Investigation of Immunogenic Gluten Peptides: Identification Using Enzymatic Tagging and HPLC-MS^n; Analysis and Quantification Using HPLC-MS/MS

Jennifer A. Sealey-Voyksner

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry.

Chapel Hill
2009

Approved by:

Advisor: Professor James W. Jorgenson
Professor Gary L. Glish
Professor Mark H. Schoenfisch
Professor Gary J. Pielak
Professor Susan T. Lord

ABSTRACT

Jennifer A. Sealey-Voyksner: Investigation of Immunogenic Gluten Peptides: Identification Using Enzymatic Tagging and HPLC-MS^n; Analysis and Quantification Using HPLC-MS/MS
(Under the direction of Dr. James W. Jorgenson)

The goal of my research was to provide some insight into a widely appreciated but poorly understood relationship between cereal grain proteins and human health. My research objectives were: (1) to identify and characterize inflammatory, physiologically relevant, wheat gluten peptides and (2) to develop a unique analytical methodology to screen commercially available food and consumer products for the quantitative detection of these peptides.

Gluten proteins comprise a very large protein family found in cereal grain seeds. This large protein family consists of hundreds of proteins ranging in size from about 30 kDa into the millions of KDa. Today's nomenclature refers to "gluten" as the water-insoluble seed storage proteins found in the Triticeae tribe of the grass (Gramineae) family that includes wheat, rye and barley. Some gluten proteins associated with grains in the Triticeae tribe (specifically: wheat, rye and barley), have been implicated in various autoimmune diseases, food allergies, intolerances and are important factors in several inflammatory diseases.

Many analytical techniques have been used to study gluten proteins and peptides. Unambiguous identification and structural characterization of such

peptides is a necessary step toward an eventual understanding of their chemical biology. Liquid chromatography-mass spectrometry (LC-MS) is an extremely powerful tool for such analyses.

Results of my research are presented here, in the following chapters of this thesis report. Data presented supports the development of a novel and effective analytical methodology, using enzymatic/chemical labeling chemistry and HPLC/MS^n; to identify and characterize seven physiologically relevant wheat gluten peptides. A sensitive and specific assay was then developed for the quantitative detection of these peptides via direct *in-vitro* proteolytic digestion and HPLC-MS/MS. This versatile methodology allows both processed and native foods, as well as consumer products, to be analyzed for the presence of wheat gluten. Continued efforts in this area will pave the way for eventual commercial application, as a service to both the celiac community and the food industry, by providing an accurate and economic means to generate much needed data for researchers developing treatments for patients with gluten sensitivities and manufacturers producing and labeling products that are safe.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. James Jorgenson, for his support and encouragement. I would also like to thank Dr. Chaitan Khosla and the Celiac Sprue Research Foundation for their guidance as well as assistance with experiments, which were invaluable to the success of this research project.

My husband, Robert, deserves volumes to be written about him. He has supported, inspired, mentored, taught as well as patiently and unconditionally loved me throughout this academic adventure. From the first moment that I asked his advice on the notion of going back to school to achieve a Ph.D, to introducing me to my future advisor, Dr. James Jorgenson; to helping me understand electronics; to teaching me how to better interpret mass spectra and mass spectrometers; to drying the tears of frustration; to brainstorming about the experimental design and interpretation of the data at 3am and on the road to the beach and in the air to yet another meeting and even over a glass of wine on our much treasured date nights; to never being too busy to review manuscripts, posters and presentations; to constantly reminding me to back up my data; to making me re-re-re-work the concentration calculations; to tirelessly trying to make me believe in myself; Robert has and always will be my knight: I love him and will be eternally grateful to him.

My sincere gratitude goes out to many of my friends, especially my dear friend Roger Mercer. Roger first inspired me as a young scientist, then truly supported, mentored and encouraged me to pursue this daunting personal goal.

Lastly, I appreciate the love and support of my parents and especially that of my brother, who has given me an unending supply of R.A.I.S.E. and ehugs.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Å	angstrom
AA	amino acid
Ac	acetyl group (CH_3CO-R)
ACN	acetonitrile
AOAC	Association of Analytical Communities
API	atmospheric pressure ionization
APC	antigen presenting cell
AU	absorbance units
BBM	brush border membrane enzyme
BPI	base peak index chromatogram
˚C	degrees Celsius
C_{18}	octadecyl silane group ($C_{18}H_{36}$) particle chemistry
$CaCl_2$	calcium chloride
CD	celiac (or coeliac) disease
CID	collision-induced dissociation
CS	celiac sprue
D	detected (value is between the LOD and the LOQ)
Da	Dalton
DPP IV	dipeptidyl peptidase IV enzyme
DTT	dithiothreitol
EDTA	ethylenediaminetetracetic acid
EIC	extracted ion chromatogram

ELISA	enzyme-linked immunosorbent assay
ESI	electrospray ionization
FAO-UN	Food and Agriculture Organization of the United Nations
FDA	Food and Drug Administration
FIA	flow injection analysis
GDE	general digestive enzyme
GEs	gluten exorphins
GS	gluten sub-unit
HCl	hydrochloric acid
HLA	human leukocyte antigen
HMW	high molecular weight
HPLC	high performance liquid chromatography
IFN $_{\gamma}$	interferon gamma
IL-15	interleukin-15
IETLs	intraepithelial lymphocytes
ITMS	ion trap mass spectrometry
kDa	kiloDalton
$KHNaPO_4$	potassium phosphate
L	liter
LC	liquid chromatography
LC-MS	liquid chromatography-mass spectrometry
LC-MS/MS	liquid chromatography-tandem mass spectrometry
LC/UV	liquid chromatography-ultraviolet

LMW	low molecular weight
LOD	limit of detection
LOQ	limit of quantification
M	moles/liter (molar)
MALDI-TOF	matrix assisted liquid dissociation ionization – time of flight
mAU	milliabsorbance unit (10^{-3} AU)
MDC	mono dansyl cadaverine
mer	number of amino acid residues in a peptide sequence (i. e. 11mer)
mg	milligram (10^{-3} gram)
MHC	major histocompatability complex
MIC-A	major histocompatability complex class I chain related A antigen
min	minute
mL	milliliter (10^{-3} liter)
mm	millimeter (10^{-3} meter)
mM	millimolar (10^{-3} molar)
mol	mole
ms	millisecond (10^{-3} second)
MOPS	3-(N-morpholino) propane-sulphonic acid
MRM	multiple reaction monitoring
MS	mass spectrometry
MS^n	mass spectrometry to the n stages
MS/MS	mass spectrometry/mass spectrometry (tandem mass spectrometry)
MW	molecular weight

m/z	mass/charge (mass to charge ratio)
µg	microgram (10^{-6} gram)
µL	microliter (10^{-6} liter)
µm	micromolar (10^{-6} molar)
µM	micrometer (10^{-6} meter or micron)
NaCl	sodium chloride
NaOH	sodium hydroxide
$NaHPO_4$	sodium phosphate
ND	not detected (value is below the LOD)
ng	nanogram (10^{-9} gram)
NKG2D	natural-killer-cell marker
nL	nanoliter (10^{-9} liter)
nm	nanometer (10^{-9} meter)
OAc	acetate (CH_3COO-R)
PBS	phosphate buffered saline
ppb	part per billion
ppm	part per million
PCR	polymerase chain reaction
PDA	photodiode array detector
PEP	prolyl endopeptidase
pg	picogram (10^{-12} gram)
PTC	pepsin / trypsin / chymotrypsin
PTCECA	pepsin / trypsin / chymotrypsin / elastase / carboxypeptidase A

QC	quality control
QQQ	triple quadrupole
R^2	correlation coefficient
RhC	tetramethyl rhodamine cadaverine
RP	reversed phase
RSD	relative standard deviation
RT	retention time
s	second
SD	standard deviation
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
T-cell	thymus-cell
TFA	trifluoroacetic acid
TIC	total ion chromatogram
TG2	type 2 tissue transglutaminase enzyme
tgases	tissue transglutaminase enzymes
Tris-HCL	tris hydroxymethylaminoethane
UV	ultraviolet
V	volt
WHO	World Health Organization

LIST OF SYMBOLS

α	alpha
β	beta
γ	gamma
λ_{abs}	absorbance wavelength
λ_{em}	emission wavelength
ω	omega
A	alanine
C	cysteine
E	glutamic acid (or glutamate)
F	phenylalanine
G	glycine
H	histidine
I	isoleucine
K	lysine
L	leucine
m	mass
M	methionine
N	asparagine
P	proline
Q	glutamine
Q→E	conversion of glutamine to glutamic acid
R	arginine

S	serine
t	time
T	threonine
V	valine
W	tryptophan
Y	tyrosine
z	integer charge
<	less than
>	greater than
%	percent
+/-	plus or minus

CHAPTER 1

Introduction to Gluten and Celiac Disease

1.1 What is gluten?

"Gluten" is a term that collectively refers to an enormous family of complex heterogeneous storage proteins found in the endosperm of cereal grain seeds. As their description suggests, these proteins have no apparent function other than the storing of nitrogen, sulphur and carbon in the developing endosperm in order to support the needs of the growing seedling [1]. Storage proteins make up approximately 50% of the total proteins in mature cereal grains. Although cereal grains only contain about 10% protein, they are a most important crop, ultimately providing much of the daily nutritional requirement of protein needed by humans and livestock worldwide. Apart from their nutritional value, the functional properties of cereal seed proteins permit them to serve an important role in food manufacturing and processing [2, 3].

Gluten is considered to be one of the most complex families of proteins found in nature. There are hundreds of proteins in this family, ranging in size from about 30 kDa into the millions of kDa [1, 3]. Since the late1700's, this sub-family of cereal grain proteins has been studied extensively and several theories have been successfully applied to further subdivide and classify the proteins in this family into

smaller sub-groups [4 - 6] according to their characteristics, amino acid sequences, locations within the grain, solubility and relationship within other grains. One of the first classification systems was developed by Osborne, who, because of his pioneering research, went on to earn the distinction of being the “father of plant protein chemistry”, in the mid 1850s. He divided gluten proteins into four groups, based on their solubility in various solvents (Figure 1.1). The albumins are soluble in water, globulins in dilute salt solutions, prolamins in alcohol and glutelins in dilute acid or base solutions [6]. Amazingly, this early classification scheme is still used today; and over the last hundred and fifty years, the nomenclature has evolved to today’s definition of “gluten”, which is generally used to describe the water-insoluble seed storage proteins found in the Triticeae tribe of the grass (Gramineae) family. Figure 1.2 delineates the taxonomy of some of the common dietary cereal grains found in the grass family. The relationship between four subfamilies, selected tribes and common names for the grains are shown in this figure [7].

A different classification scheme, developed by Shewry and Tatham [1, 3], devised a way to describe some gluten proteins of the Triticeae tribe (wheat, barley and rye). Here, the prolamins and glutelins could be grouped together, based on their amino acid composition according to the following three classifications: (1) sulphur-poor, (2) sulphur-rich and (3) high molecular weight (HMW) proteins. Such a scheme, representative of the prolamins and glutelins of wheat (i. e. gliadins and glutenins), is shown in Figure 1.3. The basis of the sulphur-containing fractions was the ability to group the proteins according to the presence of the amino acids cysteine and methionine.

A sticky, rubbery ball of “gluten” is what would remain if one washed a sample of wheat, barley or rye flour in the kitchen sink. The compositions of the hundreds of proteins that make up this sticky gluten dough ball determine the unique combination of strength and elasticity needed to ensure a properly baked food product. Although all of these proteins differ in amino acid composition, they are all characterized by a high % of proline (P) and glutamine (G) residues and a small % of methionine (M) and cysteine (C) residues. Gluten proteins from all grains are both monomeric (prolamins) and polymeric (glutelins) in structure. Cysteines are responsible for linking the monomeric protein units together via intermolecular disulphide bonds. These polymeric aggregated molecules, with molecular weights that can span into the millions of KDa, hold important functionalities in the quality of dough during the baking process [8].

In terms of solubility, the majority of gluten proteins can be generally classified into two solubility-based subgroups: (1) the aqueous alcohol-soluble prolamins and (2) the alcohol-insoluble glutelins (i.e. “gliadins” and “glutenins” - in wheat). The monomeric gliadins can be further sub-divided into α-, ß-, γ- and ω-gliadins; whereas the polymeric glutelins can be further sub-divided into low molecular weight (LMW) and high molecular weight (HMW) subunits. In the wheat kernel, the monomeric gliadins and polymeric glutenins are found in the relative ratio 65:35 and form the majority of the storage protein in wheat. Similar groups of proteins are found in barley and rye, corn and oat grains. Extensive research involving the functionality of wheat proteins in relation to baking quality, has determined that it is the gliadin proteins that are responsible for viscosity and

extensibility, while the glutenins determine elasticity. Interestingly, the functionalities of the gliadins and glutenins their relationship to each other can be thought of as somewhat analogous to a multi-component epoxy glue system [8].

In recent years, there have been many excellent and informative studies conducted, which have focused on characterizing the various classes of gluten proteins [9 - 11]. As more information has become available about the composition and structure of the many types of proteins, new insights into their functionality have become clearer. Although complete amino acid sequences have been determined for representatives from the prolamin and glutelin sub-groups of wheat, barley and rye, gaps still exist in the knowledge base for the structure, composition and functionality of all members of this enormous protein family.

1.1.1 Wheat gluten protein nomenclature and composition

Wheat gliadins are monomers and most have been found to be alcohol soluble. Early attempts to study the gliadins led to their classification into four sub-groups (α-, ß-, γ- and ω-), based on their mobility in gel electrophoresis. Further investigations have re-grouped the α- and ß-gliadins together because they are structurally very similar. The α/ß- and γ-gliadins are lower in molecular weight (30 to 60 kDA) than the ω-gliadins (up to 75 kDa). They contain repetitive sequences of 11 (for α/ß) and 7 (for γ) amino acids in the N-terminal domain and have homologous, non-repetitive C-terminal domains along with the presence of 6 - 8 cysteines which form intra-chain cross-links. Other domains are present in these proteins that are made up of long repetitive sequences of individual glutamines and sequences high in proline and glutamine. Repetitive sequences 8-10 amino acids, consisting of high

percentages of proline and glutamine make up the majority of the ω-gliadins. These proteins are the minor component of the gliadins. They contain the highest % of proline and glutamine and lowest % of sulphur containing amino acids. Secondary structural information about α-, ß-, γ- and ω-gliadins describe the N-terminal domain for each as consisting mainly of ß-turns; whereas the C-terminal domain is characterized as consisting of ß-sheets and α-helices.

Wheat glutenins are polymeric; they are aggregates of the alcohol-soluble monomeric gliadins linked together by inter-chain disulphide bonds. As such, glutenins are considered to be insoluble in aqueous alcohol. However, should their inter-chain disulphide bonds be reduced, the resulting monomeric subunits do become soluble in aqueous alcohol The low molecular weight gluten subunits (LMW-GS) constitute about 20% of the total wheat gluten protein, while the high molecular weight gluten subunits (HMW-GS) fraction represents about 10%. The LMW-GS have domains similar in amino acid composition to α-, ß- and γ-gliadins. The N-terminal domain is characterized by short repetitive sequences high in proline and glutamine, while the C-terminal domain has cysteines, which form the inter-chain disulphide bonds. The HMW wheat glutenins consist of 3-5 subunits, each of which contains 3 structural domains. The N- and C- terminal domains are both non-repetitive and contain charged residues and cysteines. The middle domain contains repetitive sequences of 4-6 amino acids. Limited information is available about the secondary structure of the HMW-GS, but what is known is that the middle domain consists of ß-reverse turns which are predicted to overlap, thus forming a spiral. The

N- and C- terminal domains are thought to contain α-helices that contribute to these domains' globular structures.

An important network of covalent bonds consisting of inter- and intra-chain disulphide bonds between and within gluten proteins, are a common structural feature of this vast and heterogeneous family of proteins. In addition, non-covalent bonds, such as hydrogen, ionic and hydrophobic bonds are also prevalent structural features. These types of bonds are responsible for the various functionalities of gluten proteins in manufacturing processes. It has been determined that it is the HMW-GS proteins that are the most influential in determining the properties of dough. One especially large wheat polymer, the "glutenin macropolymer", is well known on account of its substantial contribution to the strength and volume of baked bread [3, 8, 12 and 13].

Prolamin and glutelin proteins from wheat, barley and rye are all basically similar in structure, but do have considerable differences in their detailed structures. Comparisons of the various domains from all three groups of proteins indicate that they have a common evolutionary origin. It has been speculated that because select regions of Triticeae proteins are also found in other groups of proteins (i. e. oats, rice, sunflower, corn and castor bean), that perhaps these other proteins shared a limited evolutionary link with those of the Triticeae tribe [3, 13].

1.2 Link between gluten and human disease

Dietary gluten, from wheat, barley and rye grains has been identified as a principle trigger of a variety of immune diseases, including food allergies and intolerances [14, 15 and 16]. Some of the more common are listed as follows:

anemia, rheumatoid arthritis, carcinoma of the oropharynx, esophagus and small bowel, celiac disease; dermatitis herpetiformis, diabetes (Type 1), Down syndrome; enteropathy-associated T-cell lymphoma; IBS - irritable bowel syndrome, kidney disease, liver disease, Sjogrens syndrome, thyroid disease (autoimmune), ulcerative jejunoileitis etc. [17]. Gluten intolerance has recently been considered a factor associated with various neurological symptoms, such as, depression, migraine, headaches and learning disorders, such as, autistic spectrum of disorders and attention deficit disorders [16].

1.3 Celiac disease

Celiac (or Coeliac) disease (CD), also known as Celiac Sprue (CS), is a complex autoimmune disorder of the small intestine. Interaction between several factors, including gluten, can result in an autoimmune response which is best characterized by small-intestinal injury with severe structural damage to the villi lining of the small intestine, adversely affecting one's ability to absorb nutrients [17, 18]. Results from the inadequate absorption of nutrients from digested food in the alimentary canal, especially by the small intestine, can manifest themselves in a wide spectrum of serious problems involving various bodily systems, including the nervous system, the heart, teeth and bones. Current knowledge about the pathogeneses of CD implicates environmental, genetic and immunological factors. In those who are genetically predisposed, exposure to gluten proteins, from wheat, barley and rye, can trigger both adaptive and innate immune responses (Figure 1.4). Although all cereal grains contain prolamin and glutelin proteins, the amino acid sequences of these types of proteins in each grain are different. Only those from

wheat, barley and rye have been identified as the trigger of the immune response that affects the intestinal lining of those genetically susceptible to CD. Symptoms of CD vary greatly between patients. Symptoms may be severe, limited or even absent. Prolonged exposure to even modest quantities of dietary gluten results in severe damage to the small intestine and may ultimately result in death if left untreated [19].

Current epidemiological research reports that CD affects ~1% of the world's population [20]. Due to the wide variation in clinical manifestations of the disease, many people affected by CD have trouble receiving proper diagnosis; many remain undiagnosed. Reports have indicated that CD is now becoming widely recognized in North and South America, Europe, Africa, Australia and India. Prevalence of CD in Asia has arisen recently, due to the incorporation of wheat into the diet of the more affluent Asian citizens.

1.3.1 The genetic factor

CD will not develop in an individual unless they have inherited the necessary genetic factors that are part of the immunological response to gluten. Two of the most important genetic factors have been identified as human leukocyte antigen (HLA) DQ2 and DQ8 genes. People carrying these genes have been found to carry the highest risk factor for developing CD [21].

The major histocompatability complex (MHC) is a large gene region found on the short arm of chromosome six (Figure 1.5). Genes in this family have an important role in the immune system of vertebrates. The MHC is divided into regions: I, II and III. HLA genes are a large (4Mb) part of the MHC class II family.

The HLA region contains numerous immune and non-immune system related genes. The HLA-DQ locus is where known CD-related genes are located.

Because the involvement of these HLA genes in the susceptibility to CD has been so clearly proven, HLA-typing has been established as a successful diagnostic tool in instances where it is necessary to exclude the disease as a potential cause of sickness. Although carrying one or both of these HLA-DQ2 or DQ8 genes is necessary for the development of CD, they are not the only genetic factors involved. Approximately 30% of American citizens carry these genes, but only 1% will develop the disease. Other genetic factors must also be considered, such as those that influence the innate and adaptive immune system and physical condition of the intestinal mucosal barrier [22].

1.3.2 The gluten factor

The later stages of digestion of dietary foods in the alimentary tract occur within the villi lining on the outside surface of small intestinal cells. The purpose of these fingerlike structures (villi) is to increase (by several hundred times), the surface area of the intestinal cells, in order to maximize the ability of the cells to absorb nutrients. By the time food reaches the villi in the small intestine, it has been metabolized down to a mixture of the basic fundamental units of disaccharides, single amino acids, di- and tri-peptides, fatty acids, and monoglycerides. These components are either efficiently absorbed by the villi and passed on, or eliminated. Villus capillaries collect amino acids and simple sugars and pass them on into the blood stream; whereas villus lacteals collect triglycerides, cholesterol and amphipathic proteins and direct them to the rest of the body through the lymphatic

fluid. Figure 1.6 depicts the alimentary tract and the location of the villi which are damaged by the immune reaction to dietary gluten.

Gluten proteins contain an unusually high percentage of the amino acids proline and glutamine. The substituted and conformationally constrained amide bond of proline residues renders them considerably more resistant towards normal breakdown by proteolytic gastric, pancreatic and brush border membrane digestive processes [23, 24]. Over time, an increase in concentration of relatively stable small peptides (~4-50 amino acids in length) results in the small intestine. The immune response to these incompletely metabolized gluten peptides promotes an inflammatory reaction in the small intestine. This response is characteristically mediated by two different immune mechanisms, the adaptive and the innate mechanism [20]. This response is currently, the best characterized toxicological effect of gluten in patients with CD is this intestinal inflammation and enteropathy. Other toxicological effects of exposure to gluten have also been described [15, 25 and 26]. Of note, a small group of four and five amino acid residue peptides have been identified that result from the enzymatic digestion of dietary wheat gluten. Termed the “gluten exorphins (GEs)”, this family of opioid-like peptides, closely resembles enkephalins in structure and has been found to be able to penetrate abnormally permeable intestinal membranes (a condition analogous to that which occurs in people with CD). Once allowed to enter the blood plasma and finally reach the central nervous system, these GEs have been found to affect behavior by modulating neurotransmission functions.

1.3.3 The immunological factor

Under normal physiological conditions, passage of larger molecules across the intestinal epithelial barrier is prevented by intercellular tight junctions that keep the cells close together. In those individuals who are susceptible to CD, however, interaction between physiologically relevant undigested immunogenic gluten peptides and intestinal cells appears to trigger disassembly of these tight junctions and allows the intestinal wall to become permeable [27, 28]. The mechanism responsible for the regulation of the permeability of these intracellular tight junctions is not well understood. However, research in this area has identified a protein (zonulin) whose function appears to be to somehow mediate the regulation of intracellular permeability [29]. Zonulin has been found to be overexpressed in those with CD [30] and recent work has finally been able to characterize its biochemical nature [31].

Once the intestinal epithelial layer has become permeable to larger molecules (i. e. gluten peptides), these immunogenic gluten peptides proceed diffuse into the lamina propia of the intestinal villi and are allowed to trigger two distinct immunological pathways (innate and adaptive), resulting in the characteristic tissue damage and villous atrophy. These symptoms are those most commonly presented by the majority of those afflicted with CD.

The immune system generally acts to protect its host organism by using a layered defense of increasing specificity. The first layer of defense is simply a physical barrier to prevent pathogens (agents that cause disease or illness to its host) from entering the host organism. Should a pathogen break though this first

layer of defense, the innate immune system mediates an immediate, but non-specific response to the pathogen. If the pathogen escapes the innate response, the adaptive immune system adapts the organism's response accordingly, in order to remove the pathogen [32].

The body can act through both of these two mechanisms, in response to the presence of gluten peptides. Although only a small number of gluten peptides that can activate the immune system in this fashion have been identified to date, the list is increasing on an ongoing basis. A list of selected known immunogenic gluten epitope sequences is shown in Table 1.1 [33, 34]. The core 9-mer region of the gluten peptide-HLA binding register is listed, along with the associated HLA element that has been identified that binds to it. It is important to note that peptide sequences from both gliadin and glutenin proteins have been identified as immunogenic.

1.3.3.1 Adaptive immune response to gluten

Under certain conditions, peptides belonging to one group of structurally similar gluten peptides are allowed to enter the abnormally permeable layer of the villi's epithelium. A brief description of the adaptive immune response to these gluten peptides involves a complex which is formed between dendritic cells (antigen presenting cells [APCs] - in the intestines), the HLA DQ2/8 proteins and gluten peptides. This complex is then presented to and subsequently binds to T-helper cells. The T-cells then becomes activated and direct the release of various pro-inflammatory cytokines and anti-bodies resulting in eventual tissue damage.

1.3.3.2 Innate immune response to gluten

Another group of structurally similar gluten peptides, different than those involved in the T-cell mediated adaptive response, can activate an innate immune response in the intestinal epithelium. The mechanism of this innate immune response appears to be quite complicated and much less understood.

An overview to this mechanism likens its progression to a kind of stress response elicited by the epithelium. In this instance, undigested gluten peptides interact with the epithelium directly. This initial cell damage results in the release of interleukin-15 (IL-15), which in turn leads to the production of interferon gamma (IFN $_{\gamma}$). Intraepithelial lymphocytes (IETLs) become activated and become cytotoxic, expressing NKG2D (a natural-killer-cell marker) receptors. Mucosal damage results as cells that express the major-histocompatability-complex class I chain-related A (MIC-A) antigen on their surface are then targeted, bind to the NKG2D receptor and are then killed. MIC-A is a protein that is produced due to some kind of stress [35].

1.3.4 Other environmental factors

Exposure to gluten is considered to be the most important environmental factor involved with the pathogenesis of CD. However, according to recent research findings, there are several other environmental factors that could contribute to increasing the risk for developing CD [20]. The accessibility of breast milk appears to offer some protection to infants who are genetically predisposed to CD, either by protecting them completely from developing the disease, or in delaying its onset [36, 37]. It is known that children who are not breastfed or who are given gluten too early in life are more at risk to develop the disease [38].

Another interesting factor that has provided evidence for increasing the risk for the onset of CD is the type and rate of incidence of certain gastrointestinal infections that occur at a young age [39].

1.4 Motivation for measuring gluten

Although information about the pathogenesis of CD is increasing, there is currently no cure for the disease, no representative animal model from which to study the immune response mechanisms and little hope of changing the immunogenic nature of all Triticeae grains through bioengineering. Therefore, the only effective therapy available for people with CD is a strict, lifelong adherence to a gluten-free diet. However, maintaining a true gluten-free existence is actually quite difficult, expensive and fraught with accidental exposure to gluten, which can result in incomplete recovery and/or relapse from recovery. Several non-dietary therapeutic approaches have recently been proposed, which attempt to address occurrences of inadvertent ingestion of small amounts of gluten [40 - 42]. Notwithstanding, it is in the best interests of those who suffer with CD to provide a means of preventing accidental exposure to gluten.

The best word to describe the presence of gluten in our environment is to say that it is ubiquitous. It is used extensively, throughout the world in a variety of manufacturing arenas [43]. In food manufacturing, it is used as a flavor enhancer, thickener, fortification ingredient, filler, whitener etc. Gluten is also used in the manufacturing of personal care products, nutritional supplements and drug products. Fortunately, for those with CD or other forms of gluten sensitivities, the presence of gluten in a product is usually noted by its inclusion in the ingredient list on product

packaging (albeit referenced by a variety of names). Unfortunately, there also exist many hidden sources of gluten as a consequence of unlabeled ingredients and from cross-contamination in manufacturing processes and equipment cleaning. Therefore, people with CD are destined to a life of label-reading to ensure that products they use are safe for personal use and consumption. Several groups of researchers are working to establish if there actually exists a “safe” quantity of gluten that is safe for people with gluten sensitivities to ingest on a daily basis [44, 45]. But in the meantime, a gluten content of “zero“ is desired. Unfortunately, based on how prevalent gluten is in the environment, this “zero” tolerance is not realistic.

From the perspectives of patient disease management, manufacturing quality control and worldwide consistency in product labeling, there poses an obvious need to be able to accurately determine that commercially available foods and consumer products are gluten-free and label them as such. However, the subject of product labeling and gluten-free certification is a controversial one because there is no concise definition of “gluten-free” that is currently accepted worldwide.

1.5 Methods of gluten measurement

Increased interest, awareness and research surrounding gluten-related illnesses (such as CD) over the last two decades have resulted in a wealth of new information. New data ultimately provides health care professionals the means to be able to determine proper diagnoses for patients that present with and without gluten-related symptoms. With respect to CD, there is a growing understanding of the pathogenesis of the disease, new methods for screening patients and novel

methodologies for studying gluten proteins and screening gluten-containing foods and products.

Nonetheless, the analysis, characterization and quantification of cereal grain gluten peptides and proteins are complicated. Protein composition, profile and natural sequence fluctuations of grains are in a constant state of flux, due to variances in cultivars, areas grown, climate and the emergence of genetic engineering. Gluten protein composition can also change during various processing stages of manufacturing. Several well-established analytical techniques, such as SDS-PAGE, capillary electrophoreses, PCR, RP-HPLC and mass spectrometry, have all been used quite successfully, to study cereal grain proteins [46 - 48]. Mass spectrometry has recently become involved in both genomic and proteomic areas of gluten analysis [49]. MS-based methods have focused mainly on applications involving the qualitative detection and characterization of gluten proteins, through the use of MALDI-TOF-MS [50, 51] and HPLC-MS based methods [26, 52]. Research in the area of assessing the total gluten content and detecting gluten contamination in food has been traditionally done using immunochemical methods [15, 45, 53 - 55]. In fact, the only current commercially available methods for determining the presence of gluten in foods are the immunological antibody-based ELISA methods. Thus, immunochemistry is the only analytical technique that is currently being endorsed by both the FDA and the Codex Alimentarius, for gluten detection in commercially available foods and consumer products.

1.5.1 Measurement of gluten using HPLC-MS

To date, the analytical capabilities of HPLC-MS have found limited application in the area of quantitative detection of trace levels of physiologically relevant gluten peptides in complex matrices. My research was directed toward this end, by the proposed identification of some immunostimulatory gluten peptides, then evaluating commercially available food and consumer products for their presence by quantitative detection using HPLC-MS. It should be emphasized that this work did not focus on determining the total gluten content in samples, nor the fingerprinting of samples for a gliadin / glutenin profile, nor the identification of which gluten proteins were present in samples. My work did endeavor to develop a means to comprehensively, selectively and accurately screen complex samples in order to provide evidence of the presence of trace levels of immunogenic gluten.

High performance liquid chromatography (HPLC), atmospheric pressure electrospray ionization (ESI)-ion trap mass spectrometry (ITMS) and triple quadrupole mass spectrometry (QQQ-MS) are powerful analytical techniques. Their application to the area of proteomics in order to enable peptide and protein identification, as well as precise quantitative analysis, has been well documented [56 - 59].

Hundreds of gluten peptides are generated by the *in-vivo* gastric/pancreatic enzymatic digestion of wheat gluten. In order to study these peptides, a similar digestion process could be performed in an *in-vitro* fashion. An HPLC separation of the resulting mixture of peptides would offer a first degree of specificity needed to study the components of such a complex mixture. ESI-MS detection offers a second

degree of specificity, by providing the ability to detect some specific mass-to-charge (m/z) ions that represent the molecular weights of these gluten peptides. Multiple stages of mass spectrometry (MS^n, where n=number of stages) offer even more degrees of specificity, based on the ability to detect product ions generated from the collision-induced dissociation (CID) of certain chosen parent ions. This provides the means to gather the essential information needed to determine the partial or even the complete primary amino acid sequence, thus the identity, of some gluten peptides.

There are several configurations of commercially available instruments that can offer multiple stages of mass spectrometry and are based on combinations of quadrupoles, ion traps, time of flight and magnetic/electric sector mass analyzers. Two of the most common instrument configurations for conducting MS^n experiments are the ion trap and triple quadruplole mass spectrometers. These two instrument configurations in combination with HPLC were used in this research and are briefly described as follows.

1.5.1.1 Ion trap mass spectrometer

A schematic of the Agilent Technologies ion trap mass spectrometer is shown in Figure 1.7. The HPLC effluent is nebulized and charged in the atmospheric pressure electrospray chamber, thus producing charged droplets. These droplets eject ions which are sampled through the capillary and then directed through a series of skimmers, octopole and lenses directly into the trap, while the neutral gas molecules are pumped away. The ability to generate MS and MS^n spectra using the

ion trap is based upon computer control of voltages on the ring electrode and voltages and frequencies on the two end caps of the trap.

An ideal application of ion trap mass spectrometry is the structural elucidation of unknown compounds, such as immunogenic gluten peptides. The advantage of using this instrument in this type of application stems from the ability of the trap to accumulate all the ions from the sample being analyzed, simultaneously. This enhances the duty cycle of the ions being studied, as compared to the duty cycle of a scanning mass analyzer, such as a quadrupole. This translates into good full scan sensitivity, which is necessary for identifying and studying unknowns. Another useful feature of the ion trap system is its ability to perform a data dependent acquisition. Thus, MS^n spectra can be acquired in addition to full scan spectra, therefore providing much more information which is useful for the identification of unknown compounds.

1.5.1.2 Triple quadrupole mass spectrometer

A schematic of the Agilent Technologies triple quadrupole mass spectrometer is shown in Figure 1.8. The process of generating ions in the electrospray ionization chamber of this instrument is identical to that of the ion trap system, however, the triple quadrupole system transmits the ions directly into a first stage quadrupole mass analyzer rather than a trap analyzer. In this first quadrupole, desired target ions are selected and transported into the next stage of the system, a collision cell. This hexapole collision cell breaks apart the target ions into characteristic product ions and transports them into the final quadrupole analyzer. Here, the product ions are isolated and then directed into the detector assembly.

Triple quadrupole mass spectrometry is ideally suited for the application of quantitative detection of target compounds. Therefore, once the identity of an immunogenic gluten peptide has been determined, this analytical system is well suited to determine its presence and concentration in a food sample.

1.6 Overview of this thesis report

The goal of this research was to provide some insight into a widely appreciated but poorly understood relationship between cereal grain proteins and human health. The research objectives were: (1) to identify and characterize potential immunogenic wheat gluten peptides and (2) to develop an analytical methodology using HPLC-MS for use in screening commercially available food and consumer products for the quantitative detection of trace quantities of these peptides. The following five chapters (Chapters 2, 3, 4, 5 and 6) describe the journey I undertook in an endeavor to successfully achieve these two objectives.

1.6.1 Chapter 2

This chapter describes initial efforts toward the identification of immunogenic gluten peptides. A procedure was developed that was designed to simulate the *in-vivo* enzymatic digestive process, which could be performed in an *in-vitro* fashion. This procedure succeeded in its attempt to release similar types of gluten peptides from proteolyzed wheat gluten proteins that are released *in-vivo*. Data presented supports the proof of the principle of this concept and provides an interesting first impression of into the challenges of the analysis of food proteins by HPLC-MS.

1.6.2 Chapter 3

Hundreds of gluten peptides are released via a proteolytic digestion. This chapter describes a strategy that was developed which attempted to discriminate between those that are immunogenic and those that are not. The strategy focused on certain aspects of how the body itself determines which peptides it considers as “antigens” and how it targets them during the autoimmune responses to gluten. Following this strategy, a methodology was developed that subsequently allowed several potentially physiologically relevant immunogenic gluten peptides to be identified in proteolyzed wheat flour. Data presented describes how these peptides were identified, how their structures were determined and how they came to be considered physiologically relevant to celiac disease.

1.6.3 Chapter 4

The ability to be able to detect and quantify trace levels of dietary gluten in commercially available food and consumer products is important, because of the established link between exposure to gluten and human health. This chapter describes how the seven physiologically relevant immunogenic peptides identified in this research could be used effectively as markers for gluten toxicity in food. An analytical method was developed for the quantitative detection of the seven selected immunogenic wheat gluten peptides in such products, using HPLC-MS/MS. Experimental results describe how this method can accurately, sensitively and reproducibly detect and quantify trace levels of the immunogenic peptides that have been identified by this research.

1.6.4 Chapter 5

This chapter describes two practical applications of this developed methodology. The first application describes how the methodology was utilized in the evaluation of the quality of commonly available consumer foods and products. The data presented demonstrate how this method can detect trace levels of immunogenic gluten in products that have been labeled gluten free, as a result of possible contamination during the manufacturing process. This is essential knowledge for patients who suffer from various forms of gluten sensitivities and need to maintain a gluten-free existence. Currently, there is no concise, universally accepted definition of “gluten-free”, with respect to the gluten content of manufactured foods and products that become available for consumption by the general population. Various agencies around the world, including the United States, have proposed their own definitions, which affect how products are manufactured and labeled and how their quality is determined. The work described in this thesis report has attempted to address this subject area by providing a means to quantitatively determine the presence of immunogenic wheat gluten in a complex food matrix.

A second practical application of the developed methodology describes how it was utilized as a tool in the assessment of the proteolytic capabilities of a new orally-based enzyme. This type of oral enzyme therapy is designed to address instances of unavoidable everyday exposure to small amounts of gluten. The function of an oral enzyme of this type is primarily to target the immunogenic gluten peptides that resist gastric and pancreatic digestion and subsequently break them up into smaller, more

digestible fragments, thus preventing the characteristic inflammatory damage to the tissues of the small intestine. The data presented demonstrate that this method has the capacity to be used as a tool in research, to further our knowledge about gluten-related diseases and to aid in the development of methods that aim to treat them.

