Spring 5-2021

NOVEL SOURCES OF FOOD ALLERGENS

Lee Palmer
University of Nebraska-Lincoln, leekpalmer@yahoo.com

NOVEL SOURCES OF FOOD ALLERGENS

by

Lee Palmer

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Food Science and Technology

Under the Supervision of Professor Philip E. Johnson

Lincoln, Nebraska

May 2021

NOVEL SOURCES OF FOOD ALLERGENS

Lee Palmer, Ph.D.

University of Nebraska, 2021

Advisor: Philip E. Johnson

Technological advancement and globalization have led to the spread of foods to countries where the food does not yet have a documented history of consumption, in other words, novel foods. Novel foods also encompass truly novel foods, foods that have been processed in a novel manner, and novel means of exposure. With novel foods comes the potential of food allergens that pose an uncharacterized risk to those with food allergies. Food allergies are an increasingly important facet of public health. Therefore, a deeper understanding of novel foods as well as methods to evaluate consumers' potential risk is necessary. Literature reviews and experimental evaluations leveraging liquid chromatography-electrospray ionization-mass spectrometry were used to explore the risks posed by novel sources of food allergens. Subject sources of allergens included *Acheta domesticus*, the house cricket, *Tenebrio molitor*, the yellow mealworm; extensively thermally processed walnut hulls and peanuts, as well as smoke from the wood of tree nut trees, and vapor from E-cigarette liquids. Novel methodologies to interpret complex mass spectrometry data were developed, allowing resultant information to be used in the assessment of allergenic risk. The methodologies developed in this research expand upon the utility of mass spectrometry to evaluate potentially allergenic proteins from poorly

characterized sources. Broader characterization of the hazards and risks posed by food allergens permits stakeholders to be more adequately informed regarding the risks they wish to undertake.

ACKNOWLEDGMENTS

Thanks to everyone who believed in Dr. Lee.

PREFACE

This doctoral dissertation is organized into eight chapters providing an overview of the analysis of novel sources of food allergens using mass spectrometry and the application of risk assessment methodologies to evaluate the risks posed to consumers.

Some of the content included in Chapter 1, section 1.7 is expected to be included in a manuscript under the lead of Justin Marsh (J. T. Marsh, S. J. Koppelman, L. K. Palmer, P. E. Johnson, "*Determination of major allergen levels, isoforms, and hydroxyproline modifications among peanut market types by mass spectrometry*," (in preparation for submission to a refereed journal)).

Some of the content in included in Chapter 2 is expected to be included in a manuscript under the lead of Lee Palmer (L. K. Palmer, B. Oppert, L. C. Perkin, M. Lorenzen, A. T. Dossey, P. E. Johnson, "*Predicted allergens and quantitative proteomics from life stages of the house cricket, Acheta domesticus,"* (in preparation for submission to a refereed journal)).

Chapter 4 has been published in *LWT* (L. K. Palmer, J. T. Marsh, J. L. Baumert, P. E. Johnson, LWT **132**, 109903 (2020)).

The abstract of Chapter 5 has been submitted for publication in Allergy (L. K. Palmer, P. E. Johnson, "*Detection of Food Allergen-Derived Peptides from Extractive-Based E-Cigarette Liquids,"* (submitted for publication in Allergy, January 2021)).

Overall aims of the Dissertation Research

1. Explore novel foods, foods processed in novel ways, and novel means of exposure as novel sources of food allergens;
2. Develop methodologies for incorporating mass spectrometry into allergen risk assessment;
3. Evaluate how conventional risk assessment methodologies can be leveraged to determine the qualitative and quantitative allergic risks posed by novel sources of food allergens.

Structure of the dissertation

In **Chapter 1,** the literature regarding the relationship between food allergy and novel foods is explored.

Chapters 2 and 3 introduce insects as a seldom consumed food in the Western Hemisphere and explores *Acheta domesticus*, the house cricket, and *Tenebrio molitor*, the yellow mealworm. Developmental stages of *A. domesticus* are evaluated for differences in potential risk posed to consumers and a pair of genetically modified *T. molitor* are compared with the wildtype to determine the effects of genetic modification on levels of predicted allergens.

Chapter 4 explored extensively thermally processed peanuts to evaluate if any remnants of proteins can persist and if they have the capacity to theoretically cause reactions in peanut allergic consumers.

Chapter 5 evaluated the presence of food allergen residues in E-cigarette liquids resulting in risk assessments in terms of both food allergens and respiratory allergens.

Chapters 6 and 7 are literature-based risk assessments evaluating the use of extensively thermally processed walnut hulls for water purification and the potential risks associated with the use of tree nut wood to smoke foods.

Finally, **Chapter 8** presents a discussion of the findings including limitations, necessary assumptions, and future directions.

TABLE OF CONTENTS

LIST OF FIGURES xvii

LIST OF TABLES xx

CHAPTER 1: UNDERPINNING THE RELATIONSHIP BETWEEN FOOD ALLERGIES AND NOVEL FOODS 1

1.1 Introduction 1

1.2 Novel foods 1

1.2.1 The niche for novel foods 1

1.2.2 Brief novel food regulations of selected countries and trade blocs 4

1.3 Food allergy 17

1.3.1 Overview 17

1.3.2 Mechanisms of IgE-mediated hypersensitivity 21

1.3.3 Diagnosing food allergy 23

1.3.4 Prevalence and impact of food allergies 26

1.3.5 Properties and characteristics of allergenic food proteins 28

1.3.6 Food processing and food allergens 30

1.3.7 Novel sources of food allergens 32

1.4 Quantitative methodologies for the detection of food allergens 32

1.4.1 Polymerase chain reaction 33

1.4.2 Enzyme-link immunosorbent assays 35

1.4.3 Mass spectrometry 36

1.5 Risk analysis and food allergens 39

1.5.1 Risk assessment 40

1.5.2 Risk management 44

1.5.3 Risk communication 45

1.6 Allergic risks of consuming insects 46

1.6.1 Background 46

1.6.2 Crickets 46

1.6.3 Mealworm 50

1.6.4 Allergens of edible insects and related species 51

1.6.5 Allergen classes of concern 55

1.6.6 Allergen Carryover from Feed 59

1.6.7 Summary 60

1.7 Peanut food allergy and the effects of thermal processing 61

1.7.1 Background 61

1.7.2 Peanut allergy 62

1.7.3 Effects of thermal processing on peanut allergenicity 69

1.7.4 Summary 71

1.8 Allergic risks of exposure to food allergens by inhalation 71

1.8.1 Background 71

1.8.2 Defenses, barriers, and transport of allergens 72

1.8.3 Inhalation of food allergens....75

1.8.4 Insights into routes of exposure from biopharmaceuticals 75

1.8.5 Summary 77

1.9 Concluding remarks 77

CHAPTER 2: PREDICTED ALLERGENS AND QUANTITATIVE PROTEOMICS FROM LIFE STAGES OF THE HOUSE CRICKET, *ACHETA DOMESTICUS*....79

2.1 Abstract....79

2.2 Introduction....80

2.3 Methods and Materials 81

2.3.1 Chemicals and samples 81

2.3.2 Extraction and SDS-PAGE 82

2.3.3 Mass Spectrometry 83

2.3.4 Data Analysis....85

2.4 Results 88

2.4.1 Characteristics of samples and cricket life stages 88

2.4.2 Proteomic shifts across life stages 88

2.4.3 Proteome annotation and gene ontology enrichment....91

2.4.4 Predicted allergens across life stages 93

2.5 Discussion....97

CHAPTER 3: PROTEOMIC EVALUATION OF YELLOW MEALWORM, TENEBRIO MOLITOR, WITH EXPERIMENTAL GENETIC MODIFICATIONS ...101

3.1 Abstract....101

3.2 Introduction....102

3.3 Methods and Materials....103

3.3.1 Samples and Chemicals....103

3.3.2 Sample preparation and protein extraction....104

3.3.3 Mass Spectrometry....104

3.3.4 Data Analysis....106

3.4 Results....109

3.4.1 Proteomics of *T. molitor* samples....109

3.4.2 Proteome annotation and gene ontology enrichment....112

3.4.3 Predicted allergens of *T. molitor*....114

3.5 Discussion....118

CHAPTER 4: PERSISTENCE OF PEANUT ALLERGEN-DERIVED PEPTIDES THROUGHOUT EXCESSIVE DRY THERMAL PROCESSING....121

4.1 Abstract....121

4.2 Introduction....122

4.3 Materials and methods....123

4.3.1 Chemicals....123

4.3.2 Thermal Processing....124
4.3.3 Protein Extraction....124
4.3.4 Mass Spectrometry and data analysis....125
4.3.5 SDS-PAGE, sera, and immunoblotting....127
4.4 Results....128
4.4.1 Observable Properties of Thermally Processed Peanuts....128
4.4.2 Mass Spectrometry....129
4.4.3 Antibody Binding to Thermally Processed Peanut....133
4.5 Discussion....135
CHAPTER 5: DETECTION OF FOOD ALLERGEN-DERIVED PEPTIDES FROM EXTRACTIVE-BASED E-CIGARETTE LIQUIDS....139
5.1 Abstract....139
5.2 Introduction....140
5.3 Methods and Materials....142
5.3.1 Samples and Chemicals....142
5.3.2 Sample preparation....143
5.3.3 SDS-PAGE....143
5.3.4 Mass Spectrometry....143
5.3.5 Risk assessment....145
5.4. Results....146

5.4.1 Visualization of present proteins 146

5.4.2 Assessment of Data Sets and Peptide Assignments 150

5.4.3 Identification of proteins present in E-cigarette liquids 152

5.4.4 Quantification of proteins and evaluation of the risk of eliciting allergic reactions 152

5.5 Discussion 153

CHAPTER 6: LITERATURE-BASED RISK ASSESSMENT OF ACTIVATED CARBON MADE FROM WALNUT HULLS FOR USE IN WATER PURIFICATION SYSTEMS 157

6.1 Summary 157

6.2 Background 157

6.3 Activated Carbon Processing 158

6.4 Allergenic Hazards of Walnut Hulls 159

6.5 Effects of Thermal Processing on Allergens 160

6.6 Allergen Labeling of Activated Carbon Derived from Walnut Hulls 161

6.7 Chemical Hazards of Walnut Hulls 162

6.8 Conclusions 163

CHAPTER 7: LITERATURE-BASED RISK ASSESSMENT OF ALLERGENS IN SMOKE DERIVED FROM TREE NUT WOOD 164

7.1 Summary 164

7.2 Background....164

7.3 Wood, Trees, and Logging....165

7.4 Industrial Smoke Production, Processing, and Usage....166

7.5 Smoke as a Medium for Allergen Transport....169

7.6 Effects of Thermal Processing on Allergens....171

7.7 Allergen Labelling of Products Produced in Part with Smoke from Tree Nut Wood....173

7.8 Conclusions on the risks associated with Smoke derived from Tree Nut Wood...173

CHAPTER 8: SUMMARY AND CONCLUSIONS ON NOVEL SOURCES OF FOOD ALLERGENS....174

REFERENCES....177

APPENDIX A SUPPLEMENTARY FIGURES AND TABLES FROM CHAPTER 2: PREDICTED ALLERGENS AND QUANTITIATIVE PROTEOMICS FROM LIFE STAGES OF THE HOUSE CRICKET, *ACHETA DOMESTICUS*....204

APPENDIX B ANNOTATIONS OF QUANTIFIED PROTEINS FROM ACHETA DOMESTICUS (XLSX, 380 KB)....213

APPENDIX C SUPPLEMENTARY FIGURES AND TABLES FROM CHAPTER 3: PROTEOMIC EVALUATION OF YELLOW MEALWORM, TENEBRIO MOLITOR, WITH EXPERIMENTAL GENETIC MODIFICATIONS....214

APPENDIX D ANNOTATIONS OF QUANTIFIED PROTEINS FROM TENEBRIO MOLITOR (XLSX, 337 KB)....227

APPENDIX E SUPPLEMENTARY FIGURES AND TABLES FROM CHAPTER 4: PERSISTENCE OF PEANUT ALLERGEN-DERIVED PEPTIDES THROUGHOUT EXCESSIVE DRY THERMAL PROCESSING 228

APPENDIX F COMPLETE PEPTIDE LIST FOR ROASTED PEANUTS (XLSX, 4853 KB) 235

APPENDIX G SUPPLEMENTARY FIGURES AND TABLES FROM CHAPTER 5: DETECTION OF FOOD ALLERGEN-DERIVED PEPTIDES FROM EXTRACTIVE-BASED E-CIGARETTE LIQUIDS 236

LIST OF FIGURES

Figure 1.1. Examples of novel foods and processes involving by-product or waste valorization.....2

Figure 1.2 Mechanism of cellular activation in response to allergenic insult and subsequent recognition by IgE. 21

Figure 1.3 The constituent parts and sub-parts of an iteratively applied risk analysis.....40

Figure 2.1 Heatmaps and Pearson correlation matrices of all proteins and robust proteins 91

Figure 2.2 GO enrichment of life stages.....93

Figure 2.3 S-curves of quantifiable proteins and predicted allergens per life stage 94

Figure 2.4 Venn diagram of protein assignments and annotations 95

Figure 2.5 Robust and variable predicted allergen heatmap and characteristics of predicted allergens across life stages.....96

Figure 3.1 Heatmaps of all proteins and robustly identified proteins 111

Figure 3.2 GO enrichment of samples against all proteins or robustly identified proteins 113

Figure 3.3 S-curves of quantifiable proteins.....115

Figure 3.4 Venn diagram of protein assignments and annotations 116

Figure 3.5 Selected predicted allergen heatmap and characteristics of predicted allergens across life stages.....117

Figure 4.2 SDS-PAGE and immunoblots.....134

Figure 4.3 Dot blots.....135

Figure 5.1 SDS-PAGE gel of E-liquids.....150

Appendix A Figure 2.1.1 *A. domesticus* life stages.....204

Appendix A Figure 2.1.2 SDS-PAGE of extracts from *A. domesticus* life stages.....206

Appendix A Figure 2.1.3 Spectra of peptides with posttranslational modifications and relevant unmodified forms.....207

Appendix A Figure 2.1.4 Phylogeny of the origin species of allergen predictions 209

Appendix C Figure 3.1.1 Alignments and sequence coverage for arginine kinase and tropomyosins.....217

Appendix C Figure 3.1.2 Spectra of peptides with posttranslational modifications and relevant unmodified forms.....219

Appendix C Figure 3.1.3 Phylogeny of the origin species of allergen predictions.....221

Appendix C Figure 3.1.4 Alignment of *T. molitor* identified tropomyosins with known allergenic tropomyosins.....225

Appendix C Figure 3.1.5 Identity matrix of alignment of *T. moltior* tropomyosins with known allergenic tropomyosins 226

Appendix E Figure 4.1.1 Peanut kernel halves over thermal processing 230

Appendix E Figure 4.1.2 Extracts of peanuts heated at 176 °C and 260 °C.....231

Appendix E Figure 4.1.3 Extracts of peanuts heated at 176 °C and 260 °C.....232

Appendix E Figure 4.1.4 Peanut allergenic profile for peanuts heated at 176 °C.....232

Appendix E Figure 4.1.5 Persistence of peanut allergen peptides at 176 °C.....233

Appendix E Figure 4.1.6 Negative control immunoblots.....234

Appendix G Figure 5.1.1 Annotated Peptide spectra.....242

LIST OF TABLES

Table 1.1 Symptoms by system associated with IgE-mediated reactions to food. 19

Table 1.2 Pros and cons of quantitative methodologies to detect food allergens. 33

Table 1.3 Major and minor allergens of peanut 63

Table 2.1 Life stage quantification characteristics 89

Table 3.1 Sample quantification characteristics.....110

Table 4.1 MS quantification and summary of peanut allergens after roasting at 176 °C130

Table 4.2 Maximum length of contiguous peptides of peanut allergens after roasting for 176 °C 132

Table 5.1 Literature values and assumptions used in food allergen risk assessment of E-liquids 147

Table 5.2 Literature values and assumptions used in respiratory allergen risk assessment of E-liquids.....148

Table 5.3 E-liquid properties and volumes analyzed by SDS-PAGE and Mass Spectrometry 149

Table 5.4 Protein quantification using mass spectrometry 153

Appendix A Table 2.1.1 Life stage replicate characteristics 205

Appendix A Table 2.1.2 Mass spectrometry multi-round search metadata.....206

Appendix A Table 2.1.3 Peptide posttranslational modification characteristics 208

Appendix A Table 2.1.4 Enriched Gene Ontology (GO) terms for the top 10% most abundant proteins per life stage.....210

Appendix C Table 3.1.1 *Tribolium castaneum* tropomyosin-2 exon structure utilized to manually curate *T. molitor* protein sequences.....214

Appendix C Table 3.1.2 *T. molitor* tropomyosin-2 exon nucleotide sequences.....215

Appendix C Table 3.1.3 Mass spectrometry multi-round search metadata.....218

Appendix C Table 3.1.4 Peptide posttranslational modification characteristics 220

Appendix C Table 3.1.5 Enriched Gene Ontology (GO) terms of the top 10% most abundant proteins per sample against all proteins.....221

Appendix C Table 3.1.6 Enriched Gene Ontology (GO) terms of the top 10% most abundant proteins per sample against robust proteins.....223

Appendix E Table 4.1.1 Posttranslational modifications added to PTM search.....228

Appendix E Table 4.1.2 Lists of peptides used for quantification of peanut allergens...229

Appendix G Table 5.1.1 Comprehensive peptide identifications 237

Appendix G Table 5.1.2 Normalized sample wise peptide presence and quantity.....238

Appendix G Table 5.1.3 Comprehensive protein identifications.....243

CHAPTER 1: UNDERPINNING THE RELATIONSHIP BETWEEN FOOD ALLERGIES AND NOVEL FOODS

1.1 Introduction

Technological advancement and globalization have led to the spread of food to countries that do not have a history of safe consumption of that food, in other words novel foods. As novel foods, novel techniques to process foods, and novel means to consume foods continue to develop, so do the sources of potential food allergens. Food allergies are becoming an increasingly important facet of public health around the world. To protect the health of consumers a deeper understanding of both novel foods and the risk posed is needed. The following will review food allergy and detection of food allergens, novel foods and regulatory definitions, and the principles of food allergen risk analysis with considerations to novel sources of food allergens.

1.2 Novel foods

1.2.1 The niche for novel foods

A novel food can be broadly defined as a food without a significant history of consumption or produced by a method that had not been previously applied to food, which is subjective at the level of the individual, culture, and governmental regulatory agencies. For example, for many in the Western Hemisphere, insects are considered a novel food but elsewhere, such as Thailand, they represent a common nutrient source. Novel foods are diverse including examples such as cell-based meat [14], mycoprotein [15], insects [16], microalgae [17], and valorization of waste or by-products [18] (Figure

1.1).

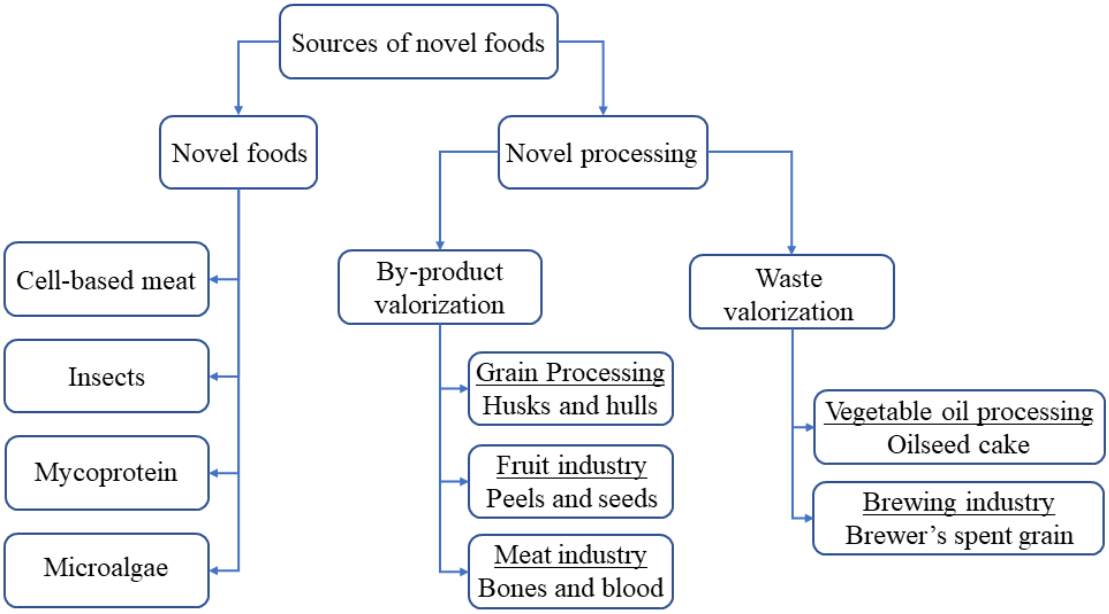


Figure 1.1. Examples of novel foods and processes involving by-product or waste valorization.

Novel processing technologies are also being researched and utilized such as cold atmospheric plasma, oscillating magnetic fields, high-pressure processing, pulsed electric fields, pulsed light, irradiation, and ultrasound, where much of the industry is focused on maintaining sensory qualities but held back by costs and a current dearth of research into the application of these technologies as applied to food [19]. Novel foods and ingredients may also be produced through the application of traditional processing techniques to foods that had not previously been processed in that manner, such as with protein isolates. Genetic modification of crops and organisms can generate novel variants of crops with improved nutritional quality as in golden rice engineered to produce β-carotene (provitamin A) and tackle vitamin A deficiencies, iron biofortified rice to

increase the iron content and combat anemia, and quality protein maize to improve the amino acid profile of corn via a higher production of essential amino acids [20].

The world population is projected as just under 10 billion people in 2050 [21], and feeding the population requires adequate food and protein. Despite farmers producing approximately 4600 kcal per person, or double what is needed, malnutrition remains a persistent threat [22]. The global spread of the Western diet, characterized by a diet rich in animal proteins, refined fats, sugars, and processed foods, has driven further animal rearing and has been viewed as a significant confounder in a path toward sustainably feeding the future population [23, 24]. Proponents suggest novel foods as a means to alleviate unsustainable consumption of animal protein requiring profound inputs of land, fresh water, and energy but also excessive environmental impacts such as generation of greenhouse gasses, deforestation, loss of biodiversity, and degraded water supplies through runoff of excess fertilizer and pesticides [23, 25]. Novel processing technologies then can be implemented for improved efficiency of current food systems by reducing food waste, where approximately one-third of food produced for human consumption is lost or wasted throughout the supply chain [25].

Comparisons to conventional animal rearing have been performed using insects as an alternative animal protein source. Insects broadly have a greater food conversion ratio, require less land to rear, require minimal water, and produce fewer greenhouse emissions than conventional livestock [26]. In terms of energy, mealworms have been investigated in the Netherlands in a year-round climate-controlled facility and found to require less energy than beef but equivalent to pork [27], but this could be improved by growing the mealworms in a more suitable climate. Although insects have been touted for their ability

to thrive on waste and by-products, crickets reared on an industrial scale do require at least an organic side-stream of processed relatively high-quality feed to thrive and preferably the current feed given to poultry [28]. Insects as a novel food demonstrate considerable improvements in sustainability compared to conventional livestock but are more appropriately understood as competition rather than wholly usurping the position of conventional livestock.

As sustainable novel foods are but a competitor within the marketplace, ensuring that novel foods are supported from both the consumer and producer sides is critical. Producers need to be appropriately compensated for practicing sustainable farming practices that support the environment [23]. Simultaneously, consumers need to be appropriately guided to overcome neophobia through the persistence of the novel food in the market but also labeling of food packages [29, 30]. Careful presentation of novel foods is key, as evidenced by long-used processing technologies such as irradiation, which is commonly regarded as a novel processing technology despite its use for decades. Further, irradiated products when presented as ‘treated with ionization’ are regarded less negatively as compared to ‘treated with irradiation’, likely through associations of irradiation with inferior goods and nuclear power. Presentation of labels is in part based on the producer’s interests but heavily driven by pertinent regulations demanded by regulatory agencies.

1.2.2 Brief novel food regulations of selected countries and trade blocs

Although which foods are categorized as novel can vary by individual and cultural standards, a key unifying factor is the guidance provided by regulatory agencies

and government mandates. Regulations regarding what constitutes a novel food commonly identifies a lack of a documented history of safe use within the purview of each regulatory agency, with considerations given for safe use in other countries. Foods deemed novel undergo risk assessments to determine if significant harm will come from consumption of the novel food. Differences among regulations are often in the form of differences of explicitly stated means that a food may be novel such as if a genetically modified organism is also a novel food.

1.2.2.1 Australia and New Zealand

The statutory authority on food safety for both Australia and New Zealand is the Food Standards Australia New Zealand (FSANZ) and is responsible for the development and administration of the Australia New Zealand Food Standards Code [31]. As of 13 April, 2017, Standard 1.5.1 – Novel foods establishes that novel foods are non-traditional and thereby require safety assessments specific to adverse effects, composition, processing, source, consumption patterns [32]. Categories of non-traditional foods are given as plants or animals and components thereof, plant or animal extracts, herbs and extracts thereof, dietary macro-components, single chemical entities, microorganisms and probiotics, and foods produced from new sources or by a process not previously applied to food [32]. Further, the standard clarifies that a non-traditional food stipulates any of a food, food-derived substance, other component of food, as well as any other substance without a history of human consumption in Australia or New Zealand is non-traditional and therefore a novel food.

Retail sale of novel foods or foods with a novel ingredient may be sold if listed in section S25-2 and stipulated conditions of use are complied with (https://www.legislation.gov.au/Details/F2017C00413) [32]. Enquiries of the novelty of a particular food or food ingredient are handled by the Advisory Committee on Novel Foods and make recommendations to FSANZ pertaining to if it is a non-traditional food, if an assessment of public health and safety should be required given intended levels and use, and if the enquirer ought make an application to request the Food Standards Code be amended by FSANZ to undertake an assessment of public health and safety [33]. FSANZ also compilate a record of views formed on a larger body of foods and food ingredients to clearly present a summary of if each were identified by FSANZ as a traditional food, novel food, the justification of the view, and comments which may include safety concerns, labeling requirements, and pertinent sections of the Food Standards that may apply [32, 34].

1.2.2.2 Brazil

In Brazil, the statutory authority on food safety is the National Agency of Sanitary Surveillance (ANVISA) whose purpose is to protect the health of the population via sanitary control of food and pharmaceuticals [31]. Regulations for novel foods include Resolution No. 16, of April 30, 1999 (Resolução nº 16, de 30 de abril de 1999) and Resolution No. 17, of April 30, 1999 (Resolução nº 17, de 30 de abril de 1999) [35, 36], which established the mandatory registration procedures for novel foods and ingredients and the guidelines for safety assessments of novel foods and ingredients, respectively. ANVISA has since released a guidance document the Food and Ingredient Safety

Evidence Guide (Guia para Comprovação da Segurança de Alimentos e Ingredientes) [37]. Novel foods were defined as foods or food ingredients for human consumption without a history of consumption in Brazil or foods containing currently consumed ingredients but at much greater levels than currently observed excluding food additives, manufacturing aids, food provided as capsules, pills, tablets, or similar; and foods with novel ingredients given an exemption [31, 37, 38]. The registrant must include information pertaining to identity, scientific evaluations of safety, and comprehensive literature regarding the food or ingredient. Novel foods then undergo a safety and risk assessment on a case-by-case basis by the Techno-scientific Advisory Committee on Functional Food and Novel Foods (CTCAF).

As of July 2020, regulatory discussion has identified that the current regulations are highly subjective regarding what constitutes a safe history of use and clarity is required into the legal definition of novel foods and ingredients as well as the transparency into the safety assessments of novel foods and ingredients [38]. The proposed definition of a novel food would include foods and ingredients from vegetables, animals, minerals, microorganisms, fungi, algae, or synthetics without a history of safe consumption in Brazil as food and further specifies that novel foods and ingredients include but are not limited to: derivative products without a history of safe use, new or intentionally modified molecules, cell cultures or tissues thereof, those produced with a process not yet applied to food, products with an altered nutritional profile, nanomaterials, nutrient sources, bioactive substances, and substances only authorized for use in dietary substances if they were to be added to food.

1.2.2.3 Canada

The regulatory authority regarding novel foods in Canada is the Food Directorate of Health Canada and is responsible for assessments of novel food safety [31]. Novel foods are regulated as described by Division 28 of Part B of the Food and Drug Regulations [39], where B.28.001 stipulates that a novel food is a substance or microorganism without a history of safe use as food, or manufactured by a process not previously applied to that particular food and additionally causes the food to undergo a major change (i.e. places the food outside the accepted limits of natural variations for that food with regard to composition, structure, nutritional quality, physiological effects, alters the manner of how it is metabolized, or affects the microbiological safety, chemical safety, or safe use of the food), and lastly the food is derived from a plant, animal, or microorganism that has been genetically modified resulting in new characteristics, removed characteristics, or characteristics are outside of natural variation. The use of the term "substance" is not defined, although a novelty determination may be requested to clarify that a particular substance is either novel or non-novel (https://www.canada.ca/en/health-canada/services/food-nutrition/genetically-modified-foods-other-novel-foods/requesting-novelty-determination.html).

Novel foods require pre-market authorization for either sale or advertisement predicated on notification of Health Canada [31]. Petitioners provide pertinent information regarding the safety of the novel food such as history of use, dietary exposure, detail of production, and considerations of nutrition, toxicology, allergy, chemistry, and microbiology. The Food Directorate encourages consultations during development of novel foods to determine which data are necessary to demonstrate safety.

The Food Directorate has released guidelines on the safety assessment of novel foods [40], as well as a document providing transparency and insight into the Food Directorate's management process for novel foods [41]. The Food Directorate maintains a list of approved novel foods and genetically modified organisms with completed safety assessments and found to be safe for human consumption (https://www.canada.ca/en/health-canada/services/food-nutrition/genetically-modified-foods-other-novel-foods/approved-products.html). Further, The Food Directorate also maintains a list of non-novel determinations (https://www.canada.ca/en/health-canada/services/food-nutrition/genetically-modified-foods-other-novel-foods/requesting-novelty-determination/list-non-novel-determinations.html).

1.2.2.4 China

The most current statutory authority overseeing inquiries on novel foods and ingredients is the National Health Commission (NHC), which superseded the National Health and Family Planning Commission (NHFPC) as of March 2018 [42]. The most current novel food regulations are the Administrative Measure of Safety Evaluation of Novel Food Ingredients as of 2013, which established that novel food ingredients are those without a history of dietary use in China including animals, plants, microorganisms, components from any of animals, plants, or microorganisms; components whose original structure has been altered, and other newly developed foods. A history of dietary use was established as more than 30 years. Novel food ingredients cannot be introduced to the market without approval and safety evaluation from the NHC. Considerations are provided to novel food ingredients that have substantial equivalence to foods or prior

approved novel food ingredients not limited to genus, species, source, biological characteristics, composition, edible parts, and level of use.

The NHC evaluates the safety of novel food ingredients and provides recommendations with the assessment [42]. The assessment includes conclusions pertaining to the safety, technical need, and social impact. If approved by the NHC, the novel food ingredient may be used in general goods as well as health functional foods. Notably, new genetically modified organisms, ingredients for the exclusive use for health functional foods, or food additives are outside of the scope of the provisions of these evaluations. Announcements and interpretations for novel food ingredients are posted on the NHC website (http://www.nhc.gov.cn/sps/s2909/new_list.shtml).

1.2.2.5 Eurasian Customs Union (EACU)

The Eurasian Economic Union (EACU) formed as of 1 January 2015 and currently comprises of Armenia, Belarus, Kazakhstan, Kyrgyzstan, and the Russian Federation to represent a single market [43]. Across the EACU, the regulations for novel foods are the Technical Regulations of the Customs Union TR CU 021/2011 [44]. Within the EACU, novel foods are defined as food products including additives or flavorings without a history of consumption by humans within the customs territory such as a new or modified molecular structure, microorganisms or isolates thereof, microscopic fungi or algae, plants, isolates of animals, genetically modified organisms, nanomaterials, or nanotechnology products (Article 4). Classes specifically excluded from this definition include food products obtained by traditional methods, are currently in circulation, and are by virtue of experience safe. Food ingredients are not specifically identified as falling

within this regulation nor are new production processes unless they specifically modify molecular structures [45].

Novel food products are subject to state registration as it is being produced within the EACU or before importing into the EACU [43], where state registration is carried about by a nationally authorized body dependent which nation the application is being filed [43, 45]. Registration requires information regarding scientific procedures to confirm the novel food product is safe for humans and information regarding its effects on humans to establish no adverse effects. After successful registration, the novel product is no longer novel and not subject to registration by other applicants, although each must conform to the initial registration. Registered applications and documents are available to the public (http://eec.eaeunion.org/).

1.2.2.6 European Union (EU)

The European Union (EU) comprises of 27 member states and with Norway, Iceland, and Lichtenstein comprise the European Economic Area (EEA). The European Commission (EC), the executive branch of the EU, acts as the regulatory authority regarding novel foods with the European Food Safety Authority (EFSA) acting as the advisory scientific body [31]. In 2015, the EC established the current definitions of novel foods per Regulation (EU) 2015/2283 that applies across the EEA [46]. Using 15 May 1997 as a reference date, novel foods were defined as any food not used for significant human consumption within the Union and is any of: a food with a new or intentionally modified molecular structure, food consisting of or isolated from microorganisms, fungi, algae, minerals, plants, animals, or nanomaterials; are from a

production process not yet used for food production in the Union and resulting in significant changes to the composition or structure of the food, nutritional value, metabolism, or level of undesirable substances; vitamins, minerals or other substances produced with a process not yet used for food production in the Union, or foods used exclusively in food supplements with intent to use in foods other than supplements. Novel foods were specifically segregated from genetically modified foods, enzymes, additives, flavorings, or extraction solvents.

Novel foods require pre-market approval predicated on an evaluated lack of risk to human health, its intended use is not misleading, and if substituted for current foods its substitution would not be a detriment to human health [46]. Further, those producing novel foods are compelled to verify the novelty of their food by consulting the country where they plan to place the novel food. The administrative and scientific requirements for novel food applications were established per Regulation (EU) 2017/2469 [47]. The applicant provides a technical dossier including their own safety assessment data, including biological or toxicological assessments, as well as their proposed conclusion from their provided data to enable the EC to consult EFSA for an opinion on the overall risk assessment of the novel food and highlighting uncertainties and limitations of the evaluation. The EC catalogues lists of novelty determinations (https://ec.europa.eu/food/safety/novel_food/catalogue/search/public/index.cfm#) as well as lists of currently authorized novel foods (https://ec.europa.eu/food/safety/novel_food/authorisations/union-list-novel-foods_en).

1.2.2.7 India

Novel foods in India are under the supervision of the Food Safety and Standards authority of India (FSSAI) [48]. Novel foods are currently regulated according to the Food Safety and Standards Regulations, 2016 stipulating that a novel food may not have a history of human consumption, may have any ingredient that may not have a history of human consumption, or made from a new technology and process resulting in significant changes in the structure to alter nutrition, metabolism, or undesirable substances [49]. Individual classes or types of novel foods are not specified within this regulation.

The pre-market approval process has been explained in the Food Safety and Standards Regulations, 2017 [50]. As per guidance, novel foods may not be manufactured or imported without prior approval, where the application includes information on the source, function and intended use, scientific analyses, regulatory status in other countries, and documents on risk and toxicological assessments. Regulation, 2017 gives more insight into what constitutes a novel food, which includes novel foods and ingredients, foods processed with novel technology, additives, processing aids and enzymes, as well as foods consisting of or isolated from microorganisms, bacteria, yeast, fungi, or algae.

1.2.2.8 Japan

In Japan, the primary statutory authority regarding food sanitation is the Ministry of Health, Labor, and Welfare (MHLW) [31, 51]. Under the Food Sanitation Act, there is no segregated definition of a novel food or ingredients nor delineated procedures for approval (Article 4, Food Sanitation Act). Rather, the Food Sanitation Act has provisions to prohibit sale of foods that have not been generally served for human consumption and

have not demonstrated a lack of risk to human health, have been served for human consumption but served in an extraordinarily manner, or where serious damage to human health has been attributed (Article 7, Food Sanitation Act). In each case, the MHLW is advised by the Pharmaceutical Affairs and Food Sanitation Council (PAFSC). Genetically modified organisms are separately assessed by the Food Safety Commission (FSC) but is also responsible for risk assessment of food hazards.

1.2.2.9 Saudi Arabia

As of the end of 2019, Saudi Arabia's Saudi Food and Drug Authority (SFDA) has proposed a set of draft requirements to regulate novel foods [52]. The draft regulations pertain to microorganisms, fungi, algae, minerals, plants, animals, cell cultures or tissues, and nanomaterials as well as foods with new or modified molecular structures, foods derived from new production processes, and foods that are traditionally consumed from other countries. The regulation would require pertinent information such as the scientific basis for safety under intended use or in the case of traditional foods from elsewhere would require information about history of safe use. The SFDA would take the information provided into consideration as well as the similarity to foods that are understood as safe.

1.2.2.10 Singapore

The Singapore Food Agency (SFA) is investigating a framework for novel foods and had presented its proposed regulatory framework for comments [53]. The proposed framework seeks to clearly define and gird the scope of novel foods, establish safety

criteria, and application processes. The proposed definition is that a novel food is one not used for human consumption in Singapore or outside of Singapore for more than 20 years and is from an unconventional source or prepared by an unconventional process. Examples provided include foods isolated from or produced from plants or animals or their parts; isolated or produced from microorganisms, fungi, algae; food with new chemical structures not previously found in food whether such food is synthesized from raw materials or manufactured from a process not conventionally used in food production; food derived from biologically synthesized substances such as genetically modified organisms, tissue culture, cell culture, or cloning; foods consisting of intentionally engineered nanomaterials. Therefore, foods with a history of safe use are not novel foods, where the history is to be considered based on length of consumption and use, extent of use, quantity used, purpose and context of use, evidence of lack of adverse health effects. Delineated examples of novel foods include plant parts without a history of safe use, insects without a history of safe use, animals without a history of safe use, isolates from insects, refined extracts from animal products, newly identified microorganisms without a history of safe use, substances synthesized using food-grade raw materials (e.g. enzymatically modified ingredients), cell-based meat, engineered nanomaterials (particle size < 100 nm).

1.2.2.11 United States of America (USA)

The regulatory authority overseeing food additives in the United States (USA) is the United States Food and Drug Administration (USFDA) [31]. Under current law, there is no delineated definition of novel foods or ingredients; however, any substance

reasonably expected to become a component of food is a food additive and subject to pre-market approval by the USFDA, unless that component is generally recognized as safe (GRAS) or meets an exclusion from the definition of food additive according to section 201(s) of the Federal Food, Drug, and Cosmetic Act. Food additives approved either by USFDA pre-market approval or GRAS is evaluated under the conditions of its intended use and the use of a food additive above that which was approved would require reevaluation [31, 54].

Substances that have not been documented as GRAS or USFDA approved, listed, or excluded require USFDA pre-market approval using a food additive petition [54]. The technical elements necessary include chemical composition, proposed use, levels of use, data regarded intended effects, quantified detection levels, estimated exposure levels, safety reports, proposed tolerances of production, and environmental impact information. The USFDA provides various guidance documents for identifying pertinent regulations as well as recommendations for chemical, toxicological, microbial, and environmental testing, evaluations, and assessments (https://www.fda.gov/food/food-ingredients-packaging/food-additives-petitions).

An alternative to a food additive petition is GRAS status, which can be determined either through demonstrated food use in the USA prior to 1958 or by scientific procedures accompanied by a general recognition of safety based on the views of qualified experts [54]. It is becoming more uncommon for food additives to be approved by a history of use prior to 1958, and therefore GRAS by scientific procedures is the principal means to attain GRAS. The data required for GRAS is like that of a food additive petition such as technical safety evidence for intended use but also a

demonstration of safety using evidence that is generally known (i.e. publicly available information). A key difference is GRAS does not have to be filed with the USFDA and providing a notification of GRAS is voluntary and so some companies convene a panel of experts as a GRAS panel to evaluate a self-affirmation of GRAS while others document the self-affirmation without a GRAS panel. Despite the voluntary notification scheme, the USFDA is interested in GRAS notifications of foods that may pose a risk to the US population. If the petitioner decides to notify the USFDA of its GRAS determination, the USFDA responds to GRAS notifications either with acceptance without questions, rejection for insufficient data or unclear safety, or stop by request of the inquirer. The USFDA maintains lists of GRAS notifications (https://www.cfsanappsexternal.fda.gov/scripts/fdcc/?set=GRASNotices), food additives (https://www.fda.gov/food/food-additives-petitions/food-additive-status-list), and color additives (https://www.fda.gov/industry/color-additive-inventories/color-additive-status-list).

1.3 Food allergy

1.3.1 Overview

Allergic disease encompasses a broad spectrum of disorders characterized by abnormal immune responses. Atopic diseases, in other words diseases with a genetic predisposition toward development of allergic disease, encompass asthma, allergic rhinitis, hay fever, atopic dermatitis, and food allergy, although the relationships between these diseases and their immune response have not been fully clarified [55]. Those with atopic diseases are commonly predisposed to greater production of immunoglobulin E

(IgE) antibodies, but many people with elevated levels of total and allergen specific IgE do not present with atopic disease. Some present with one specific atopic disease whereas others co-express several. Atopic diseases are often categorized umbrella terms such as asthma or allergic rhinitis but it is more appropriate that each be understood as a collection of diseases merely grouped by the phenotype of the disease [56].

Causative agents in Immunoglobulin E (IgE)-mediated hypersensitivity reactions that can first sensitize an individual to prime the immune system and secondly elicit a reaction and develop symptoms are broad. Such agents include outdoor allergens (grass and tree pollen) [57], indoor allergens (house dust mites, pets, cockroaches) [58], insects stings and/or bites [59], as well as food [60], where the only difference among these is the source of the allergen. Food allergy is an abnormal immunological response to foods or their components [60], typically a naturally present protein and are differentiated into antibody-mediated immediate hypersensitivity reactions, cell-mediated delayed hypersensitivity reactions, and mixed antibody-cell mediated conditions. IgE-mediated hypersensitivity reactions account for most food allergic reactions [61]. Other food sensitivities and intolerances may not be immune mediated. Food intolerances are metabolic food disorders that become evident upon consumption of certain foods (lactose intolerance), and food sensitivities are idiosyncratic reactions to specific food components occurring through unknown mechanisms (sulfite-induced asthma) [60].

Symptoms of IgE-mediated food allergy are diverse where symptoms include pruritis (itching) and numbness and objective symptoms include hives, eczema, swelling, wheezing, chest tightness, chest pain, nasal congestion, trouble breathing, abdominal pain, diarrhea, nausea, vomiting, constriction of the airways, drops in blood pressure, and

anaphylaxis (Table 1.1) [62]. Symptoms can be evident within minutes and can last for several hours. In some cases, symptoms can appear hours after ingestion as in the case of alpha-gal syndrome, a delayed syndrome related to red meat ingestion due to sensitization through tick bites [60, 61]. The severity of reactions is understood to follow a dose-response relationship [62], although this is complicated by factors such as the allergen itself, individual host factors, and interventions to control the reaction [63].

Table 1.1 Symptoms by system associated with IgE-mediated reactions to food. Adapted from [3].

System	Symptoms	
Cutaneous	• Angioedema • Flushing	• Pruritis • Urticaria
Ocular	• Conjunctival erythema • Periorbital edema	• Pruritis • Tearing
Respiratory	• Chest tightness • Coughing • Dyspnea • Hoarseness • Laryngeal edema	• Nasal congestion • Pruritis • Sneezing • Wheezing
Gastrointestinal	• Abdominal cramps • Angioedema of tongue, lip, or pharynx	• Emesis • Nausea • Oral pruritus
Cardiovascular	• Bradycardia • Ventricular dysfunction • Dizziness	• Hypotension • Tachycardia
Neurologic	• Dizziness • Sense of impending doom	• Syncope

A risk factor for development of food allergies was once thought to be early exposure to the offending allergen, but early exposure was found to be a protective factor rather than a risk factor according to the Learning Early about Peanut Allergy (LEAP) trial [64]. Environmental factors also impact the development of allergies as protective

effects have been ascribed to farm exposure, breastfeeding, and exposure to other children; particularly older siblings, but risk factors include air pollution and pet ownership [65]. Genetics are a major factor as studies on twins has shown significant heritability of phenotypes of allergic disease, specifically a higher correlation of heritability has been shown for monozygotic twins (sharing 100% of genes) than dizygotic twins (sharing 50% of genes) [66]. A broad array of genes has been identified where specific gene variants have been associated with greater susceptibility to various allergic diseases. A key example is the gene for filaggrin (FLA), which has been associated with asthma, atopic dermatitis, and food allergy as the product filaggrin is a major component of the protein-lipid envelope of the epidermis and thereby important for establishment and maintenance of the epidermal barrier. Meta-analysis has shown that there is a strong dose-dependent association between atopic dermatitis, food sensitization, and food allergy as well as evidence that atopic dermatitis precedes food sensitization and food allergy [67].

There is currently no cure for food allergies. Therefore, management of food allergies from a consumer perspective is often required and demands careful reading of food package labels in addition to dietary choices [68]. Nonetheless, relying on package labels can leave patients at risk of accidental exposure due to cross-contamination with allergenic foods or incorrect labeling [62]. To reduce the burden of food allergies, immunotherapies have been explored to increase the tolerance of allergic individuals toward their offending allergens such as oral (direct consumption), sublingual (held under the tongue), and epicutaneous (dermally applied patch) immunotherapies [69].

1.3.2 Mechanisms of IgE-mediated hypersensitivity

During sensitization, the immune response is driven by differentiation of naïve T cells into T_H2 cells as well as class switching of naïve mature B cells from production of non-IgE antibodies to IgE production (Figure 1.2) [70].

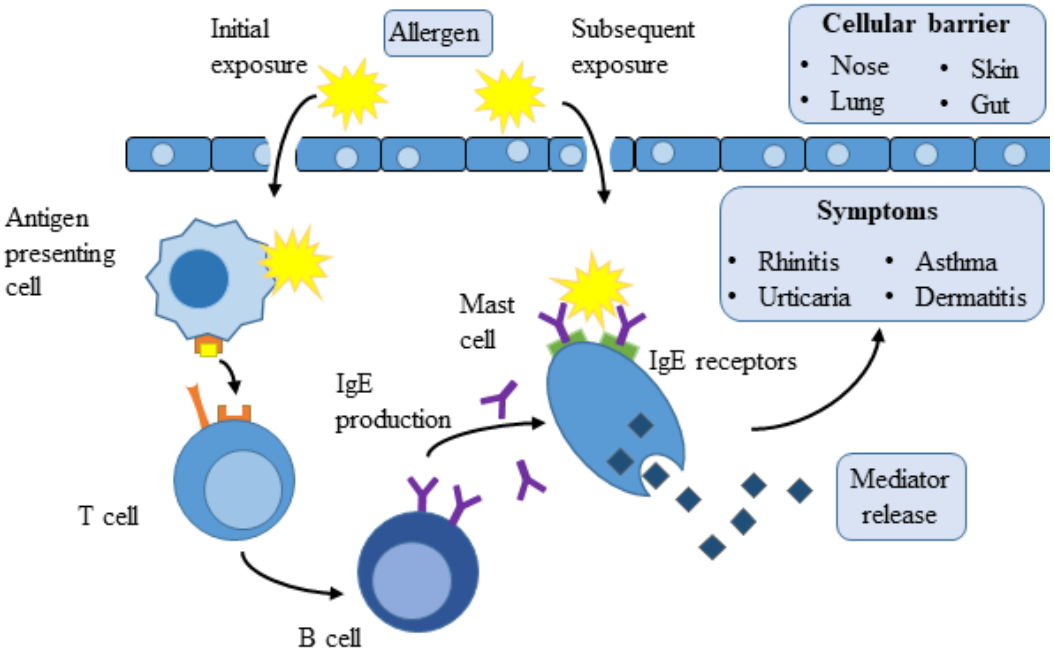


Figure 1.2 Mechanism of cellular activation in response to allergenic insult and subsequent recognition by IgE.

Adapted from [4].

Differentiation of T cells occurs as antigen presenting cells (including but not limited to dendritic cells) present an antigen-derived peptide to the T cell along with appropriate co-stimulating molecules to result in activation. Upon activation, the T_H2 cells secrete cytokines [Interleukin (IL)-4, IL-13], which encourage continued T_H2 differentiation, proliferation, as well as B-cell IgE class switching and production. Activated CD4 T-cells along with activated B-cells can result in B-cell differentiation into antibody-secreting

plasma cells and memory cells, which persist and respond to future insults with IgE production long after sensitization. Produced IgE can strongly bind to the surface of mast cell and basophil FcεRI to prime them for activation. During sensitization, there is no active immune response and therefore no symptoms.

Subsequent antigenic insults result in antigens binding to IgE and subsequent cross-linking of FcεRI resulting in both activation and degranulation of mast cells and basophils [70]. The granules release potent mediators to produce local, tissue, and systemic responses, which are responsible for the symptoms associated with immediate hypersensitivity reactions [60]. Released histamine from the granules can elicit inflammation, pruritis, and vasoconstriction of blood vessels of the gastrointestinal and respiratory tracts. Other released molecules include lipid mediators, leukotrienes, and prostaglandins, where leukotrienes are related to symptoms that develop slowly as in late-phase asthmatic reactions.

Exposure to food proteins do not typically result in the development of allergic disease nor the formation of IgE antibodies. Even in atopic individuals exposure to food proteins in the gastrointestinal tract promotes oral tolerance through the formation of protein-specific IgG, IgM, and/or IgA antibodies [60]. Other mechanisms of tolerance include clonal anergy, apoptosis of antigen specific T cells, active immune suppression, and bystander suppression. Regulatory T cells (Tregs) are involved in controlling the response of the immune system by inhibiting T cell proliferation and differentiation through the cytokines IL-10 and TGF-β [70, 71], which act as a key component of oral tolerance and maintain healthy immune responses to allergens [72].

Other types of reactions are also meaningful as in IgG mediated hypersensitivity reactions caused by IgG binding to cell surfaces resulting in cell death (penicillin allergy), immune complex mediated reactions resulting from the formation and deposition of antigen-antibody aggregates (serum sickness), and delayed hypersensitivity reactions incorporating CD8 and T_H1 cells and pathways (mosquito bites) [70]. Celiac disease is a non-IgE mediated hypersensitivity triggered by gluten proteins in various grains, predominantly wheat. Celiac is characterized by IgG and IgA responses to gluten after the effects of endogenous transglutaminase has deamidated gluten and "produced" the necessary antigen in the form a gluten-transglutaminase complexes. Resultant T cell activation and subsequent proinflammatory response is coupled with an autoantibody response causing damage to the small bowel and atrophy of the villi leading to long term damage and nutrient uptake deficiencies [73].

1.3.3 Diagnosing food allergy

Diagnosis of food allergies begins with the clinical history of the patient. Information such as the food consumed, how the food was prepared, symptoms experienced, severity of symptoms, and the interval between consumption and onset of symptoms are all incorporated to determine the likelihood of food allergy [74]. Further, background regarding atopy, current food allergies, current medications (e.g. antihistamines), and diet can refine the diagnosis and identify if any food sensitivities are present, if the manifestation is a food allergy, and if offending foods can be singled-out [75]. Clinical histories substantiating a possible food allergy are complimented with skin-prick testing, serum IgE testing, and food challenges [76].

Skin prick tests act as a relatively quick and cheap method to establish allergic sensitization. The test is performed by puncturing the skin with food extract, puncturing the suspect food and then the skin, or through intradermal injection and compared against negative (e.g. saline solution) and positive (e.g. histamine) controls [77]. A positive reaction is evaluated in terms of the resultant wheal and the size measured, where a wheal diameter greater than 3 mm is considered positive. Skin prick results must be analyzed carefully as wheal sizes vary according to age, site used for testing, skin prick technique used, and the antigen used for the test whether fresh or commercial [78]. Skin prick tests can have negative predictive values up to 95% but this varies greatly depending on the food tested and the type of extract used as commercially produced plant extracts can result in sensitivities of as low as 20% [74]. Positive skin prick tests only confirm the presence of allergic sensitization as opposed to allergic disease, although positive results coupled with a compelling clinical history are strongly incriminating [77].

Evaluation of allergen-specific IgE acts as an alternative and compliment to skin prick testing, but is more costly, time-consuming, and similarly does not discriminate allergic disease and allergic sensitization [74]. Methods to measure allergen-specific IgE include fluorescent enzyme immunoassays such as ImmunoCAP and Immuno Solid-phase Allergen Chip (ISAC). These tests are performed in a laboratory setting where serum is applied to immobilized antigens and detected by anti-IgE antibodies with fluorescent tags where greater fluorescence indicates more bound IgE. These tests are semi-quantitative, and so are not perfectly comparable across different methodologies, but are used to determine predictive values of allergic disease as increasing concentrations of measured IgE are correlated with greater likelihood of allergic disease

[79]. When measured specific-IgE are greater than or equal to 95% of the predictive value then food allergy is likely, whereas if levels are less than or equal to 50% of the predictive value then oral challenge may be carried out to exclude food allergy [78]. These tests can be further leveraged for component-resolved diagnostic testing to identify individual proteins binding a patient's serum IgE and establish patterns among foods with homologous proteins.

When clinical history and diagnostic tests are unclear, food challenges can be performed to assess clinical reactivity. Food challenges can be performed as a double-blind placebo-controlled food challenge (DBPCFC) or open food challenge (OFC). Food challenges can be expensive and necessitate a highly controlled environment and so are not performed without careful consideration of alternative means of diagnosis. Open food challenges are subject to observer bias and psychogenic factors resulting in a greater false-positive rate but are more practical in clinical settings [74, 78]. Double-blind challenges are the preferred diagnostic method, but the suspect allergen must be appropriately masked in a vehicle food to be indistinguishable from a placebo, which can be difficult in the case of easily detected allergenic foods such as apple or shrimp. In either method, the food should be given in a tiered manner given timed intervals and beginning with very low doses to determine the threshold dose resulting in symptoms. The food challenge ends upon either completion of the challenge without symptoms or discontinued upon development of symptoms and the patient treated immediately. In some cases, food challenges can result in anaphylaxis, underscoring that food challenges should be conducted by trained professionals in appropriate settings [79]. Negative

results allow the patient to introduce the challenged food into their diet, but positive results indicate that avoidance is necessary.

1.3.4 Prevalence and impact of food allergies

In 1995, the FAO-WHO Expert Consultation on Food Allergies of 1995 identified eight major food and groups associated with over 90% of food allergies: gluten containing grains, crustaceans, eggs, fish, peanuts, soybeans, milk, and tree nuts [80]. In 1999, the list was included in the Codex Alimentarius and later adopted by many countries as the basis for their laws regarding labeling of allergenic foods. There are differences in prevalence of individual food allergens as in celery allergy in Europe, sesame allergy in Israel, and buckwheat allergy in Japan [60], which is commonly reflected in differences among countries food allergen labeling laws.

Recent surveys have determined that the estimated total prevalence of convincing food allergy in the U.S. is at least 10.8% given that of those surveyed 19% self-reported a food allergy [81], which underscores the importance of clinical diagnosis and the proper education of those at risk. However, the disparity among self-reported and clinically confirmed food allergy identifies a problem that the true prevalence of food allergy is not accurately known. Large epidemiologic studies incorporating studies of questionnaires and surveys may not be accurate to the true prevalence of food allergy [56]. Considerable heterogeneity is present among studies of food allergy prevalence due to differences among how studies may incorporate likely, but unconfirmed, food allergy as opposed to prevalence of clinically or objectively confirmed food allergy. Among well designed studies, variability of questionnaires, IgE testing, skin prick testing material, and oral

challenge material can cause the prevalence of food allergy to vary considerably across studies [82].

Identifying those at risk is difficult as in the case of children born in East Asia or Africa but raised in the West being at greater risk of developing food allergy compared to Caucasian counterparts [83]. Many factors impact the prevalence of recorded food allergy including geography, diet, age, race, and ethnicity [75]. The prevalence of food allergy is generally regarded as increasing although uncertainty exists due mostly to lack of clinical confirmation of food allergy (i.e. food challenges) in favor of skin prick tests, IgE-based biomarkers, or a lack of any clinical evaluation [84].

Food allergies impose great social cost on those affected and while food allergy has been historically understood as a pediatric disease a great and increasing proportion of adults are affected [85]. The annual economic cost of food allergies has been estimated at above $4,000 per child including costs of hospitalizations, clinician visits, and medications. Those of lower socioeconomic status are most affected as they have a greater propensity to spend more on hospitalizations and emergency room visits. In addition to out-of-pocket costs, lost opportunity costs impose a great burden on those with food allergy via lost wages to take themselves or others to the hospital as well as potentially restricted career choices to take care of children with food allergies [86]. For those self-reporting food allergies, some of these costs could also be undertaken and thereby impose a burden even without the presence of disease.

1.3.5 Properties and characteristics of allergenic food proteins

Formal nomenclature of allergens is performed by the International Union of Immunological Societies (IUIS) as the IUIS Allergen Nomenclature Sub-Committee under the purview of the World Health Organization (WHO) [87]. The nomenclature of allergenic proteins is based on Linnaean binomial nomenclature identifying the genus and species of all organisms. The base format includes the abbreviated genus name, abbreviated species name, allergen number, and a four-digit number representing the isoallergen and variant numbers (e.g. Ara h 10.0102; *Arachis hypogaea* 10th allergen, isoallergen 1, variant 2). Allergen numbers were initially assigned in order of discovery but has since been assigned to organize allergenic proteins by similarity for clarity. Isoallergens are those from the same species representing similar function with a shared sequence identity of >67% and variants are those with a shared sequence identity of >90% [87].

Epitopes are specific chemical groups in an antigen that determine the specificity of an antigen [88]. Epitopes are the basic structural unit of both B and T cell receptors and antibodies and can be divided into either B or T cell epitopes according to which cells they bind and divided further into linear or conformational epitopes. Linear epitopes are made of sequentially contiguous amino acids as compared to conformational epitopes, which are composed of amino acids representing a spatially close region formed due to protein folding. Cross-linking of FcεRI requires two antibody molecules to bind to the same allergen. Epitopes adhered to antibodies or either T cell or B cell receptors are held by a binding pocket allowed by the receptor's amino acid side chains and by non-covalent interactions with the epitope [70]. Although the receptors are highly specific for

their antigen, binding is predominantly determined by anchor resides to determine the principal specificity of the binding pocket.

Cross-reactivity is recognition of multiple antigens by antibodies with a single specificity [89], in that antibodies produced against a specific protein may bind to related sequences and structures of the same or similar organisms. Cross-reactivity permits reactivity without initial exposure to the exact same epitopes, rather only highly similar epitopes. Factors affecting cross-reactivity include common proteins, epitopes, and folds [88]. In allergy, cross-reactivity results in a broader array of antigens that can result in allergic reactions, and thereby is advantageous for protection against pathogens but not in the case of dietary proteins [89]. Families of cross-reactive allergenic proteins have been identified plants including profilins, prolamins, cupins, and Bet v 1-like proteins but also animals including parvalbumins, tropomyosins, and arginine kinases [88, 90]. Cross-reactivity also leads to some odd couplings and disparities when focusing on foods rather than the cross-reactive allergens as those with cow's milk allergy are highly likely to be reactive to goat's milk but seldom reactive to horse's milk or another example of the relatively common cross-reactivity of those with latex allergy reacting to any of kiwi, banana, or avocado [91]. Cross-reactivity also bridges the gap between aero-allergens and food allergens as in pollen-food syndromes between, for example birch-apple syndrome through homology of birch Bet v 1 and apple Mal d 1 [92].

