ABSTRACT

Title of Document: STRUCTURAL STUDIES OF THE PEANUT ALLERGEN PROTEIN ARA H 2.

JINXI LI, Ph.D., 2010

Directed By: Dr. Catherine Fenselau, Department of Chemistry and Biochemistry

Presented in this dissertation are comprehensive studies of the structures of the peanut allergen protein Ara h 2 and the effect of food processing (roasting) on it.

A detailed elucidation of the primary structure and PTM of Ara h 2 from the raw peanuts has been described. Ara h 2 isoforms were purified and cleaved via microwave accelerated trypsin digestion. The peptide mixtures were analyzed by LC-MS/MS and targeted CID. *De novo* sequencing of the MS/MS spectra revealed the protein sequence of each Ara h 2 isoform. Several hydroxyproline sites have been discovered while disulfide bond structures have been partially determined.

Using anti-Ara h 2 antibodies, Western blotting of 1-D gels of the raw and dark roasted peanuts was carried out in order to characterize the changes of Ara h 2 between these two samples. The result indicates that Ara h 2 may present in a much heavier form in the roasted peanuts, possibly due to crosslinking and aggregation with other proteins. Subsequent LC-MS/MS studies of trypsin digestion of five gel pieces

(>100, 100-50, 50-25, 25-16 kDa) from 1-D gels of the raw and dark roasted peanuts suggests that roasting process causes the crosslinking of Ara h 2 with other proteins. This supports our results from the immunological studies.

STRUCTURAL STUDIES OF THE PEANUT ALLERGEN PROTEIN ARA H 2.

By

JINXI LI

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2010

Advisory Committee:
Professor Catherine Fenselau, Chair
Professor Philip DeShong
Professor Brian Bequette
Professor Utpal Pal
Dr. John Callahan (CFSAN/FDA)

Acknowledgements

First of all, I would like to thank my advisor, Dr. Catherine Fenselau. It has been a wonderful four years of training in her group that will benefit my whole life. I feel very grateful for her guidance in the scientific research and her role model for being a successful woman scientist.

I would like to thank the mass spectrometry group at CFSAN/FDA, especially Dr. John Callahan, Dr. Kevin Shefcheck, Dr. Mark Ross, Dr. Melinda McFarland, Dr. Jefferey DeGrasse, Dr. Peter Scholl, and Mr. Denis Andrzejewski. I appreciate all of their valuable advice, generous help and support.

I also need to thank my committee members, Dr. Philip DeShong, Dr. Brian Bequette, Dr. Utpal Pal, and Dr. John Callahan for the time and effort they spent for making my research projects going smoothly.

I also need to thank every fellow graduate student in Fenselau group, especially Karen Lohnes, Joe Cannon, Colin Wynne, and Waeowalee Choksawangkarn. I appreciate all the advice and our friendship.

Finally, I would like to acknowledge my parents, Fuxiang Li and Heping Luo, for their love and support and always being there for me.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

MALDI: matrix-assisted laser desorption ionization

ESI: Electrospray ionization

CID: Collision induced dissociation

ETD: Electron transfer dissociation

ECD: Electron capture dissociation

SDS PAGE: Sodium dodecyl sulfate polyacrylamide gel electrophoresis

2-D PAGE: Two dimensional polyacrylamide gel electrophoresis

HPLC: High performance liquid chromatography

LC-MS/MS: Liquid chromatography tandem mass spectrometry

PTM: Post translational modification

Q-TOF: Quadruple time of flight mass analyzer

LT: Linear trap mass analyzer

IEF: Isoelectric focusing

IPG: Immobilized pH gradient

MRM: Multiple reaction monitoring

PBS: Phosphate buffered saline

TBS: Tris buffered saline

DTT: Dithiothreitol

DEAE: Diethylaminoethyl

CHCA: α-Cyano-4-hydroxycinnamic acid

RNase A/B: Ribonuclease A/B

ACN: Acetonitrile

Chapter 1: Introduction

Background and Significance of the Study of Peanut Allergen Proteins

Food allergies are important health problems, among which peanut allergy is one of the most important food allergies. Unlike other allergies such as milk and eggs, peanut allergy is rarely outgrown after childhood. Unfortunately, there is no treatment available for the patients with peanut allergy, which makes avoidance of peanuts the only way to protect the patients. Due to the large consumption and wild application of processed peanuts in foods, there is a great possibility that peanut-free foods may be contaminated with traces of peanuts. This can be a serious health problem for the consumers with peanut allergy. About 3 million people in the United States have been diagnosed with peanut allergies. Exposure to peanuts in patients who are hypersensitive to them can cause skin rashes, nausea, and even life-threatening anaphylaxis[1].

The allergenic compounds in peanut are proteins, Ara h 1-9. Ara h 2 is one of the most abundant allergens whose amount is in the range of 5.9-9.3% of total protein contents. Additionally, it is recognized by the human immune system in over 90% patients[2, 3]. The peanut allergen proteins are recognized by immunoglobulin E (IgE) in human serum. Once the IgE binds to the epitopes of the allergen proteins, it will trigger the mast cell to release mediators including histamine and heparin. The mediators could result in all kinds of symptoms as mentioned above.

In the food industry, enzyme-linked immunosorbent assay (ELISA) is widely used for detection of peanut allergen proteins[4, 5]. This involves the application of the

antibody. However, the specificity of the antibody can be limited so ELISA may work well with some allergen proteins but not so well with some other proteins in peanuts. The sensitivity of ELISA depends on how strong the antibody binds to the allergen proteins. When the antibody binds weakly to the target allergens, the sensitivity may be too low to detect the trace amount of allergen proteins in food matrices.

Previous Work in the Literature that was Done Related to Ara h 2

Ara h 2 has not been studied in the protein level comprehensively. The protein sequences of Ara h 2 came from the deduction from two reported cDNA sequences. Ara h 2.01 and Ara h 2.02 (GenBank Accession Number FJ713110 and AY158467, respectively)[6]. One protein isoform sequence (corresponding to FJ713110) comprises 160 residues, including the signal peptide (21 amino acids), with glutamic acid and aspartic acid at residues 61 and 151. (In the numbering system used throughout this dissertation the signal peptide is not counted and glutamic acid and aspartic acid appear at residues 40 and 130, respectively). The second isoform is characterized by an insertion of 12 amino acids starting at residue 54 (75 with signal peptide), with residue 40 (61 with signal peptide) as a glycine and residue 142 (163 with signal peptide) as an aspartic acid. The mass added by insertion of this 12 amino acid piece is 1413 Da. Studies of the proteins by Yan et al[7] and Viquez et al [8] confirmed that there are amino acids variants at residues 40 and 142. The molecular masses of purified Ara h 2 from raw peanuts were measured by MALDI-TOF, which indicated two isoforms of mass 16670 Da and 18050 Da [6]. This is consistent with the

expression of the protein from two different cDNAs. The mass difference between the measured molecular weights of the two isoforms is 1380 Da, which is close to the 1413 Da proposed by the translated protein sequences based on the two cDNAs. However, the match is not perfect, which indicates post translational modifications in these isoforms of Ara h 2. In addition, the measured molecular masses of both isoforms do not match with the calculated masses based on the deduced protein sequences.

There are many cases in which foods are contaminated with peanut allergen proteins during food processing. Some foods may be labeled peanut free while containing peanut proteins. It is important to detect peanut allergen proteins confidently in food matrices for regulatory monitoring purposes. Besides the application of immunological methods in the detection of peanut allergen proteins, mass spectrometry based methods have become more and more important because of its high sensitivity. Previous characterization of Ara h 2 provided the identification two tryptic peptides [82-94] and [95-110], which were proposed as biomarker candidates for Ara h 2 detection[9]. The presence of the biomarker peptide suggests the presence of the corresponding protein in the sample. The quantity of the peptide can even be used as a relative measurement of the amount of protein.

A recent study of 2-D gel electrophoretic patterns of proteins from the Virginia type peanut identified fourteen spots that were linked to Ara h 2 by partial peptide mass fingerprinting[10]. All the peptides that were involved agree with the deduced protein sequences. However, the partial peptide mass fingerprinting of the Ara h 2 spots in the 2-D gel did not provide any result to differentiate those isoforms.

In a separate study in 1992 focused on potential post translational modifications (PTM), Burks and coworkers proposed that Ara h 2 is a glycoprotein, based on periodic acid-Schiff staining. Subsequent analysis reported that this protein contained 20% carbohydrate by mass, with galacturonic acid, arabinose, and xylose as the three most abundant sugars in the putative glycoprotein[11]. Their sample was purified from the same runner variety as used in this study, but it was from a different cultivar.

It is the objectives of our work to define the entire protein primary structure and to discover PTMs in order to support on-going studies on peanut related food allergens.

Application of Mass Spectrometry in Food Allergens

Ionization Techniques-MALDI

In 1985, Koichi Tanaka showed that proteins can be ionized without losing their structures by laser radiation when mixed with a metal powder in glycerol. This is known as soft laser desorption (SLD) technique. He was awarded one quarter of the Nobel Prize in Chemistry in 2002 for being the first scientist to ionize proteins by laser radiation[12]. However, SLD is not used for protein analysis any more. The concept of matrix-assisted laser desorption ionization (MALDI) was first introduced by Karas and Hillenkamp in1987[13]. The matrices are small organic molecules instead of the metal powder in SLD. The MALDI process is shown in Figure 1.1. In this technique, a pulsed UV laser beam is radiated at crystals of the mixture of chromaphoric matrice and analytes. The matrix, which has a high absorbance at the wavelength of the laser beam, can transfer the absorbed energy to the analytes and

trigger the ionization at the surface. Common matrices used for peptide and protein studies include dihydroxybenzoic acid (DHB), α-cyano-4-hydroxycinnamic acid (CHCA), and sinapinic acid (SA). Most of the ions generated by MALDI are singly charged ions. However, multiply charged ions can be produced occasionally. MALDI has several advantages[14]. Because of the fact that it produces mostly singly charged ions, the spectra are relatively easy to analyze without the need to worry about charge state distribution of the analytes. Additionally, MALDI tolerates a limited amount of sample contaminants, which include urea, detergents, and salts[15]. For example, MALDI can tolerate up to 0.1% detergents (except SDS which is 0.01%).

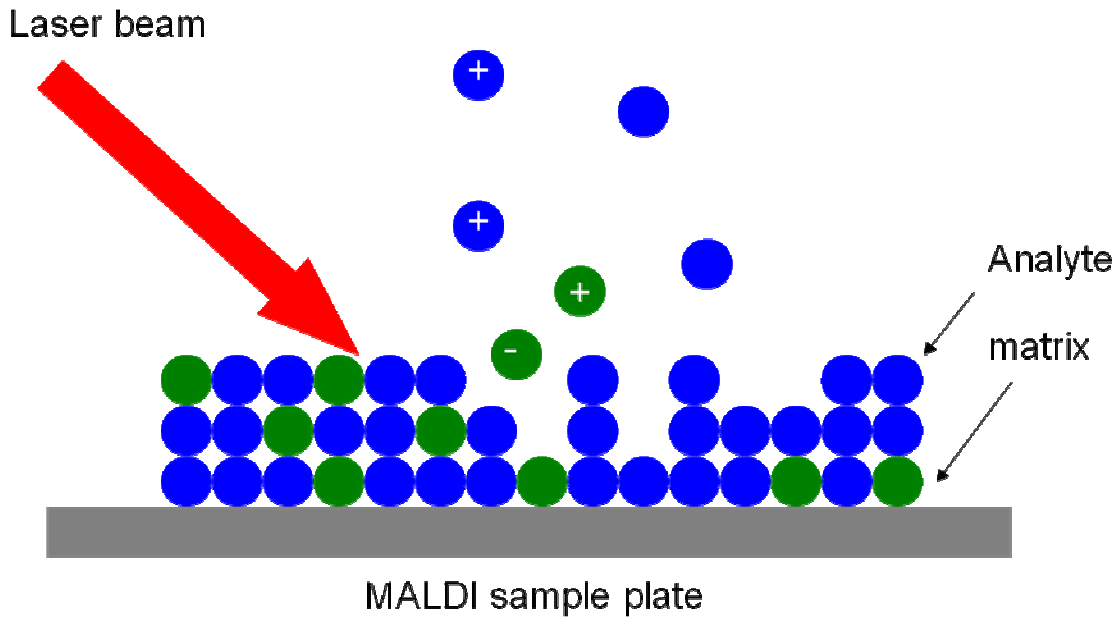


Figure 1.1: Schematic graph of MALDI.

Ionization Techniques-ESI

ESI was introduced by Yamshita and Fenn in 1984 [16] and has brought a revolution in the mass spectrometric analysis of proteins. It is a soft ionization technique because it solves the problem that some fragile polar macromolecules tend to fragment during ionization. The 2002 Nobel Prize of Chemistry was awarded to John Bennett Fenn for his contribution to the development of electrospray ionization for biological macromolecules.

In ESI (Figure 1.2), a volatile solvent that carries the analytes passes through a nozzle upon which a high voltage is applied. The solvent breaks into a fine mist of charged droplets at the tip of the nozzle. A nebulizing gas is usually applied coaxially to help with the migration and vaporization of the droplets. As these droplets travel toward the mass spectrometry inlet which is opposite in charge to the droplets, they become smaller as a result of vaporization of the solvent. Thus the surface charges become denser. The droplets break into even smaller droplets when the electrostatic repulsion of the charges overcomes the surface tension. Eventually, charged ions in gas phase are formed[17, 18]. The mass spectrometer will then measure the mass over charge (m/z) value instead of the mass directly. The m/z can be calculated via the molecular weight (MW) and the number of charges (n) that the analyte carries as shown in the equation below.

$$m/z = (MW + nH^+)/n$$

In contrast to MALDI in which singly charged ions are generated, ESI usually produces multiply charged analytes whose mass range can vary from less than 100 Da

to 1,000,000 Da. Analytes such as peptides and proteins can have a distribution of multiply charged ions because they have many potential basic amino acid side chains that may be protonated during ESI.

Samples can be directly injected to the mass spectrometer via ESI if the samples are uncomplicated and in appropriate buffer. Alternatively, the samples can be separated via liquid chromatography (LC) and then analyzed via ESI-MS. ESI can be couple with various kinds of mass spectrometers including ion trap, Q-TOF, and triple quadrupoles.

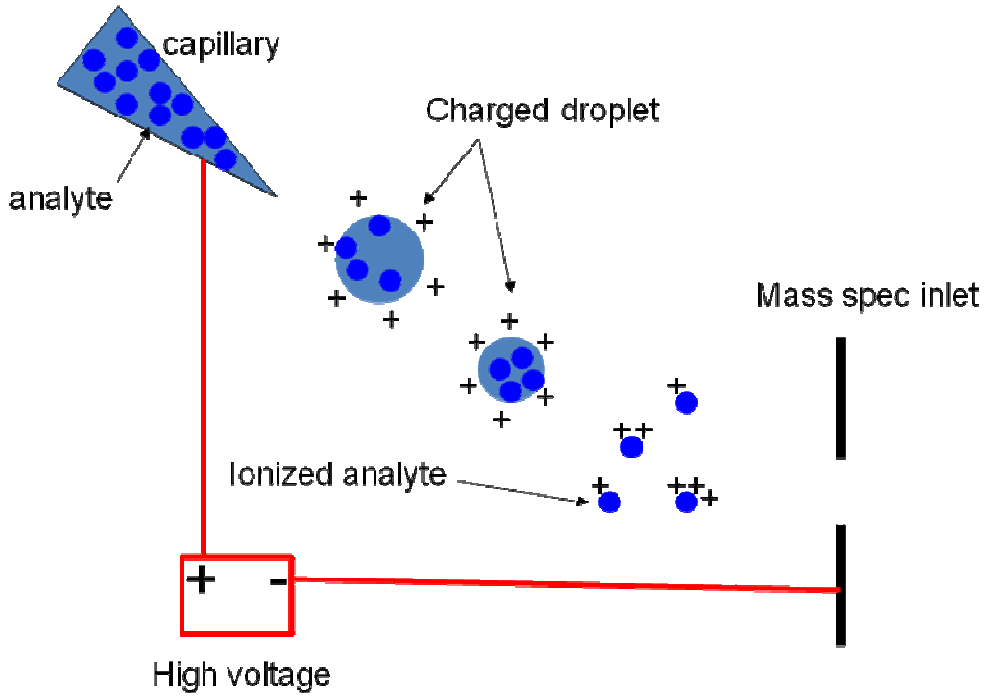


Figure 1.2: Schematic graph of ESI.

Mass Spectrometric Fragmentation Techniques -CID

For complex samples, there is a great chance that peptides of different amino acid sequences are isobaric or very close in mass. Thus, the mass measurements of the precursor ions are not enough to elucidate sequence information for peptide and protein identification confidently. Not to mention that mass measurements of the peptides alone do not provide sufficient information of post-translational modifications. Thus tandem mass spectrometry is performed in order to obtain further sequence information[19].

Collision induced dissociation (CID) of a precursor ion is performed in a collision chamber that is filled with inert collision gas, He or Ar, inside a mass spectrometer. When the precursor ion is introduced into the collision chamber, it collides with the incubation gas and breaks apart. Peptide fragmentation can happen at any place in the peptide backbone and sometimes at side chains. The fragmentation patterns and the corresponding nomenclature (a, b, c, x, y, and z ions) are shown in Figure 1.3[20, 21]. In order for an ion to be detected by mass spectrometry, the ion must carry at least one charge. For example, if the fragmentation happens at the peptide bond, and the charge is retained on the C-terminus of the peptide, the ion is called y ion. Otherwise, it is called b ion. Another kind of fragment ions that is not shown in Figure 1.3 is internal fragment which is produced by cleaving twice at the peptide backbone. All the ions in Figure 1.3 including internal fragment ions can be produced via high energy CID. However, only b and y ions are common in low energy CID. The MS/MS spectra of peptides obtained via CID can be interpreted manually or

searched against a database on a computer. This will be explained in another section in more detail.

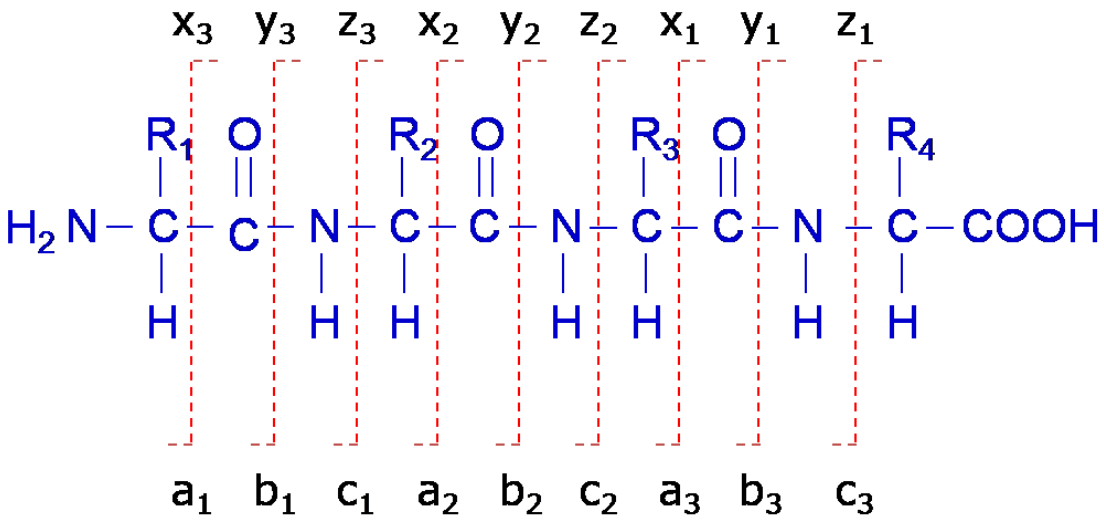


Figure 1.3: Diagram of fragmentations at peptide backbones and the corresponding fragment ions.

Mass Spectrometric Fragmentation Techniques-ETD

Electron transfer dissociation (ETD) generates mostly c, z ions via ion/ion interaction. It is a novel fragmentation method that has been recognized to be very useful in the analysis of highly charged ions (>+3), larger peptides, and even direct fragmentation of whole proteins in top-down analysis [22-26]. ETD is usually coupled with an ion trap mass spectrometer. Positively charged precursor ions of interest are injected into the ion trap followed by the injections of negatively charged radical anions. The anions are usually fluoranthene that are generated in a chemical ionization source as shown below.

$$C_{16}H_{10} + e^- \longrightarrow C_{16}H_{10}\,]^{\cdot -}$$

Then electron transfer process from the fluoranthene radical anion to positively charged sample ions occurs via an ion/ion reaction inside the trap. The ion radical products will undergo a quite different fragmentation pathway from CID (Figure 1.4). Electron capture dissociation (ECD) is a similar fragmentation technique to ETD. Instead of using anionic ions to interact with the precursors, electrons were used directly [22, 27]. Compared to collision induced decomposition (CID), which has certain preferred fragmentation channels, ETD is well known for its random cleavage along the peptide backbone while preserving modifications[28, 29].

ETD/ECD also gives an advantage over collision induced dissociation (CID) for the localization of post translational modification, because it preserves the PTM while fragmenting the peptide backbone [23, 30]. People have demonstrated the application of ETD in analysis of phosphopeptides and glycopeptides. The phosphate group and glycan were all well preserved [31]. The electron transfer product ions produced by the

ion/ion reaction is often stable, and so ETD is usually assisted by supplemental activation which is basically low energy CID [26, 32]. Theoretically, larger peptides carry more charges. Large peptides which are more suitable for ETD analysis have been generated via Lys-C, Asp-N protease, and some chemical cleavage methods including microwave-accelerated acid digestion (discussed in the section of microwave-accelerated enzymatic and chemical digestion of proteins)[33].