1.6.5 Chapter 6

The last chapter of this report provides a brief synopsis of all of the research presented in the context of this project. It reviews the goals and objectives of the research and summarizes the significance of the results obtained from the work. Some insight is also offered with respect to possible future directions for the continuation of this work and its contribution to the subject area in general.

Source protein	Epitope sequence (9-mer core)	HLA restriction
α-gliadin	PQPQLPYPQ	DQ2
α-gliadin	PFPQPQLPY	DQ2
γ-gliadin	PQQSFPQQQ	DQ2
α-gliadin	QGSFQPSQQ	DQ8
HMW-glutenin	QGYYPTSPQ	DQ8
α-gliadin	PYPQPQLPY	DQ2
γ-gliadin	IQPQQPAQL	DQ2
γ-gliadin	QQPQQPYPQ	DQ8
LMW-glutenin	PFSQQQQPV	DQ2/8
γ-gliadin	PQPQQQFPQ	DQ2
α-gliadin	FRPQQPYPQ	DQ2
LMW-glutenin	PFSQQQQPV	DQ2
γ-gliadin	IQPQQPAQL	DQ2
γ-gliadin	QQPQQPYPQ	DQ2
γ-gliadin	SQPQQQFPQ	DQ2
γ-gliadin	QQPFPQQPQ	DQ2

Table 1.1

Selected known DQ2/DQ8-restricted immunogenic gluten epitopes

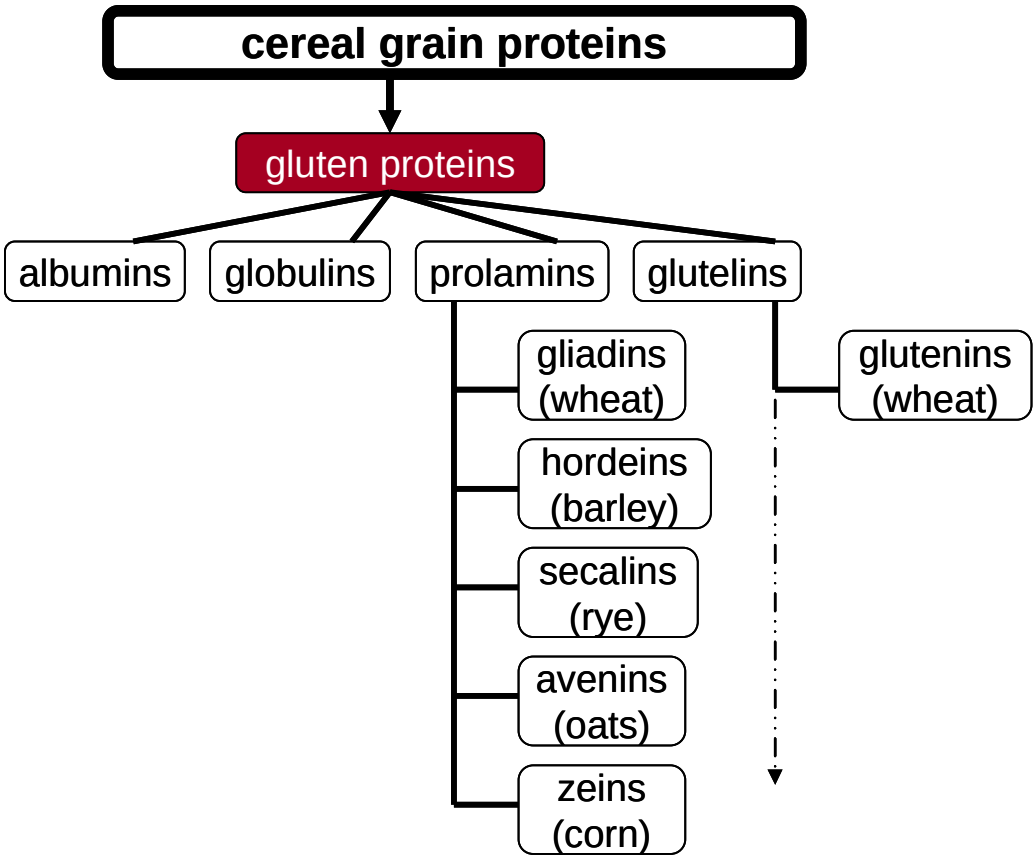


Figure 1.1

A classification of gluten proteins based on solubility

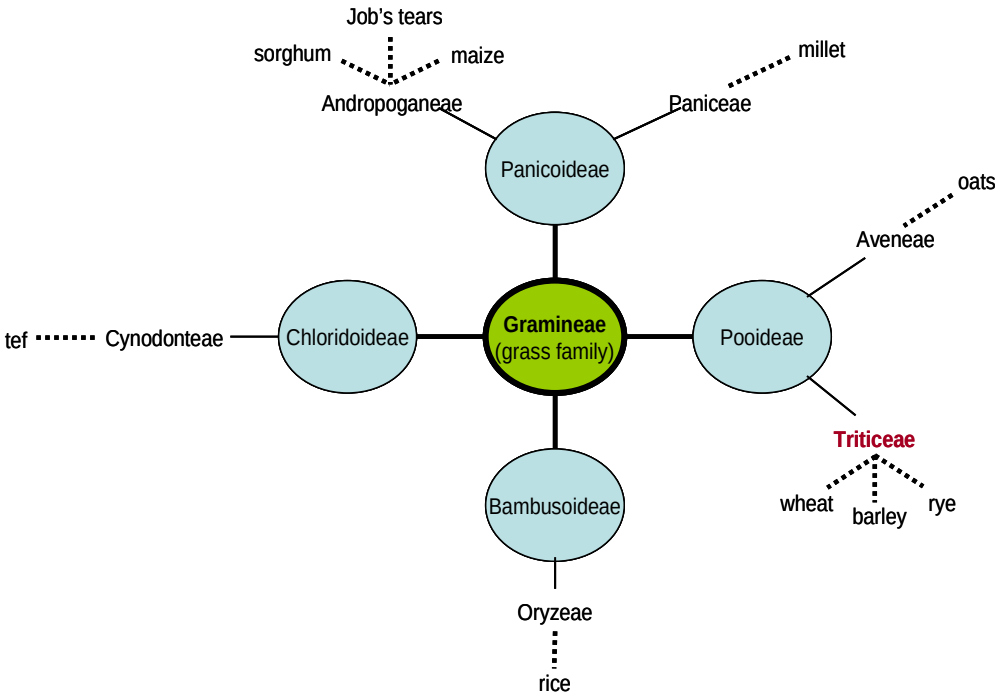


Figure 1.2

Taxonomy of some common dietary cereal grains

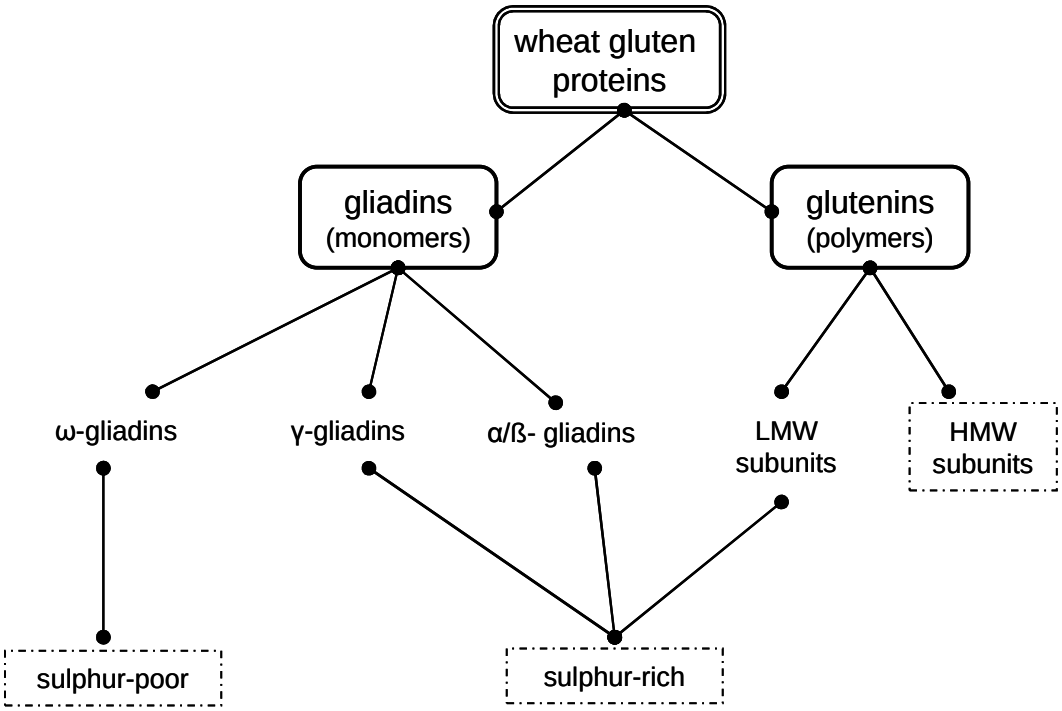


Figure 1.3

A classification of wheat gluten proteins based on amino acid composition

(adapted from Shewry [3])

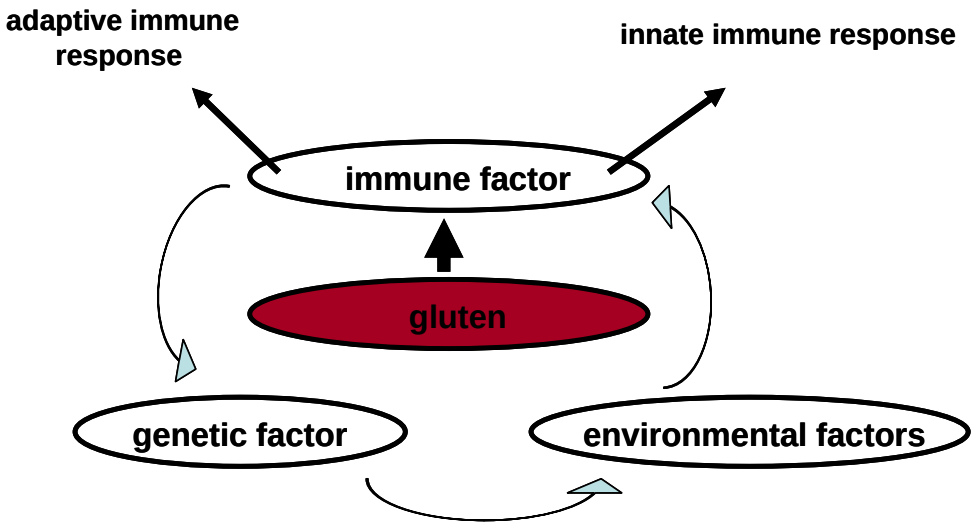


Figure 1.4

Factors involved in celiac disease (CD)

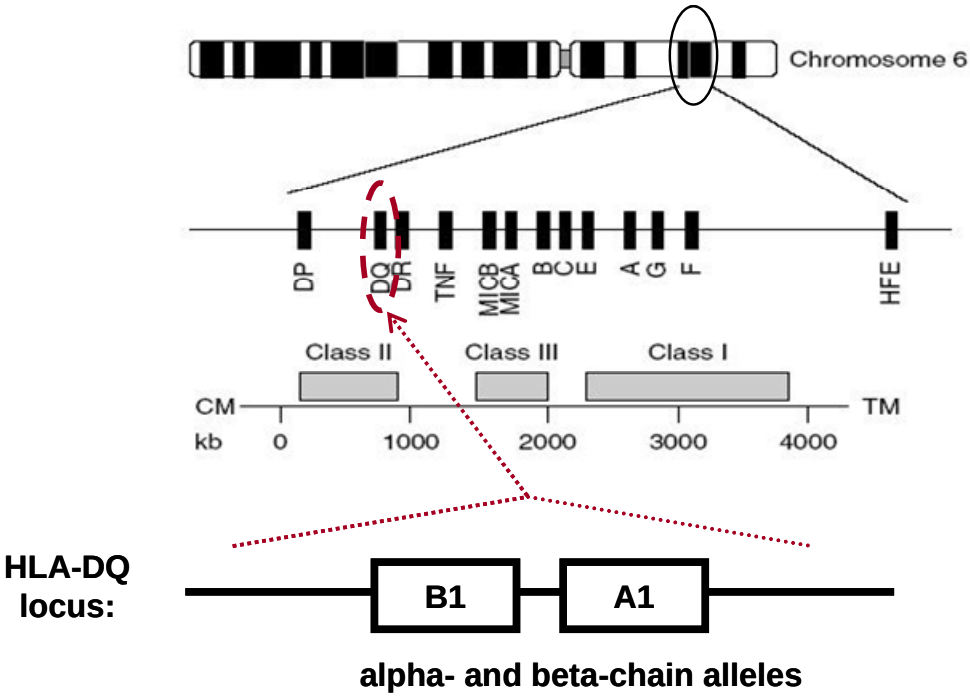


Figure 1.5

Overview of the HLA gene region of the MHC

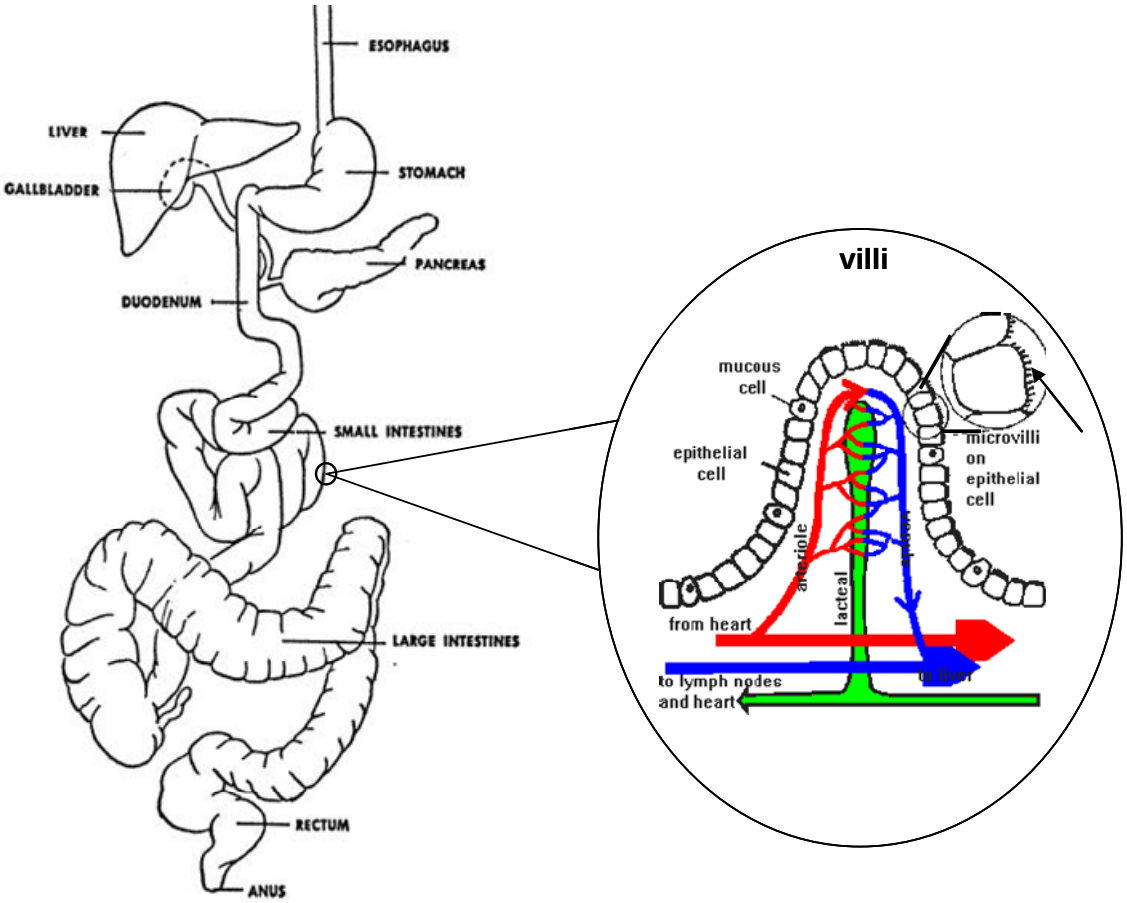


Figure 1.6

Alimentary tract

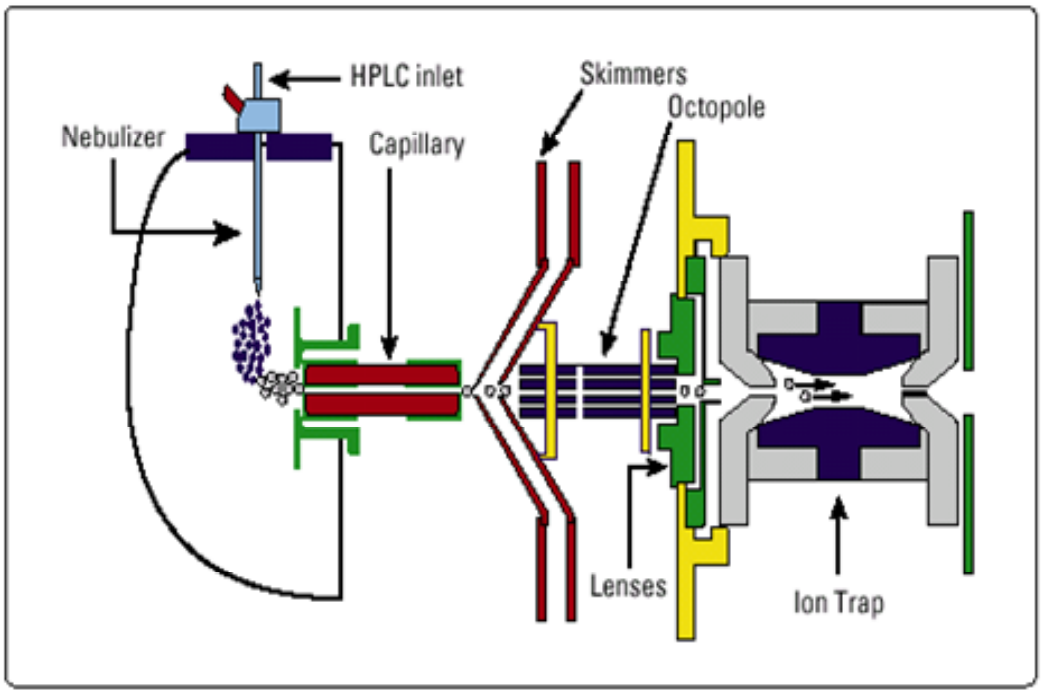


Figure 1.7

Agilent Technologies 1100 Series HPLC-MSD Ion Trap Mass Spectrometer

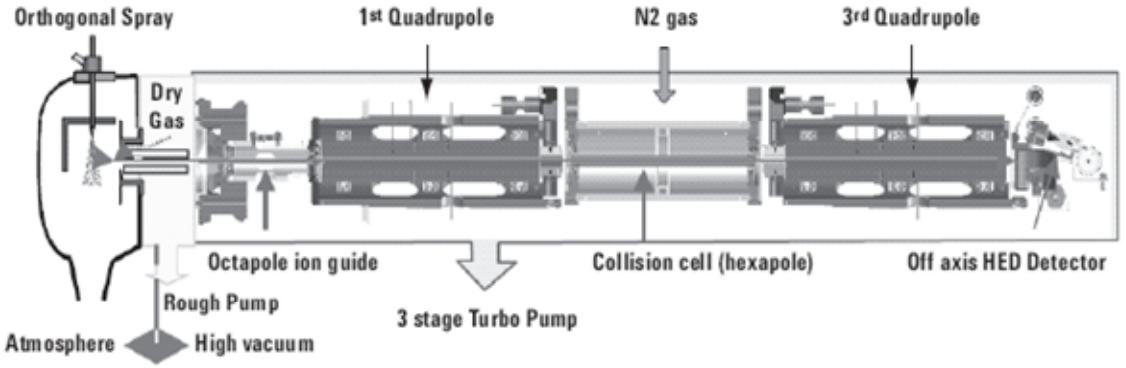


Figure 1.8
Agilent Technologies 6410 Series HPLC-Triple Quadrupole Mass Spectrometer

1.7 References

[1] Shewry, P.; Tatham, A.; Forde, J.; Kreis, M. *Journal of Cereal Science* **1986**, *4*, 97-106.

[2] Wang, J.; Zhao, M.; Zhao, Q.; Bao, Y.; Jiang, Y. *Journal of Food Science* **2007**, *72*, 203-107.

[3] Shewry, P.; Halford, N. *Journal of Experimental Botany* **2002**, *52*, 947-958.

[4] Beccari, J. *De Bononiensi Scientarium et Artium, Instututo Atque Academia Commentarii 2* **1745**, *1*, 122.

[5] Parmentier, A. *Examination Chimique des Pommes de Terre*, dans Lequel On Trait des Parties Constituents du Bli, Didot Dijeaune, Paris **1773**.

[6] Osborne, T. *The Vegetable Proteins*, 2nd edn., Longmans, Green and Co., London **1924**, 154-158.

[7] National Plant Data Center, Baton Rouge, LA 70874-4490 USA *The PLANTS Database, Version 3.5* (http://plants.usda.gov) **2005**.

[8] Weiser, H. *Food Microbiology* **2007**, *24*, 115-119.

[9] Payne, P. *Annual Revision of Plant Physiology* **1987**, *38*, 141-146.

[10] MacRitchie, F. *Advances in Food Nutrition Research* **1992**, *36*, 1-13.

[11] Ferranti, P.; Mamone, G.; Picariello, G.; Addeo, F. *Journal of Mass Spectrometry* **2007**, *42*, 1531-1536.

[12] Howdle, P. *European Journal of Gastroenterology & Hepatology* **2006,** *18*, 703-706.

[13] Shewry, P.; Tatham, A. *Journal of Biochemistry* **1990**, *267*, 1-12.

[14] Collin, P.; Kaukinen, L.; Maki, M. *Digestive Diseases* **1999**, *17*, 100-106.

[15] Shan, L.; Khosla, C. *Immunology, Endochronology and Metabolism Agents in Medicinal Chemistry* **2007**, *7*, 187-193.

[16] Ford, R. *Medical Hypotheses* **2009**, *73*, 438-440.

[17] Jabri, B.; Kasarda, D.; Green, P. *Immunological Reviews* **2005**, *206*, 219-231.

[18] Kagnoff, M. *Journal of Clinical Investigation* **2007**, *117*, 41-49.

[19] Marsh, M. *Gastroenterology* **1992**, *102*, 330-354.

[20] Green, P.; Cellier, C. *The New England Journal of Medicine* **2007**, *357*, 1731-1743.

[21] van Heel, D.; West, J. *Gut* **2006**, *55*, 1037-46.

[22] Heap, G.; van Heel, D. *Seminars in Immunology* **2009** (in press).

[23] Shan, L.; Molberg, O.; Parrot, I.; Hausch, F.; Filliz, F.; Gray, G.; Sollid, L.; Khosla, C. *Science* **2002**, *297*, 2275-2279.

[24] Hausch, F.; Shan, L.; Santiago, N. *American Journal of Physiology: Gastrointestinal & Liver Physiology* **2002**, *283*, 996-1003.

[25] Fukudome, S.; Jinsmaa, Y.; Matsukawa, T.; Sasaki, R.; Yoshikawa, M. *FEBS Letters* **1997**, *412*, 475-479.

[26] Fanciulli, G.; Azara, E.; Wood, R.; Delitala, G.; Marchetti, M. *Journal of Chromatography. B* **2007**, *852*, 485-490.

[27] Schneeberger E.; Lynch, R. *American Journal of Physiology: Cell Physiology* **2004,** *286*, 1213–1228.

[28] Fasano, A.; Catassi, C. *Best Practice & Research. Clinical Gastroenterology* **2005**, *19*, 467-478.

[29] Wang, W.; Uzzau, S.; Goldblum, S.; Fasano, A. *Journal of Cell Science* **2000**, *113*, 4435-4440.

[30] Drago, S.; El Asmar, R.; Di Pierro, M.; Grazia, C.; Tripathi, A.; Sapone, A. *Scandinavian Journal of Gastroenterology* **2006**, *41*, 408-419.

[31] Tripathi, A.; Lammers, K.; Goldblum, S.; Shea-Donohue, T.; Netzel-Arnett, S.; Buzza, M.; Antalis, T.; Vogel, S.; Zhao, A.; Yang, S.; Arrietta, M.; Meddings, J.; Fasano, A. *PNAS* **2009**, *106*, 16799-16804.

[32] *Genes and Diseases* (http://www.ncbi.nlm.nih.gov/books/).

[33] Marti, T.; Molberg, O.; Li, Q.; Gray, G.; Khosla, C.; Sollid, L. *The Journal of Pharmacology and Experimental Therapeutics* **2005**, *312*, 19-25.

[34] Qiao, S.; Sollid, L.; Blumberg, R. *Current Opinion in Immunology* **2009**, *21*, 111-117.

[35] Meresse, B.; Chen, Z.; Ciszewski, C. *Immunity* **2004**, *21*, 357-366.

[36] Persson, L.; Ivarsson, A.; Hernell, O. *Advances in Experimental Medical Biology* **2002**, *503*, 115-123.

[37] Ivarsson, A.; Hernell, O.; Stenlund, H.; Persson, L. *American Journal of Clinical Nutrition* **2002**, *75*, 914-921.

[38] Norris, J.; Barriga, K.; Hoffenberg, E. *Journal of the American Medical Association* **2005**, *293*, 2343-2351.

[39] Stene, L.; Honeyman, M.; Hoffenberg, E. *American Journal of Gastroenterology* **2006**, *101*, 2333-2340.

[40] Ehren, J.; Moron, B.; Martin, E.; Bethune, M.; Gray, G.; Khosla, C.; *PLoS ONE* **2009**, *4*, (in press: e63313).

[41] Siegel, M.; Bethune, M.; Gass, J.; Ehren, J.; Xia, J.; Johannsen, A.; Stuge, T.; Gray; G.; Lee, P.; Khosla, C. *Chemistry & Biology* **2006**, *13*, 649-658.

[42] Camarca, A.; Anderson, R.; Mamone, G.; Fierro, O.; Facchiano, A.; Gianfrani, C. *The Journal of Immunology* **2009**, *182*, 4158-4166.

[43] Wang, J.; Zhao, M.; Zhao, Q.; Z.; Bao, Y.; Jiang, M. *Journal of Food Science* **2007**, *72*, 103-109.

[44] Catassi, C.; Fabiani, E.; Iacono, G.; D'Agate, C.; Francavilla, R.; Biagi, U.Volta, F.; Accomando, S.; Picarelli, A.; DeVitis, I.; Ianelli, G.; Gesuita, R. *American Journal of Clinical Nutrition* **2007**, *85*, 160-168.

[45] Hischenhuber, C.; Crevel, R.; Jarry, B.; Makis, M.; Moneret-Vautrin, D.; Romano, A.; Troncone, R.; Ward, R. *Alimentary Pharmacological Therapeutics* **2006**, *23*, 559-563.

[46] Debnath, J.; Martin, A.; Gowda, L. *Food Research International* **2009**, *42,* 782-787.

[47] Payne, P. *Annual Review of Plant Physiology* **1987**, *38,* 141-149.

[48] MacRitchie, F. *Advances in Food Nutrition Research* **1992**, *36,* 1-10.

[49] Ferranti, P.; Mamone, G.; Picariello, G.; Addeo, F. *Journal of Mass Spectrometry* **2007**, *42*, 1531-1535.

[50] Dworschak, R.; Ens, W.; Standing, K.; Preston, R.; Marchyl, B.; Nighingale, M.; Stevenson, S.; Hatcher, D. *Journal of Mass Spectrometry* **1998**, *33*, 429-433.

[51] Amiour, N.; Merlino, M.; Leroy, P.; Branlard, G. *Proteomics* **2002**, *2*, 632-641.

[52] Mamone, G.; Ferranti, P.; Chianese, L.; Scafure, L.; Addeo, F. *Rapid Communications in Mass Spectrometry* **2000**, *14*, 897-899.

[53] Thompson, T.; Mendez, E. *Journal of the American Dietetic Association* **2008**, *108*, 1682-1687.

[54] Denery-Papini, S.; Nicolas, U.; Popineau, Y. *Journal of Cereal Science* **1999**, *30*, 121-127.

[55] Debnath, J.; Martin, A.; Gowda, L. *Food Research International* **2009**, *42*, 782-788.

[56] Tao, W.; Wu, L.; Cooks, R. *Journal of the American Society for Mass Spectrometry* **2001**, *12*, 490-476.

[57] Lin, T.; Payne, A.; Glish, G. *Journal of the American Society for Mass Spectrometry* **2001**, *12*, *497-504.*

[58] Downard, K. *Mass Spectrometry of Protein Interactions*, Wiley Interscience, Hoboken, NJ, **2007.**

[59] Matsuo, T.; Caprioli, R.; Gross, M.; Seyama, Y. *Biological Mass Spectrometry,* John Wiley & Sons, England, **1994.**

CHAPTER 2

Use of *In-Vitro* Enzymatic Digestion and Metal Ion Attachment to Selectively Study and Identify Potential Immunogenic Gluten Peptides using HPLC-MS^n

---

2.1 Introduction

During the human *in-vivo* digestion process, dietary gluten proteins from wheat, barley and rye cereal grains are broken down into very small peptides and individual amino acid residues, which are taken up into the alimentary tract or subsequently allowed to pass through the intestines. Those with celiac disease (CD) cannot do this efficiently and some small (~4-40 amino acid residue) peptides, rich in proline and glutamine, resist the normal proteolytic breakdown by gastric and duodenal enzymes and remain in the small intestine. Over time, these peptides can trigger an inflammatory response which subsequently can result in intestinal tissue damage, characteristic of celiac disease. One such wheat gluten peptide is a 33mer (LQLQPFPQPQLPYPQPQLPYPQPQLPYPQPQPF), termed "αG-33". This 33mer has recently been identified as the physiological form of immunodominant antigens, thus suggesting it is a potent trigger of the inflammatory response to gluten proteins [1] in those with CD. Amino acid sequences of other related peptides in wheat, barley and rye are theorized to be similar in size and in amino acid composition.

2.1.1 In-vitro enzymatic digestion of gluten proteins

These wheat, barley and rye gluten peptides, that have been proposed as immunogenic gluten peptides, have been found to possess the following characteristics or elements that are associated with the CD immune response: (1) contain multiple proline (P) and glutamine (Q) residues, (2) are targets for the tissue transglutaminase 2 (TG2) enzyme within the epithelial layer of the intestinal wall, (3) contain certain consensus sequences (i. e. Q-x-P or PQPQLPY) that involve specific Q residues, targeted by the TG2 enzyme and (4) originate only from wheat, barley and rye cereal grain gluten proteins [2].

In theory, it should be possible to simulate, in an *in-vitro* fashion, the approximate physiological conditions which take place in the gut using the main human digestive enzymes, in order to generate a mixture of gluten peptides similar in nature to those that are produced *in-vivo*. Store bought wheat gluten could be used as an ideal test matrix, because it contains the highest concentration of gluten proteins, relative to native wheat, barley and rye flours. HPLC-MS^n could be used to study the digested products and potentially isolate and characterize those that may be immunogenic for CD.

A strategy for immunogenic peptide identification would initially involve evaluation of the HPLC-MS data from the enzymatically proteolyzed wheat gluten to isolate those peptides with sufficient MS response so as to generate adequate MS/MS spectra. This would allow protein database searching to take place. Submission of any MS/MS spectra obtained in this manner to the MASCOT or other relevant protein database could ascertain a possible sequence of the peptide and

whether is gluten-related or not. Manual interpretation of the MS^n spectra would also need to be done to confirm any results offered by this type of database search. Once potential sequences were postulated for these peptides, confirmation of the presence of known immunogenic epitope sequences within would offer relevant evidence that the peptides are in fact immunogenic. Some known immunogenic gluten epitope sequences are listed in Table 1.1. Next, any potential target peptide sequences identified in this manner would then have to be verified, through further database searching, to confirm that they are unique to wheat and not present in any grains deemed acceptable (such as corn or rice) for consumption by people with gluten sensitivities. Therefore, this process would establish that proper physiological conditions had been met for an *in-vitro* digestion and that an appropriate procedure to identify physiologically relevant immunogenic gluten peptides was proven successful.

2.1.2 Metal ion attachment

One method that can offer improvements in the sensitivity and the quality of structural information obtained for a compound during analysis by HPLC-MS involves the formation of metal ion adducts with the compound in solution. Introducing such complexes into a mass spectrometer can offer these improvements using collision-induced decomposition (CID) during the MS/MS experiment. Metal ion adducts have been successfully used to achieve both of these improvements in performance with compounds, such as carbohydrates [3] nucleotides [4] and phospholipids [5]. There has been some research published which involved specific metal attachment to proline [6], the differentiation of isomeric di-peptides through the

formation of copper complexes [7] and C-terminal peptide sequencing of alkali cationized peptides up to 10 residues in length [8]. Also, it has been shown that the formation of 5 member rings by low energy CID of peptides (using ion trap mass spectrometry) can offer opportunities for specific metal ion attachment and unique CID product ions [9]. The choice of the correct metal ion (i. e. size) can serve to bridge these ring structures and offer unique peptide fragmentation information.

Therefore, it was theorized that this technique could be applied successfully in order to achieve improved specificity in gluten peptide analysis, which would ultimately aid in the identification and characterization of immunogenic peptides. Since it has been determined that immunogenic gluten peptides contain high concentrations of proline and glutamine, the proposed experiment would involve the infusion of solutions of a variety of metal acetate salts and selected synthetic peptides into an ion trap mass spectrometer and monitor the MS responses. Evaluation of the data from this type of experiment would determine if there are any metals that show a unique affinity for proline-containing peptides. Further experiments using on-line HPLC-ion trap MS/MS, with metal ion addition conducted via post column addition would provide a means to obtain structural information of gluten peptides.

2.2 Experimental

2.2.1 Materials and reagents

2.2.1.1 Expression and purification of recombinant prolyl endopeptidase

Prolyl endopeptidase (PEP) was expressed from Flavobacterium meningosepticum was expressed by fed-batch fermentation. Colleagues at Stanford

University, CA, USA have succeeded in isolating the PEP at high specific activity and purity and have provided the methodology under a standard Materials Transfer Agreement to the Celiac Sprue Research Foundation, CA, USA. Dr. Chaitan Khosla, of Stanford University, provided the PEP for use with this research project.

2.2.1.2 Isolation of brush border membrane enzymes

Colleagues at Stanford University, CA, USA isolated and purified brush border membrane enzymes (aminopeptidase N and dipeptidyl peptidase IV) from Sprague-Dawley rat brush border membrane cells. In brief, a section of rat jejunum was surgically removed and the mucosa scraped off. The mucosa was homogenized and processed to a final resulting pellet which was re-suspended in PBS at a pH 7.1.

2.2.1.3 Chemicals and samples

Pepsin, trypsin, chymotrypsin, carboxypeptidase A, elastase, $KHNaPO_4$, NaOH and various metal acetates were obtained from Sigma-Aldrich (St. Louis, MO, USA). HPLC-grade acetonitrile and HCl were obtained from Fischer Scientific. Water was obtained from an in-house Milli-Q water purification system (Millipore, Billerica, MA, USA). Wheat gluten was obtained from a local specialty food store (Chapel Hill, NC, USA). Synthetic gluten peptides were supplied by Dr. Chaitan Khosla, Chemistry Department, Stanford University, CA, USA.

2.2.2 Enzymatic digestion of wheat gluten flour

2.2.2.1 Treatment with gastric and pancreatic enzymes

Wheat gluten flour was treated with pepsin, trypsin, chymotrypsin, elastase and carboxypeptidase A (PTCECA) according to the following protocol: 300 mg of

wheat gluten was added to 10 mL of water with the addition of HCl to achieve a stable suspension at pH 2. Pepsin (Pepsin NF powder, 1:10000, 6.0 mg) was mixed into the gluten suspension and the gluten-pepsin mixture incubated and gently shaken at 38℃ for 120 min. Following this pepsin trea tment, 35 mg of Na_2HPO_4 was added and the pH adjusted to 6.5 by addition of 0.1 M NaOH. Trypsin (1 mL @ 1 mg/mL), chymotrypsin (1 ml @ 1 mg/mL), elastase (0.2 mL @ 1 mg/mL) and carboxypeptidase A (0.2 mL @ 1 mg/mL) were then added sequentially and each mixture was incubated and gently shaken at 38℃ for 120 min. Following each treatment, the digested gluten suspension was heated to 95℃ for 10 minutes and then allowed to cool to room temperature before the next treatment began.

Aliquots from each of the above proteolytic treatments were retained for analysis directly by HPLC-MS. The remainder of the digested mixture was treated with prolyl endopeptidase (PEP) and/or rat brush border membrane (BBM) enzymes and then analyzed by HPLC-MS.

2.2.2.2 Treatment with prolyl endopeptidase (PEP)

A portion of the wheat gluten extract that had been digested with all PTCECA enzymes was then treated with purified PEP after adjusting the pH to 6.5. The PEP was added at ratios of 30, 300 and 500 mU/mg digested gluten and then incubated at 38℃ for various reaction times ranging from 0 to 2 40 min. Following each PEP treatment, the reaction mixtures were heated to 95℃ for 10 minutes and then allowed to cool to room temperature.

2.2.2.3 Treatment with brush border membrane enzymes (BBM)

BBM enzymes were then added to both PTCECA and PTCEC+PEP treated wheat gluten extracts at a ratio of 130 mU/mg digested gluten. These mixtures were also incubated at 38℃ for various reaction times rangin g from 0 to 240 min. Following each BBM treatment, the reaction mixture was heated to 95℃ for 10 minutes and then allowed to cool to room temperature.