To determine if a particular protein has the capacity to elicit reactions, proteins can be evaluated to determine the plausibility of their allergenicity. Decision trees and weight-of-evidence approaches have been adapted from evaluations of proteins derived from biotechnology to aid in the process of evaluating potentially allergenic proteins

[93]. Individual proteins can be evaluated on the basis if there are other allergenic proteins known from the organism, if the protein is homologous to other known allergens, if sera with known reactivity have IgE that can bind the protein, enzyme digestibility assays (e.g. pepsin, trypsin, chymotrypsin), and animal models (e.g. BALB/c mice). Knowledge about the properties of the protein in question can also be informative as allergens tend to have similar properties such as maintenance of protein folds (calcium binding, lipid binding, and disulfide bonding leading to heat resistance), maintenance of primary sequence (resistance to proteolysis), repeated epitopes (repeating primary structures, multimeric quaternary structures), and unique motifs (glycosylation and glycation sites) [94], where each of these properties results in a greater capacity to distinguish the protein from self-proteins, allow for single epitopes to be bound multiple times across a single antigen, as well as maintain the allergen for the immune system to identify. Predictive tools aid in distinguishing plausible allergenic proteins, for example AllergenOnline [95], AllerCatPro [96], and BepiPred [97], to predict allergenic proteins on the basis of their sequence, structure, and B-cell epitopes. Guidelines on predicting the allergenic potential of proteins based on sequence alone include CODEX guidelines of greater than 35% identity over segments of 80 amino acids or 100% identity in 6 or 8 amino acids, but prediction based on greater than 50% identity over the full length has been regarded as the most predictive [95].

1.3.6 Food processing and food allergens

For known allergenic proteins in commonly consumed foods the question is what can be done about them. Various forms of food processing are common in preparation of

food such as thermal, high pressure, enzymatic, and fermentative processing. Thermal processing is one of the most common and preferable methods to prepare food for desirable organoleptic properties [98]. Thermally processing allergenic proteins can result in reduced IgE-binding for some proteins but can expose hidden epitopes for others. For example, Bet v 1-like allergens tend to show greatly reduced IgE binding after heat treatment but Ara h 1 shows similar IgE binding before and after baking. High-pressure processing (HPP) is an emerging nonthermal technique often used for inhibition of microbial growth. The underlying principle is that application of pressures above 200 MPa can affect higher order structures of protein without affecting covalent bonds, which allows for nutritional value, flavors, and aromas to be unaffected. The efficacy of high-pressure processing is still under investigation but shows promise as an alternative to thermal processing for microbiological hazards.

Processing methods that directly impact the sequence of allergenic proteins tend to have great effect in reducing their capacity to bind IgE. Enzymatic hydrolysis is a well-known nonthermal processing technique commonly used for foods such as milk, lentils, and peanuts, although this technique is heavily reliant on proteins that are susceptible to the proteolytic cleavage for the end result to be a greatly diminished allergenic protein [99]. Further, a major drawback of enzymatic hydrolysis are the negative impacts on functional properties and emergent bitterness and astringency from resultant small polypeptides [98]. Alternatively, fermentation is one of the oldest food preservation methods. Utilizing microbes to produce enzymes and act upon the food results in changes in texture, flavor, and functional properties [99]. Milk and wheat have consistently shown that fermentation can result in reductions in IgE binding, although the range of foods

where fermentation is applicable is limited as the fermented product is often greatly different from the starting material in terms of both function and taste [98].

1.3.7 Novel sources of food allergens

Accompanying novel foods and processing comes food allergens that may be themselves novel, processed to render a well-known allergen to be modified in a novel manner, or a well-known allergen being consumed or ingested in an alternative manner. Understanding a novel source of food allergens then demands insight into which proteins are present, how they are affected by processing, as well as how to both detect and quantify present proteins that may be hazardous to consumers. Accurate identification and evaluation of the presence and quantity of food allergens is heavily dependent on the methodology used.

1.4 Quantitative methodologies for the detection of food allergens

There are currently a wide variety of methodologies that can detect food allergens, although a smaller subset has been widely accepted and utilized for quantitative determinations. Methods for quantitative evaluation of food allergens include polymerase chain reaction, enzyme-linked immunosorbent assay, and mass spectrometry (Table 1.2).

Table 1.2 Pros and cons of quantitative methodologies to detect food allergens.

Methodology	Pro	Con
Polymerase chain reaction	• High throughput • Multiplex potential • Fast • DNA is tolerant of robust extraction	• Detects DNA rather than protein • DNA may not correlate to protein presence
Enzyme Linked Immunosorbent Assay	• High throughput • Industry standard • Fast	• Requires antibodies • Antibodies can cross-react • Requires calibration to reference materials • Extraction requires antibody tolerance
Mass spectrometry	• Multiplex potential • Can be absolutely quantitative • High sensitivity to all present proteins	• High capital inputs • High levels of expertise • Time consuming • Limited by data analysis

1.4.1 Polymerase chain reaction

Polymerase chain reaction (PCR)-based methods allow for highly specific and sensitive detection and quantification of DNA stretches, rather than protein, from allergenic sources of interest beginning with extraction and purification of present DNA, amplification of DNA, and subsequent detection of the DNA [100, 101]. PCR utilizes thermal cycling and repeated cycles of both heating and cooling to denature DNA and enzymatic replication with Taq polymerase [100]. Repeated cycles of denaturation, annealing, and extension takes place until DNA of interest is sufficiently amplified. Key factors in an effective PCR are the specificity of the stretch of DNA and the specificity of the oligonucleotides chosen as primers for the reaction [100, 101].

Real-time PCR is preferred for quantitative analysis of allergenic food- derived DNA [100]. In real-time PCR, the reaction tube additionally contains a target specific oligonucleotide probe and fluorescent reporter dye attached to a quencher. Detection of fluorescence is prevented due to proximity of the dye and quencher but when the DNA-hybridized probe is exposed to the 5' exonuclease activity of Taq polymerase, the dye is displaced into solution. As the reaction progresses, more dye is displaced into solution and the reaction can be monitored and measured as fluorescence is proportional to the amount of target DNA present.

In principle, PCR allows for a greater diversity of valid methods to be developed per allergenic food compared to protein-based methods as three nucleotides represent single codons as translated to amino acids where degenerate use of codons allows for otherwise indistinguishable proteins to be distinguished by PCR in addition to untranslated introns [101]. PCR methods only require the genetic code and so could be utilized as open-source methods as compared to the limited and often proprietary antibodies utilized by antibody-based methods. The time to run a PCR is comparable to that of enzyme-linked immunosorbent assays (ELISA) and faster than mass spectrometry (MS). In a well-designed PCR, simultaneous detection of several DNA stretches, i.e. multiplex allergen detection, is possible.

DNA is generally more tolerant toward harsh extraction procedures than protein and so PCR benefits; however, matrix impurities can impair detection more so than protein-based methods such as polysaccharides, polyphenols, fats, minerals, and enzymes [102]. PCR is not universally applicable to allergenic foods as both milk and egg are poorly detected by PCR due to low levels of present DNA [100]. Use of PCR requires

calibration curves to convert quantified allergenic food-derived DNA into total allergenic food as quantities of DNA are not universal across foods [101]; further, as PCR does not directly detect the proteins, their utility in assessment of the risk posed is limited [100].

1.4.2 Enzyme-link immunosorbent assays

Enzyme-linked immunosorbent assays (ELISA) are the most widely used tool by the food industry for detection of specific allergenic proteins [100]. The antibodies used for ELISA are typically IgG from animal sources, usually rabbits, goats, or sheep, which allow for suitable quantities of antibody with minimal variability as compared to using human serum IgE. Quantitative ELISAs come in either of sandwich or competitive formats, each with their own benefits and drawbacks. Sandwich ELISAs are the most common format for allergen detection where an IgG capture antibody is immobilized on a solid phase, such as a 96-well plate, and the food is extracted and applied to the plate and subjected to the immobilized IgG. After incubation and washing, a reporter IgG labelled with an enzyme (e.g. horseradish peroxidase) is added to sandwich present allergenic proteins. To quantify present allergenic protein a substrate is added, and the enzyme labelled IgG develops a colored product that can be measured. Color is proportional to the amount of present allergenic protein and is quantified by comparison to a standard curve.

Alternatively, competitive ELISAs are available where allergenic proteins of interest are immobilized on the solid phase and separately the sample extract is pre-incubated with allergen-specific IgG [100]. When the sample-IgG mixture is applied to the wells, IgG that are unbound will bind to the plate. After washing, a reporter IgG is

added, color developed, and colored product is measured. In competitive ELISA, the color developed is inversely proportional to present allergenic protein.

Sandwich ELISAs require two IgG epitopes per target protein whereas competitive ELISA only requires a single IgG epitope [100], which allows competitive ELISA to have an advantage detecting fermented or hydrolyzed proteins where target proteins may not be wholly intact for a sandwich ELISA to be wholly effective. ELISA can be performed relatively simply with trained personnel and using inexpensive equipment [101]. There are many commercially available ELISA kits that encompass most food allergens, although different kits for the same allergenic food may not detect the same protein fractions or allergens (e.g. the whey versus casein fractions of milk) [100].

The performance of ELISA can vary among samples from different food sources according to matrix effects and food processing effects [101]. Proteins may not be effectively extracted due to protein aggregation, denaturation, or chemical modification, which limits comparison to calibrants that may not be representative of the same matrix, processing state of the allergenic protein in the food [101, 103]. Polyclonal IgG and the calibration curves are subject to batch variation and may further vary across manufacturers, producing different results across kits detecting the same allergenic proteins [101].

1.4.3 Mass spectrometry

Mass spectrometry has historically been applied to the identification and characterization of proteins, but only more recently has been applied to quantification of

food allergens [100, 104]. The most applied quantitative workflow is called bottom-up, where the measured targets are peptides derived from enzymatic digestion of allergenic food-specific proteins [105]. A generalized quantitative workflow has four steps: 1) protein extraction, 2) proteolytic digestion, 3) analysis of identified peptides for evaluation, 4) iterative targeted evaluation of selected peptides. Extraction is often the most important for sensitive detection of target peptides as the food matrix can contain many non-target proteins and peptides as well as starches, fats, and other interfering molecules that can be a detriment to accurate quantification [106]. After extraction, the proteins can be prepared for mass spectrometry by reduction to break present disulfide bonds, alkylation to prevent reintroduction of disulfide bonds, and proteolysis that commonly features either trypsin or chymotrypsin to reproducibly cleave the proteins into peptides [100].

To analyze the complex mixture of peptides, mass spectrometers are commonly coupled with reverse phase high performance liquid chromatography (RP-HPLC) to fractionate the peptides according to their hydrophobicity and feed the peptide fractions into an electrospray ion source (ESI) [101]. Mass spectrometers can be any of an array of different analyzers such as a quadrupole (Q), ion-trap, time-of-flight (ToF), Orbitrap™, or Fourier-transform ion cyclotron (FTICR) as well as used in tandem configurations, such as a triple quadrupole (QqQ), to increase sensitivity, resolution, and accuracy [100]. The mass spectrometer measures ions according to their mass-to-charge ratio (m/z) and so first measures the whole peptides and, for tandem mass spectrometry, fragment the peptides to generate further information about the sequence of the peptide and generate further actionable information about the peptide.

An initial investigation proceeds without prior knowledge of present proteotypic peptides representing the allergenic food to identify suitable candidates. Identification of peptides is performed bioinformatically with software, such as Mascot or PEAKS, and compared with databases of proteins representing the targeted allergenic foods and background matrices [100]. Candidate peptides are screened on a wide array of criteria not limited to uniqueness against the food matrix, conservation among varieties of the allergenic food, high abundance, lack of modifications (i.e. Maillard reaction products, oxidation sites, deamidation sites), efficient extraction and digestion, reproducible and unique fragments, a predominant charge state, and of length between 6-12 amino acids [105]. Optimal peptides from the allergenic food may be appropriate in some food matrices but not others and therefore sub-optimal peptides may have to be used. Further, it is a continuous challenge that the databases assessed for peptide identifications may be incomplete for one or both of the allergenic food and matrix such that better peptides could be present and not utilized. Once suitable candidate peptides are identified and initial verification of their uniqueness and utility are established, chemically pure peptides labelled with heavy isotopes (commonly C^{13} and N^{15}) can be included with the unlabeled target peptides to act as standards for quantification [101].

Mass spectrometry enables the detection of multiple proteins per allergenic food and allergenic foods in a single analysis in addition to highly degraded proteins [101]. Mass spectrometry detects proteins from the allergenic food without the need for immunochemistry, which can be affected by food processing [100]; however, the effects of food processing must be evaluated with respect to mass spectrometry detection as protein extraction can be greatly influenced by processing. Taking a sample through the

entire process of mass spectrometry, data analysis, and quantification requires lengthy periods of time, expensive equipment, and highly trained personnel. The use of mass spectrometry is currently very well suited to confirmatory analysis, but as a young technology can grow to distinguish itself.

1.5 Risk analysis and food allergens

Risk can be defined as a function incorporating both the probability and severity of adverse health effects caused by a hazard, where a hazard is an agent that can cause adverse health effects [107]. With respect to food allergens, the hazard is the food allergen, and the risk is a function of an individual with allergy to that food allergen consuming the allergen as well as the dose consumed. Risk analysis is a three-part scientific process incorporating risk assessment, risk management, and risk communication (Figure 1.3). By virtue of new developments or considerations, risk analysis iteratively incorporates information, concerns of stakeholders, and feasibility to gauge the level of risk posed relative to the level of risk permissible. Application of risk analysis to food allergens can inform manufacturers' use of non-mandatory precautionary labeling.

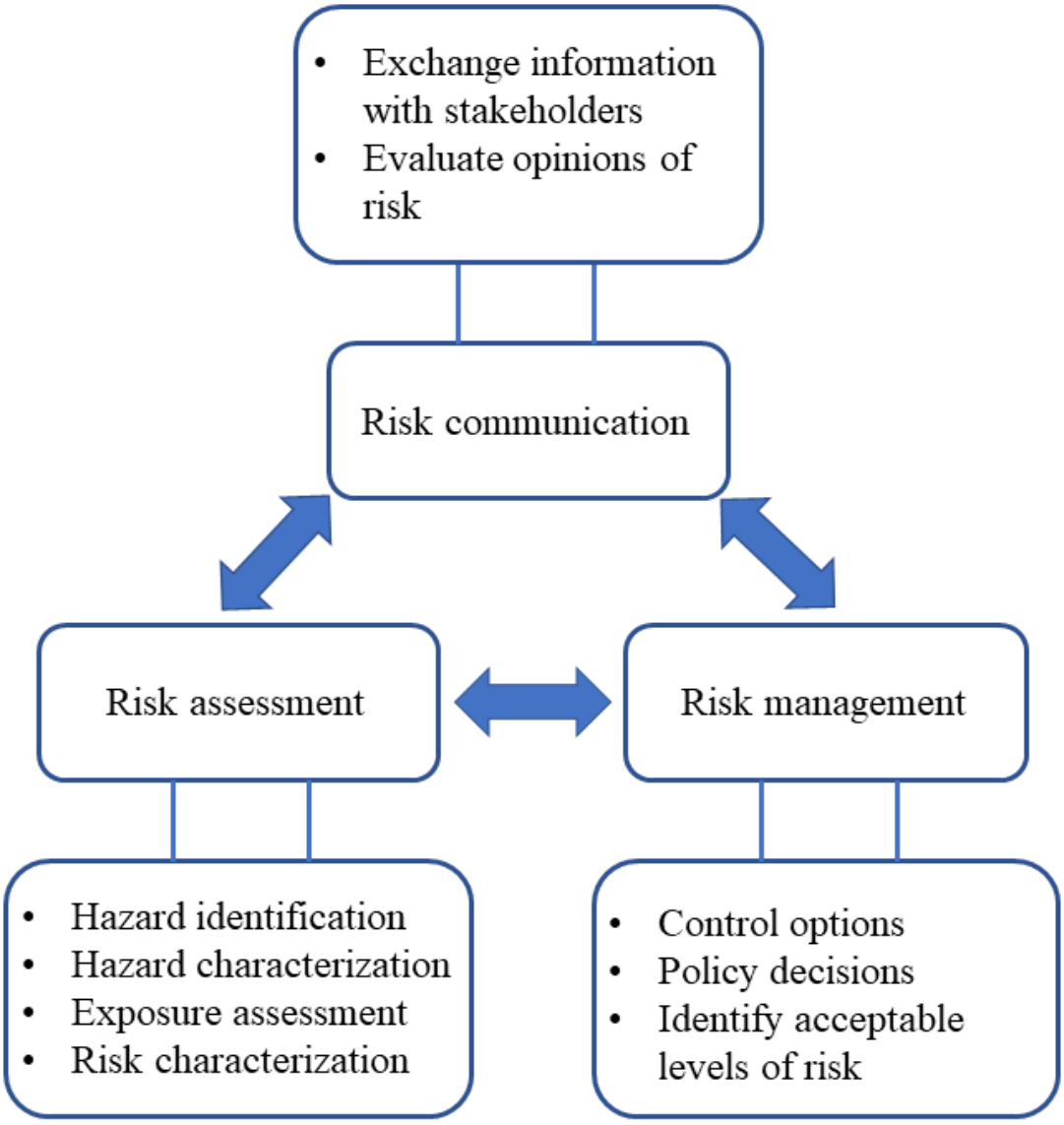


Figure 1.3 The constituent parts and sub-parts of an iteratively applied risk analysis.

1.5.1 Risk assessment

Risk assessment incorporates four steps consisting of hazard identification, hazard characterization, exposure assessment, and risk characterization [107]. For food allergens, these steps can be summarized as identifying plausible allergens of concern,

evaluation of the effects to allergic consumers, determination of likely consumption patterns, and culminating in an estimation of the risk posed to allergic consumers.

1.5.1.1 Hazard identification

Hazard identification is the identification of agents capable of causing adverse health effects [107], where for the hazard identification of a food allergen would be the range of plausible food allergens that may be present in a given food. Identification of relevant allergenic sources can be performed relative to data attributed to human consumption, including clinical studies, provided that the allergenic source was attributed correctly and judgment and/or guidance that the allergen is of significant concern to the overall population [108]. For example, strawberries can cause severe reactions in those with strawberry allergy; however, in the U.S. strawberries do not warrant delineation on package labels as an allergen nor handling as an allergenic hazard. Such ingredients may yet be judged as a hazard for a manufacturing plant or company if the packaged food were to be sold specifically to those with food allergies.

1.5.1.2 Hazard characterization

Hazard characterization is the qualitative and/or quantitative evaluation of the nature of the adverse health effects caused by agents [107]. If data is available, a dose-response assessment should be included to determine the relationship between the magnitude of exposure and the severity or frequency of adverse health effects. Non-allergen food safety (conventional) risk assessments may estimate the likelihood of the presence of a hazard on a scale of 1-3 ranging from unlikely to very likely and similarly

estimate the severity of the hazard from 1-3 ranging from minor injury to death and further multiplying these to generate an overall risk score per hazard and thereby determine a hierarchy of hazards to control [109]. Transposing such a scheme to allergens does not adequately characterize present hazards as the severity of an allergen can be further evaluated in terms of the physical nature of the allergen (liquid, powder, granules), amount of allergenic protein present (concentration), processing applied to the allergen, and clinical thresholds among allergic individuals [109]. As the severity of allergic reactions are related to the exposure dose, consideration and discrimination among allergen-containing ingredients provides a greater level of nuance. For example, a soy protein isolate as compared to soy lecithin contains a far greater amount of protein (allergens) as well as being a free-flowing powder rather than a viscous liquid where without characterizing the nuances between the two would be regarded as equally hazardous.

At the population level, allergic individuals begin to demonstrate symptoms at a wide range of doses [110]. Individual minimum eliciting doses can be determined using DBPCFC studies to determine the highest dose observed not to produce an adverse effect (No Observed Adverse Effect Level (NOAEL)) and/or the lowest does that does produce an adverse effect (Lowest Observed Adverse Effect Level (LOAEL)). Aggregation of individual minimum eliciting doses by modelling can be used to determine eliciting doses (ED) predicted to result in reactions in percentages of the population (i.e. 1% or 5% is then ED01 and ED05, respectively). This approach has been applied in the Voluntary Incidental Trace Allergen Labelling (VITAL) scheme for 14 allergens for manufacturers

to use as reference doses for manufacturers and guide risk assessment and management [7].

1.5.1.3 Exposure assessment

Exposure assessment is the qualitative and/or quantitative evaluation of the likely intake of agents and other relevant sources [107]. Exposure can be quantitatively determined by incorporating probabilities that allergic consumers purchase a product and the amount of the food consumed in a single sitting. It has been demonstrated that food allergic consumers do not differ in food intake as compared to non-allergic counterparts [111], and so general data on food intake can be used to evaluate likely consumption and exposure patterns allergic individuals will face. Country-wide dietary surveys can be used to determine values for both purchasing and consumption. Examples include the U.S. National Health and Nutrition Examination Survey (NHANES) and the Canadian Health Measures Survey (CHMS). Exposure assessments can incorporate further variables such as the propensity to heed specific wordings of precautionary allergen labeling as well as proprietary marketing data to narrow the assessment to specific products or a single company.

1.5.1.4 Risk characterization

Risk characterization is the qualitative and/or quantitative estimation including uncertainties of the probability of occurrence and severity of potential adverse health effects in a given population per hazard identification, hazard characterization, and exposure assessment [107]. In order of increasing complexity and requirements for data

quality, methods to characterize risk include safety assessments, benchmark dose [112], and probabilistic modelling [113]. Evaluation in order of increasing complexity clarifies if simpler methodologies are sufficient prior to application of more rigorous methods.

Briefly, a safety assessment is a classical toxicological approach where available data is surveyed for the lowest NOAEL and/or LOAEL and further divided by an uncertainty factor reflecting human variability to arrive at a dose that should be safe to consume in a single sitting [108]. The benchmark dose method involves constructing a dose-distribution curve from population data and determining a benchmark dose from the lower confidence limit of a dose corresponding to a predetermined increase in reactions (e.g. 1% or 10% or more) [108, 112]. Probabilistic modelling leverages distributions of minimum eliciting doses, concentration of allergen present, and consumption amounts to sample from using Monte Carlo simulations and calculate the likely number of expected allergic reactions [108, 113]. Both safety assessment and benchmark dose methods produce qualitative estimates of risk whereas probabilistic models are quantitative. Each tend to produce estimates that are risk-adverse such as probabilistic models overestimating the total number of reactions [108]. As the methodologies continue to improve coupled with higher quality data, more accurate estimates of the true risk can be produced.

1.5.2 Risk management

Risk management is weighing policies with respect to stakeholders and considerations of risk assessment to select adequate prevention and control options [107]. In context of the risk assessment for a particular hazard, risk management seeks to

actively control hazards considering all present hazards and heeding good manufacturing practices. Managing allergens seeks to make food products such that allergic consumers can make informed and safe choice [108]. Managing risk involves the entire supply chain to the consumer. Examples of steps to control include the sourced ingredients, where allergenic ingredients are stored in a facility, how manufacturing areas are cleaned and validated, and ensuring packages are accurate and reflect the food therein.

1.5.3 Risk communication

Risk communication is the interactive exchange of information and opinions throughout risk analysis including the explanation of risk assessment findings and the basis of risk management decisions [107]. Much of the communication directly between the manufacturer and consumer regarding the risk involved with foods is provided in the form of labeling. Major allergenic sources are commonly mandated to be clearly displayed such that concerned individuals can make safe choices. In the case of unintentional allergens, precautionary labeling is voluntary but ought to be used to reflect uncertainty where hazards cannot be controlled such that a preponderance of precautionary labels are not used for foods that such a warning is not warranted [108]. Ensuring that allergic consumers are well-informed is important but precautionary labeling should not be used in lieu of appropriate risk management as the result is undue restrictions on the diets on those with allergy.

1.6 Allergic risks of consuming insects

1.6.1 Background

There is increased interest and activity working toward production and use of insects as food sources in industrialized countries. Certain insects (locusts, termites, and other related arthropods such as scorpions) have been consumed for centuries in Southeast Asia and elsewhere, usually as whole identifiable organisms. New food products being developed now require processing of the insects into powders or emulsions that are added to processed foods. Thus, regulatory authorities and developers are interested in considering potential safety issues. The primary predictable risk is potential IgE allergic reactions for those already allergic to crustacean shellfish (also arthropods) or to other insects, who might react upon their first ingestion. This review is intended to consider the prevalence and identity of proteins that might cause cross-reactivity or *de novo* sensitization, but currently there is a lack of information on methods for accurately predicting *de novo* sensitization.

1.6.2 Crickets

1.6.2.1 Food allergy to crickets

Allergic reactions to crickets can occur from either airway, contact or oral exposure as discussed in a recent review [114]. Currently, there is little data regarding food allergy to crickets, with only one reference to the possibility of reactions to cricket as a food source [115]. There is little documentation of the use of insects as food sources in the United States. However, it is not uncommon to find insects used as food in some international cultures. Yet food allergy to crickets has been regarded as relatively rare in

light of notable consumption of crickets around the globe [114]. Allergy to crickets, locusts, and grasshoppers are often conflated, especially in light of the many species that are classified in each of these categories and that healthcare professionals may have difficulty in distinguishing them [114]. Published reports suggest that allergy to locusts or grasshoppers is more severe and common than cricket allergy [114, 116, 117]. Consumption of edible insects has resulted in severe reactions in Thailand and China. A Thai university hospital reported 24 cases of food-induced anaphylaxis over 6 years of which a single instance was to unspecified fried insects [118]. Another study from a tertiary care hospital found 36 food-induced anaphylaxis cases and 7 of those to fried insects, predominantly grasshoppers and crickets, over 2 years [115]. A review of Chinese allergy literature found that edible insects comprised the third leading cause of food-borne anaphylaxis in China with 7.5%, 7.5%, and 1.4% of all cases attributed to locusts, grasshoppers, and silkworm pupa respectively [119]. It should be noted that anaphylaxis is defined differently by some clinicians and some studies do not define their scoring system clearly. A common description is observable clinical signs involving at least two organ systems, such as respiratory tract (asthma or rhinitis), skin (atopic dermatitis or hives) or gastrointestinal involvement (pain, emesis, or diarrhea).

1.6.2.2 Non-food allergy to crickets

In adults, there are many reports that food allergic reactions can stem from cross-reactivity with airway allergens [120, 121], where exposure to airway allergens can sensitize individuals potentially leading to food allergy [122]. Inhalant allergies are more associated with symptoms including asthma, rhinitis, and atopic dermatitis whereas food

allergies may also include cutaneous symptoms and anaphylaxis [123]. Environmental exposure can occur outside homes as in the case of field allergy which results from exposure to swarms of insects or indoors such as exposure to cockroach matter such as hair, scales, feces, and body parts that result in generalized dust [123, 124]. Individuals with field allergy to grasshoppers and locusts have reported asthma and fatalities during large outbreaks of insects can occur [124].

Immediate hypersensitivity to cricket has been detailed only once in the literature where a teenager was using crickets as bait while fishing and exhibited lacrimation and angioedema within 30 minutes and wheeze within 5 hours [114, 125]. Chronic environmental exposure to *Acheta domesticus* in a cricket breeder resulted in a case of hypersensitivity pneumonitis with arginine kinase as the implicated allergen [126]. A case of occupational asthma with rhinoconjunctivitis due to occupational allergy to *Acheta campestris* where monosensitization to cricket was demonstrated and several IgE reactive proteins were described but not identified or characterized [127].

Cases of occupational allergy with grasshoppers and locusts have been reported in the scientific literature as early as the 1950s [124]. A case study by Rauschenberg *et al*. details a zookeeper with occupational allergy to locust (*Locusta migratoria*) with contact urticaria occurring by direct cutaneous exposure to whole dead or live locusts [128]. The patient was found to have specific IgE to locust without specific IgE to tropomyosin nor other arthropods. Lopata *et al.* investigated ten subjects exposed to locusts (*L. migratoria*) in a laboratory [129]. Six experienced symptoms of urticaria, rhinoconjunctivitis, and asthma, seven had positive skin prick tests (SPT), and five had specific IgE to *L. migratoria*. The authors suggested that a 35 kDa band identified from

L. migratoria by immunoblot was likely tropomyosin according to similar weight of reactive bands identified in cockroaches (*Blattella germanica* and *Periplaneta americana*). Further, a 70 kDa immunoreactive protein was identified in locust wings. Tee *et al*. investigated a research center breeding locusts (*L. migratoria* and *Schistrocera gergaria*) and identified multiple IgE reactive bands of between 18 to 68 kDa by western blot in extracts [130]. The source of the allergens was traced to the peritrophic membrane in the gut where the allergens were aerosolized in the excreted feces.

Cross-reactivity is not complete across members of Orthoptera including grasshoppers, locusts and crickets [114], but crickets have been suggested to be broadly cross-reactive amongst crickets [131]. An early study of allergic asthmatic children found that a significant proportion of children had IgE that bound to moth, cricket, grasshopper, and housefly extracts as determined by radioallergosorbent tests (RAST) used to measure specific IgE. Most were found to have cross-reactivity across the insects [132]. Cross-reactivity of crickets and crustaceans has been explored as in a case study by Lineres *et al*. which reported reactions upon exposure to crickets (*Gryllus campestris, Gryllus bimaculatus*, and *A. domesticus*) in a farmworker that was related to occupational rhinitis and asthma without cross-reactivity to prawns (*Penaeus sp*) or specific IgE to tropomyosins [131]. However, Srinroch *et al.* tested IgE binding to field cricket (*G. bimaculatus*) and giant freshwater prawns (*Macrobrachium rosenbergii*) with a pool of 16 prawn-allergic subjects' sera and found IgE reactivity to both cricket and prawn arginine kinases [133]. Additionally, hexamerin 1B was identified as a novel and minor allergen of *G. bimaculatus*. Phiriyangkul *et al*. used shrimp allergic sera to probe the effects of thermally processing Bombay locust (*Patanga succincta*) and identified

hexamerin, enolase, and arginine kinase as allergens in raw locust and hexamerin, pyruvate kinase, enolase, and glyceraldehyde-3-phosphate dehydrogenase as allergens in fried locust [134]. These proteins are highly conserved in function and to a great degree in amino acid sequence and structure across broad evolutionary taxa.

1.6.3 Mealworm

Mealworm (*Tenebrio molitor*) is an upcoming edible insect and currently has no listed allergens by the WHO/IUIS, however studies have begun to illustrate the specific associated risks. Verhoeckx *et al.* assayed mealworm protein fractions using sera with inhalation or food allergy to house dust mites or crustacea [135]. IgE from some subjects was found to cross-react with tropomyosin and arginine kinase and these proteins were found to be moderately stable by pepsin stability assay. Broekman *et al.* investigated thermal processing of mealworms and tested with shrimp allergic patients by SPT, Basophil activation test (BAT), and immunoblot [136]. The authors found that arginine kinase decreased in solubility upon heating whereas tropomyosin became soluble after heating, but there were no significant differences noted in allergenicity of heated compared to unheated samples despite changes in protein solubility. Further work by Broekman *et al*. utilized double blinded placebo controlled food challenges (DBPCFC) with 15 shrimp allergic patients using mealworm [137]. Confirmed oral mealworm allergy was found in 13 patients with individual protein doses resulting in objective symptoms ranging from 216 mg to 13 g with symptoms including oral allergy symptoms, urticaria, nausea, abdominal cramps, and dyspnea. Four of the patients were also challenged with shrimp and found to have similar eliciting does and severity even

compared with dose threshold data for shrimp [137, 138]. The authors concluded that mealworm allergy was not only likely in shrimp-allergic patients but also with severe outcomes despite acknowledging the limited power of the study [137]. Exploring potential cross-reactivity between primary airway and oral allergy to mealworm, Broekman *et al.* used C3H/HeOuJ mice and 4 mealworm breeders, half of which reported inhalant allergy and the other half reported food allergy to mealworm [139]. All four were found to be sensitized to mealworm by ImmunoCAP, SPT, and BAT but were all negative to oral challenge to shrimp. Only those with reported food allergy to mealworm tested positive to mealworm by DBPCFC. Immunoblots using sensitized mice and humans showed binding to tropomyosin, arginine kinase, and both myosin light and heavy chain. Further, larval cuticle proteins were identified and suggested as a major source of primary mealworm allergy, where larval cuticle proteins have been identified to bind to chitin [139, 140].

1.6.4 Allergens of edible insects and related species

A review by Barre *et al.* identified 13 likely cross-reactive allergens of edible insects: alpha-amylase, arginine kinase, chitinase, glutathione-S-transferase, triose phosphate isomerase, trypsin, hemocyanin, hexamerin, actin, sarcoplasmic calcium binding protein, myosin, tropomyosin, and troponin [141]. The list by Barre *et al.* was formed according to their ability to cross-react IgE of patient sera allergic to shrimp and dust mites. Arthropods with well-characterized allergens can be used as a measure to assess the most likely and most dangerous risks associated with edible insects. Those

related to crickets, in order of increasing taxonomic distance, include cockroaches (order Blattodea), mealworms (order Tenebrionidae), and crustaceans (subphylum Crustacea).

1.6.4.1 Cockroaches

Of Blattodea, *P. americana* (American cockroach) and *B. germanica* (German cockroach) are the primary focus of allergy research [142], however *Blatta orientalis* (Oriental cockroach) and *Supella longipalpa* (brown-banded cockroach) are also important environmental pests [143]. Cockroach is an important indoor allergen associated with severe asthma particularly in urban inner-cities where 40-60% of asthmatics have IgE to cockroach antigens [144, 145]. Cockroach allergens from feces, saliva, or bodily debris can persist in homes post-infestation and increase the risk of developing asthma as well as asthma morbidity [146-148]. Wang *et al*. investigated the correlations between IgE-medicated sensitization to shrimp, cockroach, and dust mite related to exposure in inner-city children [149]. It was found that across 504 subject sera a strong positive correlation was found between shrimp and cockroach IgE levels and higher exposure to cockroach was significantly correlated with higher shrimp and cockroach IgE.

Cockroach allergens have been well reviewed elsewhere [142, 150, 151] and a brief summary is provided here. Cockroaches currently have twelve groups of allergens listed by the by the World Health Organization/International Union of Immunological Societies (WHO/IUIS online at www.allergen.org). Group 1 allergens (*B. germanica* 1, Bla g 1; *P. americana 1*, Per a 1) are midgut and fecal proteins with a repeating tandem structure and unique fold and a hydrophobic core understood to bind phospholipids. It is

found as intertwined polymers resulting in varying molecular weights and has been shown to have cross-reactivity with fruit flies as well as mosquitos. Group 2 allergens (Bla g 2, Per a 2) are inactive aspartic proteases with a highly stable structure of 5 disulfide bonds as well as zinc binding domain and has also been found to cross-react with mosquito and fungal allergens. Group 3 allergens (Bla g 3, Per a 3) are hemocyanins, which are hexameric proteins similar to arylphorins and hexamerins and are known to cross-react between arthropods. Group 4 allergens (Bla g 4) are lipocalins and are calcium-binding proteins excreted or secreted and exclusively found in adult male reproductive tracts, but as a class these are important inhalant allergens from dog, cats, horses, and cows. Group 5 allergens (Bla g 5) are glutathione-S-transferases (GST) that have been shown to have cross-reactivity with helminth but not dust mite due to low surface similarity. Further, principal GSTs produced by *B. germanica* and *P. americana* belong to separate subclasses of GST with low sequence identity and with low cross-reactivity, but each cockroach does have homologs of each with high sequence identity. Groups 6 (Bla g 6, Per a 6) and 8 (Bla g 8) are troponin C and myosin light chain, which are EF-hand proteins associated with muscle contraction. Group 7 (Bla g 7, Per a 7) is tropomyosin, which has been regarded as pan-allergen after shrimp tropomyosin was found to have IgE cross-reactivity with crustaceans, mollusks, dust mites, and insects. Group 9 (Bla g 9, Per a 9) are arginine kinases, which was identified to cross-react between Indian meal moths and tiger shrimp, but also between shrimp and arthropods more broadly. Group 10 (Per a 10) are serine proteases identified to be a major allergen in India. Group 11 (Bla g 11, Per a 11) are alpha-amylases that were initially identified in fecal extracts. With potential cross-reactivity to mite homologs, it has been suggested as a

significant allergen in Korea. Group 12 (Per a 12) are chitinases with some identity to mite homologs and more recently identified as a major allergen in China. Lastly, chitin is known to be proinflammatory (promoting overall inflammation and Th1 responses) as well as being able to induce Th2 cytokine responses in some cases [151].

Major allergens of cockroach are regarded as groups 1-5 while groups 1, 7, and 9 are regarded as cross-reactive, but the nature of the observed cross-reactivity among cockroaches and other arthropods is not fully understood [150]. Groups 1, 2, and 10 are released into the environment through feces, whereas the remaining are proteins found within the insect bodies that can be released into the environment upon death [142]. In a review focusing on the first ten groups of allergens in Taiwan, the most abundant *P. americana* allergens in homes were groups 9, 10, and 2 [147]. Groups 6-8 are regarded as minor allergens of cockroaches due to relatively low IgE sensitization [148], but the high rate of sensitization to cockroaches makes relatively minor allergens potentially significant. However, the prevalence of specific IgE to each of the allergens is known to vary depending on regional exposure [150].

1.6.4.2 Crustaceans

Crustacean shellfish IgE cross-reactivity with insects has been well documented [152-154]. Shellfish allergy is one of the most common food allergies with self-reported rates of 1.9% in the United States [155], but also tends to persist through life [156]. Cross-reactivity is possible across airway and food allergens as in the IgE cross-reactivity across house dust mites, shrimp, and cockroaches [153, 157]. Historically the source of the cross-reactivity has been focused on tropomyosin. However other allergens have

more recently been implicated in the cross-reactivity between insects and shellfish [152, 154]. The most relevant crustacean allergens are tropomyosin (*Litopenaeus vannamei* 1, Lit v 1), Arginine kinase (Lit v 2), myosin light chain (Lit v 3), sarcoplasmic calcium binding protein (Lit v 4), and hemocyanin [158]. Of shrimp allergens, tropomyosin, sarcoplasmic calcium binding protein, and myosin light chain are suggested to be more specific to shrimp, whereas arginine kinase and hemocyanin are cross-reactive among arthropods [159]. Other allergens have been identified in crustaceans such as triose phosphate isomerase (*Crangon crangon* 8, Cra c 8) and troponin I (*Pontastacus leptodactylus* 7, Pon l 7). Consumption habits are important to consider for crustaceans compared to insects. In the United States only the tail-muscle is consumed of shrimp and lobster, while crab legs are the major tissues consumed. However, some people consume the digestive glands and abdominal parts. In other cultures it is more common to eat most of the soft-parts of these crustaceans. Insects such as crickets will be consumed whole. Some insects (silkworm and meal worm) are immature larvae, and the body composition can differ, which may alter the profile of proteins that consumers could become sensitized to.

1.6.5 Allergen classes of concern

1.6.5.1 Arginine kinase

With respect to the composition and identified allergens of crickets, as well as related species, the principal allergenic concerns are arginine kinase, tropomyosin, hexamerin, and chitin. Arginine kinase been shown to be a cross-reactive allergen across shrimp (*M. rosenbergii)* and crickets (*G. bimaculatus*) [133]. Arginine kinase from

Indianmeal moth (*Plodia interpunctella*) has been demonstrated to be a cross-reactive allergen across crustaceans and insects via serum IgE and SPT [160]. High levels of IgE prevalence have been shown against arginine kinases from cockroaches (*P. americana*, *B. germanica*) [142], and arginine kinase has been shown to be a major allergen among Thai cockroach allergic patients [161]. Further, sensitization to arginine kinase without shrimp allergy has been associated with allergy to house dust mite and/or cockroaches [159]. Cockroach (*P. americana*) arginine kinase is moderately thermostable and retains half of enzymatic activity after heating for 10 minutes at 50 °C [162]. Crayfish (*Procambarus clakrii*) arginine kinase shows low-level aggregation at above 44 °C and diminished IgE binding after boiling [163]. However, mealworm (*T. molitor*) arginine kinase was only found to decrease solubility when heated but not lose allergenicity despite lower solubility [136]. The IgE reactivity of shrimp-allergic sera to arginine kinases of other crustaceans has been shown to vary depending on denaturation [164]. Crayfish (*P. clarkii*) and mealworm (*T. molitor*) arginine kinases have been shown to be stable by pepsin stability assay [135, 163].

1.6.5.2 Tropomyosin

Tropomyosin has been identified as an inhalant allergen of cockroaches [165-167], a food allergen of crustaceans [168, 169], as well as an occupational allergen in crab processing workers [170, 171]. It has also been implicated as a cross-reactive allergen across crustaceans, dust mites, and insects [154, 172, 173]. Investigations have compared insects, dust mites, and crustaceans and shown tropomyosin has a very high sequence similarity and therefore may also act as a cross-sensitizing allergen [174].

Heating crustaceans and mollusks has been demonstrated to result in increased tropomyosin reactivity post-heating according to a monoclonal tropomyosin antibody [175]. However, heating in tandem with high pressure treatment was found to decrease allergenicity of shrimp (*L. vannamei*) tropomyosin via pooled shrimp-allergic sera and tropomyosin-sensitized BALB/c mice fed orally according to cytokines and IgE levels [176]. Purified tropomyosin from *Penaeus modon* and *Litopenaeus vannamei* have been shown to be resistant to pepsin digestion with *Litopenaeus vannamei* tropomyosin being more resistant than *Penaeus modon* [177], but crab tropomyosin (*S.paramamosain*) has been found to become more digestible after boiling and high pressure steam was found to increase digestibility as well as decrease IgE binding [178].

1.6.5.3 Hexamerin

Hexamerins are multimeric proteins and belong to the same protein family as hemocyanins and arylphorins, which have low sequence similarity but similar structure [179, 180]. Hemocyanin has been suggested to be a clinically relevant crustacean allergen due to the high prevalence of allergic individuals with IgE binding to the protein, but also because a significant proportion of those have histories of anaphylactic reactions [181]. Again, Srinroch *et al*. has identified hexamerin 1B from cricket (*G. bimaculatus*) as a novel and minor allergen [133]. Van Broekhoven *et al*. explored processing and *in vitro* digestion across mealworm species and found that IgE from crustacean or dust mite allergic patients were able to bind mealworm hexamerin 1B precursor without IgE cross-linking observed in heat processed fractions [182]. The observed hexamerin shared approximately 40% sequence identity with the cockroach arylphorin Per a 3. Giant

freshwater prawn (*M. rosenbergii)* hemocyanin was identified as heat-stable after it bound IgE of shrimp-allergic patients after thermal treatment but *Penaeus modon* was unable to inhibit *M. rosenbergii* reactivity using 13 shrimp-allergic sera [183]. However, Ayuso *et al*. found that hemocyanin may be cross-reactive amongst crustaceans and cockroach using 34 shrimp-allergic sera against recombinant hemocyanin and dot blot assays demonstrated inhibition of hemocyanin by cockroach extract [184].

1.6.5.4 Chitin

Chitin is a thermostable polysaccharide comprising insect exoskeletons, but also found in crustaceans [185-187]. Multiple studies have shown that chitin has immunomodulatory effects. De Silva *et al*. used TLR2-2/TLR-4 null, MyD88-null, and IL-17A null mice intraperitoneally injected with mixtures of chitin and ovalbumin or alum and ovalbumin with subsequent challenge to aerosolized ovalbumin [186]. The authors found that chitin induced adaptive immune responses and functioned as an adjuvant similar to alum. Dubey *et al*. used crab chitin and C57BL/6 mice sensitized by intraperitoneal injection of *Aspergillus fumigatus* culture filtrate or mixed with either alum or chitin [188]. Compared to alum, chitin resulted in lower inflammatory responses, depressed Th2 cytokines, enhanced Th1 cytokines, yet had similar total and specific IgE and IgG1 levels. Bae *et al*. used C3H/HeJ mice sensitized by intraperitoneal injection of ground whole peanut and challenged intragastrically where mice were given either α-chitin, β-chitin, or β-chitosan in their food [189]. It was found that oral administration of chitin and chitosan is protective against IgE-mediated anaphylaxis and inhibited Th2

responses. Chitin has the potential to function as an adjuvant similar to alum and can therefore contribute to *de novo* sensitization.

Oral, cutaneous, or inhalant exposure to chitin can result in sensitization to proteins bound to the chitin or other protein antigens present [190]. Therefore, the chitin content of insects could impact the potential allergenicity risk. By dry basis, crickets (*Gryllus testaceus*) have been found to comprise of 8.7% chitin [191], three Blattodea species (*Blaptica dubia*, *Blaberus discoidalis* and *Blatta lateralis*) had between 5.3 and 8.3% chitin [192], and seven grasshoppers of Orthoptera had chitin ranging from 5.3 to 8.9% [193]. The chitin content of these insects is similar and therefore risks associated with chitin from cockroaches may be comparable to Orthoptera and crickets.

1.6.6 Allergen Carryover from Feed

Insects are commonly consumed whole, as a powder or paste, or as extracts; each of which include gastrointestinal content of the insect [194]. Insects are commonly fasted prior to freezing or may be fed a nutrient-rich diet prior to freezing, referred to as gut loading. Fasting assists in mitigating the risks posed by allergens, otherwise the content of the insect's gastrointestinal tract could expose consumers to allergens from the feed [195, 196]. For example, crickets raised on insect farms are often fed with high-quality feed such as chicken feed or pet food [194, 197]. Chicken feed may contain a number of allergenic ingredients including wheat, milk, soy, peanut, or fish in forms such as wheat bran, dry whey, oilseed cake, and fishmeal [198]. A limited amount of research has been done regarding identifying consumed proteins in insects including tomato defense proteins in tobacco hornworm (*Manduca sexta*) and cabbage looper (*Trichoplusia ni*)

[199] and wheat proteins in sunn pest (*Eurygaster integriceps*) [200]. Allergic reactions to residual proteins from feed have yet to be attributed in the literature; however, the insects' diet may result in non-insect allergenic material passing to consumers to pose a risk to allergenic consumers.

1.6.7 Summary

IgE binding and IgE cross-reactivity is not a measure of allergenicity. However, it is clear that IgE binding to specific proteins is required to trigger immediate reactivity by eliciting mediator release from mast cells and basophils. There are many studies that demonstrate some IgE binding or low-levels of cross-reactivity that do not parallel allergic reactions. The studies reviewed here rarely used food challenge, much less double-blinded studies, to demonstrate clinical importance of the IgE binding. Few studies used titrated inhibition to evaluate differences in affinity and epitope recognition, however, this data is used by most for judging cross-reactivity. Some biological assays such as skin prick test also over- or under-predict allergic reactivity. This review demonstrates that there is shared IgE binding between a number of proteins that have been reported to be important IgE binders, or in a few cases clearly proven allergens. The level of risk for allergic consumers will vary by their specific sensitivities. Additional steps are being taken to evaluate the likelihood of cross-reactivity between various crustaceans and between crustaceans and insects. It is important to note that the sequence identity of two proteins and overall structure influence likely cross-reactivity as does the abundance of the protein in the source material that humans are exposed to. Some proteins also denature and in foods that are highly processed or cooked, some proteins

will aggregate and be difficult to extract and test. Others may unfold and if the IgE binding epitopes are conformational, may be markedly reduced or lost following cooking. It is important to consider those factors in judging potential cross-reactivity.

1.7 Peanut food allergy and the effects of thermal processing

1.7.1 Background

Peanuts (*Arachis hypogaea)* are a non-novel legume of the family Fabaceae shared with other legumes such as soybeans, green peas, and chickpeas. Peanuts are cultivated for a wide variety of products derived from the plants such as the peanuts themselves as well as skins, hulls, oil, and press cake. As a food, peanuts are a relatively protein rich food at approximately 25% protein per weight [201]. Four varieties are peanuts are commonly produced in the US including Virginia, Spanish, Valencia, and Runner [201]. The protein content, protein profiles, and IgE-binding profiles are largely comparable among peanut varieties in terms of the major allergens with some variation among minor allergens. The majority of U.S. peanut production is focused on either Runner peanuts, commonly used for peanut butter, or Virginia, which is often a snack peanut or for in-shell products. Peanuts are commonly processed to improve their flavor and aroma such as by boiling, frying, or roasting. This review is intended to consider the effects of thermal processing on peanut food allergy and allergens to explore how peanuts may be extensively thermally processed into a novel source of food allergens.

1.7.2 Peanut allergy

Peanuts are a known food allergen with a prevalence of 1.8% in U.S. adults and 2.2% of U.S. children and adolescents with increasing incidence [81, 202, 203]. In the U.S., peanuts are a member of the eight classes of allergenic foods comprising 90% of U.S. food allergies and use of peanuts must be signified on package labels [204]. Food allergens are considered as a major allergen if recognized by IgE of >50% of the allergic population, then minor allergens are recognized by <50% of the allergic population [1]. In peanut, the major allergens are Ara h 1, 2, 3 and 6 while the remainder are minor allergens. On a weight basis, Ara h 3 is the most abundant allergen of peanuts followed by Ara h 1, and then Ara h 2 and 6 [201]. Currently, a total of 16 peanut allergens have been characterized (Table 1.3) [1, 2].

Table 1.3 Major and minor allergens of peanut [1, 2]

Peanut allergens	Biochemical name
Major allergens	
Ara h 1	Cupin
Ara h 2	Prolamin
Ara h 3	Cupin
Ara h 6	Prolamin
Minor allergens	
Ara h 5	Profilin
Ara h 7	Prolamin
Ara h 8	Pathogenesis related protein family 10 protein (PR-10)
Ara h 9	Non-specific lipid-transfer protein (nsLTP)
Ara h 10	Oleosin
Ara h 11	Oleosin
Ara h 12	Defensin
Ara h 13	Defensin
Ara h 14	Oleosin
Ara h 15	Oleosin
Ara h 16	nsLTP
Ara h 17	nsLTP

A. hypogaea is an allotetraploid resulting from the genomic merger of the ancestral species *A. duranensis* and *A. ipaënsis* resulting in both sets of chromosomes being part of *A. hypogaea* [205]. Efforts have since mapped the genomes of the ancestral species [206] and *A. hypogaea* [207].Progress in annotating the genome of *A. hypogaea* has expanded the diversity of proteins known to be present, particularly novel isoforms of known allergens but these have yet to be evaluated as allergens [208, 209].

1.7.2.1 Classification and superfamilies of seed storage proteins

Each of the major allergens of peanut are seed storage proteins, whose principal function is as a storage reservoir for amino acids but may have other functions [210].

Seed storage proteins are classified based on their sedimentation coefficient and conditions of solubility. Solubility classifications include the water-soluble albumins, dilute saline- soluble globulins, dilute alcohol-soluble prolamins, and dilute acid or alkali-soluble glutelins. In peanut, the major storage proteins are albumins and globulins [1].

2S albumins are a major group of seed storage proteins that are synthesized as a single polypeptide and subject to proteolytic cleavage commonly at both the N- and C-termini and trimming of linker peptides [210]. 2S albumins are compact globular proteins with a conserved cysteine structure and disulfide bonds. The stability offered by the compact and tightly bound structure offers 2S albumins a high degree of resistance to gastrointestinal conditions; however, the flexible loops are solvent-exposed and predominate known IgE epitopes [211]. Coinciding allergies to peanuts and various tree nuts has been attributed in part to the homology and similar physiochemical properties shared among 2S albumins [212].

The other major group of seed storage proteins in peanuts are the globulins as either the 7S vicilin-type or 11S legumin-type [210]. In legumes, 11S legumins are the major storge protein of legumes and consist of six pairs that interact noncovalently. Each pair consists of a single polypeptide that is proteolytically cleaved after disulfide bond formation to produce an acidic and basic subunit. 7S vicilins are often trimeric proteins lacking significant cysteine residues and therefore disulfide bonds. The family of globulin proteins have dissimilar sequences but share physiochemical properties and structures where 7S globulins are trimers (3α) and 11S globulins are either trimers (3α) or dimer-trimers (3α, 3β).

1.7.2.2 Ara h 1 and Ara h 3

Ara h 1 and Ara h 3 are bicupin seed storage proteins belonging to the cupin superfamily [2]. Ara h 1 is a 7S vicilin-type cupin and Ara h 3 is an 11S legumin-type cupin. Ara h 1 has a β-barrel core domain and secondary domain predominantly made from α-helices [213], which is conserved among other legume 7S vicilins such as lentil, pea, and soy and has been leveraged to explain cross-reactive allergy among edible legumes. While Ara h 1 can be observed naturally as the trimer, it can also be found in oligomeric forms that are more resistant to enzymatic digestion [214]. Epitope mapping studies have found that the IgE binding sites of Ara h 1 include the pepsin-resistant N-terminus [214], on connecting loop domains and coils in the core region [215], as well as inside the β-barrel domain per monomer. Epitopes buried due to trimer formation have been speculated to protect these regions from enzymatic digestion and preserve their allergenicity [215, 216]

Ara h 3 is composed of multiple sub-units including an acidic and basic subunit which are proteolytically processed prior to maturation and held together by disulfide bonds [217]. Use of reducing conditions with Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) evidences a 42-45kDa acidic and 25 kDa basic subunit. Ara h 3 is known to have trypsin inhibitory functionality and so contributes to the enzymatic resistance of peanut [218]. It has been long understood that multiple Ara h 3 isoforms are present in peanut as per the depreciated Ara h 4, which has since been understood as an isoform of Ara h 3 [219].

1.7.2.3 Ara h 2 and Ara h 6

Ara h 2 and 6 belong to the 2S albumin seed storage proteins and the prolamin superfamily [2]. The structure of Ara h 2 consists of a five-helix bundle held together by four disulfide bonds and noted to have a structure similar to trypsin inhibitors [220], and has been confirmed as having trypsin inhibitory activity [221]. Ara h 2 contains multiple hydroxyprolination motifs, DPYSPS [222], which have been shown to have profound effects on IgE binding. Tscheppe *et al.* produced mutant Ara h 2 in insect cells to exclude flexible loop regions, including hydroxyprolination motifs, and further unfolded through reduction and alkylation demonstrate substantially decreased IgE binding and anaphylactic potency [223]. Notably, comparison of the mutant to wildtype was marred by lack of attention to the effects of hydroxyprolination, which may have made the decreases more drastic.

Ara h 6 has a nearly identical structure to Ara h 2 in terms of both fold and disulfide bonds [224], but does not contain the hydroxyprolination motifs of Ara h 2 [222]. Multiple forms of Ara h 6 have been characterized as both an intact polypeptide and one posttranslationally cleaved in the core [225, 226]. The two forms of Ara h 6 do not show differences in IgE binding or abundance [225]; however, the intact form is more susceptible to thermal denaturation than the cleaved form [226].

1.7.2.4 Relationships between major allergens

Ara h 1, Ara h 2, and Ara h 3 do not share significant sequence or structural similarities, but localized similarities have been implicated in IgE cross-reactivity among these allergens [227]. Ara h 2 and Ara h 6 has been observed as cross-reactive due to

shared homology [228]. Both Ara h 2 and Ara h 6 are resistant to pepsin digestion whereas Ara h 1 and Ara h 3 are readily digested by pepsin [229]. Despite a higher seroprevalence of Ara h 6 compared to Ara h 2, Ara h 6 has demonstrated weaker biological activity via human basophils than Ara h 2 [230]. This has been corroborated using rat basophil leukemia cells showing that Ara h 2 is a more potent allergen than Ara h 6, but both are far more potent than either Ara h 1 or Ara h 3 [231]. A majority of the allergenic activity attributable to crude peanut extracts have been attributed to Ara h 2, Ara h 6, and their variants [232].

Ara h 2 coupled with Ara h 6 have been cited as the most accurate in diagnosing peanut allergy as compared to other peanut allergens or specific peanut IgE [233, 234]. In children, both Ara h 2 and Ara h 6 are the most frequently recognized major peanut allergens and this pattern remains stable over time [235]. Further, co-sensitization to both Ara h 2 and Ara h 6 is associated with severe allergic reactions as opposed to mild reactions [236].

1.7.2.5 Minor peanut allergens

Profilins are one type of actin-binding proteins acting to regulate the organization and function of the actin cytoskeleton [2]. Ara h 5 is a member of the profilin family and related to other pollen allergens such as Hev b 8 (latex) and Bet v 2 (birch) [237]. The structure of Ara h 5 is a seven stranded antiparallel β-sheet with two α-helices flanking one side and another helix on the other side.

Ara h 7 is the third 2S albumin of peanut and has been found to be equally potent compared to Ara h 2 and Ara h 6, but is far less abundant [238]. Multiple isoforms of Ara

h 7 are known to exist but Ara h 7.0101 has not been demonstrated to exist as a protein [238, 239]. In those with sensitivity to any of the peanut 2S albumins, sensitivity to each of Ara h 2, Ara h 6, and Ara h 7 is most frequent.

Ara h 8 is a member of the family 10 of pathogenesis-related (PR-10) proteins and related to the cross-reactive allergen Bet v 1 (birch) [240]. Members of PR-10 are associated with sensitization to plant pollen and sensitization to food allergens [241]. Ara h 8 has a seven stranded antiparallel β-sheet with two short α-helices between the first two strands of the β-sheets with a third α-helix at the C-terminus [242]. Ara h 8 has similar structure and ligand binding to Bet v 1 including binding to various flavones including quercetin, apigenin, and daidzein, and resveratrol [2, 240]. Reactions to Ara h 8 tend to be mild and likely mediated through individuals with cross-reactivity to Bet v 1 and other homologues. Monosensitization to Ara h 8 has been associated with tolerance to peanuts rather than allergy [243].

Ara h 9 and 17 are regarded as type I non-specific lipid transfer proteins (nsLTP) and Ara h 16 is a type II nsLTP [2]. The types of nsLTP are differentiated by their molecular weight, primary sequence identity, and disulfide bond pattern where type I nsLTP are approximately 9 kDa and type II are approximately 7 kDa [244]. Functionally, nsLTPs can serve many purposes including membrane stabilization, cell wall organization, signal transduction, and assist in growth and development [2]. nsLTPs have been shown to be very stable to enzymatic digestion and thermal treatment due to a structure of four α-helices stabilized by four disulfide bonds [245]. Associations have been noted between peanut, peach and hazelnut allergies owing to homologies between

Ara h 9, Pru p 3, and Cor a 8, respectively [240], specifically those with peanut allergies for those in the Mediterranean area [245, 246].

Peanut oleosins (Ara h 10, Ara h 11, Ara h 14, and Ara h 15) function as structural proteins of oil bodies to stabilize the oil body [247]. Allergies to oleosins is relatively common but underestimated partially due to their underrepresentation in aqueous extracts owing to their lipophilic nature [248]. Oleosins show resistance to both temperature and enzymatic digestion. Peanut oleosins have been observed as multimers on SDS-PAGE [249]. Sensitization to oleosins is associated with severe reactions, although sensitization to peanut oleosins is highly associated with Ara h 2 sensitization [247].

Defensins, such as the peanut defensins Ara h 12 and Ara h 13, are a large group of small, cationic, and disulfide-rich proteins associated with the innate immune system across animals, plants, and fungi [250]. Peanut defensins have demonstrated inhibitory effects to some molds but without effects on bacteria [251]. Structurally both Ara h 12 and Ara h 13 are lipophilic proteins consisting of an alpha-beta fold and 4 disulfide bonded structure [250, 251].