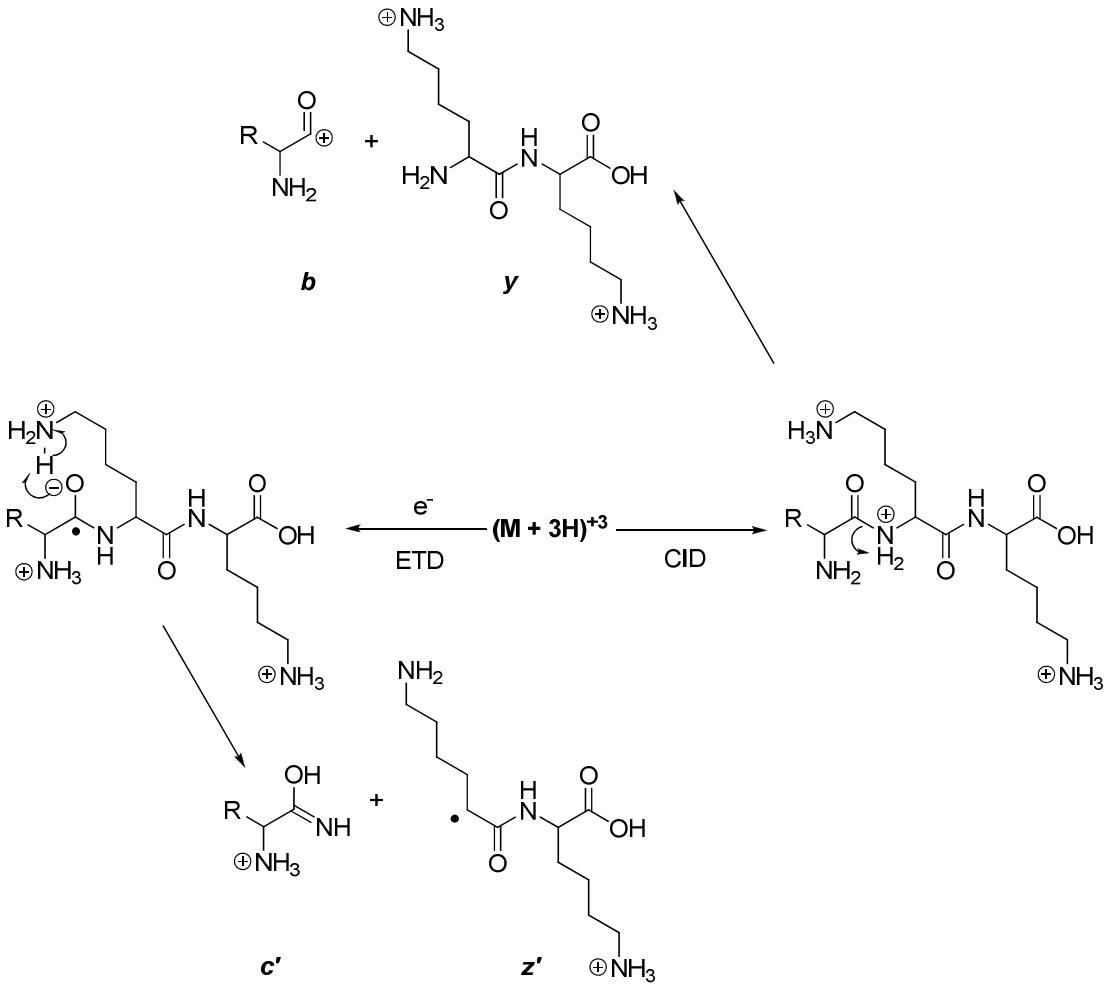


Figure 1.4: Peptide fragmentation scheme for CID and ETD. Adapted from Chi A. et al. *PNAS* 104 (2007) 2193-2198.

LC-MS/MS

LC-MS/MS is a powerful analytical tool that combines liquid chromatography with tandem mass spectrometric analysis. It involves chromatographic separation, ionization and mass measurement of the analytes. Due to the sample complexity, some kind of separation and purification before analysis is usually required for better sensitivity and selectivity. Some LC separation techniques are discussed in the section of pre-fractionation of samples.

The most important part of a mass spectrometer is the mass analyzer. The modern hybrid mass spectrometers such as LTQ-Orbitrap, Q-TOF, and TOF/TOF can have more than one mass analyzer. Tandem mass spectrometry involves two stages of mass analysis known as MS/MS. Figure 1.5 is a typical scheme of MS/MS. The ionized analytes are introduced to the first mass analyzer for separation based on their mass over charge (m/z) values. The precursor of interest is isolated and introduced to a collision chamber that is usually filled with inert gas such as He or Ar for fragmentation[34]. The generated fragment ions that contain detailed sequence information of the precursor ion are then separated via the second mass analyzer.

LC-MS/MS supports for not only high throughput peptide discovery and identifications, but also labeled or label-free quantitative analysis of proteins. It is a powerful and automated tool that generates data compatible with both *de novo* sequencing and database searching.

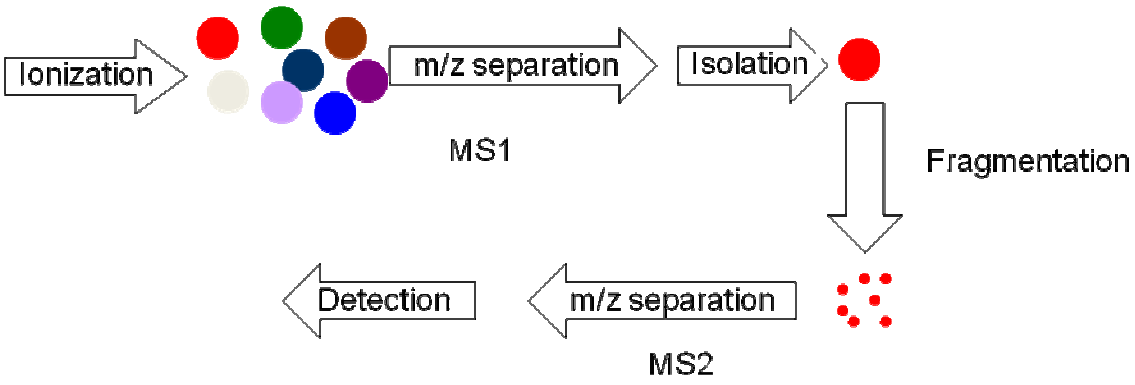


Figure 1.5: Schematic graph of tandem mass spectrometry (MS/MS).

Peptide and Protein Identification via *De novo* Sequencing

De novo sequencing of peptides is usually performed via manual interpretation of MS/MS spectra without knowing the peptide sequence initially. Most of the amino acids have their own unique masses except leucine and isoleucine. Protein sequencing via low energy CID and tandem mass spectrometry was proposed by Donald Hunt in 1986[35]. A detailed description of how to interprete low energy CID spectra of tryptic peptides has been developed through the years.

Low energy CID generates mostly b and y ions. It is common for the b and y ions to lose water (-18 Da) or ammonia (-17 Da). Peptides do not fragment equally at all the positions. This is why the peaks in MS/MS spectra are not equally abundant. For example, N-terminus of a proline is a favored fragmentation site in CID. Thus, the corresponding y ion appears to be abundant in the MS/MS spectrum. *De novo* sequencing is usually performed with MS/MS spectra of tryptic peptides because trypsin digestion tends to generate peptides of the right size for manual interpretation. The spectra will be difficult to analyze manually if the peptides are too big. When the peptides get bigger with more charges, their fragment ions are likely to carry more charges accordingly. Thus, high resolution mass spectrometers will be needed to resolve their charge states. In general, doubly charged tryptic peptides often provide nice fragmentation for *de novo* sequencing.

The tryptic peptide sequence can be deduced starting from the C-terminus since it is usually arginine or lysine. Their y_1 ions are at 175 Da or 147 Da, respectively. However, in an ion trap mass spectrometer, the low mass range of a CID spectrum is cut off at one third of the precursor's m/z. So y_1, which is very likely to fall in the low

mass region, can not be observed. This is not a problem with a Q-TOF mass spectrometer though. An alternative way is to look for the corresponding b ion. The mass of the corresponding b ion can be calculated as shown below, in which n is the number of amino acids in a peptide and m can be any number smaller than n.

b_{n-m} = peptide mass + 2 - y_m

By assuming that y_1 ion is arginine or lysine, we can look for possible y_2 candidate by subtracting the residue mass of lysine or arginine from a higher fragment mass in the MS/MS spectrum to see if the mass difference matches with any residue mass of an amino acid. In the meanwhile, labeling the corresponding b ion is always a good idea. We can keep doing this and extend the peptide sequence from the C-terminus to the N-terminus. However, it is very likely that one or more fragment ions are missing for peptide sequencing. In this case, we can use the sum of the residue masses of any two amino acids to extend the sequence instead. There might be more than one possibility. For example, LS and TV both weigh 200.12 Da. For those peaks that are not assigned in the end, we can consider the doubly charged ions (high resolution MS/MS spectrum will allow determination of the charge state, but low resolution MS/MS spectrum will not), loss of water and ammonia from b and y ions, and internal fragments that are generated by breaking a peptide twice.

It is sometime necessary to make assumptions in *de novo* sequencing. However, we do have to be careful and consider all the possible combinations. It is important to verify the deduced peptide to see if it makes sense. Of course its mass has to match with the measured mass. It is also unlikely to have an abundant y ion that is generated by cleaving the C-terminus of a proline. Manual analysis can be time-consuming.

However, automated *de novo* sequencing programs and software have become available[36]. *De novo* sequencing is still very important, while database searching is becoming widely used in proteomics because there are times we need to go back and confirm the identification of peptides or proteins of interest provided by database searching. Database search can only identify peptides in the database. When there are unknown proteins or mutated proteins that are not present in the database, *de novo* sequencing may be the only choice for analyzing MS/MS data[37, 38].

Peptide and Protein Identification via Database Searching

Database searching is an alternative way of analyzing mass spectrometric data. There are many search engines available for protein and post translational modification (PTM) identification (Mascot, X!tandem, Sequest). Mascot (www.matrixscience.com) will be discussed since it was the only search engine that was used in this study. Mascot is a program that does statistical evaluation of matches between the theoretically generated peptides and fragments and the real data. That means a peptide has to be in the database in order to be identified. There are three kinds of database searching provided by Mascot. They are peptide mass fingerprinting, sequence query, and MS/MS ion search. Peptide mass finger printing is widely used for data acquired on a MALDI-TOF mass spectrometer[39]. Peptide mixtures from an enzymatic digestion of a protein are like its fingerprints and can lead to highly confident identification of the protein. In order to perform peptide mass fingerprinting, m/z values in a mass spectrum will be submitted online with an appropriate setting of parameters (type of enzymes, number of missed cleavages in

the digestion, database to use, etc). Mascot will then generate theoretical peptides of each protein in the database based on the indicated parameters. It will match the masses of the theoretical peptides with the data. The protein with the most matched peptides is the most confident identification. Sequence query is not as commonly used as the other two. A sequence tag is submitted. The search result provides a list of all the proteins in a database that contain this peptide sequence. MS/MS ion search is probably the most used search algorithm in Mascot[40]. The data is usually acquired via LC-MS/MS and can be large files that contains hundreds of thousands MS/MS spectra. For each MS/MS spectrum, there is a precursor mass and its corresponding fragment masses. Mascot will generate theoretical peptides from proteins in the database and match those with the precursor mass. Those matched theoretical peptides will further be used to generate masses of their possible fragment masses. These will be matched with the measured fragment masses in the data file. Each assignment for a MS/MS spectrum will be given an expectation value (E value) as an evaluation of how good the match is between the spectrum and theoretical fragments. E value defines the probability of a peptide hit being random. The smaller the E value is, the less likely the peptide identification is false. In our studies, an E value cut off at 0.05 is used. Finally, all the identified peptides will be considered to generate protein identifications. Each identified protein will be given a score as well.

There are many advantages of database searching. It is much faster than manual analysis and can handle a very large data set with highly confident protein identifications. However, not all protein identifications will be true. It is still important to verify the protein assignments in some cases.

Discovery of Post Translational Modification

Post translational modifications (PTMs) are made to proteins after their translations by adding biofunctional groups to amino acids. PTM plays a very important role in the structure and biofunction of proteins. For instance, glycans at cell membranes are vital for cellular recognition and trafficking [41]; hydroxyprolines in collagens are known to stabilize the collagen's triple helix structure[42].

PTM has been widely studied using mass spectrometry related methods[43]. For PTM discovery, proteins are usually digested to peptides, and then the peptides are mapped via LC-MS/MS. For example as shown in Figure 1.6, oxidation of methionine will add an additional 16 Da to the amino acid. As a result, b_4 and y_5 ions will be shifted by 16 Da while b_3 and y_4 ions remain the same. In this way, potential PTM in proteins can be revealed confidently. Besides manual analysis, PTM discovery can be performed via Mascot database searching for the types of modifications that might present in the peptides. However, both analyses should be based on understanding of the sample from the previous work to see if the identified modification makes sense.

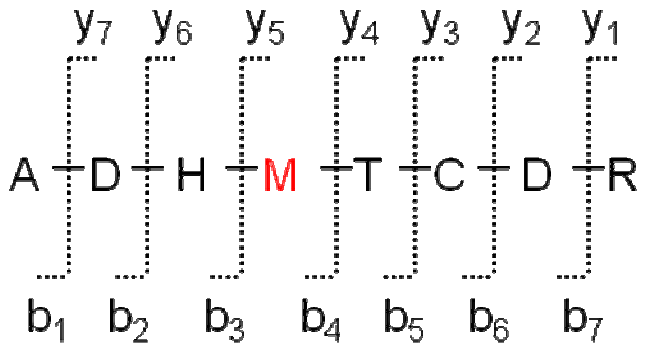


Figure 1.6: Oxidation of methionine can be located via the mass shifts at fragmentation ions y_5 and b_4.

SDS PAGE

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE) is one of the most used biochemical tools. It was introduced in the 1960s and has been further optimized by introducing SDS and Laemmli buffer [44-46]. SDS is an anionic detergent that is added to the sample buffer to quantitatively bind to proteins via its hydrophobic dodecyl unit. Each SDS molecule contributes two negative charges, which makes protein side chain charges negligible. As a result, the charges that SDS binding proteins carry are usually proportional to their sizes.

A high percentage of acrylamide results in highly cross-linked polyacrylamide gels with small pore sizes. It is important to choose a gel with the appropriate percentage depending on the mass range of your protein mixtures. In gel electrophoresis, an electric voltage is applied so the negatively charged proteins will move to the positive side (cathode) in the gel. The polyacrylamide gel creates a cross-linked environment that provides resistance to protein migration. The small proteins are able to get through faster than the big proteins. Eventually, the small proteins will be in the bottom of the gel while big ones will be on the top of the gel. The separation of proteins based on their molecular weights is achieved.

2-D Gel Electrophoresis

2-D gel electrophoresis has been introduced by O'Farrell in 1975 to separate cellular proteins[47]. It allows isoelectric focusing (IEF) of proteins in one dimension followed by regular SDS PAGE in the other dimension (Figure 1.7). In the first dimension of 2-D PAGE, the protein mixtures are separated based on their isoelectric

points via IEF. This is usually carried out by loading proteins onto an immobilized pH gradient (IPG) strip on which the pH gradient is generated by carrier ampholyte. Upon application of voltages on the ends of the IPG strip, the proteins in the strip migrate toward their pI until they are focused at their pI and become neutral. This is followed by a second dimension SDS PAGE in which the strip will be placed on top of a gel. The focused proteins will migrate into the gel from the strip and be separated based on their molecular weights as described above.

When the gel fractionation is complete, it will be stained with coomassie blue or silver stain depending on the concentration of the proteins and the purpose of the experiments[48]. Silver stain is a very sensitive staining technique that can detect proteins in 1-10 ng scale. Coomassie blue is commonly used and its detection limit is 50-100 ng. These two staining methods are both compatible with mass spectrometry given appropriate destaining steps.

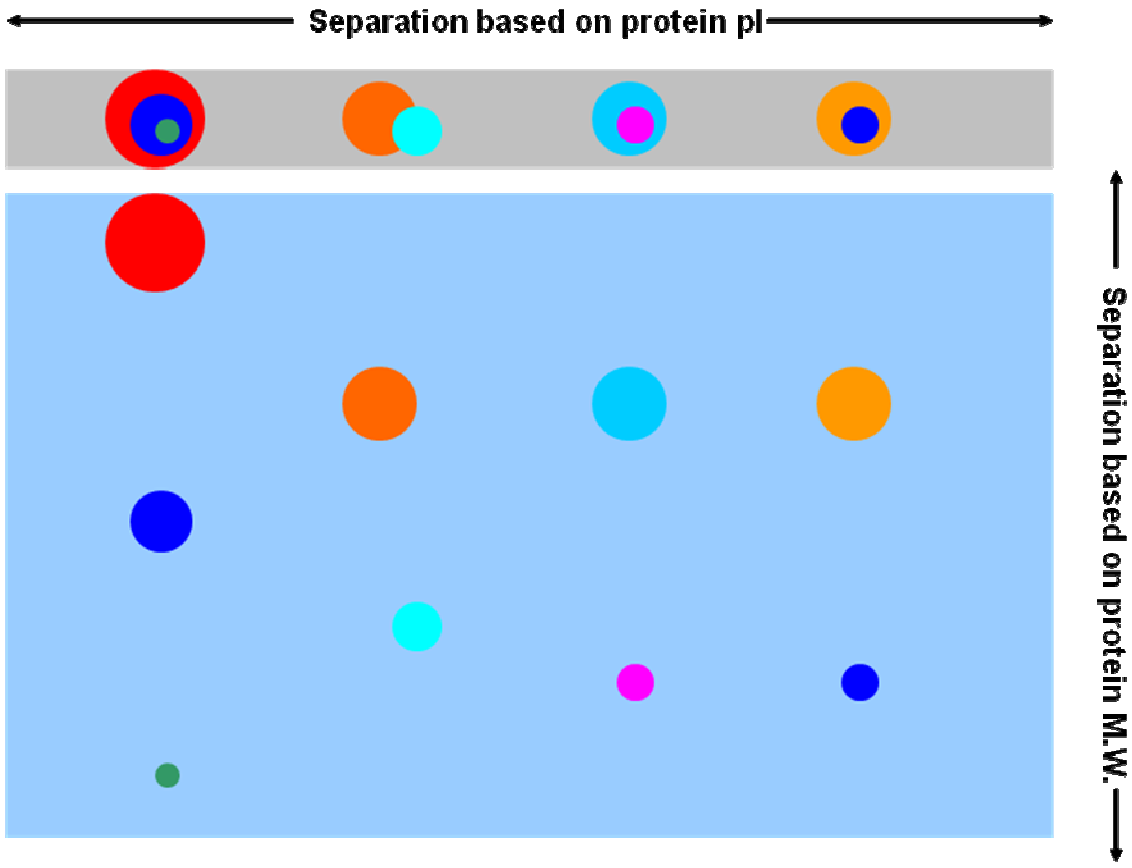


Figure 1.7: 2-D gel electrophoresis theory.

Western Blotting

Western blotting is an analytical tool named after W. Neal Burnette that uses antibodies to detect the target proteins in mixtures[49]. As shown in Figure 1.8, protein mixtures are usually separated on a gel. And then the proteins are transferred from the gel to a membrane (nitrocellulose or PVDF membrane) followed by the detection via antibodies[50]. The antibody that is specific to the antigen is called the primary antibody. A secondary antibody that is specific to the primary antibody is also needed. It is usually covalently linked with a reporter enzyme that can produce light or color when exposed to its corresponding substrates[51]. Thus the antigen can be detected on the membrane.

The most important factor is the quality of the antibody. Not all antibodies are suitable for Western blotting. The antibodies are produced by immunizing an animal (rabbit, mouse, donkey, etc.) with the target antigens. The antigens are usually injected into or fed to the animal with appropriate doses. The immunosystem of the animal will respond and produce immunoglobulins. The blood serum of the immunized animal which contains the immunoglobulins will be collected. The antisera can be used directly. Further purification can be done if highly pure antibodies are desired[52]. The concentration of antibodies used in an individual experiment can vary greatly depending on the binding efficiency of the antibody to the target protein. There are two kinds of antibodies, polyclonal and monoclonal. Polyclonal antibodies are generated by more than one kind of immune cells. They are a combination of immunoglobulin proteins that have multiple binding sites to the

antigen. Monoclonal antibodies are produced by the same type of immune cells. They only bind to a certain site (epitope) of antigens.

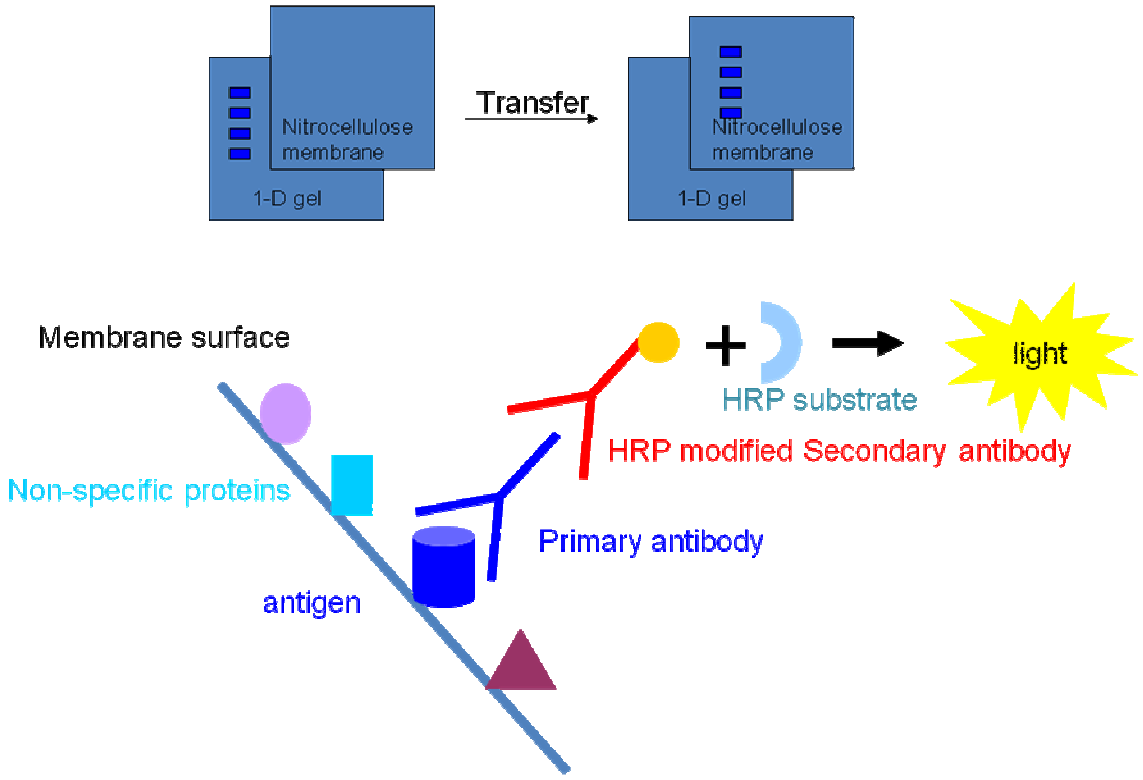


Figure 1.8: (top) proteins are transferred from a gel to a nitrocellulose membrane; (bottom) the membrane is probed by antibodies in Western blotting.

Immunoblotting and Mass Spectrometry

Mass spectrometry related methods have become more and more important in the analysis of food proteins and peptides. Their diverse applications include characterization, detection, identification, and quantitation of food proteins, which significantly help us to understand the protein structures, post translational modifications, and the biological functions of food proteins[53, 54].