2.2.2.4 Sample preparation for HPLC-MS analysis

The final extracts from all proteolysis experiments were divided into 1 mL aliquots and centrifuged. The supernatants were evaporated to dryness and reconstituted in 95% / 5% acetonitrile (ACN) / H_20 + 0.025% trifluoroacetic acid (TFA), ready for HPLC-MS analysis.

2.2.3 HPLC-MS^n analysis

2.2.3.1 Enzymatic digestion analysis

All HPLC-MS analyses were performed on an Agilent 1100 HPLC-ion trap mass spectrometer, operated in atmospheric pressure positive electrospray ionization (ESI) mode. The RP-HPLC separation was performed by injecting 1-5 µL onto a C18 RP 0.32 x 150 mm column with 3.5 µm particles. Flow rates of 6 - 20 µL/min were used along with gradient elution programming using 0-95% ACN + 0.025 % TFA over 10 - 60 minutes.

The ion trap was optimized for mass transmission over the mass range 300-2200 Da using a standard Agilent tune compound. Under “auto-MS^n” operation, both full scan and MS^n spectra were acquired in order to obtain as much information as

possible to be able to identify the peptides in the sample. MS^n mass spectra were acquired from the most intense m/z ion signals within the selected “inclusion” mass range, provided the signal was above the specified intensity threshold. In order to avoid missing ions that resulted from chromatographic co-elution, “active exclusion” was used. This feature permitted the acquisition of auto-MS^n mass spectra from sequentially lower ion intensities (those that were still in the inclusion mass range and above the intensity threshold specified). After each LC peak (~ 5-8 s wide) the active exclusion feature was reset and the most intense ion was again selected; the process was continuously repeated. Identification of peptides in the extracts of proteolyzed wheat gluten was found to be possible predominantly using auto MS^2 mass spectra. Additional stages of tandem MS (i. e. MS^3) were used to achieve further sequence information of a detected peptide in order to further increase the probability of identification.

Some key operational parameters needed to be adjusted to maximize MS response. These included: the operation under data dependent auto-MS/MS mode, MS^3, and collision induced dissociation (CID) conditions using 35% of the parent mass using a fragmentor ramp of 0.3 to 2V in 30ms.

2.2.3.2 Metal ion analysis

Initial evaluation involved infusing various solutions of metal acetate salts into the mass spectrometer using a syringe pump, along with selected synthetic peptide solutions. These experiments allowed the optimization of the choice of metal and its respective molar concentration ratio (metal to peptide). Further experiments with on-

line HPLC-MS/MS analysis using metal ions were achieved using post column addition of the metal ion solutions on the Agilent ion trap system.

2.3 Results and Discussion

2.3.1 In-Vitro enzymatic digestion of wheat gluten

As an initial attempt at isolating and identifying the undigested gluten peptides that are implicated as triggers of gluten related diseases, it was theorized that a normal human *in-vivo* enzymatic digestion of food could be simulated in an *in-vitro* fashion. This would produce a mixture of peptides from digested food similar to that found in the gut of patients with celiac disease. Store bought wheat gluten flour was used and was digested using the major human gastric and pancreatic digestive enzymes (pepsin, trypsin, chymotrypsin, elastase and carboxypeptidase A). Normal *in-vivo* physiological conditions (i. e. ionic concentrations, pH, temperature, digestion time) were simulated as closely as possible. In addition, recombinant prolyl endopeptidase (PEP) and rat brush border membrane enzymes (BBM) were employed in order to identify any proline-containing peptides that resulted from further proteolysis.

All extracts of proteolyzed gluten were analyzed using RP-HPLC with UV and ion trap-MS^n detection. The resulting chromatograms were complicated and difficult to interpret because there were so many peaks present. A representative LC-MS total ion chromatogram (TIC) of wheat gluten flour proteolyzed with PTCECA is shown in Figure 2.1. Such complicated LC chromatograms and MS spectra were not unexpected due to the known complexity of the initial gluten matrix. However, over the 60 minute analytical time scale, it was difficult to monitor individual peaks

because most peaks represented mixtures of co-eluting peptides. Overall, it was noted from data from the pepsin proteolysis that no significant differences in the patterns of gluten peptides were observed over the entire course of treatment. The trypsin and chymotrypsin digestion experiments, however, did result in the observation of a shift of LC peaks towards shorter retention times, indicating the peptides had been digested into smaller peptides. Some further shifts toward even shorter retention times were observed from the data from the treatment with elastase and carboxypeptidase A, but not nearly as much as was observed following the trypsin/chymotrypsin treatment. Upon addition of prolyl endopeptidase (PEP), even shorter retention time LC peaks appeared, along with the reduction and complete disappearance of some peptides. The profile of these peaks represents P-containing peptides. This is illustrated in Figures 2.2(A) and (B). The LC-MS total ion chromatogram shows the degradation of one particular peak, (labeled with an arrow, in Figure 2.2(A)), from a large peak to a much smaller peak (Figure 2.2(B)) after the PEP was added. As expected, data from proteolyzed extracts that had been treated with PEP and the BBM enzymes showed that the only remaining peaks eluted quickly, presumably representing very small gluten peptides.

"PQ" sequences are present in every known immunogenic epitope (Table 1.1). An interesting means by which to identify the peptides that contained "PQ" sequences was to conduct a neutral loss scan. The neutral loss experiment involved performing CID and identifying the molecular weights of peptides that have lost PQ. It was theorized that this type of MS scan could simplify this enzymatic digestion data and make it easier to detect new potentially immunogenic PQ-containing target

peptides. Figures 2.3(A) and (B) show results from a representative neutral loss scan for “PQ” in proteolyzed wheat gluten flour. The peak labeled with an arrow in Figure 2.3(A) has totally disappeared after treatment with PEP, indicating it is a PQ-containing peptide that resisted complete proteolysis until the treatment with PEP. This indicated that this peptide may be physiologically relevant. This technique was able to identify some peptides that appeared to contain PQ sequences, but the sensitivity was quite low. In addition, it appeared to only identify those peptides where the PQ sequence was at the C-terminal end of the peptide.

Upon further evaluation of the data from these experiments the feasibility of this *in-vitro* enzymatic digestion approach towards the identification of immunogenic peptides from wheat gluten was proven in two ways. First, upon screening extracted ion chromatograms (EICs) from the data from proteolyzed wheat gluten for the presence of known immunogenic epitopes and peptides, the major immunodominant gluten peptide (αG-33) was discovered to be present. This is illustrated in Figure 2.4. The LC-MS total ion chromatogram in Figure 2.4(A) shows the various peptides that were released via the proteolysis procedure. The peak which is labeled at the retention time of approximately 33 minutes corresponds to the 33mer peptide. Figure 2.4(B) shows the full scan MS spectrum of this peak, where the triply charged ion at m/z 1304.6 Da, the $[M+4H]^{4+}$ ion at m/z 978.9 Da and the $[M+5H]^{5+}$ ion at m/z 783.1 Da are all observed. Figure 2.4(C) shows the MS/MS spectrum of the most abundant ion, the $[M+3H]^{+3}$ ion at m/z 1304.8 Da. This MS/MS spectrum was sent to the MASCOT protein database which correctly identified the peptide as the 33mer, an α-gliadin peptide from wheat with the correct sequence.

Further evidence that this enzymatic digestion approach can successfully identify potentially immunostimulatory gluten peptides is shown in Figure 2.5. Figure 2.5(A) shows the LC-MS total ion current chromatogram of various digestion products from wheat gluten that had undergone PTCECA digestion. One particular peak at a retention time of approximately 18 min. was shown to degrade upon treatment with PEP. Figure 2.5(B) shows full scan MS data for this peak at 18.3 minutes. Further evaluation of data in this full scan spectrum shows that several singly charged ions and one doubly charged ion can be seen with some intensity. The ion at m/z 980.5, corresponding to a $[M+2H]^{+2}$ ion, was selected by auto MS/MS which produced a product ion spectrum, shown in Figure 2.5(C). This MS/MS data was submitted to the Mascot protein database for an identity search. The top hit from this search is shown in Figure 2.5(D) and the report indicates that the peptide is a wheat α–gliadin with the proposed sequence LQPQNPSQQQPQEQVPL, which does contain short known immunogenic epitopes.

2.3.2 Metal ion affinity for proline

2.3.2.1 Determining metal affinity for proline

Metal ion acetate solutions were prepared from all compounds listed in Table 2.1. The objectives of these experiments were to compare MS/MS-CID spectra and sequence information from any metal adduct-gluten peptide complexes that formed, in order to determine if any metals showed an affinity for proline and/or glutamine residues. In addition, it would be useful to determine if there exists any metal affinity specifically for immunogenic peptides.

Solutions of 1:1 molar ratios of metal to proline were first evaluated by direct infusion into the ion trap mass spectrometer, in order to determine which metals would form a complex with proline. The affinity of a metal to proline was measured by the intensity of the MS signal of any adduct(s) that formed between proline and the particular metal. Figure 2.6 shows an example of how Ni complexes to proline. In the case of Ni, two adducts form; a nickel-acetate adduct and a nickel-proline adduct. In addition, it is interesting to note that since nickel has five naturally occurring isotopes (the two largest being ^{58}Ni (68%) and ^{60}Ni (26%)), the spectra of Ni complexes do show extra peaks corresponding to these isotopic contributions. In general, several types of adducts were observed in the MS spectra of metals with proline (P). They include: $[P+H]^+$, $[P+metal]^+$, $[2P+metal]^+$ and $[P+acetate+metal]^+$. Each metal was not observed to form all of these adducts with proline. Under the conditions of these experiments, the results summarized in Figure 2.7, show that Cu, Mn, Co, Ni and Zn appeared to have the best affinity for proline. An interesting observation from these results notes that sodium (Na) also showed a high affinity for proline. However, in mass spectrometric analyses, sodium is an unwanted species because it has a high affinity for just about anything that is able to form adducts. Therefore, it would not be a good choice for use in this instance, where specificity for only proline-containing peptides is required.

A comparison of the metal affinity for a proline-containing peptide versus a non proline-containing peptide evaluated the αG-33 (33mer) and a myoglobin peptide (KGHHEAELKAL) respectively. The 33mer was used here because of its known immunogenicity. Similar results were obtained in the previous experiment,

where metals Mn, Zn, Ni and Fe demonstrated a significant affinity for the 33mer, as compared to the myoglobin peptide which contained no prolines (Figure 2.8). Further evidence why alkali metals (i. e. Na and K) were not chosen for further study is the apparent lack of specificity shown by the attachment of potassium (K) to the proline-containing peptide and the non proline-containing peptide. Similar results are again shown for a cocktail of various synthetic proline and glutamine containing peptides (Figure 2.9).

Optimization of various instrumental parameters and experimental conditions, such as the MS fragmentor voltage, molar ratio (metal to peptide), pH and post column flow rate, endeavored to maximize the intensity of the MS signal for each peptide-metal complex. This was accomplished using flow injection analysis (FIA) of various peptide solutions concomitant with the direct infusion of the various metal solutions. Figure 2.10 and Figure 2.11 illustrate results from the pH and post column flow rate optimization experiments. The pH was found to be an interesting condition to evaluate. Figure 2.10 shows that 33mer-zinc adduct formation was observed to be very low at a high pH (9). The reason for this is likely because peptides become anionic at high pH and can form an ion-pair with any metal cations present. This results in an overall lack of selectivity and sensitivity because the resulting complexes would be neutral, or exist at a low enough charge state whereby the mass exceeds the upper end of the operational range of the mass spectrometer (as would be the case for singly charged peptide complexes >2200 m/z). At a low pH, peptides become protonated, which increases the likelihood that binding with a particular metal cation would be much more selective. Therefore, adduct formation

under these circumstances would likely be determined by the actual affinity of the metal for the specific amino acid composition of the peptide.

Flow rate optimization, as shown in Figure 2.11, was the means by which the metal to peptide molar ratio was evaluated. The metal solutions were introduced via post column addition and adjusting the flow rate changed the actual amount of metal available to bind to the peptide. The concentration of metal in relation to that of the peptide was found to be important in the formation of adducts. As the flow rate increased it was generally observed that adduct formation also increased. However, as the relative concentration became very high (i. e. 560μL/hr or 80:1 molar ratio), multiple metal attachments became prevalent. Too many metal ions led to the condition were non-selective binding prevailed in solution. In addition, a very high flow rate resulted in a significant dilution of the peptide solution. For the 33mer and zinc, an acceptable compromise was achieved with a flow rate of 280 μL/hour, which corresponded to a metal to peptide molar ratio of 40 to 1.

2.3.2.2 Sequence information about gluten peptides using metal ion-adducts

Sequence information (primary amino acid structure) about a peptide or protein can be obtained by performing collision-induced dissociation (CID)-MS/MS experiments. When the parent ion of a compound fragments during CID, the product ions that result are representative pieces of that parent compound itself. If the compound studied is a peptide, common fragmentation patterns exist that produce product ions that can be considered characteristic parts of the peptide. These can be pieced back together to determine the original structure of the parent peptide. Figure 2.12 illustrates some common fragmentation patterns for peptides that undergo CID

in an ion trap or tandem mass spectrometer. Fragment ions occur in pairs, such as a/x type ions, b/y type ions, or c/z type ions. In positive atmospheric pressure electrospray ionization (ESI) mode, the peptide fragment detected depends on the location of the basic sites within the peptide, because these are the sites that will carry the positive charges. In low energy CID conditions, such as those that occur in an ion trap or tandem mass spectrometer, the types of fragment ions that are more commonly detected are the b/y type ions.

As an initial attempt to investigate the type of structural information that could be obtained from CID-MS/MS of peptides complexed with metal ions, a simple small synthetic hexapeptide ($YSGFd_3$-LT (689 m/z)) peptide was chosen which did not contain any proline residues. Solutions of various metal ions (such as copper and silver acetates), were combined with the hexapeptide at near equal molar ratios, in order to observe any significant differences in their respective CID-MS/MS spectra. It was observed that MS/MS product ions could be obtained from each of the $[M+H]^+$, $[M+Cu]^+$ and $[M+Ag]^+$ parent ion complexes. The b and y type fragment ions for each complex are shown in Figure 2.13(A)-(C) respectively. Figure 2.13(A) shows the sequence ions generated from MS/MS of the $[M+H]^+$ ion at m/z 690. This MS/MS spectrum shows a y5 ion at m/z 526, a weak y4 at m/z 439, a b5 ion at m/z 677 and a b4 ion at m/z 561. The formation of copper and silver complexes greatly enhanced the sensitivity and number of the b type fragment ions detected, as seen in Figures 2.13(B) and (C). The addition of copper (Figure 2.13(B)) formed an $[M+Cu]^+$ ion at m/z 751. The MS/MS spectrum shows the loss of copper at m/z 689 and b5 and b4 fragments at m/z 632 and m/z 516 respectively. The addition of silver (Figure

2.13(C)) formed an $[M+Ag]^+$ ion at m/z 796. The MS/MS spectrum of that ion predominately showed the b5, b4 and b3 peptide product ions at m/z 677, 561 and 414 respectively. In addition, the a5, a4, and a3 product ions (cleavage across the C-(C=O) versus the (C=O)-N bond) produced an ion series at m/z 649, 533 and 386 respectively. Clearly, the addition of metals like silver and copper can provide complementary structural information which can aid in identifying C-terminal sequences in some peptides (through the determination of b-type ions) to the more traditional type of information provided from proton type CID-MS/MS spectra.

Sequence information from the $[M+3H]^{3+}$ proton CID-MS/MS spectrum of the 33mer was compared with that from the 33mer-Ni adduct spectrum (Figure 2.14). Again, it was observed that the fragment ions were all of the b-type. Therefore, an interesting observation using these metal-adducts to provide information needed to determine sequence information from the CID-MS/MS analysis of peptides, is that most fragment ions appeared only to be of the b-type. This allows sequence information about the C-terminal portion of the peptide to be determined. Another interesting observation is that the fragmentation appears to cleave the peptide only at N-terminal proline sites.

Finally, target peptide 1 (proposed sequence LQPQNPSQQQPQEQVPL), that was identified as a potential immunostimulatory peptide in the enzymatic digestion experiment, was complexed with various metals in order to investigate whether similar structural information could be obtained. Results using Zn are shown in Figure 2.15. The results again show that some interesting C-terminal sequence information becomes available, whereby the fragments that were observed were b-

type ions and were again formed by breaking peptide bonds on the N-terminal side of the C-terminal prolines.

2.4 Conclusions

It appears to be feasible to proteolyze wheat gluten in an *in-vitro* fashion, under similar conditions to those which take place *in-vivo*. Under the conditions used in these experiments, wheat gluten proteins were proteolyzed until relatively small peptides remained. These peptides appeared to be resistant to further enzymatic degradation. Some of these remaining peptides were seen to degrade upon treatment with a recombinant prolyl endopeptidase, which would infer that they may be of the type of peptides that are immunostimulatory in people with gluten sensitivities. Data analysis from these experiments was difficult owing to the complexity of the chromatograms. However, one such peptide was identified by obtaining satisfactory MS/MS data and searching existing protein databases. The proposed sequence was confirmed by manual interpretation as LQPQNPSQQQPQEQVPL.

It was observed from experiments with metal ion-adducts, that proline did show an affinity for the following metals: Cu, Mn, Co, Ni, Zn and Ag. Using post column addition of metal acetate solutions, the formation of metal-peptide complexes was optimized at pH 1.5 (with 0.025 % TFA) and at a molar ratio of > 4:1 (metal to peptide). Some metals, such as: Zn, Mn, Fe, and Ni showed a higher affinity for proline containing peptides than non–proline containing peptides. Collision- induced dissociation (CID)-MS/MS of the metal-adducts favored fragmentation to b-type product ions at the N-terminal side of C-terminal proline

linkages. MS/MS of non-proline containing peptides also appeared to favor fragmentation to b-type ions. The sequence information gathered from the CID experiments was different from that obtained by proton MS/MS spectra, thus providing a complementary set of sequence information.

Based on these initial results, it was expected that with further research, this HPLC-MS/MS protocol could lead to the identification of more peptide sequences. However, an important observation noted was that there was no obvious indication as to whether any of the peptides identified in this manner would be, in fact, immunogenic in nature. An alternate strategy for the identification of peptides that are immunogenic would have to be devised.

metal	ave. metal m/z (Da)	metal acetate formula	ave. formula weight (g/mole)	metal solution conc. (mM)
Cu	63.5	$Cu(C_2H_3O_2)_2.H_20$	199.7	55.7
Ba	137.3	$Ba(C_2H_3O_2)_2$	255.43	96
Cr	51.9	$Cr_3(C_2H_3O_2)_7\ (OH)_2$	603.3	39.3
Mg	24.3	$Mg(C_2H_3O_2)_2$	142.3	163
Ca	40.1	$Ca(C_2H_3O_2)_2$	158.2	129.2
Rb	85.5	$Rb(C_2H_3O_2)$	144.5	212
Li	6.9	$Li(C_2H_3O_2).2H_2O$	102	229.2
Cs	132.9	$Cs(C_2H_3O_2)$	191.9	105.8
K	39.1	$K(C_2H_3O_2)$	98.1	215.6
Na	22.9	$Na(C_2H_3O_2).3H_2O$	136	147.6
Sn	118.7	$Sn(C_2H_3O_2)_2$	236.7	15
Ag	107.8	$Ag(C_2H_3O_2)$	166.9	52.2
Ni	58.7	$Ni(C_2H_3O_2)_2.4H_2O$	248.8	104
Mn	54.9	$Mn(C_2H_3O_2)_2$	172.9	60.7
Pd	106.4	$Pd(C_2H_3O_2)_2$	224.4	20.8
Zn	65.4	$Zn(C_2H_3O_2)_2.2H_2O$	219.5	44.2
Fe	55.8	$Fe(C_2H_3O_2)_2$	173.8	49.2
Sr	87.6	$Sr(C_2H_3O_2)_2$	205.7	115.8
Co	58.9	$Co(C_2H_3O_2)_2$	177	68

Table 2.1

Metal ion solutions used for P/Q affinity experiments

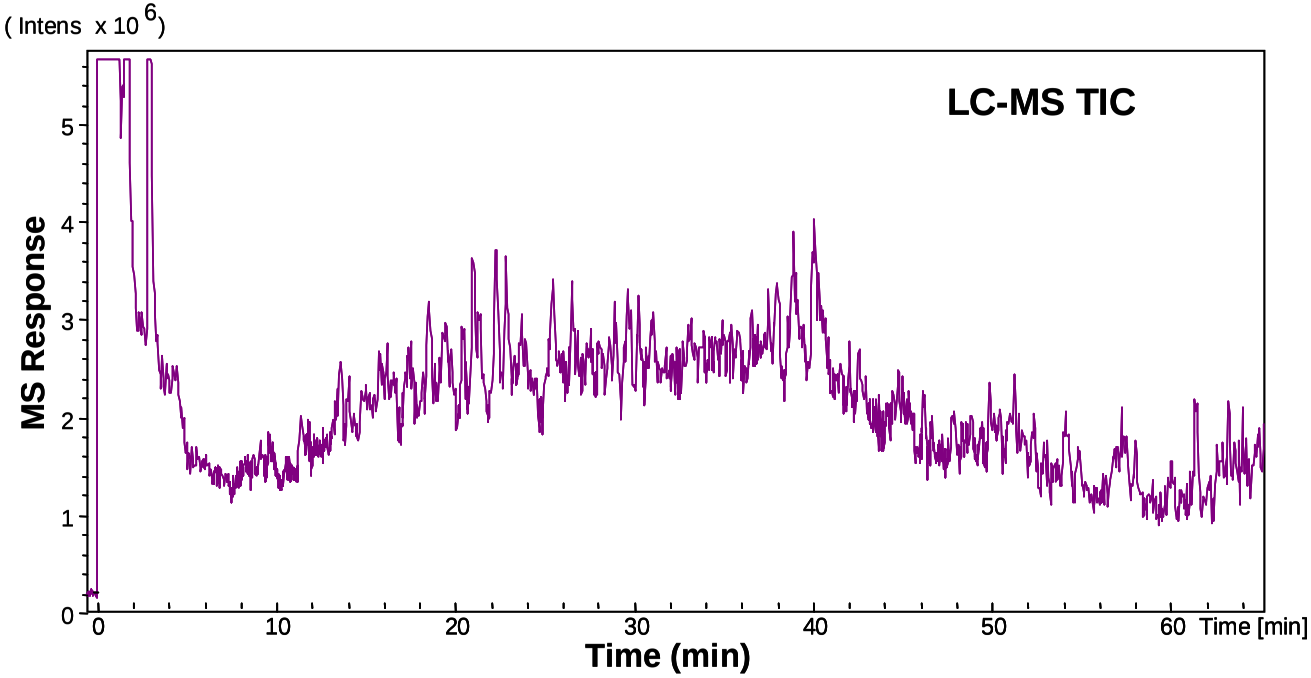


Figure 2.1

In-vitro proteolytic digestion of wheat gluten flour with PTCECA

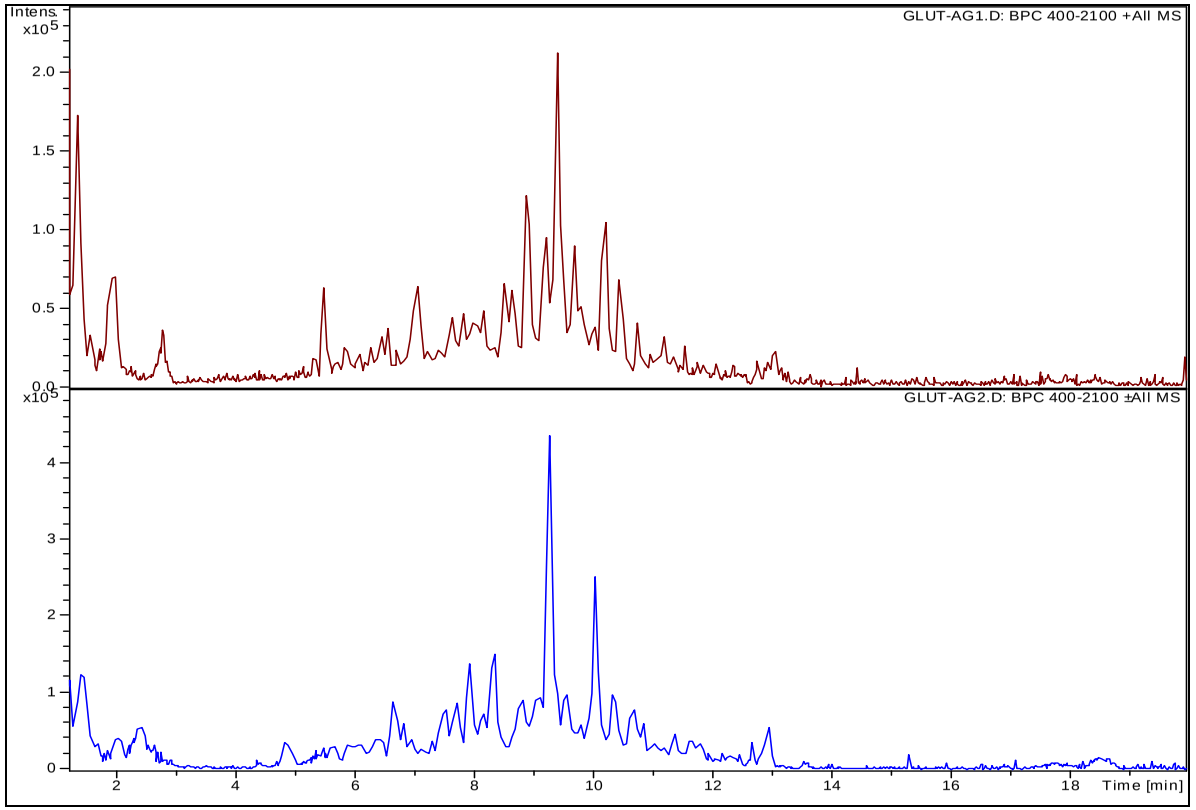


Figure 2.2

Degradation of some wheat gluten peptides following treatment of PTCECA and prolyl endopeptidase (PEP)

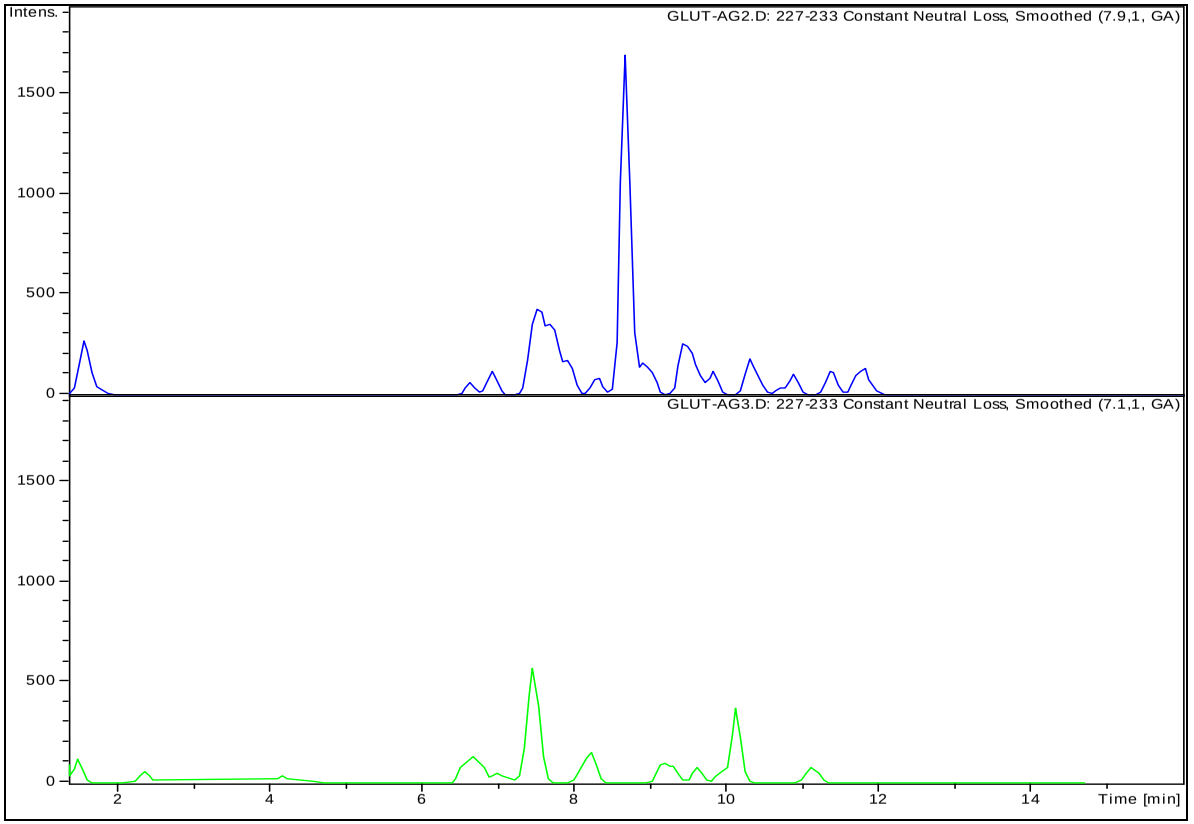


Figure 2.3

Screening for “-PQ-” containing gluten peptides from proteolyzed wheat gluten via a constant neutral loss scan

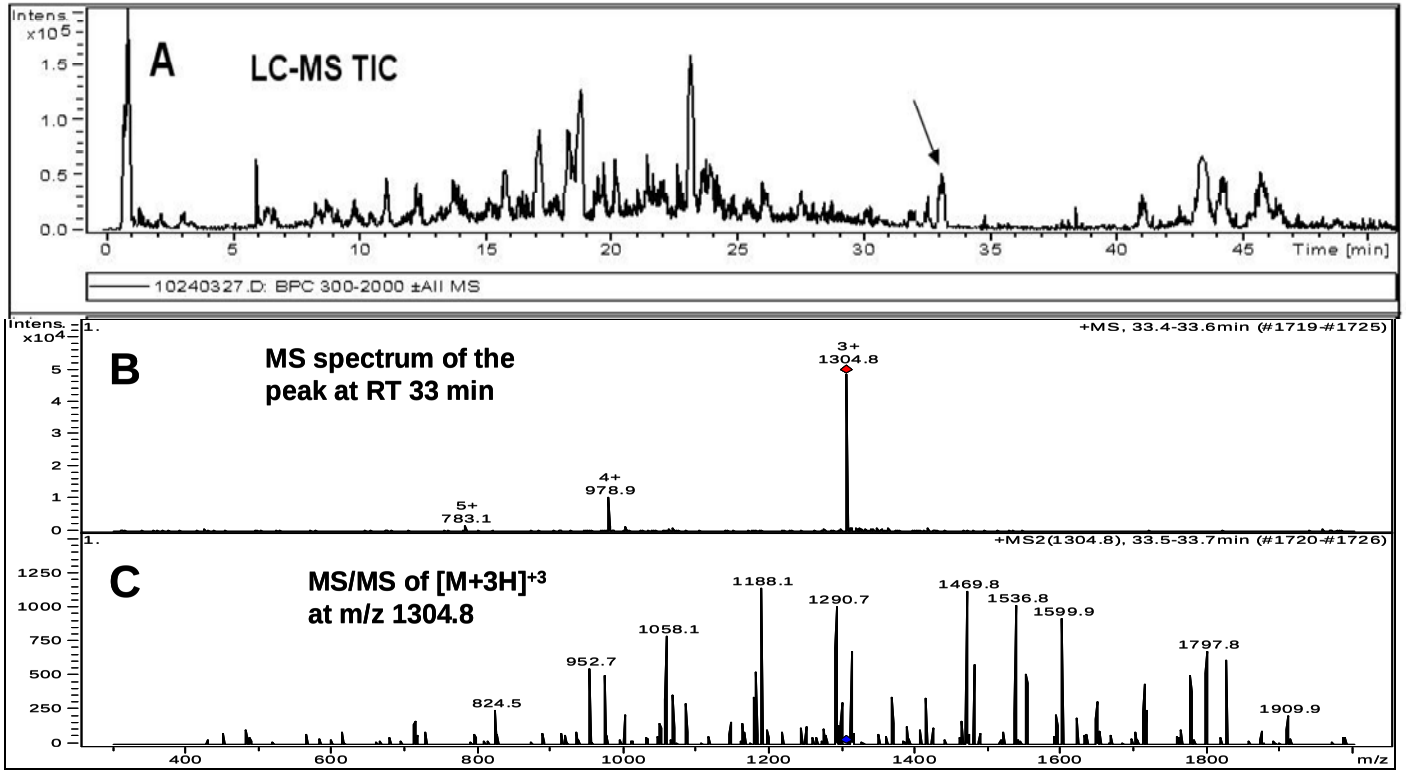
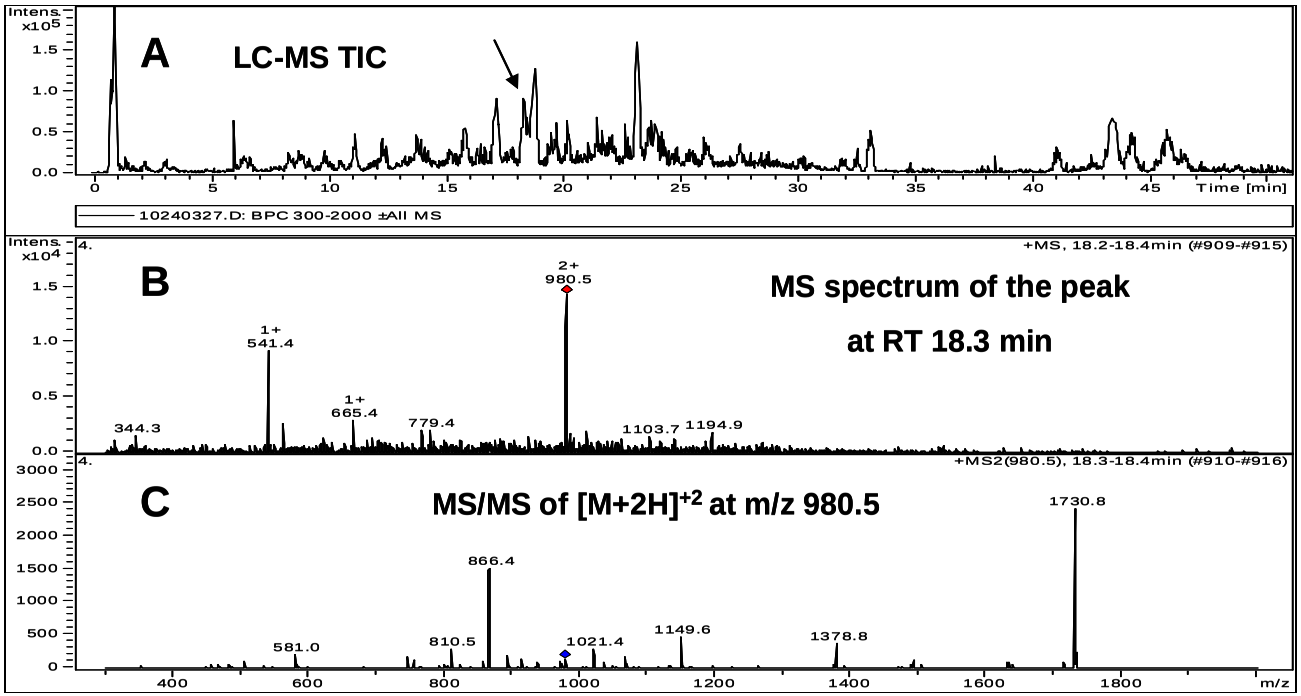


Figure 2.4

In-vitro digestion of wheat gluten – detection of the 33mer (αG-33)



D

1. P02863-00-03-00 **Mass:** 32943 **Total score:** 28 **Peptides matched:** 1
(GDAO_WHEAT) Splice isoform Displayed; Variant CLONE PW8233 & REF. 3; Conflict Displayed; from P02863 Alpha/beta-gliadin
Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Rank	Peptide
1	980.51	1959.01	1957.99	1.02	0	34	1	LQPQNPSQQQPQEQVPL

Figure 2.5

In-vitro digestion of wheat gluten – identification of potential target peptide 1

