Andreas L. Lopata [c,d,g], Michelle L. Colgrave [a,b,*]

[a] CSIRO Agriculture and Food, 306 Carmody Rd, St Lucia, QLD 4067, Australia
[b] Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science, School of Science, Edith Cowan University, Joondalup, WA 6027, Australia
[c] Molecular Allergy Research Laboratory, Australian Institute of Tropical Health and Medicine, James Cook University, Townsville, QLD, Australia
[d] Biomolecular Function Discovery Division, Bioinformatics Institute, Agency for Science, Technology and Research, Singapore
[e] IFCS Programme, Singapore Institute for Food and Biotechnology Innovation, Agency for Science, Technology and Research, Singapore
[f] Department of Biological Sciences, National University of Singapore, Singapore
[g] Tropical Futures Institute, James Cook University-, Singapore, Singapore

ARTICLE INFO

Keywords:
Insect
Black soldier fly
Proteomics
Extraction protocol
LC-MS/MS
LC-MRM-MS
Allergen
Tropomyosin

ABSTRACT

Exploration of important insect proteins — including allergens — and proteomes can be limited by protein extraction buffer selection and the complexity of the proteome. Herein, LC-MS/MS-based proteomics experiments were used to assess the protein extraction efficiencies for a suite of extraction buffers and the effect of ingredient processing on proteome and allergen detection. Discovery proteomics revealed that SDS-based buffer yields the maximum number of protein groups from three types of BSF samples. Bioinformatic analysis revealed that buffer composition and ingredient processing could influence allergen detection. Upon applying multi-level filtering criteria, 33 putative allergens were detected by comparing the detected BSF proteins to sequences from public allergen protein databases. A targeted LC-MRM-MS assay was developed for the pan-allergen tropomyosin and used to assess the influence of buffer composition and ingredient processing using peptide abundance measurements.

Significance: We demonstrated that the selection of protein extraction buffer and the processing method could influence protein yield and cross-reactive allergen detection from processed and un-processed black soldier fly (BSF) samples. In total, 33 putative allergens were detected by comparing the detected BSF proteins to sequences from public allergen protein databases. An LC-MRM-MS assay was developed for tropomyosin, indicating the importance of buffer selection and processing conditions to reduce BSF samples' allergenicity.

1. Introduction

Black soldier fly (BSF; *Hermetia illucens* [L.], Diptera: Stratiomyidae) has shown a promising outlook for growth at an industrial scale due to their ability to live on a wide range of organic waste residues. Their enriched nutritional composition and the possibility for use as food and feed makes them attractive as a component of the circular economic model [1]. In this respect, BSF larvae have been used for poultry [2],

Abbreivations: ANOVA, Analysis of Variance; BSF, Black Soldier Fly; DTT, Dithiothreitol; FDR, False Discovery Rate; GO, Gene Ontology; GRAVY, Grand Average Hydropathicity Index; Hsp, Heat Shock Protein; HCA, Hierarchical Clustering Analysis; IEDB, Immune Epitope Database and Analysis Resource; IUIS, International Union of Immunological Societies; JCU, James Cook University; LC-MRM-MS, Liquid Chromatography-Multiple Reaction Monitoring-Mass Spectrometry; LC-MS/MS, Liquid Chromatography with tandem Mass Spectrometry; µg, microgram; ORF, Open Reading Frame; pI, isoelectric point; PC, Principal Component; PCA, Principal Component Analysis; P, Protocol; SDS, Sodium Dodecyl Sulfate; SP-1, Screw Pressed-1; SP-2, Screw Pressed-2; TCA, Trichloroacetic Acid; TM, Tropomyosin; WHO, World Health Organisation..

* Corresponding author at: CSIRO Agriculture and Food, 306 Carmody Rd, St Lucia, QLD 4067, Australia.
E-mail address: michelle.colgrave@csiro.au (M.L. Colgrave).

[1] Authors contributed equally.

https://doi.org/10.1016/j.jprot.2022.104724
Received 26 March 2022; Received in revised form 25 May 2022; Accepted 6 September 2022
Available online 10 September 2022

aquaculture [3] and pet feed [4]. BSF larvae contain high-quality nutrition as they are rich in protein (37% to 63%) and have a similar amino acid profile to traditional soybean meal [5]. Additionally, BSF contains a high lipid content (15% to 49%), which can be isolated and used to prepare biodiesel. In contrast, the rest of the defatted meal could be a protein-rich source for the feed industry. Thus, feeding insects to livestock and fisheries may have a lower environmental impact that could appeal to environmentally conscious consumers and industries.

The extraction of proteins from any biological material poses challenges, and often a single method is insufficient to capture the diversity of the proteome. Mass-spectrometry (MS)-based proteomics to profile whole organisms or processed ingredients require that the samples be extracted with a suitable extraction buffer to maximise the proteome coverage of analytes. Several protein extraction buffer compositions and defatting and protein precipitation steps have been applied to obtain broad proteome coverage from BSF samples [6–9]. For instance, defatting with diethyl ether followed by protein extraction with Tris-HCl buffer (25 mM Tris-HCl pH 8.0) was used to extract proteins from BSF and mealworm [6,10]. Smets and co-workers have used sequential extraction protocols, i.e., lipid extraction followed by protein extraction using protein precipitation with 6.0 M HCl and solubilisation of the protein extracts with NaOH (pH 7.0) [7]. The complex composition of insects and the effect of processing add another layer of challenge in protein extraction that needs to be overcome to maximise protein identification and quantitation rates. As a result, a protein extraction protocol that can be easily adopted to differently processed insects with minimal cost and time is essential. Additionally, incomplete, unannotated and redundant genomic resources underscore a bottleneck for identifying proteins from non-model species, such as BSF.

Shellfish (shrimp, crab, crawfish and lobster) are the most frequent cause of adverse food allergic reactions in hypersensitive individuals [11]. Allergic reactions can also be triggered through cross-reactivity of the immune system with proteins from different food products that are structurally similar to the sensitising allergens [12]. For instance, people diagnosed with seafood allergies may experience allergic reactions following the consumption of edible insects [13–15]. The occurrence of allergies resulting from insect consumption may be based on common allergenic proteins (pan-allergens) between insects and crustaceans such as arthropod tropomyosin. The cross-reactivity of the immune system to different allergens occurs because of shared identical or similar IgE-antibody binding epitopes on the allergens. Tropomyosin is the significant protein present in arthropods as a cross-reactive target of IgE from crustacean-allergic patients [16]. Tropomyosins from three mealworm species showed 60–90% sequence similarities to other insect species and crustaceans [17]. A recent review demonstrated that IgE cross-reacts with insect tropomyosins from various edible and non-edible species and that thermal processing and digestion did not eliminate their allergenicity [12]. Thus, comprehensive studies are required to understand the diversity of insect-derived proteins as allergens to ensure food safety, particularly for patients with existing crustacean allergies.

Previous studies of BSF have aimed to measure insect proteins after extraction using a single protocol; this has left a gap in our knowledge of the extracted protein complement. Here, a liquid chromatography-mass spectrometry (LC-MS)-based discovery proteomics experiment was used to assess the efficiency of protein extraction buffers and the effect of ingredient processing. Next, a multiple reaction monitoring (MRM)-based targeted method was developed for the pan-allergen tropomyosin to identify the optimal protein extraction for quantification. Extensive bioinformatic analyses and epitope mapping were performed on the acquired discovery and targeted proteomics data to understand the physicochemical properties and functional classes enriched by each extraction process and how the peptides mapped to known IgE-binding epitopes.

2. Materials and methods

2.1. BSF samples

Three types of BSF ingredients were collected from Goterra (Canberra, Australia). The whole BSF sample was dehydrated at 65 °C for 8 h. The two types of screw pressed samples were prepared as per the whole BSF sample, followed by processing with a benchtop olive oil screw press instrument (KK Oil Press 20 F Universal, Nut Solutions, Australia). In brief, the screw press process was operated in cold press mode (no temperature applied with a fan in operation to cool the unit), wherein room temperature material was fed into the press until completely processed (SP1). The solid material was returned to the press inlet and pressed a second time until completely processed (SP2). The screw pressed-1 sample (SP-1) was prepared by screw pressing to contain less chitin and a cleaner product than sample screw pressed-2. The screw pressed-2 sample (SP-2) was prepared using the same instrument but contained more chitin.