For identification of food proteins, especially food allergen proteins, one common proteomics strategy is coupling immunoblotting with mass spectrometry. In this technique, the proteins are usually separated on a 1-D or 2-D gel followed by subsequent immunoblotting using patient sera to detect target proteins. The detected protein spots on the gel will be cut, digested by trypsin, and analyzed using either MALDI-TOF or LC-MS/MS followed by protein identification via database searching[55]. It has been reported that peptide biomarkers can be used to quantitatively detect peanut allergen proteins in food[9]. In order to do that, a high throughput LC-MS/MS analysis was performed for peptide identification. Based on this result, good peptide candidates (peptides in high abundance) were chosen as biomarkers of the corresponding proteins for subsequent multiple reaction monitoring (MRM) analysis on a triple-quadruple mass spectrometer. MRM monitors the precursor ion and its fragment ions consecutively in order to confirm the presence of the targeted peptide. Only in cases that both the correct masses of the precursor ion and its fragment ions are present are regarded as a positive identification of the peptide. With the high sensitivity of this methodology, the detection limit of proteins in food matrices can be

as low as femtomoles. Mass spectrometry based methods play a vital role in ensuring food safety and enforcing appropriate labels on food.

Sample Preparation

Protein Extraction and Purification

Raw peanut kernels are usually processed into various kinds of foods. For this purpose, raw peanut kernels are usually roasted to reduce the moisture and develop a pleasant flavor[56]. Poms has proved that increased roasting temperature in the processing of peanuts can decease the protein solubility[57]. The yield of soluble proteins that are extracted from the roasted peanuts was reported to drop by 75-80% compared to the yield of proteins extracted from the raw peanuts[56]. Various buffers have been used including phosphate buffered saline (PBS), Tris buffered saline (TBS), various extraction buffer from commercial ELISA kits, and urea. All of the tested buffers showed the same reduction of soluble protein yield from the roasted peanuts. The protein extractions that were performed at pH 8-11 seemed to be the most efficient[57].

The allergen proteins with low solubility were reported to have a higher binding capability with human IgE than the relatively more soluble peanut allergen proteins have. It is essential to make sure the more allergenic proteins with low solubility could be extracted and analyzed[58].

Peanuts are rich in nutrients. They contain many essential nutrients and phytonutrients[59]. During protein solubilization and extraction out of the peanut flours, other non-protein nutrients are very likely to be extracted at the same time. Their

presence in the crude extraction increases the complexity of the sample. Thus, further protein purification is essential for the subsequent analysis. The purification methods can either be a straightforward one step precipitation or multiple steps. The latter method usually contains some kind of chromatography.

Anion Exchange Chromatography

Anion exchange is a kind of ion exchange chromatography. The separation of analytes is based on the electrostatic attraction between the negatively charged solute ions and positively charged sites bound to the stationary phase[60]. The resin with covalently linked cations will attract mobile anions and affect their retention time. Cellulose and dextran ion exchangers (gel type) that possess large pore sizes and low charge densities are well suited for ion exchange of proteins. These kinds of columns are made by crosslinking dextrans and attaching functional groups to the hydroxyl groups of the sugars. Acidic proteins with low pI are suitable for anion exchange chromatography because they are easily negatively charged while basic proteins with high pI are best handled by cation exchange chromatography. The pI of Ara h 2 is predicted to be between 5 and 6. So anion exchange chromatography is suitable for the separation of Ara h 2 isoforms. The DEAE (diethylaminoethyl) ion exchange column that was used in the separation of Ara h 2 isoforms is made by attaching DEAE functional groups to the resin. Each functional group has its own cation exchanger. DEAE is usually compatible with intermediate base[61]. The retention time of proteins are basically determined by the protein pI, pH of mobile phase, salt concentration, and the partition of proteins between mobile and stationary phases.

Reverse Phase Chromatography

Reverse phase chromatography is contrast to normal phase chromatography. Its stationary phase is nonpolar and the mobile phase is polar. The analytes are separated based on the hydrophobic interactions between the solutes and the stationary phase[62]. Theoretically, any inert nonpolar materials can be used as packing materials. Most commonly used reverse phase packing for peptide separation is silica modified with octadecyl carbon chains. It is usually referred as C18 where the number indicates the length of the alkyl chain. The longer the chain is, the longer the analytes are retained. The strength of the mobile phase plays a vital role in the separation. It is often a mixture of aqueous and organic solvents. Common combinations include water and acetonitrile (ACN), water and methanol. The separation can be performed isocratically in which the composition of the aqueous and organic solvents remains the same through the run, or it can be performed using a gradient in which the composition of aqueous and organic solvents changes. The polarity of the mobile phase decreases upon increase of the organic solvents. The polar solutes elute first while the nonpolar compounds are retained by the column and elute later by increasing the composition of the organic solvent in the mobile phase. An ion-pairing reagent, such as trifluoroacetic acid (TFA), is usually added to the mobile phase to form ion pairs with the charged peptides in order to minimize non-hydrophobic reaction and increase the separation efficiency. However, TFA causes signal suppression in ESI-MS and is not compatible with this type of analysis[63]. Further cleanup is required, or a different acid (acetic acid and formic acid) may be used for online LC-MS/MS analysis.

Microwave Accelerated Enzymatic and Chemical Digestion of Proteins

Microwave radiation was used first in organic synthesis [64], and recently the application was expanded to peptide hydrolysis and enzymatic proteolysis [65]. It has been proven that microwave radiation itself does not cause protein degradation and backbone breakdown without the presence of enzymes [66]. Microwave-tryptic digestion was applied to highly folded bovine ubiquitin and bovine cytochrome c that is relatively resistant to enzymatic digestion [66]. Seventy percent of the protein could be digested in minutes under optimized condition.

Besides the application of microwave oven in enzymatic digestion of proteins, success has also been achieved by using microwave oven in chemical cleavage of proteins via a mechanism that has been considered previously[67, 68]. It has been demonstrated that chemical cleavage with dilute formic acid takes place specifically at both sides of aspartic acid residues using thermal incubation above 104°C when the pH $\leq$ 2.1 [68]. Fenselau group enhanced the application of acid digestion by using a microwave oven with controlled temperature and power, and showed that the reaction can be realized in seconds or minutes. This technique has proved very effective in the analysis of spore proteins [69]. Acetic acid was used instead of dilute formic acid for the first time in acid digestion of a model virus [70] and yeast ribosome proteomes [71]. The effect of acid digestion on post translational modifications (PTM) has also been discussed [71]. Methionine and cysteine were not oxidized, however detectable hydrolysis of phosphate group were observed during the process. It has also been observed that acid digestion can generate a wide mass range of peptides including some large ones which weigh over 5000 Da [67].

Because Ara h 2 is resistant to tryspin, microwave radiation could increase the efficiency of tryspin digestion. Alternatively, microwave assisted acid digestion may be suitable for the cleavage of Ara h 2.

Chapter 2: Evaluation of microwave-accelerated acid digestion of N-linked and O-linked glycoproteins

Introduction

Ara h 2 has been reported to be a potential glycoprotein [11] even though neither the location nor the structure of the glycans has been studied. Because of the fact that Ara h 2 is difficult to be cleaved via a regular trypsin digestion [72], we have evaluated the microwave-accelerated acid digestion as an alternative method to cleave glycoproteins. The objective is to figure out whether the acid treatment affects the glycan linkages or not.

We have examined the cleavage of ribonuclease A and the related N-linked glycoprotein ribonuclease B, and the O-linked glycoprotein α-crystallin A chain, using MALDI-TOF and LC-ESI-MS methods to identify the peptide products.

Ribonuclease A (RNase A) and ribonuclease B (RNase B) were used to validate the method for analysis of N-linked glycoproteins and proteins with internal disulfide bonds. RNase A is a non-glycosylated protein, which serves as a control for RNase B, as they have the same primary structure. RNase B carries an N-linked glycan at Asp 34, whose structure has been characterized as a high-mannose type with a N-acetylglycosamine core (GlcNAc) and 4 to 9 mannose residues (Man) attached to the core [73, 74]. Additionally, both of these relatively small proteins have four internal disulfide bonds, which make their structures stable and relatively difficult to digest.

It was of primary interest to determine if carbohydrate-protein linkages would be affected by the microwave-accelerated acid digestion and if the oligosaccharide

side chain would be cleaved. Carbohydrate heterogeneity in the acid cleavage products of RNase B was characterized by LC-ESI-MS mass spectrometry and compared to that of the products of a parallel tryptic digestion. α-crystallin A chain carries a single O-linked GlcNAc at Ser 162 [75], which allowed evaluation of the stability of O-linkages during acid catalyzed proteolysis

Materials and Experiments

Materials

HPLC gradients (ACN, water, formic acid) were purchased from Burdick & Jackson (Morristown, NJ), glacial acetic acid was purchased from Fisher (Fair lawn, NJ), α-cyano-hydroxycinnamic acid (CHCA) and dithiothreitol (DTT) were purchased from Sigma (St. Louis, MO), protein calibration standard kit, ribonuclease A, and ribonuclease B were purchased from Sigma. Alpha-crystallin A chain was purchased from Streegen Bioreagents (Ann Arbor, MI). Trypsin was purchased from Promega (Madison, WI). All chemicals and proteins were used without further purification.

Trypsin Digestion and Microwave-Assisted Acetic Acid Digestion

A Discover Benchmate microwave system (CEM Corp, Matthews, NC) was utilized to perform the protein digestions which allowed control of the temperature, pressure, power, and time. RNase A was dissolved in Milli-Q water to make a 2 mg/mL solution. Following the protocol that was well established before[67] , 2.5 µL of this protein solution was mixed with 6.25 µL of acetic acid and 41.25 µL of Milli-Q

water in a 300 µL micro-glass vial and exposed to microwave irradiation in the open vessel mode. The final concentration of acetic acid was 12.5%. The incubation time was 5 minute with a fixed microwave irradiation power at 300W and a maximum temperature of 140°C. The same procedure was used for RNase B.

Another experiment of RNase A and B was performed with addition of a reducing reagent into the sample. Two and a half microliter of this protein solution was mixed with 6.25 µL of acetic acid, 25 µL of 10 mM DTT (final concentration 5 mM), and 16.25 µL of Milli-Q water. The final concentration of acetic acid remains 12.5% as the previous experiment. The peptide mixture of RNase A and B were analyzed the same way.

α-Crystallin A chain came in phosphate buffer solution at 1.2mg/mL (0.15M NaCl, 0.05M phosphate, pH 7.2). Two and one tenth microliters of this protein solution was mixed 6.25µL acetic acid and 25µL of 10mM DTT, and water was added to reach a final volume of 50 µL. The final concentration of acetic acid was 12.5%. The reaction was performed the same way as described RNase B.

Tryptic digestion of RNase B and α-crystalline A chain was performed. The glycans are preserved in trypsin digestion so the subsequent glycopeptides could be used as a control to evaluate the effect of acid digestion on the glycan linkages and sidechains.

Seventy nine milligrams of ammonium bicarbonate was mixed with 10 mL of Milli-Q water to reach a final concentration of 100 mM. Trypsin solution was made by adding 1.6 mL ammonium bicarbonate solution to a vial containing 20 µg sequence grade immobilized trypsin. Ribonuclease B solution (2.5 µL) was incubated

with 30 µL 100 mM DTT for 3 hour at 57°C and allowed to cool to room temperature. Fifty microliter of trypsin solution was added to the sample, followed by incubation overnight at 37°C. For the α-crystallin A chain, 5 µL protein was mixed with 45 µL 10 mM DTT, and processed as described above. The digestions were quenched by adding acetic acids to reach a final concentration of 0.5%.

The peptide products of the crude mixture from acid digestion were analyzed directly by MALDI-TOF-MS. A Kratos Axima-CFR Plus MALDI-TOF (Shimadzu Biotech, Columbia MD) equipped with a 337 nm N2 laser was used for sample analysis. Under linear mode, the instrument was externally calibrated with angiotensin II ($[M+H]^+$=1046.5), P14R ($[M+H]^+$=1533.9), ACTH fragment 18-39 ($[M+H]^+$=2465.2), insulin oxidized B chain ($[M+H]^+$=3494.7), and insulin ($[M+H]^+$=5730.6). The laser power that was used for calibration was 85 arbitrary units, the same laser power was applied during the spectrum acquisition for the sample analysis. The matrix solution was prepared by dissolving 10 mg CHCA in 1 mL water containing 70% acetonitrile and 0.1%TFA. The sample was applied to the MALDI plate via a sandwich method. A layer of 0.5 µL matrix solution was pipetted onto the plate first, followed by 0.5 µL sample solution and another layer of 0.5 µL matrix solution. Each layer was allowed to air dry before the next layer was applied. The spectra were obtained in the linear positive ion mode, with a 337 nm N_2 laser, by collecting and summing 200 profiles from 1500-5500 Da.

HPLC-ESI-MS Analysis

The peptide mixtures of RNase B from both the tryptic digestion and the acetic acid digestion were each separated on a nanoACQUITY UPLC System (Waters, Milford, MA). Ten µL of sample solution was introduced onto a reverse phase C-18 column (1.0mm i.d.×100 mm) with 1.7 µm bridged ethyl hybrid (BEH) particle packing (Waters, Milford, MA). The peptides were eluted via a binary solvent system (solvent A: 0.1% TFA and 1% water in acetonitrile; solvent B: 0.1% TFA and 1% acetonitrile in water) using a linear gradient from 0% B to 40% B in 40 min. The flow rate was set at 50 µL/min. The HPLC system was directly coupled to the ESI-MS. All mass spectra were acquired between 400 Da-2000 Da on an LTQ XL ion trap (Thermo Electron, San Jose, CA). The electrospray voltage was 4 kV, and the ESI nebulizing gas was nitrogen.

The peptide mixtures obtained from α-crystallin A chain by acetic acid digestion was separated and analyzed using the same system. A linear gradient of 60 min from 0%-50% B was applied.

Mascot Database Search

A Mascot search (Matrix Science Ltd., London UK) was carried out using peptide mass fingerprint to analyze the MS spectra from the acetic acid digestion. All the m/z values in the spectrum in Figure 2.1 were entered except for the doubly charged ions. The search was performed against SwissProt database (all species) using “formic acid” as the enzyme with a peptide mass tolerance of 4 Dalton (Formic acid is known to cleave at the same position, aspartate, as the acetic acid). Four

missed cleavages were allowed. The average mass of $[M+H]^+$ was selected, while no modification was entered.

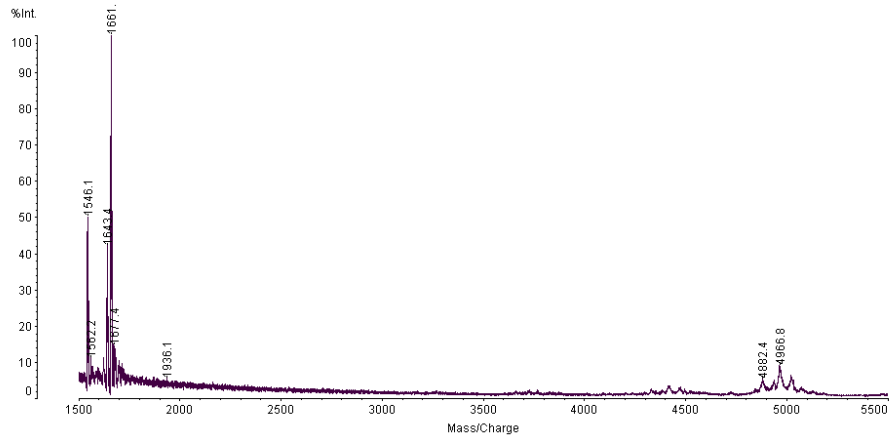
Results and Discussion

Ribonuclease A and B contain four disulfide bonds. A good digestion could not be achieved without reducing the disulfide bonds. This is demonstrated for RNase A in Figure 2.1(a). It shows that only two N-terminal peptides, [1–13] and [1–14], are released from acid digestion of the native protein. The same reaction with the addition of DTT suggests efficient reduction of disulfide bonds and cleavage of the protein backbone concurrently in 5 min, as illustrated in Figure 2.1(b) and (c). No further cleanup was needed before the MS analysis. We proposed that the inefficient digestion might due to two reasons. First, disulfide bonds might keep the protein tightly folded, so the acid can not access some parts of the protein. Second, the proteins are cleaved by acid, but the disulfide bonds are holding the peptide pieces together. In either case, a complete digestion could not be realized without the disulfide bonds being cleaved.

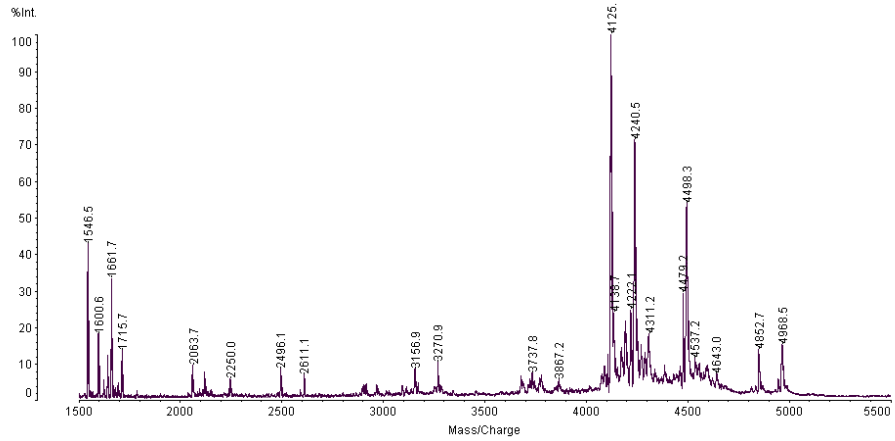
As described before, acid treatment cleaves the polyamide chain on both sides of aspartic acid. Thus, each peptide can lose or retain its aspartic acid at the terminus. It has been reported that peptides with both aspartic acids retained are very rare, and peptides with one or no aspartate are more common in acid digestion[67]. These peptide pairs (with one aspartate and no aspartate) apart by 115 Da (residue mass of aspartate) and can be easily located in the spectra.

Shown in Figure 2.1 (b) and (c) are the MALDI spectra of peptides that were generated from reduced ribonuclease A and B, respectively. Pairs of peaks that are separated by 115 Da can be observed in the spectra. Peptides that were identified provide 100% sequence coverage of RNase A, as listed in Table 2.1. For RNase B, 80% sequence coverage was achieved (Table 2.2). The missing sequence was associated with the glycopeptide [15-38] in which carbohydrate was attached to Asn34. This is due to the poor ionization efficiency of glycopeptides in CHCA matrix in MALDI. In order to quantitatively characterize this glycopeptide from acid and trypsin digestion, LC separation coupled with electrospray ionization mass spectrometry (LC-ESI-MS) was utilized.

(a)


%Int.
1661.
1546.1
1643.4
1677.4
1936.1
4882.4
4966.8
Mass/Charge

(b)


%Int.
4125.
4240.5
4498.3
1546.5
1661.7
1600.6
1715.7
2063.7
2250.0
2496.1
2611.1
3156.9
3270.9
3737.8
3867.2
4138.7
4222.1
4311.2
4479.2
4537.2
4643.0
4852.7
4968.5
Mass/Charge

(c)

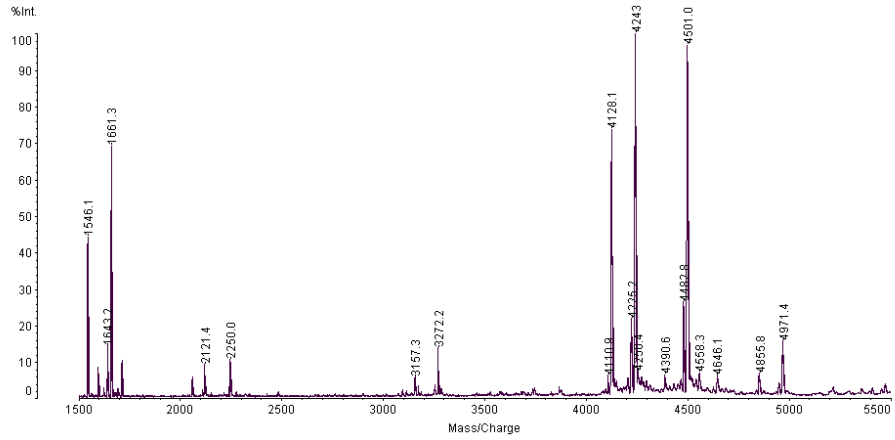

%Int.
4243
4501.0
4128.1
1661.3
1546.1
1643.2
2121.4
2250.0
3157.3
3272.2
4110.9
4225.2
4390.6
4482.8
4558.3
4646.1
4855.8
4971.4
Mass/Charge

Figure 2.1: MALDI-MS spectra of peptide products of the acetic acid digestion of (a) RNase A without the DTT (b) RNase A with 5mM DTT, and (c) ribonuclease B with 5mM DTT. Peptides identified are shown in Tables 1 and 2.

Position	Obs.	MW (expt)	MW (calc)	Delta mass	Miss cleavage	Peptide sequence
1–13	1546.5	1545.493	1546.75	–1.2570	0	-.KETAAAKFERQHM.D
1–14	1660.6	1659.593	1661.837	–2.2444	1	-.KETAAAKFERQHMD.S
1–14	1661.7	1660.693	1661.837	–1.1444	1	-.KETAAAKFERQHMD.S
15–37	2496.1	2495.093	2496.753	–1.6606	0	D.SSTSAASSSNYCNQMMKSRNLTK.D
15–38	2611.1	2610.093	2611.841	–1.7480	1	D.SSTSAASSSNYCNQMMKSRNLTKD.R
39–53	1715.7	1714.693	1715.928	–1.2349	1	D.RCKPVNTFVHESLAD.V
39–82	4852.7	4851.693	4855.447	–3.7543	2	D.RCKPVNTFVHESLADVQAVCSQKNVACK NGQTNCYQSYSTMSIT.D
39–83	4968.5	4967.493	4970.534	–3.0417	3	D.RCKPVNTFVHESLADVQAVCSQKNVACK NGQTNCYQSYSTMSITD.C
54–82	3156.9	3155.893	3157.535	–1.6421	0	D.VQAVCSQKNVACKNGQTNCYQSYSTMS IT.D
54–83	3270.9	3269.893	3272.622	–2.7295	1	D.VQAVCSQKNVACKNGQTNCYQSYSTMS ITD.C
84–120	4125	4123.993	4127.641	–3.6480	0	D.CRETGSSKYPNCAYKTTQANKHIIVACEG NPYVPVHF.D
84–121	4240.5	4239.493	4242.728	–3.2354	1	D.CRETGSSKYPNCAYKTTQANKHIIVACEG NPYVPVHFD.A

Table 2.1: Peptides from the acid digestion of ribonuclease A identified by Mascot.