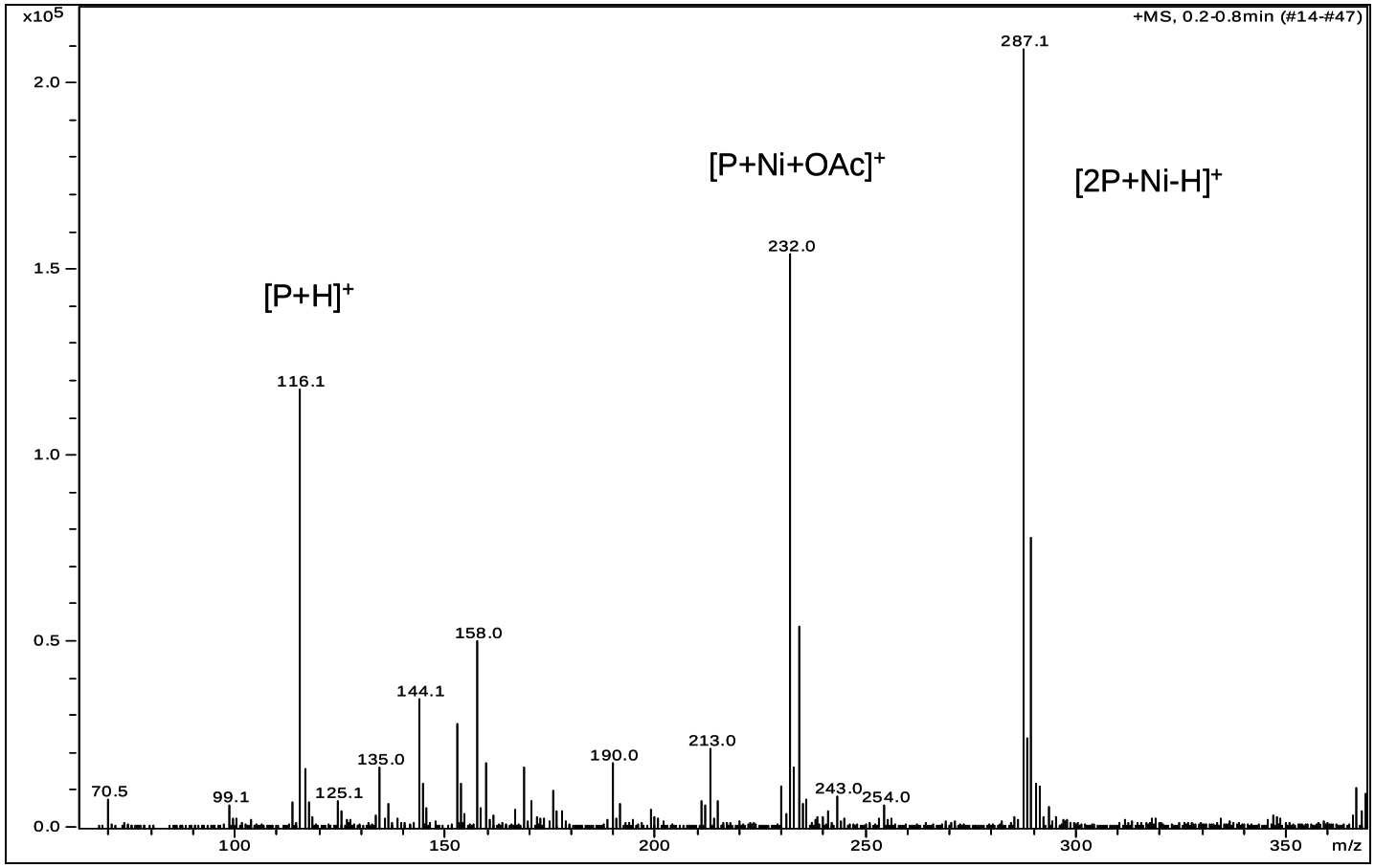


Figure 2.6

Investigation of Ni-adducts to proline

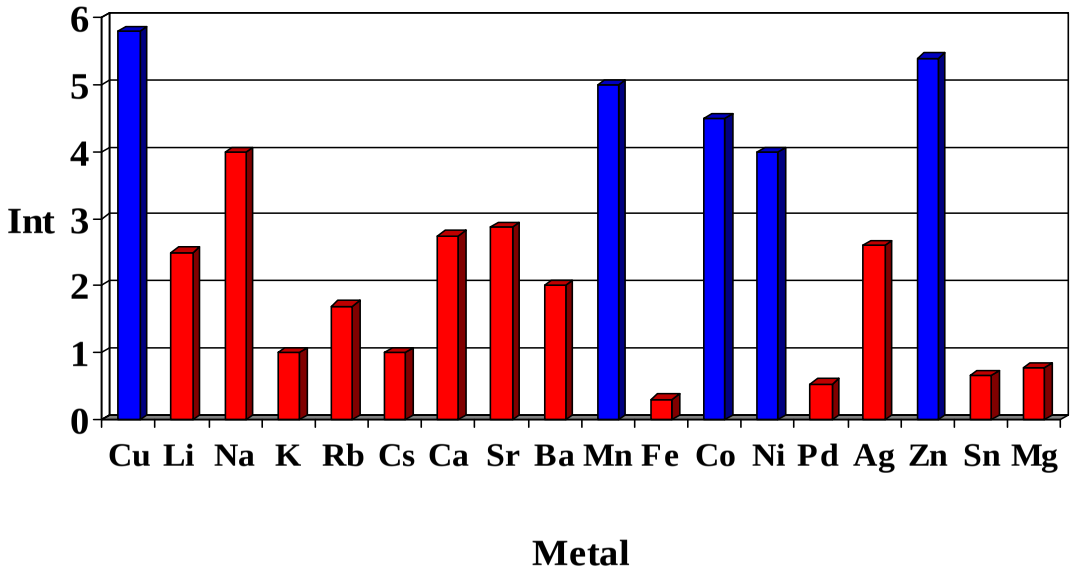


Figure 2.7

Investigation of metal Ion affinity for proline

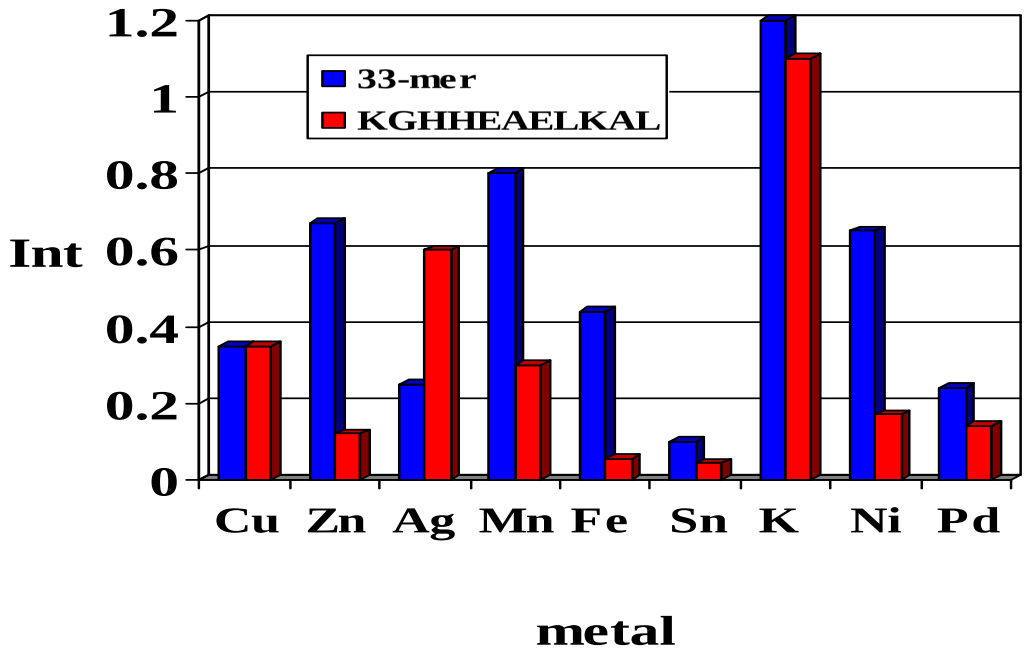


Figure 2.8

Metal ion affinity for a P-containing (αG-33)
versus non P-containing peptide

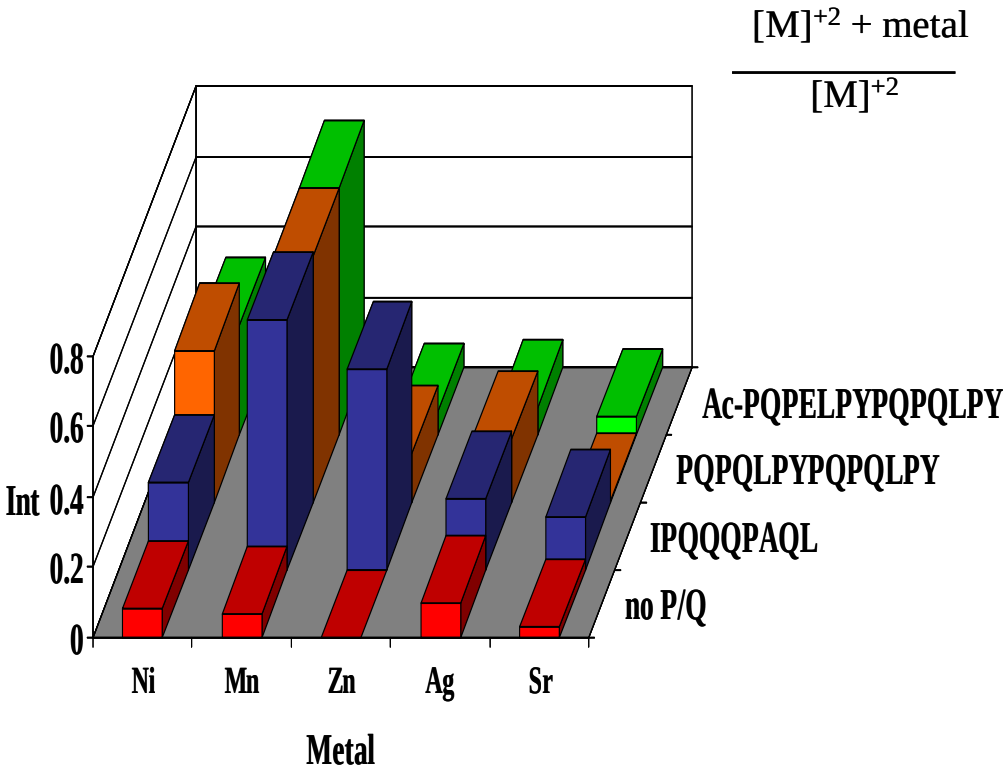


Figure 2.9

Metal ion affinity for various peptides containing P and Q

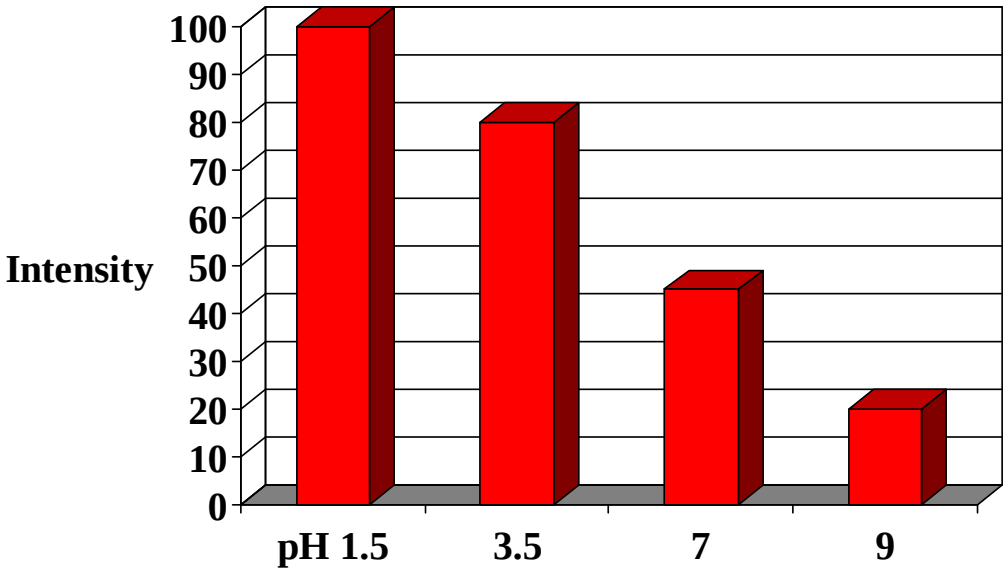


Figure 2.10

Effect of pH on peptide (33mer)-Zn adduct formation

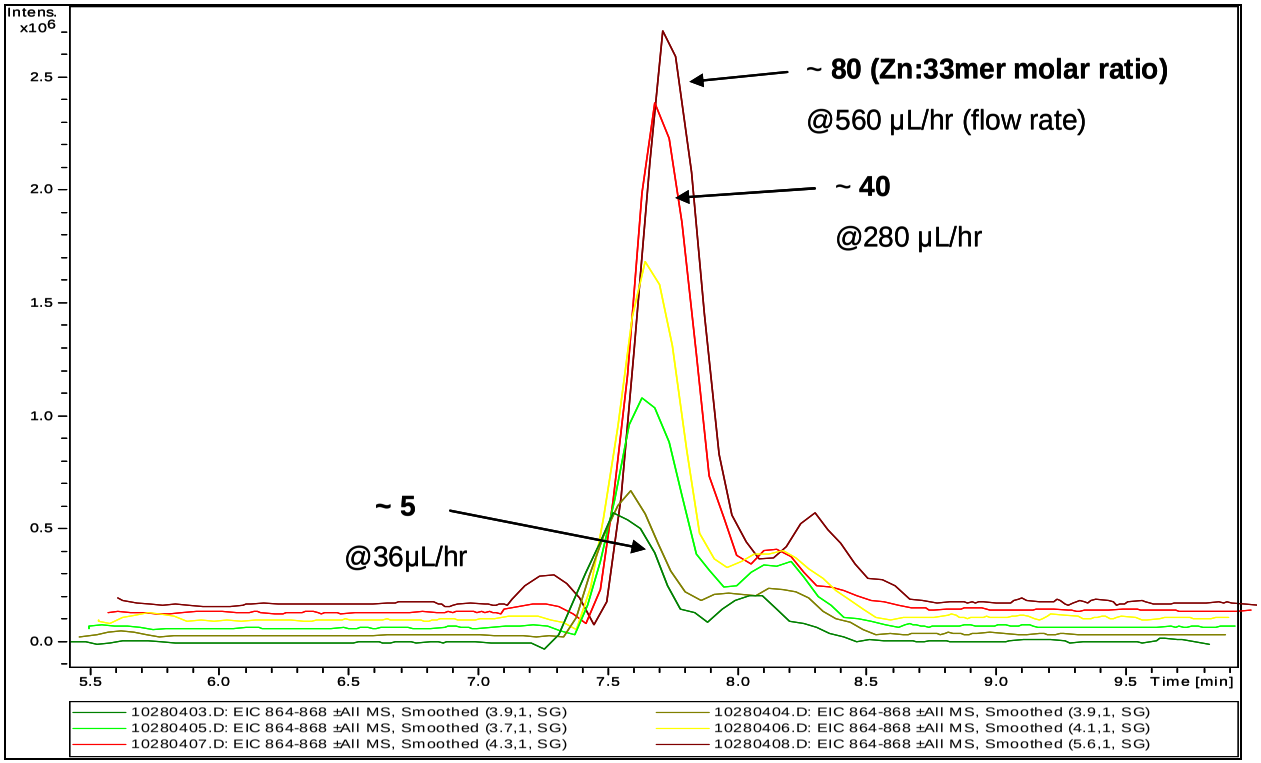


Figure 2.11

Optimization of metal:peptide molar ratio (Zn:33mer)

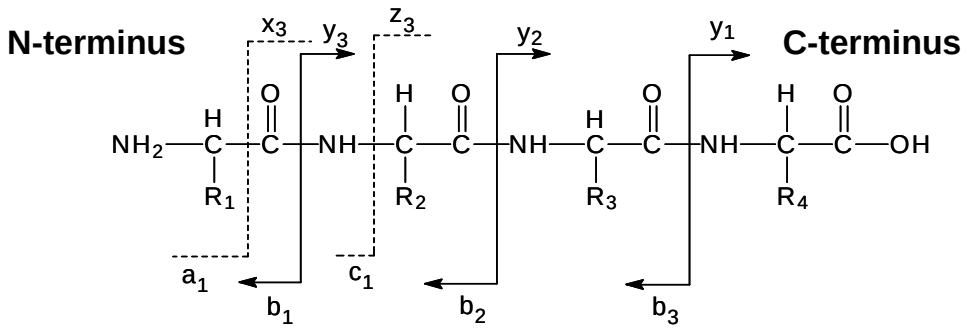


Figure 2.12

Common peptide fragmentation patterns under CID-MS/MS conditions

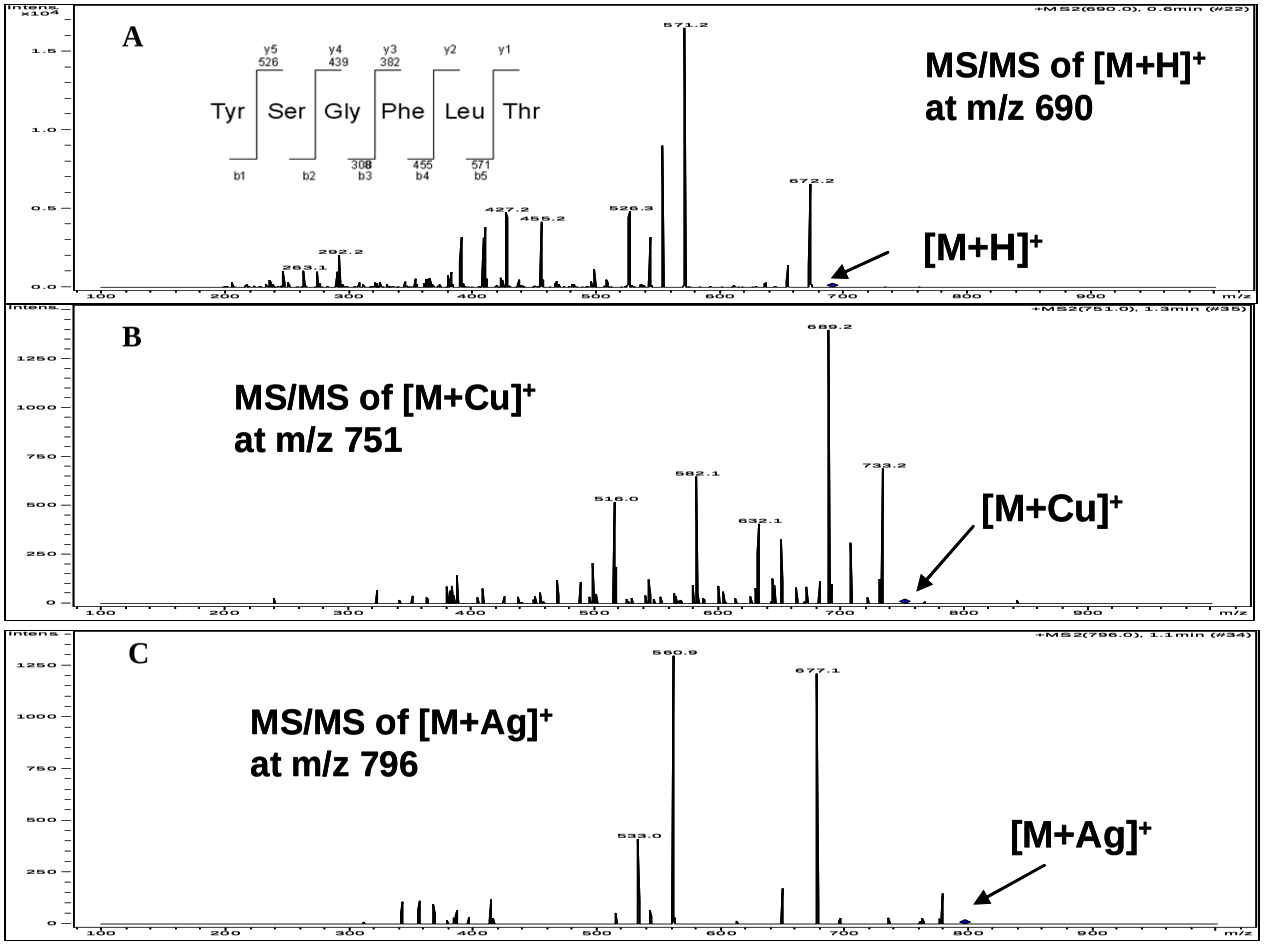


Figure 2.13
CID-MS/MS Sequencing of complexes of YSGFLT with Cu and Ag

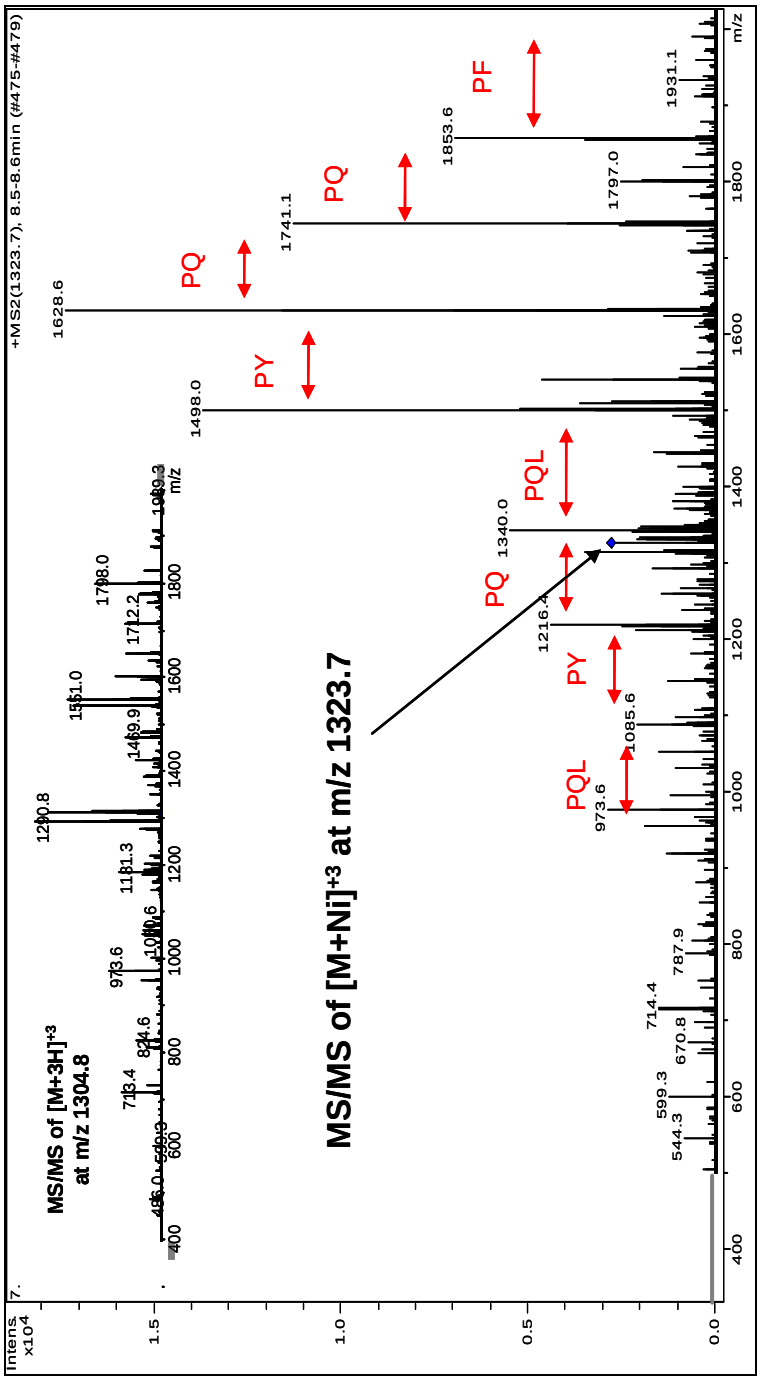


Figure 2.14
CID-MS/MS of αG-33 (LQLQPFPQPQLPYPQPQLPYPQPQPF) with Ni

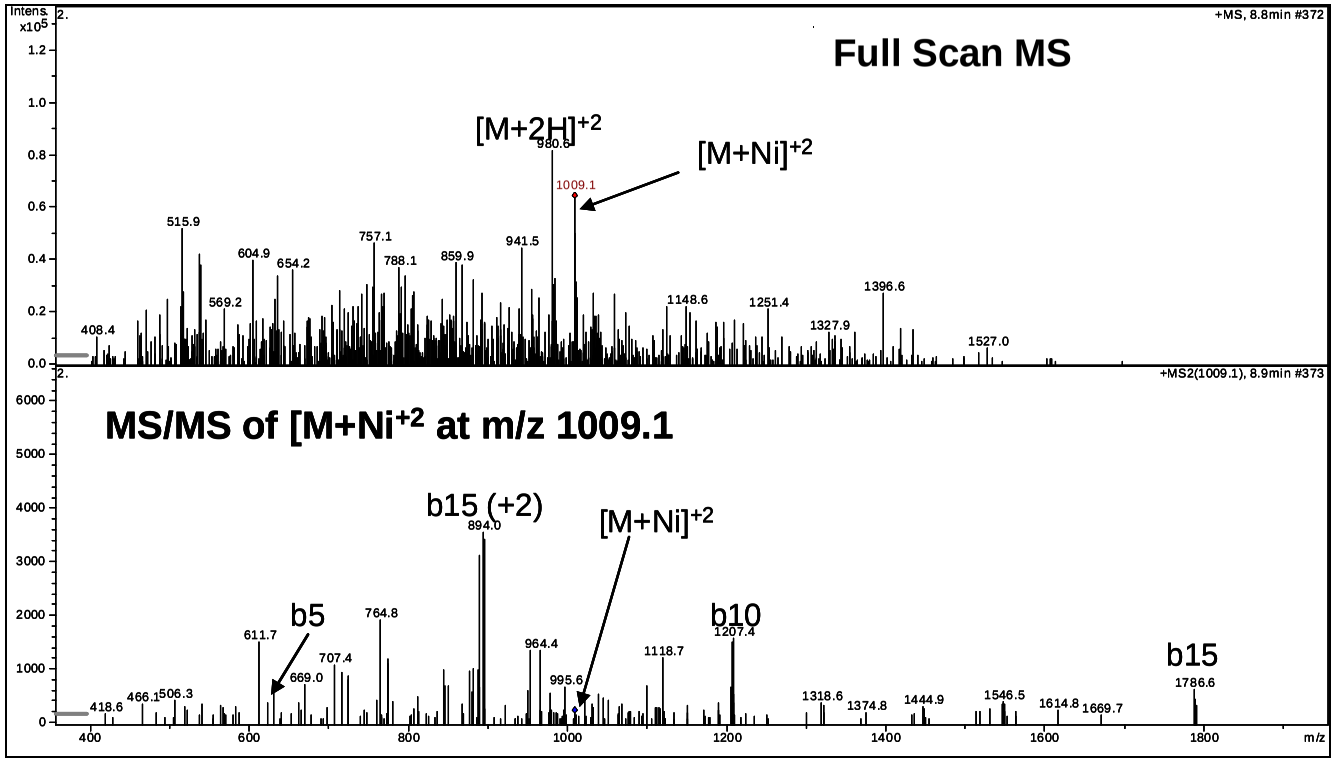


Figure 2.15

CID-MS/MS of peptide 1 (LQPQNPSQQQPQEQVPL) with Ni

2.5 References

[1] Shan, L.; Molberg, O.; Parrot, I.; Hausch, F.; Filiz, F.; Gray, G.; Sollid, L.; Khosla, C. *Science* **2002**, *297*, 2275-2279.

[2] Sollid, L. *Annual Immunological Reviews* **2000**, *82*, 53-81.

[3] Kohler, M.; Leary, J. *Analytical Chemistry* **1995**, *67*, 3501-3508.

[4] Wang, Y.; Taylor, J.; Gross, M. *Journal of the American Society for Mass Spectrometry* **2001**, *12*, 550-556.

[5] Ho, Y.; Huang, P.; Deng, K. *Rapid Communications in Mass Spectrometry* **2003**, *17*, 144-121.

[6] Park, S.; Kim, Y.; Choi, S.; Kim, I. *Korean Chemical Society Bulletin* **2001**, *22*, No. 7.

[7] Tao, W.; Wu, L.; Cooks, R. *Journal of the American Society for Mass Spectrometry* **2001**, *12*, 490-496.

[8] Lin, T.; Payne, A.; Glish, G. *Journal of the American Society for Mass Spectrometry* **2001**, *12*, 497-504.

[9] Schlosser, A.; Lehmann, W. *Journal of Mass Spectrometry* **2000***, 35*, 1382-1390.

CHAPTER 3

Identification of Immunogenic Wheat Gluten Peptides Based on Enzymatic Tagging and HPLC-UV-Fluorescence and-MS^n

---

3.1 Introduction

3.1.1 A new strategy

The hypothesis that immunogenic wheat gluten peptides could be identified solely through *in-vitro* enzymatic digestion of wheat gluten flour followed by HPLC-MS detection proved somewhat successful, because some peptides that were identified in this manner did appear to be stable to further proteolysis. It was observed that some of these peptides only degraded further upon treatment with a prolyl endopeptidase (PEP), which inferred that they may belong to the class of peptides that mediate T-cell related inflammatory reactions in those with celiac disease and other gluten sensitivities. A well studied and known potent immunostimulatory 33mer (LQLQPFPQPQLPYPQPQLPYPQPQLPYPQPQPF) [1], was one of the peptides identified from the proteolysis of wheat gluten flour in this manner. This provided evidence that the physiological conditions used for *in-vitro* procedure were appropriate and that the procedure itself was successful in its ability to release such peptides. One new potentially immunogenic peptide, a 17mer with the proposed sequence LQPQNPSQQQPQEQVPL, was also identified from these

experiments. However, further efforts toward additional peptide identification were not pursued due to the complexity of the resulting MS spectra, which made structural interpretation difficult.

The use of metal ion attachment provided some interesting observations about the structures of gluten peptides that resisted proteolytic digestion and how they respond during collision-induced decomposition (CID) in an ion trap mass spectrometer. Fragmentation of metal ion peptide adducts appeared to only produce b-type ions, which were produced by breaking the peptide bond on the N-terminal side of any proline residues present. This type of fragmentation allowed sequence information about the C-terminal end of the peptide to be interpreted. The more prolines there were in the peptide, the more b-type ion fragments there were in the MS/MS spectra. However, the bigger the peptide, the more complicated the spectra were to interpret.

From the experimental data obtained, it was discovered that hundreds of gluten peptides were released during enzymatic digestion and those with prolines, that formed complexes with the metals, were not necessarily immunogenic. From this information, there was no way to really determine which, if any, of these peptides would actually be immunostimulatory *in-vivo*. The only method that determines direct gluten immunogenicity towards a specific gluten peptide sequence utilizes the *in-vitro* stimulation of T-cells, as measured by a [3H]-thymidine incorporation assay [2]. This expensive and labor intensive assay relies solely on biopsy derived T-cell lines or clones from patients with celiac disease (CD). Thus, in

order to proceed with the identification of immunogenic gluten peptides, it was determined that it would not be feasible to continue with this particular strategy.

Therefore, another strategy was devised, whereby the focus shifted to the mechanism of the autoimmune reaction to gluten and what occurs *in-vivo* during the adaptive type immune response. It was theorized that by understanding how the body itself determines which gluten peptides are “toxic” might provide an insight into how they might be identified *in-vitro*. The new strategy for the identification of immunogenic gluten peptides focused on what was known about the relationship between certain types of gluten peptides and a tissue transglutaminase enzyme. It has been determined that they have a unique association together in the adaptive autoimmune response mechanism of CD and this has important connotations for the pathogenesis of the disease [3 - 5].

3.1.2 Molecular basis for the adaptive immune response in celiac disease

Celiac disease (CD) is a result of a deficiency in the immune system that occurs in those who are genetically predisposed. In general, a human immune system protects its host with a layered defense of increasing specificity. Normally, if a pathogen makes it though the initial layer of defense, an innate (non-specific) response is activated. Another layer of protection (adaptive immune response) is activated should antigen-specific recognition be required. It is known that both innate and adaptive immune responses are involved in celiac disease, although the mechanism of each is not fully understood [6]. Notwithstanding, the specific “non-self” antigens that are targeted during both immune responses in CD are known to be a group of structurally similar cereal grain gluten peptides.

The genetic predisposition for celiac disease is based on the inheritance of a configuration of human leukocyte antigen (HLA) genes, namely genes in the HLA-DQ locus [4]. The function of these major histocompatability complex (MHC) HLA-DQ genes is to encode antigen presenting proteins onto the surface of immune system cells. These proteins display peptide fragments, from the cell itself and/or fragments of invading microorganisms, to T-helper cells (white blood cells) by binding them to a unique CD4 receptor on the surface of the T-cell. This T-cell complex then has the capacity to kill or co-ordinate the killing of pathogens or infected/malfunctioning cells [6].

It has been determined that the configuration of alpha and beta HLA- DQ chains inherited, directly influences the risk of developing CD. More specifically, it is the HLA-DQ2/DQ8 genes that have been identified as those responsible for selectively binding certain types of small gluten peptides and presenting them to CD4+ T-cells. These T-cells are involved in further mediation of the autoimmune process. Current knowledge in this area has identified four haplotypes (a kind of well-defined genotype) that can be considered risk factors associated with the development of CD (Table 3.1) [6, 7]. The first and last haplotypes are considered sufficient to carry significant risk in the development of the disease. These are the DQA1*0501 & DQB1*0201 alleles (that encode DQ2) and the DQA1*0301 & DQB1*0302 alleles (that encode DQ8).

An overview of what is known about the adaptive immune response is illustrated in Figure 3.1 [8]. Known immunogenic gluten peptides have a specific structure which makes them attractive to bind to both the DQ molecules and T-cells.

These peptides can resist proteolysis in the gut and make their way inside the epithelial layer of the villi that line the inside of the intestinal wall. Meanwhile, inside the epithelial layer, antigen presenting cells (dendritic cells in the intestines) bind HLA DQ2/8 proteins to their surfaces. These bound proteins then selectively bind to the gluten peptides. A key step in the immune mechanism is noted here, whereby the only way that these gluten peptides can bind, is if they have been selectively modified first. This post-translational modification step is mediated by a tissue transglutaminase enzyme (TG2), present in the lamina propria of the intestinal epithelium. It has been determined that the purpose of this post-translational modification appears to be to increase the attractiveness of the gluten peptide for both the DQ molecules and the CD4+ T-cells. This APC-DQ-modified gluten peptide complex is now presented to the T-helper cells, which then bind via their CD4 receptors. This CD4 receptor is a special kind of receptor that specifically recognizes the DQ2/8 in the complex. The T-cell, now bound to the complex, becomes activated, divides and causes B-cells, which recognize gluten, to then divide. These cells subsequently produce various pro-inflammatory cytokines and anti-bodies to both TG2 and gluten. Destruction of the intestinal villi and other tissue damage is the end result of this complex adaptive immune response to gluten [3, 8].

3.1.3 TG2 catalyzed post-translational modification of wheat gluten peptides

Type 2 tissue transglutaminase (TG2) is one of a group of transglutaminase enzymes found in most living systems. Currently, there have been eight transglutaminase isoenzymes identified in mammals. Tissue transglutaminase enzymes have been studied since the 1950s [9, 10], with their involvement

documented with a number of disease states, autoimmune conditions, cancer and cellular processes [11]. Under normal physiological conditions, transglutaminase enzymes are inactive, but under disrupted physiological conditions and in the presence of high (mM) concentrations of Ca^{2+}, they can become activated. Once these enzymes become activated, they can catalyze reactions that can result in the post-translational modification of proteins/peptides. Acyl-transfer mechanisms, involving specific glutamine (Q) residues (as acyl donors) and primary amines (as acyl acceptors), generate deamidated and/or cross-linked products [12 - 14]. De-regulation of normal TG2 functions have been well documented with respect to a number of human diseases, such as neurodegeneration, neoplastic diseases, skin and tissue fibrosis related diseases, as well as autoimmune diseases such as CD.

TG2 is expressed on the surfaces of most cells and exists ubiquitously in an inactive state [15]. In the case of CD, the TG2 enzyme and Ca^{2+} are both found in high concentrations inside cells of the small intestine. It is not fully understood why this is the case nor where and how the interaction of the peptides and TG2 takes place. It is known, however, that the enzyme selectively modifies the gluten peptides by creating negatively charged sites which have been found to be required in order to allow peptide binding to the DQ molecules [13, 16] to occur.

TG2 is a 686 amino acid, 80 kDa cysteine protease and catalyzes the modification that occurs to the gluten peptides. This modification involves the creation of negatively charged sites by targeting specific glutamine (G) residues within the peptide and changing them into glutamic acid (E) residues. Normally found in the inactive state, TG2 becomes activated under such conditions that

require an unusually high concentration of Ca^{2+} is present. Once activated, TG2 catalyzes deamidations and transamidations involving these glutamines via a charge-relay mechanism in its catalytic triad (cysteine^{277} – histidine^{335} – asparagine^{358}). Transient thioester intermediates are formed with specific gluten peptidyl glutamines or lysines [17, 18]. This catalytic activity is exhibited toward γ-carboxamide groups of glutamine residues and ε-amino groups of lysines, which result in either an inter- or intra-chain iso-peptide bond [19].

Known immunogenic gluten proteins contain high percentages of the amino acids glutamine (G) and proline (P) and low percentages of negatively charged residues. Because these peptides are rich in glutamine, it permits them to be susceptible to these deamidation or transamidation reactions. As previously stated, in order to for these peptides to become bound to the groove on the HLA-DQ molecules, it has been determined that they must first be chemically modified by imparting negatively charged sites [20]. TG2 fulfills this role, by targeting specific glutamine residues and converting them into glutamic acid (E) residues [16, 21]. TG2 is actually thought to catalyze two competing pH dependent reactions, both resulting in villous atrophy [22]: (1) transamidation; where the TG2 cross-links gliadin peptides to itself or to other gliadin peptides, thus forming TG2-gliadin-gliadin peptide complexes; and (2) deamidation; where specific glutamine residues are converted to glutamic acid (E) residues with the release of NH_3. Both deamidation and transamidation reactions are known to occur, but the deamidation reaction appears to be favored under conditions of decreasing pH.

The association of TG2 with the binding of gluten peptides to DQ2 and T-cells appears to be very important in the mechanism of the adaptive autoimmune reaction in CD. The determination of the structures of peptides that become deamidated by TG2, could prove to be a useful means of identification of those peptides responsible for the inflammatory processes that occur in those sensitive to gluten. Therefore, I chose to simulate the *in-vivo* deamidation mechanism by an *in-vitro* approach. An *in-vitro* mechanism could investigate the type of peptides that are targeted by TG2, as well as how and where they are modified. Assuming that this process could take place in a similar fashion to the *in-vivo* mechanism, this could allow the structures of such physiologically relevant peptides to be identified.