2.2. Protein extraction

Details of extraction buffer preparation, composition and protein extraction methodologies have been described in our recent publication on cricket flour [18]. In brief, the buffer composition for the protein extraction protocols are as follows: Protocol 1 (P1): 100 mM Tris-HCl, 4% SDS, 50 mM dithiothreitol (DTT) (pH 7.6); Protocol 2 (P2): 200 mM Tris-HCl, 2 M urea, 50 mM DTT; Protocol 3 (P3): 20 mM Tris, 8 M urea, 2 M thiourea, 50 mM DTT; Protocol 4 (P4): 10 mM HEPES, 6 M urea, 2 M thiourea, 50 mM DTT (pH 8.0); Protocol 5 (P5): 50 mM Tris-HCl, 50 mM DTT, 100 mM KCl; Protocol 6 (P6): defatting with n-pentane followed by extraction with 100 mM Tris-HCl, 8 M urea, 50 mM DTT (pH 8.5); Protocol 7 (P7): 50 mM Tris-HCl (pH 7.2); Protocol 8 (P8): 50 mM Tris-HCl, 6 M Urea (pH 7.2). All reagents were purchased from Sigma Aldrich (Castle Hill, Australia) and buffers were prepared on the day of the extraction. Protein extraction protocols (P1-P6) were performed at the Commonwealth Scientific and Industrial Research Organisation (CSIRO) research laboratories in Brisbane, Australia. Protein extraction protocols P7 and P8 were performed at the James Cook University (JCU; Townsville, Australia) and then transferred to CSIRO laboratories in Brisbane to perform protein digestion and LC-MS data acquisition.

The abovementioned protein extraction buffers (200 μL; 10 μL/mg) were added to respective samples in 1.5 mL micro-tubes followed by vortex mixing until the powder was mixed thoroughly with the buffer solution. Sample tubes were then placed in a sonication bath for 5 min. For P1, sample tubes were placed on a thermomixer (Eppendorf, Germany) at 600 rpm; 55 °C, 30 min. For the remaining protocols (P2-P5), sample tubes were placed on a thermomixer (600 rpm, room temperature, 30 min) before centrifugation for 15 min at 20,800 ×*g*. Supernatants were used for subsequent processing.

Where required (P6), flour samples were first defatted with 200 μL (*w/v*) of n-pentane prior to placing them on a thermomixer. In brief, ~20 mg aliquots of BSF flour were weighed into 1.5 mL micro-tubes and mixed with 200 μL of n-pentane. The tubes were placed on a rotator for 15 min, followed by centrifugation (5 min) at room temperature at 20,800 ×*g*. The supernatant was discarded, and the process was repeated two more times. The pellets were air-dried and redissolved in 200 μL of extraction buffer.

For samples prepared at JCU, BSF flour was weighed and added to the P7 buffer (5 mL per gram sample). Sample tubes were incubated for 16 h at 4 °C with a tumbling motion followed by centrifugation at 10,000 ×*g* for 30 min at 4 °C before collecting the supernatant. The pellet was immersed and homogenised in P8 buffer and incubated, centrifuged, and the supernatant collected as per P7. Extracts of P7 and P8 were filtered using glass fibre filters (Sartorius Stedim Biotech, Germany).

2.3. Protein concentration estimation by Bradford assay

Protein concentration estimation of the extracts (P1-P6) was performed using a Bradford colorimetric assay. Samples were diluted 10×, 20× and 40× with water and standards were generated with BSA ranging from 7.8 μg/mL to 500 μg/mL. Sample dilutions and standards (10 μL) were added to 200 μL of Bradford Reagent (BioRad: 1:5 in H_2O), incubated for 5 min at RT and read at 595 nm.

Protein concentration of P7 and P8 extracts were estimated in triplicates using the Pierce™ BCA (bicinchoninic acid) Protein Assay kit (ThermoFisher Scientific, USA). Samples were diluted 10× with buffer and protein concentration estimation was conducted as per the manufacturer's microwell plate protocol with BSA as the protein standard ranging from 125 μg/mL to 2000 μg/mL.

2.4. Protein digestion

BSF protein extracts (100 μg) were transferred to 30 kDa molecular weight cut-off (MWCO) filters (Merck Millipore, Bayswater, Vic). Protein digestion steps were described in detail previously [18–20].

2.5. Global proteome measurement

Proteins extracted with eight different extraction protocols were analysed by LC-MS as described previously [18] with chromatographic separation using an Ekspert nanoLC415 (Eksigent, Dublin, CA, USA) coupled to a TripleTOF 6600 MS (SCIEX, Redwood City, CA, USA). Chromatographic gradient separation and mass spectrometry parameters were described previously [18].

2.6. Proteome database creation and protein identification

The BSF database was built by combining protein sequences from UniProt Diptera (accessed on 07/04/2020), translated Open Reading Frame (ORF) sequences of two available BSF genome assemblies (Genbank Assembly accessions GCA_001014895.1 and GCA_009835165.1 coded as G1 and G2) downloaded from NCBI Genbank and IUIS reference allergens for crustaceans, house mite and cockroach. These protein sequences were merged with the common repository of adventitious proteins (cRAP) sequences and the iRT pseudo-protein sequence (in total, 523,909 sequences). ProteinPilot v5.0.3 software (SCIEX) incorporating the Paragon and ProGroup algorithms was used for protein identification. The MS/MS data were searched against the above-mentioned protein database using previously described parameters [18]. The protein identification results indicate that the Tris-HCl-KCl-based protocol (P5) yielded the lowest number of proteins and was excluded from downstream comparative analysis.

A database search combining all raw BSF sample data was performed to detect all tropomyosin proteoforms. The resulting output file was searched thoroughly to identify tropomyosin sequences that fall within a 1% global FDR threshold.

2.7. Physicochemical property, multivariate and bioinformatic analyses

A custom Python script was used to determine the Grand Average Hydropathicity Index (GRAVY), aromaticity, isoelectric point (pI) and molecular weight for each protein extracted with the top five protocols. Groups were compared using ordinary one-way ANOVA and Tukey's post hoc test.

Principal Component Analysis (PCA) was performed using Biovinci version 1.1.5 (BioTutoring Inc., San Diego, California, USA), and the quantitative measurements of tropomyosins were visualised using Morpheus R package (Broad Institute, Cambridge MA, USA; https://software.broadinstitute.org/morpheus/). Gene Ontology (GO) and InterPro domain mapping were performed to obtain functional classification data using OmicsBox software version 2.0.24 (BioBam Bioinformatics). Insect proteins downloaded from UniProt (8,082,118 proteins; accessed on 05/06/2020) were used as a background for the BLAST2GO analysis.

2.8. Mapping allergen proteins identified in BSF samples

The database search results comprising all BSF samples extracted with all extraction methods were searched against the Database of Allergen Families (AllFam, Medical University Vienna, Austria; http://www-old.meduniwien.ac.at/allfam) to identify allergen domain-containing proteins. Proteins detected at a sub-1% FDR threshold from the database were also analysed using the AllerCatPro 1.7 web tool (https://allercatpro.bii.a-star.edu.sg/index17.html) to investigate their putative allergenicity using both linear and 3D structural epitope windows (Maurer-Stroh et al., 2019). The heatmap using the pairwise sequence similarity values for the alignment of reference allergens and the detected BSF allergens was generated using Morpheus software (Broad Institute). The non-redundant list of monitored tropomyosin peptides was searched against the Immune Epitope Database and Analysis Resource (IEDB) using the online version (https://www.iedb.org). Peptide sequences were searched for epitope matches (exact and BLAST 90, 80 and 70% matches) while keeping other parameters as default.

2.9. Relative quantitation of tropomyosin by MRM

Detected tropomyosin proteoforms were imported to Skyline software to identify suitable MRM transitions [21]. In silico digestion yielded 32 peptides representing six tropomyosin proteoforms. These were used initially to acquire MRM data from a pooled sample (injected three times to evaluate the reproducibility) consisting of all the BSF protein extracts. The results from these analyses were used to refine the transitions and schedule retention times. Peptides were selected that yielded intense peaks and were fully tryptic, with no variable modifications or missed cleavages.