1.7.3 Effects of thermal processing on peanut allergenicity

Thermal processing of food proteins can lead to several modifications including unfolding, aggregation, and chemical modifications [252]. Structural modifications can further affect digestion stability as well as the manner that the allergens are presented to the immune system. Through the Maillard reaction, chemical modifications of free amines with free reducing sugars, proteins can be chemically modified not limited to

cross-linking of proteins [1]. Through cross-linking, Ara h 1 and Ara h 2 can form higher molecular weight aggregates that bind IgE more effectively than unmodified forms [252]. The type of thermal processing affects resultant advanced glycation end (AGE) products as boiling generates approximately 10x fewer AGE than roasting. While peanut allergens with AGE modifications are more frequently recognized by peanut allergic patients, there is a level of bias given that consumers will typically consume peanuts that have been thermally processed.

Thermal processing affects the proteins and their allergenicity by how the processing was applied. Studies of thermally processed Ara h 1 and Ara h 3 are a challenge as they tend to become insoluble, although in roasted peanuts IgE binding shows minimal differences [252]. The trypsin inhibition activity of Ara h 2 increases after roasting [221]. Both Ara h 2 and Ara h 6 retain similar IgE reactivity after roasting and maintain their structure [253]. Ara h 8 increases IgE reactivity and resistance to enzymatic digestion with roasting, which may be associated with binding to lipophilic ligands [254]. Oleosins have also demonstrated a higher IgE binding capacity when roasted in-shell [255]. The type of thermal processing does affect the structure of Ara h 1 as boiling causes the formation of rod-like aggregates with reduced IgE binding capacity, whereas roasting causes the formation of globular aggregates with IgE reactivity similar to non-thermally processed [256]. Boiling peanuts can decrease the IgE binding capacity of Ara h 1, Ara h 2, and Ara h 3 compared to roasting, given that Ara h 2, Ara h 6, and Ara h 7 tend to leech out of the peanuts into the water [2].

Extraction conditions can also result in differences in assessing relative allergenicity. Assessments of Ara h 1 and Ara h 2 have shown that extraction procedures

incorporating defatting and the buffer of choice can affect the protein yield and sensitivity of subsequent immunoassays [257]. The extractability of peanut proteins is also interconnected with the type of processing applied [258]. Further, the extraction conditions itself can affect the protein structure and alter the evaluated allergenicity [259].

1.7.4 Summary

As a staple food for many people around the world peanuts are a nutritionally and economically valuable food, but for those with peanut allergies they are a significant hazard and source of distress. To adequately inform stakeholders of the potential risks associated with peanuts and their products, it is imperative to understand the breadth of information known about the allergenic proteins of peanut. Thermal processing has wide ranging effects on peanut proteins and assessment of the proteins can have similar effect when attempting to evaluate the potential risk they pose to consumers.

1.8 Allergic risks of exposure to food allergens by inhalation

1.8.1 Background

The principal route of exposure for food allergens is through consumption of foods that contain the allergen either as an intended ingredient or as a contaminant. Similarly, inhalant allergens are principally encountered through inhalation; however, these categories are not strict in that food allergens can be inhaled and result in elicitation of allergic reactions and visa-versa. Cases of pancake syndrome also known as oral mite anaphylaxis demonstrate that oral consumption of inhalation allergens can result in

allergic reactions [260], whereas inhalation of food allergens is most detailed in terms of occupational asthma due to handling or cooking of food and food allergens. This review is intended to introduce how allergic individuals may be exposed to food allergens through inhalation and result in novel sources of food allergens.

1.8.2 Defenses, barriers, and transport of allergens

1.8.2.1 Oral route

Consumption of food and exposure to food proteins begins in the mouth and has been identified as both a means to effectively absorb proteins via sublingual and buccal mucosa as well as a route for treatment of food allergies [261]. The oral cavity is protected by mucosal associated lymphoid tissues. After swallowing, food proteins are exposed to low pH via hydrochloric acid to assist protein denaturation as well as active pepsin to begin proteolysis of proteins [262]. Passing the bolus to the small intestine begins by partially neutralizing the pH and thereby allowing gastric proteases and peptidases such as trypsin, chymotrypsin, elastase, and carboxypeptidases to continue degradation the food proteins. The intestines include further defenses such as secretory IgA, mucus, and microbes that can influence the rates of absorption of intact proteins as well as provide extra time for proteolysis to occur [263]. Digestion of dietary proteins ends with intestinal brush border cells uptake amino acids as well as di- and tripeptides through various mechanisms [262].

Uptake or transport of partially digested or undigested proteins across the intestinal tract can vary depending on the characteristics of the protein including size, shape, and structure [263]. Transport routes for proteins include para-cellular and trans-

cellular transport, where para-cellular transport refers to the movement of compounds through the intercellular space between cells and is regulated by tight junctions, but the maintained integrity of the tight junctions in healthy adults leaves this route of lesser concern. Para-cellular transport is typically relegated to compounds that are far smaller than allergic fragments capable of causing reactions, but some evidence suggests that smaller intact proteins (insulin) may be able to pass [263, 264]. Allergic sensitization can reduce the integrity of the tight junctions via presence of mast cells and increase the permeability to intact proteins.

Several routes are possible for trans-cellular transport, but only endocytosis is viable for non-degraded proteins [263]. Enterocytes are the most abundant cells in the intestinal barrier and act as absorptive cells and endocytose soluble particles including proteins that may include both degraded and intact protein. To assess if intact proteins can be successfully transported across the intestine, Warshaw *et al.* assessed the proportion of intact tritium labelled bovine serum albumin reaching the lymph and blood circulation after duodenal infusion, finding that approximately 2% was able to be identified [265]. Immune cells have also been suggested to take part in trans-cellular transport including mast cells, dendritic cells, and macrophages [263].

1.8.2.2 Inhalation route

Compared to oral intake, inhalation has fewer strict barriers to restrict intact protein. Upon taking a breath the first barrier to inhaling particles that may contain protein is the particle size itself, where the inhalable range of proteins is less than 100 µm [266]. Larger particles, greater than 8 µm, tend to deposit in the lung further up in the

respiratory airway in a size-dependent manner from the nose to the larger bronchioles. Smaller particles, up to 3 μm, tend to diffuse throughout the lung tissue including deposition in the alveoli. Tidal breathing can also inhibit deposition of present particles by removing the particles prior to deposition upon respiratory tissues. The epithelial lining of the alveoli is principally responsible for resisting the transport of proteins across the lung [267]. The alveolar surfaces of the lung are lined principally by thin, single-cell thick, type I pneumocytes that are joined by tight junctions [267, 268].

Protein clearance from lung tissues often focus on pulmonary edema and therefore endogenous proteins rather than exogenous but can be used to inform how the mechanisms may apply to exogenous proteins. The rate of protein clearance in the alveolar spaces of the lung is approximately 1-2% per hour for albumin [267], but transport of albumin may be assisted through an albumin-specific receptor [269]. Depending on the amount of albumin present, mechanisms of transport proceed by receptor-mediated endocytosis at lower concentrations but by para-cellular transport at high concentrations. The rates that specific proteins are cleared from the lung depend on molecular size and or weight as smaller proteins clear faster than larger proteins. Further, insoluble proteins are degraded more slowly than soluble proteins. Degradation of proteins is mediated by proteases and peptidases as well as alveolar macrophage recruitment [269]. Proteases are first present to degrade proteins, but studies have not demonstrated that this is a significant means to remove present proteins as repeated studies have shown that greater than 95% of proteins remain intact and transport into the blood [269]. In the long term (>24 hours), macrophages can be recruited to the alveoli and significantly clear present proteins in a time scale from 2 to 6 days.

1.8.3 Inhalation of food allergens

Exposure and inhalation of food allergens can occur during food preparation or handling to generate aerosols and dust that can be inhaled by individuals and elicit allergic reactions [270]. Case reports have implicated many foods including fish, shellfish, seeds, soy, grains, eggs, and milk in causing allergic reactions due to inhalation of a food allergen. Examples of specific proteins implicated in inhalation based allergic reactions include shrimp tropomyosin [271] as well as hen's egg lysozyme [272].

Baker's asthma, an occupational allergic disease due to inhalation of cereal flours such as wheat flour, provides insights into comparing inhalation and oral allergies [273]. Many of the same proteins that are associated with oral wheat allergy are also implicated to some degree in inhalation allergy suggesting these proteins retain their allergenicity through either route of exposure. Notably, the serodominance of each protein allergen differs between oral and inhaled wheat allergy indicating that there are other key factors to consider such as their relative stability after aerosolization compared to consumption and capacity and ease to be aerosolized.

1.8.4 Insights into routes of exposure from biopharmaceuticals

The goal of a biopharmaceutical is to have an individual take an appropriate dose resulting in positive effects. This is contrasted with allergens causing negative effects. Studies of biopharmaceutical proteins and polypeptides can provide insight into how different routes of administration impact the amount found systemically. To better understand the differences among oral and inhalation the following examples of well

characterized biopharmaceuticals are explored in terms of their contrasting bioavailability.

Insulin is a 51 amino acid anabolic hormone with the mature form comprised of two peptide chains and therapeutically used to assist individuals with either type 1 or 2 diabetes [274]. Reviews of the bioavailability, the systemic exposure, of inhaled insulin is approximately 10% relative to injected [275], whereas the oral bioavailability has been regarded as less than 1% [276]. Then the inhaled bioavailability of insulin is approximately 10 times greater than when taken orally.

Human growth hormone is a 191 amino acid anabolic hormone comprised of a single polypeptide prescribed in cases of deficiencies or stunted growth [277]. A study in infants has measured the inhaled bioavailability of human growth hormone at approximately 3.5% [278], compared to a study in rats estimating that the oral bioavailability is 0.01% [279]. Each taken at face value, this results in a relative bioavailability of 350 times greater when inhaled compared to oral.

Using these examples, we can infer how allergens may differ depending on their route of exposure. If bioavailability limitations are due to the route of exposure, this suggests that the impact upon sequence and structure is a key factor. As inhalation of a protein does not have the same rigorous barriers to systemic entry as oral ingestion, the protein structure can be expected to be less impacted and without the excess of proteolytic enzymes similarly the primary sequence will also be kept intact. The risks of either oral or inhaled intake are not clear as these studies of biopharmaceuticals indicate that while the bioavailability is generally greater when inhaled it does not clarify typical scenarios when an allergen may be alternatively inhaled as opposed to consumed in terms

of dose. It is not realistic for an individual to inhale a similar dose that they may otherwise consume orally; however, the doses required can be expected to be lesser if they were to be inhaled as opposed to consumed orally.

1.8.5 Summary

While exposure to food allergens is predominantly through consumption, inhalation of food allergens is also possible during handling, preparation, and cooking of food and can similarly sensitize and elicit allergic reactions. The barriers and defenses against intact proteins entering the body differ greatly between oral and inhalation and may significantly impact consumers. Understating that those with a known food allergy may be at risk when inhaling food proteins is critical to adequately protect individuals.

1.9 Concluding remarks

The principles of risk analysis are sufficiently broad to be readily applicable to any hazard and keen understanding of allergens allows for specific considerations to be applied to novel allergens. Methods to characterize the potential allergenicity of novel foods is principally focused on cross-reactivity of proteins as evaluation of risk of *de novo* sensitization is limited by the limitations of *in vivo* (animal) models [280]. Further assessments of taxonomic relationships, homology of proteins to known allergens, and analysis of sera with allergies to the novel food or taxonomically similar sources can be utilized [281]. Given an intended use of the novel food, exposure may be estimated based on similar products and uncertainty incorporated reflecting information gaps to allow for characterization of the risk.

Food allergy is a present and increasing threat to public health and adequate controls are required to protect those with food allergies. Novel foods will continue to be discovered and produced; therefore, novel sources of food allergens need to be carefully scrutinized. The goal is not to unduly dissuade innovation but rather to inform consumers and producers of the risks associated with novel foods. With pertinent and wide-spread information stakeholders can more informed choices regarding novel foods and sources of novel allergens.

CHAPTER 2: PREDICTED ALLERGENS AND QUANTITATIVE PROTEOMICS FROM LIFE STAGES OF THE HOUSE CRICKET, *ACHETA DOMESTICUS*

2.1 Abstract

Background: Insects are increasingly important alternative food source to conventionally grown animal sources of protein, but consumption of insects poses a risk to those with shellfish allergy. *A. domesticus* (crickets) are commercially available in the form of several life stages with unclear differences in the potential risks posed to consumers. We evaluated life stages of crickets to determine the relative risks posed by consumption of crickets.

Methods: Eight life stages of crickets were extracted and their proteins quantified by mass spectrometry. Databases representing *ab initio* genomic predicted genes and translated transcriptomic genes were applied and identified proteins annotated for homology to known allergens. Predicted allergens with statistical differences among life stages were evaluated to compare relative risks.

Results: Ten proteins were robustly identified among life stages with significant differences among life stages and were predicted as allergens: a tropomyosin, arginine kinase, alpha-tubulin, paramyosin, two myosin light chains, and four hexamerins. Proportions of allergenic proteins to total proteins were equivalent excluding adult females, which were significantly lower.

Conclusions: Adult female *A. domesticus* were predicted to be of less risk to consumers than other life stages. Segregating adult crickets may pose an undue burden on producers posing a challenge to leverage the differences in predicted risk.

2.2 Introduction

Insects are an alternative protein source to conventional animal sources. While consumption is increasing and perceptions are improving in Western countries, entomophagy is still not widespread faces both neophobia and reports of allergic reactions associated with shellfish allergy [282]. Food allergies affect approximately 10.8% of US adults and shellfish allergy affects approximately 2.9% [81]. Evaluation of proteins present in *Acheta domesticus* crickets can help inform consumers as to the risks associated with consumption and inform producers as to what can be done to manage risks posed.

A. domesticus is a member of order Orthoptera and undergoes a hemimetabolous life cycle, where upon hatching from an embryo the nymph resembles an adult but without wings nor sexual organs and undergoes molts until finally reaching sexually dimorphic adulthood [283]. Proteomic shifts are expected to accompany the development of *A. domesticus* that may similarly result in possibly allergenic proteins changing in abundance. The phylum Arthropoda contains numerous allergens that demonstrate cross-reactivity among species including both the panallergens tropomyosin and arginine kinase [284]. The related two-spotted cricket, *Gryllus bimaculatus*, has demonstrated reactivity to hexamerin [133]. Allergens specific to *A. domesticus* are yet unknown and

so insight into potentially cross-reactive proteins can permit an evaluation of the risks present.

This study addresses the potential relative risk posed to consumers by various life stages of *A. domesticus* crickets. The potential risk is evaluated utilizing protein quantification supplied by mass spectrometry. Annotation of quantified proteins in terms of identity and associations with known allergens clarifies which proteins are of importance. We evaluated the quantifications and annotations of proteins to describe how the life stages differ and how that translates into allergenic risk to consumers. The aim was to evaluate if there are life stages of crickets that may pose relatively less risk to consumers.

2.3 Methods and Materials

2.3.1 Chemicals and samples

Tris, iodoacetamide (IAA), and rabbit glycogen phosphorylase B (GP; P6635; Uniprot P00489.3) were obtained from Sigma (St. Louis, MO, USA). Trypsin (Pierce, MS grade) was obtained from Thermo Scientific (Waltham, MA, USA). Dithiothreitol (DTT) was obtained from Arcos Organics (Geel, Belgium). Ammonium bicarbonate (ABIC) was obtained from Honeywell Fluka (Charlotte, NC, USA). Acetone, methanol, hexanes, acetic acid, acetonitrile (ACN), and formic acid were obtained from Fisher (Hampton, NH, USA). Precision plus Protein^{TM} Dual Xtra standards and Coomassie R250 were obtained from Biorad (Hercules, CA, USA). NuPage LDS sample buffer, Nupage MES SDS running buffer (20x), and 4-12% NuPage Bis-Tris gels were obtained from Invitrogen (Carlsbad, CA, USA).

Crickets (*Acheta domesticus)* were purchased from a cricket farm representing 7 developmental stages: pinhead, 1 week, 1/3 grown, 2 weeks, 1/2 grown, 2/3 grown, and adults; adults were segregated as males and females for a total of 8 samples (Appendix A Figure 2.1.1). Triplicate samples were taken per stage (n=24) to an approximate 200 mg (wet basis) of material per sample and ground by mortar and pestle under liquid nitrogen. Samples were stored -80 °C until use.

2.3.2 Extraction and SDS-PAGE

Samples were individually prepared for defatting in 2 mL Eppendorf tubes by adding 0.5 mL of acetone and vortexed to saturate the material. Defatting was performed by adding an additional 1 mL of acetone (>1:5 w/v) followed by vortexing, shaking for 10 minutes, and centrifugation (10 minutes, 17 k x g). The supernatant was discarded and acetone defatting repeated three additional times. The material was then defatted by substituting the acetone for hexanes and repeated as before four times. Pellets were dried in a fume hood overnight and were then stored at -80 °C.

Defatted samples (n=24) were extracted 1:20 (w/v) using a zwitterionic chaotropic buffer (50 mM Tris pH 8.8, 5 M urea, 2 M thiourea, 2% CHAPS, 67 mM DTT) in a heated sonicating water bath (60 °C, 30 minutes) and centrifuged (10 minutes, 17 k x g). Per extraction, half of the extraction buffer was removed as supernatant and fresh buffer added to maintain volume. Each pellet was extracted a total of three times and the three supernatants per defatted sample were pooled (n=24). Samples underwent protein quantification via 2D-Quant kit (GE healthcare, Chicago, IL, USA).

Samples were prepared for SDS-PAGE by adding sample buffer with 50 mM DTT. An equivalent of 500 µg of tissue per sample was fractionated. Gels were fixed with a solution of 50% methanol and 10% acetic acid for 1 hour and visualized with Coomassie R250. Gels were imaged using a UVP Biospectrum 815 Imaging System (Analytik Jena US LLC, Upland, CA, USA) with VisionWorks software (Version 8.20, Analytik Jena US LLC).

2.3.3 Mass Spectrometry

2.3.3.1 In-solution digestion

Sample volumes equivalent 250 µg of tissue were diluted 1:10 in zwitterionic chaotropic buffer and 4 volumes of chilled acetone added for an acetone precipitation (-80 °C, overnight). The samples were then centrifuged (4 °C, 30 minutes), decanted, and the pellet washed with 200 µL of 80% acetone and the centrifugation repeated. The pellet was then allowed to dry in a fume hood and stored at -80 °C until use.

In-solution digestion of acetone precipitated pellets was preformed according to Palmer *et al.* [285]. Briefly, proteins were reduced with DTT for 5 minutes at 95 °C, alkylated with IAA for 20 minutes in the dark, and digested twice with trypsin first for 3 hours at 37 °C and secondly overnight at 30 °C.

2.3.3.2 Preparation for mass spectrometry and running parameters

Digested peptides were cleaned, and MS preformed as previously described with modifications [285]. In brief, digests were cleaned using Pierce C-18 Spin columns (Thermo scientific), the eluate dried, and peptides resolubilized in 30 µL of 0.1% (v/v)

formic acid with 5% (v/v) ACN. Samples were aliquoted into vials with 12 µL of peptides and 2 µL of 200 fmol μL^{-1} GP. A 3.5 µL injection is equivalent to 25 µg of tissue and 100 fmol GP. Samples were injected into the MS in triplicate (N=72).

One-dimensional microscale liquid chromatography separation of tryptic peptides was performed with an UltiMate 3000RSL® liquid chromatography system (Thermo Scientific), equipped with a Hypersil Gold C18 1.9 µm, 100 x 1 mm analytical reversed phase column (Thermo Scientific). Mobile phase A consisted of 99.9% (v/v) water containing 0.1% (v/v) formic acid and mobile phase B was 99.9% (v/v) ACN containing 0.1% (v/v) formic acid. 3.5 µL of the sample was injected on-column and peptides were eluted from the analytical column and separated using a gradient of 2-40 % mobile phase B over 60 minutes at a flow rate of 60 µL/min. The analytical column temperature was maintained at 35 ºC.

Mass spectrometric analysis utilized a Q Exactive Plus™ Hybrid Quadrupole-Orbitrap™ MS (Thermo Scientific) in the data-independent mode with survey scans and fragment ion spectra acquired at a resolution of 70,000 at m/z 400. Up to the top 20 most abundant isotope patterns with charge 2 to 4 from the survey scan were selected with an isolation window of 2 m/z with a window offset of -0.4 m/z and fragmented by higher energy collisional dissociation with normalized collision energies of 27. The maximum ion injection times for the survey scan and the MS/MS scans were 100 and 240 ms, respectively, and the ion target value for scan modes were set to 3E6 and 1E5, respectively. Repeat sequencing of peptides was minimized by dynamic exclusion of the sequenced peptides for 20 s.

Injection quality was assessed via consistency of GP quantification using PEAKS 8.5 (Bioinformatics Solutions Inc., Waterloo, ON, Canada) [286]. against a Polyneoptera database (Uniprot, accessed 190403) and including GP. PEAKS search parameters were 5 ppm mass error tolerance, 0.05 da fragment mass error tolerance, no missed cleavages, and fixed carbamidomethyl cysteine. Data representing individual injections were handled independently and assessed together in PEAKS. GP was quantified via Top3 excluding peptides that were identified within the Polyneoptera database. Quantifications were tested for normality via Shapiro-Wilk Test using GraphPad Prism version 4.03 for Windows (GraphPad Software, La Jolla California USA) at a significance level of 0.05 both including and excluding outliers. Injections indicated as outliers from normality were regarded as injection failures. In this case, the injection was freshly prepared again and reinjected to replace the data from the injection failure.

2.3.4 Data Analysis

2.3.4.1 Protein identifications

Protein databases for MS analysis included *ab initio* genes predicted from the genome by Dovetail Genomics (Santa cruz, CA, USA) and an *A. domesticus* transcriptome (GenBank GHUU00000000.1) [287], translated in 6-frames by EMBOSS transeq [288]. Both databases were appended with 348 sequences representing GP and 347 decoy sequences identified from Uniprot from a search for “glycogen phosphorylase” within Insecta (obtained 18/09/2020). Data was analyzed using PEAKS 8.5 with search parameters including tryptic peptides, maximum of 3 missed cleavages, 5 ppm parent mass error tolerance, 0.05 fragment mass error tolerance, carbamidomethyl

cysteine as a fixed modification, and a contaminant database (www.thegpm.org/crap; version 3/3/19). Searches including post-translational modifications (PTM) were preformed using the default 313 modifications included by PEAKS. Data was analyzed using the multi-round search feature in four rounds in the order of *ab initio* predicted genes, the *A. domesticus* transcriptome, PTMs of *ab initio* predicted genes, and lastly PTMs of the *A. domesticus* transcriptome where *de novo* only spectra remaining from each prior search were passed to the following search.

Each search was quantified using the label-free quant feature with a 6-minute retention time shift tolerance and data adhering to a peptide-spectrum match FDR < 0.1% and at least 1 unique peptide per protein was exported for analysis. Data was aggregated from each analysis and searches without PTMs was used for quantification. PTMs were assessed relative to their unmodified forms. Proteins were evaluated for robust presence requiring that each protein be present in at least 2 technical replicates of a single biological replicate and 2 biological replicates of a single sample.

2.3.4.2 Protein quantification

Peptides were normalized to GP and converted to fmol. Proteins robustly identified within developmental stages were quantified by averaging across valid technical replicates and quantified by average of the top 3 most abundant unique peptides per protein, if possible. Data was handled in Microsoft Excel 365 and Graphpad Prism (version 9.0.0; GraphPad Software, San Diego, CA, USA). For statistical analyses, missing values were imputed with zeroes. Heatmaps are presented as $\log_2$(fmol) without imputation. Life stages and proteins were analyzed by two-way ANOVA at a significance

level of 0.05 with multiple comparisons tested by two-stage step-up method of Benjamini, Krieger, and Yekutieli at a significance level of 0.01.

2.3.4.3 Protein annotation

Proteins were queried against AllergenOnline by sliding 80mer windows with matches of greater than 35% (Version 20, Accessed 201027) [95]. Proteins with a full length E-value below 1E-7 and > 50% shared identity were regarded as predicted allergens [289]. Hits were not excluded based on taxonomy. Hexamerins identified by Oppert *et al.* were included as predicted allergens regardless of their evaluation by AllergenOnline [287]. The originating species of matched proteins were phylogenetically visualized by IcyTree [290].

Proteins identified by MS were compared with the non-redundant protein sequences database from NCBI using BLASTp-fast limited to Hexapoda (taxid: 6960) with a default E-value of 1E-3 [291]. Proteins were mapped and annotated in Blast2GO v5.2.5 (BioBam, Valencia, Spain) [292]. Sequences were also annotated by InterPro within Blast2GO with default settings and the annotations merged [293]. Gene ontology enrichment analyses were performed using Blast2GO using Fisher's exact test. Proteins representing the top 10% of proteins by quantification per life stage were submitted as the test set compared to the reference set of the remaining proteins (FDR < 0.05). Enriched terms were reduced to the most specific terms (FDR < 0.05).

2.4 Results

2.4.1 Characteristics of samples and cricket life stages

Crickets were handled to ensure at least 200 mg of sample was used during sample processing resulting in an increasing number of individuals used for earlier stages of development compared to later stages (Appendix A Table 2.1.1). There were no significant differences in protein content among life stages. Fractionation by SDS-PAGE evidenced more shifts in protein composition in adults than in earlier stages as a 35 kDa band in adult males and a band above 100 kDa in adult females but otherwise differences were indistinguishable from differences in total protein applied per lane (Appendix Figure A 2.1.2)

2.4.2 Proteomic shifts across life stages

Use of the multi-round search methodology resulted in most (97.5%) of all acquired spectra assigned to one of the databases (Appendix A Table 2.1.2). After label-free quantification requiring at least one unique peptide, the total number of valid proteins dropped to 675 encompassing 3977 peptides without and 81 peptides with non-carbamidomethyl post-translational modifications. Among life stages, there were no statistical differences in the mean peptides, mean proteins, or summed quantification of all proteins (Table 2.1). There were significant differences in the quantification of the top 10% most abundant proteins per stage and the proportion these proteins encompassed particularly in adult females.

Table 2.1 Life stage quantification characteristics

	Pinhead	1 Week	1/3 Grown	2 Weeks	1/2 Grown	2/3 Grown	Adult Male	Adult Female
Mean peptides Identified (CV%)	3132.7[a] (2.04%)	3176.0[a] (1.92%)	3264.3[a] (0.88%)	3178.7[a] (7.45%)	3213.3[a] (3.5%)	2685.3[a] (13.5%)	2567.7[a] (17.9%)	2501.3[a] (25.3%)
Mean proteins identified (CV%)	579.0[a] (1.79%)	593.3[a] (1.02%)	601.3[a] (1.58%)	598.7[a] (2.56%)	588.3[a] (1.82%)	537.7[a] (7.77%)	519.7[a] (9.96%)	479.3[a] (15.3%)
Summed fmol quantified (CV%)	2635.1[a] (27.5%)	2668.4[a] (10.4%)	3150.9[a] (4.4%)	4344.0[a] (17.0%)	3958.9[a] (23.9%)	2935.1[a] (22.9%)	3232.2[a] (26.1%)	4637.7[a] (19.6%)
Sum fmol of top 10% of proteins (CV%)	1394.2[a] (30.0%)	1406.4[a] (9.8%)	1784.3[a] (4.8%)	2623.4[ab] (18.8%)	2310.6[ab] (27.0%)	1767.8[a] (21.0%)	2065.5[a] (27.2%)	3433.9[b] (15.4%)
%fmol represented by top 10% of proteins (CV%)	52.6%[a] (2.9%)	52.7%[a] (2.8%)	56.6%[ab] (0.8%)	60.3%[bc] (3.0%)	58.1%[b] (3.5%)	60.4%[bc] (2.2%)	63.8%[c] (2.2%)	74.5%[d] (6.1%)

Values that do not share common letters are significantly different by one-way ANOVA ($p < 0.05$) and two-stage linear step-up procedure of Benjamini, Krieger, and Yekutieli (FDR < 0.01).

Comparison of quantifiable proteins across life stages resulted in most being similar with correlation coefficients between 0.45-0.99 comparing all proteins (Figure 2.1). Focusing only on proteins that were robustly identified in at least 7 of 8 life stages (74.5% of all proteins), comparison of the quantifiable proteins demonstrated correlation coefficients between. 0.84-0.99. Adult females were less correlated due to several highly abundant proteins that were uniquely identified in adult females.

Data were further assessed for biologically derived posttranslational modifications to match as much of the data possible to the databases used. A total of six modifications were identified in the form of N-terminal acetylation, acetylation of lysine, and methylation but only one, $[YPIEHGIITNWDDMEK]^{3+}$, was also identified as unmodified, which was observed almost exclusively as the modified form in later stages (85.1% in pinheads, 96.3% in adult females; Appendix A Figure 2.1.3 and Appendix A Table 2.1.3). Each peptide was observed in at least 66 injections (91.6%) with spectra that substantiated the position of the modification.

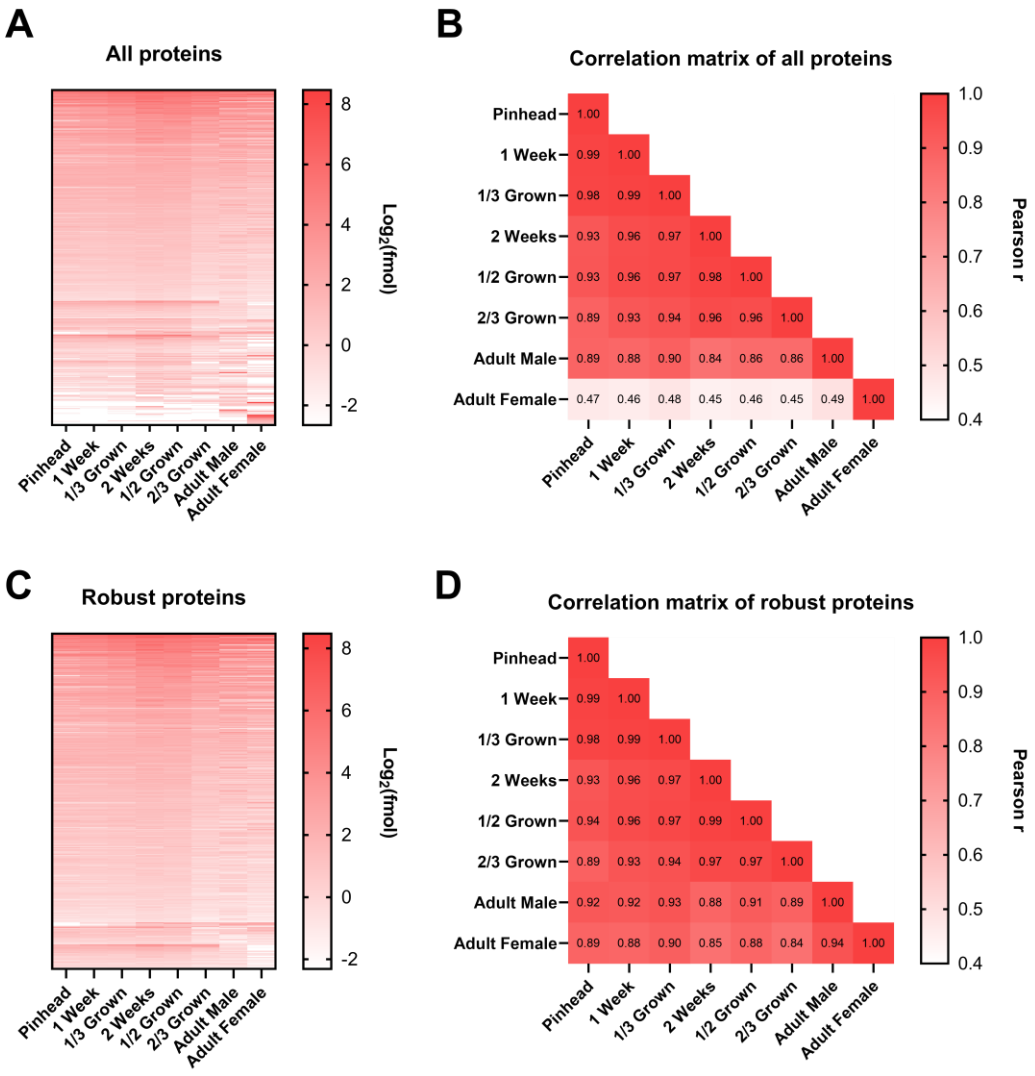


Figure 2.1 Heatmaps and Pearson correlation matrices of all proteins and robust proteins

Heatmaps (A and C) and Pearson correlation matrices (B and D) of all quantified proteins (A and B) and proteins robustly identified across at least 7 of 8 life stages (C and D). Data are presented from the Log_2 (fmol) per protein without imputation.

2.4.3 Proteome annotation and gene ontology enrichment

Of quantified proteins, 85.5% were successfully annotated, 9.5% identified without gene ontology (GO) terms, and the remaining 5% did not have significant

BLAST search results (Appendix B). A total of 64 proteins (9.5%) were predicted to be allergens, most of which (75%) had their best allergenic match to members of Arthropoda (Appendix A Figure 2.1.4). For allergen hits that were well described, most closely matched the descriptions assigned via BLAST; however, proteins that did not meet the cutoff criteria for predicted allergens had more mismatches between BLAST and Allergenonline searches.

GO enrichment was applied to evaluate the primary functional focus during each life stage (Figure 2.2, Appendix A Table 2.1.4). Life stages were similar in a focus on higher representation of proteins associated with motor activity, actin filament binding, and myosin complex terms. Prior to 2 weeks of development, terms including purine ribonucleotide and triosephosphate binding were enriched. Between 2 weeks and 2/3 grown, enriched terms indicate an increasing focus on cuticular structure and monophenol oxidase activity with relatively less focus on metabolism. The shift into monophenol oxidase activity is associated with proteins annotated as hexamerins that are highly abundant in 2 weeks and 1/2 grown crickets. Adult males and females continued with a lesser focus on metabolism while females greatly increased proteins associated with lipid transport. These lipid transport proteins were predominantly annotated as vitellogenins, which were almost exclusively identified in adult females.

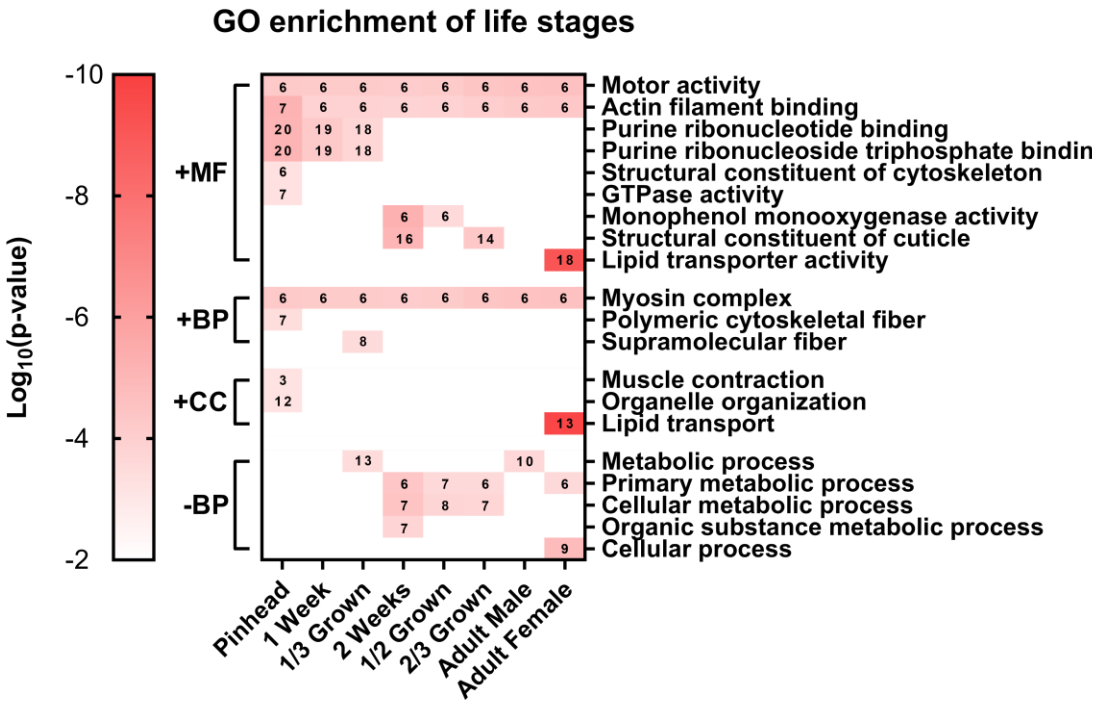


Figure 2.2 GO enrichment of life stages

Summary of GO enrichment results of top 10% of proteins per life stage of *A. domesticus*. GO terms significantly enriched (FDR<0.05) and reduced to most specific terms (FDR<0.05). Cell color represents $\log_{10}$(p-value) from a floor of p=0.01 and numbers per cell represent number of proteins with enriched term. MF: Molecular function, BP: Biological process, CC: Cellular component, +: Overrepresented term, -: Underrepresented term.

2.4.4 Predicted allergens across life stages

The quantitative distribution of predicted allergens showed no pattern and was not significantly different compared to the remaining non-predicted allergens for any of the life stages (Figure 2.3).

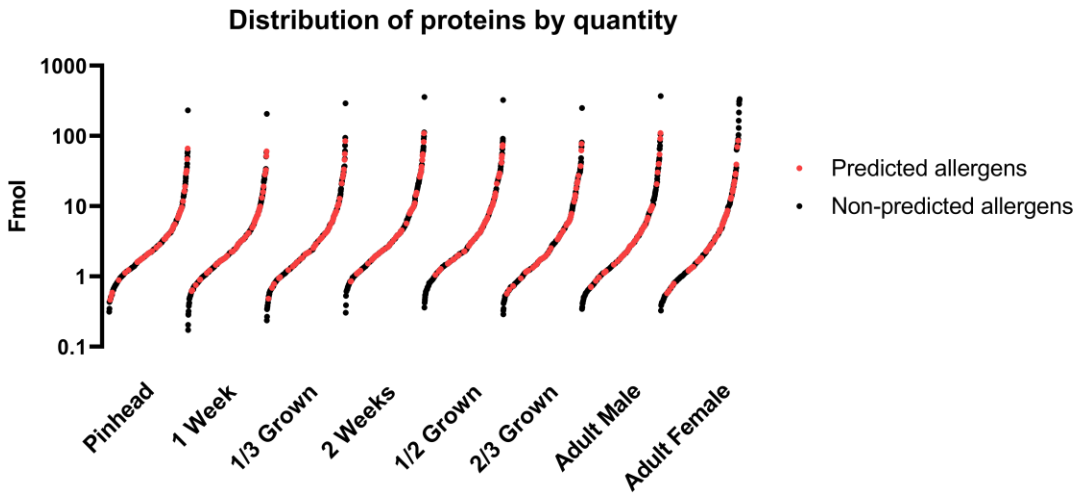


Figure 2.3 S-curves of quantifiable proteins and predicted allergens per life stage

Quantifiable proteins per life stage differentiated by predicted allergens (red) and non-predicted allergens (black). The x-axis is arbitrary and scaled. No differences were found in average quantity of predicted allergens and non-predicted allergens in any life stage by two-way ANOVA ($p < 0.05$) and two-stage linear step-up procedure of Benjamini, Krieger, and Yekutieli (FDR < 0.01).

Most proteins identified (73.0%) were robustly identified in 7 of the 8 life stages (Figure 2.4). Only a small subset of 10 proteins were robustly identified, predicted as allergens, and statistically differed among the life stages. Predicted allergens that were robustly identified comprised a vast majority of the quantifiable protein (97.3%) compared to proteins not robustly identified across life stages and therefore were the primary focus.

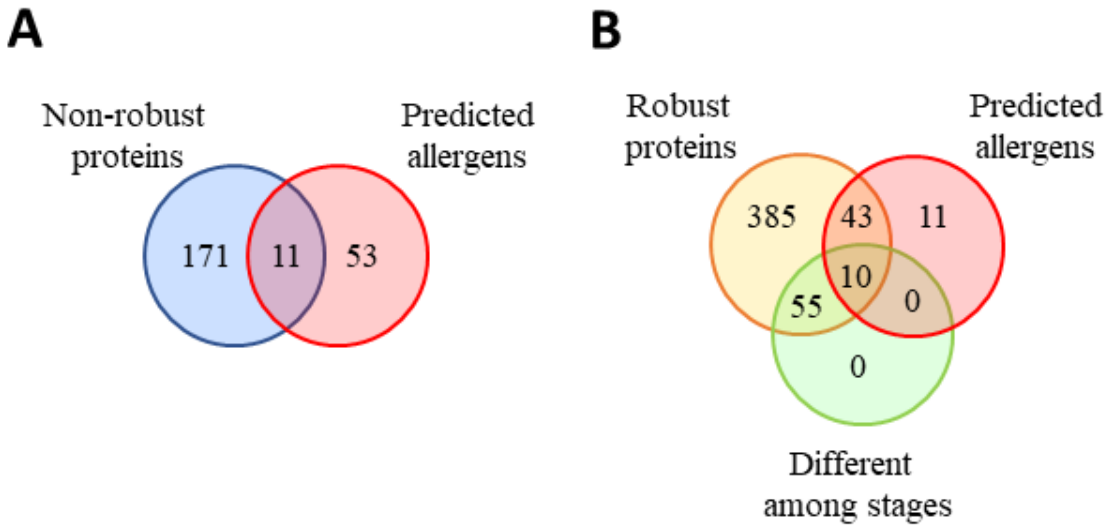


Figure 2.4 Venn diagram of protein assignments and annotations

(A) 182 proteins identified by mass spectrometry were not robustly identified across at least 7 of 8 life stages with 64 proteins predicted as allergens and 11 not robustly identified and predicted as allergens. (B) Within 493 robustly identified proteins, 65 were significantly different among life stages by two-way ANOVA ($p < 0.05$) and two-stage linear step-up procedure of Benjamini, Krieger, and Yekutieli (FDR < 0.01). 10 proteins were robustly identified, predicted as allergens, and significantly different among life stages.

Of predicted allergens that were robustly identified but found to be significantly different among life stages, four hexamerins were identified and demonstrated far greater amounts in crickets at 2 weeks, 1/2 grown, and to a lesser degree adult females (Figure 2.5 A). The remaining six predicted allergens either were variable from stage to stage or broadly increasing over development. Evaluation of the variable predicted allergens to all predicted allergens evidenced a significant and increasing trend toward a greater proportion of the allergens being encompassed by these variable predicted allergens (Figure 2.5 B). Comparison between all predicted allergens and all remaining proteins only was significant for adult females, which had significantly less predicted allergenic content compared to other life stages.

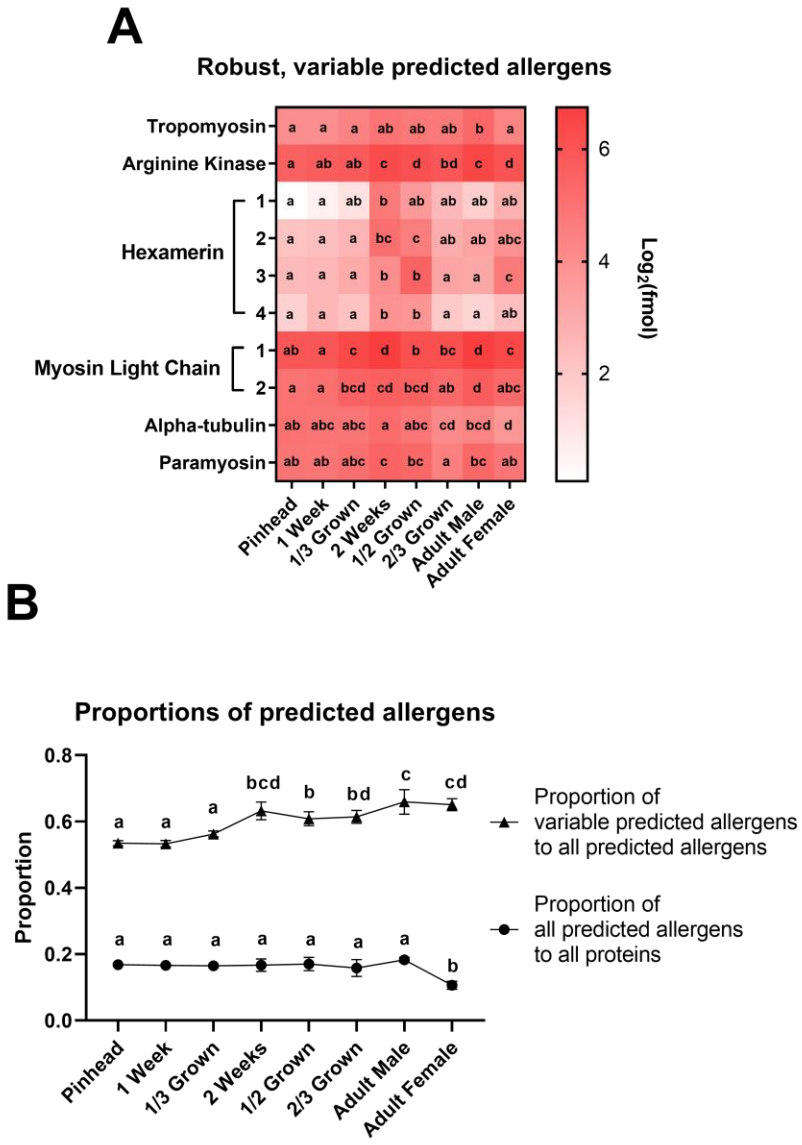


Figure 2.5 Robust and variable predicted allergen heatmap and characteristics of predicted allergens across life stages

(A) Heatmap of predicted allergens robustly identified across at least 7 of 8 life stages with significant differences across life stages presented as Log_2 of the sum of each group per life stage. (B) Proportions of quantified variable predicted allergens compared to all predicted allergens and the proportion of predicted allergens to all proteins. Values per protein or data point that do not share common letters are significantly different by two-way ANOVA ($p < 0.05$) and two-stage linear step-up procedure of Benjamini, Krieger, and Yekutieli (FDR < 0.01).

2.5 Discussion

This study was devised to follow the development of *A. domesticus* crickets from a proteomic perspective, evaluate the identity of present proteins, predict plausible allergenic proteins, and examine the protein-level differences among life stages. We sought to explain what various *A. domesticus* life stages are doing at the protein level as well as if there were any differences among life stages that could be leveraged to decrease the risk posed to consumers.

Comparison of the GO enrichment performed here based on protein abundance to that of Oppert *et al*. based on transcript abundance shows little correlation via the terms shared by similar life stages [287]. Far fewer proteins were identified here to serve as a test set for GO enrichment as opposed to highly abundant transcripts resulting in a less rich comparison of life stages at the protein level. Terms enriched at the protein level for all stages related to muscle development and similar terms can be seen sparingly in the transcriptome in the form of myosin complex, troponin complex, and calcium ion binding (e.g. tropomyosin and myosin light chain); however, terms enriched among highly abundant transcripts such as cytochrome-c oxidase activity or related to DNA were not enriched in abundant proteins. Notably, terms related to lipid transport that were enriched among highly abundant proteins were absent in adult female transcripts. The model insect, *Drosophila melanogaster*, has also evidenced a poor correlation between mRNA and protein [294]. Factors explaining the discrepancy includes protein stability as related to protein turnover as highly stable proteins do not require high levels of transcription to become abundant as well as post-transcriptional controls that maintain protein levels [295].

Protein annotation identified homologues to many known allergens of cockroaches, shellfish, and house dust mites and were predicted as allergens in *A. domesticus*. As many of the predicted allergens were identified as robust and stable among the life stages, focus was given to predicted allergens that showed variation among life stages. Given the cumulative stability in overall predicted allergen abundance it is likely that the variability observed reflects both the choice to normalize samples on a tissue basis rather than protein as well as the naturally greater noise observed in the quantity of highly abundant proteins [296]. Of the focused upon predicted allergens, both tropomyosin and arginine kinase are known as panallergens among arthropods [160, 297]. Myosin light chain is known as an allergen in cockroaches (Bla g 8), shrimp (Pen m 3), and house dust mites (Der p 26) and cross-reactivity has been speculated among invertebrates but not clearly established [92, 298]. Both alpha-tubulin (Der p 33) and paramyosin (Der p 11) are known as allergens in house dust mites. Alpha-tublin of *Tenebrio molitor*, the yellow mealworm, was identified using immunoprecipitation using sera of patients with known shellfish allergy and regarded as a putative allergen [299]. Paramyosin has been described as potentially cross-reactive among house dust mites and mollusks [92], but paramyosin from the mollusk *Halitosis discus discus*, the disc abalone, has been demonstrated to have cross-reactivity with tropomyosin [300].

Hexamerin has been identified as a novel allergen in *Gryllus bimaculatus* [133], the two-spotted cricket, although the evidence provided was limited to possible IgE-binding and 2 peptides matched to *Tribolium castaneum* Hexamerin 1B precursor but without reporting the identity of the peptide sequences matched. Hexamerins are a member of the greater hemocyanin superfamily, which comprises many similar proteins

with diverse function, and care was taken to ensure that the proteins identified were attributable to the appropriate storage hexamerins as opposed to phenoloxidases, hormone binding proteins, or hemocyanins [301]. Ten members of the superfamily were identified in part with allergen hits to either Per a 3 or *Fenneropenaeus merguiensis* hemocyanin, of these three (ANN16672.RA, ANN17126.RA, GHUU01045934.1.4) were annotated as phenoloxidases rather than storage hexamerins by Oppert *et al*. [287], and another two (ANN20571.RA, ANN20572.RA) were evaluated as juvenile hormone binding proteins via matching peptides and N-terminal sequencing according to the work of Tawfik *et al.* on *G. bimaculatus* [302]. The remaining five (ANN12312.RA, ANN12313.RA, ANN12314.RA, ANN12315.RA, GHUU01039257.1.1) were identified using contigs from Oppert *et al.* [287]. Members of the hemocyanin superfamily have been demonstrated as allergens across Arthropoda, but cross-reactivity is less clear but may be achieved through cross-reactive carbohydrate determinants [303].

Among the surveyed life stages, predicted allergenic content was relatively stable saving the exception of adult females, which had significantly less predicted allergenic content. However, adult females were highly rich in vitellogenins. Various vitellogenins are known allergens such as Api m 12 and Ves v 6, which are injection allergens and cross-reactive despite low shared identity (40%) [304]. Cockroach vitellogenins have also been demonstrated to bind IgE at relatively higher rates than other cockroach allergens [305]. Cockroach vitellogenin is distantly identical to other known allergenic vitellogenins such as Der p 14, although fragments of Der p 14 has been shown to be highly IgE reactive and speculated to act as an adjuvant to enhance IgE responses to itself and other potential allergens [306]. It is not clear if the distances in identity would result

in vitellogenins of *A. domesticus* sharing characteristics of other known allergenic vitellogenins but should be a consideration if IgE binding is identified in the future.

Many of the allergens identified are principally known as airway allergens, which suggests that farming operations should be aware of the risks posed during cricket rearing. For consumers, the risks are yet unclear. *A. domesticus* poses risks at each life stage surveyed and while current data suggests adult females may pose lesser risk the need to segregate male and female crickets to leverage this difference may pose undue burden on producers. We suggest that consumers be adequately informed of the risks they are undertaking and that crickets be cautiously consumed by those with current allergies to arthropods.

CHAPTER 3: PROTEOMIC EVALUATION OF YELLOW MEALWORM, TENEBRIO MOLITOR, WITH EXPERIMENTAL GENETIC MODIFICATIONS

3.1 Abstract

Background: *T. molitor* (yellow mealworm) larvae are an important alternative food and protein source to conventionally grown animal protein but retain risks to those with crustacean shellfish allergy. Genetic modification can result in unintended effects to the presence and quantity of other proteins present, including potential allergens. We evaluated wildtype and proprietary genetically modified variants of *T. molitor* to assess the allergenic risks posed as well as the effects of the modifications on the proteome.

Methods: Three mealworm samples including one wildtype and two genetically modified variants were extracted and proteins investigated by mass spectrometry. Databases representing *ab initio* genomic predicted genes and publicly available protein sequences were used and identified proteins annotated for homology to known allergens. Predicted allergenic proteins with statistical differences among samples were used to assess the relative risks posed.

Results: A total of 655 proteins were identified among samples with 531 identified robustly in each sample. 45 proteins were predicted as allergens representing many known allergens of Arthropods, but 2 proteins, arginine kinase and a myosin light chain, were predicted as allergens and significantly elevated in one of the genetic variants.

Conclusions: Compared to wildtype *T. molitor* one of the genetically modified samples posed no greater allergenic risk but the other posed an elevated risk. Wildtype *T. molitor* poses risks to consumers particularly those with allergies to crustacean shellfish, cockroaches, or house dust mites.

3.2 Introduction

Food allergies affect approximately 10.8% of U.S. adults with shellfish allergy affecting 2.9% [81]. Symptoms of a food allergic reaction varies but can include rash, hives, shock, and anaphylaxis. *Tenebrio molitor* has been assessed by the European Food Safety Authority (EFSA) as a safe novel food with respect to whole insects and as a powder at a maximum level of 100% in the form of snack foods [307]. The scientific opinion notes that *T. molitor* as a food contains risks of both *de novo* sensitization as well as cross-reactions with those with allergies to either crustaceans or dust mites.

Safety assessment of genetically modified organisms relies on establishing substantial equivalence relative to the unmodified organism and can be performed at the level of the metabolome, transcriptome, or proteome [308]. Proteomic evaluations rely on evaluating the effects on protein levels due to the genetic modification and so the natural variability observed in accepted unmodified organisms plays a key role in assessing genetically modified organisms. Food allergens are predominantly proteins and therefore assessment of the effects of genetic modification on allergen levels is important. Allergens of interest in mealworms include the panallergens tropomyosin and arginine kinase as well as other proteins known to be allergenic in other Arthropods such as crustaceans, cockroaches, and house dust mites [309, 310].

This study evaluated the risk posed by wildtype *T. molitor* as well as two genetically modified variants. The risk was assessed by protein quantification using mass spectrometry. Annotation of quantifiable proteins clarified what proteins were identified and searches against allergen databases predicted which proteins were of relevance. We evaluated both the quantification and annotations of proteins to evaluate the relative allergenic risks of *T. molitor* samples as well as the effects of genetic modification.

3.3 Methods and Materials

3.3.1 Samples and Chemicals

Tris, iodoacetamide (IAA), 3-[(3-Cholamidopropyl)dimethylammonio]-1-Propanesulfonate (CHAPS), and rabbit glycogen phosphorylase B (GP; P6635; Uniprot P00489.3) were obtained from Sigma (St. Louis, MO, USA). Trypsin (Pierce, MS grade) was obtained from Thermo Scientific (Waltham, MA, USA). Dithiothreitol (DTT) and Thiourea were obtained from Arcos Organics (Geel, Belgium). Ammonium bicarbonate (ABIC) was obtained from Honeywell Fluka (Charlotte, NC, USA). Acetone, acetonitrile (ACN), hexanes, LC/MS grade water, and formic acid were obtained from Fisher (Hampton, NH, USA). Urea was obtained from Biorad (Hercules, CA, USA).

Yellow mealworm larvae (*Tenebrio molitor*) were obtained from the Dossey laboratory including wildtype mealworms and two genetically modified variants denoted as WT, GM1, and GM2, respectively. Information regarding the genetic modifications was proprietary.

3.3.2 Sample preparation and protein extraction

For each type of mealworm (N=3), individual mealworms were ground by mortar and pestle under liquid nitrogen in triplicate (N=9). Ground samples were stored at -80 °C until use. Samples were individually prepared for defatting in 1.5 mL Eppendorf tubes by adding 0.5 mL of acetone and vortexed to saturate the material. Defatting was performed by adding an additional 1 mL of acetone (>1:5 w/v) followed by vortexing, shaking for 10 minutes, and centrifugation (10 minutes, 17 k x g). The supernatant was discarded and acetone defatting repeated three additional times. The material was then defatted by substituting the acetone for hexanes and repeated as before four times and the pellets dried in a fume hood overnight.

Defatted samples (N=9) were extracted 1:20 (w/v) using a zwitterionic chaotropic buffer (50 mM Tris pH 8.8, 5 M urea, 2 M thiourea, 2% CHAPS, 67 mM DTT) in a heated sonicating water bath (60 °C, 30 minutes) and centrifuged (10 minutes, 17 k x g). Per extraction, half of the extraction buffer was removed as supernatant and fresh buffer added to maintain volume. Each pellet was extracted a total of three times and the three supernatants per defatted sample were pooled (N=9).

3.3.3 Mass Spectrometry

3.3.3.1 Sample preparation for mass spectrometry

Per sample, 10 µL of each extract was diluted 1:10 in zwitterionic chaotropic buffer and to 10 µL of the diluted extract (equivalent to 1 µL undiluted) 4 volumes of chilled acetone added for an acetone precipitation (-80 °C, overnight). The samples were then centrifuged (4 °C, 30 minutes), decanted, and the pellet washed with 200 µL of 80%

acetone and the centrifugation repeated. The pellet was then allowed to dry in a fume hood and stored at -80 °C until use.

Pellets were reduced by adding 45 µL of 50 mM ABIC, 4.5 µL of 100 mM DTT, 26.7 µL of water, and 4.8 µL of 100% ACN and incubating at 95 ºC for 5 minutes. The samples were alkylated by adding 9 µL of 100 mM of IAA and incubating at room temperature in the dark for 20 minutes. The samples were digested by adding 3 µL of 100 ng μL^{-1} trypsin (Pierce, MS grade) and incubating at 37 ºC for 3 hours. A further 3 µL of 100 ng μL^{-1} trypsin was added and incubated at 30 ºC for overnight. The reaction was quenched by freezing.

Pierce C-18 Spin columns (Thermo Scientific) were used according to manufacturer's instructions excluding the final elution, which was performed using 50% ACN. Eluted samples were dried by centrifugal vacuum evaporation (RC1010, Jouan, Saint-herblain, Pays de la Loire, France). Dried samples were resolublized in 27 µL of 0.1% (v/v) formic acid and 5% (v/v) ACN. Samples were aliquoted into vials with 9 µL of resolublized peptides and 1.5 µL of 200 fmol μL^{-1} rabbit glycogen phosphorylase B (GP; Uniprot sequence P00489.3; Sigma-Aldrich #P6635) where a 3.5 µL injection would be equivalent to 100 fmol of GP. Samples were injected in triplicate (N=27).

3.3.3.2 Running parameters for mass spectrometry

One-dimensional microscale liquid chromatography separation of tryptic peptides was performed with an UltiMate 3000RSL® liquid chromatography system (Thermo Scientific), equipped with a Hypersil Gold C18 1.9 µm, 100 x 1 mm analytical reversed phase column (Thermo Scientific). Mobile phase A consisted of 99.9% (v/v) water

containing 0.1% (v/v) formic acid and mobile phase B was 99.9% (v/v) ACN containing 0.1% (v/v) formic acid. Two µL of the sample was injected on-column and peptides were eluted from the analytical column and separated using a gradient of 2-40 % mobile phase B over 60 minutes at a flow rate of 60 µL/min. The analytical column temperature was maintained at 35 ºC.