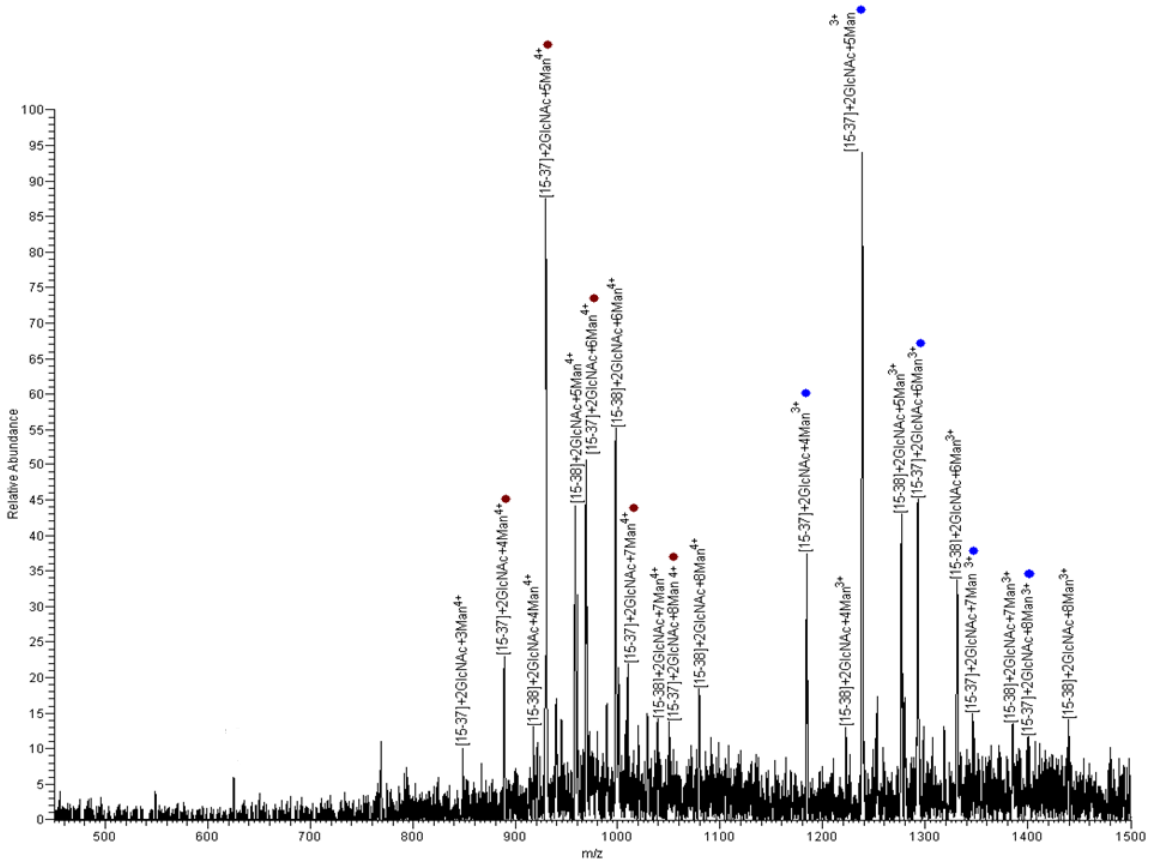
Position	Obs.	MW (expt)	MW (calc)	Delta mass	Miss cleavage	Peptide sequence
1–13	1546.1	1545.093	1546.75	–1.6570	0	-.KETAAAKFERQHM.D
1–14	1660.2	1659.193	1661.837	–2.6444	1	-.KETAAAKFERQHMD.S
1–14	1661.3	1660.293	1661.837	–1.5444	1	-.KETAAAKFERQHMD.S
39–53	1715.3	1714.293	1715.928	–1.6349	1	D.RCKPVNTFVHESLAD.V
39–82	4855.8	4854.793	4855.447	–0.6543	2	D.RCKPVNTFVHESLADVQAVCSQKNVACK NGQTNCYQSYSTMSIT.D
39–83	4971.4	4970.393	4970.534	–0.1417	3	D.RCKPVNTFVHESLADVQAVCSQKNVACK NGQTNCYQSYSTMSITD.C
54–82	3157.3	3156.293	3157.535	–1.2421	0	D.VQAVCSQKNVACKNGQTNCYQSYSTMS IT.D
54–83	3272.2	3271.193	3272.622	–1.4295	1	D.VQAVCSQKNVACKNGQTNCYQSYSTMS ITD.C
84–120	4128.1	4127.093	4127.641	–0.5430	0	D.CRETGSSKYPNCAYKTTQANKHIIVACEG NPYVPVHF.D
84–121	4243	4241.993	4242.728	–0.7354	1	D.CRETGSSKYPNCAYKTTQANKHIIVACEG NPYVPVHFD.A
84–124	4501	4499.993	4500.014	–0.0217	2	D.CRETGSSKYPNCAYKTTQANKHIIVACEG NPYVPVHFDASV.-
1–13	1546.1	1545.093	1546.75	–1.6570	0	-.KETAAAKFERQHM.D

Table 2.2: Peptides from the acid digestion of ribonuclease B identified by Mascot.

N-Linked Glycoprotein-RNase B

We compared the relative abundances of the glycoforms produced from the same sample by acid cleavage and trypsin in order to evaluate the stability of bonds within the carbohydrate chain during the short treatment with hot acid. Four small tryptic glycopeptides were observed, [34-49], [32-49], [34-37], [32-37], while two glycopeptides [15-37], [15-38] were detected from the acid digestion with molecular masses around 4000 Da (Figure 2.2). Analysis of the former was optimized for [34-39] and the latter for [15-37]. These two pieces of peptides were good candidates for quantitative analysis of glycopeptides because they carried a relatively full pattern of glycan forms. They both have 4-8 mannoses while the other peptides miss one or several variants of glycans. The latter glycopeptides weighed around 4 kDa, including the carbohydrate, and were ionized mainly in two charge states, +3 and +4. The relative abundances of glycoforms of $[34\text{-}39]^{2+}$, $[15\text{-}37]^{3+}$ and $[15\text{-}37]^{4+}$ are plotted in Figure 2.3. The charge states seem to have little effect on the distribution and the relative abundance of glycoforms because they look almost the same between the peptide [15-37] in two differenct charge states, +3 and +4 (Figure 2.3). The relative abundances of glycoforms containing 5, 6, 7 and 8 Man residues are comparable. However, a glycopeptide with Man-9 is observed only among the tryptic products, and the abundance of the Man-4 glycoforms is higher in the acid products. These differences may reflect the preferential analysis and detection of the smaller ions in the series of heavy glycopeptides or minor degradation of the carbohydrate sidechain in the acid treatment.

(a)


Relative Abundance
m/z
[15-37]+2GlcNAc+3Man4+
[15-37]+2GlcNAc+4Man4+
[15-38]+2GlcNAc+4Man4+
[15-37]+2GlcNAc+5Man4+
[15-38]+2GlcNAc+5Man4+
[15-37]+2GlcNAc+6Man4+
[15-38]+2GlcNAc+6Man4+
[15-37]+2GlcNAc+7Man4+
[15-38]+2GlcNAc+7Man4+
[15-37]+2GlcNAc+8Man 4+
[15-38]+2GlcNAc+8Man4+
[15-37]+2GlcNAc+4Man3+
[15-38]+2GlcNAc+4Man3+
[15-37]+2GlcNAc+5Man3+
[15-38]+2GlcNAc+5Man3+
[15-37]+2GlcNAc+6Man3+
[15-38]+2GlcNAc+6Man3+
[15-37]+2GlcNAc+7Man 3+
[15-38]+2GlcNAc+7Man3+
[15-37]+2GlcNAc+8Man3+
[15-38]+2GlcNAc+8Man3+

(b)

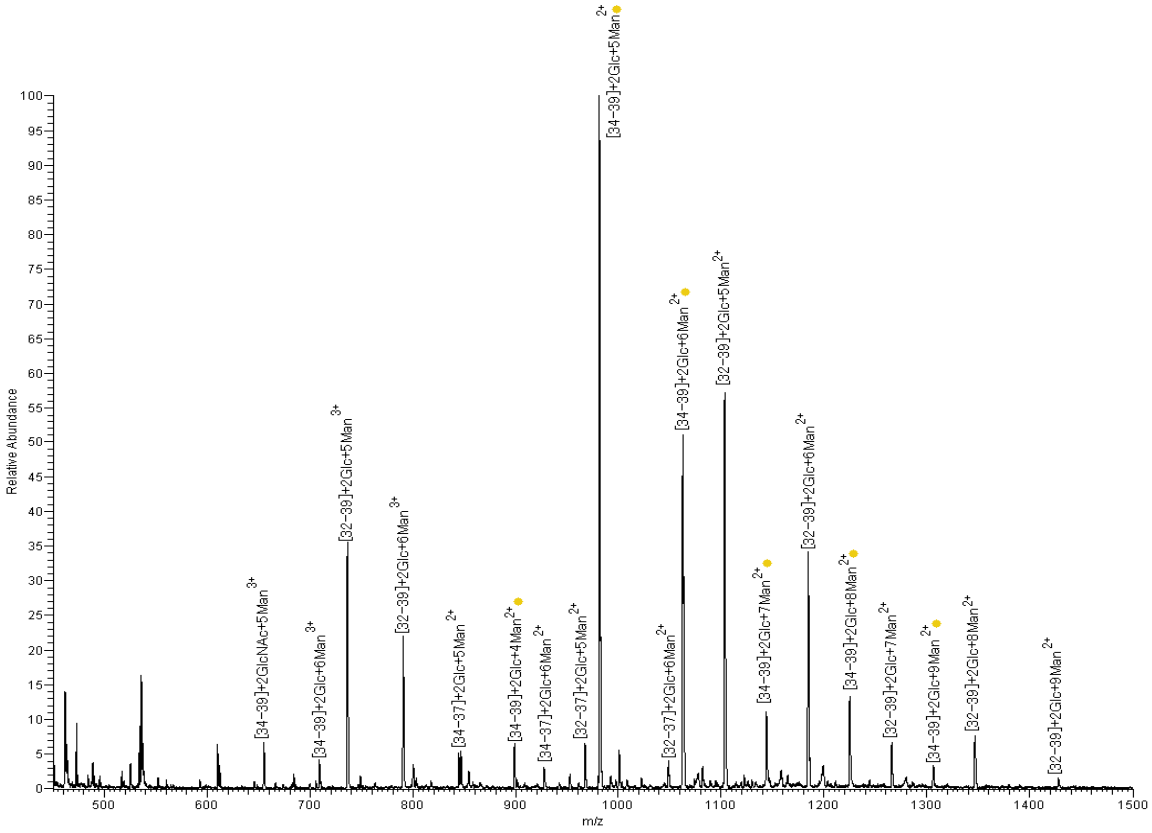


Figure 2.2: LC-MS spectra acquired on a LTQ-XL mass spectrometer of the glycopeptide products from (a) acid cleavage; (b) trypsin cleavage, showing glycan heterogeneity.

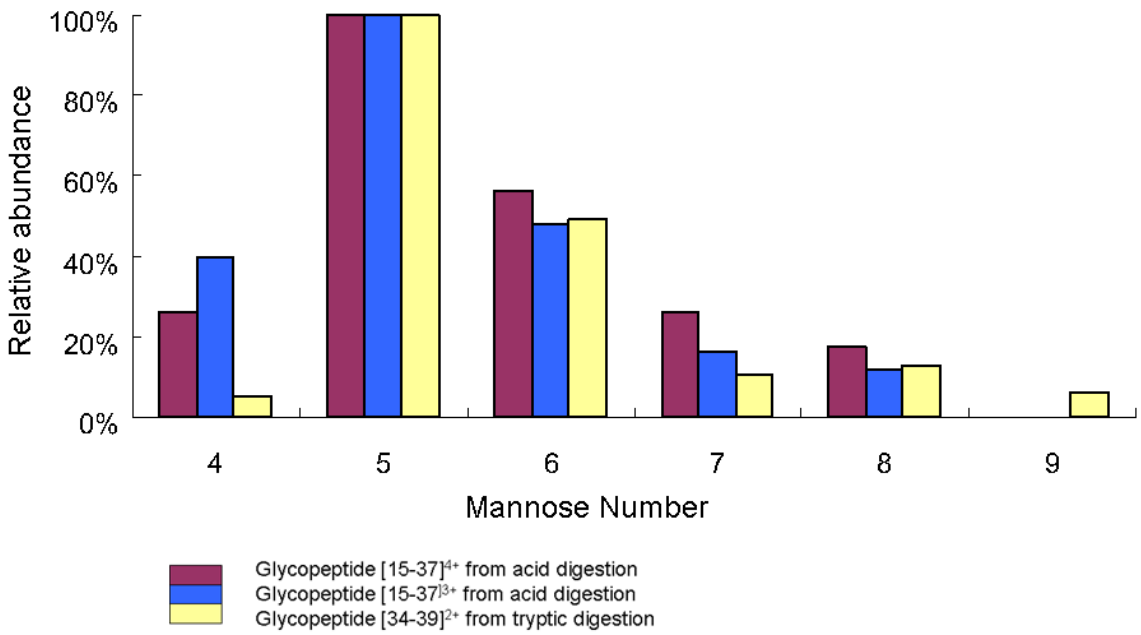


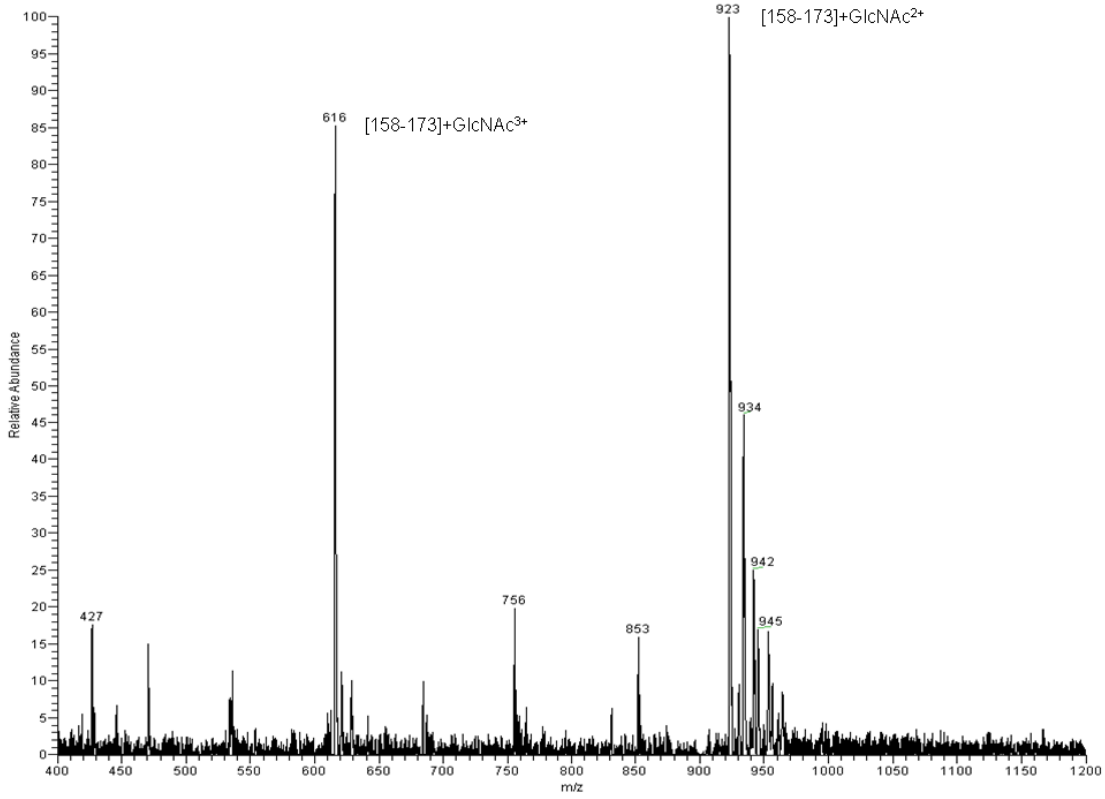
Figure 2.3: Relative abundances of glycopeptides produced by acetic acid and tryptic digestions of ribonuclease B.

O-Linked Glycoprotein-α-Crystallin A Chain

The intact tryptic glycopeptide [158-173] from O-linked α-crystallin A chain is characterized in two charge states, +2 and +3 in the electrospray as shown in Figure 2.4(a). Figure 2.4(b) presents the spectrum of peptide [152-173], recovered from the acid digestion of α-crystallin A chain. The spectrum supports the conclusion that the O-linked GlcNAc attached to Ser162 was removed by the acid treatment. The most commonly used method for O-glycan removal is β-elimination under the alkaline condition. O-linked glycans can also be detached from the peptide backbone via trifluoromethanesulfonic acid even though this reaction seems to be hard to control. Our observation of the removal of O-linked glycan by acid treatment in the microwave oven is consistent with this report[76].

The method of microwave-accelerated acid digestion was not applied to Ara h 2 in the ongoing studies because we did not know what kind of glycosylation that Ara h 2 might have. We want to keep the PTM on Ara h 2 for analysis. Microwave radiation was incorporated to regular trypsin digestion of Ara h 2 instead.

(a)


923
[158-173]+GlcNAc2+
616
[158-173]+GlcNAc3+
934
942
945
756
853
427
Relative Abundance
m/z

(b)

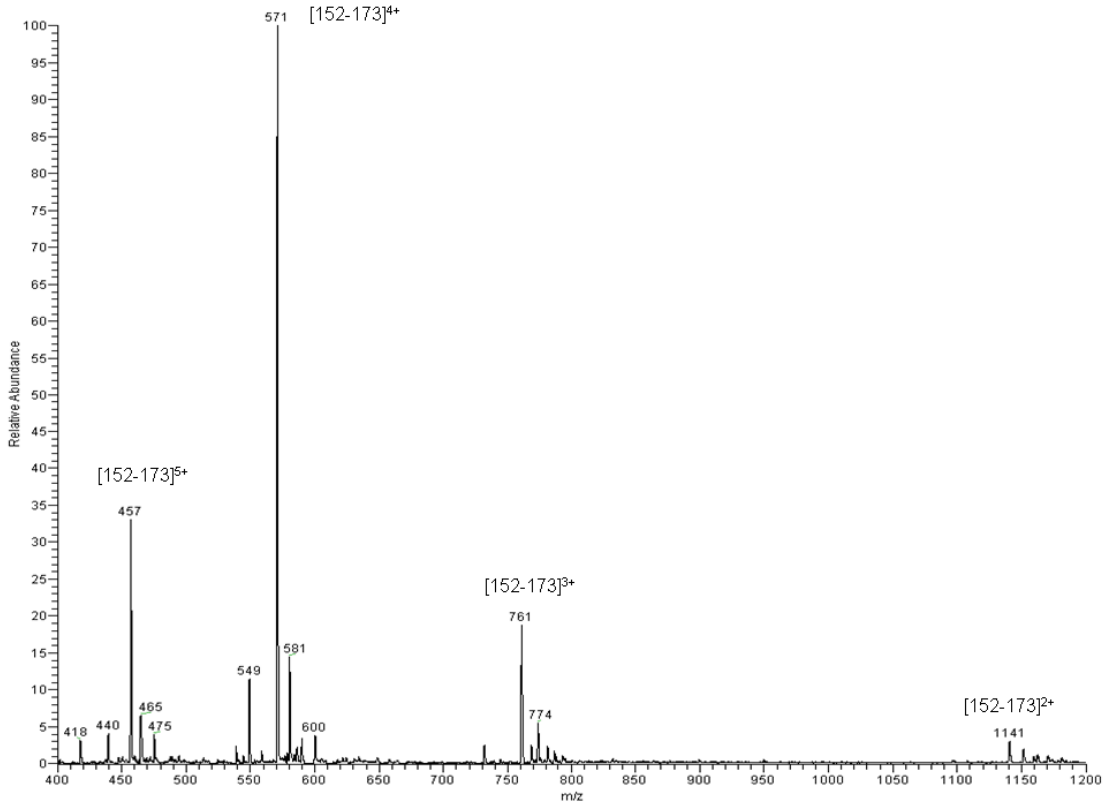


Figure 2.4: MS spectra of (a) the glycopeptide [158-173]+GlcNAc from tryptic digestion; (b) deglycosylated peptide [152-173] from acid digestion.

Chapter 3: Characterization of all four isoforms of Ara h 2 in the raw peanuts, their primary structure, PTM, and disulfide linkages

Introduction

Two cDNA sequences of Ara h 2 have been reported in the literature. However, proteins could be enzymatically modified after their translation. Ara h 2 has not been characterized completely in the protein level. There is also ambiguous information about whether Ara h 2 is glycosylated or not. In this chapter, we used mass spectrometric methods to elucidate the primary structures and post translational modifications (PTMs) of Ara h 2 isoforms. The disulfide bond pattern in Ara h 2 has been characterized. This serves as a starting point for the on-going studies of Ara h 2 in the dark roasted peanuts.

Materials and Experiments

Materials

Solvents for HPLC gradients (acetonitrile, water) were purchased from Burdick & Jackson (Morristown, NJ). Formic acid, cyanogen bromide (CNBr), and dithiothreitol (DTT) were purchased from Sigma (St. Louis, MO). Ara h 2 protein, prepared from the Runner variety without using any acid or protease, was purchased from TNO (Zeist, the Netherlands). Trypsin was purchased from Promega (Madison, WI). Trifluoro acetic acid (TFA) was purchased from J.T. Baker (Phillipsburg, NJ). All chemicals and proteins were used without further purification.

Separation of Ara h 2 isoforms

Ara h 2 isoforms were separated using an AKTA purifier FPLC (GE Healthcare Bio-Sciences Corp., Piscataway, NJ). One hundred micrograms of Ara h 2 suspended in 1 mL distilled water was loaded onto a BioSuite DEAE 2.5 µm NP anion exchange column with an ID of 2.6 mm and a length of 35 mm (Waters, Milford, MA). The isoforms were eluted by increasing solvent B from 1%-7% in 40 column volumes (cv) (solvent A: 0.1M Tris buffer, pH 7.4; solvent B: HPLC grade water) using a flow rate of 1 mL/min. Each fraction was collected via a fraction collector that would initiate/stop the collection when the UV absorption level exceeded/dropped below a specified level. Each fraction was desalted and concentrated in a 5000 Da cut off spin column (Millipore, Temecula, California). The protein concentration of the P1 isoform was measured at about 0.05 mg/mL by RC DC protein assay (Biorad, Hercules, CA). The concentrations of the other three isoforms were not determined experimentally, but they were at comparable concentration to P1 based on their comparable respective chromatographic peak areas in Figure3.1. One hundred micrograms of each isoform was loaded onto a reverse phase C18 column (1.0 mm I.D. ×100 mm) with 1.7 µm packing material on an ACQUITY LC system (Waters, Milford, MA). The gradient was linearly ramped from 0% to 100% solvent D (solvent C: 0.1% Formic acid and 1% acetonitrile in water, solvent D: 0.1% formic acid and 1% water in acetonitrile) in 40 min. The flow rate was at 100 µL/min. The molecular mass of each isoform was measured using a Q-TOF premier mass spectrometer (Waters, Milford, MA). Each molecular weight was determined by deconvoluting the MS spectrum using the MaxEnt1 of the MassLynx software (Waters, Milford, MA).