3.1.4 Identification of deamidated wheat gluten peptides by HPLC-MS

Each time a deamidation occurs, a glutamine residue is converted into a glutamic acid residue. Therefore, the corresponding molecular weight of the peptide should increase by 1 Da. These small m/z increases of 1 Da per deamidation site (Q → E), as well as the relative assumed small changes in reversed phase LC retention times, could be monitored by HPLC-MS. It has already been determined in Chapter 2 that hundreds of peptides are generated in the gastric/pancreatic *in-vitro* digestion of wheat flour proteins. The objective would entail the identification of those specific peptides that show slightly altered retention times and molecular weights upon interaction with TG2. Subsequent structural information could then be obtained through further MS experiments. It has been determined that the size of the peptides that are targeted by TG2 and that ultimately become bound to the DQ molecules are at least nine to fourteen amino acids in length [4, 17]. Therefore, it

was decided to use an ion trap mass spectrometer to attempt to identify them. Through the use of a low resolution mass analyzer, such as an ion trap, peptides less than thirty amino acids in length can usually be sequenced completely. Limitations to this technique should be noted with relevance to the size of the analyte being studied. The larger the peptide is, the more difficult it becomes to retrieve enough information in order to permit complete sequencing. Fragments that are produced from the collision-induced dissociation (CID) of larger peptides can have a number of charged sites and it becomes increasingly more difficult to determine the charge state of such product ions using a low resolution mass analyzer.

Ion trap-MS offers two main advantages over other MS instrumental configurations in the qualitative identification of unknown peptides, such as those in a peptide digest. The ion trap has the power to obtain better full scan sensitivity and specificity, through the ability to perform MS^n (n=number of stages of tandem MS). One key factor that contributes to an increased level of sensitivity in full scan mode, involves the duty cycle (% of time the ions are sampled and stored in the trap relative to the total MS scan cycle) of the ion trap instrument. The duty cycle in an ion trap is significantly higher (i. e. ~40%), as compared with the low (i. e. ~0.1%) duty cycle offered by a scanning (non-storage) quadrupole-MS instrument. In addition to the ability to spend a longer time sampling ions, the MS^n capability of an ion trap aids significantly in the identification of unknown components by offering greater specificity than just the single m/z value of the parent peptide. Additional stages of tandem MS, (i. e. MS^3), can also be used to obtain more sequence

information about a peptide, which can further improve the ability to achieve complete sequence determination.

3.2 Experimental

3.2.1 In-vitro TG2 catalyzed deamidation and transamidation

3.2.1.1 Chemicals and reagents

Recombinant human tissue transglutaminase 2 was cloned and purified at the Stanford Protein and Nucleic Acid Facility, Stanford University, Stanford, CA, USA. MOPS solution, $CaCl_2$, dithiothreitol (DTT), HPLC-grade acetonitrile, tris and HCl were obtained from Fischer Scientific. Water was obtained from an in-house Milli-Q water purification system (Millipore, Billerica, MA, USA). Putrescine, $KHNaPO_4$ and NaOH were obtained from Sigma-Aldrich (St. Louis, MO, USA). Alexa Fluor 405 cadaverine, dansyl cadaverine and tetramethylrhodamine cadaverine were all obtained from Molecular Probes, a Division of Invitrogen Detection Technologies, Carlsbad, CA, USA.

3.2.1.2 Peptide standards and samples

Synthetic peptides (PFPQPQLPYPQ, QLQPRPQPQLPY, FLQPQQPFPQQPQQPYPQQPQQPFPQ, $PF(PQPQLPY)_3PQPQP$, PQPQLPYPQPQLPY, Ac-PQPELPYPQPQLPY, $LGPGQSKVIG\text{-}CONH_2$, PFPQPQQQF, PQPELPYPQPELPY, $(PQPQLPY)_3$, $(PQPELPY)_3$, $PF(PQPELPY)_3PQPQ$, PQPQPPP, PQQPQQPY, PFSQQQQPV, SQPQQQFPQPQQPQ, $LQLQPF(PQPQLPY)_3PQPQPF$ and KGHHEAELKAL were all synthesized and purified at the Stanford Protein and Nucleic Acid Facility and

verified to be >90% pure by HPLC-MS. These were all supplied by Dr. Chaitan Khosla, Chemistry Department, Stanford University, CA, USA.

Stone ground whole grain corn flour, vital wheat gluten flour, stone ground whole wheat flour (brand 1), whole wheat flour (brand 2), were all obtained at a local supermarket (Raleigh, NC, USA).

3.2.1.3 In-vitro TG2 catalyzed deamidation procedure

Stock solutions of various synthetic peptides (1 mM, in water) were mixed with DTT (1mM) and MOPS solution (5x200mM MOPS + 5mM $CaCl_2$) and maintained at pH approximately 7.5. To this mixture, TG2 solution (1 µM in phosphate) was added and the mixture incubated at 38˚C for various time points from 1 minute to 120 minutes. The reaction was terminated by freezing the mixture in dry ice.

3.2.1.4 In-vitro TG2 catalyzed transamidation and chemical tagging procedure

Stock solutions of various synthetic peptides (1 mM in water) were mixed with DTT (1mM), MOPS solution (5x200mM MOPS + 5mM $CaCl_2$) and chemical tag solution (µM) and maintained at pH approximately 7.5. To this mixture, add TG2 solution (µM in phosphate) and incubate at 38˚C for various time points from 1 minute to 120 minutes. The reaction was terminated by freezing the mixture in dry ice.

3.2.2 In-vitro proteolytic digestion of cereal grains

3.2.2.1 Chemicals and reagents

Pepsin, trypsin, chymotrypsin, carboxypeptidase A, elastase, KHNaPO4 and NaOH were obtained from Sigma-Aldrich (St. Louis, MO, USA). HPLC-grade acetonitrile and HCl were obtained from Fischer Scientific. Water was obtained from an in-house Milli-Q water purification system (Millipore, Billerica, MA, USA). Human recombinant dipeptidyl peptidase IV (DPPIV) was obtained from ProSpec, Rehovot, Israel.

3.2.2.2 Samples

Stone ground whole grain corn flour, vital wheat gluten flour, stone ground whole wheat flour (brand 1), whole wheat flour (brand 2), rye flour, barley flour, stone ground whole grain soy and oat flour were all obtained at a local supermarket (Raleigh, NC, USA).

3.2.2.3 In-vitro proteolytic digestion procedure

Samples of native cereal grains were proteolyzed with pepsin, trypsin, chymotrypsin, elastase and carboxypeptidase A (PTCECA) according to the following protocol:

30 mg of homogenized grain sample was dissolved into 1mL pepsin solution (0.01M HCl adjusted to pH 2), to establish a 1:100 pepsin to protein ratio; heat (at 38˚C) and gently shake for 2 hours. To each sample, add 50 µL of a 50 mM phosphate buffer and 35 µL of a 0.1 M NaOH solution, to establish the pH between 7-7.5. To each sample, add 25 µL of a 1:100 trypsin/chymotrypsin to protein solution

(in 50 mM phosphate). Heat (at 38˚C) and shake for 2 hours. Heat to 95˚C for 15 minutes and cool back down to room temperature. To each sample, add 20 µL of a 1:500 elastase solution (in 50 mM Tris); adjust pH to 7.5 with HCl; heat (at 38˚C) and shake for 2 hours. Heat to 95˚C for 15 minutes and cool back down to room temperature. To each sample, add 25 µL of a 1:100 carboxypeptidase A solution (in 50 mM phosphate); heat (at 38˚C) and shake for 2 hours. Heat to 95˚C for 15 minutes and cool back down to room temperature. Spin down samples and aliquot 200 µL of the supernatant for HPLC-MS analysis.

3.2.3 Synthetic immunogenic peptide standards

Synthetic standards of peptides (LQPQNPSQQQPQEQVPL, PQQSGQGVSQSQQQSQQQ, TQQPQQPFPQQPQQPFPQ, FPLQPQQSF, VPVPQLQPQNPSQQQPQEQVPL, RPQQPYPQPQPQY, QPQQPFPQTQQPQQPFPQ, PQQQFPQTQQPQQPFPQP, QPQQPLPQPQQPF and PQQSPF) were synthesized (Thermo Electron, Ulm, Germany) and were analyzed using HPLC-MALDI-TOF mass spectrometry in order to determine exact molecular weights and purity (>90% pure).

3.2.4 HPLC-ESI-MS instrumentation and analytical conditions

HPLC-MS analyses were performed on an Agilent 1100 HPLC-ion trap mass spectrometer, operated in atmospheric pressure positive electrospray ionization (ESI) mode. RP-HPLC separations were performed by injecting 1-20 µL onto a C18 RP 0.32 x 150 mm column with 3.5 µm particles or Zorbax SB-C18 column (0.5 x 150mm using 3.5 µm particles) or a Zorbax Bonus SB-C18 column (2.1 x 33mm

using 1.5 µm particles). Flow rates of 20 to 250 uL/min were used, along with a gradient elution program of 0-95% ACN + 0.025 % TFA over 10 - 60 minutes.

The ion trap was optimized for mass transmission over the mass range of 300-2200 Da using the standard Agilent tune compound. Under "auto MS^2 " operation, both full scan and MS/MS spectra were acquired, in order to obtain as much information as possible to be able to identify the peptides in the sample. The MS^2 mass spectra were usually acquired from the most intense m/z ion signal within the mass range scanned, as long as its response was above the intensity threshold selected. In order to avoid missing ions from co-eluting peptides, active exclusion was used. This permitted the acquisition of auto MS^2 mass spectra from sequentially lower ion intensities (that are within the inclusion mass range and above the intensity threshold). After each LC peak (~ 5-8 seconds in width) the active exclusion was reset and the most intense ion again selected; this process was repeated. The auto MS^2 mass spectra, obtained from analysis of these PTCECA gluten extracts were used for the identification of possible immunogenic peptides. Additional stages of tandem MS (i. e. MS^3) were used to achieve further sequence information on a detected peptide in order to further increase the probability of its sequence determination.

Some key MS operational parameters needed to be adjusted in order to maximize response. These included: the operation under data dependent auto-MS/MS mode, MS^3, and collision-induced dissociation (CID) conditions using a 30% cutoff (ring electrode voltage) of the parent mass, with an end cap ramp of 0.3 to 2V in 30ms.

3.3 Results and Discussion

3.3.1 TG2 catalyzed deamidation

In order to investigate what types of peptides are targeted by TG2 and how they are modified, the *in-vivo* deamidation mechanism was replicated as closely as possible via an *in-vitro* procedure. The theoretical deamidation mechanism is shown in Figure 3.2. Various small synthetic peptides (listed in section 3.2.1.2), based on 9-33 amino acid sequences were evaluated as targets for the TG2 enzyme. These peptides were chosen as test peptides because they known to exist within wheat α-gliadins. The deamidation reactions were carried out according to the procedure described in section 3.2.1.3. Reaction conditions were chosen in an attempt to reproduce physiological conditions (physiological pH, mM [Ca^{2+}], concentration of peptide (mM) and TG2 (μM)). The reaction products were analyzed by LC-MS.

Each glutamine (Q) residue that was deamidated (thus converting the residue into a glutamic acid (E) with a gain in OH and a loss of a NH_2), the corresponding molecular weight of the peptide was seen to increase by 1 Da. Small m/z increases, as well as slightly longer reverse phase LC retention times, were observed for some peptides. However, the chromatograms of the digested deamidated products were complex, because each deamidation produced an additional LC peak. A simple example illustrating this is shown in Figure 3.3, which represents an LC-MS extracted ion chromatogram (EIC) of the products of the TG2 catalyzed deamidation reaction of an 11mer (PFPQPQLPYPQ). This peptide contains three glutamines. With each successive deamidation, another peak was observed. After a reaction time of 30 minutes, all three glutamines had been converted into glutamic acids. It

can be seen that the single sharp peak at a retention time of 9 minutes, corresponding to the starting material (undeamidated 11mer), had spread out to become a blob consisting of 4 peaks (corresponding to the undeamidated 11mer and 3 more peaks for each deamidated product).

As the size of the test peptides increased (the largest peptide evaluated was a 33mer), there were often more that five or six products all contained in one broad bumpy peak. In addition, small changes in mass (2 or 3 Da) that arose from select deamidations that corresponded to ions with charged states of +2, +3 or higher, were difficult to resolve using a low resolution ion trap mass spectrometer.

Overall, results from these experiments showed that the *in-vitro* deamidation reaction worked, but the yield of products was very low. Also, specificity of the TG2 to certain glutamines was not observed. This is contrary to literature, which reports that specific glutamines that are contained in certain consensus sequences (i. e. **Q**-x-P or PQP**Q**LPY) are those targeted by the TG2 [16, 23, 24]. This raised the concern that perhaps the correct physiological reaction conditions had not been achieved. Lastly, the sensitivity needed to properly sequence the peptides using the acquired MS/MS spectra was found to be insufficient. Therefore, it was concluded that the reaction with TG2 to produce selective deamidation was not selective enough to provide the means to pick out which of the peptides from the hundreds of peptides present in proteolyzed wheat gluten were immunogenic.

3.3.2 TG2 catalyzed transamidation with chemical tagging

In order to improve the selectivity of the deamidated products using HPLC-MS detection, the conditions of the deamidation reaction needed to be optimized.

Also, it was hypothesized that if the specific glutamines that are targeted by the TG2 enzyme could somehow be chemically labeled, this may allow for easier identification. Therefore, the next experiments endeavored to couple the TG2 deamidation reaction with a chemical labeling step. The hypothesis here was that the TG2 enzyme would still attack the same specific Q residues present in immunogenic peptides, but instead of being deamidated (Q → E), the targeted glutamines would be transamidated. This would produce peptides with specific Q's tagged with a chemical functionality that would offer a higher degree of selectivity and/or sensitivity needed for their identification by LC-MS. Figure 3.4 illustrates the proposed reaction scheme for this TG2 catalyzed transamidation + chemical tagging experiment. After forming a thioester bond with the TG2 enzyme in the intermediate complex, an acyl group (from the glutamine acyl donor) is then transferred to the acyl acceptor amine (chemical tag), which forms an amide bond.

The optimal chemical tag used in the chemical labeling step would be one which would possess a UV/fluorescent chromophore that would allow monitoring of an unusual $\lambda_{absorbance}$ or $\lambda_{emission}$ not normally used in peptide analysis. In addition the chemical tag would contain a functionality that would increase MS sensitivity (i. e. amines or pre-charged ions). To this end, four compounds were evaluated as chemical tags: putrescine, monodansyl cadaverine (MDC), tetramethyl rhodamine cadaverine (RhC) and Alexa Fluor 405 cadaverine. Structures and properties of each of these four compounds are listed in Figure 3.5. HPLC-MS analysis of peptides is generally performed in positive ESI mode. It was proposed to evaluate one compound as a chemical tag that would allow the tagged products to be

detected in negative ESI mode. The idea to use negative ESI was that this might provide a superior format in order to simplify the way the tagged peptides are presented. Figure 3.6 shows the proposed reaction scheme using monodansyl cadaverine (MDC) as the chemical tag. MDC has a MW 335 Da, $\lambda_{absorbance}$ = 335nm and $\lambda_{emission}$. = 526nm. The NH_2 group of the specific Q residue in the peptide is targeted by the TG2 enzyme and an intermediate complex forms between the Q, the TG2 active site and the MDC. A stable covalent bond forms between the terminal C=O of the Q residue and the aliphatic part of the MDC probe, releasing NH_3. The product peptide is now tagged with an amine functionality that also has UV/fluorescent specificity.

Synthetic wheat gluten peptides were used again as target peptides. Most experiments focused on two peptides, an 11mer [PFPQPQLPYPQ, MW 1310.7 Da] and a 28mer [PF$(PQPQLPY)_3$PQPQP, MW 3279.7 Da], because these peptides were close to the minimum and maximum molecular weights of the expected immunogenic peptides that would be released from proteolyzed wheat. The reaction was carried out as indicated in section 3.2.1.4. Initial experiments used the 11mer (PFPQPQLPYPQ) as the target peptide. This 11mer contained a known T-cell stimulatory epitope (PQPQLPY). Putrescine was chosen as a chemical tag because it is a primary amine and should improve electrospray positive ion MS sensitivity under acid conditions. Results indicated that the proposed reaction scheme did produce one tagged product [PFPQP(Q-putrescine)LPYPQ], which eluted just prior to the untagged 11mer. However, the yield of this product was less than 5% and its MS intensity was not as high as expected. Therefore, the low yield of the reaction

product coupled with the labor intensive procedure that would be involved to screen the hundreds of peptides present in proteolyzed wheat for the mass addition of putrescine, made this tag a poor choice for further evaluation. Figure 3.7 shows the products of the TG2 transamidation and chemical tagging experiment of the 11mer with putrescine. Figure 3.7(A) shows the products from the TG2 deamidation, where all three glutamines have been converted to glutamic acids (each showing a loss of 1 NH_2). Figure 3.7(C) shows the full scan MS spectrum of the most intense singly deamidated product at retention time 18.8 min. Figure 3.7(B) shows the three tagged products, where each of the glutamines that had been targeted and deamidated by TG2 has now been tagged with a putrescine. Figure 3.7(D) shows the full scan MS spectrum of most intense singly tagged product at retention time 17.6 min.

The next set of experiments evaluated tags that had a unique UV or fluorescence signature, with the idea that these tagged products would be much easier to identify than those tagged with putrescine. In addition, there would be minimal UV background at 330nm (for MDC) and 557nm (for RhC) detected in proteolyzed wheat samples. Figures 3.8(A) and (B) show the LC/UV chromatograms for a 9mer, 11mer and 28mer [PFPQPQLPYPQ, m/z 1310.7 Da; PFPQPQQQF, m/z 1115.5 Da and $PF(PQPQLPY)_3PQPQP$, m/z 3279.7 Da], using the RhC and MDC tags respectively. As expected, the RhC tagged products were about 20 times more sensitive, owing to the higher molar absorptivity coefficient of RhC. It is interesting to note, that based on the three peaks detected for the 28mer, it appears that it received multiple tags, because of the distance between each of the three peaks. Compounds with multiple deamidations elute more closely together. Looking at the

structure of the 28mer, it can be seen that there are 8 glutamines present. Interestingly, it appears that only 3 of the possible 8 glutamines have been targeted by TG2. Literature [16, 23, 24] reports that glutamines in the conserved sequences **Q**xP or **Q**xx (where x could be F, Y, W, M, L, I or V) have been known to be those that are specifically targeted by the TG2 enzyme in known immunogenic peptides. Future work involving the specificity of TG2 will attempt to determine if the glutamines targeted by TG2 in these experiments were the same as those contained in such consensus sequences.

Mass spectra of some of the products from the untagged 11mer as well as those tagged with MDC and RhC, are displayed in Figures 3.9 and 3.10. In each case, only one tag appeared to have been added to the 11mer, even though the compound contained 3 glutamines and had been deamidated twice. Figures 3.9(A) and 3.10(A) show the LC-MS extracted ion chromatograms (EICs) of these 11mer products, while Figures 3.9(B) and 3.10(B) show the full scan mass spectra of the most intense singly tagged MDC and RhC products, respectively. Similar spectra for the 28mer, did show multiple additions of the tags along with multiple deamidations. Figures 3.11(A) and 3.12(A) show the LC-MS EICs for these MDC and RhC products. In each case, three broad peaks can be seen, corresponding to the addition of 1, 2 and 3 tags. Presumably, the same glutamines were tagged in each case. Future work will investigate this further. The peaks appeared wide because each peak contained individual peaks that represented 3-5 deamidated species in addition to the tagged species. Figures 3.11(B) and 3.12(B) show full scan MS spectra of each of these three broad peaks representing the compound with one,

two and three tags. A myoglobin peptide (KGHHEAELKAL m/z 1257.7 Da), used as a control (no Q in its sequence), showed no tagged products indicating the reaction was in fact specific for Q-containing peptides.

Clearly, results from these experiments demonstrated that the chemical tagging approach could work to identify potentially immunogenic gluten peptides. The best UV/fluorescent specificity and MS sensitivity responses were achieved in positive ion ESI mode with MDC and RhC. A fourth compound (Alexa Fluor 405 cadaverine) was also evaluated, but did not succeed in producing tagged peptides for use with negative ion ESI mode.

3.3.3 TG2 catalyzed transamidation and chemical tagging of grain samples

The TG2 transamidation and chemical tagging experiments showed that synthetic celiac active peptides could be tagged *in-vitro* in a similar fashion to that which occurs *in-vivo*. In order to study complex real whole grain gluten samples, a similar approach was taken to evaluate wheat gluten flour.

The wheat gluten was first digested with gastric and pancreatic enzymes, according to the procedure outlined in section 3.2.2.3. Proteolyzed extracts were then reacted with TG2 and either MDC or RhC. Both untagged and tagged product peptides were analyzed by LC/UV-MS. The goal was to end up with unique tagged peptide products which would be identified as potential immunogenic physiologically relevant peptides. Figure 3.13(A) shows a representative LC-MS total ion chromatogram (TIC) of proteolyzed untagged wheat gluten peptides and (B) displays an LC/UV chromatogram @ 330nm, showing peptides that had been tagged with MDC. It was expected that there would be a large number of peptides resulting from

the proteolytic digestion, however, it was not expected that so many tagged products would be produced from the TG2 catalyzed transamidation and tagging reaction. Based on these observations, three questions arose which would have to be addressed in future experiments: (1) Does a lack of specificity of TG2 result in so many tagged products?; (2) Are there really this many immunogenic gluten peptides present in wheat and are there as many present in other grains? and (3) Are the *in-vitro* reaction conditions representative of the *in-vivo* adaptive immune reaction conditions?

Further experiments attempted to address these questions and evaluated other immunogenic grain flours (barley and rye), non-immunogenic grain flours (corn and soy) and oats (as a controversial grain [25, 26]), using the same *in-vitro* reaction scheme. Soy is technically not a grain, but a legume. It was used as a control in this experiment to verify that few, if any, tagged products would be produced. The 28mer synthetic peptide was spiked into all grain samples in order to verify that tagged products could be seen, thus confirming that the reaction was successful. Figure 3.14 shows representative LC-MS TICs for all grains. This figure only shows the latter portion of the chromatographic analysis, where most of the tagged peptides eluted. The highlighted area to the right of the figure (retention time ~ 60 – 73 min.) shows where the tagged 28mer products eluted. The y-axis (MS intensity) of each TIC was normalized to full scale relative to the 28mer products. From evaluating each TIC, it was observed that wheat clearly produced the most tagged peptide products. In fact, there were so many peaks present that the tagged 28mer products were not even able to be identified. Barley, rye and oats also produced many tagged

peptide products, while soy and corn produced the fewest. These TICs also show other reaction products, so it is difficult to assess how many of the peaks might correspond to actual tagged immunogenic grain peptides.

In contrast, Figure 3.15 shows representative LC/UV chromatograms of the RhC-tagged peptides from each grain. Again, 28mer tagged peptide products can easily be identified in all but the wheat flour sample. Each grain chromatogram appears to have several common peaks, probably due to other reaction products. These peaks appear at the approximate retentions times: 31, 40, 47, 48.5, 49, 51, 56 and 57 minutes, respectively. However, it can be observed that, apart from the common reaction products, soy and corn appear to have very few tagged products, while the chromatograms representing oats, rye and barley show many peaks. Wheat clearly contains the most tagged peptide products.

Two of the concerns that had been posed earlier in this section have now been addressed. It can be concluded from these observations that it does appear that other grains (barley, rye and perhaps oats) do produce tagged peptide products, with wheat clearly producing the most. The specificity of the TG2 enzyme has not been fully determined by these experiments, but it is generally apparent upon comparison of the amount of tagged peptides from corn and soy, that they do not produce anywhere near as many tagged peptide products as barley, rye and wheat. Therefore, there must be some level of specificity being maintained by TG2, under these *in-vitro* reaction conditions. Searching the NCBI protein database, it was determined that corn and soy proteins also have many glutamines and prolines, but peptides from these proteins did not appear to become tagged with RhC.

Further experiments were performed in order to evaluate the *in-vitro* reaction conditions. Various whole grains were used as samples in the MDC and RhC tagging reaction. Soy was used as a control. In particular, the amount of TG2 relative to the amount of grain, the length of time that TG2 is allowed to react with the proteolyzed peptides, pH, variations in the yield of tagged peptides with the type of TG2 (i. e. human versus guinea pig) as well as tagged peptide response in the various grains were all investigated in order to better define the specificity of TG2 *in-vitro*. The most noteworthy observation from these experiments was that for a 30 mg sample of proteolyzed grain, approximately 5 µM TG2 with mM [Ca^{2+}] and 1-2 mM RhC (MDC) were found to be sufficient to produce the desired tagged products. It was observed that excess RhC (MDC) started to appear after approximately 15 minutes of reaction time. That would infer that the reaction had become quenched and that there were no more immunogenic peptides left in the sample to tag. Data to support this observation is shown in Figure 3.16. A large peak at a retention time of 20 minutes can be seen in the LC-MS TICs of corn, soy, oats, rye and barley. Interestingly, two different brands of wheat were analyzed and an excess of the RhC chemical tag was not observed in either case. The time point (15 minutes) where the transamidation reaction appeared to be quenched is within the range of what literature reports as the approximate time (10-30min) for symptoms of gastric stress to begin to appear in an individual reacting to the ingestion of immunogenic gluten. Notwithstanding, it should be noted that *in-viv*o digestion of food varies from person to person and with how much and what type of grain proteins are ingested. Further investigation along these lines is outside the scope of this research project.

Therefore, it was sufficient to assume that under these general conditions, it is feasible to release and identify some potentially immunogenic wheat gluten peptides. Data from re-analysis of wheat and corn with MDC are displayed in Figure 3.17(A) and (B) respectively. The large peak at a retention time of 22 min. indicates the excess of unreacted MDC that had not been observed under the conditions used in previous analyses. Perhaps these optimized conditions were better suited to allow only the immunogenic peptides to become tagged.

3.3.4 Determination of potential immunogenic peptide molecular weights and sequences.

Data from the analyses of different sources of wheat flour and gluten were then examined in detail, in order to attempt to determine which UV and/or MS peaks corresponded to potential immunogenic wheat peptides and to determine their molecular weights and structures. The process began by focusing on the most intensely responding UV peaks, resulting from the proteolysis and TG2 transamidation and RhC (MDC) tagging of wheat gluten. These are shown in Figure 3.18, where (A) represents RhC-tagged wheat peptides and (B) represents the corresponding MDC-tagged wheat peptides. The sixteen most intensely responding peaks are numbered. These were the peaks that were chosen in order to try to identify the corresponding peptides. Figure 3.18(B) shows two peaks that have multiple entries [i. e. 6(1, 2) and 8(1, 2)]. These represent LC/UV peaks 6 and 8 that each consist of two co-eluting peptides. In Figure 3.18(A), it can be seen that these four peptides do not co-elute. The differences in retention characteristics between these peptides may be a result of the fact that they are tagged with different

numbers of MDCs or RhCs attached to that particular them that result in their co-elution / or separation. The synthetic 28mer peptide [PF(PQPQLPY)$_3$PQPQP] was spiked into all samples as a control to ensure that the reaction was successfully producing tagged products. Peaks corresponding to various tagged products of the 28mer were identified and are labeled in Figure 3.18(A) and (B).

In order to successfully identify potential immunogenic peptides, a complicated data analysis process was developed that required peaks from both the LC/UV chromatograms and MS/MS spectra (corresponding to both untagged and tagged peptides) to be interpreted manually, as well as through database search algorithms. A summary of the progression of this process is shown in a flowchart format in Figure 3.19. Data from the analysis of untagged proteolyzed wheat, MDC and RhC-tagged peptides and from the analysis of other grains were all used in order to confirm the presence/absence of each proposed peptide sequences in immunogenic and non-immunogenic grains.

Each of the 16 peaks was investigated individually. This process is described in detail, as follows, for the identification of the sequence of the peptide that corresponded to the LC/UV peak #3. This peak is depicted at a retention time of 35.5 minutes in Figure 3.18(B). The initial step in the identification process was to obtain MS and MS/MS (if possible) spectra of the MDC-tagged peptide. For peak #3, these are shown in Figure 3.20. Figure 3.20(A) shows the LC/UV chromatogram acquired @ 330nm, where peak # 3 is shown at a retention time of 35.5 minutes; (B) an EIC showing that there is an ion at m/z 759 which corresponds to this UV peak; (C) full scan MS of the UV peak at 35.5 min., showing a base peak at m/z 759.3 and

another very small peak at m/z. 1139 and (D) the MS/MS spectra of the m/z 759.3 ion. The two ions at m/z 759.3 and 1139 correspond to $[M+2H]^{2+}$ and $[M+3H]^{+3}$ ions. Based on these two ions, the mass of the MDC-tagged parent ion should be 2275 Da. From previous work with synthetic immunogenic peptides (33 residues or less), a compound of this MW would likely have 1, 2 or 3 MDC tags. The MW of one MDC is 317 Da. Since the number of MDC tags for this compound is not known, the theoretical molecular weights of the various ions corresponding to the untagged version of this peptide can be postulated as follows:

- $[M+H]^{+}$ of m/z 1959; $[M+2H]^{+2}$ of m/z 980 (corresponding to the peptide tagged with 1 MDC)
- $[M+H]^{+}$ of m/z 1642; $[M+2H]^{+2}$ of m/z 821.5 (corresponding to the peptide tagged with 2 MDCs)
- $[M+H]^{+}$ of m/z 1325; $[M+2H]^{+2}$ of m/z 663 (corresponding to the peptide tagged with 3 MDCs)

The next step involved mining the LC-MS data from the proteolyzed wheat samples (no tags) for one or more of these ions. Figure 3.21 shows three extracted ion chromatograms from the LC-MS data of the untagged proteolyzed wheat samples that would represent $[M+2H]^{+2}$ ions of a peptide tagged with one (A), two (B) and three (C) MDCs. The top trace (A) shows a peak at m/z 980, which corresponds to the +2 charge state for a peptide tagged with 1 MDC. Interestingly, in addition to the LC-MS data of proteolyzed wheat showing a peptide of molecular weight 980 Da ($[M+2H]^{+2}$), a peak was also observed at slightly higher m/z (in the mass range m/z 980-982) in the LC-MS data from proteolyzed wheat that had also

been treated with TG2 (resulting only in deamidated peptide products), which looked a bit broader and increased slightly in retention time (Figure 3.22). This suggests that a deamidated version of the peptide of MW 980 (RT 23.3 min,) eluted at retention time 23.5min.

Additional verification of the molecular weight of this unknown peptide was achieved by repeating the above process using the RhC data. This involved obtaining MS and MS/MS (if possible) spectra of the corresponding RhC-tagged peptide represented by peak #3 in Figure 3.18(A). Figure 3.23 shows results from RhC data that is analogous to what is depicted in Figure 3.20. Figure 3.23(A) displays an EIC showing that there was an ion detected at m/z 819, corresponding to the UV peak at a retention time of 36.6 minutes; (B) displays the full scan MS of this UV peak, showing a base peak at m/z 819.5 and another peak at m/z. 1228.4; (C) displays the MS/MS spectra of the m/z 819.5 ion. The two molecular ions at m/z 819.5 and 1228.4 correspond to $[M+3H]^{+3}$ and $[M+2H]^{+2}$ ions, respectively. Based on these two ions, the mass of the RhC-tagged parent ion should be 2456 Da. An untagged proteolyzed wheat peptide of molecular weight 1958 with one RhC would have a molecular weight of 2456 Da. Under positive electrospray ionization conditions, a peptide of that MW would most likely exhibit a +2 and /or a + 3 ion at m/z 1228 and 819 respectively, as seen in Figure 3.23. In addition, a small peak was observed at a retention time of 54.5 minutes. The MW of this peak corresponded to the addition of two RhC tags to this peptide.

Next, it was verified that this peptide is not found in corn (a non-immunogenic grain). This was accomplished by the digestion of corn flour in a similar fashion to

that done with wheat and comparing the results. Figures 3.24, 3.25 and 3.26 show the extracted ion current chromatograms of the $[M+3H]^{+3}$ ion at m/z 759 in corn and wheat (representing the peptide with one MDC tag); the $[M+2H]^{+2}$ ion at m/z 980 in corn and wheat (representing the peptide with no tag); and the $[M+3H]^{+3}$ ion at m/z 819 in corn and wheat (representing the peptide with one RhC tag), respectively. These figures demonstrate that this peptide is indeed unique to wheat and would be an excellent candidate to be considered as an immunogenic target peptide.

The final step was to sequence this candidate peptide. The MS and MS/MS spectra were first sent to the Mascot MS/MS ion search program (Matrix Science) to retrieve any possible known sequences that corresponded to this peptide in wheat. The search conditions specified the taxonomy searched to be Viridiplantae (green plants). Other parameters of note specified that the data came from an ion trap mass spectrometer; the peptide could have deamidations and no enzymatic cleavages. This allowed the program to search each protein for a subset (peptide) that best matched the data file entered. Figure 3.27(A) shows results from the Mascot protein database search for the identity of the peptide corresponding to UV peak #3, indicating the sequence that was the top hit belongs to an alpha/beta gliadin. Figure 3.27(B) and (C) show the MS and MS/MS mass spectra of the peptide along with the proposed primary sequence ions. The postulated sequence for this peptide is LQPQNPSQQQPQEQVPL and it was definitely considered as a candidate for synthesis, which would serve as final confirmation of the peptide's identity.

A summary of all the information about each of the 16 most intensely responding LC/UV peaks (corresponding to those denoted in Figure 3.18), is shown

in Table 3.2. Ten peptides from this list (sequences in bold) appeared to fit all the criteria to be considered potentially immunogenic. These peptides were subsequently synthesized in order to confirm their proposed sequences and MS responses.

Some problems were encountered with the data interpretation within this phase of the research. These problems involved reduced efficiency of the separation of the proteolyzed and tagged peptide mixture. It was found that improved LC separation of these peptides was required because the existing LC separations, such as the one shown in Figure 3.28, were not sufficient to adequately resolve all the eluting components. This made data analysis and interpretation quite difficult. MS analysis also revealed that most LC peaks were found to be composed of several co-eluting components. Figure 3.28 illustrates the type of data complexity that was encountered in the determination of the identity of peptide 4 (RPQQPYPQPQPQY), which corresponded to one of the compounds that eluted within UV peak 15 in the LC/UV chromatogram of the MDC-tagged peptides [see Figure 3.18(B)]. Part (A) of Figure 3.28 shows peak 15, which eluted at approximately 62 minutes. Part (B) shows the EIC of one compound at a m/z 912.4, which corresponded to a $[M+3H]^{+3}$ ion from a compound of MW 2914 Da. Part (C) shows the EIC of another compound at a m/z 858.7, which corresponded to a $[M+4H]^{+4}$ ion from a compound of MW 3426 Da. Part (D) shows the full scan MS spectrum of all the ions that were found under the first half of this UV peak (61.7 to 62.0 minutes). The ions in green correspond to the compound shown in (B), while those in red belong to the compound shown in (C). The two triply charge ions shown

in black belong to other compounds that co-elute under this UV peak (MW 3873 and MW 4246).

On occasion, it was not apparent which of these components were the peptides that were tagged and which were not tagged. The presence of untagged compounds reduced the ultimate usefulness of the proposed approach because they hindered the identification of the tagged peptides. Co-eluting compounds can reduce the MS signal of the tagged peptides due to electrospray ion suppression as well as by reduced ion trap accumulation times. This reduction in MS signal reduced the ability to detect the tagged peptides and resulted in less intense MS/MS spectra, making sequencing difficult and database searching inefficient.

The process of determining the peptide molecular weight in PTCECA proteolyzed wheat gluten, as compared to the corresponding peptides tagged with MDC or RhC, did not always result in a molecular weight that matched. It was discovered that when several tags were present in a compound, the tagged peptide molecular weight could increase to the high 2000's and ultimately up to 3500 Da. The electrospray ionization mass spectrum of these tagged peptides consisted of +3 or greater charged states and resulted in an average molecular weight determination for that peptide, since the isotopic peaks for that peptide could not be resolved on the ion trap. However, the monoisotopic mass was usually eventually determined, because the PTCECA digested gluten peptides usually produced +1 or +2 ions, whereby the capabilities of an trap could resolve these isotopes. This could result in a 1-3 Da mass discrepancy when comparing average molecular weight to monoisotopic molecular weight. Additionally, the chemical tagging process which

utilized the catalytic power of the TG2 enzyme, often resulted in several deamidated sites. These additional sites were found to be present as well as those that were involved in tagging with the MDC or RHC. This resulted in small increases in molecular weight of the tagged peptides, which were passed onto the postulated molecular weights of the peptides in the proteolyzed gluten. To account for this complexity when searching for peptides in proteolyzed gluten, EICs covered a 2-4 m/z mass range (versus a usual single m/z). If multiple peaks were found initially, work focused on the peak with greatest abundance first, then moved through all available ions down to the peaks with lower abundance.

3.3.5 Confirmation of proposed target peptide identity

Synthetic standards were prepared of the ten target gluten peptides [(1) LQPQNPSQQQPQEQVPL, (2) TQQPQQPFPQQPQQPFPQ, (3) VPVPQLQPQNPSQQQPQEQVPL, (4) RPQQPYPQPQPQY, (5) QPQQPFPQTQQPQQPFPQ, (6) PQQSPF, (7) QPQQPLPQPQQPF, (8) PQQQFPQTQQPQQPFPQP, (9) PQQSGQGVSQSQQQSQQQ and (10) FPLQPQQSF] identified by the enzymatic proteolysis and chemical tagging experiments. This was done in order to both confirm the identity of each peptide and its LC-MS response. The peptides were synthesized (Thermo Electron, Ulm, Germany) and then analyzed using HPLC-MALDI-TOF mass spectrometry in order to determine their exact molecular weights and purity (>90% pure). MALDI-TOF spectra, representing the purity of two of the synthetic peptides (VPVPQLQPQNPSQQQPQEQVPL and RPQQPYPQPQPQY), are shown in Figures 3.29 and 3.30.