3. Results

3.1. Detection of proteins according to extraction buffer and sample processing

A total of 474 proteins were detected at 1% FDR from the three types of BSF samples extracted with the eight extraction buffers. Each data file was searched individually to investigate the total number of proteins and peptides detected at 1% global FDR across the experimental conditions (Table S1). The SDS/Tris-HCl (P1) buffer yielded the maximum number of proteins and peptides from three types of BSF samples. The urea/thiourea buffer (P3) yielded the second-highest number of proteins and peptides. The Tris-HCl-based buffers (P2 and P7) yielded the lowest number of proteins and peptides from the three types of BSF samples. For instance, the Tris-HCl-based buffer (P7) yielded ~10 times fewer proteins and peptides than the P1 buffer for whole BSF samples. Notably, the addition of a defatting step before extraction with urea buffer (P6) yielded the highest peptides per protein ratio compared to other extraction buffers tested in the present study (Table S1). For example, the protein and peptide ratio for the whole BSF sample extracted with SDS/Tris-HCl (P1) was 4.65; whilst the protein and peptide ratio for the whole BSF sample extracted with defatting + urea (P6) was 4.72. Overall, the SDS/Tris-HCl (P1) buffer yielded more protein identifications than other buffers; however, the peptide to protein ratio was higher for the extraction protocol with defatting prior to urea-based extraction (P6).

3.2. Whole BSF

To explore the proteome diversity within the whole BSF sample, the five top-performing extraction buffer compositions (P1-P4 and P6) were

selected based on the protein yield estimation results by Bradford assay and discovery proteomics experiments. Hierarchical Clustering Analysis (HCA) of the unused protein scores (a measure of the total, unique peptide evidence related to a given protein) was performed to investigate the relationships between the extraction protocols (Fig. 1A). Using the unused protein score P1 and P3 clustered together. While P2 completely separates from the other extraction buffers; however, it is more closely related to P4 and P6. Next, to understand how each protocol affects the proteome identification yield from whole BSF, a comparative analysis was performed on the lists of proteins generated using the SCIEX Protein Alignment template. An UpSet plot shows the presence of 413 unique proteins identified after extraction with five solvents where 42 (~10%) proteins were co-extracted with all buffers (Fig. 1B). The extraction buffer P1 yielded the maximum number of total and unique proteins (310 or 75% and 49 or 12%), respectively. P2 yielded the lowest number of total (75; 18%) and unique (6; ~1.5%) proteins, respectively. Peptides detected at 1% FDR were aligned across the five extraction protocols to understand their representation within each extraction method. In total, 2997 distinct peptides were detected from the whole BSF sample where only 112 peptides (~4%) were commonly identified by all five protocols (Fig. 1C). In concordance with protein identification, P1 yielded the maximum number of unique peptides 562 (19%), while P2 yielded the lowest number of unique peptides 71 (2.4%). The physiochemical parameters such as GRAVY, aromaticity, pI and molecular weight were also found to be different based on the proteins detected from individual protocol (Fig. S1A—D).

To understand the types of protein functional classes extracted with each extraction protocol, GO analysis was performed on each unified protein identification set (Fig. 1D). Metal ion binding, oxidoreductase activity, protein binding, transferase activity and ATP binding activities were the top five molecular functions resulting from (P1, P3, P4 and P6). Though P2 yielded the lowest number of proteins in comparison to other extraction protocols, molecular functions such as calcium ion binding and transporter activities were uniquely represented within this set of detected proteins.

3.3. BSF- Screw press-1

Hierarchical clustering analysis was used to understand the similarity in proteome measurements using the unused protein scores from unified protein accessions. Similar to the whole BSF sample, the P1 buffer extraction from SP-1 yielded protein unused scores separate from other protocols (Fig. 2A). Comparative analysis of the proteins detected from the SP-1 sample reveals the presence of 344 proteins, which was <20% less than whole BSF samples (Fig. 2B). Extraction using P1 yielded the maximum number of proteins (261; 63%), while P2 yielded the lowest frequency of proteins (49; 11.86%) (Fig. 2B). All five protocols co-extracted (34; 8%) proteins, while P1 yielded the highest number of unique proteins (70; 17%) among all the protocols tested. The alignment of peptides from SP-1 extraction protocols shows the presence of 2859

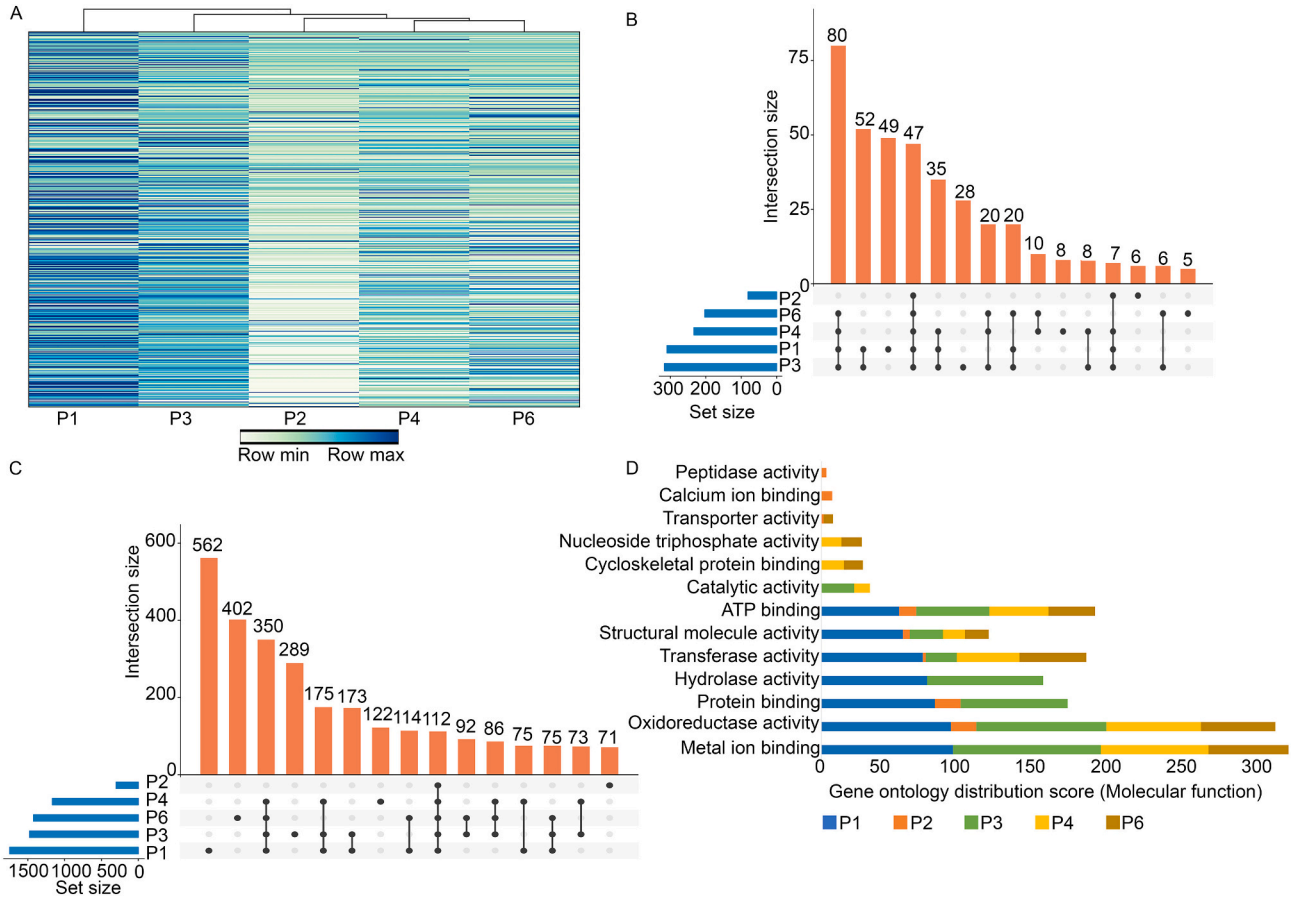

Fig. 1. Protein and peptide identifications (1% FDR) from whole BSF extracted with five extraction protocols. (A) Heatmap and HCA showing the unused protein scores of unified proteins detected following five extraction protocols: (P1) 100 mM Tris-HCl, 4% SDS, 50 mM dithiothreitol (DTT); (P2) 200 mM Tris-HCl, 2 M urea, 50 mM DTT; (P3) 20 mM Tris, 8 M urea, 2 M thiourea, 50 mM DTT; (P4) 10 mM HEPES, 6 M urea, 2 M thiourea, 50 mM DTT; (P6) Defatted with n-pentane followed by 100 mM Tris-HCl, 8 M urea, 50 mM DTT solubilisation. (B) UpSet plot of intersections between sets of proteins extracted from five extraction protocols (P1-P4 and P6), sorted by size. (C) UpSet plot of intersections between sets of peptides extracted from five extraction protocols (P1-P4 and P6), sorted by size. Dark dots in the matrix indicate sets that are part of the intersection. The bar plot at the top indicates the total number of shared protein accessions for those intersections. (D) Representative GO terms (molecular function; based on the node or score distribution) obtained from whole BSF proteins.