Proteolytic digestion of Ara h 2 isoforms

A microwave-assisted trypsin digestion was applied to Ara h 2. The Discover Benchmate microwave system (CEM Corp, Matthews, NC) was utilized to perform the protein digestion. Fifty microliters of protein solution (0.05 mg/mL) was incubated with 10 mM DTT at 60°C for 20 min and then heated in the microwave with 2 µL trypsin (0.4 µg/µL) for 30 min at 37°C. Formic acid was added to reach a final concentration of 0.1% to stop the reaction. In experiments designed to minimize missed cleavages, trypsin digestion was carried out in the microwave system for 60 min. The other experimental conditions were the same as described above.

LC-MS/MS

The peptide mixtures that were obtained from 30 and 60 min microwave accelerated trypsin digestions were separated on a Waters nanoACQUITY HPLC interfaced to an LTQ-XL ion trap mass spectrometer (Thermo Electron, San Jose CA). Ten microliters product was loaded onto a reverse-phase C18 column 1.0 mm I.D. ×100 mm) with 1.7 µm bridged ethyl hybrid (BEH) particle packing (Waters, Milford, MA). The gradient was ramped from 0% to 10% B in 5min and then ramped to 30% B in 25 min (solvent A: 0.1% formic acid and 1% acetonitrile in water, solvent B: 0.1% formic acid and 1% water in acetonitrile). Collision induced dissociation (CID) and electron transfer dissociation (ETD) were both used when necessary. For CID, 35%-40% collision energy was applied to precursor ions depending on the mass of the precursor peptide. For ETD, the reaction time was 80

milliseconds. All spectra were analyzed manually. In all cases, targeted CID was employed.

For purposes of confirmation, a peptide mixture from a 60 min digest of P1 isoform was introduced to a Q-TOF mass spectrometer (Waters, Milford, MA) using the same LC system and the same gradient as described above. CID was performed with a default collision energy profile that provided collision energy adjustment according to the precursor ion mass. The data was submitted for database search as described below in order to confirm the result from manual analysis of the protein sequence and post translational modifications.

Mascot Database Searches

The raw data from the 60 min digestion of P1 acquired on the Q-TOF mass spectrometer was processed using the default Q-TOF settings via Mascot distiller (Matrix Science Ltd., London, U.K.) to generate the Mascot generic data format that is compatible with Mascot database searching. The search was carried out using an in-house Mascot server (www.matrixscience.com). Trypsin was indicated as the proteolytic enzyme. Instrument was specified as ESI-QUAD-TOF. Both the mass tolerances of precursor ions and fragment ions were set at ±0.3 Da. No modification on cysteine was specified. Oxidation of proline was specified as a variable modification. Three missed cleavage sites were allowed. The search was carried out against the green plant database in the SwissProt database (http://ca.expasy.org/sprot/).

Fraction Collection of Peptides Containing Disulfide Bonds

P3 and P4 isoforms were digested with trypsin as described above, except that no reducing reagent was added. The peptide mixtures were separated on the same LC system as described for the LC-MS/MS. The concentration of solvent B was linearly increased from 0-50% through 50 min. All the possible combinations of disulfide linkages were considered and the corresponding masses of peptides containing interchain disulfide bonds and intrachain disulfide bonds have been calculated. When these peptides are ionized via ESI, they usually carry several charges. Thus their corresponding series of m/z values can be calculated. The retention time of these peptides can be determined easily via reconstructed ion chromatogram using their m/z values. Eventually the disulfide bond containing peptides were manually collected at their respective eluting times.

CNBr Treatment of Peptide Fractions Containing Disulfide Linkages

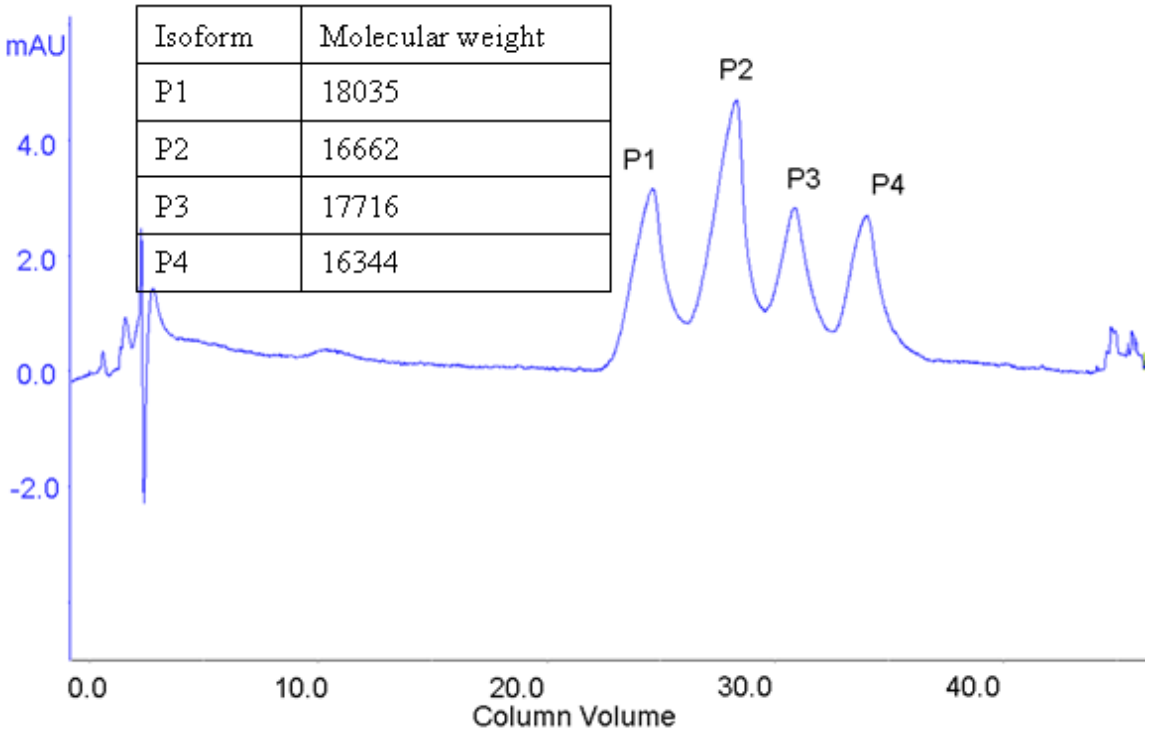
The CNBr solution was made by adding 529 mg CNBr to 1 mL ACN to reach a final concentration of 5 M. Ten microliters of each peptide fraction was mixed with 15 µL TFA and 5µL CNBr solution. The mixture was incubated at room temperature in the dark for 15 hours. Then the mixture was mixed with an equal volume of distilled water and dried down in the SpeedVac. This step was repeated once. Two microliters of CHCA solution (10 mg of CHCA in 70% ACN, 0.1% TFA) was added to suspend the sample. Half microliter sample was spotted on a MALDI plate and analyzed on a 4800 MALDI TOF/TOF mass spectrometer (Applied Biosystem,

Foster City, CA). LC-MS/MS was not used because the collected fraction is a simple mixture, including the target peptides that contain the disulfide bonds.

Results and Discussion

Determination of primary amino acid sequences of Ara h 2 isoforms

The chromatographic separation of Ara h 2 isoforms is shown in Figure 3.1. There are four chromatographic peaks, which indicate the presence of four isoforms. The molecular masses of the collected fractions (Figure 3.1 inset) were measured at 18035, 16662, 17716, and 16344 Da, respectively. These four isoforms were named P1-P4 based on their eluting sequence. The four isoforms may be divided into two pairs based on their molecular masses. P1 and P3 belong to the heavy pair, and P2 and P4 are the light pair.



Isoform	Molecular weight
P1	18035
P2	16662
P3	17716
P4	16344

Figure 3.1: Chromatogram of Ara h 2 isoform separation by anion exchange column and molecular weight of each Ara h 2 isoform which were measured by Q-TOF mass spectrometer

Collision induced dissociation (CID) spectra of the C-terminal peptides from isoforms P1, P3, and P4 are displayed in Figure 3.2 (a), (c), and (d), respectively. However, CID of the C-terminal peptide of P2 [114-139] did not provide information as useful as that produced by electron transfer dissociation (ETD). In any case, ETD of the C-terminal peptide from P2 provides twelve c ions and seven z ions, as shown in Figure 3.2 (b). The CID and ETD spectra allow for assignment of the C-terminal sequence for each of the isoforms. A previously reported amino acid variation at residue 142 is confirmed here. The b17 fragment ion in Figure 3.2 (a) indicates that there is a glutamic acid at residue 142 in the P1 isoform. Corresponding fragment ions in Figure 3.2 (b), (c), and (d) indicate that P3 also has a glutamic acid at residue 142, while P2 and P4 have aspartic acid. The results also show that the two sets of isoforms vary by two amino acids at the end of the sequence. The protein sequence of P1 isoform agrees with the deduced protein sequence from one of the reported cDNAs (GenBank Accession Number AY158467). The protein sequence of P2 isoform agrees with the deduced protein sequence from the other reported cDNA (GenBank Accession Number FJ713110). Twelve amino acids are missing in the middle of P2 compared to P1.P3 and P4 carry two fewer amino acids, Arg and Tyr, at the carboxyl termini, compared to P1 and P2 (Figure 3.3 and Figure 3.2). An artifactual origin of the truncated proteins during the commercial isolation seems less likely, because the same family of proteins has now been detected in extracts of raw peanuts of the same variety made in the our laboratory.

Determination of PTMs of Ara h 2 Isoforms

Several post translational modifications (PTM) not predicted by the cDNA have been elucidated. The 60 min tryptic digestion of P1 produced smaller peptides with fewer missed cleavages. From this digestion, PTM were identified in peptides from the central region of the protein (Figure 3.4 (a) and (b)). In Figure 3.4 (b), the CID spectrum of peptide [42-59] from P1 reveals cleavages at 15 of 17 possible backbone sites. It can be seen that the masses of prolines 46 and 53 are increased by 16 Da, while prolines 43, 50, and 57 are not. We propose that the 16 Da increment reveals proline hydroxylation, a post translational modification known to be common in plant proteins[77]. Peptide [61-70] from P1 was also mapped (Figure 3.4 (a)) via three b ions and five y ions. The masses of fragment y_5 and y_6 ions indicate that proline 65 is also hydroxylated in the heavy isoform. The same peptides [42-59] and [61-70] were identified in P3 isoform and the fragmentation patterns of those peptides are essentially the same as that of P1 isoform with minor variation in the abundance of fragment ions. Thus, we propose that proline 46, 53, and 65 are hydroxylated in the P3 isoform. In the CID spectrum of peptide [42-58] from light isoform P2, presented in Figure 3.4 (c), proline residues 46 and 53 are seen to be hydroxylated, while prolines at positions 43, 50 and 55 in P2 are not modified. Similarly, a CID spectrum of peptide [42-58] from the P4 isoform was almost the same as that from the P2 isoform with minor differences in relative peak intensities. Consequently, proline residues 46 and 53 in the P4 isoform are hydroxylated. It is worth noting that this modification occurs at analogous sites in both the heavy isoforms and the light isoforms (Figure 3.3) even the numbering of the amino acid residues is different because of the 12 amino acid insertion in the P1 and P3 isoforms.

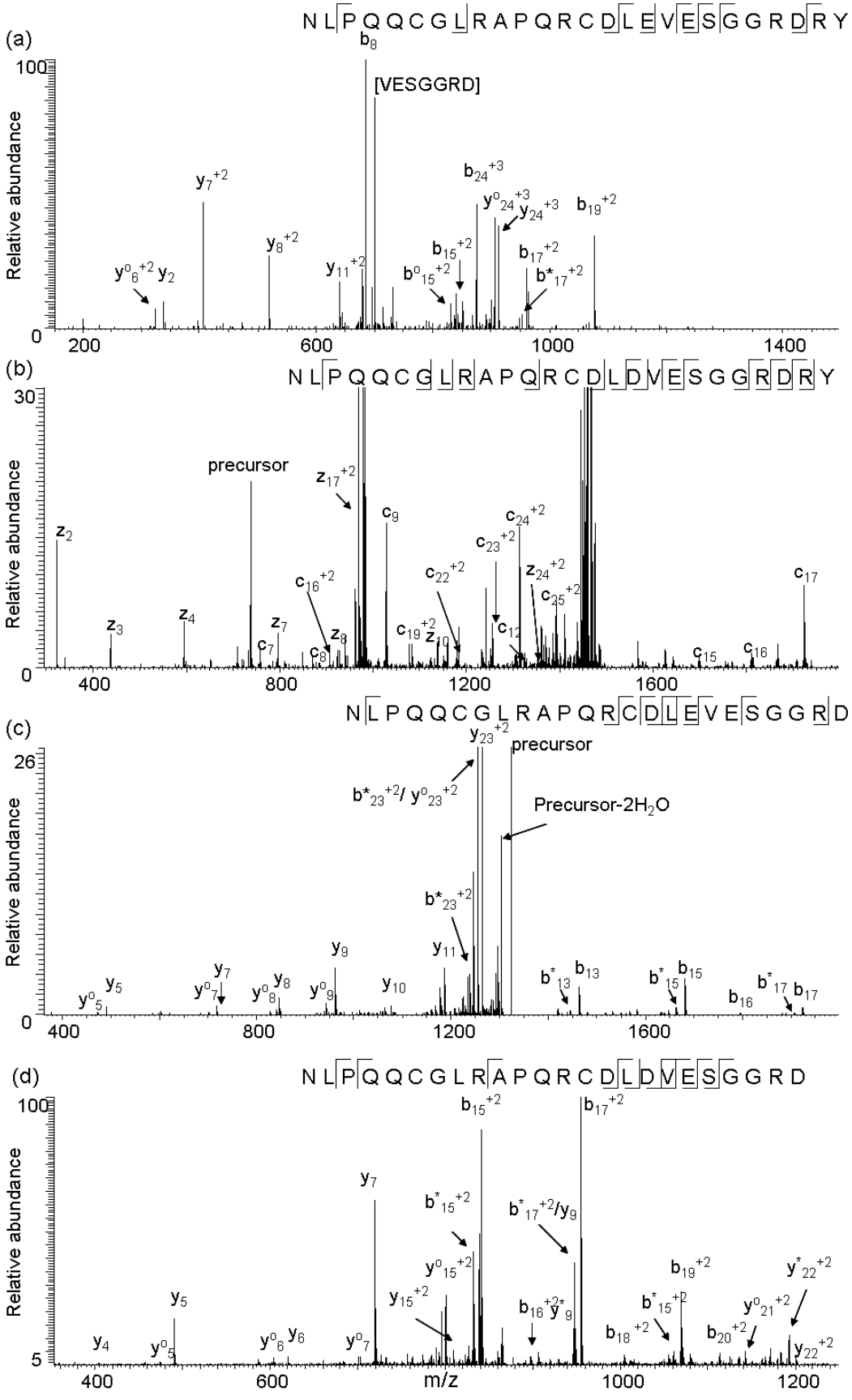

(a)
N L P Q Q C G L R A P Q R C D L E V E S G G R D R Y
[VESGGRD]
Relative abundance
(b)
N L P Q Q C G L R A P Q R C D L D V E S G G R D R Y
precursor
Relative abundance
(c)
N L P Q Q C G L R A P Q R C D L E V E S G G R D
precursor
Precursor-2H2O
Relative abundance
(d)
N L P Q Q C G L R A P Q R C D L D V E S G G R D
Relative abundance
m/z

Figure 3.2: MS/MS spectra acquired on a LTQ mass spectrometer of c-terminal peptides of each isoform of peanut allergen protein. (a) CID spectrum of C-terminal peptide [126-151] from the P1 isoform, quadruply charged ion at m/z 741 was selected as the precursor ion; (b) ETD spectrum of C-terminal peptide [114-139] from the P2 isoform, quadruply charged ion at m/z 738 was selected as the precursor ion; (c) CID spectrum of C-terminal peptide [126-149] from the P3 isoform, doubly charged ion at m/z 1322 was selected as the precursor ion; (d) CID spectrum of C-terminal peptide [114-137] from the P4 isoform, triply charged ion at m/z 877 was selected as the precursor ion. "^{0}" indicates loss of H_2O, "*" indicates loss of NH_3.

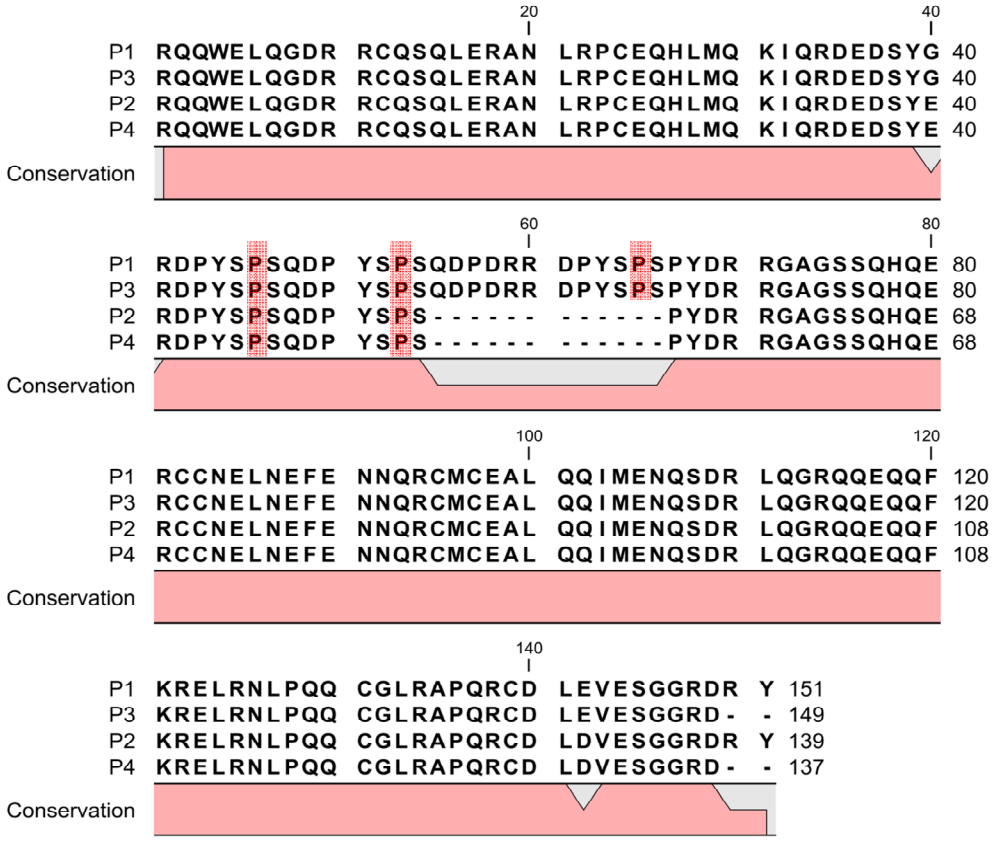


Figure 3.3: Diagram of the primary structure of four isoforms of peanut allergen protein Ara h 2. Hydroxyprolines are highlighted in red. Figure formed using CLC free workbench 3.

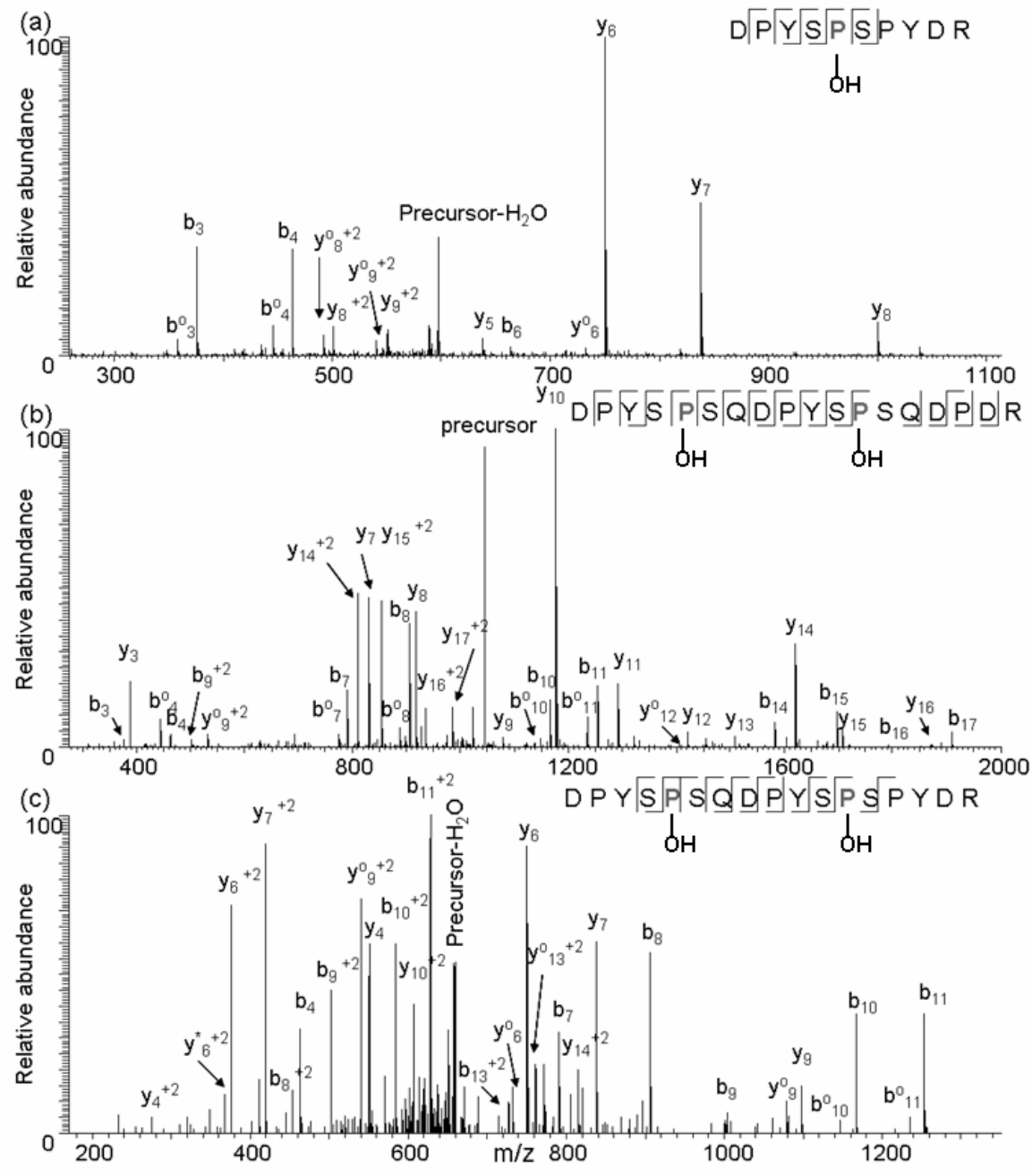


Figure 3.4: MS/MS spectra of peptides acquired on an LTQ mass spectrometer that were identified with proline hydroxylated. (a) CID spectrum of peptide [61-70] from heavy isoform P1 of Ara h 2, doubly charged ion at m/z 607 was selected as the precursor ion; (b) CID spectrum of peptide [42-59] from heavy isoform P1 of Ara h 2, doubly charged ion at m/z 1042 was selected as the precursor ion; (c) CID spectrum of peptide [42-58] from light isoform P2 of Ara h 2, triply charged ion at m/z 669 was selected as the precursor ion.