These ten synthetically prepared peptides along with proteolyzed wheat gluten samples previously analyzed, were analyzed by HPLC-MS/MS according to the same conditions that have been used to analyze all proteolyzed wheat gluten peptides thus far. Data from the analyses of the synthetic peptides were compared to that for the corresponding peptides produced via the *in-vitro* proteolysis procedure. Positive identification of the sequence composition for each potential peptide was based on: (1) LC retention time, (2) full scan MS spectrum (showing the same multiply charged ions in the same ratio) and (3) the MS/MS CID spectra (showing ions from at least two parent to product transitions, resulting in the same ion ratios), respectively.

Based on these analyses, sequences for 7 of the 10 proposed potentially immunogenic wheat gluten peptides were confirmed. These are listed in Figure 3.31(A). Part (B) shows an LC-MS/MS multiple reaction monitoring (MRM) chromatogram, where the most intense of two parent to product transitions monitored for each of the peptides 1-10 is displayed from the analysis of a solvent standard (top) and from the analysis of a proteolyzed wheat sample (bottom). Note that the sequences of peptides 1-6 were confirmed. It was discovered that the postulated sequences for peptides 8, 9 and 10 did not match those of the synthetically prepared peptides. Peptide 7 (QPQQPLPQPQQPF) was found to have been synthesized incorrectly, as QPQQPLPQPQQFF. It was re-synthesized correctly, as QPQQPLPQPQQPF, and this sequence was subsequently confirmed as a target peptide. The sequence for peptide 8 (PQQQFPQTQQPQQPFPQP) was found to be incorrect. The fact that it co-eluted with three other peptides (2, 5, and 7)

may have contributed to this misidentification of the sequence. Peptide 9 (PQQSGQGVSQSQQQSQQQ) did not chromatograph well. It was not well retained and eluted with the solvent front. No further work was directed to this peptide in lieu of the others that were successfully identified.

The peptide identification process was difficult and was definitely not fool proof. There were instances where MDC- and RhC-tagged peptides indicated a particular molecular weight that should be observed in PTCECA gluten. It was later discovered that the peak chosen that matched those molecular weights was found to be incorrect. For example, peptide 10 (MW 1091 Da), an example of a very abundant peptide, proved to be a gluten related peptide based upon the Mascot database search. The proposed sequence was FPLQPQQSF. However, only five of the nine amino acids could successfully be sequenced, based upon the MS/MS spectra. A complete sequence was not obtained because the low mass information was lost as a result of the cutoff value used for the CID process (ions below 30% of the parent ion were not detected). When the synthetic standard of peptide 10 was analyzed, it did not match the retention time of its corresponding peak in proteolyzed gluten. Upon additional investigation a very weak (5-10% relative intensity) peak, 22 Da lower was detected, indicating the original peak at m/z 1092 was a sodium adduct. Looking for the corrected molecular weight (1068.7 Da) with 1-3 MDC or RhC tags resulted in no significant peaks detected. Then a question was raised about whether the RhC- and MDC-tagged peak also contained sodium. Additional work will be required to address the situation with this peptide, in order to attempt to sequence the sodium adduct and determine if sodium adducts are observed in the

MDC or RhC mass spectra. This was an unusual observation because no other sodium adducts were observed with the other peptides. Since this peak was in the lower end of UV absorbance in the MDC chromatogram, no further work was directed to this peptide in lieu of the others that were successfully identified.

Figures 3.32 - 3.38 represent the LC-MS full scan mass spectra and LC-MS/MS CID product ion mass spectra for each of the peptides 1 though 7. The MS/MS fragmentation frequently observed was a result of cleavage at the N-terminal side of prolines, thus forming b-type product ions. The most intense b-type ions appeared to favor cleavage on the N-terminal side of “QP” sequences. The y-type product ions also favored cleavage at the N-terminal side of P, with the charge remaining on the P-containing product ion. Other weaker product ions (both b- and y-type) were observed for cleavage between Qs. These peptides can now be used as markers for wheat gluten content in the analysis of food products. It was proposed to use these peptides as target analytes and develop an LC-MS/MS assay which could be used to detect and quantify the presence of any of these peptides. This would essentially represent the presence of wheat, if any were detected in a sample of a food product.

Comparing the sequences of these seven peptides with those from several known immunogenic gluten epitopes, previously identified by various other researchers [7, 27], it is important to note that all seven peptides were found to contain partial sequences also found in many of these established DQ-restricted immunostimulatory T-cell epitopes. Some examples of these common epitopes are PYPQPQ, QPFPQQP, SQQQP, FPQQP, PQQSPF, PQQPQQP, RPQQPYPQ,

QQPQQPFPQ and QTQQPQQPFPQ. Therefore, each of these seven peptides can effectively be considered immunogenic. It has been reported that there is a high variability in the consistency and frequency of TG2 recognition of specific glutamine residues between T-cell lines investigated to date [28]. Similar sites are targeted, but these sites are not always targeted in every patient's T-cells. Further interpretation and sequencing information may provide a better insight into the specificity of TG2 *in-vitro* (i. e. which Q residues in these seven peptides are preferentially tagged by the TG2 enzyme). This will be discussed in section 3.3.7.

3.3.6 Confirmation of peptides released by direct in-vitro enzymatic digestion and HPLC-MS analysis of wheat flour

Although the seven peptides had already been proven to be released from wheat flour (previous experiments), it was felt that re-analysis was required in order to confirm that the peptides were present in various different brands of commercially available wheat flour and wheat gluten. Different brands of grain product may consist of different varieties of wheat. Corn flour was also analyzed again in order to verify the absence of the peptides. The i*n-vitro* proteolytic digestion procedure worked successfully and reproducibly, releasing all seven of the target peptides in the samples where wheat protein was present. As was found in previous analyses, all seven peptides were found to be present in wheat gluten and wheat flour, but were not found to be present in corn flour.

Another experiment tested the ability of the seven peptides to resist further proteolytic digestion. A cocktail of all seven standards were digested with pepsin, trypsin, chymotrypsin, elastase, carboxypeptidase A (PTCECA) according to same

protocol that has been used throughout these experiments. A further treatment of human brush border membrane (BBM) enzyme DPPIV was carried out in order to further assess any degree of degradation that occurred. Figure 3.39 shows LC-MS/MS MRM chromatograms for peptides 1-7 in: (A) 10 ng/mg solvent standard before proteolysis and (B) after being digested with the enzymes for 30 minutes. Interestingly, the peptides did not show much degradation. One of the characteristics of known immunogenic peptides is that they tend to remain intact after their respective proteins have been digested down to basic amino acids, di- and tri-peptides. From Figure 3.39(B), it is observed that peptides 2, 4 and 6 did slightly degrade, however, none of the peptides showed any significant (>20%) degradation. This observation bodes well to support the theory that these seven peptides are in fact immunogenic.

3.3.7 Further investigation involving the specificity of TG2

Following the successful identification of seven immunogenic gluten peptides, questions remained with respect to the level of specificity that TG2 shows when exposed to these peptides. Knowing the variability in the consistency and frequency of TG2 recognition of specific glutamine residues between T-cell lines, it was interesting to investigate the underlying relationship between TG2 and the glutamine residues within the context of this research project. Detailed specificity studies on TG2 and its ability to target certain glutamines in certain consensus sequences [16, 22 and 29], report that the reaction is somewhat specific for certain sequences, but not uniquely specific. Specificity also has been shown to vary between individuals. One short sequence, (**Q**-x-P in epitopes such as PQP**Q**LPY) has been reported to

have shown preferential specificity by the TG2 enzyme for the indicated **Q** residue contained in some T-cell stimulatory epitopes, but these were not chosen exclusively.

Experiments attempted to determine the location of the MDC and RhC tags within the identified seven peptides in order to obtain a better understanding of the selective process that TG2 followed under the experimental conditions of this research. Initial experiments demonstrated that the TG2 catalyzed transamidation reaction did show a level of specificity for certain glutamines contained in known immunogenic epitopes of various synthetic peptides. Figure 3.40 shows RhC-tagged products for several of these reference peptides. The extracted ion chromatograms for all peptides except the non Q-containing myoglobin peptide (bottom) showed RhC-tagged and deamidated products. It is interesting to note that the two glutamic acid (E) residues contained in the myoglobin peptide did not become tagged. Yet, this experiment does not pinpoint the exact site or amino acid sequences that were deamidated and/or tagged in the Q containing peptides. These reference peptides, as well as the peptides identified form this research, contained more Q's than were tagged. Typically only 1-3 Q's were tagged in the peptides (<33 amino acids in length) studied here.

These MDC- and RhC-tagged peptides were then sequenced, in order to attempt to determine the site of tagging in each, and if these sites were consistent within other peptides. Figures 3.41 and 3.42 each show the full scan mass spectra and MS/MS mass spectra for the 9mer (PFPQPQQQF) peptide tagged with one MDC or RhC tag respectively. Based upon the sequencing information provided

from the MS/MS mass spectra, it can be seen that although this peptide has 3 glutamine residues, in both cases the same site became tagged. This site was identified as the Q shown in bold PFPQP**Q**QQF. There appeared to be no difference in the specificity when the reaction was carried out with the MDC or RhC tag.

Figures 3.43 and 3.44 each show the full scan mass spectra and MS/MS mass spectra for the 11mer (PFPQPQLPYPQ) peptide tagged with one MDC or RhC tag. There are also 3 glutamine residues in this peptide and again the same site was tagged in each case. The site that was tagged was identified as the Q shown in bold PFPQP**Q**LPYPQ. Allowing the reaction to continue long enough (over 2 hours), results in all three Q's eventually becoming tagged (Figure 3.43). However, a large preference for the middle Q is clearly indicated.

A larger 28mer peptide was also studied, because it is near the upper end of the mass range (MW 3282 Da) typical of known immunogenic wheat gluten peptides. Even though the sequence of this 28mer contains 8 glutamines [PF(PQPQLPY)$_3$PQPQP], it only appeared to yield 3 tagged products corresponding to the addition of 1, 2 and 3 tags. Figures 3.45 - 3.47 show the MS/MS mass spectra for the sequencing of the 28mer with 1, 2 and 3 RhC tags, respectively. The 28mer sequence contains three PQPQLPY epitopes. Each successive RhC tag attached to the same Q, (PQP**Q**LPY). The first RhC tag conjugated to the middle PQPQLPY sequence, the second tag to the C-terminal epitope and the third to the N-terminal epitope.

The 7 peptides identified in sections 3.3.4 and confirmed in section 3.3.5 were also investigated further in order to determine their site(s) of tagging. Figure 3.48

shows the sequencing of peptide 6 (PQQSPF) tagged with RhC. Upon evaluation of the MS/MS spectra in Figure 3.48(C), it was determined that the **Q** in bold was the site that was tagged in peptide 6 (P**Q**QSPF). Figure 3.49 shows: (A) EIC from proteolyzed wheat flour, that represents $[M+3H]^{+3}$ ion for peptide 2 with 2 MDCs (MW 2786); (B) full scan mass spectrum for peptide 2 (TQQPQQPFPQQPQQPFPQ) tagged with 2 MDC tags and (C) MS/MS product ion spectrum of the $[M+2MDC]^{+3}$ ion of m/z 929.9. The sequencing revealed that the two **Q**'s in bold were tagged sites (TQQP**Q**QPFP**Q**QPQQPFPQ). Figure 3.50(A) shows the full scan mass spectrum for peptide 3 at a retention time of 55 minutes, in an LC/UV chromatogram of proteolyzed wheat gluten. The sequence information, as seen in part (B), indicated the **Q**'s in bold (VPVPQLQP**Q**NPSQQQP**Q**EQVPL) were those that were tagged with the MDCs.

Table 3.3 summarizes the findings about the specificity of TG2. Sites targeted by TG2 are shown in bold. It appeared that the reaction conditions and perhaps lower steric hindrance for the MDC relative to RhC, favored more MDCs binding to the peptides relative to RhCs. However, despite the fact that more MDC s tended to bind than RhCs, those sites where both where found to bind, were identical in each peptide.

Literature has reported that the position of prolines and glutamine residues within gluten peptides is important for determining whether the peptide will be a good substrate for DQ binding [12, 16 and 29]. In fact, a recent study [29] has compiled a list of characteristic binding patterns from known immunogenic DQ-restricted peptides. Figure 3.51 displays these common binding signatures for DQ2 and DQ8

and also shows that several of the seven peptides identified in this research share many of these DQ-binding characteristics. This provides further evidence to suggest that they would be good candidates to bind to DQ molecules.

Upon further review of the sites tagged with RhC / MDC, several common 7 amino acid motifs could be recognized. Table 3.4 lists these common motifs and the peptides that contain them. Two interesting observations can be made. First, the most common site observed that was targeted for chemical tagging was "P_1 x_1 P **Q** Q P_2 x_2 ", where P_1 is usually a proline, but can be a glutamine. P2 can be a proline or a glutamine and can be an E when P_1 is a Q. x_1 can be a Q or F and x_2 can be a Q, F, Y or V. Second, the other common sequence which was observed was "P_3 Q P **Q** Z_1", where P_3 can be an L or absent; Z_1 can be the tri-peptide PLY, NPS or QPL.

Strong tagged product responses were also observed from peptides with no Q-x-P sequences, such as QP**Q**, **Q**Qx and **Q**QQ. It is believed that these consensus sequences have yet to be reported in literature. Additional experiments comparing human and guinea pig TG2 were conducted under the same reaction conditions to determine if an animal TG2 model could be adapted to study this human immune mechanism. Results were not as successful using the guinea pig TG2 because the guinea pig TG2 did not show the same level of specificity.

3.4 Conclusions

Previous work with recombinant gliadin proteins, by various researchers, has led to the discovery of two highly inflammatory, physiologically relevant, multivalent 33- and 26-residue gluten oligopeptides [1, 24]. In turn, those peptides have proven to be useful markers of gluten toxicity and they continue to be involved in on-going

studies by a number of other researchers. Motivated by those findings, I wished to identify additional physiologically relevant gluten peptides that could be used as markers in complex foods. The ability to determine the presence/absence of such peptides with high sensitivity and specificity in complex foods would provide an overall indication of the toxicity of the product to those with gluten sensitivities.

To this end I endeavored to study native cereal grains, using an innovative approach involving chemical tagging of wheat gluten peptides with tissue transglutaminase 2. This approach used LC-ESI-ion trap-MS^n to identify and characterize several potential immunogenic peptides. Complete sequences and sites of TG2 deamidation for 7 wheat gluten peptides were successfully determined using this novel enzymatic digestion/chemical tagging methodology. Confirmation of the identity of each of the peptides was determined by comparing structure and molecular weight with synthetically prepared peptides. HPLC-MS analysis as well as protein database searches confirmed that these peptides were all present in wheat and not present in non-immunogenic grains (i. e. corn and rice). Figure 3.52 shows results from one protein database search where the sequence corresponding to peptide 1 was found in three different species of wheat and in different classes of wheat gluten proteins. Interestingly, this sequence for peptide 1 is conserved in each case (residues 28 - 44). These 7 peptides were detected by LC-MS/MS with better sensitivity than the αG-33 (immunodominant wheat gluten 33mer peptide), which may have implications for future applications of this work to analyze food for possible immunogenic wheat gluten contamination.

In order to be able to perform quantitative detection of trace quantities of these peptides in native and processed food, an analytical method would now have to be developed that would provide the best possible sensitivity, accuracy and specificity. In order to attain these requirements, the existing HPLC-ion trap MS method will be transferred to a triple quadrupole mass spectrometer, which has the capabilities to provide better accuracy and sensitivity than an ion trap, when quantitative detection is required.

DRB1	DQA1	DQB1	serological haplotype	celiac heterodimer
03	0501	0201	DR3-DQ2	DRB1*03-DQB1*0201-DQA1*0501
07	0201	0202	DR7-DQ2	DRB1*07-DQB1*0202-DQA1*0201
11/12	0505	0301	DR5/DQ7	DRB1*11/12-DQB1*0301-DQA1*0505
04	0301	0303	DR4-DQ8	DRB1*04-DQB1*0302-DQA1-0301

Table 3.1

HLA haplotypes for the development of CD

	MDC-tagged wheat			RhC-tagged wheat			Proteolyzed (untagged) wheat					
UV peak#	RT (min)	# MDC	MW (Da)	RT (min)	# RhC	MW (Da)	RT (min)	m/z	%Rel Int	peptide #	peptide MW	proposed sequence
1	24.2	1	967	ND	ND	ND	ND	ND	ND		ND	NS
2	27.5	1	990	27.9	1	1170	ND	ND	ND		671	NS
3	35.5	1	2275	36.5	1	2457	28	980.6(+2)	71.2	1	1959	**LQPQNPSQQQPQEQVPL**
	62	3	2914	54.5	2	2955	28	980.6(+2)	ND		1959	LQPQNPSQQQPQEQVPL
4	38.5	1	1516	43.7	1	1695	17.5	1198.6(+1) / 599.3(+2)	2.7		1197	QQPQQTFPQP / TQQPQQPFPQ
5	39.4	1	1021	34	1	1200	22.8	703(+1)	17.1	6	703	**PQQSPF**
6(1)	46.6	1	1309	40	1	1488	ND	ND	ND		991	NS
6(2)	46.6	2	1121	38.7	1	982	ND	ND	ND		485	NS
7	48.3	2	2594	46.7	2	2953	23.3	979.1(+2)	77.4	9	1957	**PQQSGQGVSQSQQQSQQQ**
8(1)	49.4	2	2786*	42.1	1	2648	31.3	1075.8(+2)	15.4	2	2150	**TQQPQQPFPQQPQQPFPQ**
8(1)	49.4	2	2786*	42.1	1	2648	31.3	1075.8(+2)	15.4	5		**QPQQPFPQTQQPQQPFPQ ****
8(1)	49.4	2	2786*	42.1	1	2648	31.3	1075.8(+2)	15.4	8		**PQQQFPQTQQPQQPFPQP ****
8(2)	49.4	2	1857*	41.7	1	2036	28.6	767.2 (+2)	7.6	7	1532	**QPQQPLPQPQQPF**
9	51	2	1730	36.8	1	1592	38.4	1091.7(+1)	100	10	1091	**FPLQPQQSF**
10	52.9	2	1966	50.9	2	2325	ND	ND	ND		1329	NS
11	54.7	3	1563	35.3	2	1604	ND	ND	ND		610	NS
12	55.4	2	1748	55.9	2	2107	ND	ND	ND		1112	NS
13	60.4	2	2084	57.6	2	2443	43	725.2(+2); 1449 (+1)	4.9		1448	VQQQIPVVQPSIL
14	62	3	3434	54.5	2	3476	39.8	827.7(+3); 1240.8(+2)	69.5	3	2478	**VPVPQLQPQNPSQQQPQEQVPL**
15	62.4	3	2576	48.4	2	2624	22	814.3(+2); 1627(+1)	28.3	4	1626	**RPQQPYPQPQPQY**
16	48.2	1	1728	43.2	1	1907	43.8	705.7(+2)	6.5		1410	SQQPQQPFPQPQ

NOTE:
* observed m/z 929.8 (MW 1857) assumes +2 charges ion and MW 2786 assumes a +3 charged ion
** other possible sequences for UV peak 8. There appeared to be at least 2 or 3 coeluting peptides in this peak
NS = sequence was not able to be determined
UV peak#s 6(1) + 6(2) and 8(1) + (2) represent co-eluting peptides
sequences in **BOLD** represent peptides that met all criteria to be considered immunogenic
ND = not detected

Table 3.2

Summary of potentially immunogenic peptides identified in proteolyzed wheat

Synthetic Peptide Standards	Sequence	# MDC	# RhC
9 mer	IQP**Q**QPAQI	1	1
9 mer	PFPQP**Q**QQF	1	1
11 mer	PFPQP**Q**LPYPQ	1	1
12 mer	PFPPFPQP**Q**QQF	1	1
26 mer	FLQP**Q**QPFP**Q**QPQQPYPQQP**Q**QPFPQ	3	3
28 mer	PF(PQP**Q**LPY)3PQPQP	3	3
Myoglobin	KGHHEAELKAL	0	0
Identified peptides			
1	LQP**Q**NPSQQQP**Q**EQVPL	3	2
2	TQQP**Q**QPFP**Q**QPQQPFPQ	2	1
3	VPVPQLQP**Q**NPSQQQP**Q**EQVPL	2	1
4	RP**Q**QPYPQP**Q**PQY	3	2
5	QP**Q**QPFPQTQQP**Q**QPFPQ	2	1
6	P**Q**QSPF	1	1
7	QP**Q**QPLPQP**Q**QPF	2	1

note: **Q** (in bold) represent sites tagged with MDC or RhC

Table 3.3

Specificity of TG2 in wheat gluten peptides

Amino acid sequence	Peptides with this sequence *
IQP**Q**QPA	9 mer
PQP**Q**QQF	PFP
PQP**Q**PLY	11 mer, 28 mer
PFP**Q**QPQ	26 mer, peptide 2
QQP**Q**QPF	26 mer, peptide 2 and 5
LQP**Q**QPF	26 mer
QQP**Q**QPF	peptide 5
LQP**Q**NPS	peptide 1 and 3
QQP**Q**EQV	peptide 1 and 3
PQP**Q**PQY	peptide 4
RP**Q**QPYP	peptide 4
PQP**Q**QPF	peptide 7
QP**Q**QPLP	peptide7

Note:

Q (in bold) are those targeted by TG2 and tagged by MDC and RhC

* refer to Table 3.3 for complete sequences

Table 3.4

Amino acid sequences that show sites of TG2 deamidation and tagging with MDC and RhC

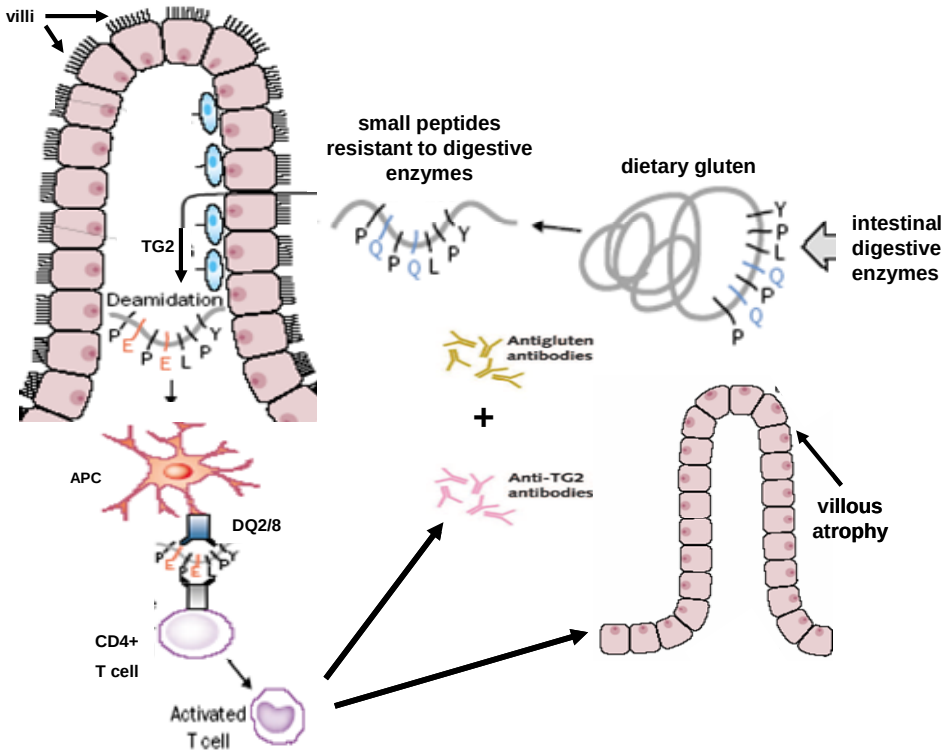
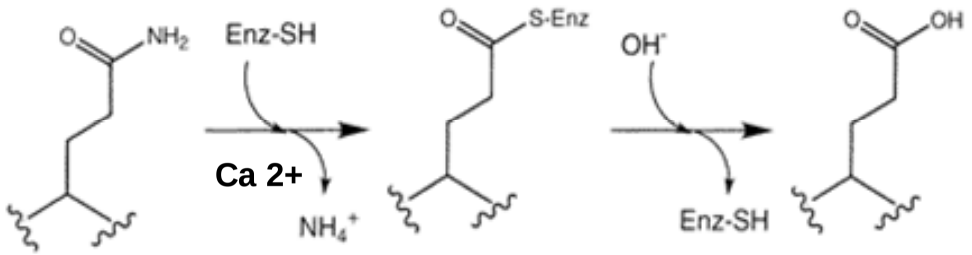


Figure 3.1

Overview of the adaptive immune reaction mechanism in celiac disease

(adapted from Mowat [8])

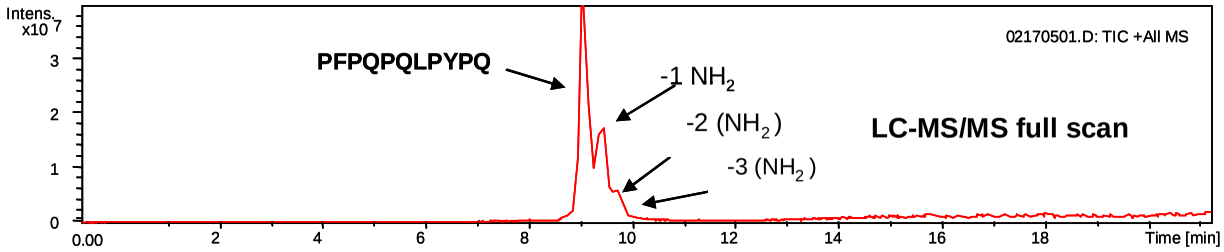
deamidation reaction : glutamine (Q) → glutamic acid (E)



glutamine **glutamic acid**

Figure 3.2

Deamidation of glutamine by TG2



Peptide	# of deamidations	Retention Time (min)	$[M+H]^+$ (Da)
PFPQPQLPYPQ	0	9.0	1311.7
PFPQP**E**LPYPQ	1	9.4	1312.8
PFP**E**P**E**LPYPQ	2	9.7	1313.7
PFP**E**P**E**LPYP**E**	3	9.9	1314.7

Note:

E (in bold) are sites of TG2 catalyzed deamidation

Figure 3.3

TG2 catalyzed deamidation of PFPQPQLPYPQ (11mer)

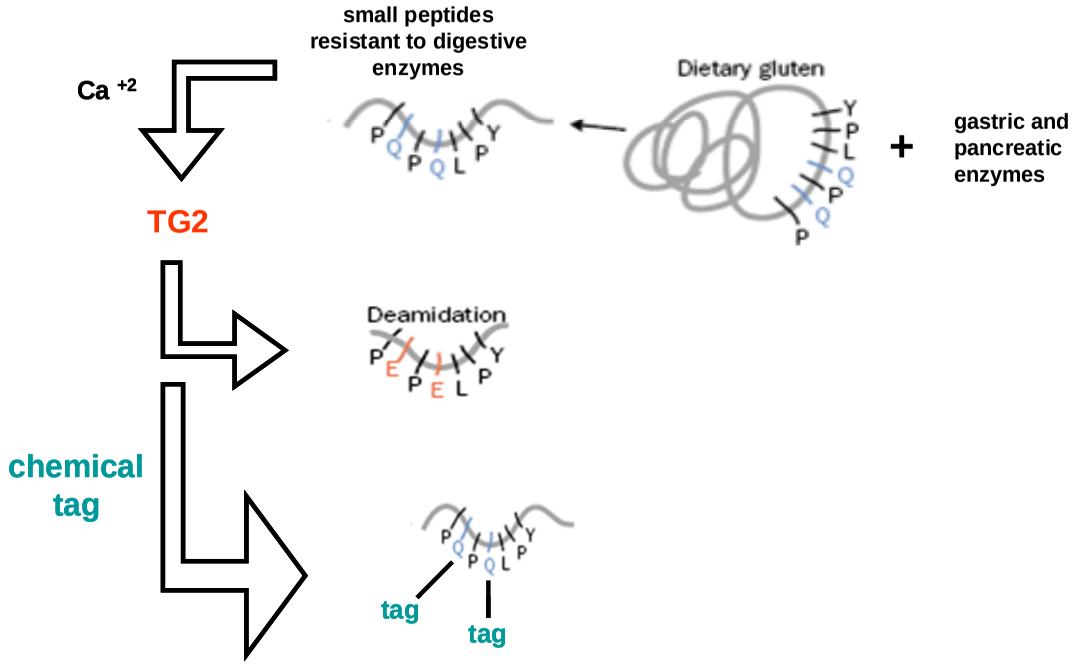
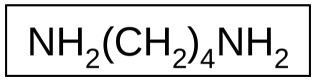


Figure 3.4

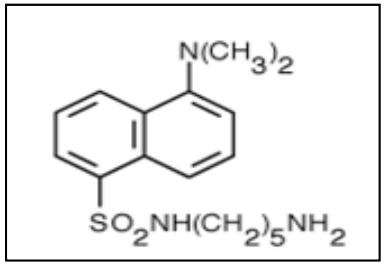
Proposed TG2 catalyzed transamidation and chemical tagging reaction

putrescine



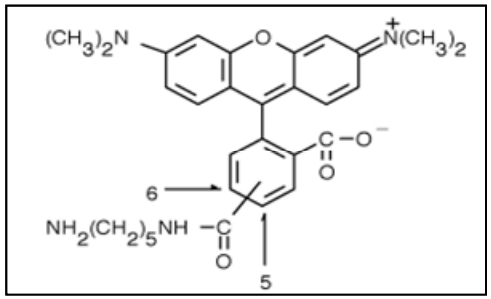
MW 88 Da
no chromophore

monodansyl cadaverine (MDC)



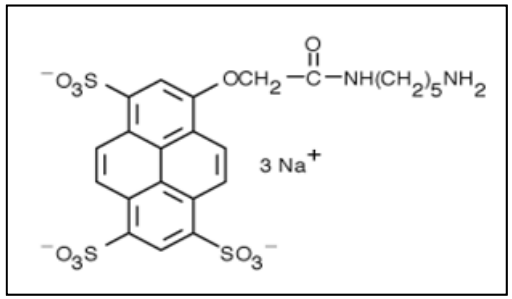
MW 335 Da **UV λ_{abs} 335nm**
Fluor λ_{em} 526nm

tetramethyl rhodamine cadaverine (RhC)



MW 515 Da **UV λ_{abs} 557nm**
Fluor λ_{em} 581nm

Alexa Fluor 405 cadaverine



MW 667 Da **UV λ_{abs} 402nm**
Fluor λ_{em} 421nm

Figure 3.5

Compounds evaluated as chemical tags in the TG2 catalyzed transamidation reaction experiments

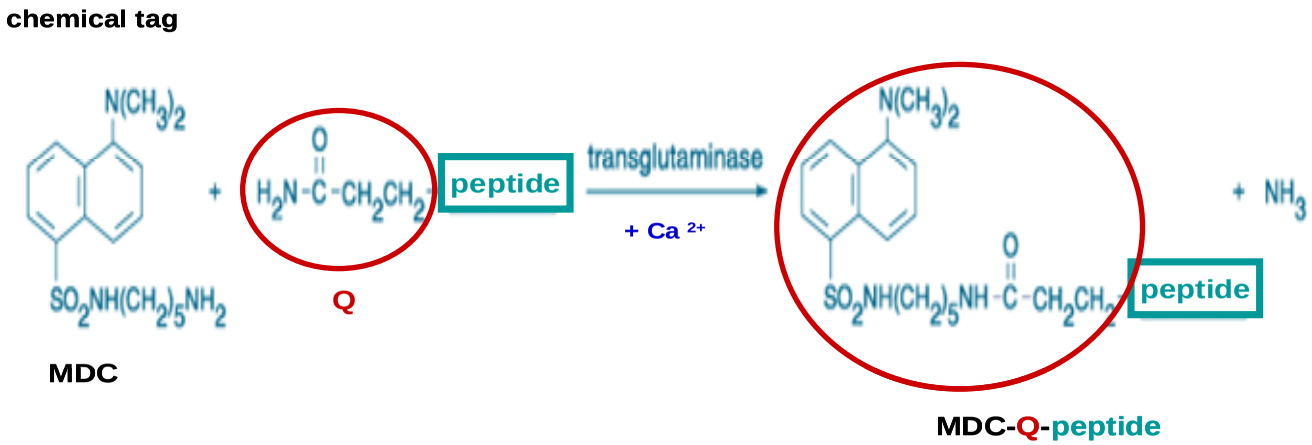


Figure 3.6

TG2 catalyzed transamidation + chemical tagging reaction mechanism with MDC

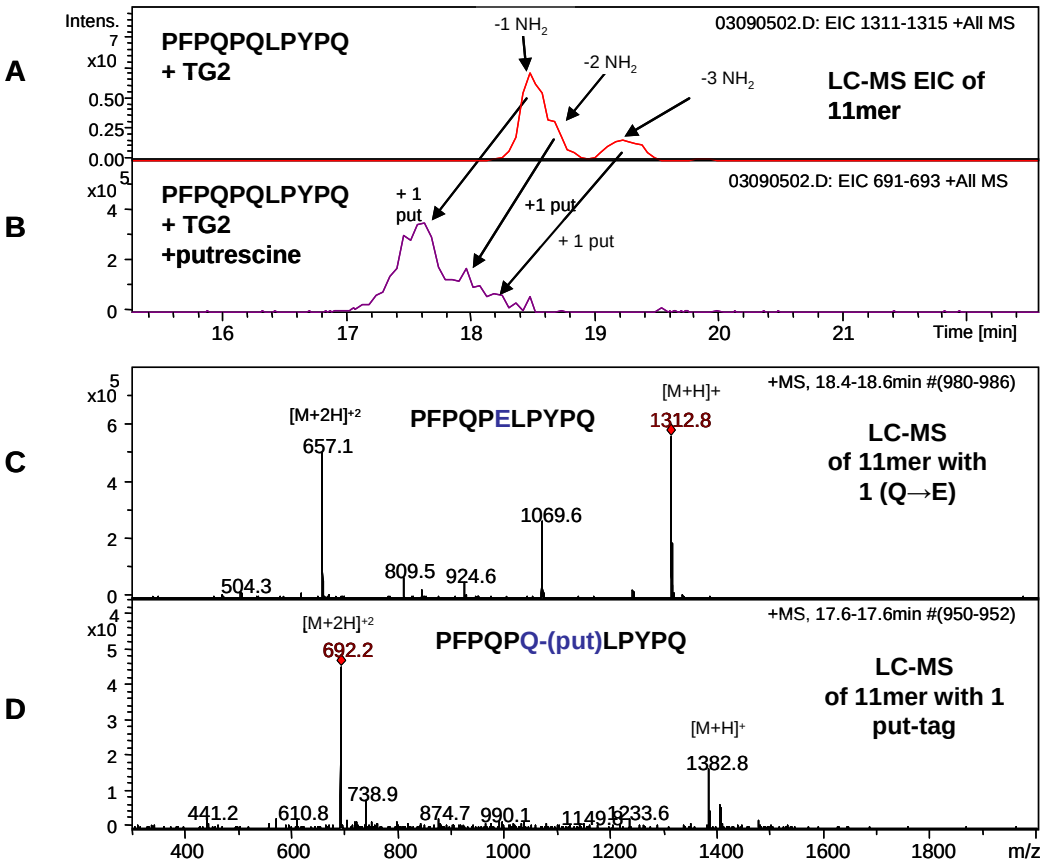


Figure 3.7

TG2 catalyzed transamidation of PFPQPQLPYPQ with putrescine

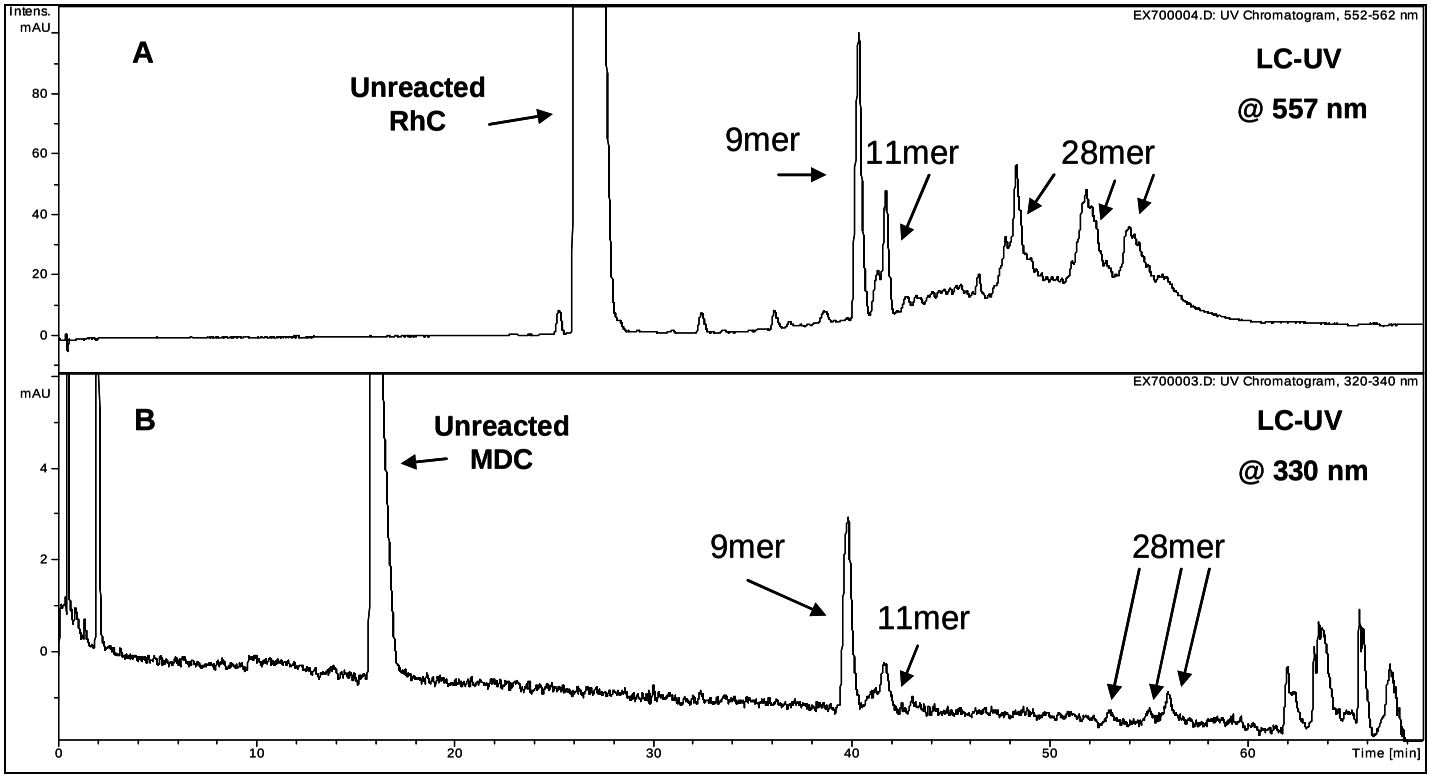


Figure 3.8

Products of the TG2 catalyzed transamidation of a 9mer, 11mer and 28mer with monodansyl (MDC) and tetramethyl rhodamine (RhC) cadaverines