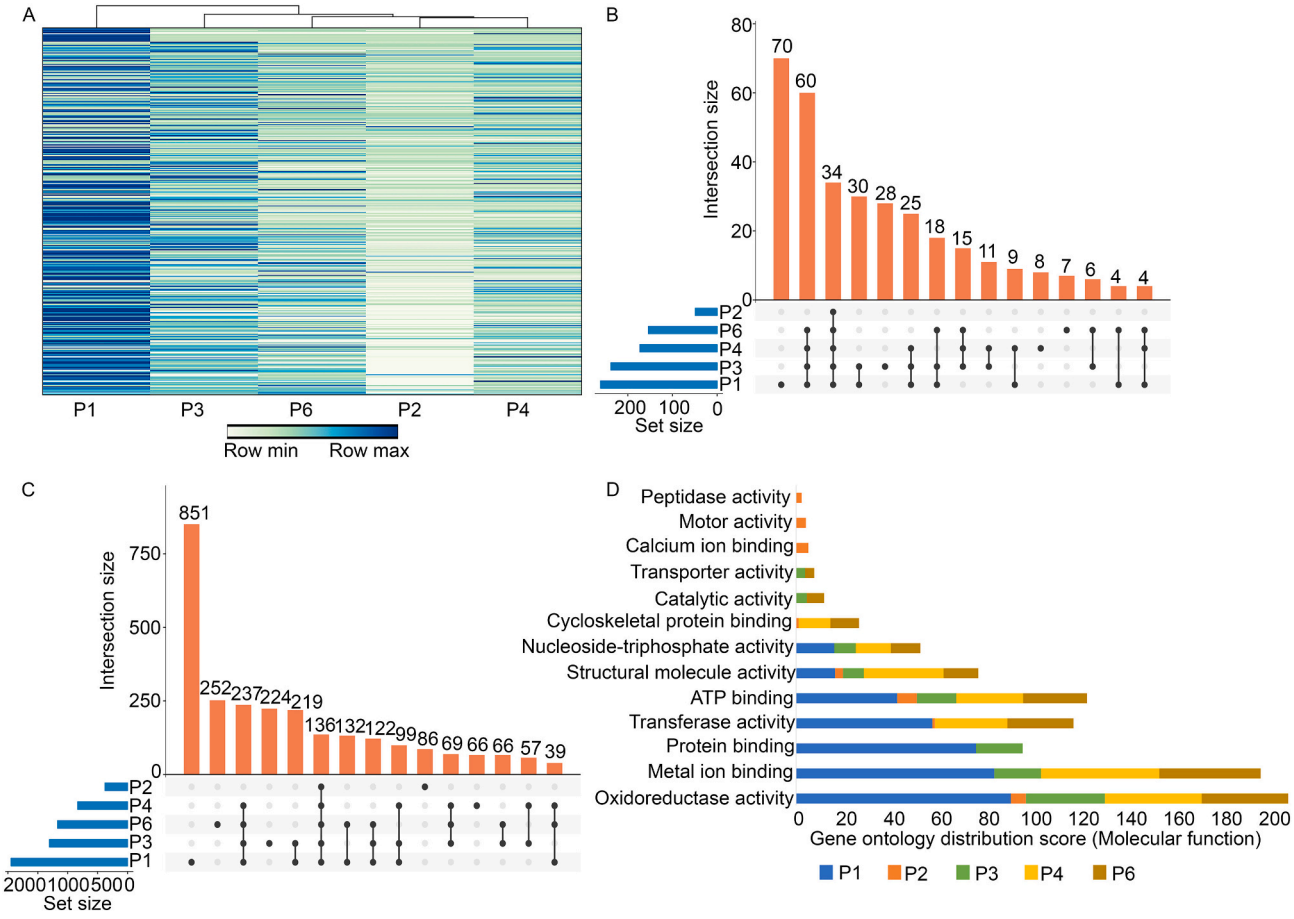

Fig. 2. Protein and peptide identifications (1% FDR) from screw pressed 1 (SP-1) BSF samples extracted with five extraction protocols. (A) Heatmap and HCA showing the unused protein scores of unified proteins detected following five extraction protocols: (P1) 100 mM Tris-HCl, 4% SDS, 50 mM dithiothreitol (DTT); (P2) 200 mM Tris-HCl, 2 M urea, 50 mM DTT; (P3) 20 mM Tris, 8 M urea, 2 M thiourea, 50 mM DTT; (P4) 10 mM HEPES, 6 M urea, 2 M thiourea, 50 mM DTT; (P6) Defatted with n-pentane followed by 100 mM Tris-HCl, 8 M urea, 50 mM DTT solubilisation. (B) UpSet plot of intersections between sets of proteins extracted using five extraction protocols (P1-P4 and P6), sorted by size. (C) UpSet plot of intersections between sets of peptides extracted from five extraction protocols (P1-P4 and P6), sorted by size. Dark dots in the matrix indicate sets that are part of the intersection. The bar plot at the top indicates the total number of shared protein accessions for those cluster intersections. (D) Representative GO terms (molecular function; based on the node or score distribution) obtained from SP-1 BSF proteins extracted with five protocols.

unique peptides, where (136; 5%) peptides were commonly extracted with all five buffers (Fig. 2C). In concordance with protein identification, P1 yielded the maximum number of total (1947; 68%) and unique (850; 30%) peptides, respectively. Although the lowest protein identifications were detected from P2, P4 yielded the least unique peptides (65; 2%). Similar to the whole BSF sample, extraction protocol-dependent physiochemical property alterations were also detected (Fig. S2A—D).

The functional annotation of proteins detected from each extraction protocol from SP-1 shows that oxidoreductase activity, metal ion binding, protein binding, transferase activity and ATP binding capacities were the top functional groups associated with the proteins (Fig. 2D). Although the number of protein and peptides were the lowest when using P2, this extraction buffer extracted proteins associated with unique molecular functions such as calcium ion binding, motor activity and peptidase activities.

3.4. BSF- Screw press-2

A comparative analysis was performed on the BSF SP-2 samples extracted with five buffers (P1-P4 and P6) to investigate the effect of extraction buffers on processed BSF samples. The HCA plot generated from the unused protein scores from five extraction protocols have shown P1 and P3 separates from the other three extraction buffers (Fig. 3A). In total, 316 proteins were detected from BSF SP-2 using the protein alignment template. Unlike whole BSF and BSF SP-1, P3 yielded the maximum number of proteins (226) in comparison to the other extraction buffers. The comparative analysis of proteins extracted with different extraction buffers indicated that 21 (7%) proteins were commonly extracted using the five extraction protocols (Fig. 3B). Protocol-3 yielded the maximum number of unique proteins 28 (9%), while the P2 yielded the lowest number of proteins 2 (<1%). The peptide level comparisons showed the presence of 2997 peptides across the five extraction protocols, where P1 and P3 yield 1746 and 1476 peptides, respectively (Fig. 3C). Although P3 yielded the maximum number of unique proteins, P1 yielded the maximum number of unique (562; 19%) peptides. Notably, P4 delivered the second-highest number of unique peptides (402; 13%), highlighting the influence of protein extraction methods on protein and peptide detection. Akin to whole BSF and SP-1, physicochemical properties were calculated for BSF SP-2 samples (Fig. S3A—D).

Functional characterisation was performed on the detected proteins resulting from each extraction buffer (Fig. 3D). The top five molecular functions: oxidoreductase activity, protein binding, metal ion binding, hydrolase activity and transferase activities were associated with all five extraction protocols. Notably, the P2 yields unique functional classes