In order to confirm the manual analysis of the PTM sites, a Mascot search was performed using the LC-MS/MS data of tryptic peptides from P1 acquired on a Q-TOF mass spectrometer. It returned a result with CONG7_ARAHY (P1) identified. Using an E value cut off ≤ 0.05, the peptides identified in an automated manner covered 47% of the protein sequence excluding the signal peptides. Among those peptides, [35-60], [42-60], and [61-70] were identified with 2, 2, and 1 hydroxyprolines. The E values were 0.00061, 0.0045, and 0.0071, respectively, indicating highly reliable identifications. The localized hydroxylation sites are identical to those described above for the heavy isoform P1. This supports the conclusion from *de novo* sequencing that there are hydroxyprolines at positions 46, 53 and 65 in the P1 isoform.

The site-specific hydroxylation of proline residues occurs only on the second proline of repeated peptide motifs PYSPS in the middle of the Ara h 2 isoforms, as is the case for amino acids [43-47] and [50-54] in the P2 isoform (Figure 3.3). Previous studies have identified the immunodominant IgE binding linear epitopes in Ara h 2 as being overlapped regions at amino acids [39-48] and [47-56][8]. Thus, both of these peptides contain the PYSPS motif. However, hydroxylation of prolines was not known to occur when these epitopes were identified, and it is not clear how the presence of hydroxyprolines affects the allergenicity of Ara h 2, Moreover, there is an additional PYSPS motif in isoforms P1 and P3 and it is not known if this additional modified motif can potentially differentiate the allergenicity among these four isoforms.

Peptides identified via *de novo* sequencing covered 100% of the protein sequence of each isoform, and accounted for the molecular masses. The primary structures of all the isoforms are shown in Figure 3.3. Assuming all eight cysteines are involved in disulfide linkages, the calculated average molecular weights of the P1-P4 isoforms, as shown, are 18033, 16661, 17714, and 16341 Da, respectively. The calculated protein masses are very close to the measured protein masses that are 18035, 16662, 17716, and 16344 (Figure 3.1 inset). The signal peptide is confirmed to be removed, and there is no evidence of carbohydrate side chains.

Determination of Disulfide Linkages

As noted above, one of our interests in the study of Ara h 2 is whether and how its structure can be modified by thermal processing. Therefore, an understanding of the disulfide bonds in the native protein from the unroasted peanuts is important. The general approach for evaluating the crosslinking structure in this protein was through examination of the tryptic digestion products obtained from non-reduced protein, as compared to the products obtained through the standard digestion process where a reducing agent is added. Fractions containing the cross-linked products were collected from the P4 isoform and analyzed. All the tryptic peptides that contain cysteines are given a letter in Table 3.1. Evaluation of the fractions revealed that there were no products corresponding to single cross-linked peptides (e.g. A-B, E-F combinations). However, among all the seven tryptic peptides in the P4 isoform that contain cysteine residues (Table 3.1), there are two peptides, C and D, that contain two cysteines. One possibility, that C and D are bridged by two cross-linkers (C+D),

was also ruled out by the absence of a product with a molecular weight corresponding to that combination. Another possibility, the product, C+D+X+Y (where X and Y are one of the single-cysteine peptides A, B, E, F or G), was again not observed, since no ion was detected with a mass greater than 5000 Da. However, C and D could each be linked to two single-cysteine peptides to create a cross-linked tripeptide (e.g. C+X+Y, where X is A, B, E, F or G). We have considered all the possible disulfide bond linkages for both C and D peptides (Table 3.2 and 2.5). There are eight possible ways to form disulfide bond linkages for the C peptide which include one intrachain linkage and seven interchain linkages (Table 3.2). Similarly, there are eight possible ways for the D peptide (Table 3.3).

For the P4 isoforms, a peptide with a mass of 4204 Da (Fraction 1) matches the mass of C+B+E (pattern 5 in Table 3.2). In fraction 2, there were two co-eluting peptides, with masses of 3807 Da and 3922 Da, respectively. The former mass, 3807 Da, agrees with the mass of D+A+F (pattern 4 in Table 3.3). The latter mass, 3922 Da, corresponds to D+A+F with an additional aspartic acid at the end of F, resulting from a missed cleavage at the C-terminal peptide. This results in an additional product with a mass 115 Da larger than C+A+F (this product will be denoted as C+A+F'). The 3922 Da peak was dominant so the following analysis of Fraction 2 was based on it.

To confirm these assignments, the fractions containing cross-linked peptides were treated with CNBr (which cleaves on the C-terminal side of methionine) and analyzed by MALDI (Figure 3.5). In Fraction 2, the 3922 Da peptide (C+A+F') has two cysteines in D that are at residues 83 and 85, separated by a methionine. The

products from CNBr cleavage were calculated for two possibilities of disulfide bond linkages within this disulfide bond linked tripeptide as shown in Figure 3.6. Two peptides, with masses of 1065 Da and 1935 Da, were observed and are in agreement with the values predicted in the left column in Table 4. This indicates that cysteine 12 in peptide A is connected with cysteine 83 in D, and cysteine 85 in D is connected with cysteine 127 in F' (Table 3.4).

In Fraction 1, the C+B+E tripeptide could not be similarly characterized. Cysteines70 and 71 in peptide C are adjacent, not separated by a methionine as in D. Thus, we have not found a definitive way to characterize the disulfide linkages in this fraction. Reduction without alkylation generated three peptides with masses of 1625 Da, 1727 Da, and 1086 Da (data not shown), corresponding to peptides B, C, and E, and further confirming the composition of the C+B+E tripeptide. Thus, we know that Cys25 either links to Cys70 or 71 while Cys119 connects with Cys71 or 70. The possibility cannot be eliminated that two patterns of adjacent S-S linkages co-exist in Fraction 1.

The same analysis was carried out with the peptide fractions from the P3 isoform. Figure 3.5 (c) and (d) shows the collected tri-peptide fraction before and after CNBr treatment. Connectivity between Cys12-Cys95 and between Cys97-Cys139 has been determined. Cys25 is linked either to Cys91 or Cys92, and Cys131 is linked to Cys 92 or Cys91. Although the insertion of 12 amino acids in the P3 isoform shifts the cysteine residue numbers, all crosslinking occurs at comparable positions in isoforms P3 and P4. The disulfide linkages of P1 and P2 isoforms were not studied here, but they are expected to be comparable to P3 and P4, respectively,

differing by 2 more amino acids at the C-termini. These results are consistent with an earlier proposal that the disulfide bonds in Ara h 2 are homologous to those in the bifunctional α-amylase/trypsin inhibitor RBI protein from ragi seed [72], which in turn suggests that there is no disulfide bond scrambling in the process of purifying Ara h 2.

To summarize, the mass spectrometry studies of Ara h 2 in the commercial sample reported here identify a heavy isoform pair and a light isoform pair that are differentiated by a 12 amino acid insertion in the middle of the sequence, and amino acid variations at residues 40 and 142 in the heavy isoforms and residues 40 and 130 in the light isoforms (Figure 3.3). The heavy isoform pair has glycine and glutamate at residues 40 and 142, while the light isoform pair carries glutamate and aspartate at residues 40 and 130, respectively. Within each pair, there is a variation in the C-terminus. The C-termini assigned are consistent with a mass difference 319 Da. The two heavy isoforms both contain three hydroxyproline residues, while the two light isoforms have two hydroxyprolines in homologous positions, as shown in Figure 3.3. No glycosylation was found, in contrast to past literature reports. There are four disulfide bonds in each isoform of Ara h 2. Two out of the four disulfide bonds have been fully defined and the remaining two are localized to two adjacent Cys residues. The 12 amino acid insertion and amino acid variations at residues 40 and 142 in the heavy isoforms (40 and 130 in the light isoforms) are consistent with previously reported sequences translated from the cDNA sequences[6, 7]. Determinations of site-specific proline hydroxylation, disulfide linkages and C-terminal variation are reported here for the first time.

Knowledge of the full sequence and modifications enables on-going studies of the way in which the protein is modified by thermal processing (such as roasting) and provides information that may be useful in the development of sensitive and automated assays, such as multiple reaction monitoring (MRM) using a triple quadruple mass spectrometer for detecting Ara h 2 in food. The discovery and definition of post-translational modifications is expected to facilitate continued studies of factors that may affect the structural determinants of allergenicity in the Ara h 2 family, particularly where processing may produce modification of the protein.

Table 3.1: List of all the tryptic peptides that have cysteine residues in P4 isoform.

Tryptic peptide that contains cysteine residues	Sequence	Residue number	Monoisotopic mass (Da)
A	CQSQLER	[12-18]	862.4
B	ANLRPCEQHLMQK	[19-31]	1566.8
C	CCNELNEFENNQR	[70-82]	1611.6
D	CMCEALQQIMENQSDR	[83-98]	1897.8
E	NLPQQCGLR	[114-122]	1027.5
F	CDLDVESGGR	[127-136]	1049.5
F’	CDLDVESGGRD	[127-136]	1164.5

Table 3.2: Possible disulfide bond linkages of C peptide of P4 isoform. Mass ≥3000Da are in average mass.

Pattern	Possible linkages of C peptide	Mass (Da)
1	C intrachain linkages	1609.6
2	C+A+B	4039.5
3	C+A+E	3499.8
4	C+A+F	3521.7
5	C+B+E	4204.8
6	C+B+F	4226.7
7	C+E+F	3687.0
8	C+D	3507.9

Table 3.3: Possible disulfide bond linkages of D peptide of P4 isoform. Mass ≥3000Da are in average mass.

Pattern	Possible linkages of D peptide	Mass (Da)
1	D intrachain linkages	1895.8
2	D+A+B	4326.0
3	D+A+E	3786.3
4	D+A+F	3808.2
5	D+B+E	4491.3
6	D+B+F	4513.2
7	D+E+F	3973.5
8	C+D	3507.9

Table 3.4: List of masses of the product peptides from CNBr cleavage of disulfide bond linked tri-peptide (Fraction 2) from P4 isoform.

Disulfide bond linked tri-peptide	D+A+F’	
Possible disulfide bond linkages within the cluster	Cys12-Cys83; Cys85-Cys127	Cys12-Cys85; Cys83-Cys127
Masses of the products from CNBr cleavage (Da)	1064.4	1746.7
	1933.9	1251.4
	747.0	747.0

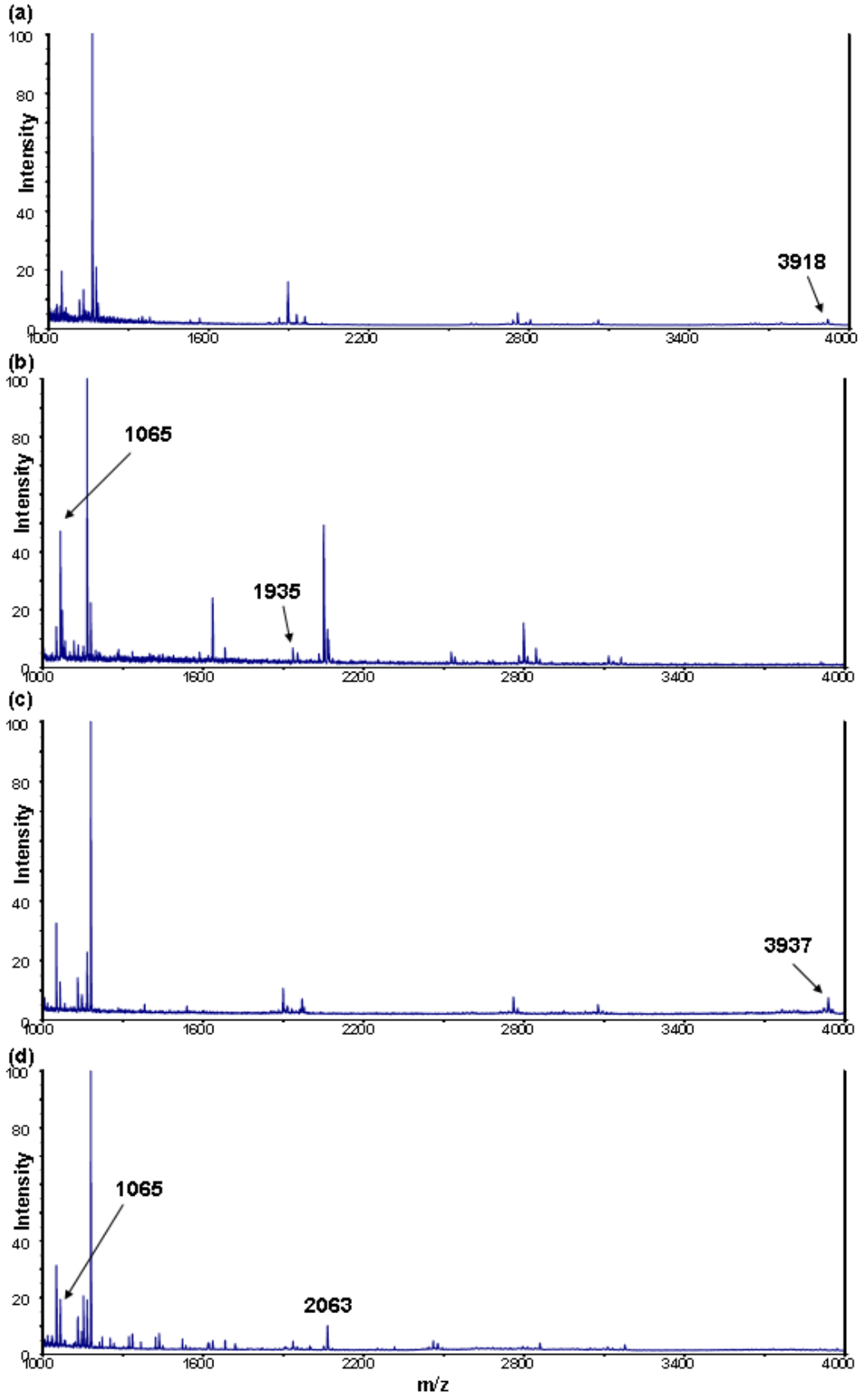

(a)
Intensity
100
80
40
20
0
1000
1600
2200
2800
3400
4000
3918
(b)
Intensity
1065
1935
(c)
Intensity
3937
(d)
Intensity
1065
2063
m/z

Figure 3.5: MALDI spectra of collected peptide fractions from (a) P4 isoform before CNBr cleavage; (b) P4 isoform after the CNBr cleavage; (c) P3 isoform before CNBr cleavage; and (d) P3 isoform after CNBr cleavage. All spectra were acquired on ABI 4000 MALDI-TOFTOF instrument.

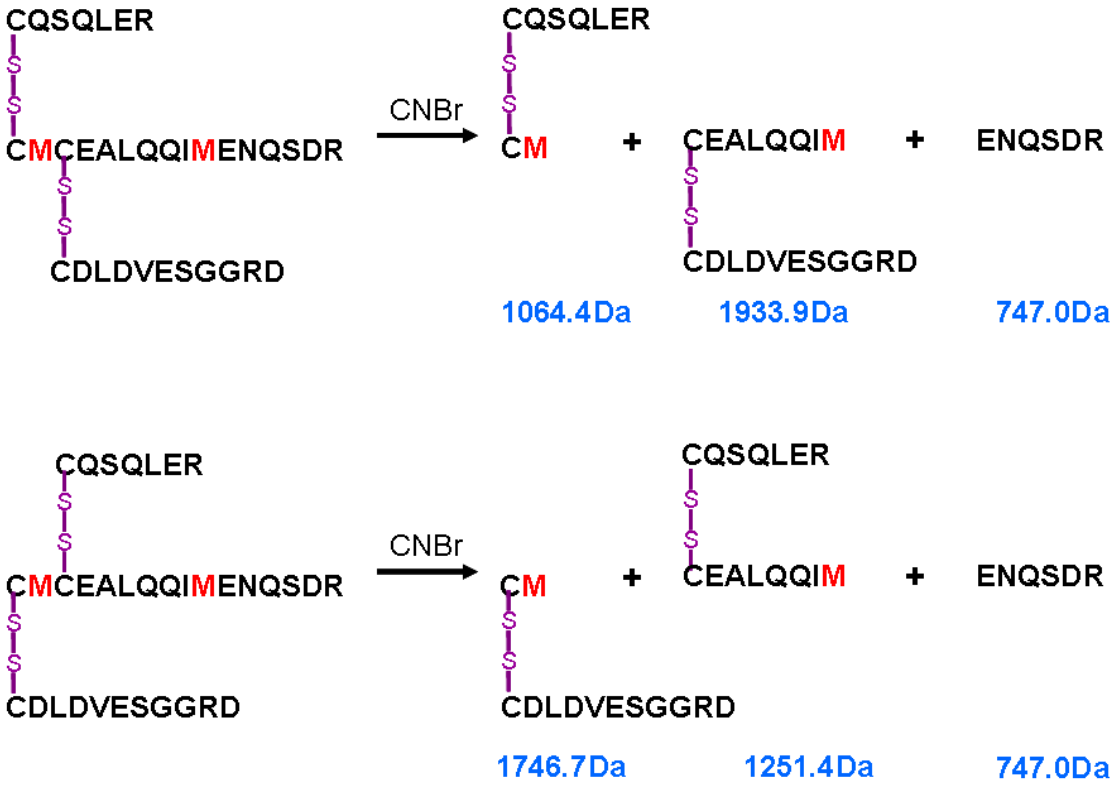


Figure 3.6: Two possibilities of the disulfide bond patterns of Fraction D+A+F' from P4 isoform and their corresponding products from CNBr cleavages.

Chapter 4: Characterization of Ara h 2 related products in roasted peanuts

Introduction

It is known that patients with peanut allergies tend to be more allergic to the roasted peanuts than the raw peanuts. Immunological studies using blood serum from hypersensitive patients suggest stronger binding of the antibodies with the roasted peanuts compared to the raw peanuts [78]. It is not clear how the roasting process affects Ara h 2 and other allergen proteins. What contributes to the increased allergenicity of the roasted peanuts is unknown either. This chapter focuses on the difference between the raw and dark roasted peanuts, especially the effect of thermal processing on the allergen protein Ara h 2. Immunological studies using anti-Ara h 2 polyclonal antibodies were performed to characterize the change of Ara h 2 as a result of the roasting process. Subsequent LC-MS/MS was utilized to confirm the immunological results.

Materials and Experiments

Materials

All the materials were purchased from Biorad (Biorad, Hercules, CA) unless indicated. Raw peanut kernels with skins and partially defatted dark roasted peanut flour (roasting condition was 255°F for 10min, and then 345°F for 10min) were kind gifts from Golden Peanut Company (Alpharetta, GA).

Protein Extraction and Purification

The raw peanuts with skins were grinded with liquid nitrogen to a fine powder. Then they were partially defatted in hexane with vigorous shaking for half an hour. This defatting process was repeated three times. The raw peanut flour was then left in a hood to dry overnight. The dark roasted peanut flour was used directly. The roasting process was carried out at 255°F for 10min, and then 345°F for 10min. A hundred milligram of the raw and dark roasted peanut flours were weighed out and suspended in 1 mL extraction buffer (7 M urea, 2 M thiourea, 1% ASB-14, 40 mM Tris, 0.001% bromophenol blue), respectively. The sample tube was rotated for 3 hours at room temperature. Solids were centrifuged down for 20 min. The supernatants were collected. Protein concentration was determined via RC DC protein assay. For the raw peanuts, the protein concentration is ~40 mg/mL. For the dark roasted peanuts, the concentration is ~23 mg/mL. The concentration variations were minor from extraction to extraction.

Proteins were purified from the crude extraction of raw and dark roasted peanuts via ReadyPrep 2-D cleanup kit following the manual in the kit. Purified protein pellets were ready for 1-D and 2-D gel electrophoresis.

2-D Gel Electrophoresis

Peanut proteins are mainly acidic proteins. Therefore, the first dimension of 2-D gel electrophoresis was performed with a pH 4-7 immobilized pH gradient (IPG) strip. The purified protein pellets corresponding to 100 µg raw peanut proteins and 200 µg dark roasted peanut proteins were redissolved in 200 µL sample rehydration

buffer (8 M urea, 2% CHAPS, 50 mM DTT, 0.2% Bio-Lyte 3/10 ampholyte, 0.001% Bromophenol Blue). The samples were transferred to rehydration tray. The IPG strip (11cm) was gently placed face down in the rehydration tray to absorb the samples. Care was taken to make sure there was no air bubble trapped below the strip. Two milliliters of mineral oil was overlaid on top of the strip to prevent evaporation during the following rehydration process. Then the strip was left at room temperature for rehydration for 12 hours.

Paper wicks were wet by 7 µL distiller water and placed on the wire electrodes. The rehydrated strip was removed from the rehydration tray. Excess mineral oil was drained to remove unabsorbed trace proteins from the surface of the strip and reduced horizontal streaking. The strip was then placed in the IEF tray facing down so the strip touched the wire electrodes on both sides.

The unit temperature was set at 20°C. The focusing conditions included four steps. First, the voltage applied to each IPG strip was rapidly increased from 0 to 250 V in 15 min. Second, the voltage was slowly increased from 250 to 8,000 V in an hour. Third, the voltage was maintained at 8,000 V while the total Volt-Hours ramped to 20,000 V-hr rapidly. Finally, the voltage dropped back to 250 V and kept the proteins focused.

The IPG strip was then transferred to a clean tray facing up. The strip was equilibrated by 4 mL equilibration buffer I (375 mM Tris-HCl, pH 8.8, 6 M urea, 2% SDS, 2% w/v DTT) and 4 mL equilibration buffer II (375 mM Tris-HCl, pH 8.8, 6 M urea, 2% SDS, 2.5% w/v iodoacetamide. Iodoacetamide was added right before use) for 10 min each on an orbital rotator with gentle shaking. The IPG strip was washed

gently with the SDS PAGE running buffer (25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3) and loaded to the IPG well of a 8–16% Tris-HCl polyacrylamide gel (13.3 x 8.7 cm W x L). The overlay agarose was melted in a microwave oven. Then it was added to the IPG well to cover the IPG strip. Efforts were taken to remove any bubbled trapped below the IPG strip. The gel was mounted to the Criterion cell unit for the second dimension electrophoresis. The condition of the voltage and current were the same as described below.