Mass spectrometric analysis utilized a Q Exactive Plus™ Hybrid Quadrupole-Orbitrap™ MS (Thermo Scientific) in the data-independent mode with survey scans and fragment ion spectra acquired at a resolution of 70,000 at m/z 400. Up to the top 20 most abundant isotope patterns with charge 2 to 4 from the survey scan were selected with an isolation window of 2 m/z with a window offset of -0.4 m/z and fragmented by higher energy collisional dissociation with normalized collision energies of 27. The maximum ion injection times for the survey scan and the MS/MS scans were 100 and 240 ms, respectively, and the ion target value for scan modes were set to 3E6 and 1E5, respectively. Repeat sequencing of peptides was minimized by dynamic exclusion of the sequenced peptides for 20 s.

3.3.4 Data Analysis

3.3.4.1 Protein identification

Protein databases for mass spectrometry data analysis were constructed using translated genes predicted using Augustus (v 3.4.0) against a *T. molitor* genome acquired from NCBI (WGS Project JABDTM01) [311, 312]. Parameters for Augustus included softmasking, the species set to tribolium2012, alternatives from evidence set to true, and a hints file generated using Peptimprove with default settings and utilizing the prior

genome along with raw mass spectrometric data generated from wildtype *T. molitor* [313]. A total of 26806 translated genes were predicted using Augustus. The database appended with 348 sequences representing GP and 347 decoy sequences identified from Uniprot from a search for "glycogen phosphorylase" within Insecta (obtained 18/09/2020) as well as 13 tropomyosin sequences of *T. molitor* manually curated by comparison of genomic contigs to the tropomyosin-2 gene structure of *Tribolium castaneum* (NCBI LOC656914; Appendix C Table 3.1.1 and Appendix C Table 3.1.2). Translated genes representing tropomyosin-1 and arginine kinase were also compared with *T. castaneum* gene structures (NCBI LOC656904 tropomyosin-1 and LOC660479 arginine kinase) but predictions were of sufficient quality and were not altered (Appendix C Figure 3.1.1). A second database of 626 *T. molitor* proteins was acquired from Uniprot (Obtained 10/01/2021).

Data was analyzed using PEAKS 8.5 with search parameters including tryptic peptides, maximum of 1 missed cleavage, carbaminomethyl cysteine as a fixed modification and a contaminant database (www.thegpm.org/crap; version 03/03/2019). Searches including post-translational modifications (PTM) were preformed using the default 313 modifications included by PEAKS. Data was analyzed using the multi-round search feature in four rounds in the order of the concatenated Augustus predicted gene database, Uniprot database, PTMs of the concatenated Augustus database, and lastly PTMs of the Uniprot database where *de novo* only spectra remaining from each prior search were passed to each following search.

Each search was quantified using the label-free quant feature with a 6-minute retention time shift tolerance and data adhering to a peptide-spectrum match FDR < 0.1%

and at least 1 unique peptide per protein was exported for analysis. Data was aggregated from each analysis and searches without PTMS used for quantification. PTMs were assessed relative to any unmodified forms. Proteins were evaluated for robust presence requiring that each protein be present in at least 2 technical replicates of a single biological replicate and 2 biological replicates of a single sample.

3.4.4.2 Protein quantification

Peptides were normalized to GP and converted to fmol. Proteins robustly identified within samples were quantified by averaging across valid technical replicates and quantified by average of the top 3 most abundant unique peptides per protein, if possible. Data was handled in Microsoft Excel 365 and Graphpad prism (version 9.0.0; GraphPad Software, San Diego, CA, USA). For statistical analyses, missing values with imputed with zeroes. Heatmaps are presented as $\log_2$(frmol) without imputation. Samples and proteins were analyzed by two-way ANOVA at a significant level of 0.05 with multiple comparisons tested by two-stage step-up method of Benjamini, Krieger, and Yekutieli at a significance level of 0.01.

3.4.4.3 Protein annotation

Proteins were queried against Allergenonline by Full Fasta 36 with an E-value cutoff of 1 and 20 max alignments shown (Version 20, accessed 13/01/2021) [95]. Proteins with a full length E-value below 1E-7 and > 50% shared identity were regarded as predicted allergens and the top hit retained for annotation [289]. Hits were not

excluded based on taxonomy. The originating species of matched proteins were phylogenetically visualized by IceTree [290].

Proteins identified by mass spectrometry were compared with the non-redundant protein sequences database from NCBI using BLASTp-fast limited to Hexapoda (taxid: 6960) with a default E-value of 1E-3 [291]. Proteins were mapped and annotated in Blast2GO v5.2.5 (BioBam, Valencia, Spain) [292]. Sequences were also annotated by InterPro within Blast2GO with default settings and the annotations merged [293]. Gene ontology enrichment analyses were performed using Blast2GO using Fisher's exact test. Proteins representing the top 10% of proteins by quantification per life stage were submitted as the test set compared to the reference set of the remaining proteins (FDR < 0.05). Enriched terms were reduced to the most specific terms (FDR< 0.05).

3.4 Results

3.4.1 Proteomics of *T. molitor* samples

The multi-round search methodology allowed for most (93.8%) of acquired spectra assigned to any of the four databases (Appendix C Table 3.1.3). After filtering the label-free quantification results by requiring at least one unique peptide, the pool of valid proteins decreased to 655 encompassing 4154 peptides without and 129 peptides with non-carbamidomethyl post-translational modifications. There were no statistical differences identified among mealworm samples regarding quantification metadata such as the mean number of peptides, proteins, or quantification of proteins (Table 3.1).

Table 3.1 Sample quantification characteristics

		WT	GM 1	GM 2
	Mean peptides identified (CV%)	3637.67 (4.54%)	3819.67 (2.11%)	3789.33 (2.57%)
	Mean proteins identified (CV%)	576.67 (2.08%)	594.33 (0.54%)	600.33 (1.93%)
All identified proteins	Summed fmol quantified (CV%)	7112.26 (18.11%)	7721.74 (7.66%)	8120.95 (7.16%)
All identified proteins	Summed fmol of top 10% (CV%)	5135.38 (17.54%)	5493.03 (8.19%)	5870.05 (7.26%)
All identified proteins	%fmol represented by top 10% (CV%)	72.2% (1.12%)	71.1% (2.30%)	72.3% (2.66%)
Robustly identified proteins	Summed fmol quantified (CV%)	6960.38 (17.99%)	7501.65 (7.44%)	7966.67 (6.87%)
Robustly identified proteins	Summed fmol of top 10% (CV%)	4987.32 (17.22%)	5265.57 (8.11%)	5682.36 (7.15%)
Robustly identified proteins	%fmol represented by top 10% (CV%)	71.7% (0.94%)	70.2% (2.30%)	71.3% (2.87%)
	Proportion of fmol from robustly identified proteins to all identified proteins (CV%)	97.9% (0.79%)	97.2% (0.45%)	98.1% (0.39%)

Values per category were assessed by one-way ANOVA ($p < 0.05$) and none were significantly different. Robustly identified proteins were those identified as present in each of the three samples as opposed to all proteins that were found in a single sample independent of other samples. WT: Wildtype *Tenebrio molitor*, GM: genetically modified *Tenebrio molitor*.

Assessing quantifiable proteins among mealworm samples evidenced minimal differences assessing either all proteins or only proteins quantified in each of the three samples (Figure 3.1). Each of the samples were highly comparable regardless of the set of proteins used as Pearson correlation coefficients were each above 0.97. Proteins that were robustly identified across samples were of principal focus as they embodied most proteins identified (79.5%) and encompassed most all the quantifiable protein (Table 3.1).

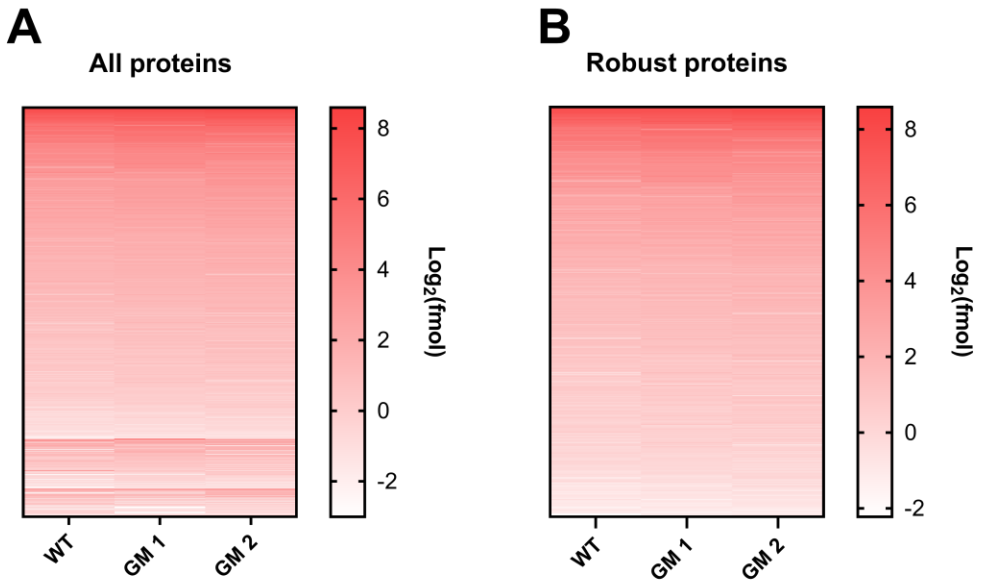


Figure 3.1 Heatmaps of all proteins and robustly identified proteins

Heatmaps of all quantified proteins (A) and proteins robustly identified across each of the three mealworm samples (B). Pearson correlation coefficients were greater than 0.97 for each comparison per data set ($p < 0.001$). Data are presented from the $\log_2$(fmol) per protein without imputation. WT: Wildtype *Tenebrio molitor*, GM: genetically modified *Tenebrio molitor*.

Data were evaluated for biologically derived posttranslational modifications to make the most of the data. A total of four modifications were identified as N-terminal acetylation and methylation of serine and histidine. Only one of these peptides, $[NLLISASSQYTK]^{2+}$ was also identified as unmodified, although only in WT (Appendix C Figure 3.1.2 and Appendix C Table 3.1.4). Excluding the unmodified $[NLLISASSQYTK]^{2+}$, each of the modified peptides were observed in at least 22 injections (81.5%) with spectra substantiating the position and identity of the modifications.

3.4.2 Proteome annotation and gene ontology enrichment

Among quantified proteins, 90.0% were annotated or assigned gene ontology (GO) terms, 8.6% were not assigned GO terms, and the remaining 1.4% had no significant BLAST results (Appendix D). Proteins predicted to be allergens encompassed 45 proteins (6.9%) with most (73.3%) having their best allergenic match to members of Arthropoda (Appendix C Figure 3.1.3). Matches to allergenic proteins closely resembled assigned annotations.

GO enrichment was utilized to evaluate differences in functionality among *T. molitor* samples in terms of either all proteins or only proteins robustly identified among each sample (Figure 3.2, Appendix C Table 3.1.5 and Appendix C Table 3.1.6). GO enrichment did not show large differences through assessment of all proteins as opposed to robust proteins indicating that the cutoff of the top 10% of quantified proteins was a contributor to differences among samples. WT mealworms evidenced an enrichment of proteins associated with the structure of the cuticle among highly abundant proteins as well as a lack of representation of terms associated with metabolism. Both GM mealworms showed similar trends compared to WT mealworms with some differences in the particular underrepresentation of terms associated with biological processes.

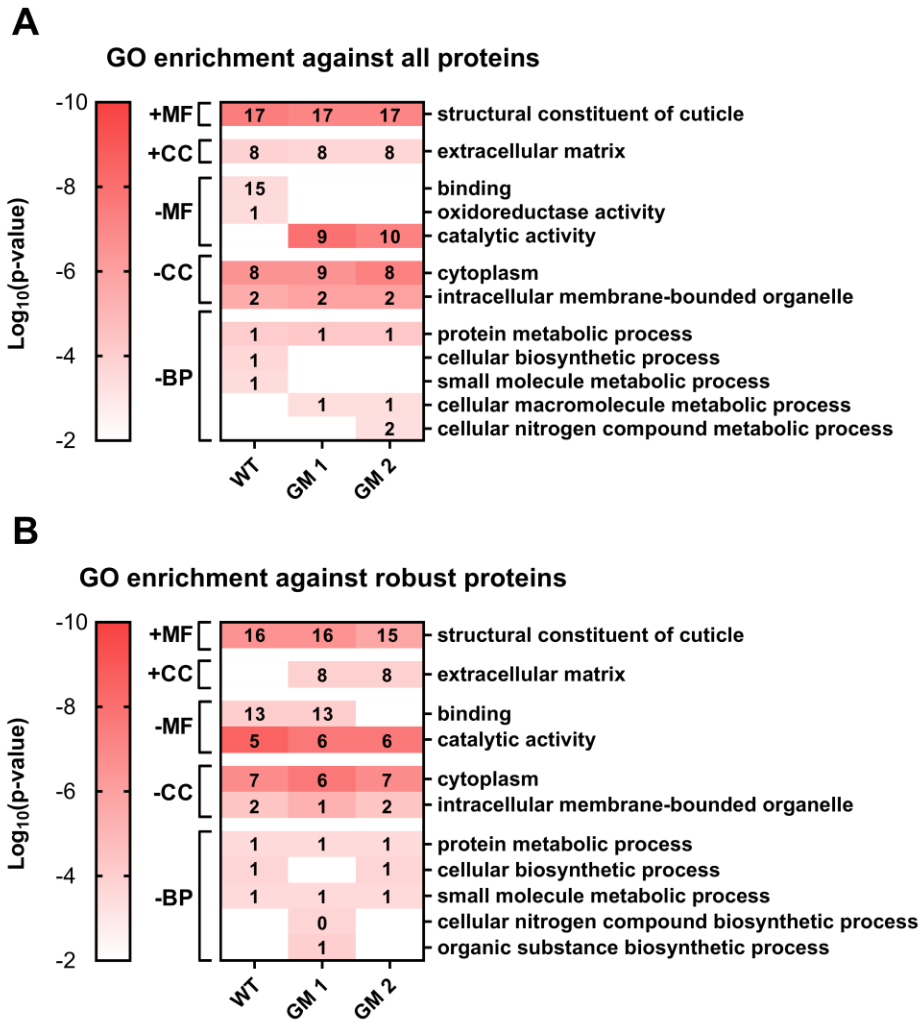


Figure 3.2 GO enrichment of samples against all proteins or robustly identified proteins

Summary of GO enrichment results of top 10% of proteins of wildtype and genetically modified *T. molitor* using (A) all quantifiable proteins or (B) all proteins robustly quantifiable across samples as a reference set. GO terms significantly enriched (FDR<0.05) and reduced to most specific terms (FDR<0.05). Cell color represents $\log_{10}$ (p-value) from a floor of p=0.01 and numbers per cell represent number of proteins with enriched term. MF: Molecular function, BP: Biological process, CC: Cellular component, +: Overrepresented term, -: Underrepresented term. WT: Wildtype *Tenebrio molitor*, GM: genetically modified *Tenebrio molitor*.

3.4.3 Predicted allergens of *T. molitor*

Quantitative distributions of predicted allergens did not differ compared to remaining non-predicted allergens of any of the samples (Figure 3.3); however, on average proteins predicted as allergens were significantly more abundant than non-predicted allergens for each of the three samples without significant differences across samples.

A majority of proteins identified were robustly identified among mealworm samples (81.1%; Figure 3.4). Every predicted allergen was robustly identified among mealworm samples but only two also statistically differed among samples: g5623t1 – arginine kinase and g4621t1 – myosin light chain alkali (Figure 3.5 A). These variable predicted allergens were exclusively significantly different in GM 2 with WT and GM 1 being not significantly different from each other. This significant difference transferred to the proportion of these two variable predicted allergens compared to the remaining 43 predicted allergens (Figure 3.5 B). With respect to the proportion of predicted allergens to non-predicted allergens there were no significant differences identified among samples.

Evaluation of the tropomyosins present found that the sequence present in AllergenOnline for *T. molitor* (g11661t1) is less identical to other known allergenic Arthropod tropomyosins than the sequence manually curated (ManualTM1.2.n; Appendix C Figure 3.1.4 and Appendix C Figure 3.1.5). While both sequences were relatively highly abundant, ManualTM1.2.n was approximately twice as abundant than g11661t1.

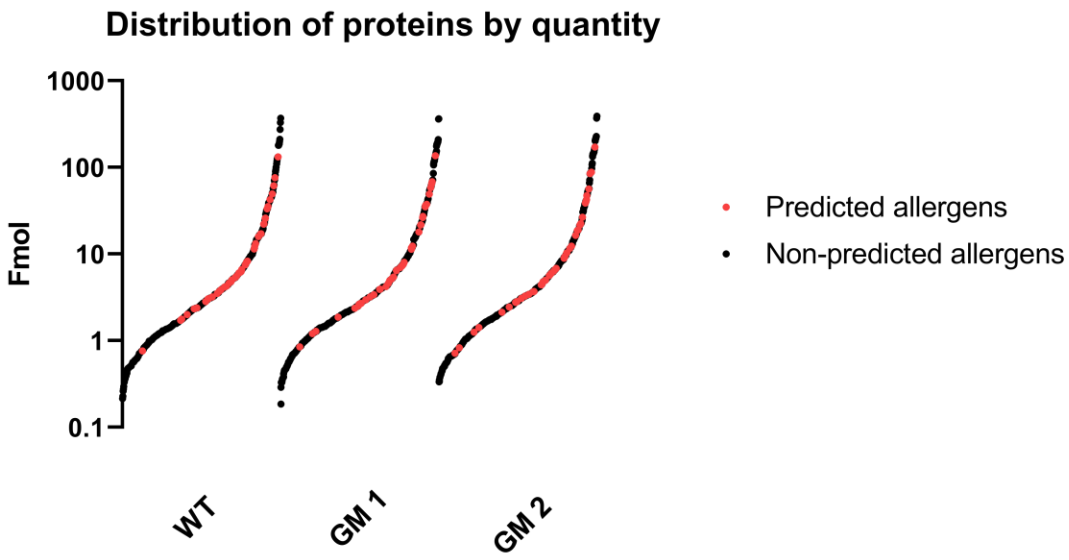


Figure 3.3 S-curves of quantifiable proteins

Quantifiable proteins per sample differentiated by predicted allergens (red) and non-predicted allergens (black). The x-axis is arbitrary and scaled. On average, predicted allergens were significantly more abundant than non-predicted allergens for each oof the three *T. molitor* samples by two-way ANOVA ($p < 0.05$) and two-stage linear step-up procedure of Benjamini, Krieger, and Yekutieli (FDR < 0.01). WT: Wildtype *Tenebrio molitor*, GM: genetically modified *Tenebrio molitor*.

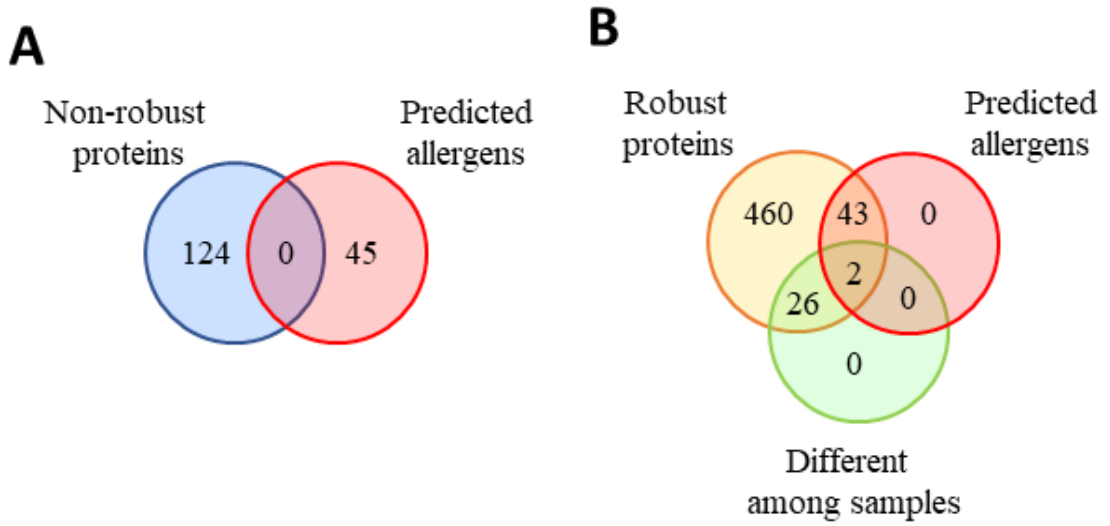


Figure 3.4 Venn diagram of protein assignments and annotations

(A) 124 proteins identified by mass spectrometry were not robustly identified across *T. molitor* samples with 45 proteins predicted as allergens and 0 not robustly identified and predicted as allergens. (B) Within 513 robustly identified proteins, 28 were significantly different among samples by two-way ANOVA ($p < 0.05$) and two-stage linear step-up procedure of Benjamini, Krieger, and Yekutieli (FDR < 0.01). 2 proteins were robustly identified, predicted as allergens, and significantly different among samples.

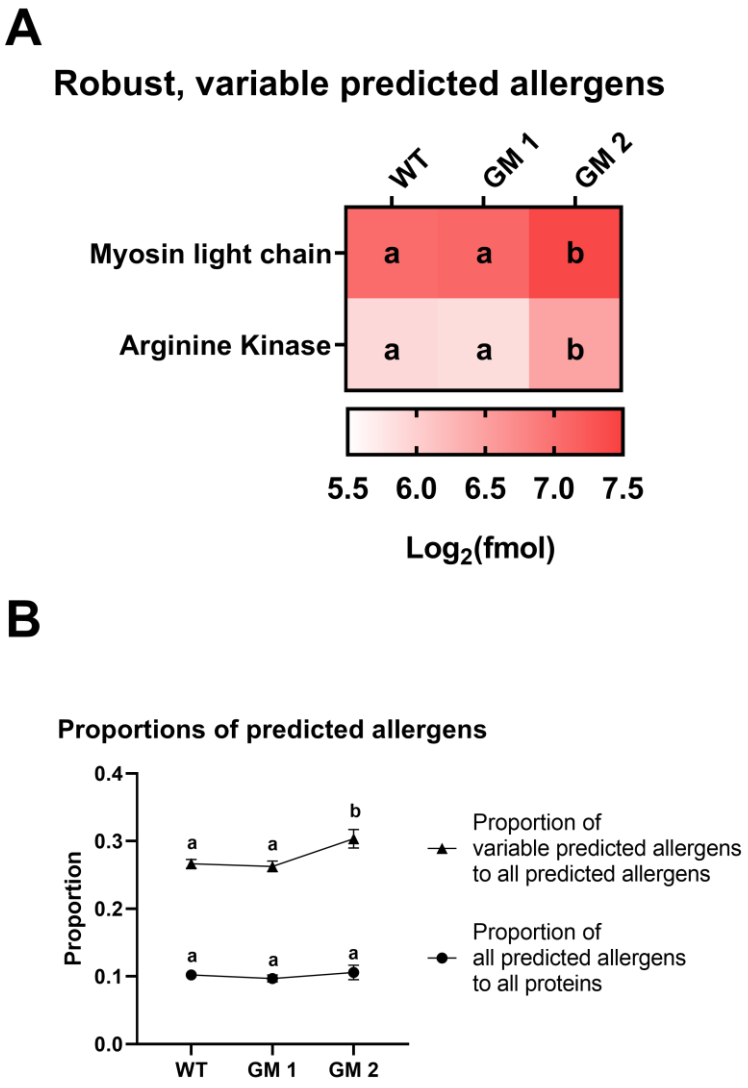


Figure 3.5 Selected predicted allergen heatmap and characteristics of predicted allergens across life stages

(A) Heatmap of predicted allergens arginine kinase (g5623t1) and myosin light chain (g4621t1) that were robustly identified in all *T. molitor* samples with significant differences among samples presented as Log_2 of the sum of each protein. (B) Proportions of quantified variable predicted allergens compared to all predicted allergens and the proportion of predicted allergens to all proteins. Values per protein or data point that do not share common letters are significantly different by two-way ANOVA ($p < 0.05$) and two-stage linear step-up procedure of Benjamini, Krieger, and Yekutieli (FDR < 0.01).

3.5 Discussion

This study was designed to examine the proteome of *T. molitor* yellow mealworm larvae, annotate proteins, predict the allergenicity of present proteins, and lastly evaluate the effects of genetic modification on two proprietary samples of mealworms. In effect, this study first evaluated *T. molitor* and the various proteins present to then compare wildtype and genetically modified variants to frame both what typical proteins and possible allergens are present to then judge if these genetically modified variants could pose a modified level of risk to consumer relative to the wildtype.

The use of Augustus for *ab initio* prediction was to both compare with the predictions by Eriksson *et al.* as well as incorporate further data into the prediction [311]. Predictions here was preformed including alternative predictions according to evidence provided as the wildtype mealworm raw data and partial gene models allowed as opposed to Eriksson *et al.* only incorporating partial gene models. An additional 32% of complete and incomplete gene models were predicted here compared to Eriksson *et al.* despite using the same prediction software and reference gene database; however, a newer version of Augustus was used here. Augustus is highly accurate compared to other *ab initio* gene prediction tools, although such tools are not yet perfect and still have difficulty with correctly identifying coding regions for proteins [314]. Notably, the sequence for tropomyosin-2 was predicted with numerous errors and inconsistent with that of *T. castaneum* and manually added.

Homologues to many known Arthropod allergens were identified particularly among shellfish, house dust mites, and cockroaches and were predicted as allergens in *T. molitor*. Barre *et al.* identified a number of *T. molitor* proteins that bound IgE of shrimp-

allergic patients including heat shock protein 70, α-amylase, arginine kinase, tropomyosin, apolipophorin-III, larval cuticle protein, and a 12 kDa hemolymph protein [309], where many of these proteins were predicted here as allergens excluding apolipophorin-III, larval cuticle protein, and the 12 kDa hemolymph protein. Both apolipophorin-III and the 12 kDa hemolymph protein were identified in part as g19557t1 and g5735t1, respectively; however, many proteins were identified and annotated as larval cuticle proteins and so it is unclear which protein was identified with IgE binding. Further, the total number of proteins identified by Barre *et al.* was just 106 whereas 655 were identified here with stringent sample-wise filtration requiring repeated identifications for acceptance, which makes it unclear if the proteins identified as IgE binding were in fact the only proteins present. They did similarly note the presence of both a tropomyosin-1 and a tropomyosin-2 as was identified here. Tropomyosins are known pan-allergens of Arthropoda [297], and have been implicated in cross-reactivity between mealworms and shellfish [310]. With respect to multiple tropomyosins present, it may be speculated that these are important for reactivity and cross-reactivity.

Functional annotation of *T. molitor* unigenes by Liu *et al.* provides insights into the GO enrichment based on protein abundance performed here [315]. Many *T. molitor* genes have been annotated principally with terms including binding, cellular process, and metabolic process, which were found to be underrepresented in the highly abundant proteins as opposed to extracellular matrix, which was only assigned to approximately 1% of genes by Liu *et al.* but was overrepresented in the highly abundant proteins. These exhibit an inverse relationship suggesting that a greater diversity of proteins are needed for binding, cellular processes, and metabolic processes that may discourage any

particular proteins from becoming highly abundant. These proteins may also have significant differences in protein stability and therefore protein turnover as highly stable proteins do not require high levels of transcription to become abundant [295].

Some differences were observed among wildtype and genetically modified variants. The observed differences in GO enrichment were likely associated with the limited number of samples and the arbitrary cutoff of the top 10% most abundant proteins. The relatively higher content of arginine kinase and myosin light chain content identified in GM2 would indicate that these pose a potentially higher risk to individuals with clinical reactivity to these proteins. GM1 did not show any significant differences in abundance compared to WT and therefore from this assessment would not be of greater risk to consumers. Both arginine kinase and myosin light chain are recognized as food and airway allergens depending on the source and therefore care should be observed either with respect to production or use of any of these mealworms.

Risk is present regardless of the trait effect of genetic modification and allergenic proteins were predicted to be present in all samples assessed, including wildtype *T. molitor*. The risk posed may be higher in GM2 mealworms, but a greater number of samples would assist in clarifying the magnitude of the difference in risk. We suggest that stakeholders be informed of the mealworms they are handling or using to allow for individuals to make informed choices as to the risks they are willing to face.

CHAPTER 4: PERSISTENCE OF PEANUT ALLERGEN-DERIVED PEPTIDES THROUGHOUT EXCESSIVE DRY THERMAL PROCESSING

4.1 Abstract

Scope: Commercial dry roasting of peanuts can result in accumulation of extensively heated peanut residue in ovens. These residues could pose a potential allergenic risk if residues are transferred to food products subsequently processed in the oven. Peanut residues can be detected in food products using antibody-based methods, but detection is greatly affected by thermal processing. We investigated the detectability of peanut allergens after excessive thermal processing using mass spectrometry (MS).

Methods and results: Peanut kernel halves were roasted in a muffle furnace, ground, robustly extracted, and probed by immunoblot with sera from patients with peanut allergy (anti-peanut IgE) and anti-peanut IgG. Extracts were further analyzed by MS. After 8 hours at 176 °C, detectable allergenic protein content, by MS, decreased 54.6-fold, with surviving contiguous regions sufficient to bind IgE to Ara h 3.

Conclusions: Failure to detect peanut residues with antibody-based methods should not be regarded as an indication of the absence of peanut residues as robust extractions coupled with MS were able to identify peanut from some of these processed samples. Peanut residues should be physically removed prior to thermal processing as reliable existing antibody-based methods are inadequate to detect peanut residues after extensive heating of products.

Keywords: Food allergen, Mass spectrometry, Peanut, Thermal processing

4.2 Introduction

Food allergies affect approximately 10.8% of US adults and peanut allergy affects approximately 1.8% [81]. Symptoms of food allergies range from rashes, hives, and anaphylaxis where harmful doses of allergenic protein can be as low as sub milligram quantities [316]. Representative quantification of present allergenic residues is required to support allergen management, which is required to adequately inform allergic consumers of the potential risk via labeling.

Residues of commercially dry roasted peanuts may accumulate in ovens, possibly resulting in harmful levels of exposure to peanut-allergic consumers. Thermal processing can hinder or even negate the detection of peanut residues when using enzyme-linked immunosorbent assays (ELISA) [209, 317]. Peanuts are commonly oven roasted at 160-180 °C, but higher temperatures are used for cleaning. Detection of peanut residues in thermally processed material may be compromised by decreased solubility of peanut protein especially given that extraction buffers must be compatible with antibody binding. Thermal processing may also result in aggregation of, and chemical or structural modifications to, proteins that can further complicate accurate detection and quantification [318]. Thermal processing affects both the detectability of allergens and their allergenicity to variable extents dependent upon the type of thermal processing [319]. Inaccurate assessment of allergenic residues following thermal processing presents a challenge to industry and could pose a risk to peanut-allergic consumers.

This study addresses the identification of peanut allergens persisting after up to 8 hours of dry thermal processing. The quantity of allergen present at each stage represents the potential risk and is addressed with antibody binding and quantitative mass

spectrometry (MS). The use of MS allows for the detection of many targets (peptides and modifications thereof) derived from peanut allergens. This contrasts to ELISA methods where only one measure of peanut presence is determined. We further describe the effects of thermal processing by examining the survival of potentially contiguous regions of allergens using MS. The aim was to determine if peanut residues with the capability to bind to peanut-specific antibodies would remain after excessive dry thermal processing.

4.3 Materials and methods

4.3.1 Chemicals

Acetone, hexane, methanol, sodium chloride, disodium phosphate, monopotassium phosphate, formic acid, water, ACN were obtained from Fisher (Hampton, NH, USA). Mass spectrometry solvents were Optima grade. Thiourea, CHAPS, Tris, and Iodoacetamide were obtained from Sigma Aldrich (St. Louis, MO, USA). DTT was obtained from Acros (Geel, Belgium). BSA and PVDF were obtained from Millipore (Burlington, MA, USA). Nupage 4-12% Bis-Tris SDS-PAGE gels, NuPage MES SDS running buffer, and LDS sample buffer were obtained from Invitrogen (Carlsbad, CA, USA). Urea, Tween 20, Coomassie Blue R-250, Precision Plus Protein™ Dual Xtra standards, and 10x Tris Glycine buffer were obtained from Biorad (Hercules, CA, USA). Potassium chloride was obtained from Labchem (Zelienople, PA, USA). Pierce C18 spin columns, Pierce MS grade Trypsin, and SuperSignal West Dura Extended Release Substrate were obtained from Thermo Scientific (Waltham, MA, USA). MassPREP Phosphorylase b Standard was obtained from Waters (Milford, MA, USA). Milli-Q purified water (18.2 MΩ cm) was used throughout this study.

4.3.2 Thermal Processing

Raw, blanched (deskinned) peanuts were obtained from the Golden Peanut Company (Alpharetta, GA, USA) and stored at 4 °C until use. Groups of 10 peanut kernel halves were placed into foil-lined crucibles. Crucibles were heated inside a Thermolyne muffle furnace (Model F30420C, Thermo Scientific, Waltham, MA, USA) preheated to either 176 °C or 260 °C. Crucibles were removed at 15, 30, 60, 120, 240, and 480 minutes. Non-heated controls were also prepared and appear as time=0 in results. Five replicates per time point and temperature combination were prepared.

4.3.3 Protein Extraction

Per replicate, pairs of peanut kernel halves were placed into a 2 mL microfuge tube and manually homogenized by metal spatula and glass rod in the presence of 500 µL of acetone. An additional 1 mL of acetone was added, samples vortexed, rocked for 15 minutes, centrifuged at 17,000*g* for 15 minutes, and decanted. This was repeated three more times. The samples were then defatted with 1 mL of hexane, a total of four times as described for the acetone procedure. Samples were dried overnight at room temperature, vortexed to produce a homogenous powder, and then stored at -20 °C.

Samples were extracted as modified from [285, 320]. In brief, 50 mg of each defatted sample was extracted in 1 mL of zwitterionic chaotropic buffer (50 mmol/L Tris, pH 8.8 containing 50 mmol/L 1,4-Dithio-D-threitol (DTT), 5 mol/L urea, 2 mol/L thiourea, and 32.5 mmol/L 3-[(3-Cholamidopropyl)dimethylammonio]-1-Propanesulfonate (CHAPS)). Suspensions were vortexed, incubated in a heated

sonicating water bath (60 °C, 20 min), and extracts were clarified by centrifugation (17,000*g*, 20 min). The supernatants (500 µL) were taken and an equal amount of extraction buffer replaced. Each sample was sequentially extracted three times and pooled as a single replicate. Extracts were stored at -20 °C until further analysis.

4.3.4 Mass Spectrometry and data analysis

MS was performed as previously described [285] with modifications. Briefly, 1 µL of extracts were precipitated with excess acetone and the pellet was reduced, alkylated, digested using trypsin, cleaned via C18 Spin columns, vacuum evaporated, and resolubilized with 28 µL of 0.96 mol/L acetonitrile (ACN), 26.5 mmol/L formic acid. Prior to injection, 1.5 µL of 200 fmol/µL rabbit glycogen phosphorylase B was added to 9 µL of each sample and 3.5 µL were injected into Thermo Q Exactive Plus™ Hybrid Quadrupole-Orbitrap™MS coupled to UltiMate 3000RSL® liquid chromatography (UPLC) system (Thermo Scientific™). Each sample was injected twice as technical replicates.

Data was analyzed using PEAKS 8.5 (Bioinformatics Solutions Inc., Waterloo, ON, Canada) using a peanut allergen database was generated by Marsh *et al*. including rabbit glycogen phosphorylase b (Uniprot.org P00489.3) and the common Repository of Adventitious Proteins contaminant database (thegpm.org/crap; obtained 190121) [209, 286]. PEAKS search parameters were 5 ppm (mg/kg) mass error tolerance and 0.05 Da fragment mass error tolerance with no modifications of amino acids. Secondary PEAKS PTM search was performed with 3 missed cleavages per peptide and performed with fixed carbamidomethyl cysteine and 3 variable modifications against a base list of 313

modifications with 31 additional Maillard reaction and lipid peroxidation modifications representing 46 different potential modifications as well as a modification reflecting acetone aldol condensation (Appendix E Table 4.1.1) [321-324]. The base list was curated to remove overlap with the added modifications for a total of 337 modifications. Peptides identifications with both <1% FDR, and each PTM identified per peptide had an ambiguity score of 20 or greater ($p < 0.01$) were accepted. Peptides with potentially artefactual modifications, not limited to ion substitutions, carbamylation, or in-source decay, were disregarded from analysis.

For peptide level data, peptides were filtered by requiring presence in both technical replicates of at least two biological replicates within a single time point. For peptide persistence, allergen-derived peptides were also filtered to only evaluate peptides that met requirements for inclusion for two time points. Contiguous peptides were evaluated per allergenic sequence and reported as the maximum contiguous length per allergenic family (e.g. Ara h 1) across all replicates. The longest contiguous peptide was determined using three methods: 1) only observed peptides, 2) extending observed peptides by the first contiguous tryptic peptide under 6 amino acids, 3) extending observed peptides by all contiguous tryptic peptides under 6 amino acids. PTMs of interest were curated relative to robust parent unmodified peptides. Peptides including a missed tryptic cleavage were included as unmodified. PTMs of interest included deamidation, pyroglutamate, and Maillard products. Peptides were grouped according to modification of interest and evaluated for trends.

Protein quantification evaluated two lists of peptides with either a focus on robustness in raw peanuts or robustness to dry thermal processing (Appendix E Table

4.1.2). Only Ara h 1, 2, 3, and 6 were investigated as the remaining allergens were not robust through dry thermal processing, were not highly abundant in raw samples, and/or prohibitively diverse. Proteins were regarded as present if both technical replicates had two peptides observed, quantified by averaging up to the top 3 peptides, and the quantities were averaged across replicates.

Data were normalized via the top three predominantly present and abundant peptides from rabbit glycogen phosphorylase. Statistical analyses were performed using GraphPad Prism version 8.3.0 for Windows (GraphPad Software, La Jolla California USA) via two-way ANOVA with significant p-values at 0.05 and multiple comparisons via two-stage linear step-up method of Benjamini, Krieger, and Yekutieli with an FDR set at 0.05. Non-detect values were treated as zeroes.

4.3.5 SDS-PAGE, sera, and immunoblotting

Protein contents of raw peanut extracts were determined using a 2D Quant-Kit™ per manufacturer's instructions (GE Healthcare Life Sciences, Chicago, IL, USA). Volumetric aliquots of samples equivalent to 30 µg of raw peanut protein were separated by SDS-PAGE and visualized with Coomassie staining.

Immunoblotting was performed as described elsewhere with the immunoblot transfer performed at 100 V for 60 minutes and dot blots were produced by applying extracts to activated polyvinylidene difluoride (PVDF) membrane and dried prior to treating them similarly [325]. Enzyme conjugated secondary anti-peanut antibody was obtained from a Morinaga Peanut ELISA Kit (Cat# M2104, Yokohama, Japan). Pooled sera from six peanut-allergic individuals were used. All individuals have positive skin

prick tests to peanut and compelling histories of allergic reactions upon peanut ingestion including laryngeal edema, angioedema of face/eyes, urticaria, vomiting, diarrhea, hypotension, and respiratory symptoms. All individuals had positive Immunocap scores ranging from 3.13-85.5 kIU L^{-1}. Each individual provided informed consent as stipulated by The University of Nebraska-Lincoln IRB Approval 200305289FB, Project ID 6029. Dot blots included positive control Morinaga enzyme conjugated antibody and negative control bovine serum albumin that had been heated for 480 minutes at 260 °C (Extracted as in 4.3.3). A solution of 1 g of non-fat dry milk in 40 mL phosphate buffered saline, 450 µM tween 20 (PBST) was used to dilute Morinaga antibody (1:20), human sera (1:10), and secondary mouse anti-human IgE Fc-HRP (1:1,000; Southern Biotech, Birmingham, AL, USA). Immunoblots were visualized by chemiluminescence using SuperSignal West Dura Extended Release Substrate and detected using a UVP Biospectrum 815 Imaging System (Analytik Jena US LLC, Upland, CA, USA) with VisionWorks software (Version 8.20, Analytik Jena US LLC). Maximum exposure time was 5 minutes and images taken were inverted with intensity range set to auto.

4.4 Results

4.4.1 Observable Properties of Thermally Processed Peanuts

Peanuts roasted at 176 °C developed off-odors and were visibly darkened after 60 minutes (Appendix E Figure 4.1.1). At 120 minutes and beyond, the peanuts were blackened and smelled burnt. Peanuts roasted at 260 °C for 15 minutes had the same blackened appearance as peanuts roasted at 176 °C for 2-8 hours; however, the burnt odor was noxious. After 30 minutes, the peanuts were smoking and produced an oily resin.

Defatted powders prepared from the roasted peanut halves had colors matching the peanut halves (Appendix E Figure 4.1.2). Extracts of the peanuts roasted at 176 °C reflected the same color progression as the powders but extracts of the peanuts roasted at 260 °C darkened through 120 minutes before lightening for the 240 and 480-minute time points (Appendix E Figure 4.1.3).

4.4.2 Mass Spectrometry

4.4.2.1 Quantification of Peanut Allergens over Time

Peanut allergens were quantified in extracts from peanuts roasted for the 176 °C time series but not for the 260 °C time series as data was insufficient to be quantified. Quantification of the major peanut allergens (Ara h 1, 2, 3, and 6) after roasting at 176 °C is shown in Table 4.1. The variation observed reflects differences among peanuts and the effects of the entire workflow. The most abundant allergens present in the extracts was Ara h 3, followed by Ara h 1, 2, and 6. Significant decreases in the detected quantity of all allergens were observed during roasting at 176°C. Ara h 1 and 6 decreased consistently, whereas Ara h 2 and 3 showed a small, statistically significant, increase at 15 minutes followed by consistent decrease. After 120 minutes, further decreases in detectable allergen content were not statistically significant, excluding Ara h 1. Ara h 3 detection was least susceptible to heating, followed by Ara h 2, 6 then 1. The relative composition of detectable proteins therefore changes over time, with Ara h 3 accounting for 59% of quantified allergens at 0 min, and 87% of quantified allergens at 480 minutes (Appendix E Figure 4.1.4).

Table 4.1 MS quantification and summary of peanut allergens after roasting at 176 °C

	Time (min)							Fold Loss
	0	15	30	60	120	240	480	0 vs 480
Ara h 1	1775.4 (54.2)[a]	2061.0 (52.5)[a]	1572.0 (39.8)[a]	359.6 (57.7)[b]	38.5 (118.3)[cd]	11.7 (65.7)[c]	4.4 (97.6)[d]	405.49
Ara h 2	2065.0 (37.4)[ab]	2523.7 (35.1)[a]	1891.5 (35.8)[b]	1167.2 (48.0)[b]	214.3 (110.5)[c]	56.4 (122.2)[c]	19.9 (148.8)[c]	103.69
Ara h 3	8550.1 (36.1)[ab]	10859.9 (41.3)[a]	8295.9 (24.6)[ab]	5842.6 (36.8)[b]	980.2 (82.3)[c]	516.5 (65.1)[c]	232.7 (71.5)[c]	36.74
Ara h 6	2081.6 (9.2)[a]	2063.5 (58.6)[ab]	1329.4 (52.7)[ab]	903.2 (30.4)[b]	103.5 (135.3)[c]	32.0 (123.0)[c]	8.1 (150.9)[c]	258.31
Sum of 1, 2, 3, 6	14472.0 (32.9)[ab]	17508.0 (40.0)[a]	13088.8 (27.0)[ab]	8272.6 (36.4)[b]	1336.6 (91.3)[c]	616.6 (71.3)[c]	265.1 (68.9)[c]	54.60

Allergen quantifications are reported in fmol as mean (CV%) of five biological replicates with different superscript letters denoting significantly different individual allergen quantities over time via two-way ANOVA ($p<0.05$) and two-stage linear step-up method of Benjamini, Krieger, and Yekutieli (FDR<0.05).

4.4.2.2 Identification of Thermally Induced PTMs and Processing Biomarkers

We examined the presence of PTMs in all time points of peanuts heated at both temperatures using label-free quantitation to evaluate chemical changes in peptides as a result of thermal processing (see Appendix F complete peptide list). At the peptide level per allergen, unmodified and modified peptides followed similar decreasing trends with a large range of values per time (Appendix E Figure 4.1.5). Focused assessment of PTMs used exemplar, robust peptides with high-quality identification of unmodified peptide as well as deamidation, pyroglutamate and Maillard products to assess prevalence of these modifications during heating (Figure 4.1). Heating at 176 °C dataset utilized the peptides [QQPEENACQFQR], [IESEGGYIETWNPNNQEFQCAGVALSR], and [FNLAGNHEQEFLR], all derived from Ara h 3. No equivalent, robust peptides were found for other allergenic proteins. Deamidation was observed in raw samples with

further increases occurring with roasting until 120 minutes when a plateau was reached. Pyroglutamate products were also present prior to processing with a large increase occurring after 60 minutes of roasting and increasing thereafter. Maillard products were not observed in raw peanuts and showed gradual increases before peaking at 120 minutes of roasting but then decreased afterward. Interestingly a deamidated Ara h 3 peptide [FNLAGN(+0.98)HEQEFLR] was found to be abundant, robust toward heating, and present in all samples (including unheated peanut) (data not shown). Further, BLAST search of the peptide with substituted Asn (N) to Asp (D) yielded no hits, indicating that the deamidation is the likely source of the peptide. Heating at 260 °C data set resulted in only one peptide detection. The pyroglutamate modified, non-tryptic peptide [E(-18.01)LQEGHVLVVPQ] was only detected at 15 minutes. It is possible that this peptide resulted from β-elimination induced by heating, which did not occur after roasting at 176ºC.

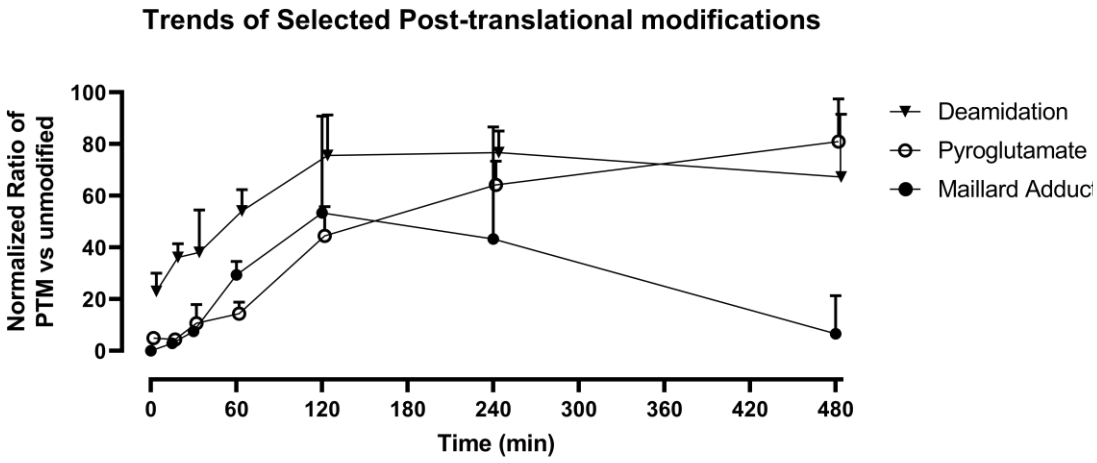


Figure 4.1 Trends of selected post-translational modifications in peanuts heated at 176 °C.

Quantity of deamidation, pyroglutamate, and Maillard adduct (sum of the abundance of all modified peptides to the abundance of unmodified peptide) for selected peptides (n=5).

4.4.2.3 Evaluation of Maximally Contiguous Peanut Peptides

The maximum length of contiguous peanut peptides from extracts of peanuts roasted at 176 °C was assessed based on three scenarios to account for bias in the MS data (Table 4.2). Across the three scenarios, decreasing maximum peptide lengths were noted over time predominantly occurring at 120 minutes and later. Ara h 2 and 6 were both consistently present and decreased at 120 minutes whereas Ara h 1 and 3 had a more gradual decrease over time.

Table 4.2 Maximum length of contiguous peptides of peanut allergens after roasting for 176 °C

	Contiguous Peptide Strategy	Time (min) 0	15	30	60	120	240	480
Ara h 1	1	87	147	146	85	61	26	25
	2	247	269	179	179	98	44	49
	3	314	360	288	253	103	49	49
Ara h 2	1	41	70	57	70	68	9	9
	2	70	70	69	70	68	12	16
	3	70	70	70	70	69	15	17
Ara h 3	1	138	155	155	155	138	99	69
	2	188	253	215	209	163	108	108
	3	509	509	506	509	319	113	113
Ara h 6	1	67	67	67	67	34	29	14
	2	67	67	67	67	34	53	28
	3	67	67	67	67	67	67	28

The maximum contiguous length of peptides (number of residues) was calculated using three methodologies: 1) only observed peptides, 2) extending observed peptides by the first contiguous tryptic peptide under 6 amino acids, 3) extending observed peptides by all contiguous tryptic peptides under 6 amino acids.

4.4.3 Antibody Binding to Thermally Processed Peanut

As shown in Figure 4.2, proteins in extracts of peanuts roasted at 176 °C showed evidence of aggregation after 15-60 minutes of roasting. The aggregated proteins were able to bind both IgE and IgG antibodies, with a gradual decrease in band intensity across the time points resulting in very little visible protein by 120 minutes. Extracts of peanuts roasted at 260 °C had no clear protein bands present in SDS-PAGE or immunoblotting, excluding smearing of lanes. Dot blotting of extracts of peanuts roasted at 176 °C showed antibody binding of the anti-peanut IgG through 60 minutes and allergic sera IgE through 120 minutes (Figure 4.3). No non-specific binding was observed to peanut proteins on negative control blots (Appendix E Figure 4.1.6).

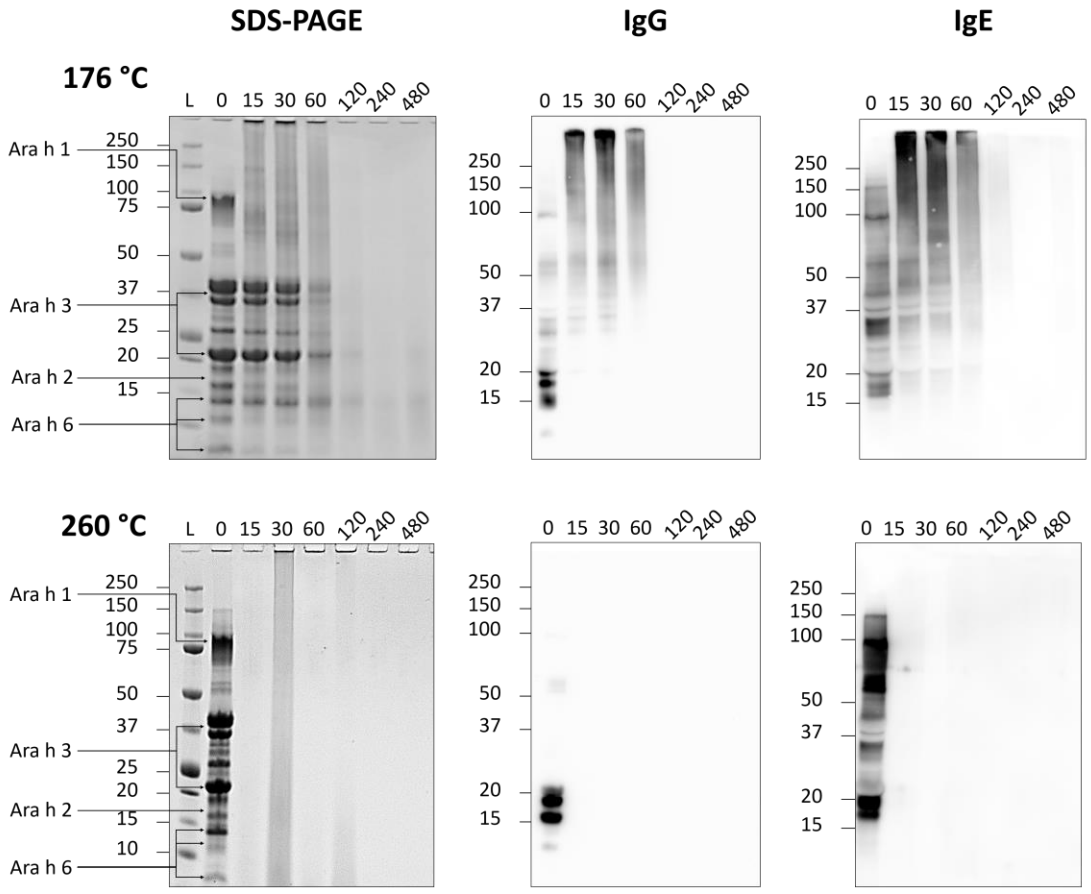


Figure 4.2 SDS-PAGE and immunoblots

SDS-PAGE (left), Morinaga anti-peanut IgG (center), and pooled peanut allergic sera IgE (right) separated according to the 176 °C (top) and the 260 °C (bottom) roasting times. Lanes were loaded volumetrically relative to 30 µg of raw peanut (approximately 1.7 µL). Molecular weight of protein standards (kDa) indicated to left of each panel.

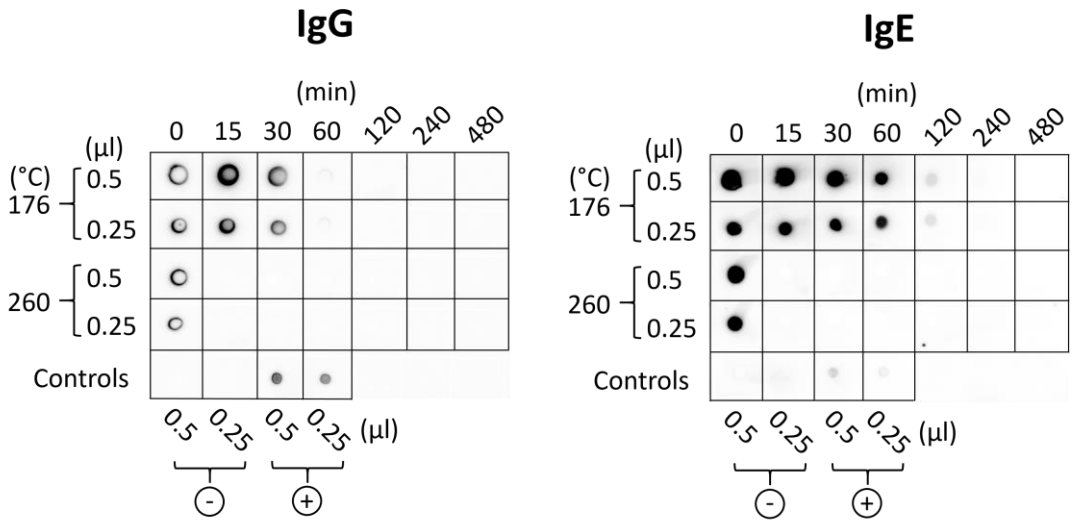


Figure 4.3 Dot blots

Dot blots probed with either Morinaga anti-peanut IgG (left) or pooled peanut allergic sera IgE (right). Dot blots were prepared by applying either 0.5 or 0.25 µL of extracts of peanuts heated at 176 or 260 °C. Negative control was bovine serum albumin heated at 260 °C for 8 hours and extracted while positive control was directly applied Morinaga anti-peanut IgG.

4.5 Discussion

This study was designed to assess the persistence of peanut allergens after excessive dry thermal processing, evaluate the fate of the individual major peanut allergens, and examine the possibility for detection of peanut after extensive heating. From a practical perspective, would peanut allergens survive in dry baking ovens for extended time periods? And could peanut allergens potentially be removed by using high oven burnout temperatures over extended time periods?

Clear differences were observed in identifiable proteins present after roasting at 176 °C and 260 °C. Work by Fu *et al.* using commercial ELISA kits found that assays of thermally processed peanut flours were hindered by decreased solubility and immunoreactivity of target proteins [317]. A key difference between this study and that of Fu *et al.* is that here we used a more robust, sequential extraction incorporating zwitterionic chaotropic buffer while the previous study was limited to an ELISA-compatible heated PBS extraction. Even with more complete extraction, and multi-analyte detection of proteins (with MS), quantitative detection of extensively heated peanut allergens was not possible.

IgE binding of extracted peanut allergens was observed extending to 120 minutes for peanuts roasted at 176 °C. Regardless of which bioinformatics method is used to assess the presence of contiguous peptides, Ara h 1, 2, 3 and 6 peptides of sufficient length to elicit IgE reactivity likely survive. Although estimates of minimum required peptide length to elicit IgE reactivity vary, Huby *et al.* estimated that 30 residues were sufficient for cross-linking [326]. After 120 minutes, IgE binding was not apparent. Interestingly, this decrease corresponds with a conspicuous decrease in longest contiguous peptide length, particularly with respect to Ara h 2. The ability to demonstrate detectable peptide length after proteolysis may be a useful predictor of potential IgE binding. It should be noted that IgE binding is not a biological measure of allergen reactivity. Much additional work would be required to establish the existence of a useful correlation between detectable contiguous peptides and clinically relevant reactivity in different scenarios.

Quantification of present residues demonstrated that robustly occurring peptides were present in extracts of peanuts roasted at 176 °C, but not in extracts of peanuts roasted at 260 °C. These peptides represented each of the major peanut allergens including Ara h 1 and 3, which are regarded as becoming relatively less detectable after thermal processing due to their tendency to denature and aggregate [327]. However, even with chaotropic extraction conditions, Ara h 1 detection was most sensitive to thermal processing, as has been observed previously [209]. The robust extraction method used here likely contributes to detectable peptides leading to Ara h 3 acting as an abundant and robust target. Further, in contrast to previous studies, we find that Ara h 1, 2, and 6 followed similar decreasing trends over time [328]. Continuous thermal processing can break disulfide bonds and denature tightly packed albumins [329, 330]. In effect, this may cause these allergens to extract similarly. Therefore, Ara h 3 represents an increasing proportion of detectable protein after thermal processing.

IgE binding and peptide identifications decrease with increasing thermal processing times and temperatures, but it is unclear what happens to the proteins and peptides as they undergo thermal processing. Modifications are a potential reason for loss of identifiable allergen peptides. Deamidation and pyroglutamate modifications may function as processing biomarkers for evaluating the degree of processing [331]. Other potential modifications including crosslinking, Maillard, and peroxidative adducts make detection of residues more difficult as these products progressively react, degrade, and differentiate. Nonenzymatic cleavage was observed in conjunction with pyroglutamate-containing peptides commonly following Gln, Glu, Asn, and Asp residues, which are known to cleave C-terminally [332, 333]. The only robust peptide identified from extracts

of peanuts roasted at 260 °C was doubly non-tryptic flanked by Glu and Gln and this peptide was not observed in extracts of peanuts roasted at 176 °C.

Allergenic peanut proteins clearly persist, even after extended heating, at typical roasting (176 °C) conditions. Interestingly, even short periods of higher temperature (260 °C) are dramatically more effective at diminishing detectable allergen-derived peptides. The robust, chaotropic, repeated extractions used in this study, combined with the inherent capability of MS to detect many, diverse analytes, represents likely the most comprehensive method for detection of peanut allergens in heated samples. However, the failure of detection should not be regarded as absence of allergenic risk. We suggest that biological or clinical models may be used to assess the reactivity of heavily thermally processed allergens.

Current (e.g. ELISA) or even specifically designed MS detection methods are not capable of reliably detecting heavily heated peanut residues. This is unlikely to change using currently available technology. Care should therefore be taken in interpreting allergen analysis after the possible introduction of heavily heated peanut material. We suggest that the risk of product contamination by oven residue be controlled by cleaning, as high temperature heating may be insufficient.