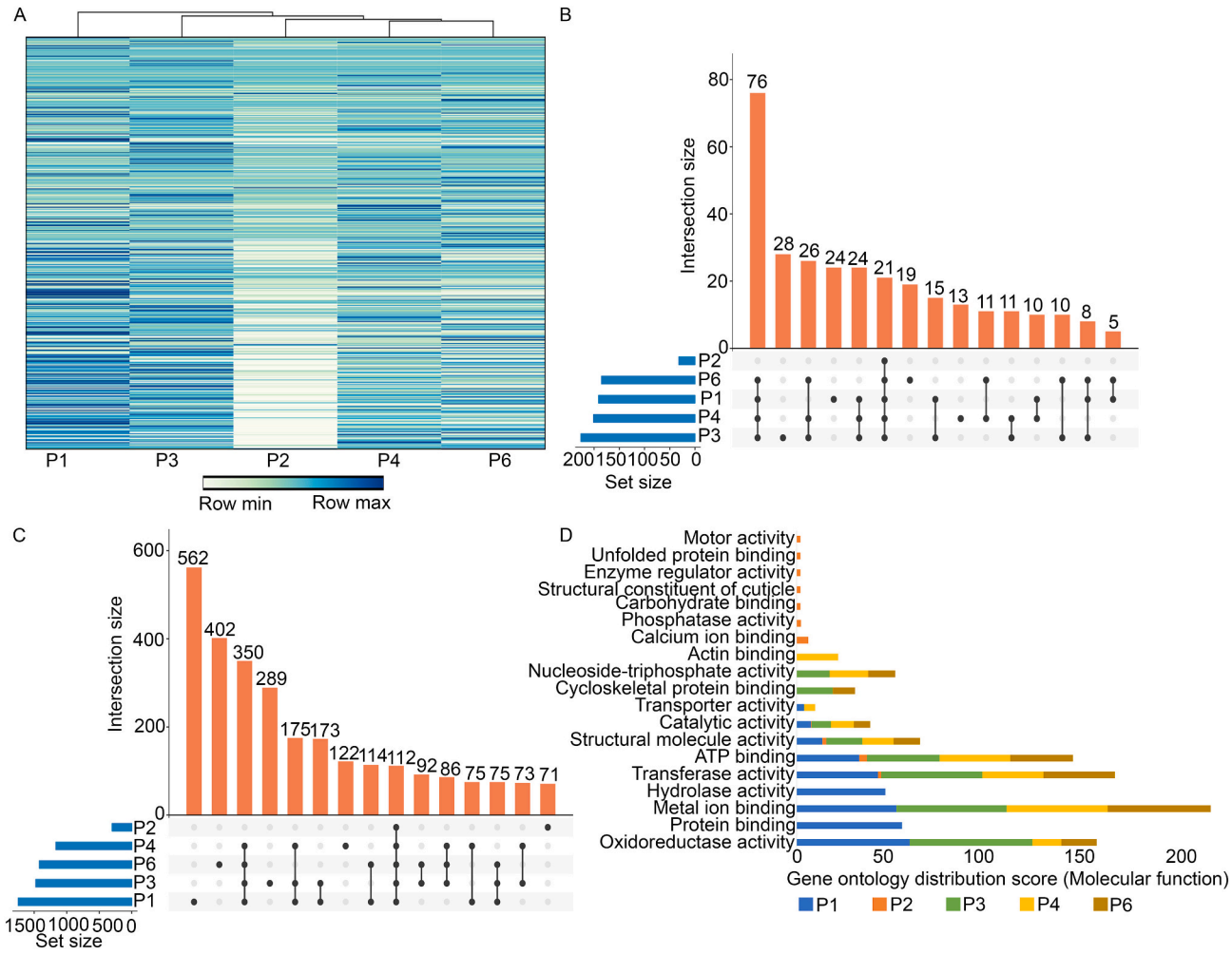

Fig. 3. Protein and peptide identifications (1% FDR) from screw pressed 2 (SP-2) BSF samples extracted with five extraction protocols. (A) Heatmap and HCA showing the unused protein scores of unified proteins detected following five extraction protocols: (P1) 100 mM Tris-HCl, 4% SDS, 50 mM dithiothreitol (DTT); (P2) 200 mM Tris-HCl, 2 M urea, 50 mM DTT; (P3) 20 mM Tris, 8 M urea, 2 M thiourea, 50 mM DTT; (P4) 10 mM HEPES, 6 M urea, 2 M thiourea, 50 mM DTT; (P6) Defatted with n-pentane followed by 100 mM Tris-HCl, 8 M urea, 50 mM DTT solubilisation. (B) UpSet plot of intersections between sets of proteins extracted from five extraction protocols (P1-P4 and P6), sorted by size. (C) UpSet plot of intersections between sets of peptides extracted from five extraction protocols (P1-P4 and P6), sorted by size. Dark dots in the matrix indicate sets that are part of the intersection. The bar plot at the top indicates the total number of shared protein accessions for those cluster intersections. (D) Representative GO terms (molecular function; based on the node or score distribution) obtained from SP-2 BSF proteins extracted with five protocols.

such as calcium ion binding, phosphatase activity, carbohydrate-binding, structural constituent of cuticle, enzyme regulator activity, unfolded protein binding and motor activities from BSF SP-2 samples.

3.5. Identification of tentative and putative allergens in the detected BSF proteins

A combined database search was performed using all the raw data acquired from discovery proteomics to identify the potential cross-reactive allergenic IgE epitopes present in the BSF proteome. Detected proteins were searched using the AllerCatPro webserver to detect the tentative cross-reactive allergens and IgE epitopes [22]. In total, 84 tentative allergens were detected using the AllerCatPro webserver with strong evidence and were used for downstream allergen mapping with the AllFam database. The allergen evidence was calculated based on the percent identity using a linear 80 amino acid window (>35%) and/or percent identity 3D epitope for the protein (>93%). To this end, 84 potential allergenic protein sequences — representing 42 allergen families — were detected, of which 20 (24%) proteins represented sequences similar to three sub-types of mite allergens, including, group 2 mite allergen [AllFam ID: AF111], group 7 mite allergen [AF195], group 5/21 mite allergen [AF156] and TM [AF054] (20 proteoforms; 24%) (Fig. S5). Notably, the majority of the TM-like proteins from BSF show high sequence similarities to 16 crustacean allergen isoforms, while five TMs were more similar to known cockroach and house dust mite allergens. Eight EF-hand proteins [AF007] detected from BSF samples were mapped to the AllFam database wherein only one protein mapped to the cockroach allergen Bla g 8. In comparison, seven proteins were aligned to crustacean allergens.

To further refine the list of putative allergens detected within BSF samples, the 84 tentatively detected allergens using AllerCatPro were sorted based on the percent identity in a linear 80 amino acid window (≥85%) and the evidence for percent identity to the 3D epitope (≥94%). The use of strict selection criteria revealed the presence of 33 putative allergens in BSF (Fig. 4). Using these stringent criteria, 6 proteins were detected as TMs from BSF samples mapped to multiple crustacean allergens. Taken together, mining of mass spectrometry data using in silico tools led to the detection of 33 proteins that contain high allergenic potential; however, future IgE-binding assays on crustacean allergic patients' sera would be essential to validate their immunological reactivity and pave the way for food safety assurance.