1-D Gel Electrophoresis

For the 1-D gel that was used for Western blotting, 40 µg proteins of raw and roasted peanuts were purified by precipitation (Biorad 2-D cleanup kit) from ~0.68 µL crude extract of raw peanuts and 1.17 µL of roasted peanuts, respectively. The protein pellets were dissolved in 20 µL Laemmli buffer (2% SDS, 10% glycerol, 0.002% bromophenol blue, and 0.0625 M Tris-HCl, pH ~6.8) that contained 10 mM DTT. For the 1-D gel that was used for LC-MS/MS analysis, 1 µL crude extract of raw peanuts (58 µg proteins) was reduced in 10 mM DTT in the presence of 4 M urea at 56°C for 45 min. Then it was alkylated in the presence of 55 mM iodoacetamide for 30 min in the dark. Then the proteins were precipitated using the 2-D cleanup kit (Biorad). The protein pellet was redissolved in 20 µL Laemmli buffer (2% SDS, 10% glycerol, 0.002% bromophenol blue, and 0.0625 M Tris-HCl, pH ~6.8) containing 10 mM DTT. The protein pellet was broken up by pipetting and vortexing. Then the sample was heated at 90°C for 10 min and was allowed to cool to room temperature. For the dark roasted peanut, 1.7 µL crude protein extract (58 µg proteins) was

processed the same way as described above. All samples were loaded to the Criterion Tris-HCl gel (8–16% polyacrylamide gel, 1 mm, 12+2 well, 13.3 x 8.7 cm (W x L)). Molecular weight markers were also loaded when necessary.

The electrophoresis was performed at 200V with a current limit at 100 mA for 60 min. When the separation was complete, the gel was removed from the cassette and washed by distilled water twice for 5 min each to remove the SDS and salts on the gel surface. About 50 mL Coomassie dye (GelCode Blue Stain Reagent, Pierce, Rockford, IL) was used for the gel staining. The incubation time was 1 h with gentle rotation followed by an additional hour of incubation with water for enhanced clarity. If the gels were not used immediately, they were stored in shallow water at 4°C before use.

Western Blotting

The transfer buffer was made by mixing 100 mL of 10× Tris/Glycine buffer with 700 mL distilled H2O and 200 mL of methanol (final concentrations: 25 mM Tris, 192 mM glycine, 20% v/v methanol, pH 8.3). It was pre-chilled in 4°C refrigerator beforehand to help with the heat dissipation during the transfer process. The gel used for transferring was pre-equilibrated in the transfer buffer for 15 min. The nitrocellulose membrane (0.2 µm pore size, 8.5×13.5 cm) was soaked in the transfer buffer before use.

A sandwich style transfer stack was made as follows. A fiber pad was soaked completely in the transfer buffer. A piece of filter paper was placed on top of the fiber pad. It got wet immediately. The pre-equilibrated gel was carefully placed on top of

the filter paper. The pre-soaked nitrocellulose membrane was placed on top of the gel followed by another piece of wet filter paper. Trapped air bubbles in the stack were rolled out by a roller. A second fiber pad was wetted by the transfer buffer and placed on top of the filter paper. This sandwich was placed in a cassette. The membrane should be facing the red side of the cassette.

The Criterion Blotter tank was filled to about 50% with the transfer buffer. Then it was placed on a magnetic stirrer with a magnetic stir bar inside so the heat can dissipate well. An ice block was put inside the tank to keep the transfer buffer cold. The cassette containing the sandwich stack was put in the tank with its red side facing the red side of the tank. The membrane was facing the cathode side right now. The tank was then completely filled with the transfer buffer.

Biorad PowerPAC 1000 was used as the power supply. The voltage was set at 100 V, the current was limited to 500 mA, and the power was limited to 250 W. The transfer was performed for 38 min with stirring.

When the nitrocellulose membrane was ready, the following Western blotting was performed with the fast Western blot kit from Thermo Pierce (Rockford, IL). This kit contains antibody diluent, 10×washing buffer, rabbit optimized HRP reagent, luminol/enhancer solution, and stable peroxide solution. The 10×washing buffer was diluted 10 times with distilled water. The membrane was washed with 20 mL washing buffer for 5 min on an orbital rotator. One microliter anti-Ara h 2 blood serum (from a collaborator at the FDA, raised in rabbit, polyclonal IgG) was diluted by 10 mL antibody diluent (1: 10,000 dilutions). The membrane was incubated with this primary antibody solution at room temperature for 30 min with gentle shaking. The

blot was removed from the primary antibody solution and placed in another clean incubation tray. The HRP solution was prepared right before use by diluting 1 mL rabbit optimized HRP reagent in 9 mL antibody diluent. The membrane was incubated with the HRP solution for 10 min at room temperature with shaking. The membrane was removed to a clean incubation tray and washed with 20 mL washing buffer for 5 min. This step was repeated three times. The membrane was transferred to a clean incubation tray after the washing. Five milliliter luminol/enhancer solution was mixed with 5 mL stable peroxide solution right before use. The membrane was incubated with this mixed solution for 3 min and was quickly placed on the platform of an image station 2000R (Kodak, Rochester, NY). An image of the membrane was generated by exposing for 3-10 min in dark.

In-Gel Trypsin Digestion

The following solutions were made in advance: 100 mM ammonium bicarbonate; destaining buffer (50% acetonitrile, 50 mM ammonium bicarbonate); 10 mM dithiothreitol in 25 mM ammonium bicarbonate. Fifty millimolar iodoacetamide in 25 mM ammonium bicarbonate was made fresh.

In order to confirm the results of immunological studies, the 1-D gels correspond to raw and roasted peanuts were cut into 5 pieces (>100, 100-50, 50-25, 25-15, <15 kDa). Gel bands were excised from Coomassie stained 1-D or 2-D gels. They were cut into 1 mm cubes using a razor blade and transferred to Eppendorf tubes. Five hundred microliter destaining buffer was added. The gel pieces were allowed to destain at room temperature for 15 min. The supernatant was removed and

500 µL distilled water was added. After 15 min incubation, the supernatant was removed. This destaining step was repeated until the stain was gone. Then the gel pieces were dehydrated with 500 µL acetonitrile. The gel pieces were further dried in SpeedVac for 5 min. The dried gel pieces were rehydrated in 500 µL 10 mM DTT and incubated at 56°C for 45 min. The DTT solution was removed and 500 µL 55 mM iodoacetamide was added. Alkylation was performed in dark at room temperature for 30 min.

After the reduction and alkylation, the gel pieces were washed by 200 µL water followed by two times 500 µL destaining buffer. They were dehydrated in 500 µL acetonitrile and further dried down in SpeedVac for 5 min. Five hundred microliter trypsin buffer (1 µg sequencing grade modified porcine trypsin in 500 µL 25 mM ammonium bicarbonate, trypsin was from Promega, Madison, WI) was added to swell the gel particles at 4°C for an hour. Additional 25 mM ammonia bicarbonate can be added if the gel pieces were not all covered by solution after rehydration. The sample was then incubated overnight at 37°C.

The supernatants of the samples were recovered. A hundred microliter 50% ACN in 1% acetic acid was added to each sample for peptide extraction. The extraction was performed at 45 °C for 20min. The supernatants were collected and combined with previously collected supernatants. The peptide extraction was repeated once. Then the samples were dried down in the SpeedVac and resuspended in 10 µL 1% acetic acid solutions.

LC-MS/MS

The peptide mixtures from trypsin digestions were analyzed by LC-MS/MS on a Waters nanoACQUITY HPLC interfaced with an LTQ-XL ion trap mass spectrometer. Ten microliter peptide mixtures were separated on a reverse phase C18 column (150 µm I.D.×100mm) with 1.7 µm bridged ethyl hybrid (BEH) particle packing (Waters, Milford, MA). The gradient started at 10% solvent B. It was ramped to 20% B in 5 min linearly, and eventually reached 40% B in another 50 min linearly. Then the column was equilibrated for 15 min at 100% B (solvent A: 0.6% acetic acid in water, solvent B: 0.6% acetic acid in acetonitrile). MS/MS was carried out via data dependent acquisition in which top 10 most abundant ions were isolate for collision induced dissociation (CID) with 35% collision energy. If the same ions have been sampled twice in 6 seconds it will be excluded for 15 seconds for MS/MS fragmentation.

Mascot Database Search

LC-MS/MS data was searched was against a customized peanut database (contains ~1,000 proteins, searched using the key word “peanut” as an organism name) via Mascot MS/MS ion search. Trypsin was indicated as the used enzyme with 1 possible miss cleavage. Peptide and fragment mass tolerance was set at 0.8 and 0.6 Da, respectively. Fixed modification included carbamidomethyl cysteines because the proteins were treated with iodoacetamide during the digestion. Variable modifications were cyclization of N-terminal glutamine, oxidation of methionine, hydroxylated proline.

Post-Search Data Process via Scaffold

The Mascot search results were imported to the Scaffold software. It further validates the search results using a different algorithm. Scaffold also provides a comprehensive comparison of proteins and peptides between samples in table and graph views. We have manually verified all the fragmentation spectra of the identified peptides of Ara h 2. Those bad fragmentation spectra were excluded even though they were regarded as confident identifications by the software. A rough quantitation (spectrum counting) of Ara h 2 in all the fractions of the raw and roasted peanuts was performed via Scaffold. It used the normalized number of assigned MS/MS spectra to measure the quantity of the corresponding protein. The probability of proteins was set at 99% while the peptide probability was set at 95%.

Results and Discussion

1-D and 2-D Gel Electrophoresis

As shown in Figure 4.1 (left), the raw peanuts have many well defined bands while the dark roasted peanuts have vertical smearing through the whole lane. This is consistent with the 2-D gel of the roasted peanuts on which more horizontal and vertical streaking (Figure 4.2 (c)) was observed compared to that of the raw peanuts (Figure 4.2 (b)). Many protein spots have lost their well defined round shapes in the 2-D gel of dark roasted peanuts compared. The background of the roasted peanut gel is also darker. This is understandable because roasting process may cause complicated chemical modification or degradations of proteins. It is also possible that

crosslinking reactions occur between proteins or between proteins and other non-protein nutrients. All these factors can affect the moving of proteins inside a gel.

In Figure 4.1 (left), Ara h 2 was localized as two well defined bands that correspond to the heavy and light isoforms in the blue rectangle in the 1-D gel of the raw peanuts. However, in the same molecular weight region in the roasted peanut gel, the two bands of Ara h 2 were smeared and hard to define. We have run the 2-D gel of the purified Ara h 2 (sample from TNO company) as shown in Figure 4.2 (a). It is consistent with our previous reports that Ara h 2 has mainly four isoforms, P1-P4. Based on their pI and molecular weights, P1-P4 were indicated in the 2-D gels of raw and dark roasted peanuts, respectively (Figure 4.2 (b) and (c)). Even the doublets of Ara h 2 in the 1-D gel of roasted peanuts can not be differentiated, the four spots corresponding to the main isoforms of Ara h 2 in the 2-D gel of roasted peanuts are well defined. However, the sizes of the spots are smaller compared to those in the 2-D gel of raw peanuts which may indicate the loss of native Ara h 2 due to the roasting process.

Western Blotting

In order to discover potential changes of Ara h 2 as a result of roasting, Western blotting of 1-D gel using rabbit serum that contains anti-Ara h 2 polyclonal IgG was performed as described in the experimental section. Immunological detection (Figure 4.1 (right)) of the 1-D gel of raw peanut extract gave well defined bands at ~20kDa and an additional faint band at ~66 kDa of Ara h 2, even though the latter signal was very weak. Based on the molecular weight, this signal corresponds to Ara h 1 protein

which appeared as a single band at ~60 kDa in the left picture in Figure 4.1. This was confirmed by studying the peptide mixtures from in-gel trypsin digestion of this band (data not shown). The latter signal at ~20 kDa corresponds to Ara h 2 proteins. The molecular weights of the four Ara h 2 isoforms range from 16~19 kDa. The bands appeared to be wide which agrees well with the mass distribution of the Ara h 2 isoforms. For the dark roasted peanuts, Western blotting showed smearing >50 kDa. There is no signal at ~20 kDa in the gel of dark roasted peanuts. Because the antibody was raised against the Ara h 2 from raw peanuts, it may not be entirely specific for the detection of Ara h 2 from the dark roasted peanuts.

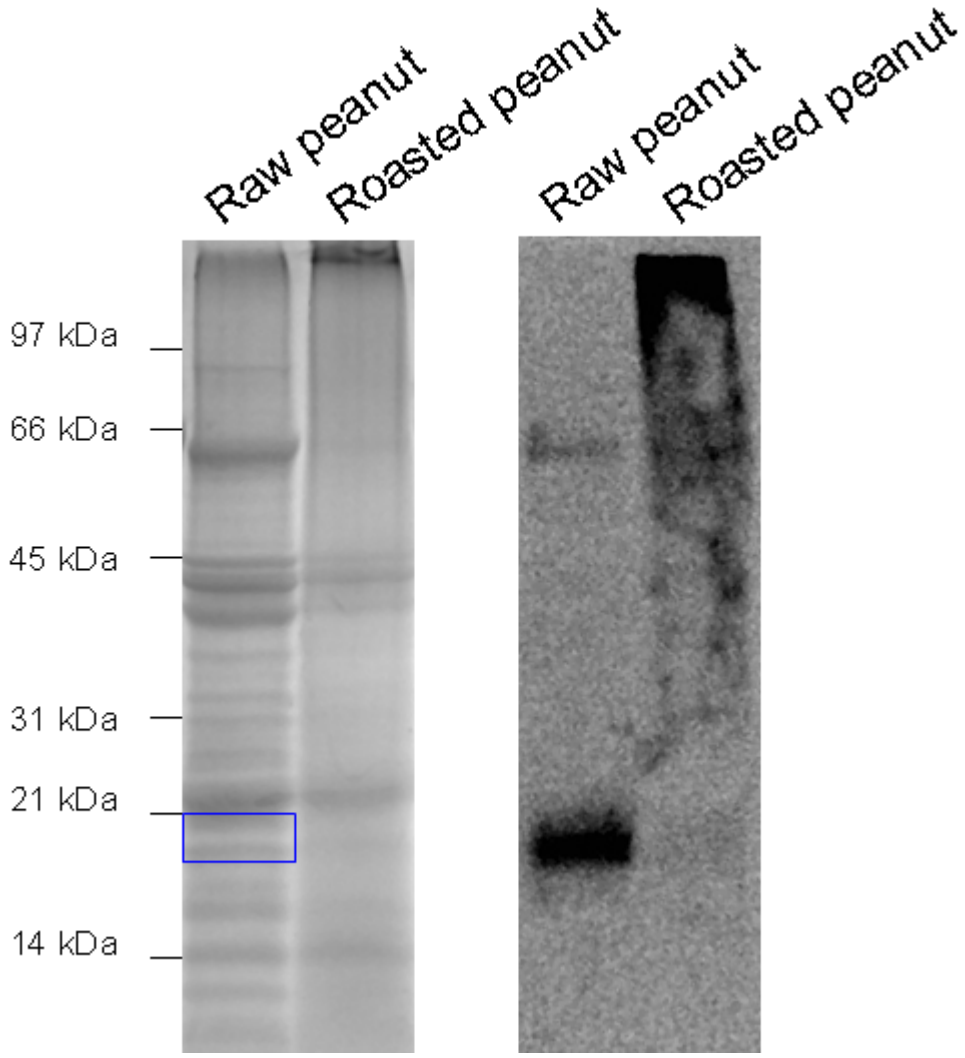


Figure 4.1: 1-D gel of raw and dark roasted peanuts (left) and Western blotting result (right) using anti-Ara h 2 antibodies.

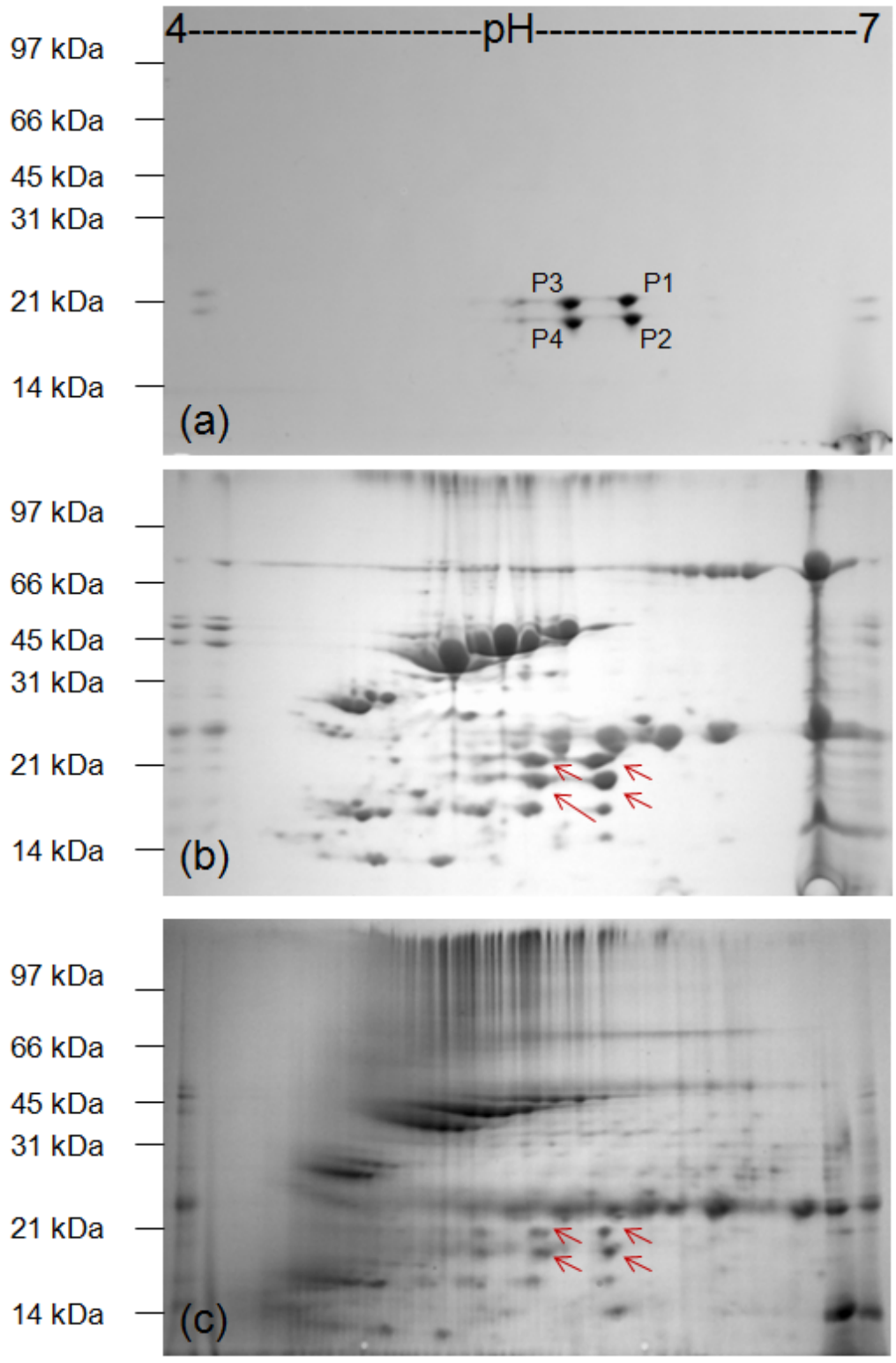


Figure 4.2: 2-D gels of (a) purified Ara h 2 from TNO company; (b) raw peanuts; (c) dark roasted peanuts.

LC-MS/MS Analysis

By studying the peptides from the gel fractions of the raw and roasted peanuts, we identified Ara h 2 peptides in every fraction (Figure 4.3 and Table 4.2) of the raw and dark roasted peanuts in different quantities. Table 4.1 and 4.2 show the peptides from Ara h 2 that were identified via Mascot database search against the customized peanut database. We have BLAST (basic local alignment search tool) searched the identified Ara h 2 peptides against NCBI database (http://www.ncbi.nlm.nih.gov/). The BLAST results show that these peptides are present in five other peanut proteins that only differ from Ara h 2 by one or two amino acid. These Ara h 2 like proteins, however, are all deduced from nucleotides.

Based on our previous work, there are mainly four isoforms of Ara h 2. Due to the high similarity in protein sequences among these isoforms, most peptides that were identified via database search are shared peptides among these isoforms. Several unique peptides were also identified that are associated with the light isoforms of Ara h 2 (P2 and P4). They are “CDLDVESGGR” that is in fraction 2 of the raw peanuts and fraction 3 of the dark roasted peanuts and “DEDSYERDPYSPSQDPYSPSPYDR” that is in fraction 1 of the raw peanuts and fraction 2, 3, 4, and 5 of the dark roasted peanuts. Even though this indicates the presence of light isoforms in those samples, there is no evidence that the heavy isoforms are missing in those samples because the identified shared peptides might be generated from them.

A relative quantitation was performed via spectrum counting of Scaffold software (Figure 4.4). It uses the number of assigned fragmentation spectra of the Ara

h 2 peptides to measure the quantity of Ara h 2. We need to point out that this is a rough quantitation estimation that has a low dynamic range. However, spectrum counting can help us understand the distribution of Ara h 2 in different mass ranges. The relative amounts of Ara h 2 were plotted across all the samples of the raw and roasted peanuts. The monomer of Ara h 2 weighs 16-18 kDa which falls in fraction 4 in both the raw and dark roasted peanuts.

Among the five samples of the raw peanuts, fraction 4 contains the most Ara h 2 peptides followed by fraction 1, 2, 5, and 3. Ara h 2 monomers are supposed to be in fraction 4. It has not been reported before that Ara h 2 peptides are present in a high molecular weight form in the raw peanuts in fraction 1 (>100 kDa) and 2 (100-50 kDa). However, it is possible that other proteins contain the same peptides as Ara h 2 does. Due to possible natural or proteolytic degradation of Ara h 2 during the storage of the peanut kernels or flours, it is not surprising that a small amount of Ara h 2 is present in fraction 5 that contains proteins with molecular masses less than 16 kDa. Fraction 3 only contains one identified peptide with a single assigned MS/MS spectrum, which indicates the lowest amount of Ara h 2 among the five fractions. This identification in fraction 3 is of low reliability.