CHAPTER 5: DETECTION OF FOOD ALLERGEN-DERIVED PEPTIDES FROM EXTRACTIVE-BASED E-CIGARETTE LIQUIDS

5.1 Abstract

Background: E-cigarette liquids may be flavored using either synthetic flavors, or those derived from plant material by extraction. Plant materials used for flavoring include many known and potent allergens. Although proteins are not significant flavor contributors to these liquids, the use of plant-derived flavors raises the possibility of presence of protein allergens in the resultant flavors and therefore in the inhaled vapors. The presence of such allergens may pose a risk of *de novo* sensitization or eliciting existing allergies.

We investigated the possible presence of proteins in flavored vape fluids using SDS-PAGE and protein mass spectrometry. Further, we estimate the risk that inhalation of the observed levels of allergenic protein may pose to consumers.

Methods: Extractive-based E-liquids were purchased from online vendors based on potential presence of plant-derived flavors. We selected apple, cherry, peach, almond, coconut, hazelnut, pecan, and peanut brittle flavored liquids. Protein from E-liquids were dialyzed, precipitated, and analyzed by SDS-PAGE with subsequent in-gel digestion and LC-MS/MS, and by in-solution digestion and LC-MS/MS. Proteins were identified by comparison to relevant sequence databases, and relative quantitation performed using label-free quantitation.

Results: SDS-PAGE analysis showed the presence of a single protein in almond, peach, and coconut E-liquids. This protein was identified by LC-MS/MS as the basic subunit of 11S globulin (Pru du 6 in almond) in each case. Comparison of the observed allergen concentration with VITAL®3.0 reference dose for almond of 0.1 mg of almond protein (based on the reference dose established for hazelnut), and using conservative but reasonable assumptions on dosing, suggests little risk in an ingestion scenario. However, estimation of inhalation risk is more difficult. We describe potential routes for estimating risk of inhalation of food allergens. Based on these assessments it is likely that some E-liquids pose a credible risk to food allergic individuals.

Conclusions: Some plant extract-based E-liquids likely pose a risk to allergic individuals. We emphasize the difficulty in estimating this risk due to the different route of exposure. We suggest that inhalation of proteins in E-liquids may also play a role in sensitization. We suggest that such liquids should be clearly labelled to mitigate risk posed to sensitized individuals.

5.2 Introduction

Electronic nicotine delivery systems (ENDS) are a common alternative to cigarettes with an increasing proportion of young adults in the U.S. using ENDS with rates between 6-9% [334], and of these a positive association between use of non-tobacco and non-menthol flavors has been found with higher frequency and amounts consumed [335]. A mixture of nicotine, propylene glycol, vegetable glycerin, flavorings, and other ingredients, herein referred to as an E-liquid, are loaded into ENDS to be aerosolized and

inhaled. Flavorings included in E-liquids can be synthetic or extractives, where extractive-based E-liquids may include extracts of allergenic foods.

Food allergies affect approximately 10.8% of U.S. adults with symptoms ranging from rhinitis to anaphylaxis [81]. The doses of protein required to elicit allergic reactions can be as low as sub-milligram quantities [316]. Allergenic sources vary greatly in the amounts of inhaled allergens to result in allergic disease [336], but inhaled allergic responses tend to be less severe than responses to ingested allergens [337]. However, aerosolized proteins pose a risk of sensitization through inhalation [338]. As yet there are no cases in the literature reporting allergic reactions associated with ENDS use attributed to proteinaceous material in E-liquid [339]. Allergic consumers of extractive-based E-liquids may be at risk for eliciting allergic reactions and both allergic and non-allergic consumers may be at risk of *de novo* sensitization.

This study addresses the identification of food allergens present in extractive-based E-liquids. The quantity of allergenic proteins represents the potential risk and is examined using quantitative mass spectrometry (MS). Use of MS allows for broad identification of allergen-derived peptides without the need for specific antibodies that may not be specific to the proteins present in the extractive-based E-liquids. The aim was to determine if allergenic protein residues were present in extractive-based E-liquids and evaluate the risks posed to allergic and non-allergic consumers.

5.3 Methods and Materials

5.3.1 Samples and Chemicals

Tris, iodoacetamide (IAA), and rabbit glycogen phosphorylase B (GP; P6635; Uniprot P00489.3) were obtained from Sigma (St. Louis, MO, USA). Trypsin (Pierce, MS grade), 3.5 kDa molecular weight cut-off dialysis tubing (Snakeskin), and tris(2-carboxyethyl)phosphine were obtained from Thermo Scientific (Waltham, MA, USA). Dithiothreitol (DTT) was obtained from Arcos Organics (Geel, Belgium). Ammonium bicarbonate (ABIC) was obtained from Honeywell Fluka (Charlotte, NC, USA). Acetone, methanol, acetic acid, acetonitrile (ACN), and formic acid were obtained from Fisher (Hampton, NH, USA). Precision plus Protein^{TM} Dual Xtra standards and Coomassie R250 were obtained from Biorad (Hercules, CA, USA). NuPage LDS sample buffer, Nupage MES SDS running buffer (20x), and 4-12% NuPage Bis-Tris gels were obtained from Invitrogen (Carlsbad, CA, USA).

E-liquids were obtained based on either “may contains” labeling or marketed as an extractive of an allergenic food and purchased without nicotine or propylene glycol to simplify sample processing. The E-liquid flavors were labeled as green apple (apple), cherry, peach, sweet almond (almond), coconut, hazelnut, pecan, and peanut brittle (peanut). Of the E-liquids, peanut brittle was purchased from Vape Organics (Riverside, CA, USA) and the remainder were purchased from Velvet Vapors (Tucson, AZ, USA). The density of each liquid was measured three times and the mean reported.

5.3.2 Sample preparation

For each E-liquid, 2 mL were dialyzed against 1 L of 50 mM Tris, pH 8 with three changes of buffer. Dialyzed E-liquids were divided in half and up to 3 mL per sample were used for acetone precipitation. Samples were precipitated by adding 4 volumes of acetone chilled at -20 °C and precipitated at -80 °C overnight. Samples were centrifuged at 17,000 x g for 30 minutes at 4 °C and the pellet washed with 80% acetone twice.

5.3.3 SDS-PAGE

Dried precipitated samples were prepared with 40 µL of SDS-PAGE sample buffer with 50 mM DTT and 20 µL of each sample was fractionated. SDS-PAGE gels were fixed with a solution of 50% methanol and 10% acetic acid for 1 hour and visualized with Coomassie R250. Images were taken using a Kodak Gel Logic 440 Image station (Kodak, Rochester, NY).

5.3.4 Mass Spectrometry

5.3.4.1 In-gel digestion

Protein bands were removed from the SDS-PAGE gel using a clean razor and slices were fixed with a solution of 50% methanol and 10% acetic acid for 2 hours. The gel slices were destained with 200 µL of destaining solution (50% ACN, 25 mM ABIC) and incubated at 37 °C for 30 minutes while shaking, the destain removed, and the wash repeated. Proteins were reduced with 30 µL of 50 mM tris(2-carboxyethyl)phosphine in 25 mM ABIC and incubating at 60 °C for 10 minutes. Proteins were alkylated with 30 µL

of 100 mM IAA in 25 mM ABIC for 1 hour in the dark. Gel slices were destained twice before shrinking with 50 µL of 100% ACN for 15 minutes and air drying for 10 minutes. Proteins were digested by adding 10 µL of 10 ng μL^{-1} trypsin in 25 mM ABIC and 15 minutes later adding 25 µL of 25 mM ABIC before incubating at 37 °C overnight. The slices were votexed, 30 µL of supernatant removed, and 1.6 µL of 100% ACN added.

5.3.4.2 In-solution digestion

In-solution digestion of acetone precipitated pellets was preformed according to Palmer *et al.* [285]. Briefly, pelleted proteins were reduced with DTT for 5 minutes at 95 °C, alkylated with IAA for 20 minutes in the dark, and digested twice with trypsin first for 3 hours at 37 °C and then overnight at 30 °C.

5.3.4.3 Preparation for mass spectrometry and running parameters

Digested peptides were cleaned, and MS was performed as previously described with modifications [285]. Briefly, digests were cleaned using Pierce C-18 Spin columns (Thermo Scientific), the eluate dried, and dried in-solution digests and in-gel digests resolubilized in 20 and 30 µL, respectively, of 5% ACN and 0.1 % formic acid. Prior to injection, 1 µL of 500 fmol μL^{-1} GP was added to 9 µL of each sample and 2 µL were injected into a Thermo Q Exactive Plus™ Hybrid Quadrupole-Orbitrap™MS coupled to UltiMate 3000RSL® liquid chromatography (UPLC) system (Thermo Scientific™).

5.3.4.4 Data Analysis

Data was analyzed using PEAKS 8.5 (Bioinformatics Solutions Inc., Waterloo, ON, Canada) against a concatenated database including *Arachis hypogaea, Carya illinoinensis, Cocos nucifera, Corylus americana, Prunus avium, Prunus dulcis, Prunus persica, Malus domestica,* and GP (Uniprot, 225786 sequences, accessed 08/11/2020) with a contaminant database (thegpm.org, accessed 08/11/2020, version last modified 03/03/2019) [286]. PEAKS search parameters were 5 ppm mass error tolerance, 0.05 fragment mass error tolerance, fixed carbamidomethylation of cysteines, and the protease set to no enzyme or trypsin with up to 3 missed cleavages. PEAKS searches were further analyzed with a post-translational modification (PTM) search including up to 3 variable modification per peptide from a curated list of 336 possible modifications including a base list of 312 PTMs provided by peaks [340]. Peptide false discovery rate (FDR) was set at 0.1% and protein identifications required a -10lgP score of ≥30, at least two peptides identified, and at least 1 unique peptide. Identified peptides underwent BLAST searches (Uniprot.org) to verify their species of origin. For quantitation, data were normalized via the top three GP peptides and proteins were quantified according to the top three peptides if possible.

5.3.5 Risk assessment

Protein identifications and quantifications were compared with literature values and utilized assumptions to evaluate the allergic risk to consumers.

5.3.5.1 Food allergen risk assessment

Assumptions and literature values used for food allergen risk assessment are presented in Table 5.1. Exposure doses were calculated from the concentration of proteins identified, the hourly rate of E-liquid consumption, a maximum time limit to accumulate the dose, and a safety factor. Per allergenic source, exposure dose was compared to reference doses to determine if the risk of eliciting an allergic reaction was present.

5.3.5.2 Respiratory allergen risk assessment

Assumptions and literature values used for respiratory allergen risk assessment are presented in Table 5.2. A framework and associated experiments from Costigan *et al.* was utilized incorporating a generalized derived minimum-effect level established for enzyme protein respiratory allergens and a safety factor but without adjustment for enzyme content instead substituting total quantified protein content [5].

5.4. Results

5.4.1 Visualization of present proteins

Coomassie staining was performed to evaluate the presence of proteins in E-liquids. Samples were dialyzed and precipitated with the relationship of used volumes to original volumes as seen in Table 5.3. Visualization of samples via SDS-PAGE (Figure 5.1) demonstrates the presence of proteinaceous matter in the almond, coconut, and peach E-liquids as a band between 20 and 25 kDa, which was of identical among the three samples. The bands were excised for later in-gel digestion and analysis by MS.

Table 5.1 Literature values and assumptions used in food allergen risk assessment of E-liquids

Literature values	Associated value(s)	Reference(s)
Reference dose of ingested almond protein	0.1 mg total almond protein	[6, 7]
Prunin fraction of total almond protein	Total almond protein is 70% prunins	[9, 10]
High end of published E-liquid consumption	30 mL day^{-1}	[11, 12]
High end of published ENDS aerosol collection efficiency	91%	[13]

Assumptions	Associated value(s)	Rationale
Maximum duration of dose accumulation	2 hours	Excessive time to have inhalation rates correspond to elicitation via ingestion
Perfect transfer of dose from E-liquid	100%	Limitations of literature on transfer of protein through ENDS systems
Average waking day for an adult to consume E-liquid	16 hours day^{-1}	8 hours of sleep per night
Total safety factor	100	Limitations of study and uncertainty of relationship of eliciting dose via inhalation compared with ingestion

Table 5.2 Literature values and assumptions used in respiratory allergen risk assessment of E-liquids

Literature values	Associated value(s)	Reference(s)
Transfer ratio of cocoa shell extract proteins from E-liquid to filter pads	1 / 12,500,000	[5]
Proposed highest tolerable derived no effect level for enzyme respiratory allergens	15 ng m^{-3}	[8]
Prunin fraction of total almond protein	Total almond protein is 70% prunins	[9, 10]
Safety factor for efficient lung delivery	10	[5]
Assumptions	**Associated value(s)**	**Rationale**
Cocoa proteins aerosolize from E-liquid equivalently to almond proteins		Required to substitute quantified protein values
Total protein reported as ppm - mg protein per liter E-liquid		Conservative assumption to maximize theoretical aerosolized allergenic dose
E-liquids assessed in Costigan *et al.* 2017 are equivalent to those assessed here		Required to substitute quantified protein values
Silver stained SDS-PAGE and standard BSA accurately reflect transferred protein		Required to substitute quantified protein values

Table 5.3 E-liquid properties and volumes analyzed by SDS-PAGE and Mass Spectrometry

#	Sample	Post-dialysis volume (mL)	% acetone precipitated	Equivalent mL digested for MS	Equivalent mL on SDS-PAGE	Density (g mL^{-1})
1	Green apple	4	100	1	0.5	1.1995
2	Cherry	3	100	1	0.5	1.1692
3	Peach	7	83	0.83	0.415	1.1905
4	Sweet almond	5.5	100	1	0.5	1.1914
5	Peanut brittle	6.5	89	0.89	0.445	1.2027
6	Coconut	6	100	1	0.5	1.1521
7	Hazelnut	7.5	83	0.83	0.415	1.1246
8	Pecan	7.5	83	0.83	0.415	1.1502

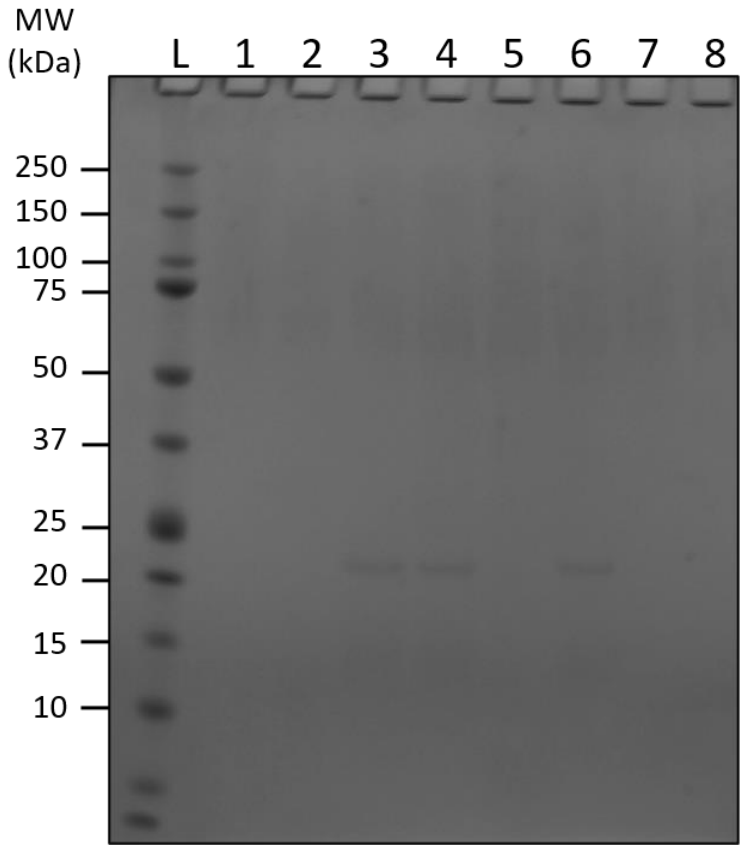


Figure 5.1 SDS-PAGE gel of E-liquids

Dialyzed and acetone precipitated E-liquids visualized using SDS-PAGE and Coomassie stain with lanes as follows: (L) molecular weight standards, (1) green apple , (2) cherry, (3) peach, (4) sweet almond, (5) peanut brittle, (6) coconut, (7) hazelnut, and (8) pecan.

5.4.2 Assessment of Data Sets and Peptide Assignments

To identify proteins present in the E-cigarette liquids, the samples were proteolytically digested and analyzed by MS. Data was initially assessed with protease set to no enzyme or trypsin both with secondary PTM searches, but initial assessment found that compared to the tryptic search without additional PTM search that the remainder had inferior peptide or protein identifications without significant

improvements in the breadth of data added and so only the base tryptic search was utilized. Identified peptides were scrutinized by individual BLAST searches on Uniprot to verify their origin. Peptides identified in almond E-liquid were attributable to *P. dulcis* and peptides identified in peach and coconut E-liquids were attributable to either *P. dulcis* or *P. persica*. Other species hits per peptide did not reach an alternative consensus species of origin.

Of the samples, only almond, coconut, and peach had at least 2 peptide matches to a single protein within a single MS injection and so the remaining samples were disregarded from further analysis. Coconut did not meet peptide identification criteria via in-solution digest but did meet criteria via in-gel digest and so was included for analysis. Among the remaining samples, a total of 14 peptides were identified (Appendix G Table 5.1) and quantified (Appendix G Table 5.2)

Peptide spectra matched across injections for peptides that were identified in multiple injections (Appendix G Figure 5.1). Overall, two peptides identified from almond were identified both in-solution and in-gel, but only one peptide from peach overlapped between in-gel and in-solution digests, and further no peptides overlapped between digests from coconut. Of specific peptides that matched across injections, [GNLDFVQPPR] was identified in both the in-solution digests of almond and peach but in neither in-gel digest. [TEENAFINTLAGR] was identified in each of the in-solution digests of almond, coconut, and peach as well as the in-gel digest of peach but was not identified in the in-gel digest of almond or coconut. Both [ADIFSPR] and [QETIALSSSQQR] were identified in each of the three in-gel digests of almond, coconut, and peach as well as the in-solution digest of almond. Only

[ALPDEVLANAYQISR] was not likewise identified in the in-solution almond digest; however, this peptide identified in the in-gel peach digest without identified in the in-solution digest.

5.4.3 Identification of proteins present in E-cigarette liquids

Proteins identified among all samples were attributable to two protein groups headed by Q43607 and A0A5E4FK23 (Appendix G Table 5.3), which both are themselves or contain proteins with descriptions regarding the almond allergen Pru du 6. Only the almond in-solution digest identified the protein group A0A5E4FK23, whereas all remaining samples identified Q43607. Peptide matches from in-solution digests of almond and peach spanned the length of Pru du 6 whereas in-gel digests from almond, coconut, and peach exclusively had peptides representing the basic chain of Pru du 6 proteins

5.4.4 Quantification of proteins and evaluation of the risk of eliciting allergic reactions

Quantified protein groups are presented in Table 5.4. For use in risk assessment, protein quantifications from in-gel digests were not used due to lack of identifications from both basic and acidic chains of Pru du 6 proteins. For protein group Q43607, the annotated signal peptide was ignored, and the molecular weight of the mature protein used to determine a concentration of 2.30 and 136.56 ng mL^{-1} present in peach and almond E-liquids, respectively. For protein group A0A5E4FK23, the signal peptide was not well annotated for any proteins within the group and so Q43608 was used with the

end of the signal peptide inferred from identity to Q43607 to determine a concentration of 4.40 ng mL^{-1} present in almond E-liquid and a summed protein concentration of 140.96 ng mL^{-1} prunins or 201.37 ng mL^{-1} total almond protein. The peach E-liquid protein quantification was not similarly converted due to unclear protein origin. According to food allergen risk assessment, the exposure doses were calculated as 687.17 ng and 7.84 ng for almond and peach E-liquids, respectively, and both were under the safety factor adjusted reference dose for almond of 1 µg. For respiratory allergen risk assessment, the exposure doses were calculated as 16.110 ng m^{-3} and 0.184 ng m^{-3} for almond and peach E-liquids, respectively. Compared to the safety factor-adjusted reference dose of 1.5 ng m^{-3}, almond E-liquid exceeded the reference whereas peach did not.

Table 5.4 Protein quantification using mass spectrometry

	In-solution digest (fmol)			In-gel digest (fmol)			Mature molecular mass (Da)
Accession	Almond	Coconut	Peach	Almond	Coconut	Peach	
Q43607	201.89	0.00	2.82	28.33	19.53	19.56	60874.45
A0A5E4FK23	7.09	0.00	0.00	0.00	0.00	0.00	55890.21

5.5 Discussion

This study sought to assess the presence of proteins present in a variety of extractive-based E-liquids, possible biological activity, evaluate their quantity, and examine the risk posed. In other words, are allergenic proteins present in extractive based E-liquids? And do these proteins pose a risk to allergic consumers?

Peptides representing Pru du 6 or a nearly identical homologue were identified in almond, peach, and coconut E-liquids. In decreasing order of both evidence and quantity

were almond, peach, and coconut. Both digests of almond confirmed the presence of almond peptides, peach identified either almond or peach peptides but was confirmed by both digests, but coconut only identified either almond or peach peptides by SDS-PAGE and corresponding digestion. The peptides identified in peach digests may have originated from peach, but this is unclear as seed storage proteins of genus *Prunus* are highly identical [341]; however, protein databases for species other than *P. dulcis* or *P. persica* are relatively small and are not similarly representative. The source of the almond peptides present in the coconut E-liquid is less clear, but we speculate that either the almond flavoring was mixed with coconut in preparation of the E-liquid or cross-contact of the flavoring itself or during mixing.

Each of the sequences identified are potentially allergenic as they have been either been themselves regarded as an allergen or nearly identical (>96%) to a known allergen. SDS-PAGE and its associated digestion of each of almond, peach, and coconut identified the presence of the basic subunit of Pru du 6 without accompanying acidic subunit [342]; however, in-solution digests of almond and peach did simultaneously identify both subunits. Linear epitopes have been identified across both subunits of Pru du 6 isoforms [343], as well as conformational epitopes to recombinantly produced whole or either independently produced subunit [344]. Peptides were identified for almond and peach that in-part encompassed epitopes and so the protein, if present in part or whole, may have the capacity to elicit allergic responses.

Inhalation of almond proteins, as opposed to ingestion, is atypical and not well documented in literature; therefore, risk was evaluated through both food allergen and respiratory risk assessments. As food allergens, assumptions were made such as

excessive E-liquid consumption as observed in a subset of ENDS users using heating coils rated to less than 1 Ohm of resistance [345], which can produce greater amounts of aerosol compared to conventional ENDS. These assumptions in factor of gross accumulation of allergenic protein did not permit a determination of risk of elicitation in allergic consumers. Alternatively, as respiratory allergens the primary assumptions were to allow substitution of observed total protein values to produce total aerosolized protein and therefore dose, which were weighed against the choice of threshold. Use of the highest proposed derived no effect level was still low enough to identify almond E-liquid as a sensitization risk, but not the peach E-liquid; however, Basketter *et al.* also reported levels for safe use of other respiratory sensitizers as low as 0.01 ng m^{-3} (1500x less than the generalized threshold used) [8], therefore the peach E-liquid may still pose a risk of *de novo* sensitization to consumers.

Food allergen aerosols have been assessed for egg [346], fish [347], and peanut [348], as well as wheat in the form of bakery dust [336]. Parallels can be drawn between extractive-based E-liquids and wheat allergies as those with oral wheat allergy commonly also have reactions to inhaled wheat [349], whereas those with inhaled wheat allergy are commonly observed to be tolerant to ingested wheat [350]. This divergence has been speculated to be due to differences in the epitopes where inhaled allergy to raw flour would utilize epitopes, likely conformational, that are not well represented in cooked wheat-based foods [349]; however, many of the same allergenic proteins implicated in oral wheat allergy have also been implicated in inhaled wheat allergy [351]. As the E-liquids are directly inhaled it is likely that any proteins present in the E-liquid would pose

a risk of sensitization although the sensitization is likely specific to the inhaled aerosolized proteins as opposed to ingested food.

Beyond proteins, the ENDS itself may also be involved in allergic sensitization. The size of particulate matter influences the immune response, where fine particulate matter (0.1 – 2.5 μm) tends to encourage Th2 responses [352], and most of the aerosol produced by ENDS fall within this range [353]. Large proportions of the vapor produced by ENDS are of the size to penetrate the tracheobronchial and bronchoalveolar regions of the lung [354], and such particulate matter has been associated with the development of asthma, particularly those with atopy [355]. Aerosolized E-liquid itself has been found to be proinflammatory [356], which constitutes an impaired barrier for allergic sensitization [357]. Use of extractive-based E-liquids is likely to encourage allergic sensitization to proteins present from the extract.

This work was designed as an initial investigation and therefore individual techniques were not well replicated, but rather used in tandem to identify proteins present in E-liquids. Inclusion of further extractive-based E-liquids, controls per allergenic food source, as well as replication is required to further establish the risk posed. Further controls regarding ENDS in terms of design, aerosolization temperatures, power, and a standardized protein-containing E-liquid are required to correlate allergen presence in E-liquids to dose inhaled. Lastly, biological testing is needed to establish the effects of inhaling these allergenic proteins on *de novo* sensitization as well as elicitation of current allergies.

CHAPTER 6: LITERATURE-BASED RISK ASSESSMENT OF ACTIVATED CARBON MADE FROM WALNUT HULLS FOR USE IN WATER PURIFICATION SYSTEMS

6.1 Summary

Activated carbon (AC) is a common material in water purification, which can be made from various materials including walnut hulls. Unprocessed walnut hulls contain several hazards including food allergens, naphthoquinones, and ellagitannins. When properly processed, the present hazards are expected to be rendered inert to make the resultant AC not a significant risk to consumers.

6.2 Background

AC is a porous material with a high surface area with multiple uses including the purification of drinking water and is often used in sequence with several other materials, where AC is included for its ability to adsorb organic compounds [358-360]. Production of AC proceeds though dehydration, pyrolysis, and carbonization using high heat and is subsequently activated by using hot steam, acid, base, or salt [360, 361]. AC is made from carbonaceous materials such as walnut shells, coconut husks, wood, or coal [358, 362, 363]. Walnut hulls are a byproduct of walnut shelling and can be used as a potential value-added source material for the production of AC. Walnut hulls form the outer husk around the shells and kernels and do not have a principal use beyond production of animal feed, applications as a dye, or as an herbal medicine [364-367]. In the U.S. there are two walnut species of economic importance: *Juglans nigra*, the Eastern black walnut,

and *Juglans regia*, the English walnut [368]. Walnuts contain a number of hazards such as food allergens and toxins, which may contaminate walnut hulls and may persist through processing [369, 370].

6.3 Activated Carbon Processing

Production of AC uses heat to cause dehydration, pyrolysis, graphene nucleation, and carbonization via heating at or above 600 °C under an inert atmosphere to increase the solid carbon content to more than 90% [361]. During carbonization solid, liquid, and gaseous byproducts are formed and can be separated to leave only carbonized biochar [371, 372]. Afterwards, the biochar is activated either by chemical means by adding inorganic acids, bases, or salts and heated for hours above 300 °C or alternatively by physical means using oxidizing gasses such as steam above 800 °C [360, 361, 373-375]. Diffusion of oxidizing agents into the biochar breaks the aromatic carbon structure and increases the porosity and surface area of the carbon [375]. For lignocellulosic biomass above 250 °C and up to 350 °C pyrolysis and depolymerization of cellulose occurs, above 350 °C aromatic carbon is seen with growing graphene sheets, and above 600 °C the remaining non-carbon is expelled and the graphene structures coalesce to form the carbonized matter [361]. As Ioannidou *et al.* has explained, AC is often heated for excessive amounts of time (1-8 hours) at temperatures exceeding 600 °C, optimized per material, and this allows the assumption that the heating is homogenous across the biochar and has undergone both pyrolytic and carbonization temperatures [376].

Activated carbon itself is a minor hazard, particularly powders due to combustibility, and it can also be an irritant. As an irritant it does have exposure limits,

however oral rat studies have shown that the LD50 was >10,000 mg/kg indicating that it is a minimal hazard because it is likely that a human would either see or taste the AC in the water prior to being able to consume enough for it to be a major health hazard [377, 378]. However, AC is not intended to become part of the product, rather it is a food-contact surface [379]. Therefore, if AC were continuously seeping into the water it was purifying that would be a major defect of the water purifier or a mistake of the granularity of the AC supplied. From this, AC itself is neither a likely or harmful hazard.

6.4 Allergenic Hazards of Walnut Hulls

Walnuts are a major allergen and part of eight classes of allergenic foods that comprise 90% of US food allergies as designated by the Food Allergen Labeling and Consumer Protection Act of 2004 [204]. Tree nuts can cause IgE-mediated reactions with symptoms ranging from oral allergy syndrome to anaphylaxis and are a major source of food allergy related fatalities [380, 381]. Among tree nuts, walnut allergy is the most common [382]. Allergic cross-reactivity can be seen across walnut cultivars and across tree nuts [383, 384]. Walnut hulls contain approximately 15% protein by dry wright [385]. Threshold doses for the walnut allergic population to have an allergic reaction have been determined such that 3.1 mg of walnut protein would be sufficient to cause an allergic reaction in 5% of the walnut allergic population (ED_{05}) [386]. Assuming a soldier drinks 11.4 liters of water per day, a protein concentration of 0.28 µg/mL would be theoretically meet the ED_{05} for walnut allergic individuals while holding the over-conservative assumptions that all walnut proteins are equivalent and that an allergic reaction could be elicited by accumulation across an entire day [387].

6.5 Effects of Thermal Processing on Allergens

Allergenic proteins elicit symptoms through a pair antibody binding epitopes on the protein, which are either bound to a conformational or linear epitope of the allergen [70]. Therefore, both unfolding and destruction of peptide bonds are necessary to eliminate allergenic potential. Protein denaturation temperatures vary by protein but can occur as high as 108 °C [388]. Protein pyrolysis and decomposition occur above 200 °C with volatile nitrogenous decomposition continuing through 400 °C [388, 389].

Thermogravimetric analysis has been applied to peptides [390], milk [391], collagen [392], and microalgae [393] to show at what temperatures that gross decomposition of proteins occurs. Amino acid decomposition has been found to follow first-order rates [394]. Synthesized peptides were largely found to represent proteins where temperatures below 200 °C evaporated surrounding water and between 200 °C and 500 °C degraded the peptides through decarboxylation, deamination, and depolymerization with rate maxima at 235 °C and 299 °C [390]. Camel and cow's milk were indicated to have combined degradation of protein and fat around 380 °C [391]. Collagen decomposition has been indicated to occur between 220 and 380 °C with maximum rates between 285 and 300 °C with combustion occurring at 350 °C [392]. Another investigation of collagen found that through 100 °C absorbed water is evaporated and between 280 °C and 400 °C bound water is released [395]. Modeling theoretical thermogravimetry of microalgae has demonstrated that protein largely degrade upon heating to 350 °C [393]. However, decomposition has been found to proceed at a lower rate than hydrolysis and therefore hydrolysis can be expected to occur first [396].

Destruction of peptide bonds has been investigated in terms of spontaneous hydrolysis. Hydrolysis of peptides has been found to be a first-order process and has been calculated to have a half-life of approximately 6 minutes at 250 °C and neutral pH [396]. Although pH can affect the rate of hydrolysis, the rate only increases in acid or alkali [397]. The minimum length of peptide needed to elicit clinical symptoms has been suggested to be as little as 29 amino acids long [398]. Under the aforementioned conditions, it would theoretically take 66 minutes to reduce titin (approximately 35,000 residues) to having no peptides over 29 amino acids. Properly processed AC is unlikely to have sufficient epitopes nor free amino acids remaining to cause allergic reactions, even if processed at relatively low temperatures.

6.6 Allergen Labeling of Activated Carbon Derived from Walnut Hulls

The Federal Food, Drug, and Cosmetic Act in the U.S. under section 409 (h) states “the term ‘food contact substance’ is any substance intended for use as a component of materials used in manufacturing, packing, packaging, transporting, or holding food if such use is not intended to have any technical effect in such food” [204]. The Food and Drug Administration has also clarified that AC is a food-contact substance as defined in section 409. Lastly, because the Food Allergen Labeling and Consumer Protection Act in the U.S. section 403 (w) requires labeling of products with ingredients that contain tree nuts it would typically require labeling of products made after use with the AC, however because by definition the carbon cannot become part of the food itself, it is not an ingredient and not subject to allergen labeling [204].

6.7 Chemical Hazards of Walnut Hulls

Walnut hulls contain a number of phenolic compounds including ellagitannins and 1,4-naphthoquinones [399, 400]. The ellagitannin ellagic acid has been described as being present in walnut kernels up to 18% by dry weight and the 1,4-naphthoquinone juglone has been described as being present in walnut hulls up to 1.5% by dry weight [400-402].

Walnut hulls contain several similar 1,4-naphthoquinones including juglone, plumbagin, and 1,4-naphthoquinone. Naphthoquinones, including juglone and plumbagin, are known to be cytotoxic whereby they will bind DNA and inhibit replication, inhibit enzymes, and disturb membranes and are able to reduce cell viability in vitro at a concentration of 1 µM [403, 404]. Lawsone, an isomer of juglone, has an identified no-observed-adverse-effect level (NOAEL) of 2 mg/kg which corresponds to an acceptable daily intake of 1.4 mg for a 70 kg individual and assuming a soldier's water intake of 11.4 L/day we find that 0.12 µg/ml or 130 ppb of total 1,4-naphthoquinones can be in the water and not be expected to be harmful [387, 405]. Data for decomposition temperatures are not available for juglone, however it is noted that it sublimes at 155 °C and therefore could be removed with the volatile gasses during production [406]. Similarly, plumbagin melts at 76 °C and 1,4-naphthoquinone melts at 121 °C, which generally indicates that the class of 1,4-naphthoquinones present could be separated out from the biochar and not be present in the produced AC [407, 408].

Walnut hulls also contain a number of ellagitannins including ellagic acid [400, 409]. Ellagic acid has been subject of a 90-day sub-chronic oral rat (F344) study and showed that female rats had a NOAEL level of 3254 mg/kg while male rats were

estimated to have a no-observed-effect level of 3011mg/kg (NOEL) where the only major effect was a decrease in female bodyweight and no changes in histopathology were observed [409, 410]. Using the NOAEL for ellagic acid, we can derive an ADI of 2278 mg/day for a 70 kg individual and taking a soldier's maximum daily water intake of 11.4 L/day then we find that 0.2 mg/ml or 200 ppm of ellagic acid can be in the water and not be expected to harmful [387]. However, this assumes that ellagic acid is not removed or destroyed by heating. Ellagic acid melts at 300 °C, so this compound could be separated from the biochar and not be present in processed AC [411].

6.8 Conclusions

Use of walnut hulls to make AC is not a significant risk provided that the hulls are processed to at least 400 °C with a holding time to result in nitrogenous decomposition. Through a heating of this magnitude, allergenic and chemical hazards will have been destroyed or removed as liquids or gasses such that no significant hazards will be present from the walnut hulls. Therefore, if AC made from walnut hulls is produced to the standard of a food-contact substance, it is unlikely to contain any significant risks.

CHAPTER 7: LITERATURE-BASED RISK ASSESSMENT OF ALLERGENS IN SMOKE DERIVED FROM TREE NUT WOOD

7.1 Summary

Smoking is a time-tested means to preserve food as well as improve flavor. Wood used for smoking includes wood from trees that produce tree nuts, which may pose a risk to tree nut allergic individuals. Due to differences in protein profiles between wood and tree nuts there is no demonstrated presence of tree nut allergens in the wood. Tree nut wood smoke is not a significant source of allergic risk.

7.2 Background

Smoke is an aerosol emitted upon combustion or pyrolysis of a material and is composed of water, gasses, and particulate matter [412, 413]. Wood smoke has desirable effects on food including color, flavor, and has antioxidant and antimicrobial effects [414]. Common woods for smoking include mesquite, fruitwood (e.g. apple, cherry), and tree nut wood (e.g. pecan, hickory) [415]. In this work, “tree nut wood” refers to wood derived from trees that produce tree nuts defined as major food allergens by the Food Allergen Labeling and Consumer Protection Act of 2004 with further clarification by the U.S. Food & Drug Administration: almond, beech, Brazil, butternut, cashew, chestnut, chinquapin, coconut, hazelnut, ginko, hickory, lichee, macadamia, pecan, pine, pili, pistachio, shea, and walnut [204, 416]. Tree nuts are part of the eight classes of allergenic foods that comprise 90% of U.S. food allergies and the presence of which in food must be labelled [204]. Tree nuts can cause IgE-mediated reactions with symptoms ranging from

oral allergy syndrome to anaphylaxis and are a major source of food allergy related fatalities [380, 381]. It is unclear if the use of tree nut wood to produce smoke is a risk to tree nut allergic individuals.

7.3 Wood, Trees, and Logging

Wood is formed from the secondary xylem of woody plants [417]. The wood and bark are segregated by the vascular cambium, which wraps around the secondary xylem. Bark is then all tissues surrounding the vascular cambium including the phloem. Wood is predominantly formed from cellulose, hemicellulose, lignin, and extractives that include a number of classes of compounds such as flavonoids, isoprenoids, and tannins [418]. Trees are divided into angiosperms and gymnosperms called hardwoods and softwoods respectively [419]. These differ in their lignin, hemicellulose, bark thickness, and seasonal senescence and abscission of foliage. Hardwood lignin subunits have a single methoxyl group as opposed to two of softwood, both have the hemicellulose xylan but hardwoods lack galactoglucomannan, hardwood bark tends to be thinner, and hardwoods tend to lose their foliage seasonally.

Members of the genus *Populus* are hardwoods that are regarded as model trees [420-422] and includes poplar, aspen, and cottonwood species [423, 424]. In *Populus*, nitrogen in the tree varies seasonally and is stored in the form of proteins and free amino acids principally as arginine, glutamine, and asparagine [425]. In autumn, leaf proteins are broken down and amino acids are transported to bark and wood to produce bark storage proteins. In spring, bark storage proteins are broken down and moved to reproductive tissues, buds, and leaves. Amino acids have been regarded as a major means

of nitrogen both storage and transport in *Populus* [426], but also peach trees [427]. In peach trees, total nitrogen and soluble protein in wood is low and seasonally stable whereas the bark stores most of both nitrogen and soluble protein. Wood soluble protein was found to be approximately 1 $mg.g^{-1}$ dry weight year-round whereas phloem soluble protein ranged from approximately 5 and 25 $mg.g^{-1}$ dry weight varying by season. Broadly, the nitrogen content of wood has been found to vary among tree species, has been found to be generally proteinaceous, and a sampling of a variety of trees were found to average approximately 0.1% nitrogen in wood [428]. The protein composition of wood is not similar to that of nuts, with woods lacking seed storage proteins associated with commonly consumed plant parts [429-431].

Processing wood for smoking begins with standing trees. Trees are felled and they may optionally undergo limbing and bucking where the whole trees have their limbs removed and are then segmented into logs respectively [432]. If limbed, the branches are predominantly left by the stump and so not used [433]. Logs may then be debarked and either used as is or processed into wood chips [432]. Wood chips and sawdust are also obtained as byproducts of processing lumber [434]. Both wood chips and sawdust are commonly used in industrial smoke production [435].

7.4 Industrial Smoke Production, Processing, and Usage

The type of wood, temperature of burning, and method of burning can alter the qualities of resultant smoke [436]. Hardwoods are preferred over softwoods to produce smoke as softwoods contain higher levels of resin acids causing the smoke to impart more soot and acidic flavor [437]. Hardwood typically consists of 40-60% cellulose and

20-30% of hemicellulose and lignin each where the balance of these primary components affects the pyrolysis product balance of the smoke. The specific wood used for smoking can alter the color and flavor imparted on the food [435], but the smoke composition can also vary with the temperature and availability of oxygen [438].

Upon heating, the wood will dry up to 170 °C and hemicellulose, cellulose, and lignin generally undergo pyrolysis between 200-260 °C, 260-320 °C, and 310-500 °C respectively [414, 437]. The desirable products stem from both primary pyrolysis and secondary breakdown of primary products. Incorporation of some oxygen during smoldering is important to high quality smoke as oxygen assists with secondary reactions [414]. Temperatures that are too low will not efficiently smoke whereas too high increases production of toxic polycyclic aromatic hydrocarbons [413]. However, if the wood is allowed too much oxygen at high enough temperature, then self-sustaining combustion or self-ignition of wood can result [439, 440], which results in a loss of consistency as well as quality.

Industrial production of smoke most commonly uses smoldering generators where a tray is heated to around 350 °C, fed either wood chips or sawdust, and sustained with air blown from underneath to feed oxygen, control temperatures, and prevent self-sustaining combustion [435]. The temperature of the wood varies according to the temperature of the plate, air flow, and the size of the wood used. The wood is often added continuously, so keeping the wood in a particular shape (e.g. cone) maintains a constant air flow to wood surface area to improve the consistency of the smoke and further hinder self-ignition [437].

Alternative smoke generators include friction generators and steam condensers. Friction generators grind wood against a rotating barrel for short bursts resulting in contact temperatures up to 380 °C [441]. Steam condensers pass sawdust into overheated steam held between 300 and 450 °C to generate smoke, which is subsequently cooled [413].

Uses of produced smoke can be divided into direct smoking, where emanating smoke is used on exposed product, or indirect smoking, where smoke is passed though water, distilled, refined, and concentrated and applied as a dip or mist [437]. Direct smoking can be separated into hot smoking, where products are smoked while cooked, or cold smoking, where the products are smoked without cooking [436]. Indirect smoking aims to reduce tar and soot content through washing and filtration steps as well as concentrating desirable flavor [414]. Smoke produced by passing it through water as in indirect smoking is referred to as liquid smoke. Cooling and condensing the water can be used with distillation or gravity to induce phase separation into oily resin, water soluble, and water insoluble phases [437]. The water soluble and insoluble phases can be then used as source materials for functional compounds to produce liquid smoke depending on product identity and desired qualities.

Liquid smoke can be further processed into a dry powder either by plating or spray-drying [442]. Plating uses a carrier molecule to embed a liquid flavor onto a solid carrier such as salt, lactose, starch, or maltodextrin. Spray-drying incorporates the liquid flavor and solid carrier into a slurry and atomizes the mixture where volatiles and water evaporate.

7.5 Smoke as a Medium for Allergen Transport

Although a number of tree nuts have been investigated, many do not have quantitative data regarding eliciting doses that cause reactions [443]. In light of this, hazelnut has been suggested as a placeholder for tree nuts in these cases. The reference dose for hazelnut via quantitative risk assessment was set at 0.1 mg of hazelnut protein based from the eliciting dose of 1% of the population (ED01), where this reference dose was set to protect 99% of the allergic population from objective reactions [316].

Inhalant allergens have been identified in the cedar wood (*Juniperus ashei*), however allergens were not identified in wood smoke by SDS-PAGE, cedar-pollen sensitive IgE, nor monoclonal antibody [444]. Mesquite wood smoke (genus *Prosopis*) has been investigated using SDS-PAGE and IgE immunoblots using sera from mesquite-pollen SPT positive and self-reported mesquite-smoke-sensitive patients [445]. From an undescribed quantity of mesquite smoke passed through water and concentrated, More *et al.* identified protein bands by SDS-PAGE and immunoreactive bands were found in both sets of sera. This indicates that some amount of proteins and allergens from wood can be aerosolized and retain reactivity.

The largest body of work regarding smoke, its composition, and effects on humans has been though tobacco cigarettes. In a study by Becker *et al.*, an 18 kDa glycoprotein was purified from cured tobacco leaves and injected to a mix of smoker and non-smokers and demonstrated positive skin tests in a subset of both groups [446]. The purified protein was analyzed by polyacrylamide gel along with extracts purified similarly but starting with either cigarette smoke condensate (tar) or cigarette smoke saline (cigarette smoke passed through phosphate-buffered saline). The purified proteins

were found to be similar according to hemagglutination inhibition assay and were also cross-reactive. Similarly cross-reactive proteins were also found to be present in other members of *Solanaceae* including eggplant, green pepper, potato, and tomato. In the tar, the concentration of the protein was found to be between 1.8 and 3.6 mg/g and was used to extrapolate a range of 720 and 1440 µg of antigen per packet of 20 cigarettes assuming 20 mg of tar per cigarette. Concentration of antigen in cigarette smoke saline was not determined. This study suggests that a quantity of proteins can be found and purified from smoke and therefore smoke can act as a medium for allergen transport. Neither the degree of transport relative to source material nor how ubiquitous the aerosolization is across proteins is not addressed.

A study by Voisine *et al.* investigated the transfer of protein into tobacco smoke using cigarettes spiked with the subtilisin enzyme savinase [447], where subtilisins are used as detergent enzymes known to aerosolize and cause IgE-mediated allergy [448]. Savinase was detected by immunostaining method allowing detection of savinsae transfer down to 0.009% in mainstream smoke (directly smoked) or 0.054% in sidestream smoke (second hand). Detectible transfer was not found in any scenario and it was concluded that there was no significant protein transfer via smoke. This contradicts Becker *et al.* to a degree, however together these suggest that there is a low level of protein transfer. Liu *et al.* contextualized the study by Voisine *et al.* and explains that the upper limit of savinase transfer into mainstream smoke was found to be <0.009%, however a more rigorous smoking regimen to increase tar yield (from 15 mg to 35 mg tar) could increase the transfer to <0.021% or at most 4.4 ng savinase per cigarette with 700 mg of tobacco spiked at 6000 ppm savinase [449].

A number of assumptions would be required to appropriately apply this information to the transfer of tree nut allergens using smoke to food, which are not limited to: wood is a perfect substitute for tobacco, conditions to burn wood are functionally identical to experimental conditions to transfer savinase in tobacco smoke, no functional differences between the aerosolizability of wood proteins vs savinase, and tree nut allergens or immunoreactive homologues are present in wood at approximately 6000 ppm. Wood and seeds do not share the same protein profile and so the main source of tree nut allergens in the wood would be tree nuts themselves, but processed wood would not carry tree nuts after limbing and debarking. The key assumption of allergen presence is not met in processed wood that would be used for smoking food.

7.6 Effects of Thermal Processing on Allergens

Allergenic proteins elicit symptoms through a pair of antibody binding epitopes on the protein, which are either bound to a conformational or linear epitope of the allergen [70]. Therefore, both unfolding and destruction of peptide bonds are necessary to eliminate allergenic potential. Protein denaturation temperatures vary by protein but can occur as high as 108 °C [388]. Protein pyrolysis and decomposition occur above 200 °C with volatile nitrogenous decomposition continuing through 400 °C [388, 389].

Thermogravimetric analysis has been applied to peptides [390], milk [391], collagen [392], and microalgae [393] to show at what temperatures that gross decomposition of proteins occurs. Amino acid decomposition has been found to follow first-order rates [396]. Synthesized peptides were largely found to represent proteins where temperatures below 200 °C evaporated surrounding water and between 200 °C and

500 °C degraded the peptides through decarboxylation, deamination, and depolymerization with rate maxima at 235 °C and 299 °C [390]. Camel and cow's milk were indicated to have combined degradation of protein and fat around 380 °C [391]. Collagen decomposition has been indicated to occur between 220 and 380 °C with maximum rates between 285 and 300 °C with combustion occurring at 350 °C [392]. Another investigation of collagen found that through 100 °C absorbed water is evaporated and between 280 °C and 400 °C bound water is released [395]. Modeling theoretical thermogravimetry of microalgae has demonstrated that proteins largely decompose upon heating to 350 °C [393]. However, decomposition has been found to proceed at a slower rate than hydrolysis and therefore hydrolysis can be expected to occur first [396].

Destruction of peptide bonds has been investigated in terms of spontaneous hydrolysis. Hydrolysis of peptides has been found to be a first-order process and has been calculated to have a half-life of approximately 6 minutes at 250 °C and neutral pH [396]. Although pH can affect the rate of hydrolysis, the rate only increases in acid or alkali [397]. The minimum length of peptide needed to elicit clinical symptoms has been suggested to be as low as 29 amino acids long [398]. Under the aforementioned conditions, it would theoretically take 66 minutes to reduce titin (approximately 35,000 residues) to having no peptides over 29 amino acids.

In context of burning wood, some of the protein can be expected to be partially degraded as the wood burns. However, it cannot be assumed that all of the protein will be degraded. For example, it is known that after smoke has been produced it will rapidly cool to the surrounding temperatures [414]. This further indicates that the focus on the

study of Voisine *et al*. controls for conditions such as protein degradation during protein transfer and therefore is not necessary to account for in the prior calculations.

7.7 Allergen Labelling of Products Produced in Part with Smoke from Tree Nut Wood

The Federal Food, Drug, and Cosmetic Act in the U.S. under section 201 (qq) defines major food allergens includes tree nuts and also stipulates labeling as defined in section 403 (w) [204]. Notably the requirement is that labeling is stipulated on the food ingredient containing protein derived from a major allergen, which requires both that tree nut protein be included in the wood and that tree nut protein transfer to the final product. It is not unreasonable to assume some tree nut is included in the wood, however here we have failed to demonstrate that transfer of protein in smoke is significant. Therefore, labeling products smoked with wood derived from trees that produce tree nuts is not necessary.

7.8 Conclusions on the risks associated with Smoke derived from Tree Nut Wood

Use of wood from tree nut trees is not a significant risk to allergic consumers of smoked foods. There is evidence that a low level of protein transfer can occur through smoke, however wood proteins differ from tree nut allergens and precludes concerns of tree nut allergens transferring from the wood in smoke. Labeling is unnecessary because of the very low levels of expected protein transfer.

CHAPTER 8: SUMMARY AND CONCLUSIONS ON NOVEL SOURCES OF FOOD ALLERGENS

Novel foods and novel sources of food allergens require investigation to evaluate present proteins and their associations with known allergens. Mass spectrometry has the capacity to evaluate novel sources of food allergens and can inform qualitative and quantitative assessments of risk. This set of works had the following aims:

1. Explore novel foods, foods processed in novel ways, and novel means of exposure as novel sources of food allergens;
2. Develop methodologies for incorporating mass spectrometry into allergen risk assessment;
3. Evaluate how conventional risk assessment methodologies can be leveraged to determine the qualitative and quantitative allergic risks posed by novel sources of food allergens.

The first aim was accomplished using literature reviews and experimental evaluations using liquid chromatography-electrospray ionization-mass spectrometry to explore *Acheta domesticus*, the house cricket, and *Tenebrio molitor*, the yellow mealworm; extensively thermally processed walnut hulls and peanuts, as well as smoke from the wood of tree nut trees and E-cigarette liquids. While many foods are well-characterized, there are many non-novel foods that are not well-characterized despite associations with food allergies. Novel foods and novel sources of food allergens are generally not well-characterized and therefore contain many unknowns such as the

identities of present proteins, the quantity of those proteins, and the associations those proteins with known allergens. To understand risks associated, the hazard must first be identified and characterized.

The second aim was accomplished during experimental investigation of subject novel sources of food allergens during the analysis of mass spectrometry data to annotate and predict the allergenic proteins present in both *A. domesticus* and *T. molitor*, determine theoretically viable IgE-binding lengths present in extensively thermally processed peanuts, and analysis of E-liquids in terms of both food allergen and respiratory risk assessments. Mass spectrometry is currently growing as a field of analysis particularly regarding data analysis. Unlike other methods of food allergen detection and quantification, such as ELISA that produces a single measure of allergen presence, mass spectrometry is far more open-ended regarding the amount of data that can be generated without clear boundaries regarding which data are meaningful. However, mass spectrometry also affords the opportunity for use of this greater amount of information generated to inform more complex questions in so far as the data analysis methodology is validated. Currently mass spectrometry is limited by a lack of clear data analysis workflows relative to the inherent integration of ELISA data, which forces mass spectrometry to act as secondary validation of ELISA as opposed to a direct alternative analytical technique.

The third aim was accomplished by incorporating assumptions into conventional risk assessment workflows to allow for the use of mass spectrometry data and inform risk. Assumptions are a critical point of risk assessment. A hazard that is impossible to encounter is no risk. Framing the circumstances that a hazard may become a risk is

therefore an important feature of risk assessment. In mass spectrometry, a critical assumption in data analysis is that individual peptide identifications indicate the presence of their progenitor protein and/or allergen. Incorporation of excessive assumptions into an analysis increases the likelihood that the resultant answer be regarded as insufficient, incorrect, or flawed. To ensure the future utility of mass spectrometry in a risk assessment workflow, a consensus among stakeholders is necessary on what conditions, circumstances, and assumptions are permissible.

Novel foods and ingredients will continue to be developed and discovered as will novel sources of food allergens. Stakeholders need to be sufficiently informed to make well-reasoned decisions as to which risks they are willing to undertake. Mass spectrometry will continue to improve and generate large amounts of data and therefore there is a need for further research into how mass spectrometry can be utilized in the workflow of allergen risk assessment.

REFERENCES

1. Mueller, G.A., S.J. Maleki, and L.C. Pedersen, *The Molecular Basis of Peanut Allergy.* Current Allergy and Asthma Reports, 2014. **14**(5): p. 429.
2. Palladino, C. and H. Breiteneder, *Peanut allergens.* Molecular Immunology, 2018. **100**: p. 58-70.
3. Bird, J.A., *Food Allergy, an Issue of Immunology and Allergy Clinics of North America, E-Book*. 2017: Elsevier Health Sciences.
4. Jeebhay, M.F. and A.L. Lopata, *Chapter 2 - Occupational Allergies in Seafood-Processing Workers*, in *Advances in Food and Nutrition Research*, J. Henry, Editor. 2012, Academic Press. p. 47-73.
5. Costigan, S. and J. Lopez-Belmonte, *An approach to allergy risk assessments for e-liquid ingredients.* Regulatory Toxicology and Pharmacology, 2017. **87**: p. 1-8.
6. Bureau, A. *Summary of the 2019 VITAL scientific expert Panel recommendations*. 2019; Available from: http://allergenbureau.net/wp-content/uploads/2019/09/VSEP-2019-Summary-Recommendations_FINAL_Sept2019.pdf.
7. Remington, B.C., et al., *Updated population minimal eliciting dose distributions for use in risk assessment of 14 priority food allergens.* Food and Chemical Toxicology, 2020: p. 111259.
8. Basketter, D.A., et al., *Defining occupational and consumer exposure limits for enzyme protein respiratory allergens under REACH.* Toxicology, 2010. **268**(3): p. 165-170.
9. De Angelis, E., et al., *Heat and pressure treatments on almond protein stability and change in immunoreactivity after simulated human digestion.* Nutrients, 2018. **10**(11): p. 1679.
10. Esteban, R.M., F.J. López-Andréu, and O. Carpena, *Protein extractability of almond (Prunus amygdalus, Batsch) seed.* Journal of the Science of Food and Agriculture, 1985. **36**(6): p. 485-490.
11. Jabba, S.V. and S.-E. Jordt, *Estimating Fluid Consumption Volumes in Electronic Cigarette Use—Reply.* JAMA Internal Medicine, 2020. **180**(3): p. 468-469.
12. Smets, J., et al., *When less is more: vaping low-nicotine vs. high-nicotine e-liquid is compensated by increased wattage and higher liquid consumption.* International journal of environmental research and public health, 2019. **16**(5): p. 723.
13. Oldham, M.J., et al., *Particle size distribution of selected electronic nicotine delivery system products.* Food and Chemical Toxicology, 2018. **113**: p. 236-240.
14. Bhat, Z.F., H. Bhat, and S. Kumar, *Chapter 73 - Cultured meat—a humane meat production system*, in *Principles of Tissue Engineering (Fifth Edition)*, R. Lanza, et al., Editors. 2020, Academic Press. p. 1369-1388.
15. Finnigan, T., L. Needham, and C. Abbott, *Chapter 19 - Mycoprotein: A Healthy New Protein With a Low Environmental Impact*, in *Sustainable Protein Sources*, S.R. Nadathur, J.P.D. Wanasundara, and L. Scanlin, Editors. 2017, Academic Press: San Diego. p. 305-325.
16. Gahukar, R.T., *Chapter 4 - Edible Insects Farming: Efficiency and Impact on Family Livelihood, Food Security, and Environment Compared With Livestock*

and Crops, in *Insects as Sustainable Food Ingredients*, A.T. Dossey, J.A. Morales-Ramos, and M.G. Rojas, Editors. 2016, Academic Press: San Diego. p. 85-111.
17. Petkova, M., *Chapter 5 - Edible insects and related products*, in *The Role of Alternative and Innovative Food Ingredients and Products in Consumer Wellness*, C.M. Galanakis, Editor. 2019, Academic Press. p. 139-165.
18. Mateos-Aparicio, I. and A. Matias, *Chapter 9 - Food industry processing by-products in foods*, in *The Role of Alternative and Innovative Food Ingredients and Products in Consumer Wellness*, C.M. Galanakis, Editor. 2019, Academic Press. p. 239-281.
19. Khouryieh, H., *Novel and emerging technologies used by the US food processing industry.* Innovative Food Science & Emerging Technologies, 2020: p. 102559.
20. Satya, P. and D. Sarkar, *Chapter 4 - Plant Biotechnology and Crop Improvement*, in *Biotechnology for Sustainable Agriculture*, R.L. Singh and S. Mondal, Editors. 2018, Woodhead Publishing. p. 93-140.
21. United Nations, *World Population Prospects 2019: Highlights (ST/ESA/SER.A/423).* Department of Economic and Social Affairs Population Division, Editor. 2019.
22. Umesha, S., H.M.G. Manukumar, and B. Chandrasekhar, *Chapter 3 - Sustainable Agriculture and Food Security*, in *Biotechnology for Sustainable Agriculture*, R.L. Singh and S. Mondal, Editors. 2018, Woodhead Publishing. p. 67-92.
23. Nadathur, S.R., J.P.D. Wanasundara, and L. Scanlin, *Chapter 1 - Proteins in the Diet: Challenges in Feeding the Global Population*, in *Sustainable Protein Sources*, S.R. Nadathur, J.P.D. Wanasundara, and L. Scanlin, Editors. 2017, Academic Press: San Diego. p. 1-19.
24. Nadathur, S.R., J.P.D. Wanasundara, and L. Scanlin, *Chapter 25 - Feeding the Globe Nutritious Food in 2050: Obligations and Ethical Choices*, in *Sustainable Protein Sources*, S.R. Nadathur, J.P.D. Wanasundara, and L. Scanlin, Editors. 2017, Academic Press: San Diego. p. 409-421.
25. Ohlsson, T., *Chapter 43 - Sustainability and Food Production*, in *Food Safety Management*, Y. Motarjemi and H. Lelieveld, Editors. 2014, Academic Press: San Diego. p. 1085-1097.
26. Huis, A.v., et al., *Edible insects: future prospects for food and feed security*. FAO Forestry paper, 2013(171).
27. Oonincx, D.G. and I.J. De Boer, *Environmental impact of the production of mealworms as a protein source for humans–a life cycle assessment.* PloS one, 2012. **7**(12): p. e51145.
28. Lundy, M.E. and M.P. Parrella, *Crickets are not a free lunch: protein capture from scalable organic side-streams via high-density populations of Acheta domesticus.* PloS one, 2015. **10**(4): p. e0118785.
29. Barrena, R. and M. Sánchez, *Neophobia, personal consumer values and novel food acceptance.* Food Quality and Preference, 2013. **27**(1): p. 72-84.
30. Siegrist, M. and C. Hartmann, *Consumer acceptance of novel food technologies.* Nature Food, 2020. **1**(6): p. 343-350.