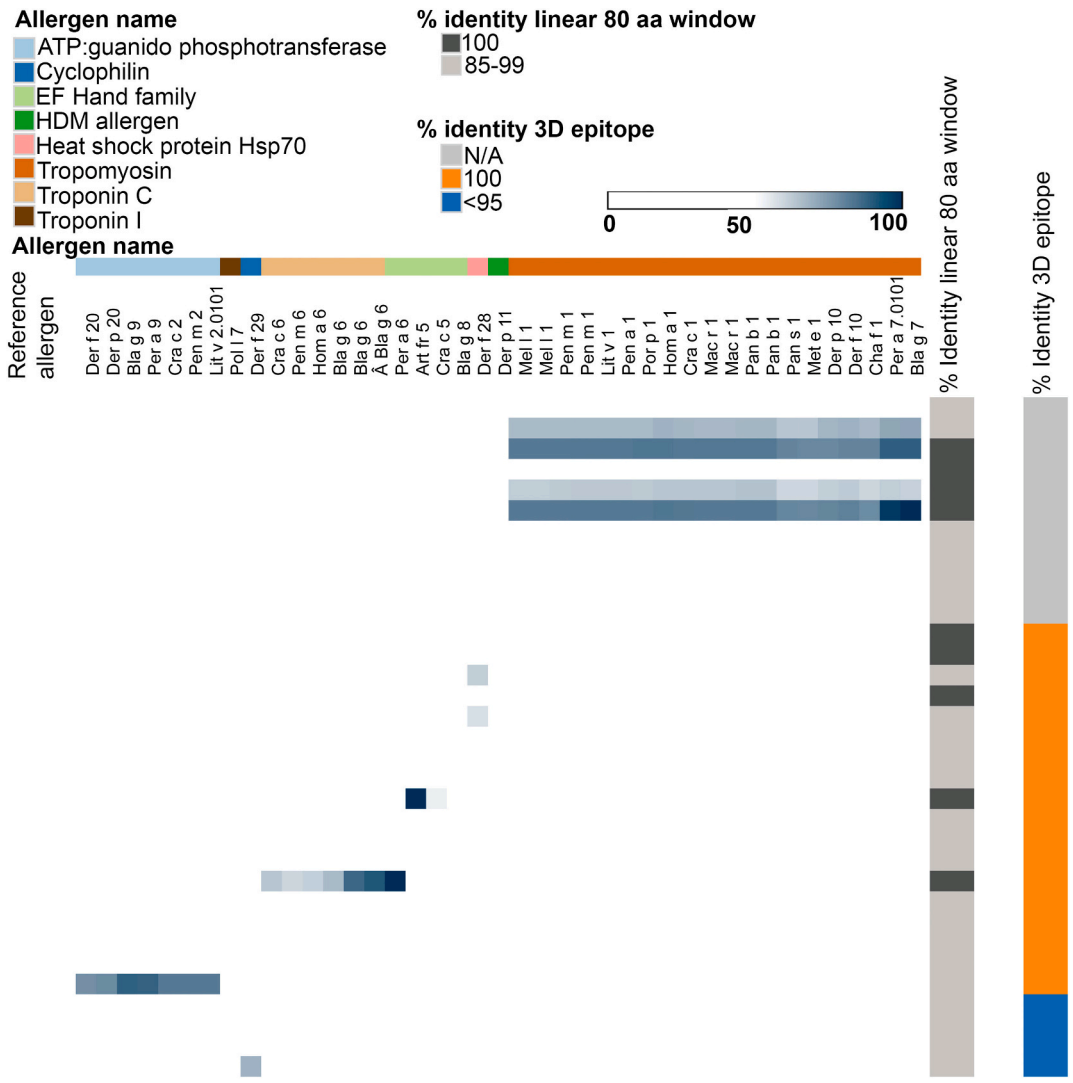

Fig. 4. Mapping the potential allergens from BSF using AllerCatPro and AllFam database mapping. The heatmap represents sequence similarity scores for the 41 reference allergen isoforms (columns) versus the 33 identified BSF proteins (rows). The reference allergens comprise both ingestion and inhalation-related insect and crustacean allergen families from the AllFam database. The 33 putative allergens were selected from the 84 allergen proteins using AllerCatPro software with the strict selection criteria: % identity linear 80 aa window ≥85% and/or % identity 3D epitope ≥94%. The resulting percent identity for both linear 80 aa sequences (85–100%) and percent identity of 3D epitopes (100 and < 95%) are shown.

3.6. Assessment of protocol-dependent tropomyosin extraction from BSF samples using LC-MRM-MS

A combined database search was performed to identify TM proteoforms in the BSF samples, and the TM identifications aligned with the AllerCatPro output. To this end, six TM proteoforms were detected at 1% FDR and unique peptide evidence from three samples extracted with seven different buffer compositions (Table S3; File S1). Although two UniProt accessions, UniProt ID: A0A1J1HVN3 and T1PHG4, were primarily identified as TMs during the combined database search, further annotation and analysis revealed these two proteins to be part of the muscle or membrane and not qualify as TMs based on their sequence similarity and 3D structure. As a result, four TMs were selected to build an MRM assay to monitor their relative variation across the extraction buffers and sample processing methods. Notably, one TM detected in the present study — UniProt ID: Q9NG56 — was registered as an allergen within the WHO/IUIS database, Bla g 7 – a tropomyosin allergen found in the German cockroach *Blattella germanica*. In total, 8 peptides were monitored for this allergen which covers ~29% of the total protein (Table S3). The IEDB database search revealed that seven out of eight monitored peptides for this protein were matched >90% to IgE epitopes from crustaceans (Table S3). Upon initial screening of TM sequences by LC-MRM-MS, 31 peptides were selected from four TMs where 10 (32%) peptides were shared between the four TMs (Fig. S6). Finally, for the final data representation 21 non-redundant peptides were selected where prioritisation was given to the peptides for the WHO/IUIS registered protein, UniProt ID: Q9NG56.

Heatmap visualisation and hierarchical clustering analysis was performed on the relative abundances of the 21 TM peptides to explore the relationship between sample processing and extraction buffers (Fig. 5A-C). The whole BSF samples showed a higher MRM response for TM peptides (Fig. 5A) in comparison to SP-1 and SP-2 (Fig. 5B-C). Protocols 3 and 4 show a similar capacity to extract TM peptides from whole BSF and SP-1 while a unique pattern of TM peptide responses were observed for SP-2 extracted with protocol 3. Although P2 resulted in the overall lowest response for peptides across the three samples, one replicate from P1 extraction of whole BSF co-clustered with P2. The protein column represents the four TM proteins detected and measured from BSF. TM peptides were also searched against the IEDB daatabase to identify the sequence similarity of BSF-derived TMs with known IgE epitopes present

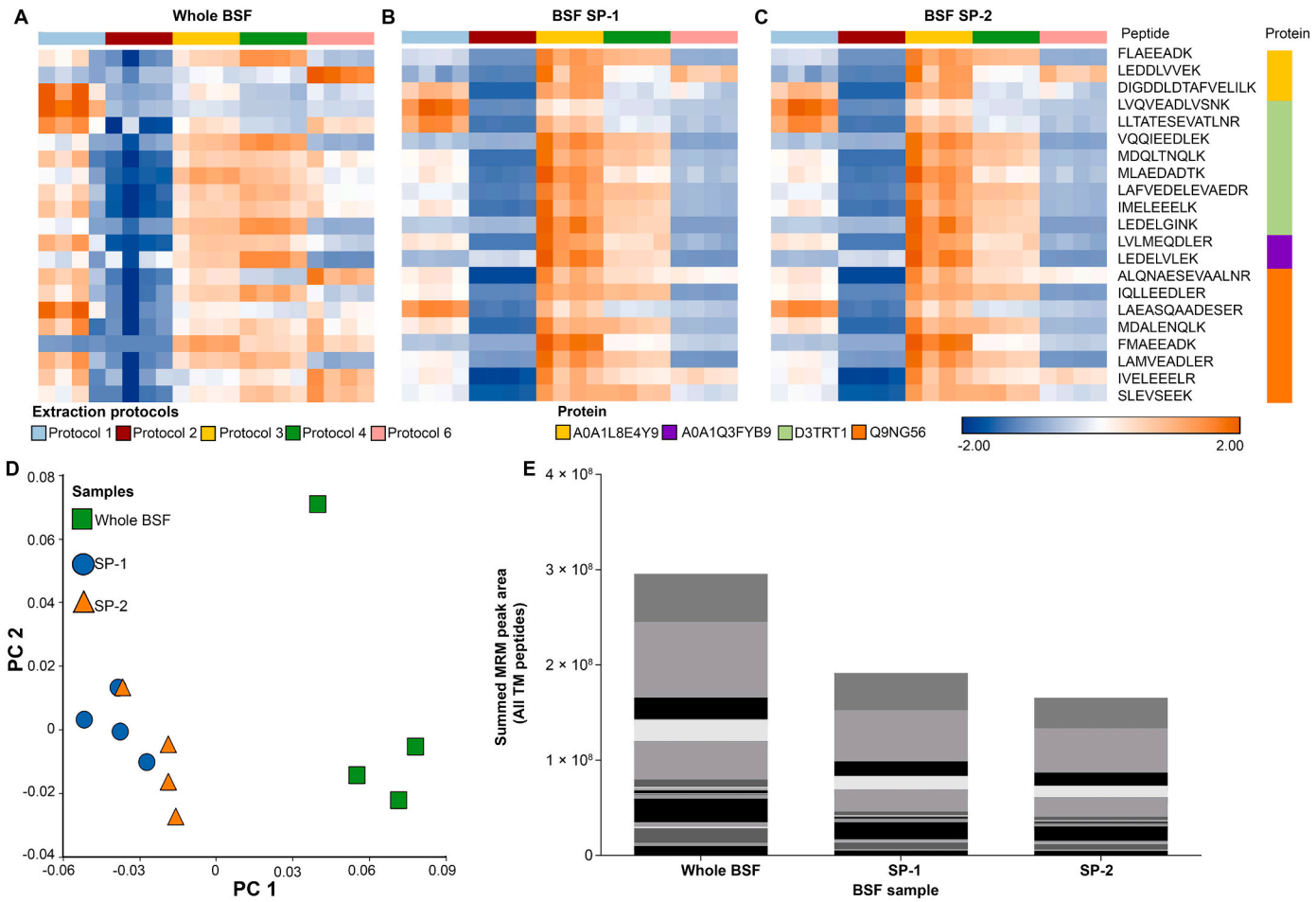

Fig. 5. Quantitation of tropomyosins by LC-MRM-MS across three BSF sample types. Heatmap visualisation shows the relative abundances (z-score) of tropomyosin peptides and the extraction protocols for BSF sample types for (A) whole; (B) screw press-1 and (C) screw press-2. The colour in each cell of the heatmap depicts the relative peak area of each peptide from row min (low) to row max (high); peptides measured by LC-MRM-MS are listed to the right. The protein column on the heatmap indicates the protein origin of each peptide. The subphylum column shows the mapping of cross-reactive peptides detected from BSF mapped to Crustacea, Chelicerata and Hexapoda. (D) A PCA score plot depicting PC1 vs PC2, showing the variation in the TM peptide profiles of three types of BSF samples. Each symbol represents a single type of BSF sample described by all monitored TM peptides. (E) Summed MRM peak area for monitored TM peptides across the three BSF sample types.

in other species. The IEDB search revealed that seven out of eight monitored peptides for this protein were matched >90% to IgE epitopes from crustaceans (Table S3).