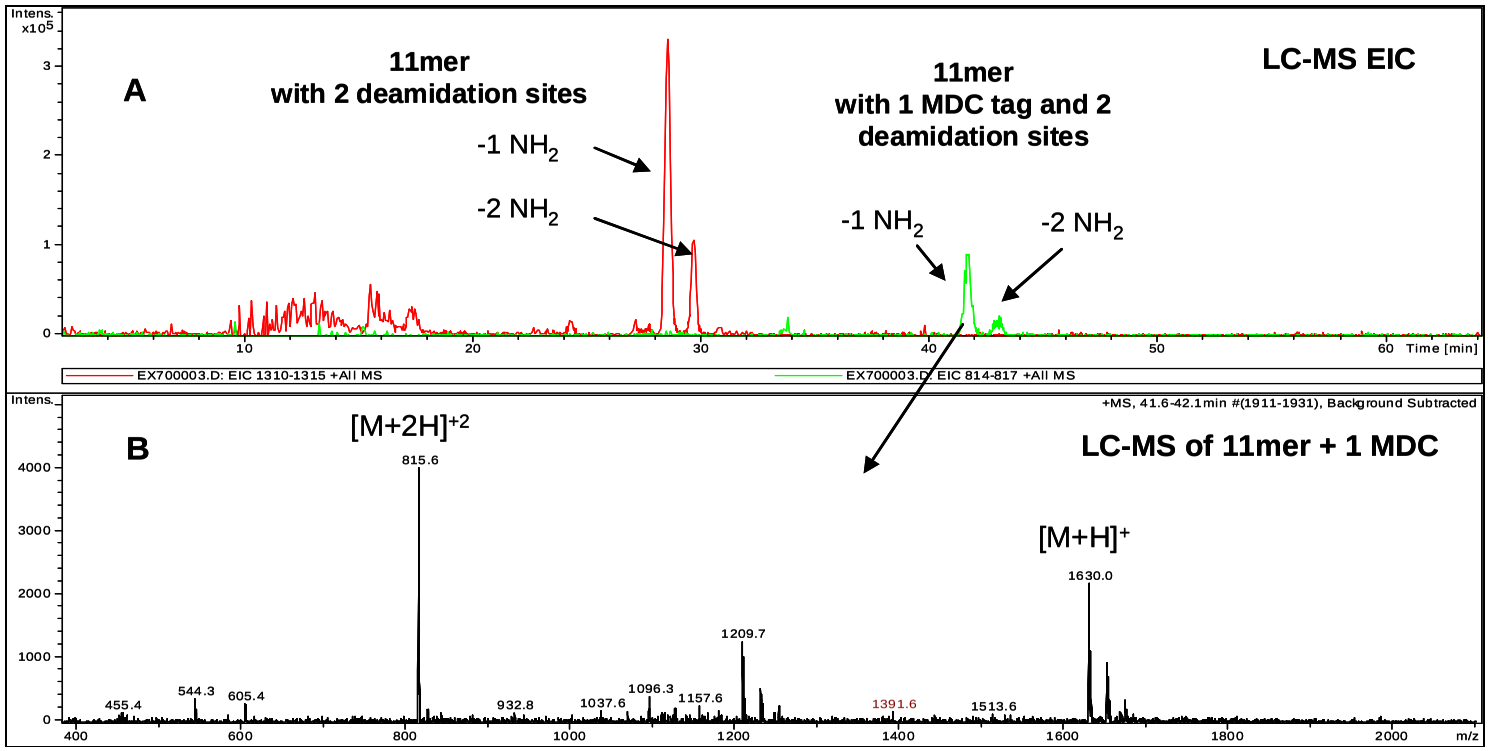


Figure 3.9

11mer products from deamidation and transamidation with MDC

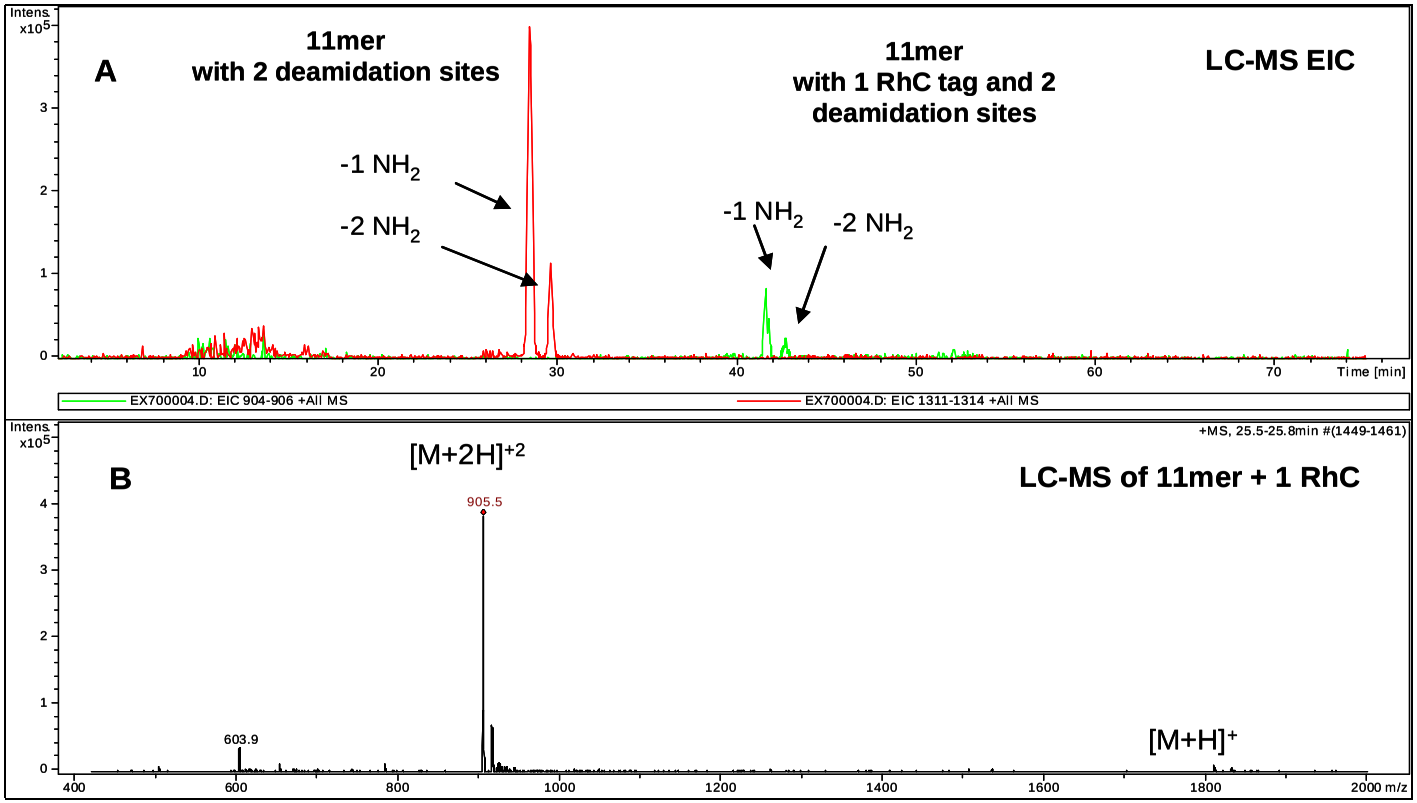


Figure 3.10

11mer products from deamidation and transamidation with RhC

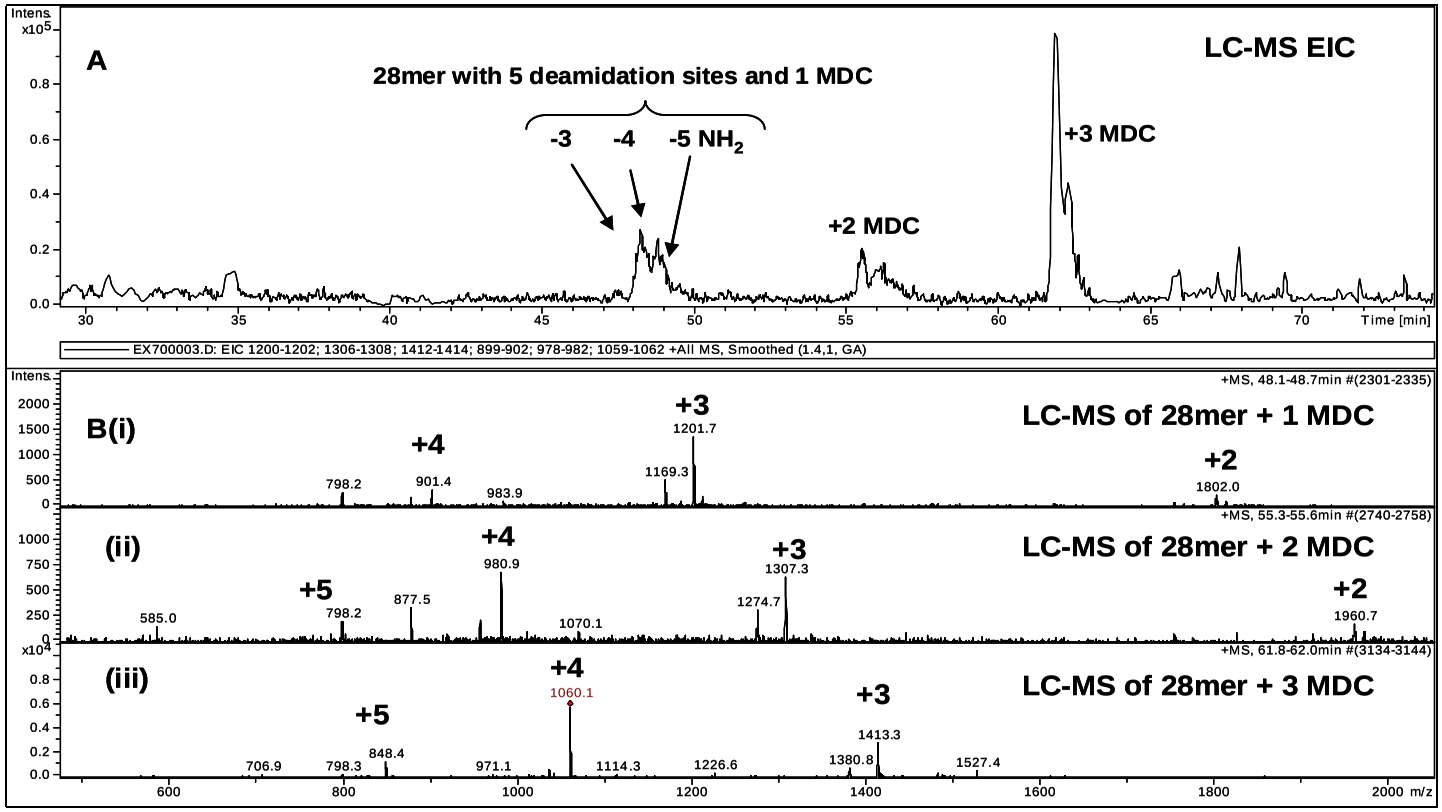


Figure 3.11

28mer products from deamidation and transamidation with MDC

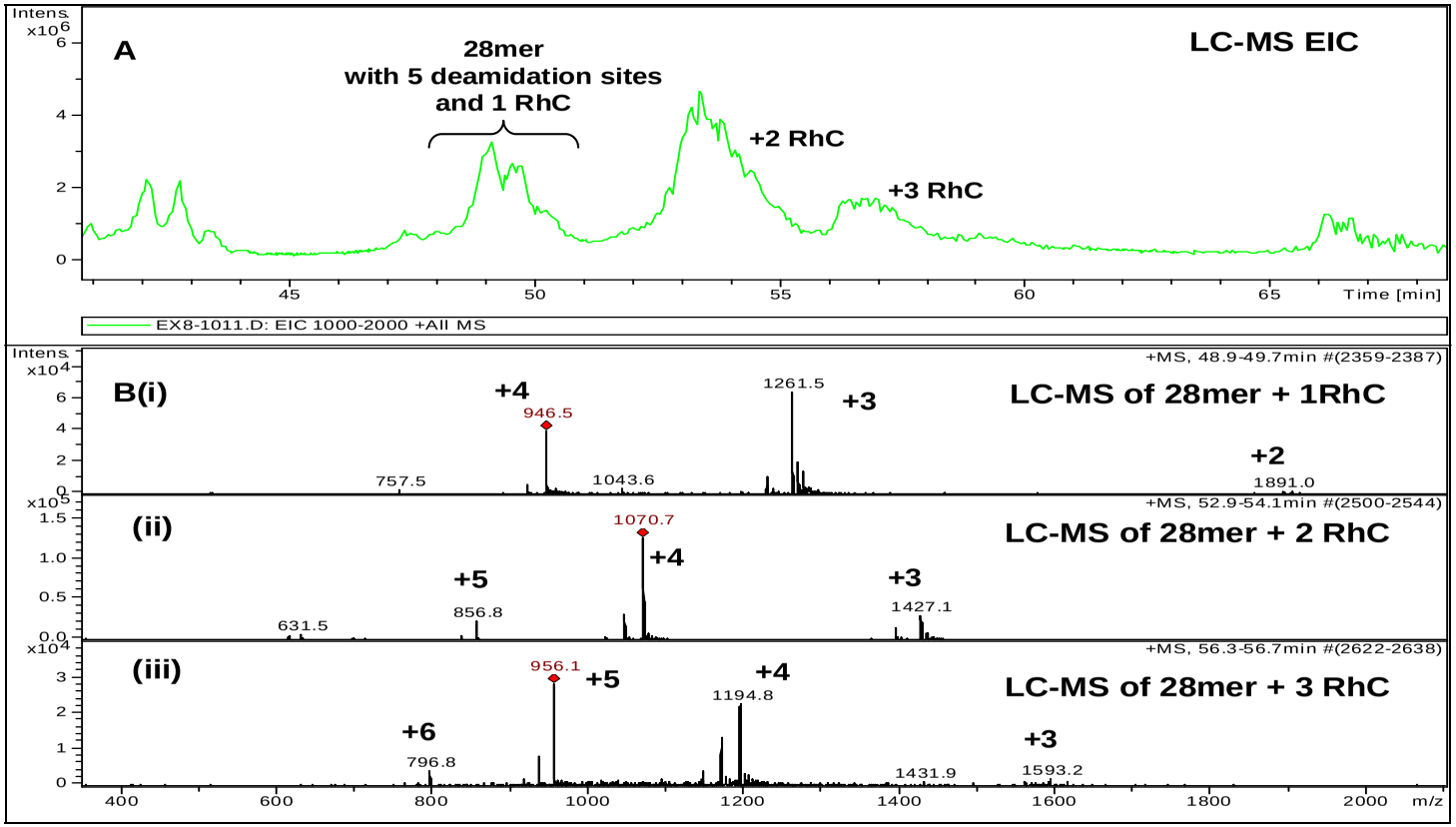


Figure 3.12

28mer products from deamidation and transamidation with RhC

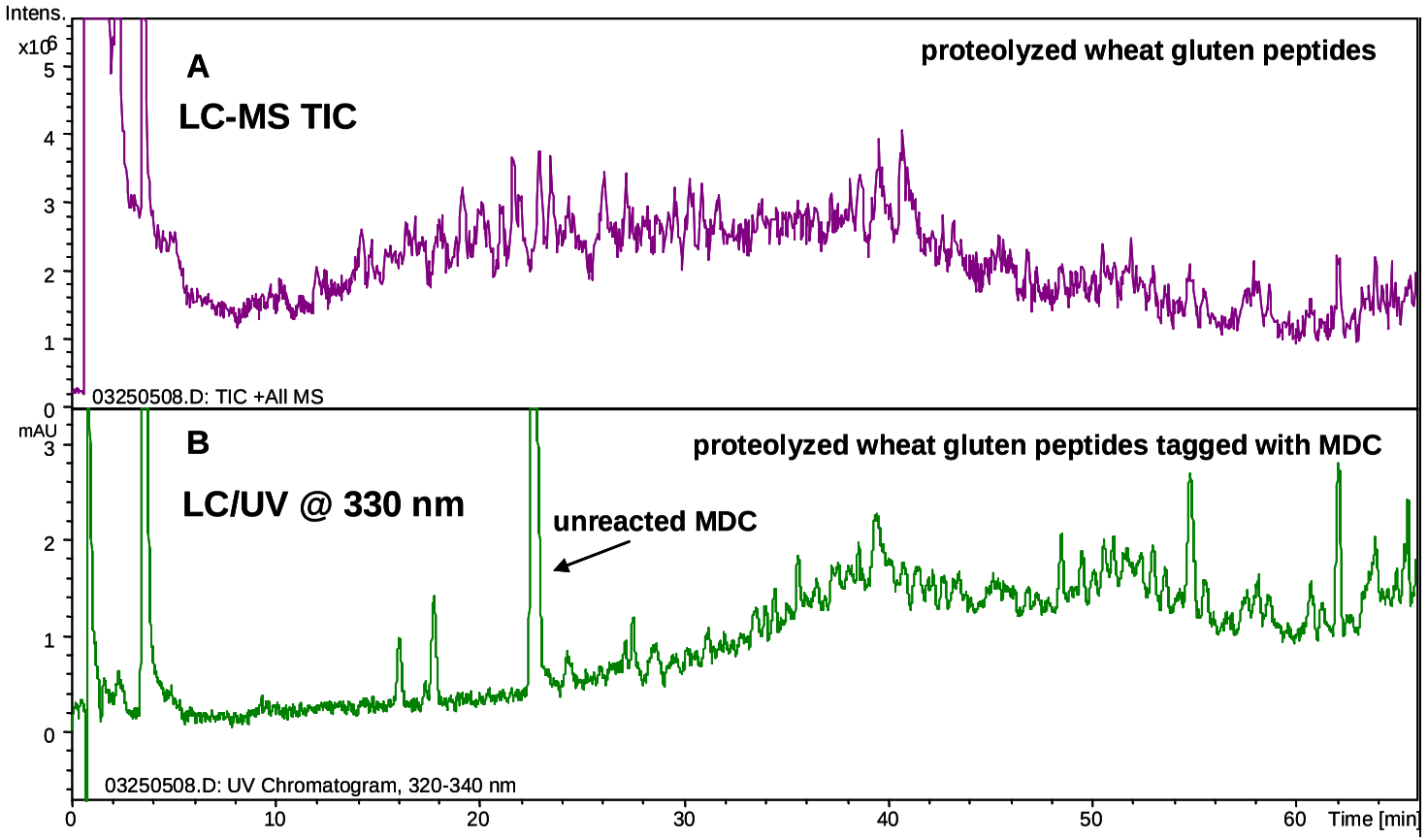


Figure 3.13

HPLC/UV-MS analysis of proteolyzed wheat gluten

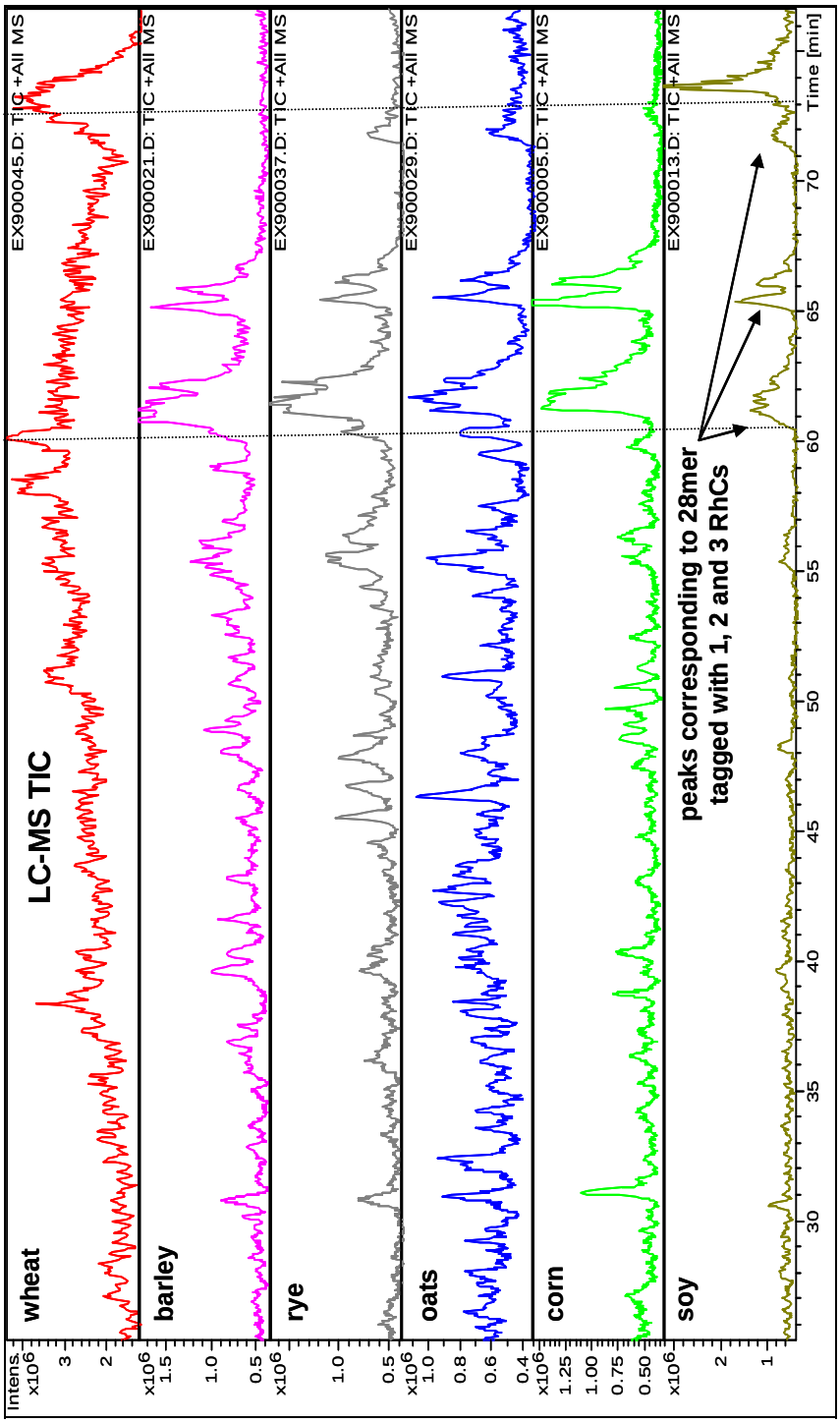


Figure 3.14

LC-MS analysis of RhC-tagged peptides from various proteolyzed grains

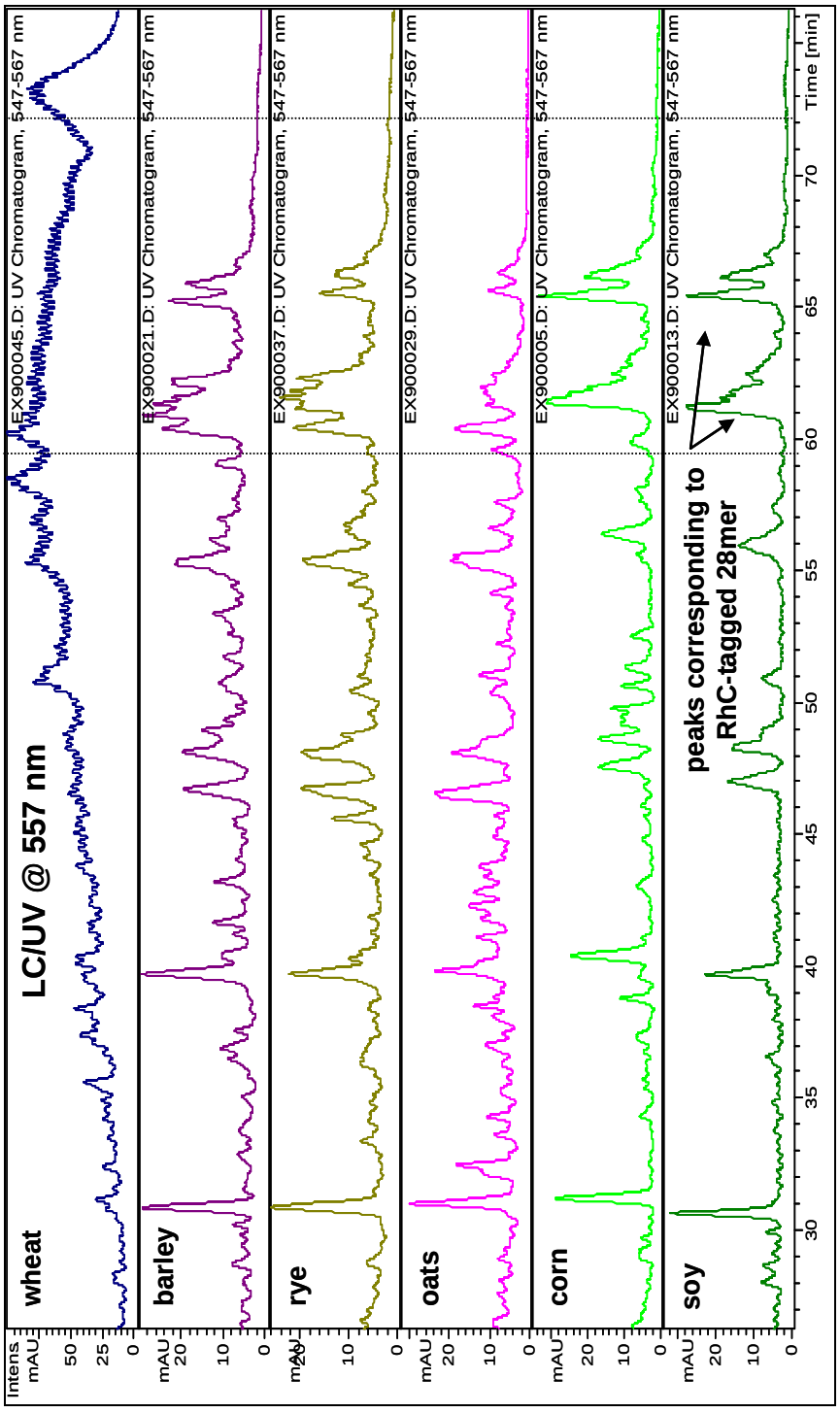


Figure 3.15

LC/UV analysis of RhC-tagged peptides from various proteolyzed grains

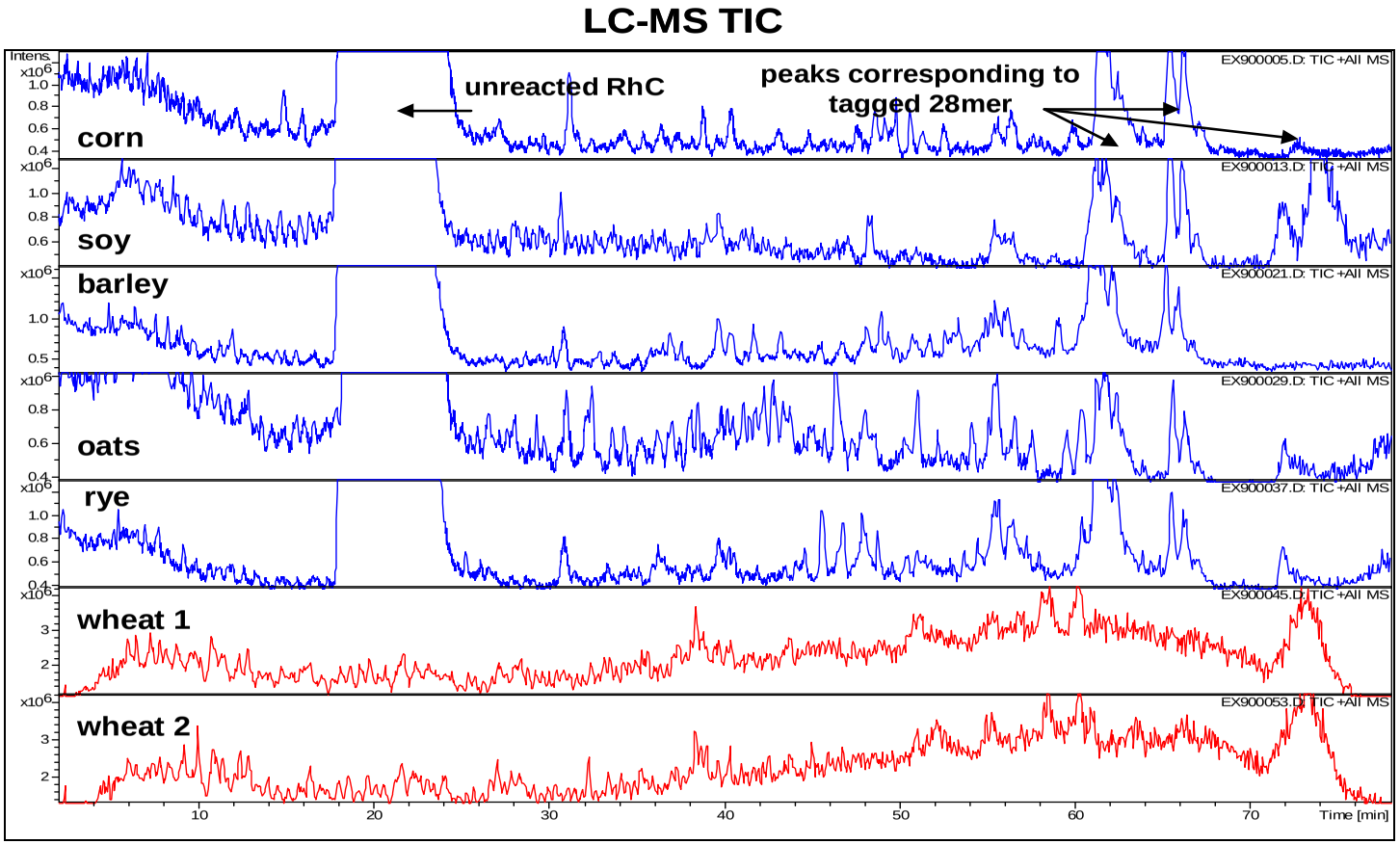


Figure 3.16

LC-MS analysis of RhC-tagged peptides from various proteolyzed grains

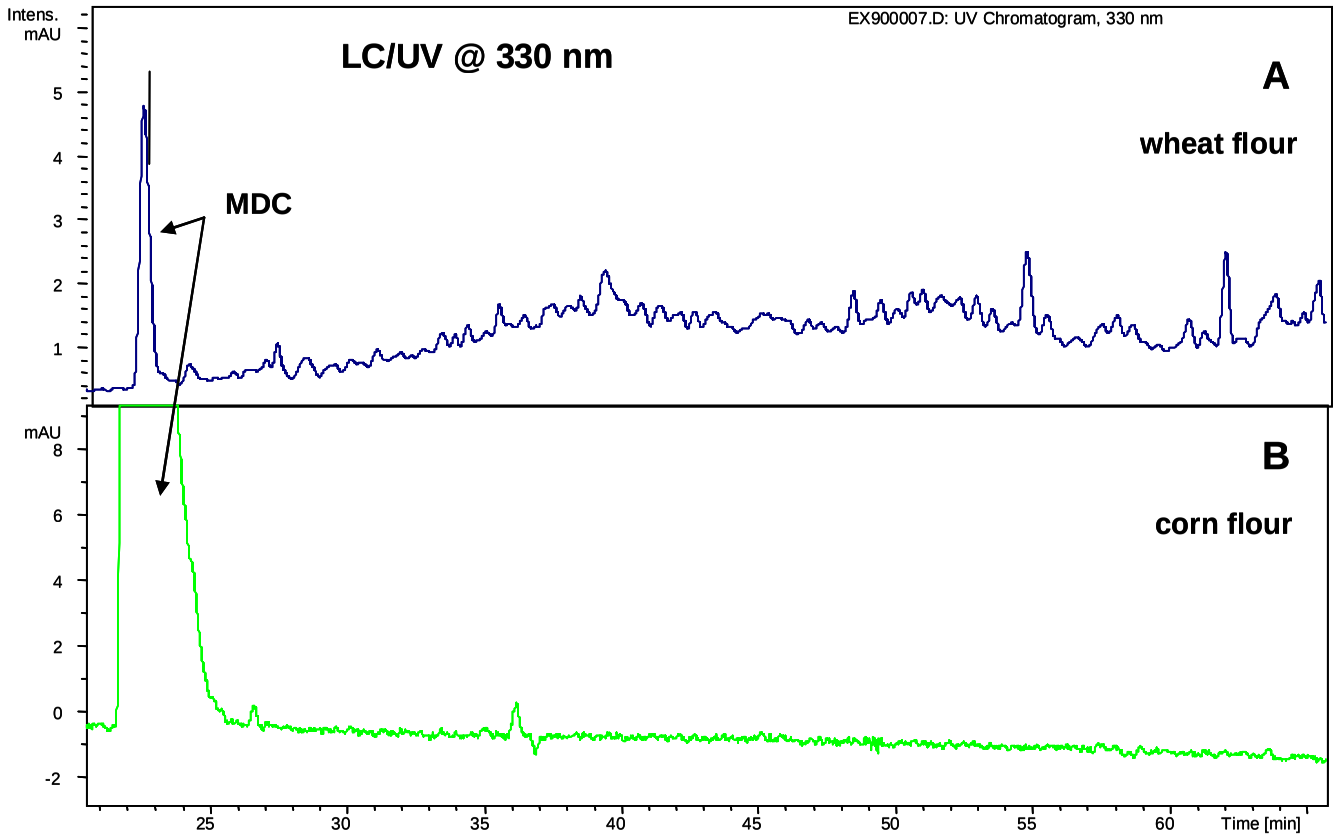


Figure 3.17

LC/UV analysis of MDC-tagged gluten peptides from proteolyzed wheat and corn flours