31. Magnuson, B., et al., *Review of the regulation and safety assessment of food substances in various countries and jurisdictions.* Food Additives & Contaminants: Part A, 2013. **30**(7): p. 1147-1220.
32. [FSANZ] Food Standards Austalia New Zealand. *Food Standards Code* 2020 201127 [cited 2020 201127]; Available from: https://www.foodstandards.gov.au/code/Pages/default.aspx.
33. [FSANZ] Food Standards Austalia New Zealand. *Advisory Committee Novel Foods* 2020 201127 [cited 2020 201127]; Available from: https://www.foodstandards.gov.au/industry/novel/novelcommittee/pages/default.aspx.
34. [FSANZ] Food Standards Austalia New Zealand. *Record of views formed in response to inquiries*. 2020 201119 [cited 2020 201127]; Available from: https://www.foodstandards.gov.au/industry/novel/novelrecs/Documents/record-of-views-19nov20.pdf.
35. [ANVISA] Agência Nacional de Vigilância Sanitária. *RESOLUÇÃO Nº 16, DE 30 DE ABRIL DE 1999*. 1999 [cited 2020 201127]; Available from: https://www.gov.br/agricultura/pt-br/assuntos/inspecao/produtos-vegetal/legislacao-1/biblioteca-de-normas-vinhos-e-bebidas/resolucao-no-16-de-30-de-abril-de-1999.pdf/@@download/file/resolucao-no-16-de-30-de-abril-de-1999.pdf.
36. [ANVISA] Agência Nacional de Vigilância Sanitária. *RESOLUÇÃO Nº 17, DE 30 DE ABRIL DE 1999*. 1999 [cited 2020 201127]; Available from: https://www.gov.br/agricultura/pt-br/assuntos/inspecao/produtos-vegetal/legislacao-1/biblioteca-de-normas-vinhos-e-bebidas/resolucao-no-17-de-30-de-abril-de-1999.pdf/@@download/file/resolucao-no-17-de-30-de-abril-de-1999.pdf.
37. [ANVISA] Agência Nacional de Vigilância Sanitária. *Guia para Comprovação da Segurança de Alimentos e Ingredientes*. 2013 [cited 2020 201127]; Available from: https://alimentosprocessados.com.br/arquivos/Seguranca-e-qualidade-dos-alimentos/Guia-para-Comprovacao-da-Seguranca-de-Alimentos-e-Ingredientes-ANVISA.pdf.
38. [ANVISA] Agência Nacional de Vigilância Sanitária. *Novos alimentos e ingredientes: Documento de base para discussão regulatória*. 2020 [cited 2020 201127]; Available from: http://antigo.anvisa.gov.br/documents/33880/5833856/Documento+de+base+sobre+novos+alimentos/ed783550-fc93-42c2-91cc-ccbd02c36fc9.
39. Department of Justice Canada. *Consolidation; Food and Drug Regulations (C.R.C., c. 870).* . 2020; Available from: https://laws.justice.gc.ca/PDF/C.R.C., c. 870.pdf.
40. Food Directorate Health Products and Food Branch of Health Canada. *Guidelines for the Safety Assessment of Novel Foods*. 2006 [cited 2020 201129]; Available from: https://www.canada.ca/content/dam/hc-sc/migration/hc-sc/fn-an/alt_formats/hpfb-dgpsa/pdf/gmf-agm/guidelines-lignesdirectrices-eng.pdf.
41. Food Directorate Health Products and Food Branch of Health Canada. *The Food Directorate's Pre-Market Submission Management Process for Food Additives, Infant Formulas and Novel Foods*. 2016 [cited 2020 201129]; Available from:

https://www.canada.ca/content/dam/hc-sc/migration/hc-sc/fn-an/alt_formats/pdf/consult/additives-additifs-eng.pdf.

42. Hu, C., *Chapter 30 - Change in raw materials and claims of health food regulations in China 2013–17*, in *Nutraceutical and Functional Food Regulations in the United States and around the World (Third Edition)*, D. Bagchi, Editor. 2019, Academic Press. p. 465-495.
43. Urazbaeva, A.N., *RUSSIAN FOOD LAW, Legal Systems of the Russian Federation and the Eurasian Economic Union (EAEU), and Food Safety Technical Regulation* in *LAW AND GOVERNANCE GROUP*. 2018, WAGENINGEN UNIVERSITY AND RESEARCH. p. 228.
44. [EEC] Eurasian Economic Commission. *Технический регламент Таможенного союза ТР ТС 021/2011*. 2019 [cited 2020 201128]; Available from: http://docs.cntd.ru/document/902320560.
45. International Finance Corporation, *Comparative Analysis of Certain Requirements of Food Legislation in the European Union and the Customs Union of Russia, Belarus, and Kazakhstan.* 2015.
46. *Regulation (EU) 2015/2283 of the European Parliament and of the Council of 25 November 2015 on novel foods, amending Regulation (EU) No 1169/2011 of the European Parliament and of the Council and repealing Regulation (EC) No 258/97 of the European Parliament and of the Council and Commission Regulation (EC) No 1852/2001*. Official Journal of the European Union 2015. **327**: p. 1-22.
47. *Commission Implementing Regulation (EU) 2017/2469 of 20 December 2017 laying down administrative and scientific requirements for applications referred to in Article 10 of Regulation (EU) 2015/2283 of the European Parliament and of the Council on novel foods.* Official Journal of the European Union, 2017. **351**: p. 64-71.
48. Chakraborty, P., *Chapter 27 - Road map to enter Indian nutraceuticals and health supplements business*, in *Nutraceutical and Functional Food Regulations in the United States and around the World (Third Edition)*, D. Bagchi, Editor. 2019, Academic Press. p. 417-430.
49. *Food Safety and Standards (Health Supplements, Nutraceuticals, Food for Special Dietary Use, Food for Special Medical Purpose, Functional Food and Novel Food) Regulations, 2016.* 2016.
50. *Food Safety and Standards (Approval of non-specified food and food ingredients) Regulations, 2017*. 2017.
51. *Food Sanitation Act*. 2020 [cited 2020 201128]; Available from: http://www.japaneselawtranslation.go.jp/law/detail_main?id=12&vm=2&re.
52. *SFDA.FD xxx/2019 - General Requirements for Novel Foods*. 2019.
53. *FIRST PUBLIC CONSULTATION ON PROPOSED REGULATORY FRAMEWORK FOR NOVEL FOOD AND NOVEL FOOD INGREDIENTS IN SINGAPORE*. 2019.
54. Frestedt, J.L., *Chapter 16 - Foods, Food Additives, and Generally Regarded as Safe (GRAS) Food Assessments*, in *Food Control and Biosecurity*, A.M. Holban and A.M. Grumezescu, Editors. 2018, Academic Press. p. 543-565.

55. von Mutius, E., *1 - Epidemiology of Allergic Diseases*, in *Pediatric Allergy: Principles and Practice (Third Edition)*, D.Y.M. Leung, et al., Editors. 2016, Elsevier: London. p. 1-6.e2.
56. Custovic, A., *Chapter 3 - Epidemiology of Allergic Diseases*, in *Middleton's Allergy Essentials*, R.E. O'Hehir, S.T. Holgate, and A. Sheikh, Editors. 2017, Elsevier. p. 51-72.
57. Weber, R.W., *20 - Outdoor Allergens*, in *Pediatric Allergy: Principles and Practice (Third Edition)*, D.Y.M. Leung, et al., Editors. 2016, Elsevier: London. p. 185-190.e1.
58. Sheehan, W.J. and W. Phipatanakul, *21 - Indoor Allergens*, in *Pediatric Allergy: Principles and Practice (Third Edition)*, D.Y.M. Leung, et al., Editors. 2016, Elsevier: London. p. 191-195.e2.
59. Golden, D.B.K., *Chapter 15 - Insect Allergy*, in *Middleton's Allergy Essentials*, R.E. O'Hehir, S.T. Holgate, and A. Sheikh, Editors. 2017, Elsevier. p. 377-393.
60. Taylor, S.L. and J.L. Baumert, *Chapter 29 - Food allergies, sensitivities, and intolerances*, in *Present Knowledge in Nutrition (Eleventh Edition)*, B.P. Marriott, et al., Editors. 2020, Academic Press. p. 535-548.
61. Bird, J.A., S. Jones, and W. Burks, *45 - Food allergy*, in *Clinical Immunology*. 2019, Elsevier. p. 625-631. e1.
62. Ward, R.K., *1 - Introduction to food allergy*, in *Handbook of Food Allergen Detection and Control*, S. Flanagan, Editor. 2015, Woodhead Publishing. p. 1-15.
63. Dubois, A.E., et al., *How does dose impact on the severity of food-induced allergic reactions, and can this improve risk assessment for allergenic foods? Report from an ILSI Europe Food Allergy Task Force Expert Group and Workshop.* Allergy, 2018. **73**(7): p. 1383-1392.
64. Du Toit, G., et al., *Randomized trial of peanut consumption in infants at risk for peanut allergy.* N Engl J Med, 2015. **372**: p. 803-813.
65. Nowak, E. and B. Schaub, *Chapter 11 - Prevention of Allergies*, in *Implementing Precision Medicine in Best Practices of Chronic Airway Diseases*, I. Agache and P. Hellings, Editors. 2019, Academic Press. p. 63-71.
66. Holloway, J.W. and S.L. Prescott, *Chapter 2 - The Origins of Allergic Disease*, in *Middleton's Allergy Essentials*, R.E. O'Hehir, S.T. Holgate, and A. Sheikh, Editors. 2017, Elsevier. p. 29-50.
67. Tsakok, T., et al., *Does atopic dermatitis cause food allergy? A systematic review.* Journal of Allergy and Clinical Immunology, 2016. **137**(4): p. 1071-1078.
68. Groetch, M. and H.A. Sampson, *48 - Management of Food Allergy*, in *Pediatric Allergy: Principles and Practice (Third Edition)*, D.Y.M. Leung, et al., Editors. 2016, Elsevier: London. p. 420-429.e1.
69. Jones, S.M. and A.W. Burks, *49 - Immunotherapeutic Approaches to the Treatment of Food Allergy*, in *Pediatric Allergy: Principles and Practice (Third Edition)*, D.Y.M. Leung, et al., Editors. 2016, Elsevier: London. p. 430-437.e3.
70. Murphy, K. and C. Weaver, *Janeway's Immunobiology, 9th edition*. 2016.
71. Sakaguchi, S., et al., *Regulatory T cells and immune tolerance.* Cell, 2008. **133**(5): p. 775-787.
72. Palomares, O., *The role of regulatory T cells in IgE-mediated food allergy.* J Investig Allergol Clin Immunol, 2013. **23**(6): p. 371-82.

73. Wieser, H., P. Koehler, and K.A. Scherf, *Chapter 8 - Celiac disease*, in *Wheat - An Exceptional Crop*, H. Wieser, P. Koehler, and K.A. Scherf, Editors. 2020, Woodhead Publishing. p. 161-220.
74. Fernández-Rivas, M. and B. Ballmer-Weber, *1 - Food allergy: current diagnosis and management*, in *Managing Allergens in Food*, C. Mills, H. Wichers, and K. Hoffmann-Sommergruber, Editors. 2007, Woodhead Publishing. p. 3-28.
75. Sicherer, S.H. and H.A. Sampson, *Food allergy: a review and update on epidemiology, pathogenesis, diagnosis, prevention, and management.* Journal of Allergy and Clinical Immunology, 2018. **141**(1): p. 41-58.
76. Hamilton, R.G., *18 - Laboratory Diagnosis of Human Allergic Disease*, in *Pediatric Allergy: Principles and Practice (Third Edition)*, D.Y.M. Leung, et al., Editors. 2016, Elsevier: London. p. 167-176.e1.
77. Chiriac, A.M., J. Bousquet, and P. Demoly, *Chapter 5 - Principles of Allergy Diagnosis*, in *Middleton's Allergy Essentials*, R.E. O'Hehir, S.T. Holgate, and A. Sheikh, Editors. 2017, Elsevier. p. 117-131.
78. Gupta, M., et al., *Diagnosis of food allergy.* Immunology and Allergy Clinics, 2018. **38**(1): p. 39-52.
79. Yue, D., et al., *Food allergy and anaphylaxis.* Journal of Asthma and Allergy, 2018. **11**: p. 111.
80. Crevel, R. and S. Cochrane, *Chapter 4 - Allergens*, in *Food Safety Management*, Y. Motarjemi and H. Lelieveld, Editors. 2014, Academic Press: San Diego. p. 59-82.
81. Gupta, R.S., et al., *Prevalence and severity of food allergies among US adults.* JAMA network open, 2019. **2**(1): p. e185630-e185630.
82. Burney, P., et al., *Chapter 3 - The Epidemiology of Food Allergy*, in *Risk Management for Food Allergy*, C.B. Madsen, et al., Editors. 2014, Academic Press: San Diego. p. 45-64.
83. Loh, W. and M.L. Tang, *The epidemiology of food allergy in the global context.* International journal of environmental research and public health, 2018. **15**(9): p. 2043.
84. Tang, M.L. and R.J. Mullins, *Food allergy: is prevalence increasing?* Internal medicine journal, 2017. **47**(3): p. 256-261.
85. Warren, C.M., J. Jiang, and R.S. Gupta, *Epidemiology and Burden of Food Allergy.* Current allergy and asthma reports, 2020. **20**(2): p. 1-9.
86. Bilaver, L.A., et al., *Economic burden of food allergy: a systematic review.* Annals of Allergy, Asthma & Immunology, 2019. **122**(4): p. 373-380. e1.
87. Radauer, C., *Nomenclature of Food Allergens*, in *Food Allergy*. 2017, CRC Press. p. 30-40.
88. Fu, L., et al., *Species and Structure of Food Allergens: Epitopes and Cross-Reactivity*, in *Food Allergy*. 2019, Springer. p. 13-39.
89. van Ree, R., *5 - Cross-reactivity of plant allergens*, in *Managing Allergens in Food*, C. Mills, H. Wichers, and K. Hoffmann-Sommergruber, Editors. 2007, Woodhead Publishing. p. 83-90.
90. Breiteneder, H., *Biomolecular and Clinical Aspects of Food Allergy*, in *Food Allergy*. 2017, CRC Press. p. 1-29.

91. Sicherer, S.H., *Clinical implications of cross-reactive food allergens.* Journal of Allergy and Clinical Immunology, 2001. **108**(6): p. 881-890.
92. Popescu, F.-D., *Cross-reactivity between aeroallergens and food allergens.* World journal of methodology, 2015. **5**(2): p. 31.
93. Fu, L., et al., *Allergenicity Evaluation of Food Proteins*, in *Food Allergy*. 2019, Springer. p. 93-122.
94. Breiteneder, H. and E.N.C. Mills, *Molecular properties of food allergens.* Journal of Allergy and Clinical Immunology, 2005. **115**(1): p. 14-23.
95. Goodman, R.E., et al., *AllergenOnline: a peer-reviewed, curated allergen database to assess novel food proteins for potential cross-reactivity.* Molecular nutrition & food research, 2016. **60**(5): p. 1183-1198.
96. Maurer-Stroh, S., et al., *AllerCatPro—prediction of protein allergenicity potential from the protein sequence.* Bioinformatics, 2019. **35**(17): p. 3020-3027.
97. Jespersen, M.C., et al., *BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes.* Nucleic acids research, 2017. **45**(W1): p. W24-W29.
98. Fu, L., et al., *Food Processing to Eliminate Food Allergens and Development of Hypoallergenic Foods*, in *Food Allergy*. 2019, Springer. p. 123-146.
99. Okolie, C.L., A.N.A. Aryee, and C.C. Udenigwe, *14 - Detection and deactivation of allergens in food*, in *Proteins in Food Processing (Second Edition)*, R.Y. Yada, Editor. 2018, Woodhead Publishing. p. 367-387.
100. Baumert, J.L., *Chapter 13 - Detecting and Measuring Allergens in Food*, in *Risk Management for Food Allergy*, C.B. Madsen, et al., Editors. 2014, Academic Press: San Diego. p. 215-226.
101. Holzhauser, T., et al., *Are current analytical methods suitable to verify VITAL® 2.0/3.0 Allergen Reference doses for EU Allergens in Foods?* Food and Chemical Toxicology, 2020: p. 111709.
102. Holzhauser, T. and M. Röder, *Polymerase chain reaction (PCR) methods for detecting allergens in foods*, in *Handbook of food allergen detection and control*. 2015, Elsevier. p. 245-263.
103. Monaci, L., et al., *Food allergens: Classification, molecular properties, characterization, and detection in food sources.* Advances in Food and Nutrition Research, 2020. **93**: p. 113-146.
104. Fu, L., et al., *Detection and Quantification Methods for Food Allergens*, in *Food Allergy*. 2019, Springer. p. 69-91.
105. Pilolli, R., E. De Angelis, and L. Monaci, *19 - Advances in MS methods for food allergens detection*, in *Chemical Analysis of Food (Second Edition)*, Y. Pico, Editor. 2020, Academic Press. p. 787-811.
106. Marzano, V., et al., *Perusal of food allergens analysis by mass spectrometry-based proteomics.* Journal of Proteomics, 2020. **215**: p. 103636.
107. FAO and WHO, *Codex Alimentarius Commission – Procedural Manual twenty-seventh edition*. 27th ed. 2019, Rome. 254.
108. Crevel, R.W.R., *3 - Food allergen risk assessment and management*, in *Handbook of Food Allergen Detection and Control*, S. Flanagan, Editor. 2015, Woodhead Publishing. p. 41-66.

109. Flanagan, S., *4 - Assessment and communication of allergen risks in the food chain*, in *Handbook of Food Allergen Detection and Control*, S. Flanagan, Editor. 2015, Woodhead Publishing. p. 67-87.
110. Crevel, R.W.R., et al., *Chapter 5 - Thresholds or 'How Much Is Too Much?'*, in *Risk Management for Food Allergy*, C.B. Madsen, et al., Editors. 2014, Academic Press: San Diego. p. 77-99.
111. Blom, W.M., et al., *Allergen risk assessment: Food intake levels of the general population represent those of food allergic patients.* Food and Chemical Toxicology, 2020. **146**: p. 111781.
112. Crump, K.S., *A new method for determining allowable daily intakes.* Toxicological Sciences, 1984. **4**(5): p. 854-871.
113. Spanjersberg, M., et al., *Risk assessment and food allergy: the probabilistic model applied to allergens.* Food and chemical toxicology, 2007. **45**(1): p. 49-54.
114. Pener, M.P., *Allergy to Crickets: A Review.* Journal of Orthoptera Research, 2016. **25**(2): p. 91-95.
115. Piromrat, K., S. Chinratanapisit, and S. Trathong, *Anaphylaxis in an emergency department: a 2-year study in a tertiary-care hospital.* Asian Pac J Allergy Immunol, 2008. **26**(2-3): p. 121-8.
116. Prasad, R., et al., *Study of skin sensitivity to various allergens by intradermal test in patients of bronchial asthma.* Indian J Allergy Asthma Immunol, 2001. **15**(1): p. 17-21.
117. Prasad, R., et al., *A study of skin sensitivity to various allergens by skin prick test in patients of nasobronchial allergy.* Lung India : Official Organ of Indian Chest Society, 2009. **26**(3): p. 70-73.
118. Jirapongsananuruk, O., et al., *Features of patients with anaphylaxis admitted to a university hospital.* Ann Allergy Asthma Immunol, 2007. **98**(2): p. 157-62.
119. Ji, K., et al., *Anaphylactic shock and lethal anaphylaxis caused by food consumption in China.* Trends in Food Science & Technology, 2009. **20**(5): p. 227-231.
120. Popescu, F.D., *Cross-reactivity between aeroallergens and food allergens.* World J Methodol, 2015. **5**(2): p. 31-50.
121. Werfel, T., *Food allergy.* J Dtsch Dermatol Ges, 2008. **6**(7): p. 573-83.
122. Ramirez, D.A. and S.L. Bahna, *Food hypersensitivity by inhalation.* Clinical and Molecular Allergy : CMA, 2009. **7**: p. 4-4.
123. Adkinson, N.F., et al., *Middleton's Allergy E-Book: Principles and Practice*. 2013: Elsevier Health Sciences.
124. Pener, M.P., *Allergy to Locusts and Acridid Grasshoppers: a Review.* Journal of Orthoptera Research, 2014. **23**(1): p. 59-67.
125. Harfi, H.A., *Immediate hypersensitivity to cricket.* Annals of Allergy, 1980. **44**(3): p. 162-163.
126. Park, M., et al., *Hypersensitivity Pneumonitis Caused by House Cricket, Acheta domesticus.* Journal of Clinical & Cellular Immunology, 2014. **5**: p. 248.
127. Bartra, J., et al., *Occupational asthma and rhinoconjunctivitis caused by cricket allergy.* J Investig Allergol Clin Immunol, 2008. **18**(2): p. 141-2.

128. Rauschenberg, R., et al., *Occupational immediate-type allergy to locusts in a zookeeper.* JDDG: Journal der Deutschen Dermatologischen Gesellschaft, 2015. **13**(2): p. 157-158.
129. Lopata, A.L., et al., *Occupational allergy in laboratory workers caused by the African migratory grasshopper Locusta migratoria.* Allergy, 2005. **60**(2): p. 200-5.
130. Tee, R.D., et al., *Occupational allergy to locusts: an investigation of the sources of the allergen.* J Allergy Clin Immunol, 1988. **81**(3): p. 517-25.
131. Linares, T., D. Hernandez, and B. Bartolome, *Occupational rhinitis and asthma due to crickets.* Annals of Allergy, Asthma & Immunology, 2008. **100**(6): p. 566-569.
132. Lierl, M.B., M.M. Riordan, and T.J. Fischer, *Prevalence of insect allergen-specific IgE in allergic asthmatic children in Cincinnati, Ohio.* Ann Allergy, 1994. **72**(1): p. 45-50.
133. Srinroch, C., et al., *Identification of novel allergen in edible insect, Gryllus bimaculatus and its cross-reactivity with Macrobrachium spp. allergens.* Food Chemistry, 2015. **184**: p. 160-166.
134. Phiriyangkul, P., et al., *Effect of Food Thermal Processing on Allergenicity Proteins in Bombay Locust (Patanga Succincta).* ETP International Journal of Food Engineering, 2015.
135. Verhoeckx, K.C., et al., *House dust mite (Der p 10) and crustacean allergic patients may react to food containing Yellow mealworm proteins.* Food Chem Toxicol, 2014. **65**: p. 364-73.
136. Broekman, H., et al., *Effect of thermal processing on mealworm allergenicity.* Molecular Nutrition & Food Research, 2015. **59**(9): p. 1855-1864.
137. Broekman, H., et al., *Majority of shrimp-allergic patients are allergic to mealworm.* Journal of Allergy and Clinical Immunology, 2016. **137**(4): p. 1261-1263.
138. Taylor, S.L., et al., *Establishment of Reference Doses for residues of allergenic foods: Report of the VITAL Expert Panel.* Food and Chemical Toxicology, 2014. **63**: p. 9-17.
139. Broekman, H.C.H.P., et al., *Primary respiratory and food allergy to mealworm.* Journal of Allergy and Clinical Immunology, 2017.
140. Andersen, S.O., et al., *Comparison of larval and pupal cuticular proteins in Tenebrio molitor.* Insect Biochemistry and Molecular Biology, 1995. **25**(2): p. 177-187.
141. Barre, A., et al., *Les allergènes croisants des insectes comestibles.* Revue Française d'Allergologie, 2016. **56**(7–8): p. 522-532.
142. Pomés, A., et al., *New Insights into Cockroach Allergens.* Current Allergy and Asthma Reports, 2017. **17**(4): p. 25.
143. Shahraki, G.H., S. Parhizkar, and A.R.S. Nejad, *Cockroach Infestation and Factors Affecting the Estimation of Cockroach Population in Urban Communities.* International Journal of Zoology, 2013. **2013**: p. 6.
144. Lombardi, C., et al., *Is allergic sensitization relevant in severe asthma? Which allergens may be culprits?* World Allergy Organization Journal, 2017. **10**(1): p. 2.

145. Sohn, M.H. and K.E. Kim, *The cockroach and allergic diseases.* Allergy Asthma Immunol Res, 2012. **4**(5): p. 264-9.
146. Bassirpour, G. and E. Zoratti, *Cockroach allergy and allergen-specific immunotherapy in asthma: Potential and Pitfalls.* Current opinion in allergy and clinical immunology, 2014. **14**(6): p. 535-541.
147. Lee, M.F., et al., *Analysis of 10 Environmental Allergen Components of the American Cockroach in Taiwan.* Ann Allergy Asthma Immunol, 2016. **117**(5): p. 535-541.e1.
148. Pomés, A. and L.K. Arruda, *Investigating cockroach allergens: Aiming to improve diagnosis and treatment of cockroach allergic patients.* Methods, 2014. **66**(1): p. 75-85.
149. Wang, J., et al., *Correlation of specific IgE to shrimp with cockroach and dust mite exposure and sensitization in an inner-city population.* J Allergy Clin Immunol, 2011. **128**(4): p. 834-7.
150. Hilger, C., et al., *Cockroach, Tick, Storage Mite, and Other Arthropod Allergies: Molecular Aspects,* in *Molecular Allergy Diagnostics: Innovation for a Better Patient Management*, J. Kleine-Tebbe and T. Jakob, Editors. 2017, Springer International Publishing: Cham. p. 429-444.
151. Portnoy, J., et al., *Environmental assessment and exposure reduction of cockroaches: A practice parameter.* The Journal of allergy and clinical immunology, 2013. **132**(4): p. 10.1016/j.jaci.2013.04.061.
152. Ayuso, R., et al., *Molecular basis of arthropod cross-reactivity: IgE-binding cross-reactive epitopes of shrimp, house dust mite and cockroach tropomyosins.* Int Arch Allergy Immunol, 2002. **129**(1): p. 38-48.
153. Crespo, J.F., et al., *Cross-reactivity of IgE-binding components between boiled Atlantic shrimp and German cockroach.* Allergy, 1995. **50**(11): p. 918-24.
154. Witteman, A.M., et al., *Identification of a cross-reactive allergen (presumably tropomyosin) in shrimp, mite and insects.* Int Arch Allergy Immunol, 1994. **105**(1): p. 56-61.
155. Sicherer, S.H. and H.A. Sampson, *Food allergy: Epidemiology, pathogenesis, diagnosis, and treatment.* J Allergy Clin Immunol, 2014. **133**(2): p. 291-307; quiz 308.
156. Waserman, S. and W. Watson, *Food allergy.* Allergy Asthma Clin Immunol, 2011. **7 Suppl 1**: p. S7.
157. Fernandes, J., et al., *Immunoglobulin E antibody reactivity to the major shrimp allergen, tropomyosin, in unexposed Orthodox Jews.* Clinical & Experimental Allergy, 2003. **33**(7): p. 956-961.
158. Faber, M.A., et al., *Shellfish allergens: tropomyosin and beyond.* Allergy, 2017. **72**(6): p. 842-848.
159. Pascal, M., et al., *Molecular Diagnosis of Shrimp Allergy: Efficiency of Several Allergens to Predict Clinical Reactivity.* J Allergy Clin Immunol Pract, 2015. **3**(4): p. 521-9.e10.
160. Binder, M., et al., *Molecular and immunological characterization of arginine kinase from the Indianmeal moth, Plodia interpunctella, a novel cross-reactive invertebrate pan-allergen.* The Journal of Immunology, 2001. **167**(9): p. 5470-5477.

161. Sookrung, N., et al., *Periplaneta americana Arginine Kinase as a Major Cockroach Allergen among Thai Patients with Major Cockroach Allergies.* Environmental Health Perspectives, 2006. **114**(6): p. 875-880.
162. Brown, A.E., R.M. France, and S.H. Grossman, *Purification and characterization of arginine kinase from the American cockroach (Periplaneta americana).* Arch Insect Biochem Physiol, 2004. **56**(2): p. 51-60.
163. Chen, H.L., et al., *Purification, physicochemical and immunological characterization of arginine kinase, an allergen of crayfish (Procambarus clarkii).* Food Chem Toxicol, 2013. **62**: p. 475-84.
164. Yu, C.J., et al., *Proteomics and immunological analysis of a novel shrimp allergen, Pen m 2.* J Immunol, 2003. **170**(1): p. 445-53.
165. Asturias, J.A., et al., *Molecular characterization of American cockroach tropomyosin (Periplaneta americana allergen 7), a cross-reactive allergen.* J Immunol, 1999. **162**(7): p. 4342-8.
166. Santos, A.B., et al., *Cockroach allergens and asthma in Brazil: identification of tropomyosin as a major allergen with potential cross-reactivity with mite and shrimp allergens.* J Allergy Clin Immunol, 1999. **104**(2 Pt 1): p. 329-37.
167. Jeong, K.Y., et al., *Allergenicity of recombinant Bla g 7, German cockroach tropomyosin.* Allergy, 2003. **58**(10): p. 1059-1063.
168. Shanti, K.N., et al., *Identification of tropomyosin as the major shrimp allergen and characterization of its IgE-binding epitopes.* The Journal of Immunology, 1993. **151**(10): p. 5354-63.
169. Reese, G., et al., *IgE and monoclonal antibody reactivities to the major shrimp allergen Pen a 1 (tropomyosin) and vertebrate tropomyosins.* Adv Exp Med Biol, 1996. **409**: p. 225-30.
170. Abdel Rahman, A.M., et al., *Absolute quantification method and validation of airborne snow crab allergen tropomyosin using tandem mass spectrometry.* Anal Chim Acta, 2010. **681**(1-2): p. 49-55.
171. Lopata, A.L., R.E. O'Hehir, and S.B. Lehrer, *Shellfish allergy.* Clin Exp Allergy, 2010. **40**(6): p. 850-8.
172. Leung, P.S.C., et al., *IgE reactivity against a cross-reactive allergen in crustacea and mollusca: Evidence for tropomyosin as the common allergen.* Journal of Allergy and Clinical Immunology, 1996. **98**(5, Part 1): p. 954-961.
173. Reese, G., R. Ayuso, and S.B. Lehrer, *Tropomyosin: an invertebrate pan-allergen.* Int Arch Allergy Immunol, 1999. **119**(4): p. 247-58.
174. Belluco, S., et al., *Edible Insects in a Food Safety and Nutritional Perspective: A Critical Review.* Comprehensive Reviews in Food Science and Food Safety, 2013. **12**(3): p. 296-313.
175. Kamath, S.D., et al., *Impact of heat processing on the detection of the major shellfish allergen tropomyosin in crustaceans and molluscs using specific monoclonal antibodies.* Food Chem, 2013. **141**(4): p. 4031-9.
176. Long, F., et al., *Effects of combined high pressure and thermal treatments on the allergenic potential of shrimp (Litopenaeus vannamei) tropomyosin in a mouse model of allergy.* Innovative Food Science & Emerging Technologies, 2015. **29**: p. 119-124.

177. Liu, G.-M., et al., *Comparative study of in vitro digestibility of major allergen, tropomyosin and other proteins between Grass prawn (Penaeus monodon) and Pacific white shrimp (Litopenaeus vannamei).* Journal of the Science of Food and Agriculture, 2011. **91**(1): p. 163-170.
178. Yu, H.L., et al., *Effects of different processing methods on digestibility of Scylla paramamosain allergen (tropomyosin).* Food Chem Toxicol, 2011. **49**(4): p. 791-8.
179. Burmester, T., *Identification, Molecular Cloning, and Phylogenetic Analysis of a Non-respiratory Pseudo-hemocyanin of Homarus americanus.* Journal of Biological Chemistry, 1999. **274**(19): p. 13217-13222.
180. Hughes, A.L., *Evolution of the arthropod prophenoloxidase/hexamerin protein family.* Immunogenetics, 1999. **49**(2): p. 106-14.
181. Giuffrida, M.G., et al., *Shrimp allergy beyond Tropomyosin in Italy: clinical relevance of Arginine Kinase, Sarcoplasmic calcium binding protein and Hemocyanin.* Eur Ann Allergy Clin Immunol, 2014. **46**(5): p. 172-7.
182. van Broekhoven, S., et al., *Influence of processing and in vitro digestion on the allergic cross-reactivity of three mealworm species.* Food Chemistry, 2016. **196**: p. 1075-1083.
183. Piboonpocanun, S., et al., *Identification of hemocyanin as a novel non-cross-reactive allergen from the giant freshwater shrimp Macrobrachium rosenbergii.* Mol Nutr Food Res, 2011. **55**(10): p. 1492-8.
184. Ayuso, R., et al., *Hemocyanin, Troponin C and Fatty Acid-binding Protein (FABP) May be Cross-reactive Allergens Between Crustaceans, Cockroach and Dust Mites.* Journal of Allergy and Clinical Immunology, 2011. **127**(2): p. AB235.
185. Bertelsen, R.J., et al., *Pulmonary illness as a consequence of occupational exposure to shrimp shell powder.* Environ Res, 2016. **148**: p. 491-9.
186. Da Silva, C.A., et al., *Chitin particles are multifaceted immune adjuvants.* Am J Respir Crit Care Med, 2010. **182**(12): p. 1482-91.
187. Nam, Y.S., et al., *Effect of the degree of deacetylation on the thermal decomposition of chitin and chitosan nanofibers.* Carbohydrate Polymers, 2010. **80**(1): p. 291-295.
188. Dubey, L.K., et al., *Chitin enhances serum IgE in Aspergillus fumigatus induced allergy in mice.* Immunobiology, 2015. **220**(6): p. 714-21.
189. Bae, M.J., et al., *Oral administration of chitin and chitosan prevents peanut-induced anaphylaxis in a murine food allergy model.* Int J Biol Macromol, 2013. **61**: p. 164-8.
190. Patel, S. and A. Goyal, *Chitin and chitinase: Role in pathogenicity, allergenicity and health.* International Journal of Biological Macromolecules, 2017. **97**: p. 331-338.
191. Wang, D., et al., *NUTRITIONAL VALUE OF THE FIELD CRICKET (GRYLLUS TESTACEUS WALKER).* Insect Science, 2004. **11**(4): p. 275-283.
192. Kulma, M., et al., *Nutritional value of three Blattodea species used as feed for animals.* Journal of Animal and Feed Sciences, 2016. **25**(4): p. 354-360.

193. Kaya, M., et al., *Comparison of chitin structures isolated from seven Orthoptera species.* International Journal of Biological Macromolecules, 2015. **72**: p. 797-805.
194. FAO, *Edible Insects: Future Prospects for Food and Feed Security*. 2013: Food and Agriculture Organization of the United Nations.
195. Finke, M.D., *Gut loading to enhance the nutrient content of insects as food for reptiles: A mathematical approach.* Zoo Biology, 2003. **22**(2): p. 147-162.
196. Yates-Doerr, E., *The world in a box? Food security, edible insects, and "One World, One Health" collaboration.* Soc Sci Med, 2015. **129**: p. 106-12.
197. Hanboonsong, Y., T. Jamjanya, and P.B. Durst, *Six-legged livestock: edible insect farming, collection and marketing in Thailand*. 2013.
198. Rich, K., *Industry & Trade Summary Animal Feeds*, U.S.I.T. Commission, Editor. 2000.
199. Chen, H., et al., *Stability of plant defense proteins in the gut of insect herbivores.* Plant Physiol, 2007. **143**(4): p. 1954-67.
200. Saadati, M. and M. Toorchi, *The study of plant protein accumulation in gut of insect using proteomics technique: Wheat–sunn pest interaction.* Journal of the Saudi Society of Agricultural Sciences, 2017. **16**(3): p. 205-209.
201. Koppelman, S.J., et al., *Allergenicity attributes of different peanut market types.* Food and Chemical Toxicology, 2016. **91**: p. 82-90.
202. Jiang, J., et al., *Updated pediatric peanut allergy prevalence in the United States.* Annals of Allergy, Asthma & Immunology, 2018. **121**(5): p. S14.
203. Lieberman, J., et al., *Increased incidence and prevalence of peanut allergy in children and adolescents in the United States.* Annals of Allergy, Asthma & Immunology, 2018. **121**(5): p. S13.
204. *Food Allergen Labeling and Consumer Protection Act of 2004*, in *21*. 2004.
205. Dash, S., et al., *PeanutBase and other bioinformatic resources for peanut*, in *Peanuts*. 2016, Elsevier. p. 241-252.
206. Bertioli, D.J., et al., *The genome sequences of Arachis duranensis and Arachis ipaensis, the diploid ancestors of cultivated peanut.* Nature Genetics, 2016. **48**(4): p. 438-446.
207. Bertioli, D.J., et al., *The genome sequence of segmental allotetraploid peanut Arachis hypogaea.* Nature Genetics, 2019. **51**(5): p. 877-884.
208. Li, H., et al., *Improving Gene Annotation of the Peanut Genome by Integrated Proteogenomics Workflow.* Journal of Proteome Research, 2020. **19**(6): p. 2226-2235.
209. Marsh, J.T., et al., *Thermal processing of peanut impacts detection by current analytical techniques.* Food Chemistry, 2020. **313**: p. 126019.
210. Shewry, P.R., J.A. Napier, and A.S. Tatham, *Seed storage proteins: structures and biosynthesis.* The Plant cell, 1995. **7**(7): p. 945-956.
211. Moreno, F.J. and A. Clemente, *2S Albumin Storage Proteins: What Makes them Food Allergens?* The open biochemistry journal, 2008. **2**: p. 16-28.
212. Nesbit, J.B., et al., *Epitopes with similar physicochemical properties contribute to cross reactivity between peanut and tree nuts.* Molecular Immunology, 2020. **122**: p. 223-231.

213. Barre, A., J.-P. Borges, and P. Rougé, *Molecular modelling of the major peanut allergen Ara h 1 and other homotrimeric allergens of the cupin superfamily: a structural basis for their IgE-binding cross-reactivity.* Biochimie, 2005. **87**(6): p. 499-506.
214. van Boxtel, E.L., et al., *Determination of Pepsin-Susceptible and Pepsin-Resistant Epitopes in Native and Heat-Treated Peanut Allergen Ara h 1.* Journal of Agricultural and Food Chemistry, 2008. **56**(6): p. 2223-2230.
215. Cabanos, C., et al., *Crystal structure of the major peanut allergen Ara h 1.* Molecular Immunology, 2011. **49**(1): p. 115-123.
216. Maleki, S.J., et al., *Structure of the Major Peanut Allergen Ara h 1 May Protect IgE-Binding Epitopes from Degradation.* The Journal of Immunology, 2000. **164**(11): p. 5844-5849.
217. Koppelman, S.J., et al., *Peanut allergen Ara h 3: Isolation from peanuts and biochemical characterization.* Allergy, 2003. **58**(11): p. 1144-1151.
218. Dodo, H.W., et al., *cDNA Clone of a Putative Peanut (Arachis hypogaea L.) Trypsin Inhibitor Has Homology with Peanut Allergens Ara h 3 and Ara h 4.* Journal of Agricultural and Food Chemistry, 2004. **52**(5): p. 1404-1409.
219. Piersma, S.R., et al., *Proteolytic processing of the peanut allergen Ara h 3.* Molecular Nutrition & Food Research, 2005. **49**(8): p. 744-755.
220. Mueller, G.A., et al., *Ara h 2: crystal structure and IgE binding distinguish two subpopulations of peanut allergic patients by epitope diversity.* Allergy, 2011. **66**(7): p. 878-885.
221. Maleki, S.J., et al., *The major peanut allergen, Ara h 2, functions as a trypsin inhibitor, and roasting enhances this function.* Journal of Allergy and Clinical Immunology, 2003. **112**(1): p. 190-195.
222. Bernard, H., et al., *Allergenicity of peanut component Ara h 2: Contribution of conformational versus linear hydroxyproline-containing epitopes.* Journal of Allergy and Clinical Immunology, 2015. **135**(5): p. 1267-1274.e8.
223. Tscheppe, A., et al., *Development of a novel Ara h 2 hypoallergen with no IgE binding or anaphylactogenic activity.* Journal of Allergy and Clinical Immunology, 2020. **145**(1): p. 229-238.
224. Lehmann, K., et al., *Structure and stability of 2S albumin-type peanut allergens: implications for the severity of peanut allergic reactions.* Biochemical Journal, 2006. **395**(3): p. 463-472.
225. de Jong, G.A.H., et al., *Purification and Characterization of Naturally Occurring Post-Translationally Cleaved Ara h 6, an Allergen That Contributes Substantially to the Allergenic Potency of Peanut.* Journal of Agricultural and Food Chemistry, 2018. **66**(41): p. 10855-10863.
226. de Jongh, H.H.J., et al., *Effect of heat treatment on the conformational stability of intact and cleaved forms of the peanut allergen Ara h 6 in relation to its IgE-binding potency.* Food Chemistry, 2020. **326**: p. 127027.
227. Bublin, M., et al., *IgE cross-reactivity between the major peanut allergen Ara h 2 and the nonhomologous allergens Ara h 1 and Ara h 3.* Journal of Allergy and Clinical Immunology, 2013. **132**(1): p. 118-124.e12.

228. Koppelman, S.J., et al., *Purification and immunoglobulin E-binding properties of peanut allergen Ara h 6: evidence for cross-reactivity with Ara h 2.* Clinical & Experimental Allergy, 2005. **35**(4): p. 490-497.
229. Koppelman, S.J., et al., *Digestion of peanut allergens Ara h 1, Ara h 2, Ara h 3, and Ara h 6: A comparative in vitro study and partial characterization of digestion-resistant peptides.* Molecular Nutrition & Food Research, 2010. **54**(12): p. 1711-1721.
230. Koid, A.E., et al., *Ara h 6 Complements Ara h 2 as an Important Marker for IgE Reactivity to Peanut.* Journal of Agricultural and Food Chemistry, 2014. **62**(1): p. 206-213.
231. Blanc, F., et al., *Capacity of purified peanut allergens to induce degranulation in a functional in vitro assay: Ara h 2 and Ara h 6 are the most efficient elicitors.* Clinical & Experimental Allergy, 2009. **39**(8): p. 1277-1285.
232. Porterfield, H.S., et al., *Effector activity of peanut allergens: a critical role for Ara h 2, Ara h 6, and their variants.* Clinical & Experimental Allergy, 2009. **39**(7): p. 1099-1108.
233. Dang, T.D., et al., *Increasing the accuracy of peanut allergy diagnosis by using Ara h 2.* Journal of Allergy and Clinical Immunology, 2012. **129**(4): p. 1056-1063.
234. Hemmings, O., et al., *Ara h 2 is the dominant peanut allergen despite similarities with Ara h 6.* Journal of Allergy and Clinical Immunology, 2020. **146**(3): p. 621-630.e5.
235. Flinterman, A.E., et al., *Children with peanut allergy recognize predominantly Ara h2 and Ara h6, which remains stable over time.* Clinical & Experimental Allergy, 2007. **37**(8): p. 1221-1228.
236. Kukkonen, A.K., et al., *Ara h 2 and Ara 6 are the best predictors of severe peanut allergy: a double-blind placebo-controlled study.* Allergy, 2015. **70**(10): p. 1239-1245.
237. Wang, Y., et al., *Crystal Structure of Peanut (Arachis hypogaea) Allergen Ara h 5.* Journal of Agricultural and Food Chemistry, 2013. **61**(7): p. 1573-1578.
238. Hayen, S.M., et al., *2S protein Ara h 7.0201 has unique epitopes compared to other Ara h 7 isoforms and is comparable to 2S proteins Ara h 2 and 6 in basophil degranulation capacity.* Clinical & Experimental Allergy, 2018. **48**(7): p. 890-897.
239. Schmidt, H., et al., *Detection and Structural Characterization of Natural Ara h 7, the Third Peanut Allergen of the 2S Albumin Family.* Journal of Proteome Research, 2010. **9**(7): p. 3701-3709.
240. Bublin, M. and H. Breiteneder, *Cross-Reactivity of Peanut Allergens.* Current Allergy and Asthma Reports, 2014. **14**(4): p. 426.
241. Riecken, S., et al., *Purification and characterization of natural Ara h 8, the Bet v 1 homologous allergen from peanut, provides a novel isoform.* Biological Chemistry, 2008. **389**(4): p. 415-423.
242. Hurlburt, B.K., et al., *Structure and Function of the Peanut Panallergen Ara h 8.* Journal of Biological Chemistry, 2013. **288**(52): p. 36890-36901.
243. Asarnoj, A., et al., *Peanut component Ara h 8 sensitization and tolerance to peanut.* Journal of Allergy and Clinical Immunology, 2012. **130**(2): p. 468-472.

244. Fleury, C., et al., *Comprehensive classification of the plant non-specific lipid transfer protein superfamily towards its sequence-structure-function analysis.* PeerJ, 2019. **7**: p. e7504-e7504.
245. Krause, S., et al., *Lipid transfer protein (Ara h 9) as a new peanut allergen relevant for a Mediterranean allergic population.* Journal of Allergy and Clinical Immunology, 2009. **124**(4): p. 771-778.e5.
246. Lauer, I., et al., *The non-specific lipid transfer protein, Ara h 9, is an important allergen in peanut.* Clinical & Experimental Allergy, 2009. **39**(9): p. 1427-1437.
247. Schwager, C., et al., *Peanut oleosins associated with severe peanut allergy—Importance of lipophilic allergens for comprehensive allergy diagnostics.* Journal of Allergy and Clinical Immunology, 2017. **140**(5): p. 1331-1338. e8.
248. Majsiak, E., et al., *Oleosins: A Short Allergy Review*. 2020, Springer US: New York, NY. p. 1-5.
249. Pons, L., et al., *The 18 kDa peanut oleosin is a candidate allergen for IgE-mediated reactions to peanuts.* Allergy, 2002. **57**(s72): p. 88-93.
250. Ozias-Akins, P. and H. Breiteneder, *The functional biology of peanut allergens and possible links to their allergenicity.* Allergy, 2019. **74**(5): p. 888-898.
251. Petersen, A., et al., *Peanut defensins: Novel allergens isolated from lipophilic peanut extract.* Journal of Allergy and Clinical Immunology, 2015. **136**(5): p. 1295-1301.e5.
252. Mills, E.N.C., et al., *Impact of food processing on the structural and allergenic properties of food allergens.* Molecular Nutrition & Food Research, 2009. **53**(8): p. 963-969.
253. Vissers, Y.M., et al., *Effect of Heating and Glycation on the Allergenicity of 2S Albumins (Ara h 2/6) from Peanut.* PLOS ONE, 2011. **6**(8): p. e23998.
254. Petersen, A., et al., *Roasting and lipid binding provide allergenic and proteolytic stability to the peanut allergen Ara h 8.* Biological Chemistry, 2014. **395**(2): p. 239-250.
255. Jappe, U. and C. Schwager, *Relevance of Lipophilic Allergens in Food Allergy Diagnosis.* Current Allergy and Asthma Reports, 2017. **17**(9): p. 61.
256. Blanc, F., et al., *Boiling peanut Ara h 1 results in the formation of aggregates with reduced allergenicity.* Molecular Nutrition & Food Research, 2011. **55**(12): p. 1887-1894.
257. Walczyk, N.E., et al., *Peanut protein extraction conditions strongly influence yield of allergens Ara h 1 and 2 and sensitivity of immunoassays.* Food Chemistry, 2017. **221**: p. 335-344.
258. Sayers, R., et al., *The effect of thermal processing on the behaviour of peanut allergen peptide targets used in multiple reaction monitoring mass spectrometry experiments.* Analyst, 2016. **141**(13): p. 4130-4141.
259. Zhang, Y., et al., *Allergenicity assessment on thermally processed peanut influenced by extraction and assessment methods.* Food Chemistry, 2019. **281**: p. 130-139.
260. Sánchez-Borges, M., et al., *Pancake Syndrome (Oral Mite Anaphylaxis).* World Allergy Organization Journal, 2009. **2**(5): p. 91-96.
261. Untersmayr, E. and E. Jensen-Jarolim, *Mechanisms of type I food allergy.* Pharmacology & Therapeutics, 2006. **112**(3): p. 787-798.

262. Erickson, R.H. and Y.S. Kim, *Digestion and absorption of dietary protein.* Annual review of medicine, 1990. **41**: p. 133-139.
263. Reitsma, M., et al., *Protein transport across the small intestine in food allergy.* Molecular Nutrition & Food Research, 2014. **58**(1): p. 194-205.
264. Morishita, M. and N.A. Peppas, *Is the oral route possible for peptide and protein drug delivery?* Drug Discovery Today, 2006. **11**(19): p. 905-910.
265. Warshaw, A.L., W.A. Walker, and K.J. Isselbacher, *Protein Uptake by the Intestine: Evidence for Absorption of Intact Macromolecules.* Gastroenterology, 1974. **66**(5): p. 987-992.
266. Thomas, R.J., *Particle size and pathogenicity in the respiratory tract.* Virulence, 2013. **4**(8): p. 847-858.
267. Kwang-Jin, K. and A.B. Malik, *Protein transport across the lung epithelial barrier.* American Journal of Physiology-Lung Cellular and Molecular Physiology, 2003. **284**(2): p. L247-L259.
268. Morgenroth, K. and M. Ebsen, *CHAPTER 8 - Anatomy*, in *Mechanical Ventilation*, P.J. Papadakos, B. Lachmann, and L. Visser-Isles, Editors. 2008, W.B. Saunders: Philadelphia. p. 69-85.
269. Hastings, R.H., H.G. Folkesson, and M.A. Matthay, *Mechanisms of alveolar protein clearance in the intact lung.* American Journal of Physiology-Lung Cellular and Molecular Physiology, 2004. **286**(4): p. L679-L689.
270. James, J.M. and J.F. Crespo, *Allergic reactions to foods by inhalation.* Current allergy and asthma reports, 2007. **7**(3): p. 167-174.
271. Asero, R., et al., *A case of allergy to airborne, heat-labile shrimp allergens.* Journal of allergy and clinical immunology, 2002. **109**(2): p. 371.
272. Leser, C., et al., *The" egg-egg" syndrome: occupational respiratory allergy to airborne egg proteins with consecutive ingestive egg allergy in the bakery and confectionery industry.* Journal of investigational allergology & clinical immunology, 2001. **11**(2): p. 89-93.
273. Salcedo, G., S. Quirce, and A. Diaz-Perales, *Wheat Allergens Associated With Baker's Asthma.* Journal of Investigational Allergology and Clinical Immunology, 2011. **21**(2): p. 81.
274. Leahy, J.L. and W.T. Cefalu, *Insulin Therapy, An Issue of Endocrinology and Metabolism Clinics - E-Book.* 2012: Elsevier Health Sciences.
275. Patton, J.S., J.G. Bukar, and M.A. Eldon, *Clinical Pharmacokinetics and Pharmacodynamics of Inhaled Insulin.* Clinical Pharmacokinetics, 2004. **43**(12): p. 781-801.
276. Gowthamarajan, K. and G.T. Kulkarni, *Oral Insulin—Fact or Fiction?* Resonance, 2003. **8**(5): p. 38-46.
277. Melmed, S., et al., *Williams Textbook of Endocrinology E-Book*. 2011: Elsevier Health Sciences.
278. Walvoord, E.C., et al., *Inhaled Growth Hormone (GH) Compared with Subcutaneous GH in Children with GH Deficiency: Pharmacokinetics, Pharmacodynamics, and Safety.* The Journal of Clinical Endocrinology & Metabolism, 2009. **94**(6): p. 2052-2059.
279. Parmentier, J., et al., *Improved Oral Bioavailability of Human Growth Hormone by a Combination of Liposomes Containing Bio-Enhancers and Tetraether Lipids*

and Omeprazole. Journal of Pharmaceutical Sciences, 2014. **103**(12): p. 3985-3993.
280. Remington, B., et al., *Approaches to assess IgE mediated allergy risks (sensitization and cross-reactivity) from new or modified dietary proteins.* Food and Chemical Toxicology, 2018. **112**: p. 97-107.
281. Verhoeckx, K., et al., *Allergenicity assessment strategy for novel food proteins and protein sources.* Regulatory Toxicology and Pharmacology, 2016. **79**: p. 118-124.
282. Van Huis, A. and F.V. Dunkel, *Chapter 21 - Edible Insects: A Neglected and Promising Food Source*, in *Sustainable Protein Sources*, S.R. Nadathur, J.P.D. Wanasundara, and L. Scanlin, Editors. 2017, Academic Press: San Diego. p. 341-355.
283. Wanninger, A., *Evolutionary Developmental Biology of Invertebrates 5: Ecdysozoa III: Hexapoda*. 2015: Springer Vienna.
284. Downs, M., P. Johnson, and M. Zeece, *Chapter 9 - Insects and Their Connection to Food Allergy*, in *Insects as Sustainable Food Ingredients*, A.T. Dossey, J.A. Morales-Ramos, and M.G. Rojas, Editors. 2016, Academic Press: San Diego. p. 255-272.
285. Palmer, L.K., et al., *Shellfish Tropomyosin IgE Cross-reactivity Differs among Edible Insect Species.* Molecular Nutrition and Food Research, 2020. **In print**.
286. Zhang, J., et al., *PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification.* Molecular & cellular proteomics, 2012. **11**(4).
287. Oppert, B., et al., *Transcriptome analysis of life stages of the house cricket, Acheta domesticus, to improve insect crop production.* Scientific Reports, 2020. **10**(1): p. 1-13.
288. Madeira, F., et al., *The EMBL-EBI search and sequence analysis tools APIs in 2019.* Nucleic acids research, 2019. **47**(W1): p. W636-W641.
289. Nugraha, R., et al., *Rapid and comprehensive discovery of unreported shellfish allergens using large-scale transcriptomic and proteomic resources.* Journal of Allergy and Clinical Immunology, 2018. **141**(4): p. 1501-1504. e8.
290. Vaughan, T.G., *IcyTree: rapid browser-based visualization for phylogenetic trees and networks.* Bioinformatics, 2017. **33**(15): p. 2392-2394.
291. Altschul, S.F., et al., *Basic local alignment search tool.* Journal of molecular biology, 1990. **215**(3): p. 403-410.
292. Conesa, A., et al., *Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.* Bioinformatics, 2005. **21**(18): p. 3674-3676.
293. Mitchell, A.L., et al., *InterPro in 2019: improving coverage, classification and access to protein sequence annotations.* Nucleic acids research, 2019. **47**(D1): p. D351-D360.
294. Casas-Vila, N., et al., *The developmental proteome of Drosophila melanogaster.* Genome research, 2017. **27**(7): p. 1273-1285.
295. Vogel, C. and E.M. Marcotte, *Insights into the regulation of protein abundance from proteomic and transcriptomic analyses.* Nature reviews. Genetics, 2012. **13**(4): p. 227-232.

296. Bar-Even, A., et al., *Noise in protein expression scales with natural protein abundance.* Nature genetics, 2006. **38**(6): p. 636-643.
297. Reese, G., R. Ayuso, and S.B. Lehrer, *Tropomyosin: an invertebrate pan–allergen.* International archives of allergy and immunology, 1999. **119**(4): p. 247-258.
298. Lopata, A.L., J. Kleine-Tebbe, and S.D. Kamath, *Allergens and molecular diagnostics of shellfish allergy.* Allergo journal, 2016. **25**(7): p. 24-32.
299. Verhoeckx, K.C.M., et al., *House dust mite (Der p 10) and crustacean allergic patients may react to food containing Yellow mealworm proteins.* Food and Chemical Toxicology, 2014. **65**: p. 364-373.
300. Suzuki, M., et al., *Paramyosin of the disc abalone Haliotis discus discus: Identification as a new allergen and cross-reactivity with tropomyosin.* Food Chemistry, 2011. **124**(3): p. 921-926.
301. Burmester, T., *Molecular evolution of the arthropod hemocyanin superfamily.* Molecular Biology and Evolution, 2001. **18**(2): p. 184-195.
302. Tawfik, A.I., et al., *Purification, characterisation and titre of the haemolymph juvenile hormone binding proteins from Schistocerca gregaria and Gryllus bimaculatus.* Journal of insect physiology, 2006. **52**(3): p. 255-268.
303. El-Faham, M.H., et al., *Antigenic cross-reactivity between Schistosoma mansoni and allergenic invertebrates putatively due to shared glycanic epitopes.* Scientific reports, 2020. **10**(1): p. 1-12.
304. Blank, S., et al., *Vitellogenins Are New High Molecular Weight Components and Allergens (Api m 12 and Ves v 6) of Apis mellifera and Vespula vulgaris Venom.* PLOS ONE, 2013. **8**(4): p. e62009.
305. Chuang, J.-G., et al., *Proteome mining for novel IgE-binding proteins from the German cockroach (Blattella germanica) and allergen profiling of patients.* PROTEOMICS, 2010. **10**(21): p. 3854-3867.
306. Huntley, J.F., et al., *Identification of tropomyosin, paramyosin and apolipophorin/vitellogenin as three major allergens of the sheep scab mite, Psoroptes ovis.* Parasite Immunology, 2004. **26**(8-9): p. 335-342.
307. EFSA Panel on Nutrition, N.F., et al., *Safety of dried yellow mealworm (Tenebrio molitor larva) as a novel food pursuant to Regulation (EU) 2015/2283.* EFSA Journal, 2021. **19**(1): p. e06343.
308. Gong, C.Y. and T. Wang, *Proteomic evaluation of genetically modified crops: current status and challenges.* Frontiers in plant science, 2013. **4**: p. 41.
309. Barre, A., et al., *Insights into the Allergenic Potential of the Edible Yellow Mealworm (Tenebrio molitor).* Foods, 2019. **8**(10): p. 515.
310. Broekman, H.C., et al., *Is mealworm or shrimp allergy indicative for food allergy to insects?* Molecular nutrition & food research, 2017. **61**(9): p. 1601061.
311. Eriksson, T., et al., *The yellow mealworm (Tenebrio molitor) genome: a resource for the emerging insects as food and feed industry.* Journal of Insects as Food and Feed, 2020. **6**(5): p. 445-455.
312. Stanke, M., et al., *Using native and syntenically mapped cDNA alignments to improve de novo gene finding.* Bioinformatics, 2008. **24**(5): p. 637-644.
313. Lorenz, L.J., *A pipeline for prokaryotic genome annotation with peptide data from MS/MS experiments*. 2019, University of Griefswald.