The PCA based on the TM peptide intensities measured using LC-MRM-MS across the three types of processed BSF sample shows that whole BSF samples are different in terms of peptide intensities than the responses acquired for SP-1 and SP-2 highlighting the influence of sample processing on allergen content signal in the products (Fig. 5D). The whole BSF samples on PC1 explains 73.5% variance while PC2 captures 22% variance due to sample processing generated from SP-1 and SP-2 (Fig. 5D). To further measure the influence of sample processing, the peak areas were summed for all peptides measured across all samples to reveal that the whole BSF sample type had the highest TM signal per μg of total protein (Fig. 5E). The analysis of normalised peak area for TM peptides shows that SP-2 processing leads to a 61% lower signal for TM compared to the whole BSF sample ($p < 0.0001$; Fig. 5E). However, no significant difference was observed between SP-1 and SP-2 in TM peptide response. Overall, BSF processing was shown to significantly reduce the signal relating to allergen content.

4. Discussion

Complementary proteins from sustainable sources such as insects have emerged within the food and feed sector. Herein, three types of BSF samples were collected from an Australian company to assess the impact of buffer composition on total proteome detection and quantification of putative allergens using both discovery and targeted measurements. The discovery proteomics study revealed that the SDS/Tris-HCl-based buffer (P1) yielded the maximum number of proteins from the three types of BSF samples (Fig. 1-3). SDS is a strong anionic denaturing detergent that disrupts the lipid membranes and denatures proteins by breaking protein-protein interactions. Though the removal of SDS from the sample can be challenging, the use of filter-aided sample preparation (FASP) prior to protein digestion can be effective in its removal [23].

The selection of extraction buffers has been shown to influence BSF samples' protein content and proteome coverage. For instance, a defatting step followed by Tris-HCl-based extraction enabled the detection of 20 proteins through a mass-spectrometry-based study of BSF larvae [6]. Likewise, trichloroacetic acid (TCA) and acetone-based protein precipitation steps prior to re-solubilising proteins in a urea (6 M) / thiourea (2 M) / Tris-HCl (30 mM) buffer yielded 531 BSF protein identifications [9]. Although protein precipitation has previously led to high numbers of identified proteins from fresh larvae [9], no information is available on how precipitation affects the processed larvae. Protein precipitation methods using TCA/acetone are used to clean up complex samples before protein extraction [24]; however, this method has several

shortcomings in terms of: (1) substantial time commitment; (2) protein loss during re-solubilisation and multiple wash steps; and, (3) incompatibility with a range of sample types [19]. In the current study, robust statistical cut-offs were used (1% global FDR for proteins; 95% confidence on peptides) to report the number of proteins detected from BSF samples (Table S1). Additionally, the database was prepared using two recently published genomes [25,26], Diptera sequences and IUIS/WHO registered allergen sequences for insects and crustaceans. The results presented herein were based on the accession numbers reported following database searching. The high degree of protein sequence homology and the uncurated and unannotated nature of the insect protein database may generate a different suite of protein accessions when using different extraction buffers. As a result, a combined database search was performed to conservatively report protein identities and subsequently permit robust comparative analyses on these unified protein identities (Fig. 1-3). The higher number of protein and peptide identifications reported herein indicate the importance of searching a comprehensive database — especially for non-model species such as BSF — to identify the maximum number of proteins and minimise false positive results.

Allergic reactions can develop where the immune system cross-reacts with proteins that are highly similar — but not the same — to those already targeted by the immune system [27]. For instance, people with seafood allergy can experience allergic reactions after consuming edible insects, such as mealworm [17], cricket [13,15] and BSF [6,14]. In this regard, shellfish and crustaceans are closely related to insects, where crustaceans make up one of the main subphyla of the phylum Arthropoda. Due to the close phylogenetic relationship between crustaceans and insects [28,29], the potential of allergies from dietary insect consumption is high due to the presence of common arthropod allergens (pan-allergens) [11]. Extensive bioinformatic analyses revealed the presence of 33 putative cross-reactive allergens detected in the BSF samples (Fig. 4). Two recent studies using IgG- and IgE-immunoblotting experiments in combination with bottom-up proteomics have shown that patients allergic to crustaceans have cross-reactivity towards lesser mealworm and BSF tropomyosin [6,14]. Though the current study has detected several putative allergens from BSF, such immunoassay-guided proteomics will be necessary to investigate and validate additional putative cross-reactive allergens such as arginine kinase, triosephosphate isomerase and glutathione S-transferase as identified herein for BSF.

The processing of food ingredients can impart large changes to the composition of finished products and their detectable allergen content. Heat treatment has been shown to reduce the allergenicity of arginine kinases from crustaceans and insects [15,18]. We have shown that screw pressed samples from BSF contain reduced TM signal in comparison to the whole BSF sample type (Fig. 5). Similarly, arginine kinase signal was decreased upon heat treatment in cricket [15,18]. Importantly, heat processing as well as in vitro digestion was shown to reduce — but not eliminate — house dust mite or tropomyosin IgE cross-reactivity when crustacean- and house dust mite-allergic patients with cross-reactivity to mealworm tropomyosin (as well as a-amylase, hexamerin 1B precursor and muscle myosin) were orally challenged with treated and untreated mealworm flour [17,30]. In the future, the allergenicity of BSF proteins will have to be confirmed in vivo by oral food challenges, such as a double-blind placebo-controlled food challenge. In addition, food processing techniques other than heat processing, such as fermentation or hydrolysis, can be explored for their ability to reduce the allergenicity of BSF protein.

5. Conclusion

Looking forward, extraction protocols that efficiently extract target proteins (i.e., allergens) yet comprehensively cover the BSF proteome are warranted. Ideally, such an optimised method will have the potential to identify novel allergens and reveal a full and continuous view of the landscape of putative allergens as it unfolds and be applicable to a broad range of food products that may contain these insect ingredients. The methods deployed herein enable a rapid assessment of extraction buffers for allergen protein yields and how these vary according to sample processing. The integration of immuno-assays and clinical studies alongside proteomic experiments employing allergen-specific extraction protocols will paint a complete picture of the allergenic potential of insect-based foods. Moreover, this combination of techniques will allow the optimisation of food processing parameters to potentially reduce or remove allergenicity, which could see novel insect-based products included in mainstream foods with safe use for all.

Author contribution

The present work carried out in collaboration between all authors from four research institutions: Commonwealth Scientific and Industrial Research Organisation (CSIRO; Australia), Edith Cowan University (ECU; Australia), James Cook University (JCU; Australia) and Agency for Science, Technology and Research (A*STAR; Singapore). Conceptualization: M.C., A.L.L. and S.M.S.; Funding acquisition: M.C., A.L.L. and S.M.S.; Data Curation and Formal Analysis: U.B., S.K., U.B. J.B., K.B. E.B.J., A.J. and S.S. Methodology: U.B., S.K., J.B., V.L. and A.J. Resources: M.C., A.L.L., J.B. and S.M.S. Writing - original draft: U.B., J.B. and A.J. Writing - review & editing: all authors.

Data availability

Raw and processed LC-MS data files have been uploaded to https://doi.org/10.25919/kfqg-fh11

Declaration of Competing Interest

Authors have declared no competing financial interests and personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The project was funded by the 1st Singapore-Australia Bilateral Program on Innovations in Food for Precision Health, 2019 (Grant number: SG-AUS2019_191D4) encompassing a partnership between The Commonwealth Scientific and Industrial Research Organisation (Australia), James Cook University (Australia and Singapore) and Agency for Science, Technology and Research (A*STAR, Singapore). We also would like to thank to Olympia Yarger (GOTERRA, Australia) for supplying three types of BSF sample.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jprot.2022.104724.