The distribution of Ara h 2 peptides in the five samples of the dark roasted peanuts is different from that of the raw peanuts. The most were not found in fraction 4 as in the raw peanuts. Instead, fraction 2 contains the most Ara h 2 peptides. The quantity in fraction 4 is a little lower than fraction 2 followed by fraction 5, 1, and 3, which suggests decreased Ara h 2 monomers in the roasted peanuts. It is noticeable that the amount of Ara h 2 peptides in fraction 5 of the roasted peanuts is much more

than that in fraction 5 of the raw peanuts. This indicates that roasting may increase the degradation of native Ara h 2, so the small degraded protein pieces migrated into the low molecular weight region in the gel. The number of Ara h 2 peptides identified in fraction 1 (>100 kDa) mainly stays the same between the raw and roasted peanuts. However, the quantities of Ara h 2 in fraction 2 and 3 of the roasted peanuts are more than the corresponding fractions in the raw peanuts. It is possible that the native Ara h 2s that are supposed to be in fraction 4 cross-reacts with other proteins and forms protein aggregates that weigh over 50 kDa due to the roasting process. As mentioned in the previous chapter, Ara h 2 contains eight cysteines that are involved in intramolecular disulfide bonds in the raw peanuts. Due to the heat in the roasting process, these disulfide bonds might rearrange and form intermolecular linkages with other proteins. This hypothesis can not be ruled out in our gel based studies since the protein mixtures were reduced before they were loaded to the gel. The possibility that roasting process may cause degradation and aggregation of Ara h 2 at the same time would also explain the decrease of Ara h 2 monomers in fraction 4 of the roasted peanuts compared to that of the raw peanuts.

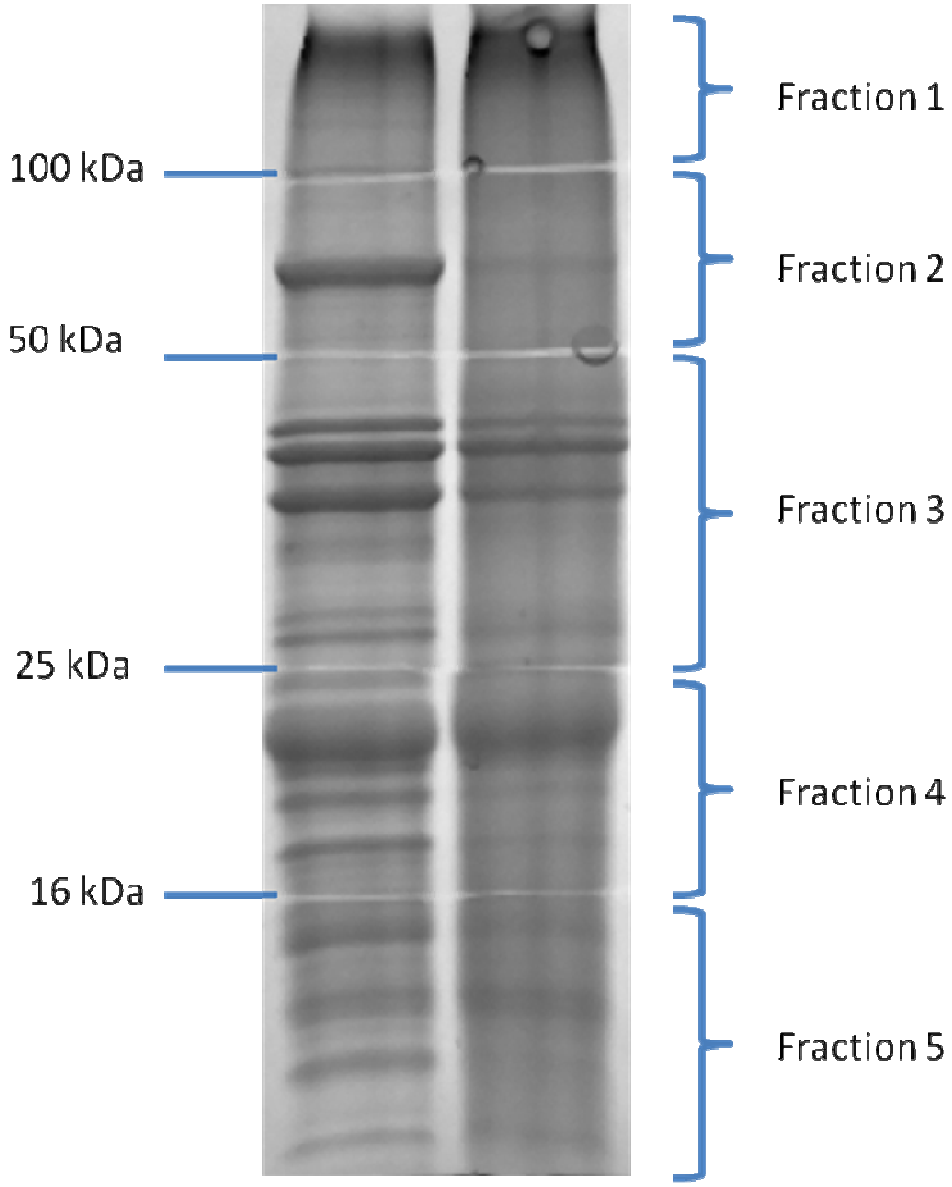


Figure 4.3: The 1-D gels of the raw and dark roasted peanuts were cut into five pieces from top to bottom. Their molecular weight ranges are >100, 100-50, 50-25, 25-16, <16 kDa, respectively.

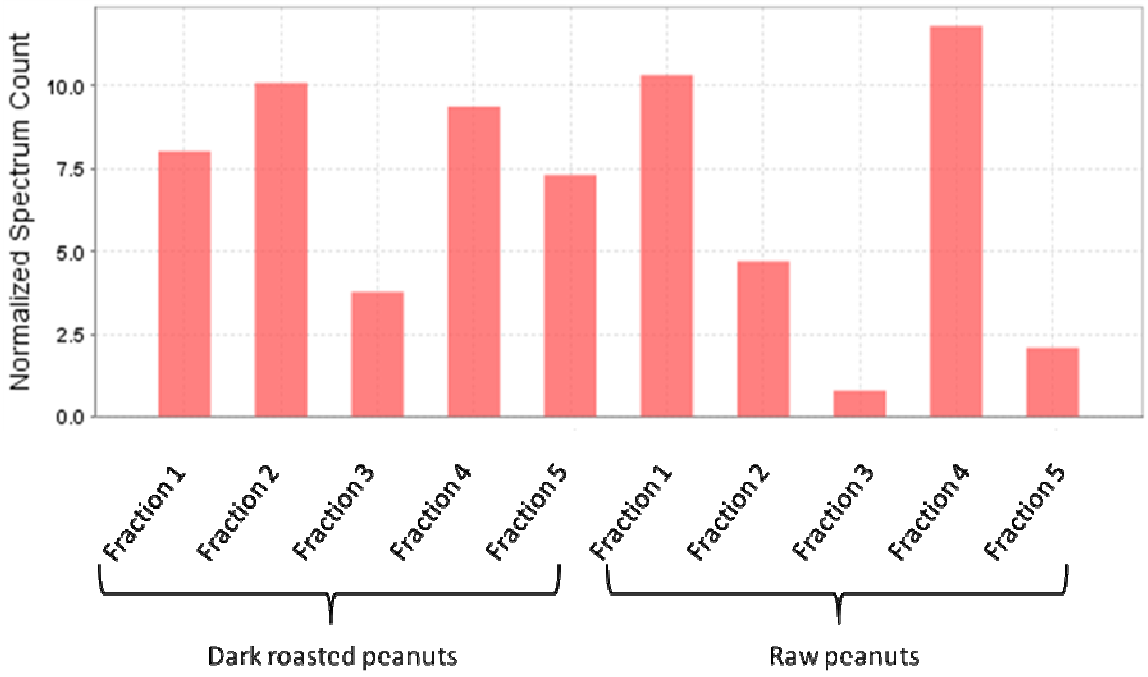


Figure 4.4: Relative quantitation of Ara h 2 in each fraction of the raw and dark roasted peanuts.

Table 4.1: Identified Ara h 2 tryptic peptides and the number of MS/MS spectra that are assigned to each peptide in gel fractions of the raw peanuts.

Sample	Fraction	Peptide sequence	Number of identified spectra
Raw peanuts	1	CCNELNEFENNQR	2
		CMCEALQQIMENQSDR	2
		DEDSYERDPYSPSQDPYSPSPYDR	2
		NLPQQCGLR	3
	2	CCNELNEFENNQR	1
		CDLDVESGGR	1
		NLPQQCGLR	1
	3	NLPQQCGLR	1
	4	CCNELNEFENNQR	2
		CMCEALQQIMENQSDR	2
		CMCEALQQIMENQSDRLQGR	1
		NLPQQCGLR	4
		QQEQQFK	1
	5	CCNELNEFENNQR	1
		CMCEALQQIMENQSDR	1

Table 4.2: Identified Ara h 2 tryptic peptides and the number of MS/MS spectra that are assigned to each peptide in gel fractions of the dark roasted peanuts.

Sample	Fraction	Peptide sequence	Number of identified spectra
Dark roasted peanuts	1	CMCEALQQIMENQSDR	6
		NLPQQCGLR	3
	2	CCNELNEFENNQR	1
		CMCEALQQIMENQSDR	4
		DEDSYERDPYSPSQDPYSPSPYDR	3
		NLPQQCGLR	3
	3	CDLDVESGGR	2
		DEDSYERDPYSPSQDPYSPSPYDR	1
		NLPQQCGLR	2
	4	CCNELNEFENNQR	2
		CMCEALQQIMENQSDR	3
		DEDSYERDPYSPSQDPYSPSPYDR	1
		NLPQQCGLR	3
	5	CCNELNEFENNQR	1
		CMCEALQQIMENQSDR	3
		CMCEALQQIMENQSDRLQGR	1
		DEDSYERDPYSPSQDPYSPSPYDR	1
		NLPQQCGLR	1

Chapter 5: Conclusions

Four isoforms of Ara h 2 have been characterized, respectively. They are named P1 to P4 based on their eluting sequences in the separation via anion exchange chromatography. The protein sequences of P1 and P2 agree with the reported cDNAs while P3 and P4 only differ from P1 and P2 by two amino acids at the C-terminus, respectively. The determination of the primary structures and PTMs of the Ara h 2 isoforms may lay the foundation for analytical method design for peanut allergen detection in foods. It may also help people to select potential biomarker peptides from the raw peanuts.

Three hydroxyproline sites were discovered in P1 and P3 while two hydroxyproline sites were found in P2 and P4. The hydroxyprolines are located at a known epitope. However, this epitope was determined using a synthesized peptide without the prolines hydroxylated. Our results may motivate researchers to map the epitopes again with the hydroxyproline containing peptides because these modifications might have an effect on the allergenicity of this allergen protein.

All eight cysteines in Ara h 2 are involved in four disulfide linkages. Two of the disulfide linkages have been determined while the other two have been partially characterized. Our results show that the disulfide bond patterns are the same among these four isoforms.

Our immunological studies suggest that Ara h 2 may aggregate or cross-linked with other proteins in the dark roasted peanuts. However, the type of crosslinking, the

crosslinking sites, and the profound modifications caused by thermal processing have yet to be defined.

We have demonstrated that mass spectrometry based methods can be applied successfully to the characterization of food allergen proteins. The application of mass spectrometry in food proteomics is still at the starting stage. However, mass spectrometry has quickly become an important method for understanding the impact of the food composition on human health. It is unbiased and highly sensitive compared to the immunological methods that mainly rely on antibodies. We believe that mass spectrometry will be more and more widely used as an analytical tool for the identification, detection, quantitation, and structural studies of food proteins.

Bibliography

(1) Sampson, H. A. *J. Am. Med. Assoc.* **1997**, *278*, 1888-1894.

(2) Koppelman, S. J.; Vlooswijk, R. A.; Knippels, L. M.; Hessing, M.; Knol, E. F.; van Reijsen, F. C.; Bruijnzeel-Koomen, C. A. *Allergy* **2001**, *56*, 132-137.

(3) Carr, W. W. *Allergy Asthma Proc* **2005**, *26*, 145-147.

(4) Goodwin, P. R. *J. AOAC Int.* **2004**, *87*, 1383-1390.

(5) Whitaker, T. B.; Williams, K. M.; Trucksess, M. W.; Slate, A. B. *J. AOAC Int.* **2005**, *88*, 161-174.

(6) Chatel, J. M.; Bernard, H.; Orson, F. M. *Int Arch Allergy Immunol* **2003**, *131*, 14-18.

(7) Yan YS, L. X., Zhang YS, Wang L, Wu K, Huang SZ *Plant Science* **2005**, *169*, 439-445.

(8) Viquez, O. M.; Summer, C. G.; Dodo, H. W. *J Allergy Clin Immunol* **2001**, *107*, 713-717.

(9) Careri, M.; Costa, A.; Elviri, L.; Lagos, J. B.; Mangia, A.; Terenghi, M.; Cereti, A.; Garoffo, L. P. *Anal Bioanal Chem* **2007**, *389*, 1901-1907.

(10) Schmidt, H.; Gelhaus, C.; Latendorf, T.; Nebendahl, M.; Petersen, A.; Krause, S.; Leippe, M.; Becker, W. M.; Janssen, O. *Proteomics* **2009**, *9*, 3507-3521.

(11) Burks, A. W.; Williams, L. W.; Connaughton, C.; Cockrell, G.; O'Brien, T. J.; Helm, R. M. *J Allergy Clin Immunol* **1992**, *90*, 962-969.

(12) Koichi Tanaka, H. W., Yutaka Ido, Satoshi Akita, Yoshikazu Yoshida and Tamio Yoshida *Rapid Commun Mass Spectrom* **1988**, *2*, 151-153.

(13) Karas, M.; Hillenkamp, F. *Anal Chem* **1988**, *60*, 2299-2301.

(14) Fenselau, C. *Anal Chem* **1997**, *69*, 661A-665A.

(15) Bornsen, K. O. *Methods Mol Biol* **2000**, *146*, 387-404.

(16) Yamshita, M. F., J. B. *J Chem Phys* **1984**, *88*, 4451-4459.

(17) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64-71.

(18) Kebarle, P.; Verkerk, U. H. *Mass Spectrom Rev* **2009**, *28*, 898-917.

(19) Biemann, K. *Biochem Soc Trans* **1989**, *17*, 237-243.

(20) Biemann, K. *Int J Mass Spec Ion Proc* **1982**, *45*, 183-194.

(21) Johnson, R. S.; Martin, S. A.; Biemann, K.; Stults, J. T.; Watson, J. T. *Analytical Chemistry* **1987**, *59*, 2621-2625.

(22) Chi, A.; Huttenhower, C.; Geer, L. Y.; Coon, J. J.; Syka, J. E.; Bai, D. L.; Shabanowitz, J.; Burke, D. J.; Troyanskaya, O. G.; Hunt, D. F. *Proc Natl Acad Sci U S A* **2007**, *104*, 2193-2198.

(23) Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E.; Shabanowitz, J.; Hunt, D. F. *Biochim Biophys Acta* **2006**, *1764*, 1811-1822.

(24) Bunger, M. K.; Cargile, B. J.; Ngunjiri, A.; Bundy, J. L.; Stephenson, J. L., Jr. *Anal Chem* **2008**, *80*, 1459-1467.

(25) Stephenson, J. L.; McLuckey, S. A.; Reid, G. E.; Wells, J. M.; Bundy, J. L. *Curr Opin Biotechnol* **2002**, *13*, 57-64.

(26) Swaney, D. L.; McAlister, G. C.; Wirtala, M.; Schwartz, J. C.; Syka, J. E.; Coon, J. J. *Anal Chem* **2007**, *79*, 477-485.

(27) Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. *Anal Chem* **2000**, *72*, 563-573.
(28) McAlister, G. C.; Phanstiel, D.; Good, D. M.; Berggren, W. T.; Coon, J. J. *Anal Chem* **2007**, *79*, 3525-3534.
(29) Macek, B.; Waanders, L. F.; Olsen, J. V.; Mann, M. *Mol Cell Proteomics* **2006**, *5*, 949-958.
(30) Shi, S. D.; Hemling, M. E.; Carr, S. A.; Horn, D. M.; Lindh, I.; McLafferty, F. W. *Anal Chem* **2001**, *73*, 19-22.
(31) Hogan, J. M.; Pitteri, S. J.; Chrisman, P. A.; McLuckey, S. A. *J Proteome Res* **2005**, *4*, 628-632.
(32) Wu, S. L.; Huhmer, A. F.; Hao, Z.; Karger, B. L. *J Proteome Res* **2007**, *6*, 4230-4244.
(33) Good, D. M.; Wirtala, M.; McAlister, G. C.; Coon, J. J. *Mol Cell Proteomics* **2007**, *6*, 1942-1951.
(34) Bordas-Nagy, J.; Despeyroux, D.; Jennings, K. R. *Journal of the American Society for Mass Spectrometry* **1992**, *3*, 502-514.
(35) Hunt, D. F.; Yates, J. R., 3rd; Shabanowitz, J.; Winston, S.; Hauer, C. R. *Proc Natl Acad Sci U S A* **1986**, *83*, 6233-6237.
(36) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. *Proc Natl Acad Sci U S A* **2000**, *97*, 10313-10317.
(37) Shen, Y.; Tolic, N.; Hixson, K. K.; Purvine, S. O.; Anderson, G. A.; Smith, R. D. *Anal Chem* **2008**, *80*, 7742-7754.
(38) Tannu, N.; Hemby, S., 2007; Vol. 8, pp 270.
(39) Aitken, A. In *The Proteomics Protocols Handbook*, 2005, pp 355-365.
(40) Fenyo, D.; Qin, J.; Chait, B. T. *Electrophoresis* **1998**, *19*, 998-1005.
(41) Naka, R.; Kamoda, S.; Ishizuka, A.; Kinoshita, M.; Kakehi, K. *Journal of Proteome Research* **2005**, *5*, 88-97.
(42) Kotch, F. W.; Guzei, I. A.; Raines, R. T. *J Am Chem Soc* **2008**, *130*, 2952-2953.
(43) Meng, F.; Forbes, A. J.; Miller, L. M.; Kelleher, N. L. *Mass Spectrom Rev* **2005**, *24*, 126-134.
(44) Ornstein, L. *Ann N Y Acad Sci* **1964**, *121*, 321-349.
(45) Shapiro, A. L.; Vinuela, E.; Maizel, J. V., Jr. *Biochem Biophys Res Commun* **1967**, *28*, 815-820.
(46) Laemmli, U. K. *Nature* **1970**, *227*, 680-685.
(47) O'Farrell, P. H. *J Biol Chem* **1975**, *250*, 4007-4021.
(48) Merril, C. R. *Methods Enzymol* **1990**, *182*, 477-488.
(49) Burnette, W. N. *Anal Biochem* **1981**, *112*, 195-203.
(50) Towbin, H.; Staehelin, T.; Gordon, J. *Biotechnology* **1992**, *24*, 145-149.
(51) Renart, J.; Reiser, J.; Stark, G. R. *Proc Natl Acad Sci U S A* **1979**, *76*, 3116-3120.
(52) Hayashi, T.; Niiya, K.; Sakuragawa, N.; Ichihara, K.; Yamazaki, T.; Watanabe, A. *Rinsho Ketsueki* **1991**, *32*, 945-950.
(53) Leitner, A.; Lindner, W. *Methods Mol Biol* **2009**, *519*, 83-101.

(54) Ferranti, B. P.; Pizzano, R.; Garro, G.; Caira, S.; Chianese, L.; Addeo, F. *J Dairy Res* **2001**, *68*, 35-51.
(55) Natale, M.; Bisson, C.; Monti, G.; Peltran, A.; Garoffo, L. P.; Valentini, S.; Fabris, C.; Bertino, E.; Coscia, A.; Conti, A. *Mol Nutr Food Res* **2004**, *48*, 363-369.
(56) Patil, U. G.; Chavan, J. K.; Kadam, S. S.; Salunkhe, D. K. *Plant Foods for Human Nutrition (Formerly Qualitas Plantarum)* **1993**, *43*, 157-162.
(57) Poms, R. E.; Capelletti, C.; Anklam, E. *Mol Nutr Food Res* **2004**, *48*, 459-464.
(58) Kopper, R. A.; Odum, N. J.; Sen, M.; Helm, R. M.; Stanley, J. S.; Burks, A. W. *Int Arch Allergy Immunol* **2005**, *136*, 16-22.
(59) Allen, L. H. *J Nutr* **2008**, *138*, 1763S-1765S.
(60) Kenney, A. C. In *Practical Protein Chromatography*, 1992, pp 249-258.
(61) Qi, Y.; Yan, Z.; Huang, J. *J Biochem Biophys Methods* **2001**, *49*, 263-273.
(62) Guzzetta, A., 2001.
(63) Annesley, T. M. *Clinical Chemistry* **2003**, *49*, 1041–1044.
(64) Caddick, S. *Tetrahedron* **1995**, *51*, 10403-10432.
(65) Juan, H. F.; Chang, S. C.; Huang, H. C.; Chen, S. T. *Proteomics* **2005**, *5*, 840-842.
(66) Pramanik, B. N.; Mirza, U. A.; Ing, Y. H.; Liu, Y. H.; Bartner, P. L.; Weber, P. C.; Bose, A. K. *Protein Sci* **2002**, *11*, 2676-2687.
(67) Swatkoski, S.; Gutierrez, P.; Wynne, C.; Petrov, A.; Dinman, J. D.; Edwards, N.; Fenselau, C. *J Proteome Res* **2008**, *7*, 579-586.
(68) Li, A.; Sowder, R. C.; Henderson, L. E.; Moore, S. P.; Garfinkel, D. J.; Fisher, R. J. *Anal Chem* **2001**, *73*, 5395-5402.
(69) Swatkoski, S.; Russell, S. C.; Edwards, N.; Fenselau, C. *Anal Chem* **2006**, *78*, 181-188.
(70) Swatkoski, S.; Russell, S.; Edwards, N.; Fenselau, C. *Anal Chem* **2007**, *79*, 654-658.
(71) Swatkoski, S.; Gutierrez, P.; Ginter, J.; Petrov, A.; Dinman, J. D.; Edwards, N.; Fenselau, C. *J Proteome Res* **2007**, *6*, 4525-4527.
(72) Maleki, S. J.; Viquez, O.; Jacks, T.; Dodo, H.; Champagne, E. T.; Chung, S. Y.; Landry, S. J. *J Allergy Clin Immunol* **2003**, *112*, 190-195.
(73) An, H. J.; Peavy, T. R.; Hedrick, J. L.; Lebrilla, C. B. *Anal Chem* **2003**, *75*, 5628-5637.
(74) Pitchayawasin, S.; Isobe, M. *Biosci Biotechnol Biochem* **2004**, *68*, 1424-1433.
(75) Roquemore, E. P.; Dell, A.; Morris, H. R.; Panico, M.; Reason, A. J.; Savoy, L. A.; Wistow, G. J.; Zigler, J. S., Jr.; Earles, B. J.; Hart, G. W. *J Biol Chem* **1992**, *267*, 555-563.
(76) Gerken, T. A.; Gupta, R.; Jentoft, N. *Biochem* **1992**, *31*, 639-648.
(77) Lamport, D. T. A. *Nature* **1967**, *216*, 1322-1324.
(78) Schmitt, D. A.; Nesbit, J. B.; Hurlburt, B. K.; Cheng, H.; Maleki, S. J. *J Agric Food Chem*, **2010**, *58*, 1138-1143.