314. Scalzitti, N., et al., *A benchmark study of ab initio gene prediction methods in diverse eukaryotic organisms.* BMC Genomics, 2020. **21**(1): p. 293.
315. Liu, S., et al., *De novo analysis of the Tenebrio molitor (Coleoptera: Tenebrionidae) transcriptome and identification of putative glutathione S-transferase genes.* Applied Entomology and Zoology, 2015. **50**(1): p. 63-71.
316. Taylor, S.L., et al., *Establishment of reference doses for residues of allergenic foods: report of the VITAL expert panel.* Food and Chemical Toxicology, 2014. **63**: p. 9-17.
317. Fu, T.-J. and N. Maks, *Impact of thermal processing on ELISA detection of peanut allergens.* Journal of agricultural and food chemistry, 2013. **61**(24): p. 5649-5658.
318. Hebling, C.M., et al., *Global proteomic screening of protein allergens and advanced glycation endproducts in thermally processed peanuts.* Journal of agricultural and food chemistry, 2013. **61**(24): p. 5638-5648.
319. Zhang, T., et al., *Different thermal processing effects on peanut allergenicity.* Journal of the Science of Food and Agriculture, 2019. **99**(5): p. 2321-2328.
320. Zhou, N., et al., *Sequential extractions: A new way for protein quantification—data from peanut allergens.* Analytical biochemistry, 2015. **484**: p. 31-36.
321. Arena, S., et al., *Dairy products and the Maillard reaction: A promising future for extensive food characterization by integrated proteomics studies.* Food Chemistry, 2017. **219**: p. 477-489.
322. Güray, M.Z., S. Zheng, and A.A. Doucette, *Mass spectrometry of intact proteins reveals+ 98 u chemical artifacts following precipitation in acetone.* Journal of proteome research, 2017. **16**(2): p. 889-897.
323. Mol, M., et al., *Advanced lipoxidation end products (ALEs) as RAGE binders: Mass spectrometric and computational studies to explain the reasons why.* Redox biology, 2019. **23**: p. 101083.
324. Nursten, H.E., *Other Physiological Aspects*, in *The Maillard Reaction: Chemistry, Biochemistry and Implications*. 2007, Royal Society of Chemistry: Cambridge, UK.
325. De Jongh, H.H., et al., *Digestibility and IgE-binding of glycosylated codfish parvalbumin.* BioMed research international, 2013. **2013**.
326. Huby, R.D., R.J. Dearman, and I. Kimber, *Why are some proteins allergens?* Toxicological Sciences, 2000. **55**(2): p. 235-246.
327. Schmitt, D.A., et al., *Processing can alter the properties of peanut extract preparations.* Journal of Agricultural and Food Chemistry, 2010. **58**(2): p. 1138-1143.
328. Comstock, S.S., S.J. Maleki, and S.S. Teuber, *Boiling and frying peanuts decreases soluble peanut (Arachis hypogaea) allergens Ara h 1 and Ara h 2 but does not generate hypoallergenic peanuts.* PloS one, 2016. **11**(6).
329. Sen, M., et al., *Protein structure plays a critical role in peanut allergen stability and may determine immunodominant IgE-binding epitopes.* The journal of immunology, 2002. **169**(2): p. 882-887.
330. Tomita, S. and K. Shiraki, *Why do solution additives suppress the heat-induced inactivation of proteins? Inhibition of chemical modifications.* Biotechnology progress, 2011. **27**(3): p. 855-862.

331. Gavage, M., et al., *High-resolution mass spectrometry-based selection of peanut peptide biomarkers considering food processing and market type variation.* Food chemistry, 2020. **304**: p. 125428.
332. Catak, S., et al., *Computational study on nonenzymatic peptide bond cleavage at asparagine and aspartic acid.* The Journal of Physical Chemistry A, 2008. **112**(37): p. 8752-8761.
333. van Boekel, M.A., *Heat-induced deamidation, dephosphorylation and breakdown of caseinate.* International Dairy Journal, 1999. **9**(3-6): p. 237-241.
334. Dai, H. and A.M. Leventhal, *Prevalence of e-cigarette use among adults in the United States, 2014-2018.* Jama, 2019. **322**(18): p. 1824-1827.
335. Chen, J.C., *Flavored e-cigarette use and cigarette smoking reduction and cessation—a large national study among young adult smokers.* Substance use & misuse, 2018. **53**(12): p. 2017-2031.
336. Baur, X., Z. Chen, and V. Liebers, *Exposure-response relationships of occupational inhalative allergens.* Clinical and Experimental Allergy, 1998. **28**(5): p. 537-544.
337. O'B Hourihane, J., *The threshold concept in food safety and its applicability to food allergy.* Allergy, 2001. **56**: p. 86-90.
338. Ramirez, D.A. and S.L. Bahna, *Food hypersensitivity by inhalation.* Clinical and Molecular Allergy, 2009. **7**(1): p. 1-6.
339. Tzortzi, A., et al., *A systematic literature review of e-cigarette-related illness and injury: not just for the respirologist.* International journal of environmental research and public health, 2020. **17**(7): p. 2248.
340. Palmer, L.K., et al., *Persistence of peanut allergen-derived peptides throughout excessive dry thermal processing.* LWT, 2020: p. 109903.
341. Noble, K.A., et al., *A Cherry Seed-Derived Spice, Mahleb, is Recognized by Anti-Almond Antibodies Including Almond-Allergic Patient IgE.* Journal of food science, 2017. **82**(8): p. 1786-1791.
342. Albillos, S.M., et al., *Purification, crystallization and preliminary X-ray characterization of prunin-1, a major component of the almond (Prunus dulcis) allergen amandin.* Journal of agricultural and food chemistry, 2008. **56**(13): p. 5352-5358.
343. Willison, L.N., et al., *Cloning, expression and patient IgE reactivity of recombinant Pru du 6, an 11S globulin from almond.* International archives of allergy and immunology, 2011. **156**(3): p. 267-281.
344. Willison, L.N., et al., *Conformational epitope mapping of Pru du 6, a major allergen from almond nut.* Molecular immunology, 2013. **55**(3-4): p. 253-263.
345. Talih, S., et al., *"Juice Monsters": Sub-ohm vaping and toxic volatile aldehyde emissions.* Chemical Research in Toxicology, 2017. **30**(10): p. 1791-1793.
346. Boeniger, M.F., et al., *Exposure to protein aeroallergens in egg processing facilities.* Applied occupational and environmental hygiene, 2001. **16**(6): p. 660-670.
347. Taylor, A.V., et al., *Detection and quantitation of raw fish aeroallergens from an open-air fish market.* Journal of allergy and clinical immunology, 2000. **105**(1): p. 166-169.

348. Brough, H.A., et al., *Distribution of peanut protein in the home environment.* Journal of allergy and clinical immunology, 2013. **132**(3): p. 623-629.
349. Salvatori, N., et al., *Asthma induced by inhalation of flour in adults with food allergy to wheat.* Clinical & Experimental Allergy, 2008. **38**(8): p. 1349-1356.
350. Armentia, A., et al., *Why can patients with baker's asthma tolerate wheat flour ingestion? Is wheat pollen allergy relevant?* Allergologia et immunopathologia, 2009. **37**(4): p. 203-204.
351. Salcedo, G., S. Quirce, and A. Diaz-Perales, *1 Wheat Allergens Associated With Baker's Asthma.* Journal of Investigational Allergology and Clinical Immunology, 2011. **21**(2): p. 81.
352. Wu, J.-Z., et al., *Effects of particulate matter on allergic respiratory diseases.* Chronic diseases and translational medicine, 2018. **4**(2): p. 95-102.
353. Pourchez, J., et al., *Impact of power level and refill liquid composition on the aerosol output and particle size distribution generated by a new-generation e-cigarette device.* Aerosol Science and technology, 2018. **52**(4): p. 359-369.
354. Son, Y., et al., *Investigating E-Cigarette Particle Emissions and Human Airway Depositions under Various E-Cigarette-Use Conditions.* Chemical Research in Toxicology, 2019. **33**(2): p. 343-352.
355. Baldacci, S., et al., *Allergy and asthma: Effects of the exposure to particulate matter and biological allergens.* Respiratory medicine, 2015. **109**(9): p. 1089-1104.
356. Scott, A., et al., *Pro-inflammatory effects of e-cigarette vapour condensate on human alveolar macrophages.* Thorax, 2018. **73**(12): p. 1161-1169.
357. Brough, H.A., et al., *Atopic dermatitis increases the effect of exposure to peanut antigen in dust on peanut sensitization and likely peanut allergy.* Journal of Allergy and Clinical Immunology, 2015. **135**(1): p. 164-170. e4.
358. Ahmedna, M., et al., *Potential of agricultural by-product-based activated carbons for use in raw sugar decolourisation.* Journal of the Science of Food and Agriculture, 1997. **75**(1): p. 117-124.
359. Nieto-Delgado, C., M. Terrones, and J.R. Rangel-Mendez, *Development of highly microporous activated carbon from the alcoholic beverage industry organic by-products.* Biomass and Bioenergy, 2011. **35**(1): p. 103-112.
360. Pal, P., *Industrial Water Treatment Process Technology*. 2017: Elsevier Science.
361. Lehmann, J. and S. Joseph, *Biochar for Environmental Management: Science and Technology*. 2012: Earthscan.
362. Bae, W., J. Kim, and J. Chung, *Production of granular activated carbon from food-processing wastes (walnut shells and jujube seeds) and its adsorptive properties.* Journal of the Air & Waste Management Association, 2014. **64**(8): p. 879-886.
363. Satya Sai, P.M., J. Ahmed, and K. Krishnaiah, *Production of Activated Carbon from Coconut Shell Char in a Fluidized Bed Reactor.* Industrial & Engineering Chemistry Research, 1997. **36**(9): p. 3625-3630.
364. Buyukada, M., *Uncertainty estimation by Bayesian approach in thermochemical conversion of walnut hull and lignite coal blends.* Bioresour Technol, 2017. **232**: p. 87-92.

365. Duerr, S., *The Handbook of Natural Plant Dyes: Personalize Your Craft with Organic Colors from Acorns, Blackberries, Coffee, and Other Everyday Ingredients*. 2010: Timber Press.
366. Meuninck, J., *Basic Illustrated Medicinal Plants*. 2014: Falcon Guides.
367. Wolfe, D. and R.A. Gauthier, *Longevity Now: A Comprehensive Approach to Healthy Hormones, Detoxification, Super Immunity, Reversing Calcification, and Total Rejuvenation*. 2013: North Atlantic Books.
368. USDA, U.S.D.o.A. *Germplasm Resources Information Network*. 151128 [cited 2017 April 20]; Available from: http://www.ars-grin.gov/.
369. Costa, J., et al., *Walnut allergens: molecular characterization, detection and clinical relevance.* Clin Exp Allergy, 2014. **44**(3): p. 319-41.
370. Plumlee, K., *Clinical Veterinary Toxicology - E-Book*. 2003: Elsevier Health Sciences.
371. Mohan, D.P., C.U.; Steele, P.H., *Pyrolysis of Wood/Biomass for Bio-oil: A Critical Review*. Energy & Fuels, 2006. **20**: p. 848-889.
372. Pütün, A.E., et al., *Fixed-bed pyrolysis of cotton stalk for liquid and solid products.* Fuel Processing Technology, 2005. **86**(11): p. 1207-1219.
373. Azargohar, R. and A.K. Dalai, *Steam and KOH activation of biochar: Experimental and modeling studies.* Microporous and Mesoporous Materials, 2008. **110**(2–3): p. 413-421.
374. del-Campo, B.G., et al., *Optimizing the production of activated carbon from fast pyrolysis char.* TECHNOLOGY, 2015. **03**(02n03): p. 104-113.
375. Pendyal, B., et al., *Removal of sugar colorants by granular activated carbons made from binders and agricultural by-products.* Bioresource Technology, 1999. **69**(1): p. 45-51.
376. Ioannidou, O. and A. Zabaniotou, *Agricultural residues as precursors for activated carbon production—A review.* Renewable and Sustainable Energy Reviews, 2007. **11**(9): p. 1966-2005.
377. ScienceLab.com, *Material Safety Data Sheet Charcoal, Acitvated, Powder MSDS*. 2013.
378. Corp., G.C., *Safety Data Sheet: GC Activated Carbon*. 2017.
379. Food and Drug Administration, *RE: Interin Final Rule - Registration of Food Facilities Under the Public Health Security and Bioterrorism Preparedness and Response Ace of 2002 [Docket No. 02N-0276]*, I.A.C.M. Association, Editor. 2003: Division of Dockets Management (HFA-305).
380. Bock, S.A., A. Muñoz-Furlong, and H.A. Sampson, *Further fatalities caused by anaphylactic reactions to food, 2001-2006.* Journal of Allergy and Clinical Immunology, 2007. **119**(4): p. 1016-1018.
381. Burks, A.W., et al., *ICON: food allergy.* J Allergy Clin Immunol, 2012. **129**(4): p. 906-20.
382. Sicherer, S.H., et al., *A voluntary registry for peanut and tree nut allergy: Characteristics of the first 5149 registrants.* Journal of Allergy and Clinical Immunology, 2001. **108**(1): p. 128-132.
383. Comstock, S.S., et al., *Extensive in vitro cross-reactivity to seed storage proteins is present among walnut (Juglans) cultivars and species.* Clin Exp Allergy, 2004. **34**(10): p. 1583-90.

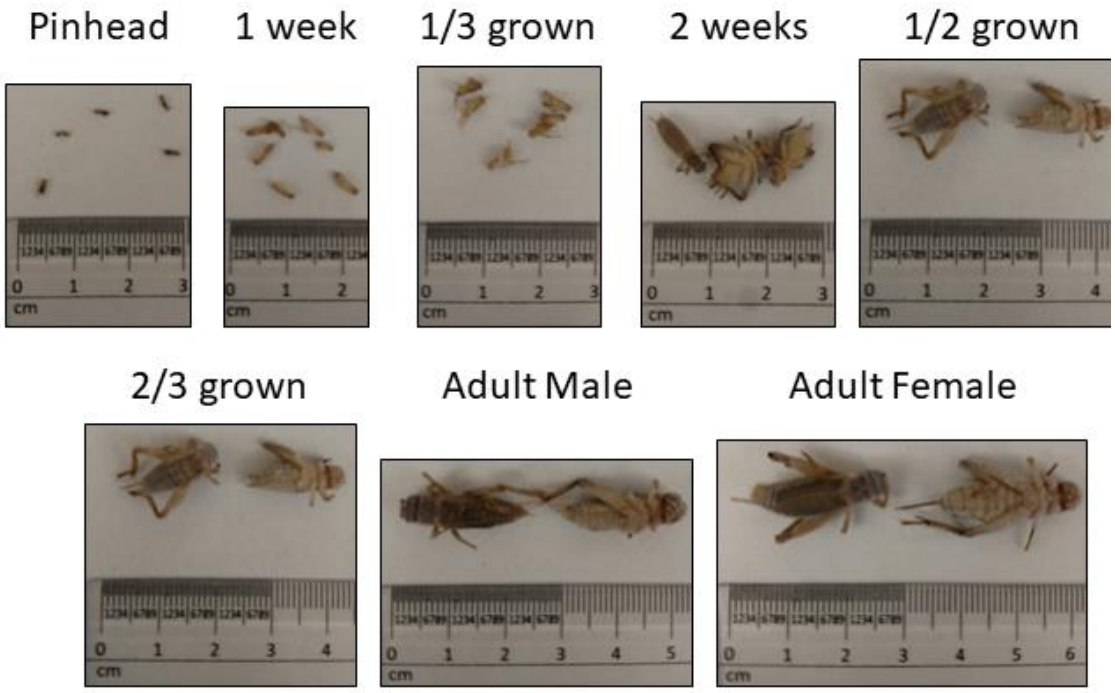
384. Wallowitz, M.L., et al., *Cross-reactivity of walnut, cashew, and hazelnut legumin proteins in tree nut allergic patients.* Journal of Allergy and Clinical Immunology. **113**(2): p. S156.
385. Ramos, D.E., *Walnut Production Manual*. 1997: University of California, Division of Agriculture and Natural Resources, Communication Services--Publications.
386. Blankestijn, M.A., et al., *Threshold Dose Distribution in Walnut Allergy.* J Allergy Clin Immunol Pract, 2017. **5**(2): p. 376-380.
387. Montain, S.J.E., M.;, *Water Requirements and Soldier Hydration*, in *Borden Institute Monograph Series*, W.R.F. Santee, K.E.;, Editor. 2012, Borden Institute.
388. Damodaran, S., K.L. Parkin, and O.R. Fennema, *Fennema's Food Chemistry, Fourth Edition.* 2007: CRC Press.
389. Russell, J.D., et al., *Thermal decomposition of protein in soil organic matter.* Geoderma, 1974. **11**(1): p. 63-66.
390. Dandurand, J., et al., *Conformational and thermal characterization of a synthetic peptidic fragment inspired from human tropoelastin: Signature of the amyloid fibers.* Pathol Biol (Paris), 2014. **62**(2): p. 100-7.
391. Sunooj, K., et al., *Thermal degradation and decomposition kinetics of freeze dried cow and camel milk as well as their constituents.* Journal of Food Science and Engineering, 2011. **1**(2): p. 77.
392. Samouillan, V., et al., *Thermal analysis characterization of aortic tissues for cardiac valve bioprostheses.* Journal of Biomedical Materials Research, 1999. **46**(4): p. 531-538.
393. Bach, Q.-V. and W.-H. Chen, *A comprehensive study on pyrolysis kinetics of microalgal biomass.* Energy Conversion and Management, 2017. **131**: p. 109-116.
394. Qian, Y., et al., *Kinetics of peptide hydrolysis and amino acid decomposition at high temperature.* Geochimica et Cosmochimica Acta, 1993. **57**(14): p. 3281-3293.
395. Sionkowska, A., et al., *Chemical and thermal cross-linking of collagen and elastin hydrolysates.* Int J Biol Macromol, 2010. **47**(4): p. 570-7.
396. Qian, Y., et al., *Kinetics of peptide hydrolysis and amino acid decomposition at high temperature.* Geochim Cosmochim Acta, 1993. **57**(14): p. 3281-93.
397. Smith, R.M.H., D.E., *The pH-Rate Profile for the Hydrolysis of a Peptide Bond.* Journal of the American Chemical Association, 1988. **120**: p. 8910-8913.
398. Zempleni, J. and H. Daniel, *Molecular Nutrition*. 2003: CABI Pub.
399. Cosmulescu, S.T., Ion; Achim, Gheorghe; Botu, Mihai; Baciu, Adrian; Gruia, Marius, *Phenolics of Green Husk in Mature Walnut Fruits.* Notulae Botanicae Horti Agrobotanici Cluj-Napoca, 2010.
400. Shahidi, F. and M. Naczk, *Phenolics in Food and Nutraceuticals*. 2003: CRC Press.
401. Alasalvar, C. and F. Shahidi, *Tree Nuts: Composition, Phytochemicals, and Health Effects*. 2008: CRC Press.
402. Stampar, F., et al., *Traditional walnut liqueur – cocktail of phenolics.* Food Chemistry, 2006. **95**(4): p. 627-631.
403. Babula, P., et al., *[Naphthoquinones and their pharmacological properties].* Ceska Slov Farm, 2007. **56**(3): p. 114-20.

404. Inbaraj, J.J. and C.F. Chignell, *Cytotoxic Action of Juglone and Plumbagin: A Mechanistic Study Using HaCaT Keratinocytes.* Chemical Research in Toxicology, 2004. **17**(1): p. 55-62.
405. SCCNFP (Scientific Committee On Cosmetic Products And Non-Food Products), *Evaluation and opinion on lawsone*. 2002: SCCNFP/0583/02.
406. Santa Cruz Biotechnology Inc., *Juglone Material Safety Data Sheet*. 2010, Santa Cruz Biotechnology, Inc.
407. Santa Cruz Biotechnology Inc., *Plumbagin Material Safety Data Sheet*. 2016, Santa Cruz Biotechnology, Inc.
408. Santa Cruz Biotechnology Inc., *1,4 Naphthoquinone Material Safety Data Sheet*. 2010, Santa Cruz Biotechnology, Inc.
409. Ismail, T., et al., *Ellagitannins in Cancer Chemoprevention and Therapy.* Toxins (Basel), 2016. **8**(5).
410. Tasaki, M., et al., *Safety assessment of ellagic acid, a food additive, in a subchronic toxicity study using F344 rats.* Food and Chemical Toxicology, 2008. **46**(3): p. 1119-1124.
411. Santa Cruz Biotechnology Inc., *Ellagic acid Material Safety Data Sheet*. 2016, Santa Cruz Biotechnology, Inc.
412. Centers for Disease Control and Prevention (US), National Center for Chronic Disease Prevention and Health Promotion (US),, Office on Smoking and Health (US),,, *Chemistry and Toxicology of Cigarette Smoke and Biomarkers of Exposure and Harm*, in *How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking-Attributable Disease: A Report of the Surgeon General*. 2010, Centers for Disease Control and Prevention (US): Atlanta, GA.
413. Feiner, G., *Meat Products Handbook: Practical Science and Technology*. 2006: Elsevier Science.
414. Chichester, C.O., E.M. Mrak, and G.F. Stewart, *Advances in Food Research*. 1984: Elsevier Science.
415. Devine, C. and M. Dikeman, *Encyclopedia of Meat Sciences*. 2014: Elsevier Science.
416. U.S. Food & Drug Administration. *Section 201(qq) of the Act defines the term "major food allergen" to include "tree nuts." In addition to the three examples provided in section 201(qq) (almonds, pecans, and walnuts), what nuts are considered "tree nuts?"*. FDA Basics For Industry 2016 05/22/2016 [cited 2017 06/26/2017].
417. Mauseth, J.D., *Botany*. 2016: Jones & Bartlett Learning.
418. Hon, D.N.S. and N. Shiraishi, *Wood and Cellulosic Chemistry, Second Edition, Revised, and Expanded*. 2000: Taylor & Francis.
419. Barton, G., *Definition of biomass samples involving wood, bark and foliage.* Biomass, 1984. **4**(4): p. 311-314.
420. Cronk, Q.C.B., *Plant eco-devo: the potential of poplar as a model organism.* New Phytologist, 2005. **166**(1): p. 39-48.
421. Ellis, B., et al., *Why and How Populus Became a "Model Tree"*, in *Genetics and Genomics of Populus*, S. Jansson, R. Bhalerao, and A. Groover, Editors. 2010, Springer New York: New York, NY. p. 3-14.

422. Jansson, S. and C.J. Douglas, *Populus: a model system for plant biology.* Annu. Rev. Plant Biol., 2007. **58**: p. 435-458.
423. Lexer, C., et al., *Barrier to gene flow between two ecologically divergent Populus species, P. alba (white poplar) and P. tremula (European aspen): the role of ecology and life history in gene introgression.* Molecular ecology, 2005. **14**(4): p. 1045-1057.
424. Tuskan, G.A., et al., *The genome of black cottonwood, Populus trichocarpa (Torr. & Gray).* science, 2006. **313**(5793): p. 1596-1604.
425. Rennenberg, H., H. Wildhagen, and B. Ehlting, *Nitrogen nutrition of poplar trees.* Plant Biology, 2010. **12**(2): p. 275-291.
426. Couturier, J., et al., *Glutamine, arginine and the amino acid transporter Pt-CAT11 play important roles during senescence in poplar.* Annals of Botany, 2010. **105**(7): p. 1159-1169.
427. Gomez, L. and M. Faurobert, *Contribution of vegetative storage proteins to seasonal nitrogen variations in the young shoots of peach trees (Prunus persica L. Batsch).* Journal of Experimental Botany, 2002. **53**(379): p. 2431-2439.
428. Veverka, P.J., et al., *On the form of nitrogen in wood and its fate during kraft pulping.* 1993.
429. Arora, R., *Adaptations and Responses of Woody Plants to Environmental Stresses*. 2004: Taylor & Francis.
430. Beardmore, T., et al., *Characterization of seed storage proteins in Populus and their homology with Populus vegetative storage proteins.* Tree physiology, 1996. **16**(10): p. 833-840.
431. Zhu, B. and G.D. Coleman, *The poplar bark storage protein gene (Bspa) promoter is responsive to photoperiod and nitrogen in transgenic poplar and active in floral tissues, immature seeds and germinating seeds of transgenic tobacco.* Plant molecular biology, 2001. **46**(4): p. 383-394.
432. Grebner, D.L., P. Bettinger, and J.P. Siry, *Introduction to Forestry and Natural Resources*. 2012: Elsevier Science.
433. Vogt, D.J., et al., *Forest Certification: Roots, Issues, Challenges, and Benefits*. 1999: Taylor & Francis.
434. Fonseca, M.A., *The Measurement of Roundwood: Methodologies and Conversion Ratios*. 2005: CABI.
435. Toldrá, F., *Handbook of Meat Processing*. 2010: Wiley.
436. Hui, Y.H., W.K. Nip, and R. Rogers, *Meat Science and Applications*. 2001: CRC Press.
437. Hui, Y.H., *Handbook of Meat and Meat Processing, Second Edition*. 2012: Taylor & Francis.
438. Purser, D., R. Maynard, and J. Wakefield, *Toxicology, Survival and Health Hazards of Combustion Products*. 2015: Royal Society of Chemistry.
439. Beji, T., *Fluid Mechanics Aspects of Fire and Smoke Dynamics in Enclosures*. 2016: CRC Press.
440. Klass, D.L., *Biomass for Renewable Energy, Fuels, and Chemicals*. 1998: Elsevier Science.
441. Pandey, A., et al., *Current Developments in Biotechnology and Bioengineering: Food and Beverages Industry*. 2016: Elsevier Science.

442. Berger, R.G., *Flavours and Fragrances: Chemistry, Bioprocessing and Sustainability*. 2007: Springer Berlin Heidelberg.
443. Blankestijn, M.A., et al., *Threshold Dose Distribution in Walnut Allergy.* The Journal of Allergy and Clinical Immunology: In Practice, 2017. **5**(2): p. 376-380.
444. Goetz, D.W., M.A. Goetz, and B.A. Whisman, *Mountain cedar allergens found in nonpollen tree parts.* Ann Allergy Asthma Immunol, 1995. **75**(3): p. 256-60.
445. More, D., et al., *Identification of specific IgE to mesquite wood smoke in individuals with mesquite pollen allergy.* Journal of allergy and clinical immunology, 2002. **110**(5): p. 814-816.
446. Becker, C., T. Dubin, and H.P. Wiedemann, *Hypersensitivity to tobacco antigen.* Proceedings of the National Academy of Sciences, 1976. **73**(5): p. 1712-1716.
447. Voisine, R., et al., *Protein transfer in mainstream and sidestream cigarette smoke.* Beiträge zur Tabakforschung/Contributions to Tobacco Research, 2004. **21**(1): p. 9-14.
448. Nielsen, G.D., et al., *Experiences from Occupational Exposure Limits Set on Aerosols Containing Allergenic Proteins.* The Annals of Occupational Hygiene, 2012. **56**(8): p. 888-900.
449. Liu, C., et al., *The use of a novel tobacco treatment process to reduce toxicant yields in cigarette smoke.* Food and Chemical Toxicology, 2011. **49**(9): p. 1904-1917.

APPENDIX A SUPPLEMENTARY FIGURES AND TABLES FROM CHAPTER 2: PREDICTED ALLERGENS AND QUANTITIATIVE PROTEOMICS FROM LIFE STAGES OF THE HOUSE CRICKET, *ACHETA DOMESTICUS*

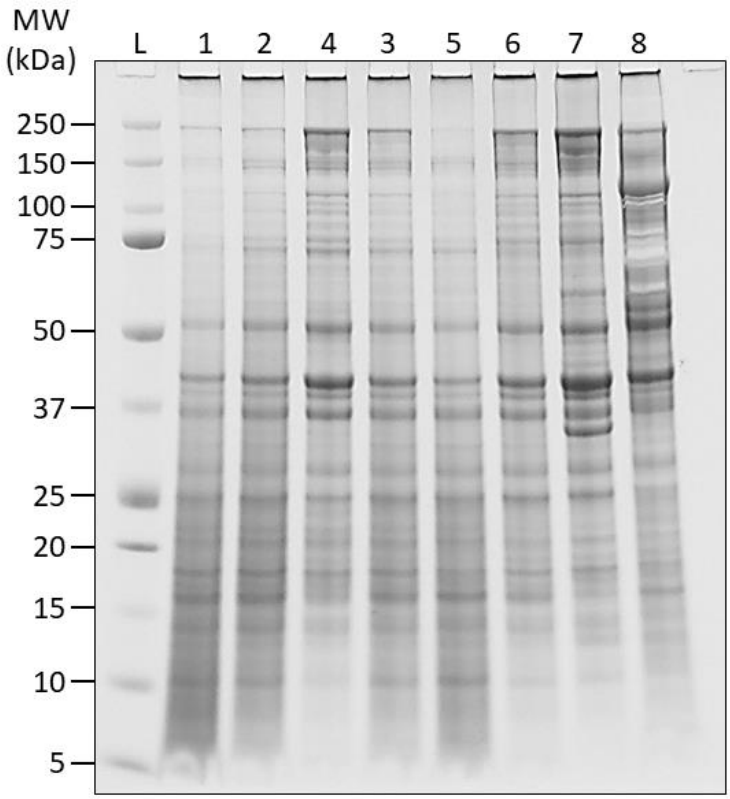


Appendix A Figure 2.1.1 *A. domesticus* life stages

Images of *A. domesticus* life stages in order of development from pinheads to adults.

Appendix A Table 2.1.1 Life stage replicate characteristics

Life stage	Replicate	#specimens	Replicate sample weight (mg)	µg protein per 25 µg wet tissue	Percentage protein (wet basis)
Pinhead	1	>50	215	1.06	5.64
	2	>50	211	1.19	6.37
	3	>50	202	1.17	6.26
1 Week	1	50	208	1.15	6.13
	2	50	207	1.32	7.05
	3	50	216	1.01	5.39
1/3 Grown	1	25	211	1.10	5.26
	2	25	214	1.04	6.46
	3	25	204	0.98	6.02
2 Weeks	1	2	217	0.99	5.84
	2	2	178	1.21	5.55
	3	2	186	1.13	5.20
1/2 Grown	1	2	234	0.88	4.72
	2	2	373	1.11	5.94
	3	2	327	1.34	7.17
2/3 Grown	1	1	209	0.75	4.02
	2	1	203	0.99	5.28
	3	1	257	1.29	6.87
Adult Male	1	1	376	1.23	6.56
	2	1	382	0.93	4.97
	3	1	416	1.45	7.74
Adult Female	1	1	557	1.51	8.06
	2	1	556	1.93	10.29
	3	1	606	1.41	7.53

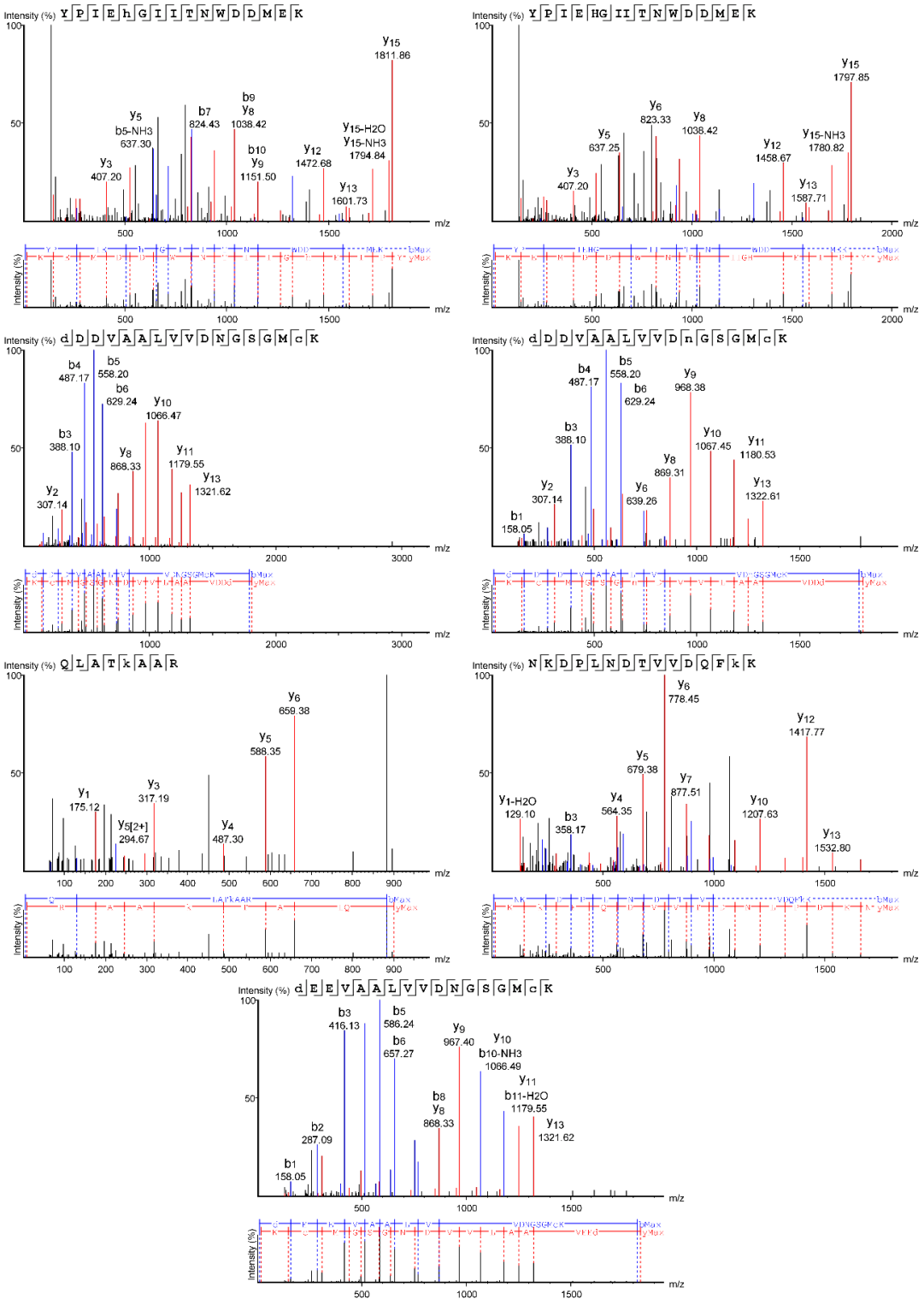


Appendix A Figure 2.1.2 SDS-PAGE of extracts from *A. domesticus* life stages

SDS-PAGE of 8 life stages of *A. domesticus.* Lanes were loaded with protein equivalent to 500 µg of wet tissue per life stage. Lanes were as follows: (L) molecular weight standards, (1) Pinhead, (2) 1 week, (3) 1/3 grown, (4) 2 weeks, (5) 1/2 grown, (6) 2/3 grown, (7) Adult male, (8) Adult female. Numbers are in order of development rather than lane loading order.

Appendix A Table 2.1.2 Mass spectrometry multi-round search metadata

	Round 1	Round 2	Round 3	Round 4
# MS scans	926192	926192	926192	926192
#MS/MS scans	425738	40737	14808	12807
Peptide-Spectrum matches at FDR < 0.01	145671	20087	1955	290
Peptides	4219	1161	287	71
Peptide FDR	0.8%	0.8%	1.0%	<0.1%
Proteins	5446	609	66	8
Protein Groups	481	340	46	8
Protein FDR	6.5%	1.5%	4.3%	<0.1%
# *De novo* only scans	40737	14808	12807	12514

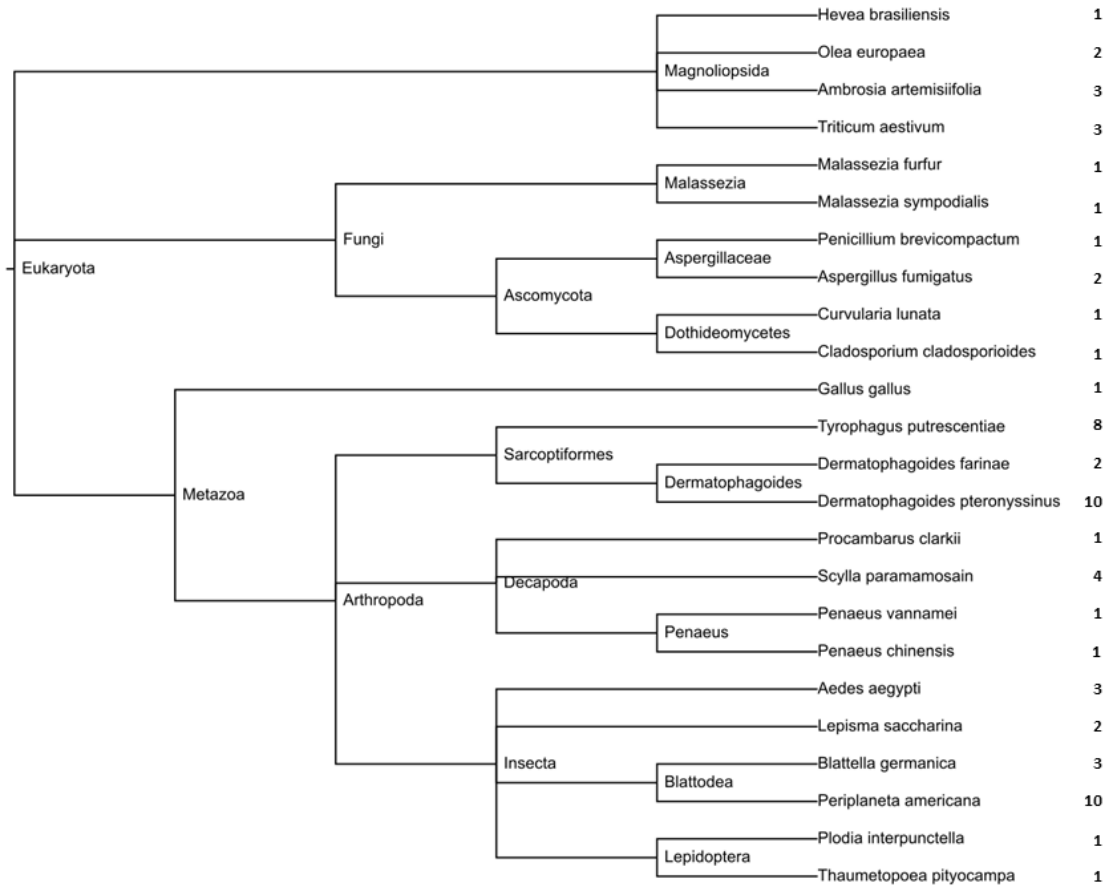


Appendix A Figure 2.1.3 Spectra of peptides with posttranslational modifications and relevant unmodified forms

Appendix A Table 2.1.3 Peptide posttranslational modification characteristics

	m/z	z	RT	-10lgP	ppm	PTM*	Accession	Description
YPIEHGIITNWDDMEK	654.3091	3	49.56	71.65	1		ANN23215.RA	Actin, muscle
YPIEH(+14.02)GIITNWDDMEK	658.9813	3	50.69	71.68	1.3	H_{74} Methylation	ANN23215.RA	Actin, muscle
D(+42.01)DDVAALVVDNGSGMC(+57.02)K	904.3942	2	51.52	62.58	-1.4	D_3 Acetylation, C_{18} Carbamidomethylation	ANN23215.RA	Actin, muscle
D(+42.01)DDVAALVVDN(+.98)GSGMC(+57.02)K	940.8854	2	51.94	59.89	-2.2	D_3 Acetylation, N_{13} Deamidation, C_{18} Carbamidomethylation	ANN23215.RA	Actin, muscle
QLATK(+42.01)AAR	450.7672	2	28.93	31.86	1	K_{24} Acetylation	ANN25863.RA	Histone H3
D(+42.01)EEVAALVVDNGSGMC(+57.02)K	918.4099	2	52.13	51.59	-1.4	D_3 Acetylation, C_{18} Carbamidomethylation	ANN06657.RA	Actin-5C
NKDPLNDTVVDQFK(+14.02)K	592.3174	3	45.6	59.6	2.5	K_{880} Methylation	ANN17471.RA	Myosin heavy chain, Muscle isoform X15

*PTM: Posttranslational modification



Appendix A Figure 2.1.4 Phylogeny of the origin species of allergen predictions

Phylogenetic tree of species produced with IcyTree whose allergenic sequences were significant matches to *A. domesticus* proteins. Branch length is arbitrary. Numbers to the right of each species represents the number of hits attributed to the species.

Appendix A Table 2.1.4 Enriched Gene Ontology (GO) terms for the top 10% most abundant proteins per life stage

Pinhead

GO ID	GO name	Go Category	FDR	P-value	Nr Test	Nr Reference
GO:0035639	purine ribonucleoside triphosphate binding	MF	1.12E-03	6.84E-06	20	67
GO:0051015	actin filament binding	MF	1.12E-03	6.13E-06	7	4
GO:0032555	purine ribonucleotide binding	MF	1.12E-03	6.84E-06	20	67
GO:0003774	motor activity	MF	5.96E-03	5.02E-05	6	4
GO:0016459	myosin complex	CC	5.96E-03	5.02E-05	6	4
GO:0099513	polymeric cytoskeletal fiber	CC	3.70E-02	3.64E-04	7	11
GO:0006936	muscle contraction	BP	4.78E-02	6.05E-04	3	0
GO:0006996	organelle organization	BP	4.78E-02	6.03E-04	12	39
GO:0003924	GTPase activity	MF	4.78E-02	5.38E-04	7	12
GO:0005200	structural constituent of cytoskeleton	MF	4.78E-02	5.45E-04	6	8

1 Week

GO ID	GO name	Go Category	FDR	P-value	Nr Test	Nr Reference
GO:0003774	motor activity	MF	1.09E-02	6.14E-05	6	4
GO:0035639	purine ribonucleoside triphosphate binding	MF	1.09E-02	5.05E-05	19	68
GO:0032555	purine ribonucleotide binding	MF	1.09E-02	5.05E-05	19	68
GO:0016459	myosin complex	CC	1.09E-02	6.14E-05	6	4
GO:0051015	actin filament binding	MF	2.07E-02	1.26E-04	6	5

1/3 Grown

GO ID	GO name	Go Category	FDR	P-value	Nr Test	Nr Reference
GO:0003774	motor activity	MF	1.64E-02	6.14E-05	6	4
GO:0016459	myosin complex	CC	1.64E-02	6.14E-05	6	4
GO:0051015	actin filament binding	MF	2.98E-02	1.26E-04	6	5
GO:0035639	purine ribonucleoside triphosphate binding	MF	3.33E-02	1.87E-04	18	69
GO:0032555	purine ribonucleotide binding	MF	3.33E-02	1.87E-04	18	69
GO:0099512	supramolecular fiber	CC	4.15E-02	2.91E-04	8	14
GO:0008152	metabolic process	BP	4.53E-02	3.39E-04	13	279

2 Weeks

GO ID	GO name	Go Category	FDR	P-value	Nr Test	Nr Reference

GO:0004503	monophenol monooxygenase activity	MF	2.76E-03	5.17E-06	5	0
GO:0042302	structural constituent of cuticle	MF	3.20E-03	7.49E-06	16	40
GO:0044237	cellular metabolic process	BP	5.68E-03	2.97E-05	7	228
GO:0044238	primary metabolic process	BP	6.50E-03	3.96E-05	6	210
GO:0003774	motor activity	MF	9.20E-03	6.78E-05	6	4
GO:0016459	myosin complex	CC	9.20E-03	6.78E-05	6	4
GO:0051015	actin filament binding	MF	1.64E-02	1.39E-04	6	5
GO:0071704	organic substance metabolic process	BP	1.66E-02	1.48E-04	7	213

1/2 Grown

GO ID	GO name	Go Category	FDR	P-value	Nr Test	Nr Reference
GO:0003774	motor activity	MF	1.64E-02	6.14E-05	6	4
GO:0016459	myosin complex	CC	1.64E-02	6.14E-05	6	4
GO:0051015	actin filament binding	MF	2.69E-02	1.26E-04	6	5
GO:0044237	cellular metabolic process	BP	3.36E-02	1.73E-04	8	227
GO:0044238	primary metabolic process	BP	3.78E-02	2.33E-04	7	209
GO:0004503	monophenol monooxygenase activity	MF	3.78E-02	2.65E-04	4	1

2/3 Grown

GO ID	GO name	Go Category	FDR	P-value	Nr Test	Nr Reference
GO:0003774	motor activity	MF	1.30E-02	3.66E-05	6	4
GO:0016459	myosin complex	CC	1.30E-02	3.66E-05	6	4
GO:0042302	structural constituent of cuticle	MF	1.45E-02	4.76E-05	14	42
GO:0051015	actin filament binding	MF	2.01E-02	7.54E-05	6	5
GO:0044237	cellular metabolic process	BP	4.31E-02	1.82E-04	7	228
GO:0044238	primary metabolic process	BP	4.77E-02	2.41E-04	6	210

Adult Male

GO ID	GO name	Go Category	FDR	P-value	Nr Test	Nr Reference
GO:0003774	motor activity	MF	1.39E-02	2.61E-05	6	4
GO:0016459	myosin complex	CC	1.39E-02	2.61E-05	6	4
GO:0051015	actin filament binding	MF	2.31E-02	5.40E-05	6	5
GO:0008152	metabolic process	BP	4.26E-02	2.19E-04	10	282

Adult Female

GO ID	GO name	Go Category	FDR	P-value	Nr Test	Nr Reference
GO:0006869	lipid transport	BP	3.57E-07	1.67E-10	13	9
GO:0005319	lipid transporter activity	MF	5.37E-07	7.55E-10	12	8

GO:0009987	cellular process	BP	3.67E-03	1.38E-05	9	306
GO:0003774	motor activity	MF	4.13E-03	2.32E-05	6	4
GO:0016459	myosin complex	CC	4.13E-03	2.32E-05	6	4
GO:0051015	actin filament binding	MF	7.34E-03	4.81E-05	6	5
GO:0044238	primary metabolic process	BP	3.07E-02	2.44E-04	5	211

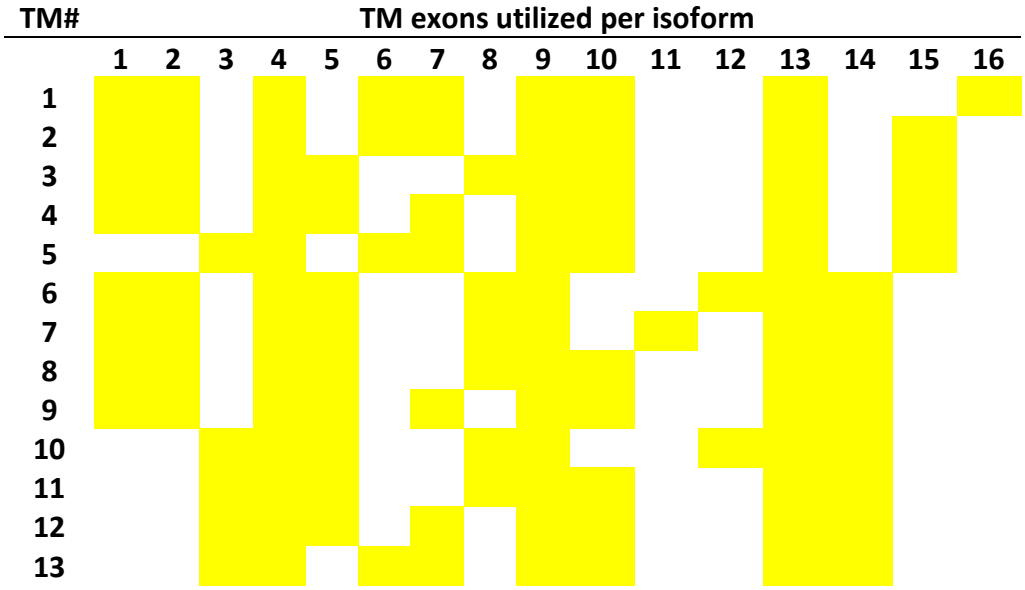
Go categories: MF, Molecular function; BP, Biological Process; CC, Cellular Component. Red and green highlighted GO IDs indicate over and underrepresented in test group, respectively.

APPENDIX B ANNOTATIONS OF QUANTIFIED PROTEINS FROM ACHETA DOMESTICUS (XLSX, 380 KB)

Annotations of quantified proteins

APPENDIX C SUPPLEMENTARY FIGURES AND TABLES FROM CHAPTER 3: PROTEOMIC EVALUATION OF YELLOW MEALWORM, TENEBRIO MOLITOR, WITH EXPERIMENTAL GENETIC MODIFICATIONS

Appendix C Table 3.1.1 *Tribolium castaneum* tropomyosin-2 exon structure utilized to manually curate *T. molitor* protein sequences



Appendix C Table 3.1.2 *T. molitor* tropomyosin-2 exon nucleotide sequences

Exon #	Exon nucleotide sequence
1	ATGGACGCGATCAAGAAGAAGATGCAAGCGATGAAGCTTGAGAAGGATAACGCTTTGGACCGTGCCATCCAGAACGAACAGCAAGC CAAAGACGCCAACCTTCGCGGAGAAAAA
2	CTCGAAGAGGAGGCCCGCACTCTCCAGAAGAAGATCCAGACCATTGAGAATGAACTCGATCAGACCCAGGAACAGCTCACCCAAGT CAACGGCAAACTGGAAGAGAAGGAGAAGGCCCTCCAGACC
3	ATGACGACGAACATACAGCAGGGCTCCCTCCTGGACGTgctcaagaaaaaaatgaggCAGACCAAGGAGGAGATGGAACGTTACAA GGACGAGTGCGAAGAGTACAACAAACGACTTCATGCGGAATGCATGCGGAGAGAGGAA
4	GCCGAGTCAGAAGTTGCGGCCCTCAACCGCCGCATCCAACTCCTGGAGGAAGACTTGGAGCGTTCGGAAGAACGCTTGGCCACTGC CACCGCCAAATTGGCCGAAGCTTCCGCCGCCGCCGACGAGAGCGAAAG
5	GATACGCAAAGCCTTAGAAAACAGGACCAACATGGAAGACGATAGAGTAGCCGTCTTGGAAAGTCAACTGTCCCAAGCGAAACTAA TAGCCGAAGAGGCCGACAAAAAATACGAAGAG
6	GCAACGCAAGGTTCTCGAGAACCGCTCCCTCGCCGATGAAGAGCGCATGGATGCCCTAGAAAACCAGCTAAAGGAAGCCCGTTTCT TGGCTGAAGAAGCCGATAAAAAATACGATGAG
7	GTAGCTCGTAAATTGGCCATGGTTGAAGCCGACTTGGAGAGAGCAGAAGAACGCGCCGAAACAGGAGAATC
8	GTTGCCCGAAAATTGGTTCTTATGGAACAAGACTTAGAACGAGCCGAAGAACGCGCTGAACAGAGCGAGAG
9	CAAAATCGTAGAGCTTGAGGAGGAACTTCGCGTCGTTGGCAACAACTTGAAGTCCCTAGAAGTGTCCGAGGAAAAG
10	GCCAACCAACGCGAAGAAGAgtacaaaaatcaaattaagaACTTGACCACCCGTCTAAAGGAG
11	GCGGCTGTTACCAGAGAACATAGCGAGGATAAAATCCGTTCCATCTCAGATAAACTGCGTGAA
12	GCTACGTTAAAAGAGGAGGAATATTCGGTTACCCTGAAACAGGTGGATCAACGATTGCAAGAG
13	GCTGAGGCTCGCGCCGAGTTTGCCGAACGTTCGGTACAGAAACTCCAGAAGGAGGTCGACAGACTAGAAG
14	ACGATCTCTTGGCAGAAAAGGAAAAGAACAAACTGTTATCCGACGAGATGGAAGCTACATTGCATGATATCCAAAATATG
15	ATGAGCTCGTCGCCGAAAAGGAGCGTTACAAGGAAATCGGCGACGACTTGGACACGGCTTTCGTCGAACTCATCTTG
16	ATGAACTAGTCGACGAGAAGGGCAAGTACAAGGCCATCGCCGACGAGATGGACCAGACGATGGCCGACTTGGCTGGATAT

Nucleotide sequences are from the reverse strand of contigs JABDTM010019790.1 and JABDTM010018031.1.

Appendix C Figure 3.1.1 Alignments and sequence coverage for arginine kinase and tropomyosins

A. Arginine kinase

```
XP_971800.2      MVDAAVLEKLEAGFKKLEASDSKSLLKKYLTRELFDKLKTKKTSFGSTLLDVIQSGLENH   60
g5623t1          MVDAAVLEKLEAGFKKLEASDSKSLLKKYLTRELFDNLKTKKTSFGSTLLDVIQSGLENH   60
                 ************************************:***********************

XP_971800.2      DSGIGIYAPDADSYSVFADLFDPIIEDYHGGFKKTDKHPPKDWGDVNAFGNLDPAGEFVV   120
g5623t1          DSGIGIYAPDAEAYSVFSDLFDPIIEDYHGGFKKSDKHPPKNWGDTSVFGNLDPAGEYIV   120
                 ***********::****:****************:******:***...*********::*

XP_971800.2      STRVRCGRSLEGYPFNPCLTEEQYKEMEQKVSSTLSGLEGELKGTFYPLTGMSKEVQQKL   180
g5623t1          STRVRCGRSLEGYPFNPCLTEEQYKEMEQKVSSTLSGLEGELKGTFYPLTGMSKEVQQKL   180
                 ************************************************************

XP_971800.2      IDDHFLFKEGDRFLQAANACRFWPTGRGIFHNDAKTFLVWCNEEDHLRIISMQMGGDLGQ   240
g5623t1          IDDHFLFKEGDRFLQAANACRFWPSGRGIFHNDAKTFLVWCNEEDHLRIISMQMGGDLGQ   240
                 ************************:***********************************

XP_971800.2      VYRRLVTGVNDIEKRLPFSHSDRFGFLTFCPTNLGTTVRASVHIKVPKLAANKAKLDEVA   300
g5623t1          VYRRLVTAVNDIEKRIPFSHSDRFGFLTFCPTNLGTTVRASVHIKVPKLSANKAKLDEVA   300
                 *******.*******:************************************:**********

XP_971800.2      AKFNLQVRGTRGEHTEAEGGVYDISNKRRMGLTEFDAVKEMYDGISEIIKMEKEL    355
g5623t1          GKFNLQVRGTRGEHTEAEGGVYDISNKRRMGLTEYDAVKEMYDGIAEIIKIEKEL    355
                 .**********************************:**********:****:****
```

B. Tropomyosin-1

```
g11661t1         MDAIKKKMQAMKLEKDNAQDKADAMEGQAKDANLRVEKLNEELRDLQKKLAQVEGDFSST   60
XP_967128.1      MDAIKKKMQAMKLEKDNAQDKADAMEGQAKDANLRVEKLNEELRELQKKLSQVEGDLTTT   60
                 *********************************************:*****:*****:::*

g11661t1         KNNLEQANKDLEEKEKTLTNAESEMAALNRKVQLIEEDLERSEERLTTATTKLAEASQAA   120
XP_967128.1      KNSLEQANKDLEEKEKTLTNAEAEMASLNRKVQTIEEDLERSEERLATATTKLAEASQAA   120
                 **.******************:***:****** ************:**************

g11661t1         DESFRMCKVLENRSQQDEERMDQLTNQLKEARLLAEDADNKSDEVSRKLAFVEDELEVAE   180
XP_967128.1      DESSRMCKVLENRSQQDEERMDQLTNQLKEARLLAEDADNKSDEVSRKLAFVEDELEVAE   180
                 *** ************************************************************

g11661t1         DRVKGGDAKIMELEEELKVVGNSLKSLEVSEEKANQRVEEFKKQLKTLTVKLKEAEARAE   240
XP_967128.1      DRVKGGDAKIMELEEELKVVGNSLKSLEVSEEKANQRVEEFKKQLKTLTVKLKEAEARAE   240
                 ************************************************************

g11661t1         YAEKTVKKLQKEVDRLEDELGINKDRYKSLADEMDSTFAELAGY      284
XP_967128.1      YAEKTVKKLQKEVDRLEDELGINKDRYKSLADEMDSTFAELAGY      284
                 *******************************************
```

C. Tropomyosin-2 isoform

```
ManualTM1.2.n    MDAIKKKMQAMKLEKDNALDRAIQNEQQAKDANLRGEKLEEEARTLQKKIQTIENELDQT   60
```

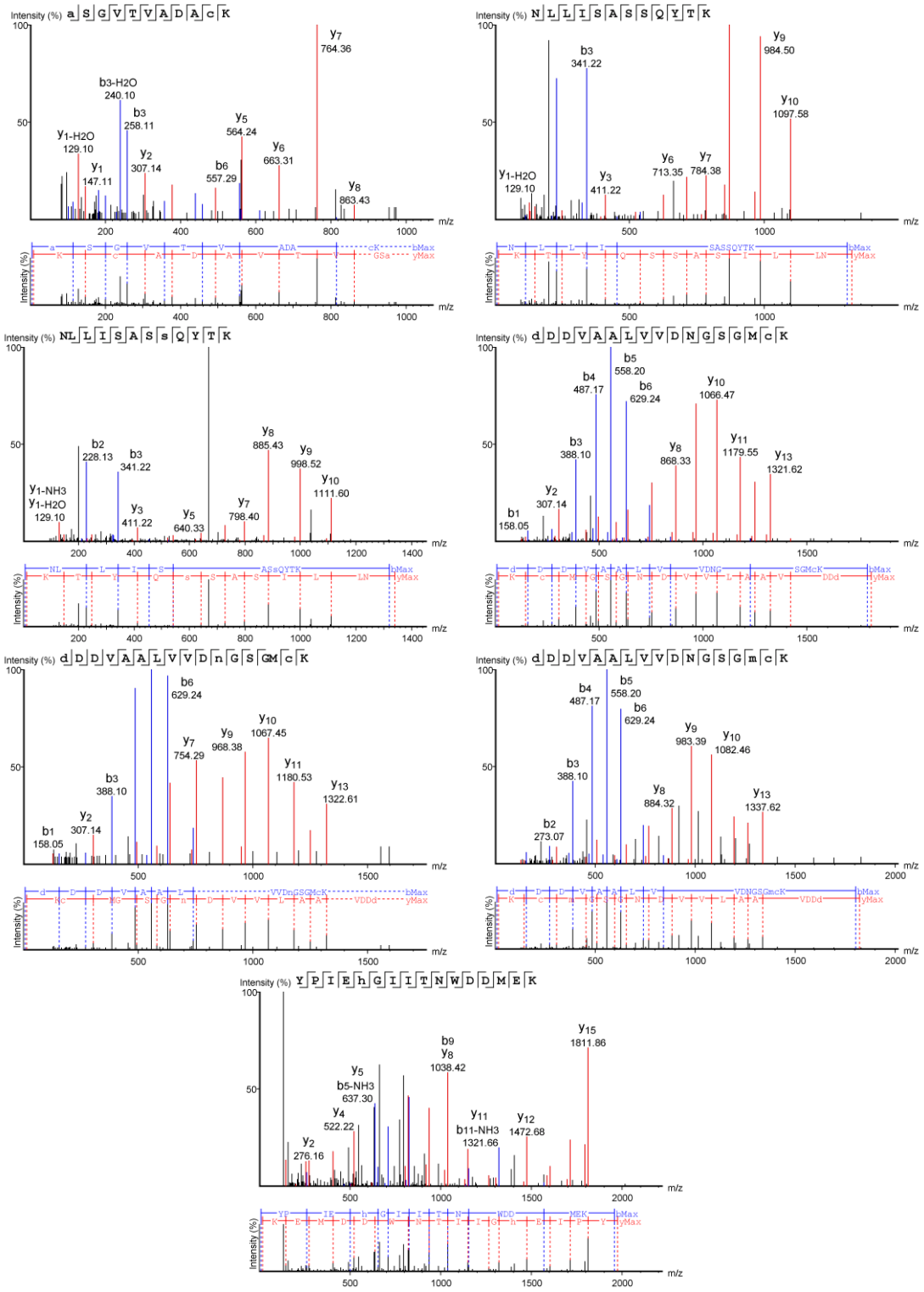
```
XP_008198924.1      MDAIKKKMQAMKLEKDNALDRAIFNEQQAKDANLRAEKLEEEARTLQKKIQTIENELDQT      60
                    ********************** ***********.************************

ManualTM1.2.n       QEQLTQVNGKLEEKEKALQTAESEVAALNRRIQLLEEDLERSEERLATATAKLAEASAAA      120
XP_008198924.1      QEQLTQVNGKLEEKEKALQTAESEVAALNRRIQLLEEDLERSEERLATATAKLAEASAAA      120
                    ************************************************************

ManualTM1.2.n       DESERQRKVLENRSLADEERMDALENQLKEARFLAEEADKKYDEVARKLAMVEADLERAE      180
XP_008198924.1      DESERQRKVLENRSLADEERMDALENQLKEARFLAEEADKKYDEVARKLAMVEADLERAE      180
                    ************************************************************

ManualTM1.2.n       ERAETGESKIVELEEELRVVGNNLKSLEVSEEKANQREEEYKNQIKNLTTRLKEAEARAE      240
XP_008198924.1      ERAEAGESKIVELEEELRVVGNNLKSLEVSEEKANQREEEYKNQIKNLTTRLKEAEARAE      240
                    ****:*******************************************************

ManualTM1.2.n       FAERSVQKLQKEVDRLEDELVAEKERYKEIGDDLDTAFVELIL    283
XP_008198924.1      FAERSVQKLQKEVDRLEDELVAEKERYKEIGDDLDTAFVELIL    283
                    *******************************************
```

Appendix C Figure 3.1.1 Alignments and sequence coverage for arginine kinase and tropomyosins

Alignments of identified *T. molitor* (A) arginine kinase and (B) tropomyosin 1 predicted by Augustus and (C) a manually curated tropomyosin 2 isoform aligned to the top hits identified from *Tribolium castaneum* using Clustal Omega (12.4). Peptide coverage per protein is represented by red font.