References

[1] E. Schmitt, I. Belghit, J. Johansen, R. Leushuis, E.-J. Lock, D. Melsen, R. Kathirampatti Ramasamy Shanmugam, J. Van Loon, A. Paul, Growth and safety assessment of feed streams for black soldier fly larvae: a case study with aquaculture sludge, Animals 9 (4) (2019) 189.

[2] S. Dabbou, F. Gai, I. Biasato, M.T. Capucchio, E. Biasibetti, D. Dezzutto, M. Meneguz, I. Plachà, L. Gasco, A. Schiavone, Black soldier fly defatted meal as a dietary protein source for broiler chickens: effects on growth performance, blood traits, gut morphology and histological features, J. Anim. Sci. Biotechnol. 9 (1) (2018) 49.

[3] I. Belghit, N.S. Liland, P. Gjesdal, I. Biancarosa, E. Menchetti, Y. Li, R. Waagbø, Å. Krogdahl, E.-J. Lock, Black soldier fly larvae meal can replace fish meal in diets of sea-water phase Atlantic salmon (*Salmo salar*), Aquaculture 503 (2019) 609–619.

[4] L.F. Meyer, M. Koelln, J. Kamphues, Dog food containing insects? Studies on compound feed with larvae of the black soldier fly as the main protein source, Kleintierpraxis 64 (3) (2019) 124–135.

[5] K.B. Barragan-Fonseca, M. Dicke, J.J. van Loon, Nutritional value of the black soldier fly (*Hermetia illucens* L.) and its suitability as animal feed–a review, J. Insects Food Feed 3 (2) (2017) 105–120.
[6] G. Leni, T. Tedeschi, A. Faccini, F. Pratesi, C. Folli, I. Puxeddu, P. Migliorini, N. Gianotten, J. Jacobs, S. Depraetere, Shotgun proteomics, in-silico evaluation and immunoblotting assays for allergenicity assessment of lesser mealworm, black soldier fly and their protein hydrolysates, Sci. Rep. 10 (1) (2020) 1–10.
[7] R. Smets, B. Verbinnen, I. Van De Voorde, G. Aerts, J. Claes, M. Van Der Borght, Sequential extraction and characterisation of lipids, proteins, and chitin from black soldier fly (*Hermetia illucens*) larvae, prepupae, and pupae, Waste Biomass Valorization 11 (2020) 6455–6466.
[8] B.D. Choi, N.A. Wong, J.-H. Auh, Defatting and sonication enhances protein extraction from edible insects, Korean J. Food Sci. Anim. Resour. 37 (6) (2017) 955.
[9] V. Rabani, H. Cheatsazan, S. Davani, Proteomics and lipidomics of black soldier fly (Diptera: Stratiomyidae) and blow fly (Diptera: Calliphoridae) larvae, J. Insect Sci. 19 (3) (2019) 29.
[10] A. Caligiani, A. Marseglia, G. Leni, S. Baldassarre, L. Maistrello, A. Dossena, S. Sforza, Composition of black soldier fly prepupae and systematic approaches for extraction and fractionation of proteins, lipids and chitin, Food Res. Int. 105 (2018) 812–820.
[11] T. Ruethers, A.C. Taki, E.B. Johnston, R. Nugraha, T.T. Le, T. Kalic, T.R. McLean, S. D. Kamath, A.L. Lopata, Seafood allergy: a comprehensive review of fish and shellfish allergens, Mol. Immunol. 100 (2018) 28–57.
[12] S. de Gier, K. Verhoeckx, Insect (food) allergy and allergens, Mol. Immunol. 100 (2018) 82–106.
[13] N. Kamemura, M. Sugimoto, N. Tamehiro, R. Adachi, S. Tomonari, T. Watanabe, T. Mito, Cross-allergenicity of crustacean and the edible insect *Gryllus bimaculatus* in patients with shrimp allergy, Mol. Immunol. 106 (2019) 127–134.
[14] L.K. Palmer, J.T. Marsh, M. Lu, R.E. Goodman, M.G. Zeece, P.E. Johnson, Shellfish tropomyosin IgE cross-reactivity differs among edible insect species, Mol. Nutr. Food Res. 64 (8) (2020) 1900923.
[15] C. Srinroch, C. Srisomsap, D. Chokchaichamnankit, P. Punyarit, P. Phiriyangkul, Identification of novel allergen in edible insect, *Gryllus bimaculatus* and its cross-reactivity with *Macrobrachium* spp. allergens, Food Chem. 184 (2015) 160–166.
[16] G. Reese, R. Ayuso, S.B. Lehrer, Tropomyosin: an invertebrate pan–allergen, Int. Arch. Allergy Immunol. 119 (4) (1999) 247–258.
[17] S. Van Broekhoven, S. Bastiaan-Net, N.W. de Jong, H.J. Wichers, Influence of processing and in vitro digestion on the allergic cross-reactivity of three mealworm species, Food Chem. 196 (2016) 1075–1083.
[18] U. Bose, J.A. Broadbent, A. Juhász, S. Karnaneedi, E.B. Johnston, S. Stockwell, K. Byrne, V. Limviphuvadh, S. Maurer-Stroh, A.L. Lopata, Protein extraction protocols for optimal proteome measurement and arginine kinase quantitation from cricket *Acheta domesticus* for food safety assessment, Food Chem. 348 (2021), 129110.
[19] U. Bose, J. Broadbent, K. Byrne, S. Hasan, C.A. Howit, M.L. Colgrave, Optimisation of protein extraction for in-depth profiling of the cereal grain proteome, J. Proteome 197 (2019) 23–33.
[20] U. Bose, K. Byrne, C. Howit, M.L. Colgrave, Targeted proteomics to monitor the extraction efficiency and levels of barley α-amylase trypsin inhibitors that are implicated in non-coeliac gluten sensitivity, J. Chromatogr. A 1600 (2019) 55–64.
[21] B. MacLean, D.M. Tomazela, N. Shulman, M. Chambers, G.L. Finney, B. Frewen, R. Kern, D.L. Tabb, D.C. Liebler, M.J. MacCoss, Skyline: an open source document editor for creating and analyzing targeted proteomics experiments, Bioinformatics 26 (7) (2010) 966–968.
[22] S. Maurer-Stroh, N.L. Krutz, P.S. Kern, V. Gunalan, M.N. Nguyen, V. Limviphuvadh, F. Eisenhaber, G.F. Gerberick, AllerCatPro—prediction of protein allergenicity potential from the protein sequence, Bioinformatics 35 (17) (2019) 3020–3027.
[23] J.R. Wiśniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for proteome analysis, Nat. Methods 6 (5) (2009) 359.
[24] X. Wu, E. Xiong, W. Wang, M. Scali, M. Cresti, Universal sample preparation method integrating trichloroacetic acid/acetone precipitation with phenol extraction for crop proteomic analysis, Nat. Protoc. 9 (2) (2014) 362–374.
[25] S. Zhan, G. Fang, M. Cai, Z. Kou, J. Xu, Y. Cao, L. Bai, Y. Zhang, Y. Jiang, X. Luo, Genomic landscape and genetic manipulation of the black soldier fly *Hermetia illucens*, a natural waste recycler, Cell Res. 30 (1) (2020) 50–60.
[26] T.N. Generalovic, S.A. McCarthy, I.A. Warren, J.M. Wood, J. Torrance, Y. Sims, M. Quail, K. Howe, M. Pipan, R. Durbin, A high-quality, chromosome-level genome assembly of the Black Soldier Fly (*Hermetia Illucens* L.), G3 11 (5) (2020) jkab085.
[27] H. van der Fels-Klerx, L. Camenzuli, S. Belluco, N. Meijer, A. Ricci, Food safety issues related to uses of insects for feeds and foods, Compr. Rev. Food Sci. Food Saf. 17 (5) (2018) 1172–1183.
[28] J. Ribeiro, B. Sousa-Pinto, J. Fonseca, S.C. Fonseca, L. Cunha, Edible insects and food safety: allergy, J. Insects Food Feed 7 (5) (2021) 833–847.
[29] R. Nugraha, S.D. Kamath, E. Johnston, S. Karnaneedi, T. Ruethers, A.L. Lopata, Conservation analysis of B-cell allergen epitopes to predict clinical cross-reactivity between shellfish and inhalant invertebrate allergens, Front. Immunol. 10 (2019) 2676.
[30] C. Lamberti, S. Nebbia, S. Cirrincione, L. Brussino, V. Giorgis, A. Romito, C. Marchese, M. Manfredi, E. Marengo, M.G. Giuffrida, Thermal processing of insect allergens and IgE cross-recognition in Italian patients allergic to shrimp, house dust mite and mealworm, Food Res. Int. 148 (2021), 110567.