Appendix C Table 3.1.3 Mass spectrometry multi-round search metadata

	Round 1	Round 2	Round 3	Round 4
# MS scans	154096	154096	154096	154096
#MS/MS scans	208046	24772	19134	17486
Peptide-Spectrum matches at FDR < 0.01	76377	1473	1334	760
Peptides	5478	144	159	119
Peptide FDR	0.7%	0.7%	0.6%	<0.1%
Proteins	1010	149	217	118
Protein Groups	873	130	172	106
Protein FDR	12.3%	59.3%	14.3%	66.7%
# *De novo* only scans	22772	19134	17486	16723

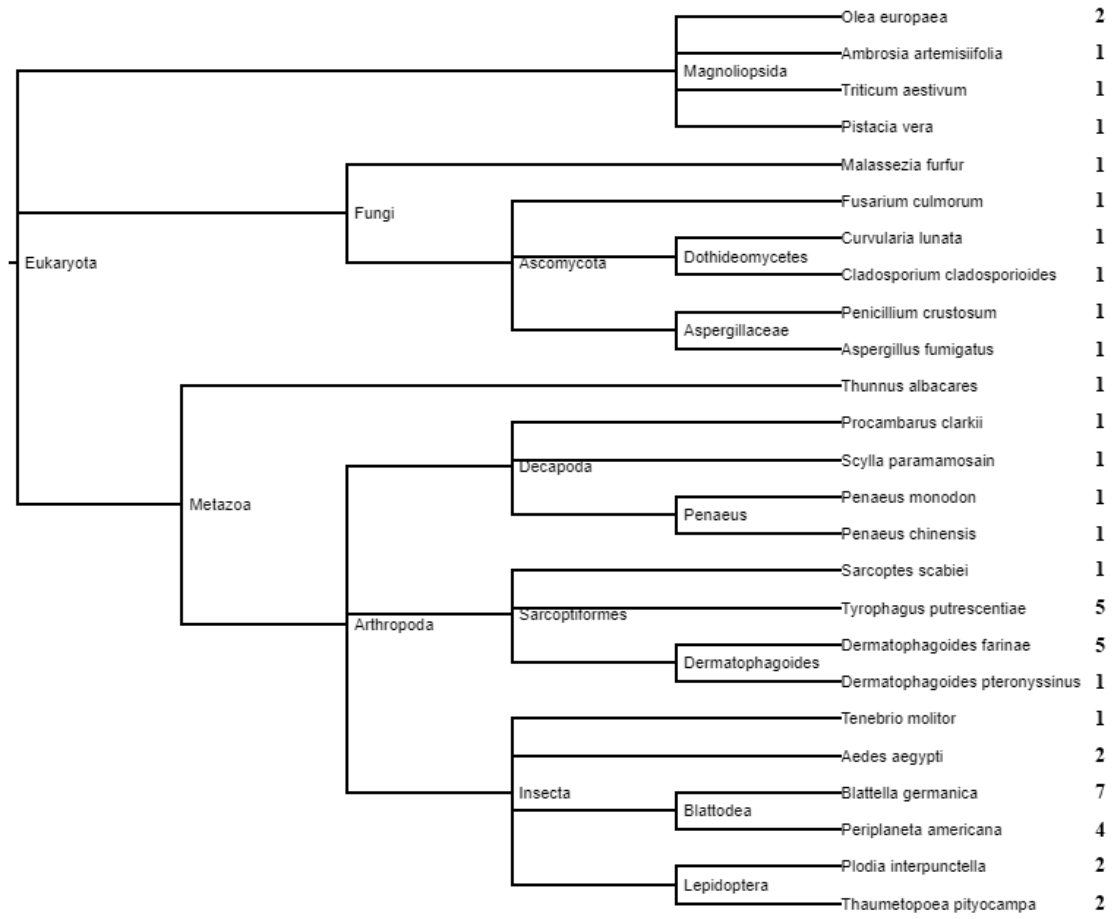


Appendix C Figure 3.1.2 Spectra of peptides with posttranslational modifications and relevant unmodified forms

Appendix C Table 3.1.4 Peptide posttranslational modification characteristics

Peptide	m/z	z	RT	-10lgP	ppm	PTM*	Accession	Description
A(+42.01)SGVTVADAC(+57.02)K	560.7686	2	33.96	75.95	-0.3	A_2 Acetylation, C_{11} Carbamidomethyl	A0A678P951	Cofilin/actin-depolymerizing factor
NLLISASSQYTK	662.8601	2	41.28	51.47	0.3		g13695t1	Transferrin
NLLISASS(+14.02)QYTK	669.8672	2	41.8	51.17	0.5	S_{666} Methylation	g13695t1	Transferrin
D(+42.01)DDVAALVVDNGSGMC(+57.02)K	904.3957	2	48.86	82.71	0.3	D_3 Acetylation, C_{18} Carbamidomethyl	g4024t1	Actin, muscle
D(+42.01)DDVAALVVDN(+.98)GSGMC(+57.02)K	904.8878	2	47.38	71.38	0.4	D_3 Acetylation, N_{13} Deamidation, C_{18} Carbamidomethyl	g4024t1	Actin, muscle
D(+42.01)DDVAALVVDNGSGM(+15.99)C(+57.02)K	912.3936	2	45.86	74.65	0.7	D_3 Acetylation, M_{17} Oxidation, C_{18} Carbamidomethyl	g4024t1	Actin, muscle
YPIEH(+14.02)GIITNWDDMEK	658.9808	3	45.94	85.75	0.7	H_{74} Methylation	g4024t1	Actin, muscle

*PTM: Posttranslational modification



Appendix C Figure 3.1.3 Phylogeny of the origin species of allergen predictions

Phylogenetic tree of species produced with IcyTree whose allergenic sequences were significant matches to *A. domesticus* proteins. Branch length is arbitrary. Numbers to the right of each species represents the number of hits attributed to the species.

Appendix C Table 3.1.5 Enriched Gene Ontology (GO) terms of the top 10% most abundant proteins per sample against all proteins

Wildtype

GO ID	Go Name	Go Category	FDR	P-value	NR test	NR reference
GO:0042302	structural constituent of cuticle	MF	9.97E-06	3.17E-08	17	29
GO:0005737	cytoplasm	CC	7.94E-05	2.84E-07	8	284
GO:0043231	intracellular membrane-bounded organelle	CC	5.58E-04	3.11E-06	2	171
GO:0019538	protein metabolic process	BP	0.009885	7.08E-05	1	124
GO:0031012	extracellular matrix	CC	0.01602	1.27E-04	8	12
GO:0044249	cellular biosynthetic process	BP	0.024143	2.02E-04	1	115

GO:0005488	binding	MF	0.033049	2.89E-04	15	304
GO:0016491	oxidoreductase activity	MF	0.034324	3.41E-04	1	111
GO:0044281	small molecule metabolic process	BP	0.034324	3.41E-04	1	111

Genetically modified 1

GO ID	Go Name	Go Category	FDR	P-value	NR test	NR reference
GO:0003824	catalytic activity	MF	6.18E-06	9.83E-09	9	311
GO:0042302	structural constituent of cuticle	MF	2.36E-05	7.50E-08	17	29
GO:0005737	cytoplasm	CC	8.35E-05	2.99E-07	9	283
GO:0043231	intracellular membrane-bounded organelle	CC	2.61E-04	1.14E-06	2	171
GO:0019538	protein metabolic process	BP	0.006375	4.56E-05	1	124
GO:0031012	extracellular matrix	CC	0.023341	1.86E-04	8	12
GO:0044260	cellular macromolecule metabolic process	BP	0.041788	3.49E-04	1	105

Genetically modified 2

GO ID	Go Name	Go Category	FDR	P-value	NR test	NR reference
GO:0005737	cytoplasm	CC	2.04E-05	5.53E-08	8	284
GO:0003824	catalytic activity	MF	2.04E-05	5.69E-08	10	310
GO:0042302	structural constituent of cuticle	MF	2.09E-05	7.50E-08	17	29
GO:0043231	intracellular membrane-bounded organelle	CC	2.61E-04	1.14E-06	2	171
GO:0019538	protein metabolic process	BP	6.75E-03	4.56E-05	1	124
GO:0031012	extracellular matrix	CC	2.33E-02	1.86E-04	8	12
GO:0044260	cellular macromolecule metabolic process	BP	4.18E-02	3.49E-04	1	105
GO:0034641	cellular nitrogen compound metabolic process	BP	4.76E-02	4.16E-04	2	120

Go categories: MF, Molecular function; BP, Biological Process; CC, Cellular Component. Red and green highlighted GO IDs indicate over and underrepresented in test group, respectively.

Appendix C Table 3.1.6 Enriched Gene Ontology (GO) terms of the top 10% most abundant proteins per sample against robust proteins

Wildtype

GO ID	Go Name	Go Category	FDR	P-value	NR test	NR reference
GO:0003824	catalytic activity	MF	1.55E-06	2.69E-09	5	241
GO:0005737	cytoplasm	CC	4.18E-05	1.45E-07	7	238
GO:0042302	structural constituent of cuticle	MF	6.51E-05	2.54E-07	16	26
GO:0043231	intracellular membrane-bounded organelle	CC	0.005361	3.71E-05	2	129
GO:0005488	binding	MF	0.010576	7.78E-05	13	253
GO:0044249	cellular biosynthetic process	BP	0.02062	1.61E-04	1	100
GO:0019538	protein metabolic process	BP	0.030571	2.78E-04	1	99
GO:0044281	small molecule metabolic process	BP	0.030571	2.77E-04	1	95

Genetically modified 1

GO ID	Go Name	Go Category	FDR	P-value	NR test	NR reference
GO:0003824	catalytic activity	MF	6.28E-06	2.09E-08	6	240
GO:0005737	cytoplasm	CC	6.28E-06	2.18E-08	6	239
GO:0042302	structural constituent of cuticle	MF	6.51E-05	2.54E-07	16	26
GO:0043231	intracellular membrane-bounded organelle	CC	8.29E-04	5.38E-06	1	130
GO:0005488	binding	MF	0.00899	7.78E-05	13	253
GO:1901576	organic substance biosynthetic process	BP	0.009744	9.28E-05	1	105
GO:0031012	extracellular matrix	CC	0.01131	1.17E-04	8	10
GO:0044271	cellular nitrogen compound biosynthetic process	BP	0.01562	1.69E-04	0	82
GO:0019538	protein metabolic process	BP	0.023778	2.78E-04	1	99
GO:0044281	small molecule metabolic process	BP	0.023778	2.77E-04	1	95

Genetically modified 2

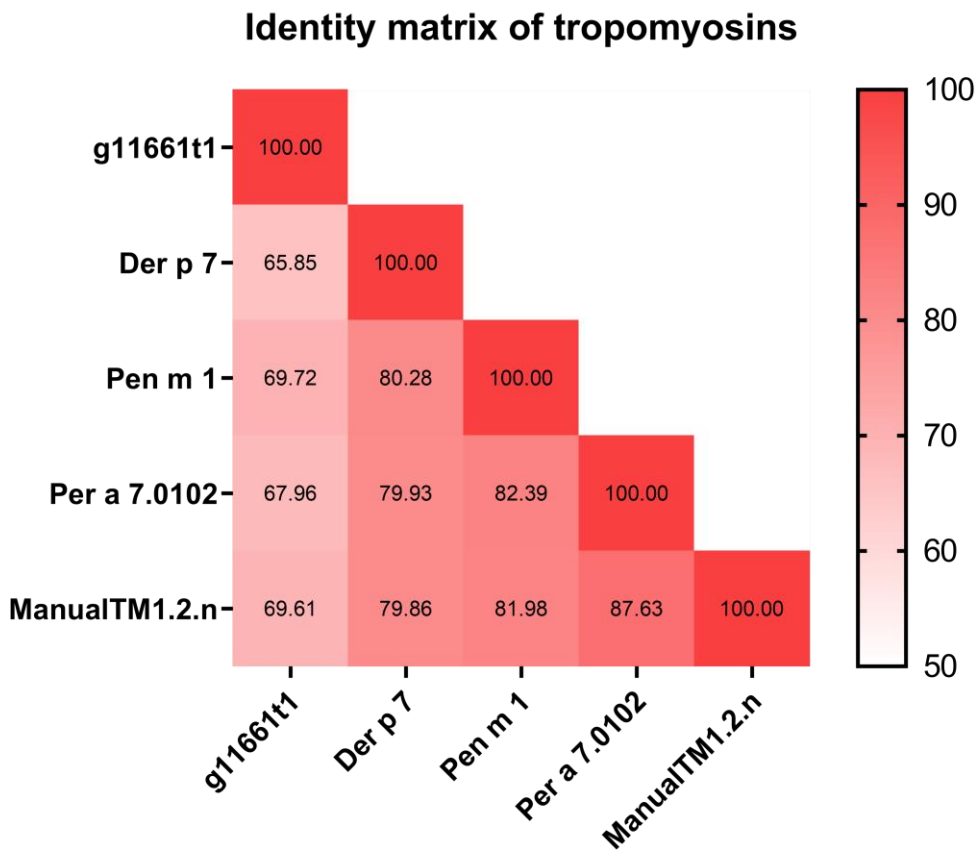
GO ID	Go Name	Go Category	FDR	P-value	NR test	NR reference
GO:0003824	catalytic activity	MF	8.03E-06	2.09E-08	6	240
GO:0005737	cytoplasm	CC	4.18E-05	1.45E-07	7	238
GO:0042302	structural constituent of cuticle	MF	4.68E-04	1.82E-06	15	27
GO:0043231	intracellular membrane-bounded organelle	CC	0.005361	3.71E-05	2	129
GO:0031012	extracellular matrix	CC	0.01508	1.17E-04	8	10
GO:0044249	cellular biosynthetic process	BP	0.019534	1.61E-04	1	100
GO:0019538	protein metabolic process	BP	0.029182	2.78E-04	1	99

GO:0044281	small molecule metabolic process	BP	0.029182	2.77E-04	1	95

Go categories: MF, Molecular function; BP, Biological Process; CC, Cellular Component. Red and green highlighted GO IDs indicate over and underrepresented in test group, respectively.

Appendix C Figure 3.1.4 Alignment of *T. molitor* identified tropomyosins with known allergenic tropomyosins

T. moltior tropomyosins (g11661t1 and ManualTM1.2.n) were aligned with allergenic tropomyosins from shrimp (*Penaeus modon,* Pen m 1, Uniprot A1KYZ2), house dust mite (*Dermatophagoides pteronyssinus,* Der p 10, Uniprot O18416), and cockroach (*Periplaneta americana,* Per a 7.0102, Uniprot P0DSM7) using Clustal Omega (1.2.4).



Appendix C Figure 3.1.5 Identity matrix of alignment of *T. moltior* tropomyosins with known allergenic tropomyosins

T. moltior tropomyosins (g11661t1 and ManualTM1.2.n) were aligned with allergenic tropomyosins from shrimp (*Penaeus modon*, Pen m 1, Uniprot A1KYZ2), house dust mite (*Dermatophagoides pteronyssinus*, Der p 10, Uniprot O18416), and cockroach (*Periplaneta americana*, Per a 7.0102, Uniprot P0DSM7) using Clustal Omega (1.2.4).

APPENDIX D ANNOTATIONS OF QUANTIFIED PROTEINS FROM TENEBRIO MOLITOR (XLSX, 337 KB)

Annotations of quantified proteins

APPENDIX E SUPPLEMENTARY FIGURES AND TABLES FROM CHAPTER 4: PERSISTENCE OF PEANUT ALLERGEN-DERIVED PEPTIDES THROUGHOUT EXCESSIVE DRY THERMAL PROCESSING

Appendix E Table 4.1.1 Posttranslational modifications added to PTM search

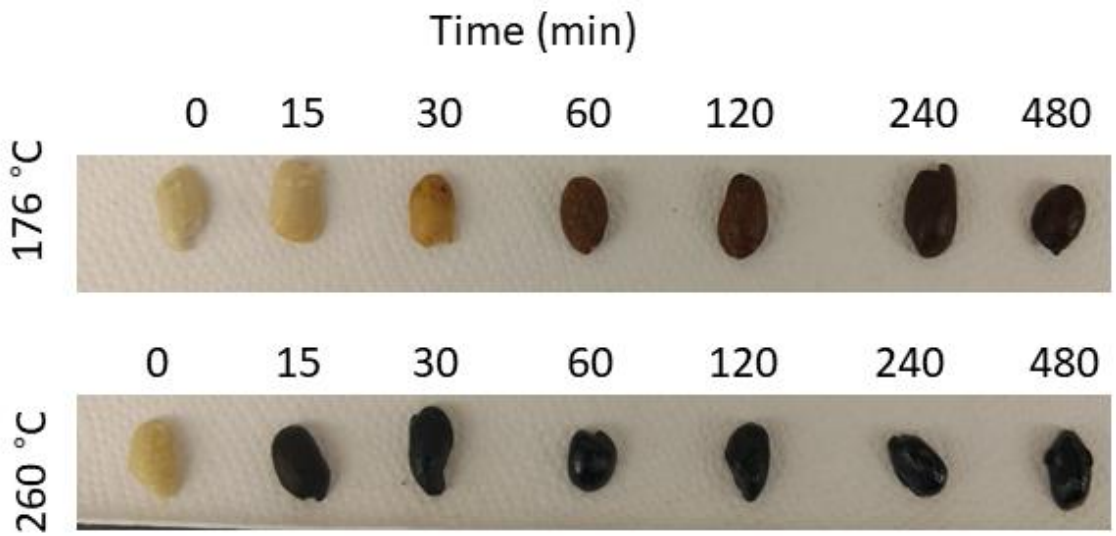
Name	Monoisotopic mass	Specificity	Formula	C	H	O
Hydroxynonenal - Michael adduct	156.11500	CHKR	C9H16O2	9	16	2
Hydroxynonenal - 2-pentilpyrrole	120.09390	RK	C9H12O0	9	12	0
Hydroxynonenal - Dehydropentylfuran (CHRK) Hydroxynonenal - Shiff base (RK)	138.10450	CHKR	C9H14O1	9	14	1
Malondialdehyde - Dihydropyridine-lysine	134.03680	K	C8H6O2	8	6	2
Malondialdehyde - Argpyrimidine	36.00000	R	C3H0O0	3	0	0
Malondialdehyde – N-propenal-Lysine	54.01056	K	C3H2O1	3	2	1
Acrolein - Michael Adduct	56.02621	CHK	C3H4O1	3	4	1
Acrolein - N-2 (4 hydroxy-tetrahydro-pyrimidyl) Ornitine	56.02621	R	C3H4O1	3	4	1
Acrolein - Double Michael Adduct	112.05240	K	C6H8O2	6	8	2
Acrolein - Schiff Base	38.01565	K	C3H2O0	3	2	0
Acrolein - Formyl-dehydro-piperidyl-lysine	94.04186	K	C6H6O1	6	6	1
Acrolein - Methylpyridine-lysine	77.03913	K	C6H5O0	6	5	0
Glyoxal-derived Hydroimidazolone	39.99491	R	C2H0O1	2	0	1
Methylglyoxal-derived Hydroimidazolone Imidazolonylornithine	54.01056	R	C3H2O1	3	2	1
Glyoxal-derived Hemiaminal (RK) Glyoxal-derived Dihydroxyimidazoline (R) N^{ω}-carboxymethylarginine (R) Carboxymethyl Lysine (K) Glycollolysine (K)	58.00548	RK	C2H2O2	2	2	2
Triosone Hydroimisazolone	70.00548	R	C3H2O2	3	2	2
Methylglyoxal-derived Hemiaminal (RK) Carboxyethyllysine (K) Methylglyoxal-derived Dihydroxyimidazoline (R)	72.02113	RK	C3H4O2	3	4	2
Argpyrimidine	80.02621	R	C5H4O1	5	4	1
Triosone-derived Hemiaminal (RK) Triosone-derived Dihydroxyimidazoline (R)	88.01604	RK	C3H4O3	3	4	3
Triosidine-carbaldehyde	93.03404	K	C6H5O1	6	5	1
Pyrraline	108.02110	K	C6H4O2	6	4	2
3-Deoxy-pentosone Hydroimidazolone	114.03170	R	C5H6O3	5	6	3
Trihydroxy-triosidine	125.02390	K	C6H5O3	6	5	3
Dihydropyrimidine Ornithine	126.03170	R	C6H6O3	6	6	3
Pronyl-lysine	126.03170	K	C6H6O3	6	6	3

Hexose Didehydrate						
3-Deoxypentosone-derived hemiaminal (RK)						
3-Deoxypentosone-derived Dihydroxyimidazoline (R)	132.04230	RK	C5H8O4	5	8	4
N^{ε}-(5,6-dihydroxy-2,3-dioxohexyl)-Lys Hexose Dehydrate	144.04230	K	C6H8O4	6	8	4
3-deoxyglucosone-derived Hydroimidazolone Tetrahydropyrimidine Ornithine	144.04230	R	C6H8O4	6	8	4
3-deoxyglucosone-derived Hemiaminal (RK)						
3-deoxyglucosone-derived Dihydroxyimidazoline (R) Hexose (K)	162.05280	RK	C6H10O5	6	10	5
Glucosone-derived Hemiaminal	178.04770	RK	C6H10O6	6	10	6
Alkyl Formyl Diglycosyl Pyrrole	270.07400	K	C12H14O7	12	14	7
Acetone aldol condensation	98.07320	KHR	C6H10O1	6	10	1

Appendix E Table 4.1.2 Lists of peptides used for quantification of peanut allergens

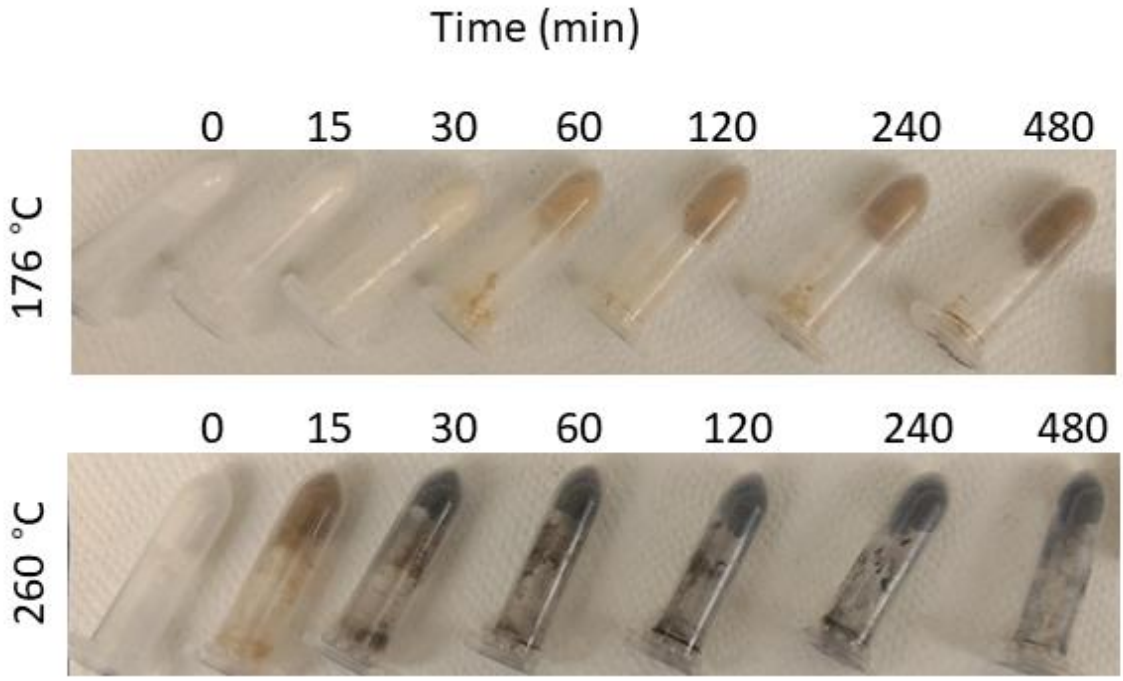
Allergen or isoform	Peptide	Robust in: Raw peanut	Robust in: Roasted peanut	Mass
Ara h 1	NNPFYFPSR		X	1140.5352
Isoforms	IPSGFISYILNR		X	1378.7609
1/2	SFNLDEGHALR	X	X	1257.6101
	NTLEAAFNAEFNEIR	X		1737.8322
	GTGNLELVAVR	X		1127.6299
Ara h 2	C(+57.02)MC(+57.02)EALQQIMENQSDR		X	2011.822
Isoforms	NLPQQC(+57.02)GLR	X	X	1084.5448
1/2	C(+57.02)C(+57.02)NELNEFENNQR	X	X	1725.6835
	ANLRPC(+57.02)EQHLMQK	X		1623.7974
Ara h 6	C(+57.02)DLDVSGGR	X	X	977.4236
Isoforms	ELMNLPQQC(+57.02)NFR	X	X	1548.7177
1/2	VNLKPC(+57.02)EQHIMQR	X	X	1651.8286
Ara h 3	WLGLSAEYGNLYR		X	1540.7673
Isoforms	RPFYSNAPQEIFIQQGR	X	X	2050.0383
4/5/10/	FNLAGNHEQEFLR	X	X	1573.7637
13/17/20	SPDIYNPQAGSLK	X		1388.6936
Ara h 3	LNALTPDNR		X	1012.5302
Isoforms	GIPADVLINAFGLR		X	1454.8245
1/11	EGQILLVPQNFAVGK	X	X	1611.8984
	IESQGGITETWNSNHPELR	X		2167.0293
	FYLAGNPEEEHPETQQQQPQTR	X		2626.2048
Ara h 3	GGHITSLNTPNMAVLQYLQLGLDR		X	2610.3589
Isoforms	EAQEGNVFSGLALETLIGSFNVQR		X	2578.3027
2/12	TSDNPIINTLAGELSLVR	X	X	1912.0265
	GVMEIVVTGC(+57.02)R	X		1219.6053
	AGSDAFDWVAIK	X		1278.6244
Ara h 3	VLPVDVVANMYQVSR		X	1688.892
Isoforms	LPILADLQLSAER	X	X	1437.8191
8/15	NIVMVEGGLDVVRPEPGSR	X	X	2023.052
	FYIAGNTEDEHGEGGR	X		1750.7546
Ara h 3	VFDEELQEGHVLVLPQNFAVAVR		X	2608.3647

Isoform	ILSPEREEFDGR		X	1446.7102
18	EIVQNLR	X	X	870.4923
	IDSEGGFIETWNPK	X		1591.7518
	QEQEFLQYQHQHGGPR	X		1980.9191
Ara h 3	TVNELDLPILNR	X	X	1395.7721
Isoform	SQSEHFLYVAFK	X	X	1454.7194
3/6/19	LGLSAEYGSIHR	X	X	1301.6727
Ara h 3	NAMFVPHYTLNAH		X	1513.7136
Isoform	SSNPDIYNPQAGSLR	X	X	1617.7747
7/16	VYDEELQEGHVLVVPQNFAVAAK	X	X	2554.3066
	AQSENYEYLAFK	X		1461.6776
Ara h 3	GLLLPHYINAPR	X	X	1362.7771
Isoform	EGQILIVPQQFVVAK	X	X	1667.9609
9/14	NDQFQC(+57.02)VGVSALR	X	X	1492.7092



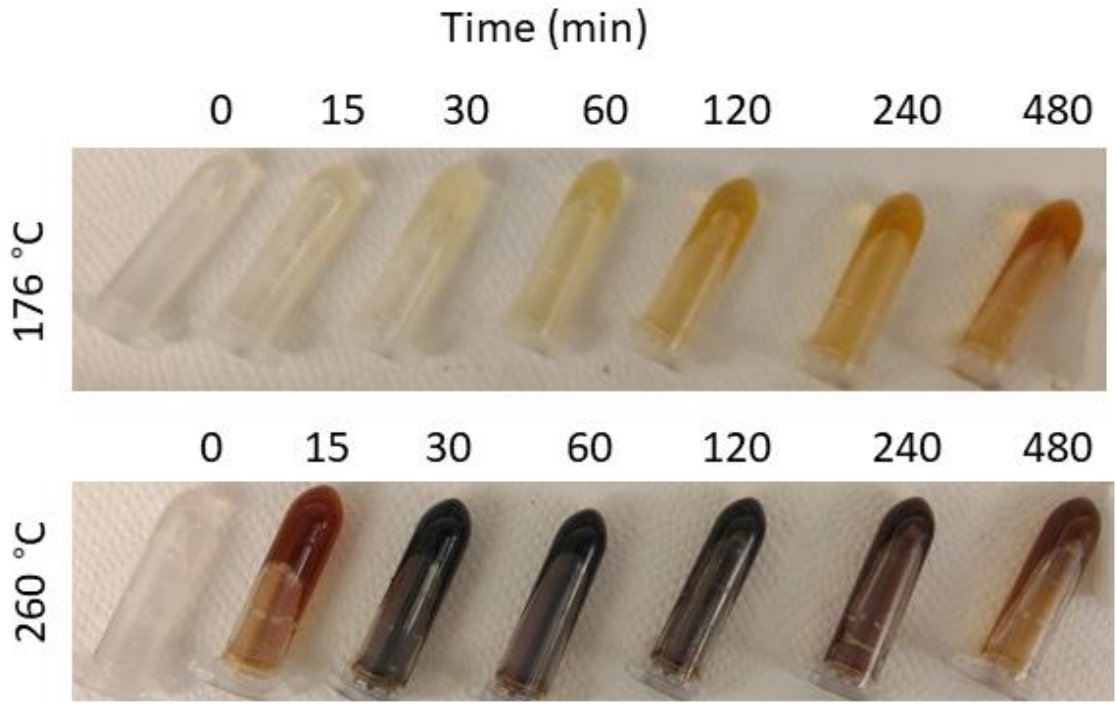
Appendix E Figure 4.1.1 Peanut kernel halves over thermal processing

Peanut kernel halves roasted for times between 0 and 480 minutes at 176 °C (top) and 260 °C (bottom).



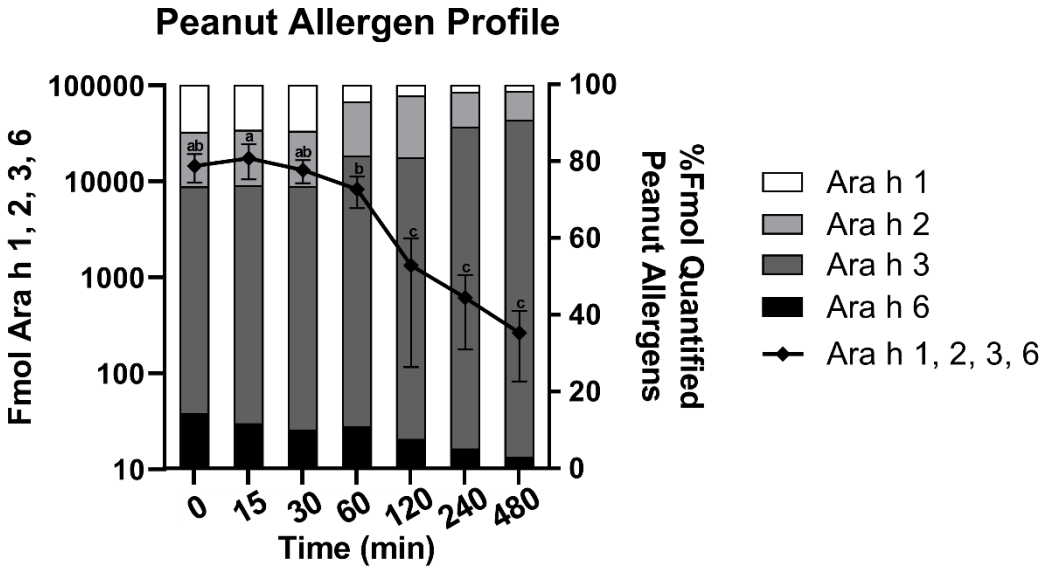
Appendix E Figure 4.1.2 Extracts of peanuts heated at 176 °C and 260 °C

Resultant defatted powders from peanut kernel halves roasted at 176 °C (top) and 260 °C (bottom) for times between 0 and 480 minutes.



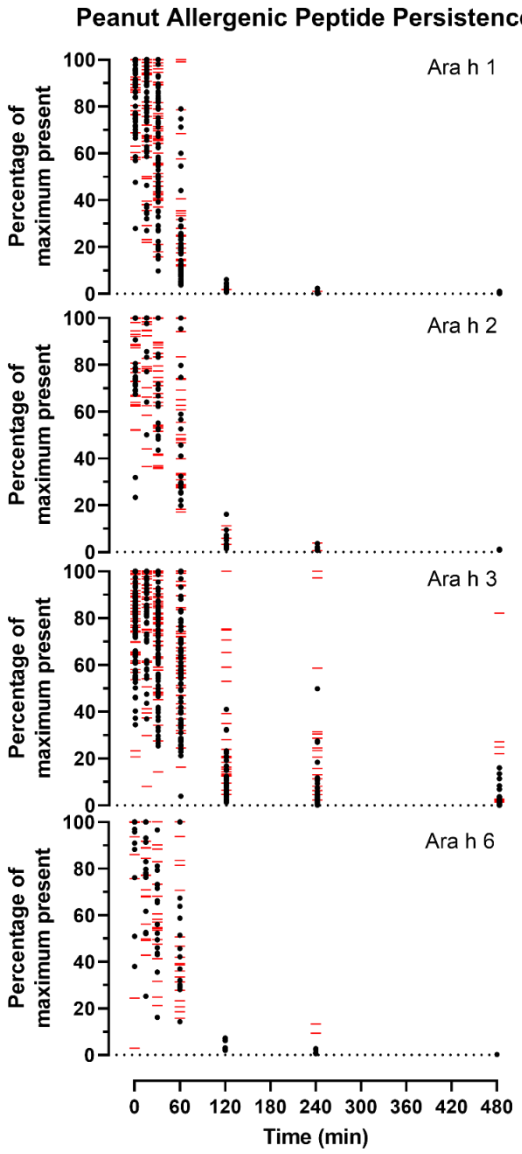
Appendix E Figure 4.1.3 Extracts of peanuts heated at 176 °C and 260 °C

Extracts produced from peanut kernel halves roasted at 176 °C (top) and 260 °C (bottom) for times between 0 and 480 minutes.



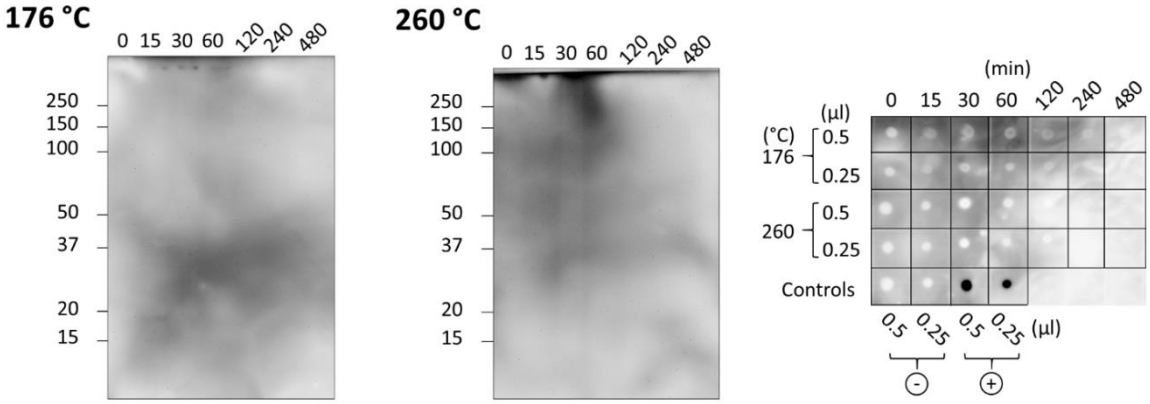
Appendix E Figure 4.1.4 Peanut allergenic profile for peanuts heated at 176 °C

The sum of peanut allergens Ara h 1, 2, 3, 6 per time point are graphed to the left Y-axis and the proportion of each allergen are graphed to the right Y-axis (N=5). For clarity, the peanut allergen profile is presented without CVs but is tabulated elsewhere. Sums of allergenic proteins that have different letters are significantly different via two-way ANOVA ($p<0.05$) and two-stage linear step-up method of Benjamini, Krieger, and Yekutieli (FDR<0.05).



Appendix E Figure 4.1.5 Persistence of peanut allergen peptides at 176 °C

Unmodified peptides (circles) and post-translationally modified peptides (bars) per Ara h 1, 2, 3, and 6 scaled to each peptide's maximum value observed over time. Data are represented as averages of 5 replicates without CV for clarity. Average %, pooled CV, and N of peptides per time point: 0 (85.3%, 21.1%, 358), 15 (87.2%, 22.4%, 390), 30 (68.1%, 33.1%, 364), 60 (54.3%, 56.3%, 312), 120 (13.5%, 130.3%, 112), 240 (9.4%, 187.8%, 84), and 480 (6.0%, 232.6%, 42).



Appendix E Figure 4.1.6 Negative control immunoblots

Control Immunoblots probed using non-atopic unsensitized serum. Transferred immunoblots were loaded volumetrically relative to 30 µg of raw peanut (approximately 1.7 µl) and dot blots were prepared with either 0.5 or 0.25 µl of extracts. Dot blot negative control was bovine serum albumin heated at 260 °C for 8 hours and extracted while positive control was directly applied Morinaga anti-peanut IgG.

APPENDIX F COMPLETE PEPTIDE LIST FOR ROASTED PEANUTS (XLSX, 4853 KB)

Complete peptide
list for roasted pean

APPENDIX G SUPPLEMENTARY FIGURES AND TABLES FROM CHAPTER 5: DETECTION OF FOOD ALLERGEN-DERIVED PEPTIDES FROM EXTRACTIVE-BASED E-CIGARETTE LIQUIDS

Appendix G Table 5.1.1 Comprehensive peptide identifications

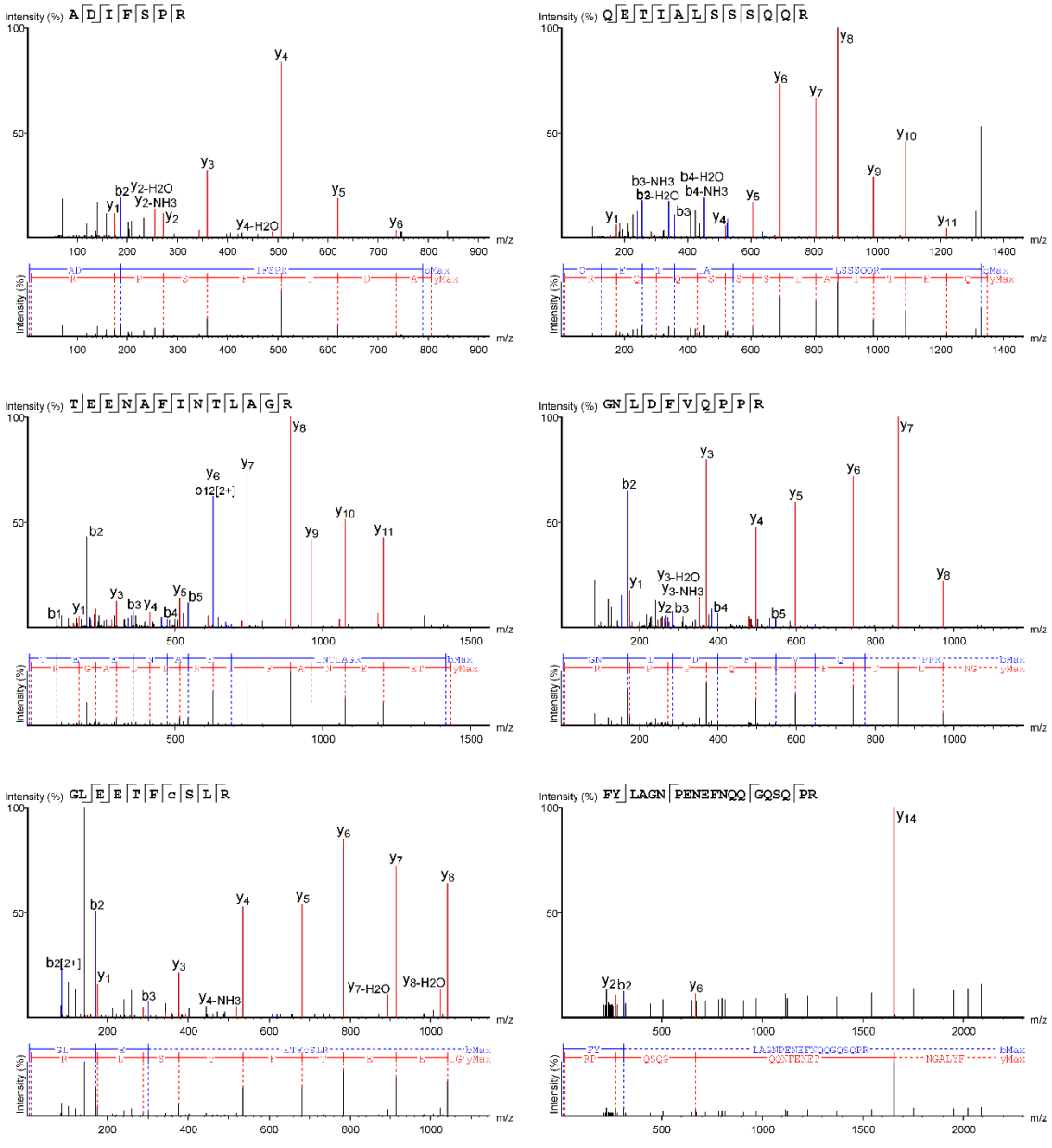
Sequence	Score	Mass	Error (ppm)	Z	RT	Start	End	PTMs
R.QSQLSPQNQC(+57.02)QLNQLQAR.E	71.68	2140.0444	-0.3	3	40.51	23	40	C10: Carbamidomethyl
K.FYLAGNPENEFNQQGQSQPR.Q	29.66	2323.0618	-1.4	2	46.05	242	261	
R.NQIIQVR.G	27.57	869.5083	0.3	2	36.61	322	328	
R.GNLDFVQPPR.G	58.15	1141.588	0.9	2	44.02	329	338	
N.GLEETFC(+57.02)SLR.L	31.42	1210.5652	0.4	2	46.56	368	377	C7: Carbamidomethyl
R.ADIFSPR.A	34.08	804.413	0.5	2	40.06	388	394	
R.ISTLNSHNLPILR.F	57.19	1476.8412	0.4	3	46.52	398	410	
K.TEENAFINTLAGR.T	62.1	1434.7103	-0.5	2	48.46	493	505	
R.ALPDEVLANAYQISR.E	31.14	1658.8628	-0.4	2	52.77	511	525	
R.QETIALSSSQQR.R	51.9	1346.679	0.2	2	35.23	536	547	
R.NLQGQNDNR.N	27.01	1057.4901	1.4	2	14.73	63	71	
R.LSQNIGDPSR.A	36.64	1085.5465	1.0	2	31.9	138	147	
R.VQVVNENGDPILNDEVR.E	32.49	1908.9541	0.1	2	45.09	208	224	
R.TDENGFTNTLAGR.T	39.98	1394.6426	-0.5	2	40.35	253	265	

Appendix G Table 5.1.2 Normalized sample wise peptide presence and quantity

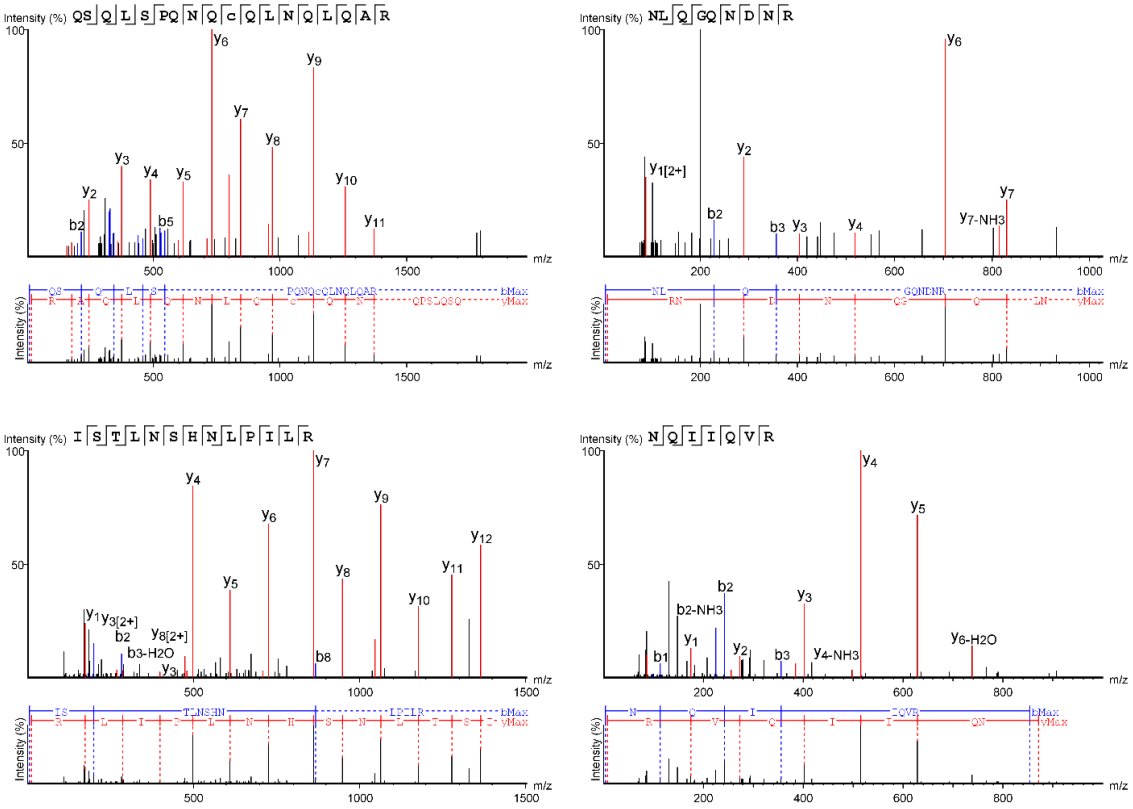
	In-solution digest (fmol)			In-gel digest (fmol)			
Sequence	Almond	Coconut	Peach	Almond	Coconut	Peach	Accession
R.QSQLSPQNQC(+57.02)QLNQLQAR.E	59.92						Q43607
K.FYLAGNPENEFNQQGQSQPR.Q	4.11						Q43607
R.NQIIQVR.G	197.78						Q43607
R.GNLDFVQPPR.G	129.63		3.49				Q43607
N.GLEETFC(+57.02)SLR.L	117.89						Q43607
R.ADIFSPR.A	206.01			42.39	29.69	25.43	Q43607
R.ISTLNSHNLPILR.F	109.27						Q43607
K.TEENAFINTLAGR.T	160.18	4.26	2.15			10.11	Q43607
R.ALPDEVLANAYQISR.E						3.58	Q43607
R.QETIALSSSQQR.R	91.25			14.28	9.37	13.69	Q43607
R.NLQGQNDNR.N	12.96						Q43607, A0A5E4FK23
R.LSQNIGDPSR.A	6.50			3.19			A0A5E4FK23
R.VQVVNENGDPILNDEVR.E	4.43						A0A5E4FK23
R.TDENGFTNTLAGR.T	7.68						A0A5E4FK23

Appendix G Figure 5.1.1 Annotated Peptide spectra

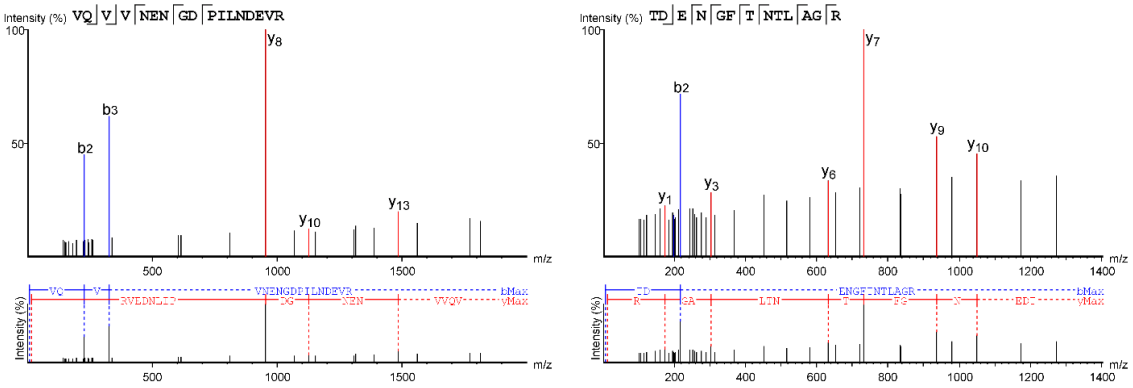
In-solution digest of Almond, protein Q43607



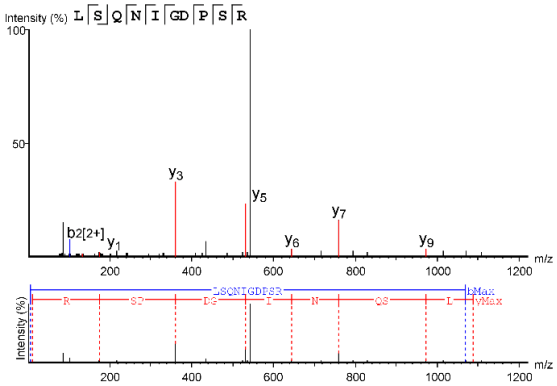
Appendix G Figure 5.1.1 Annotated Peptide spectra



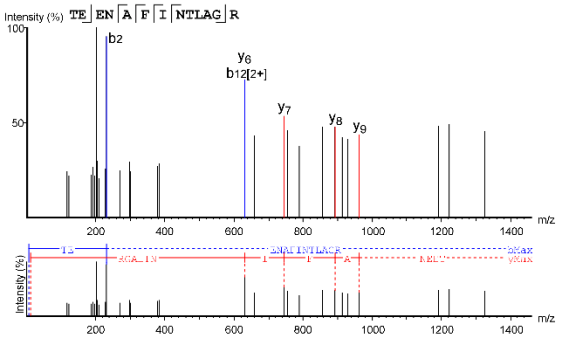
In-solution digest of Almond, protein A0A5E4FK23

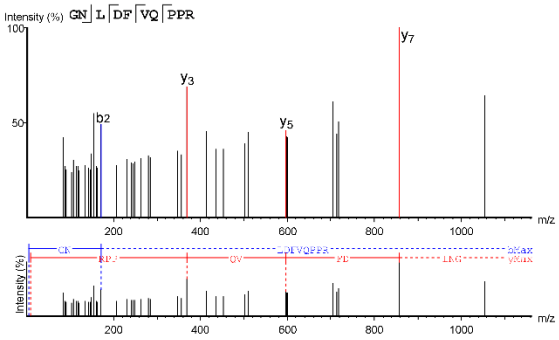


Appendix G Figure 5.1.1 Annotated Peptide spectra

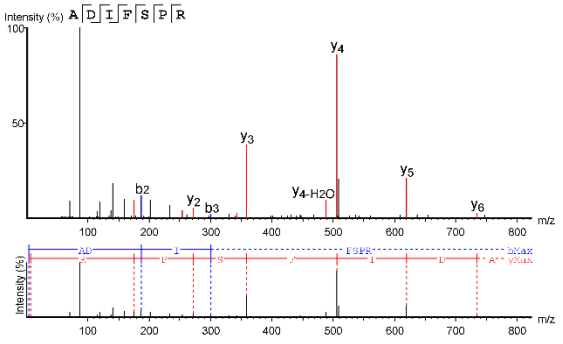


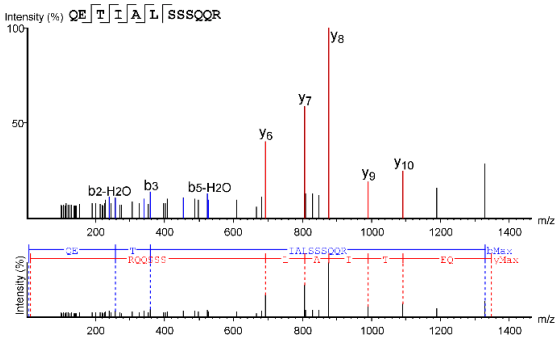
In-solution digest of Peach, protein Q43607



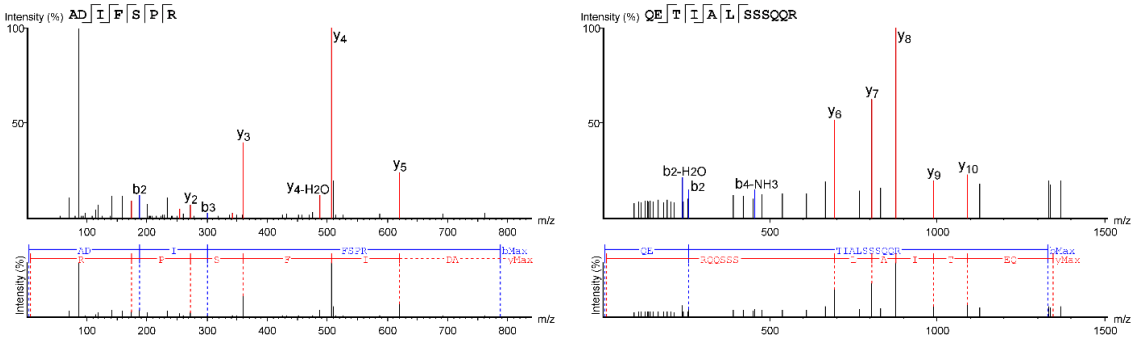


In-gel digestion of Almond, protein Q43607

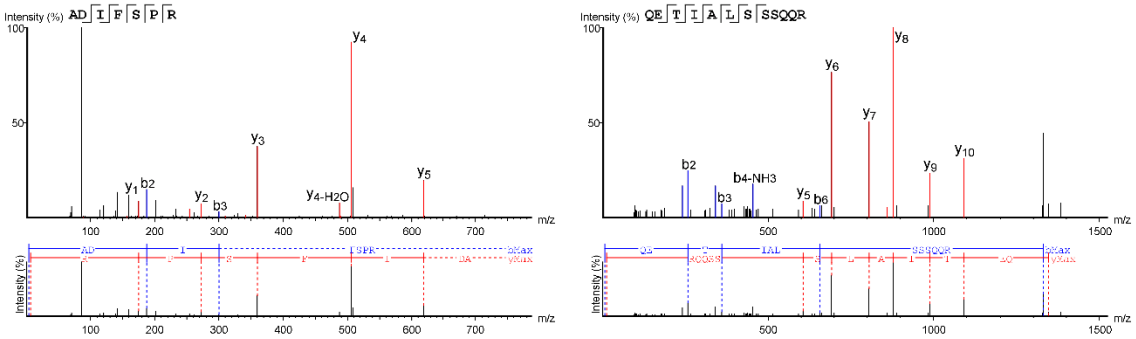


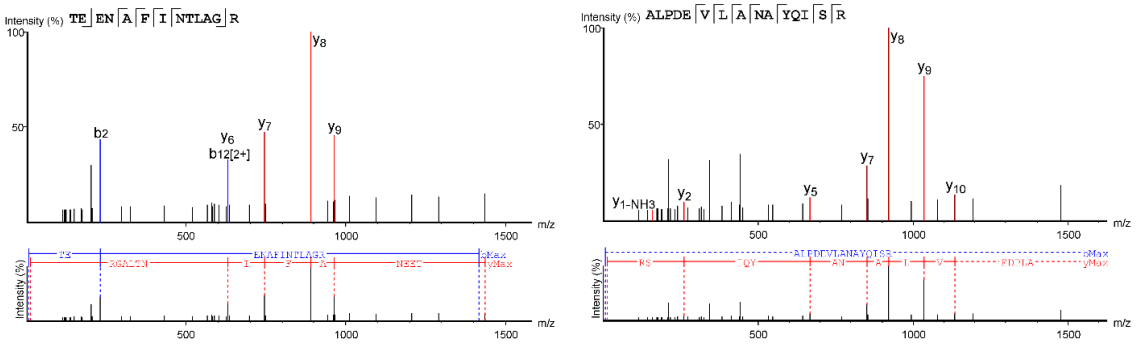


In-gel digestion of Coconut, protein Q43607



In-gel digestion of Peach, protein Q43607





Appendix G Figure 5.1.1 Annotated Peptide spectra

Appendix G Table 5.1.3 Comprehensive protein identifications

Accession	Alternative accession	Score	In-solution digest protein coverage % (Peptides/unique)			In-gel digest protein coverage % (Peptides/unique)			Accession description
			Almond	Coconut	Peach	Almond	Coconut	Peach	
Q43607		174.97	22% (10/9)	2% (1/1)	4% (2/2)	3% (2/2)	3% (2/2)	9% (4/4)	Prunin 1 Pru du 6
	A0A5E4FFS0	174.97	22% (10/9)	2% (1/1)	4% (2/2)	3% (2/2)	3% (2/2)	9% (4/4)	Predicted legumin
	E3SH28	172.92	20% (9/8)	2% (1/1)	4% (2/2)	3% (2/2)	3% (2/2)	9% (4/4)	Prunin 1 Pru du 6.0101
	A0A251N1T8	164.92	19% (8/7)	2% (1/1)	4% (2/2)	3% (2/2)	3% (2/2)	9% (4/4)	Uncharacterized protein
	M5W7Y2	164.92	18% (8/7)	2% (1/1)	4% (2/2)	3% (2/2)	3% (2/2)	9% (4/4)	Uncharacterized protein
A0A5E4FK23		90.28	15% (4/3)	0% (0/0)	0% (0/0)	3% (1/1)	0% (0/0)	0% (0/0)	Predicted legumin
	A0A4Y1S2I9	84.49	6% (3/2)	0% (0/0)	0% (0/0)	2% (1/1)	0% (0/0)	0% (0/0)	RmlC-like cupins superfamily protein
	Q43608	83.65	8% (3/3)	0% (0/0)	0% (0/0)	2% (1/1)	0% (0/0)	0% (0/0)	Pru2 protein (fragment)
	E3SH29	79.71	5% (2/2)	0% (0/0)	0% (0/0)	2% (1/1)	0% (0/0)	0% (0/0)	Prunin 2 (fragment)
	M5VL85	79.06	7% (3/2)	0% (0/0)	0% (0/0)	2% (1/1)	0% (0/0)	0% (0/0)	Uncharacterized protein