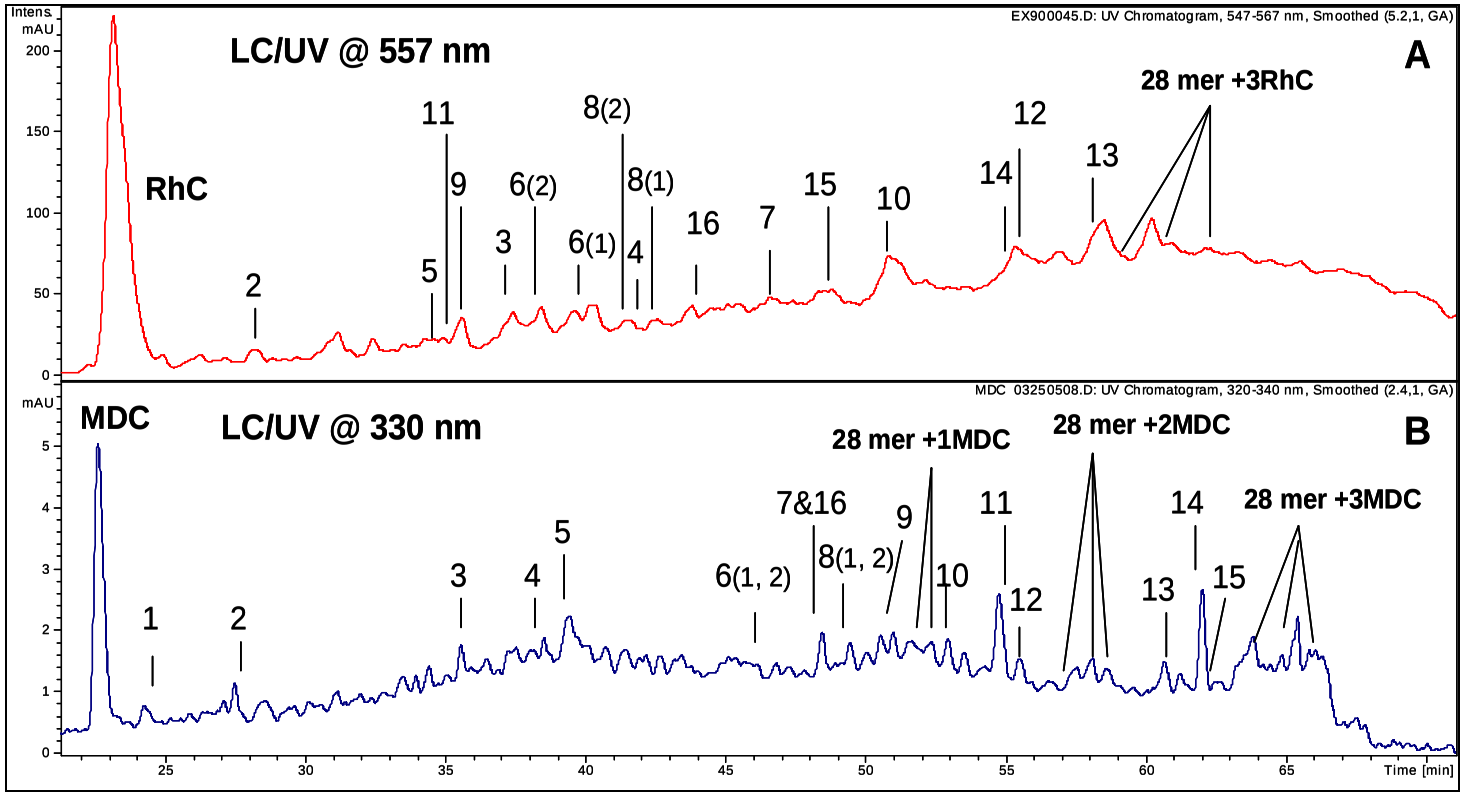


Figure 3.18

LC/UV analysis of MDC- and RhC- tagged wheat gluten peptides

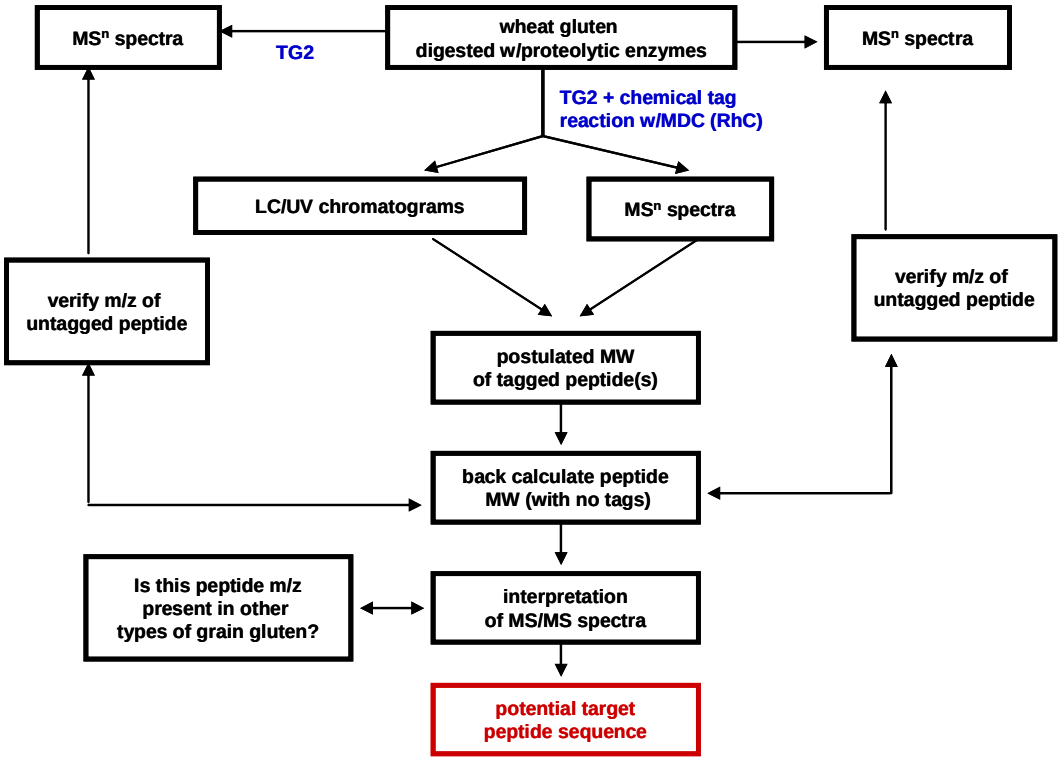


Figure 3.19

Summary of the data analysis process developed for the identification of potentially immunogenic wheat gluten peptides

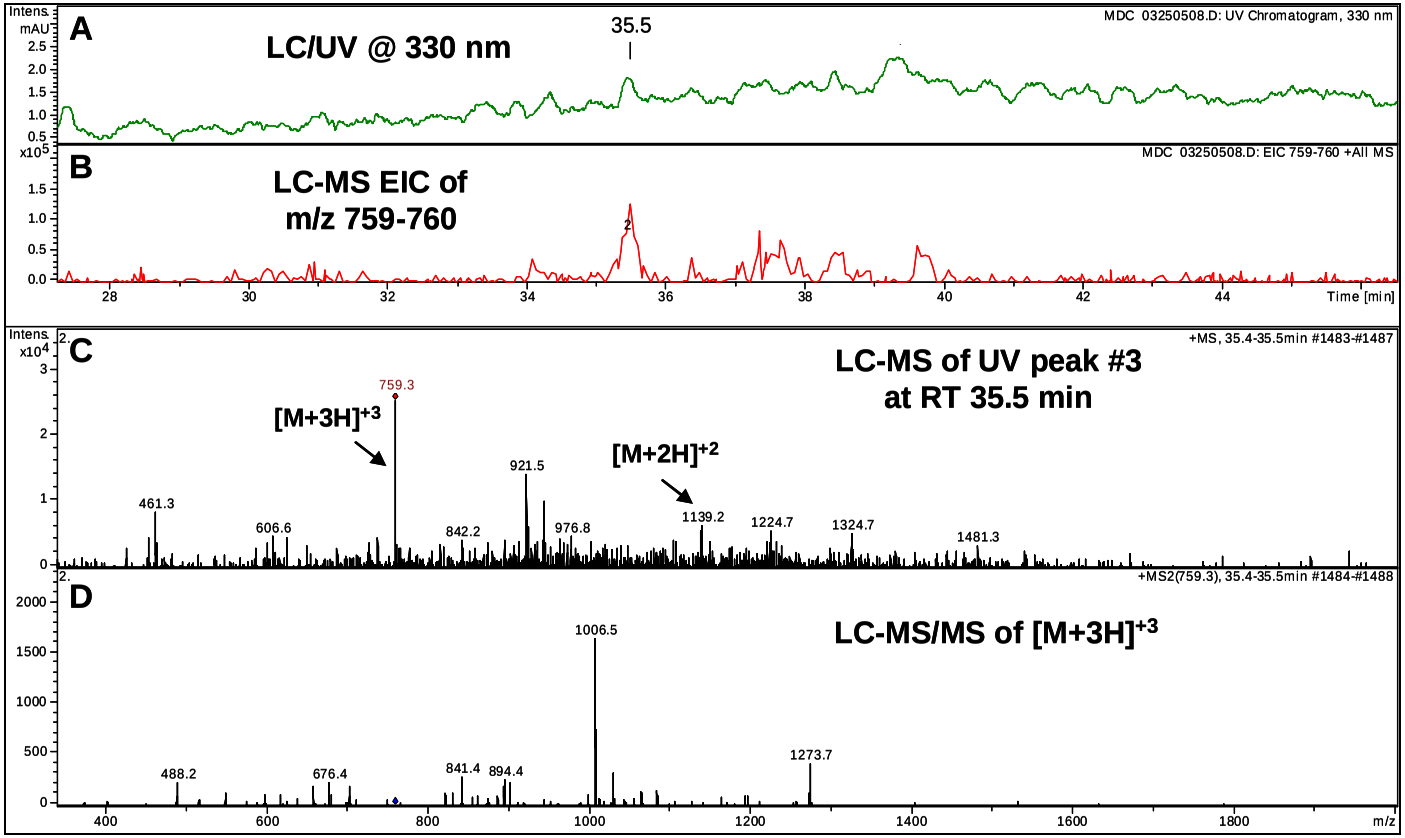


Figure 3.20

Data analysis – LC/UV and MS spectra from one MDC-tagged peptide
[LC/UV peak #3 from Figure 3.18(B)]

LC-MS EICs from proteolyzed wheat

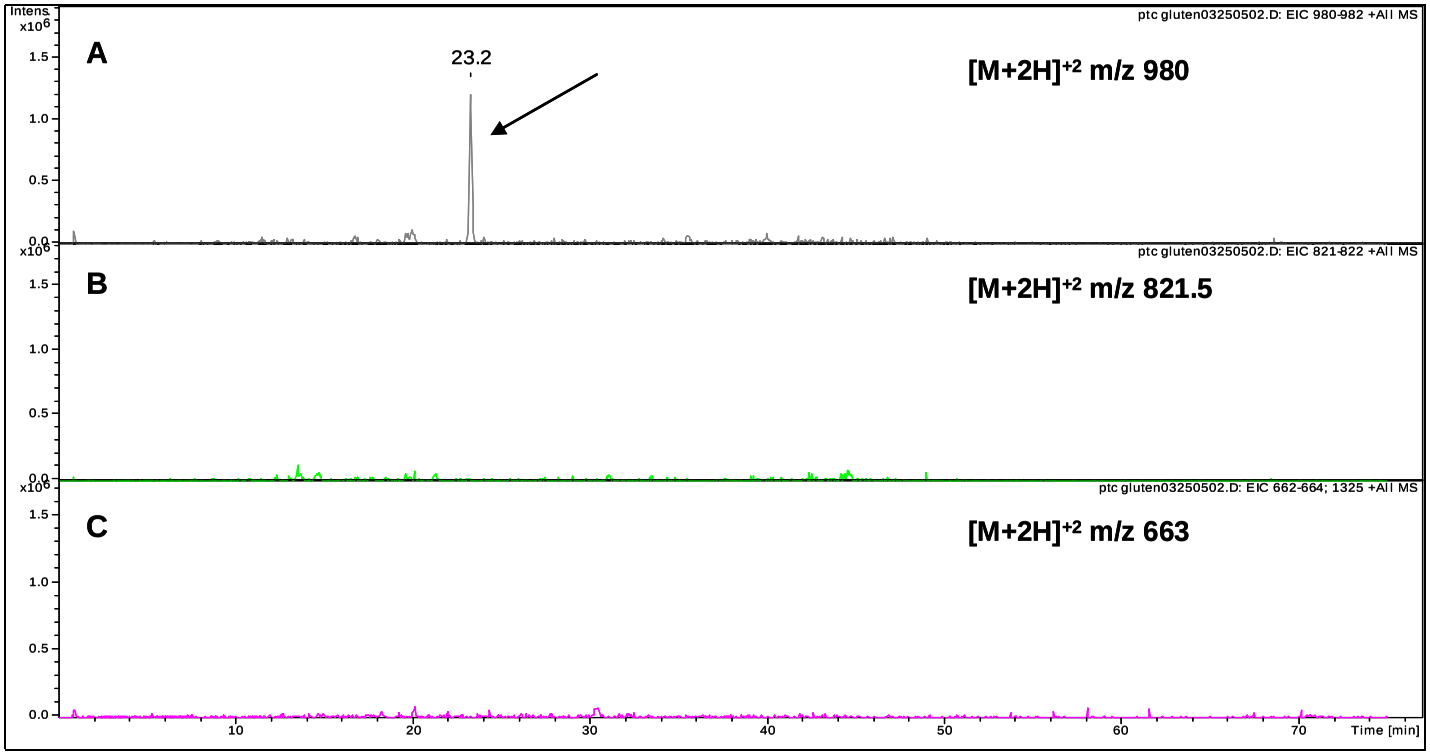


Figure 3.21

Data analysis – mining the LC-MS data of proteolyzed wheat (untagged) for specific ions from untagged peptides corresponding to ions from the MDC-tagged peptide in UV peak #3

LC-MS EIC

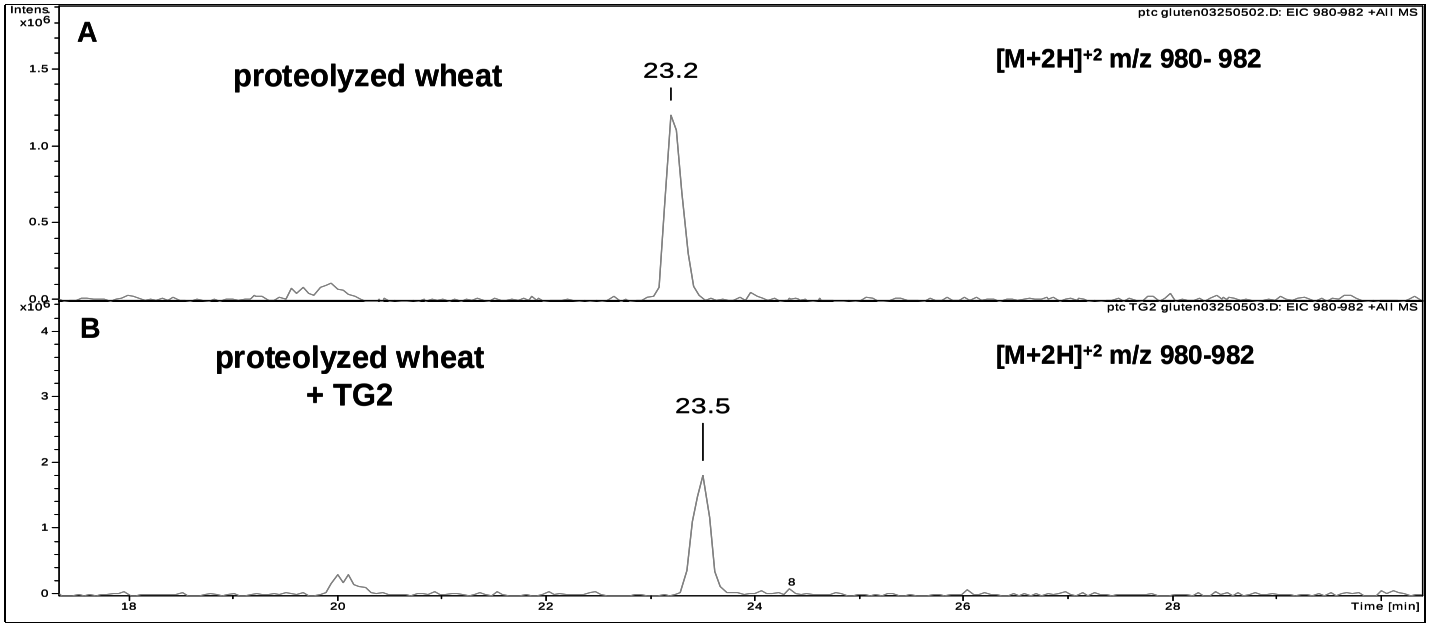


Figure 3.22

Data analysis - mining the LC-MS data of proteolyzed wheat (untagged and deamidated) for specific m/z ions corresponding to those of the related deamidated untagged peptide

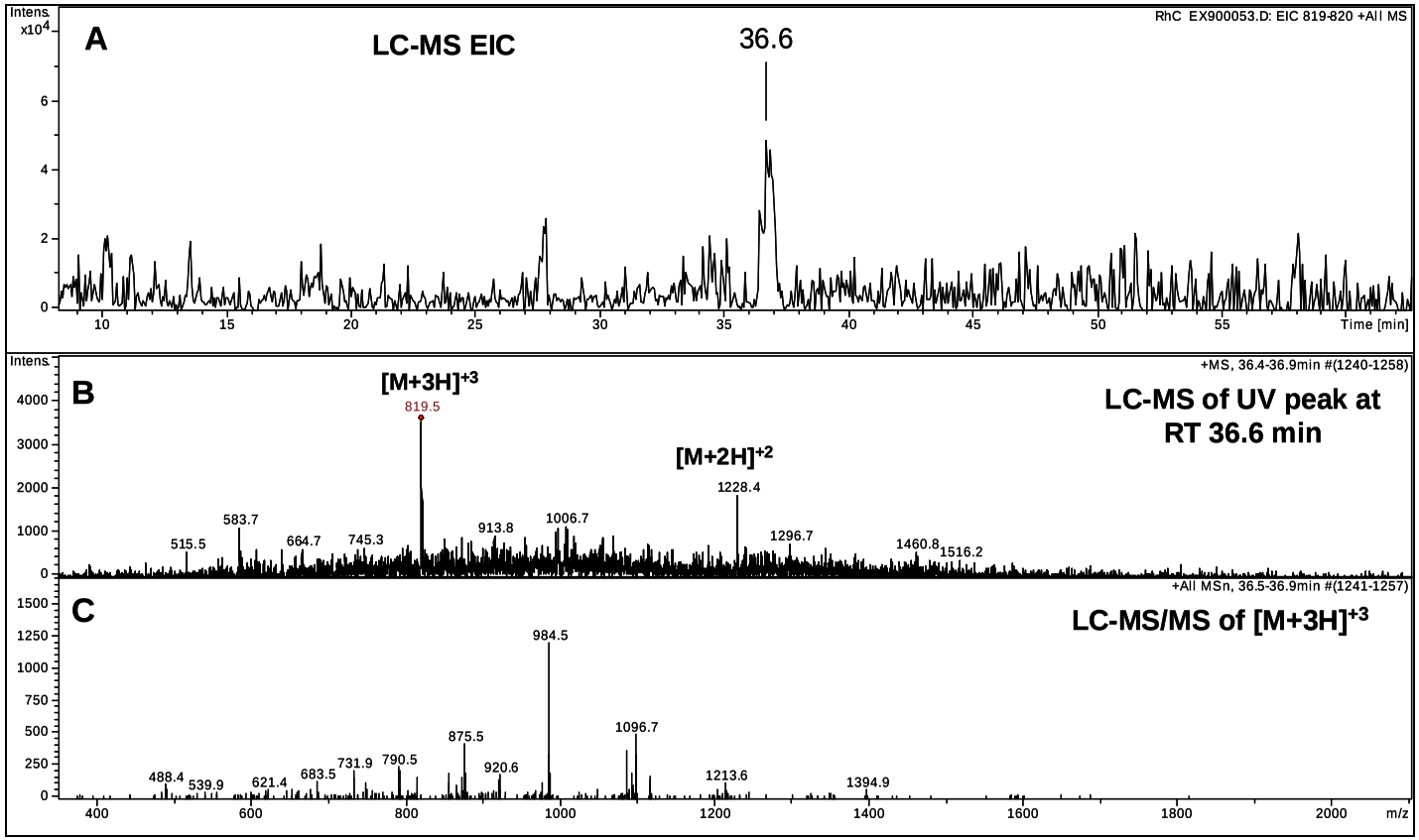


Figure 3.23

Data analysis – interpretation of the mass spectra from the RhC-tagged peptide [LC/UV peak #3 from Figure 3.18(A)]

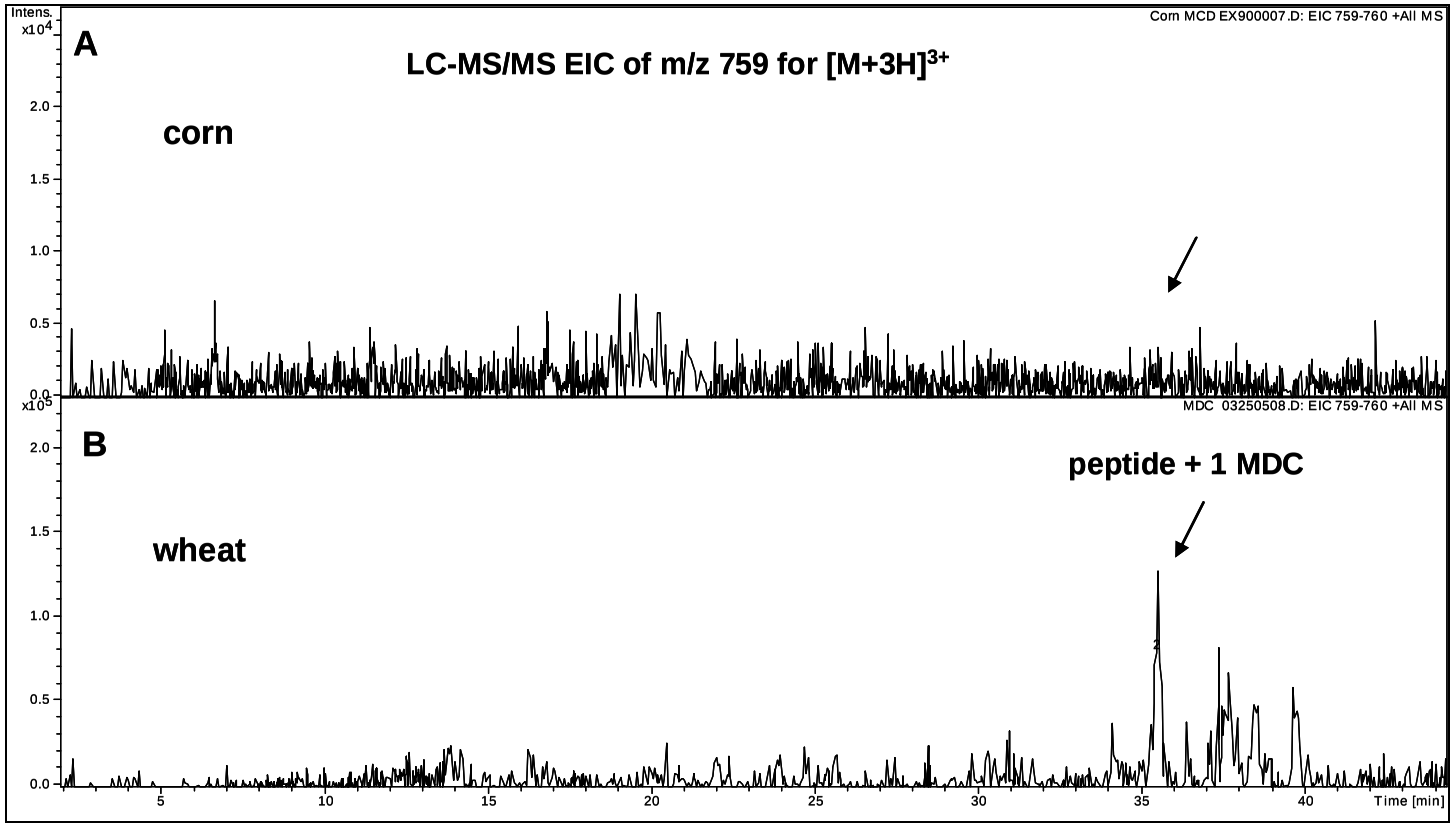


Figure 3.24

Data analysis – presence/absence of the $[M+3H]^{3+}$ ion (from LC/UV peak#3) in the mass spectra of proteolyzed and MDC-tagged corn and wheat flour

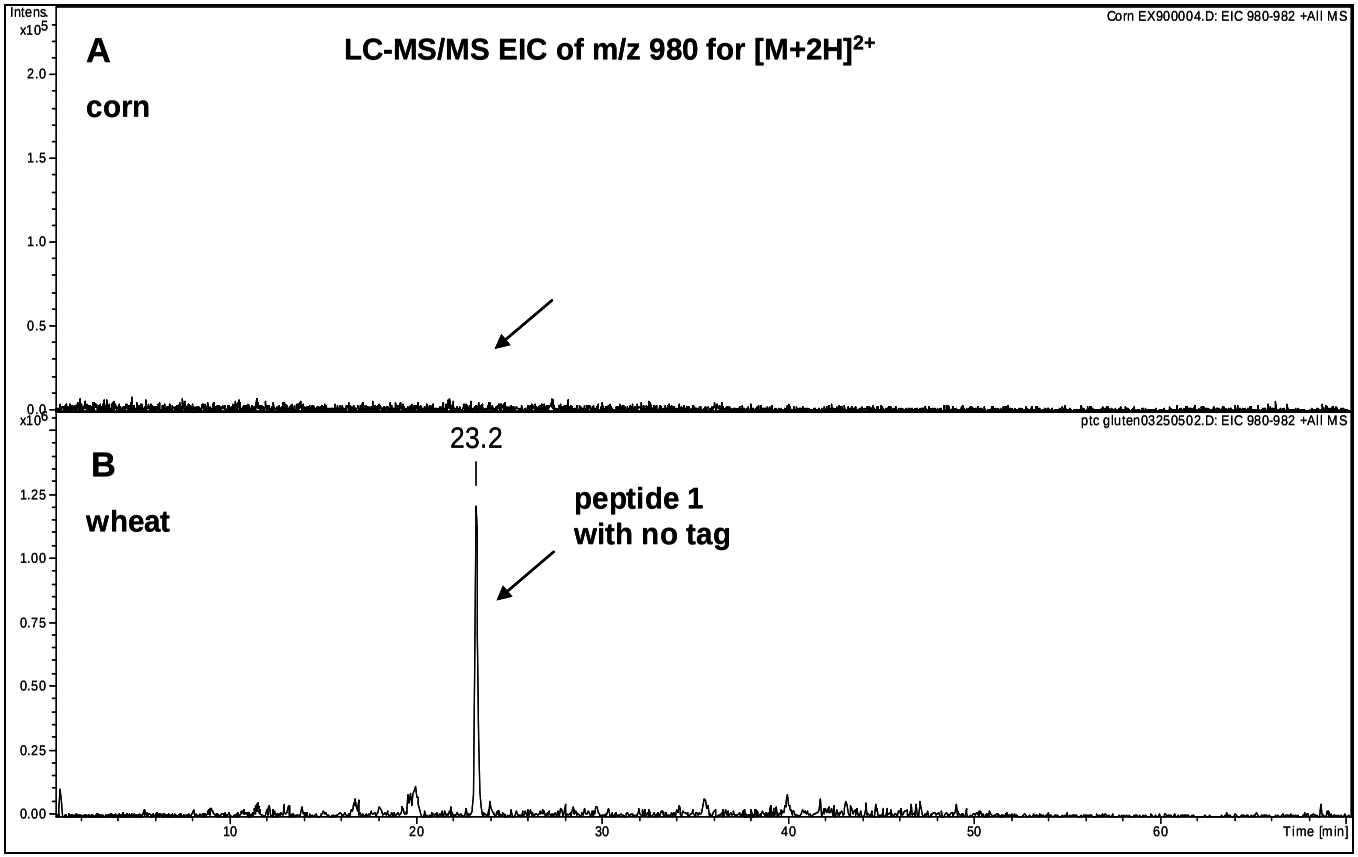


Figure 3.25

Data analysis – presence/absence of the $[M+2H]^{2+}$ ion (from LC/UV peak#3) in the mass spectra of proteolyzed and untagged corn and wheat flour wheat flour

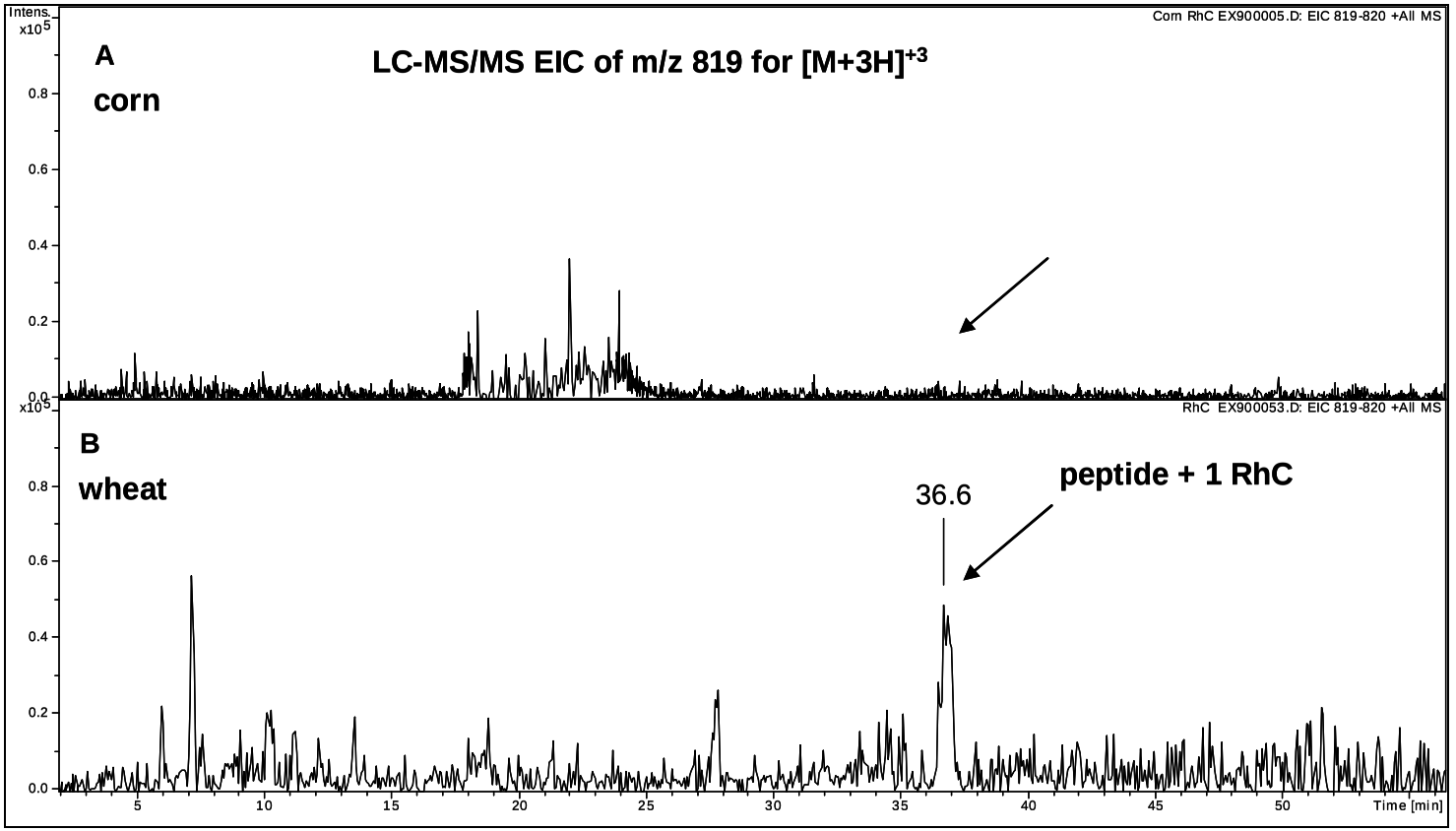


Figure 3.26

Data analysis – presence/absence of the $[M+3H]^{+3}$ ion (from LC/UV peak#3) in the mass spectra of proteolyzed and RhC-tagged corn and wheat flour

A

P04723 **Mass:** 32216 **Score:** 38 **Queries matched:** 1

(GDA3_WHEAT) Alpha/beta-gliadin A-III precursor (Prolamin) Alpha/beta-gliadin A-III precursor (Prolamin)

☐ Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
1	980.32	1958.63	1958.97	-0.34	0	51	0.17	2	Q.LQPQNPSQQQPQEQVPL.M + Deamidation (NQ)

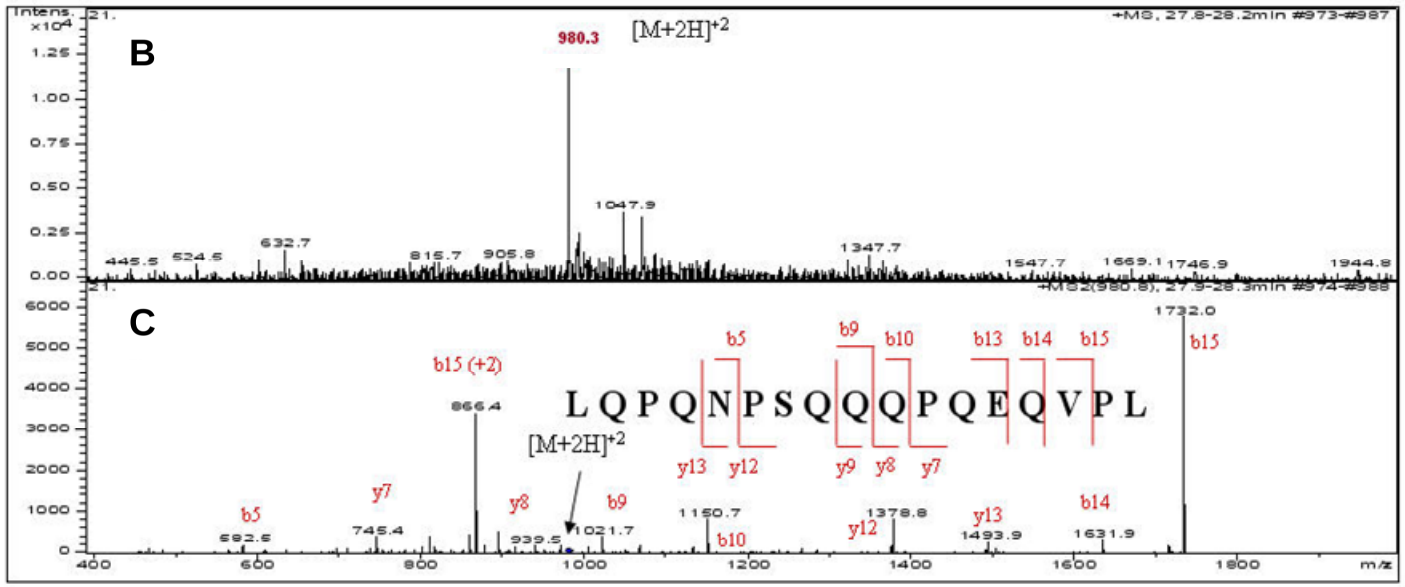


Figure 3.27

Data analysis – protein database search using experimentally generated MS/MS spectra in order to identify LC/UV peak #3

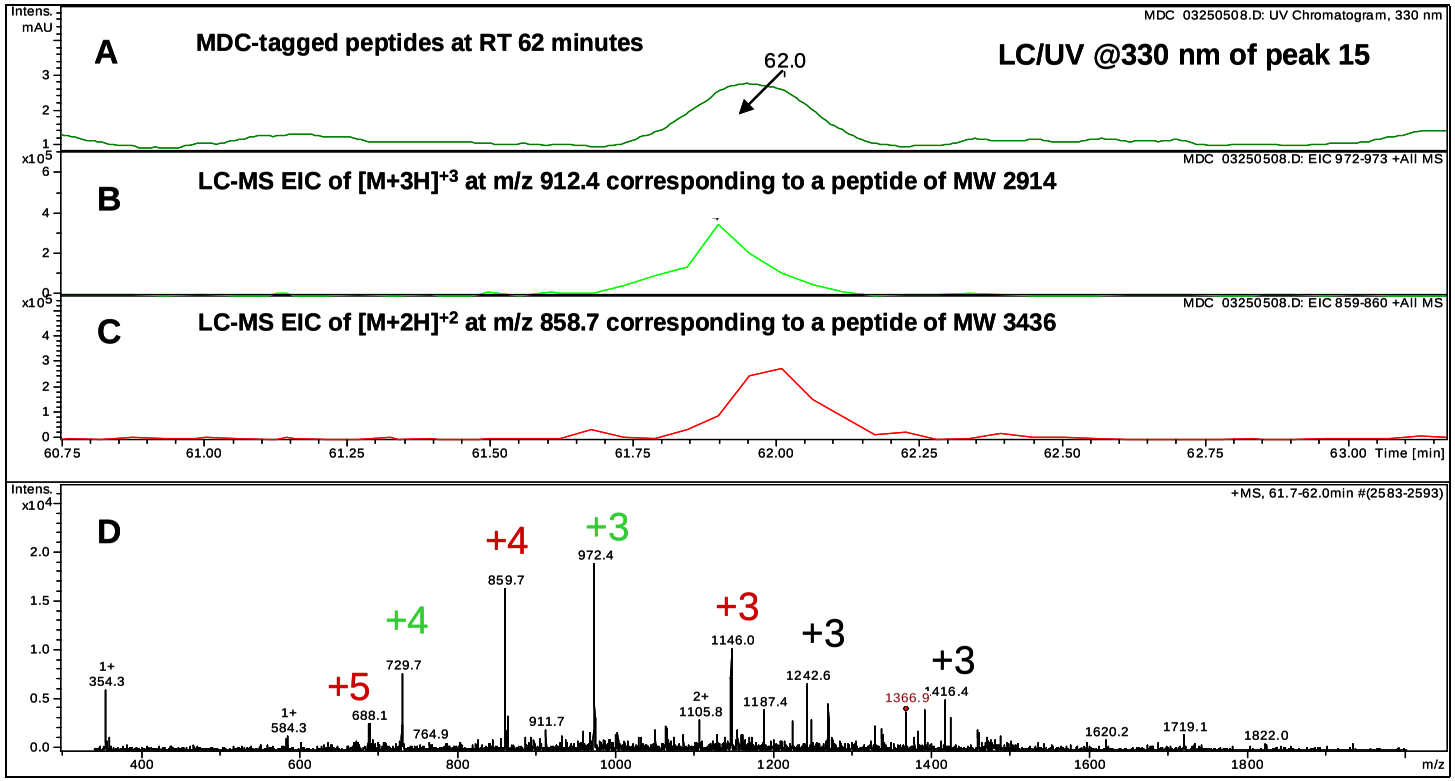


Figure 3.28

Data complexity that affected the positive identification of immunogenic peptides in proteolyzed wheat

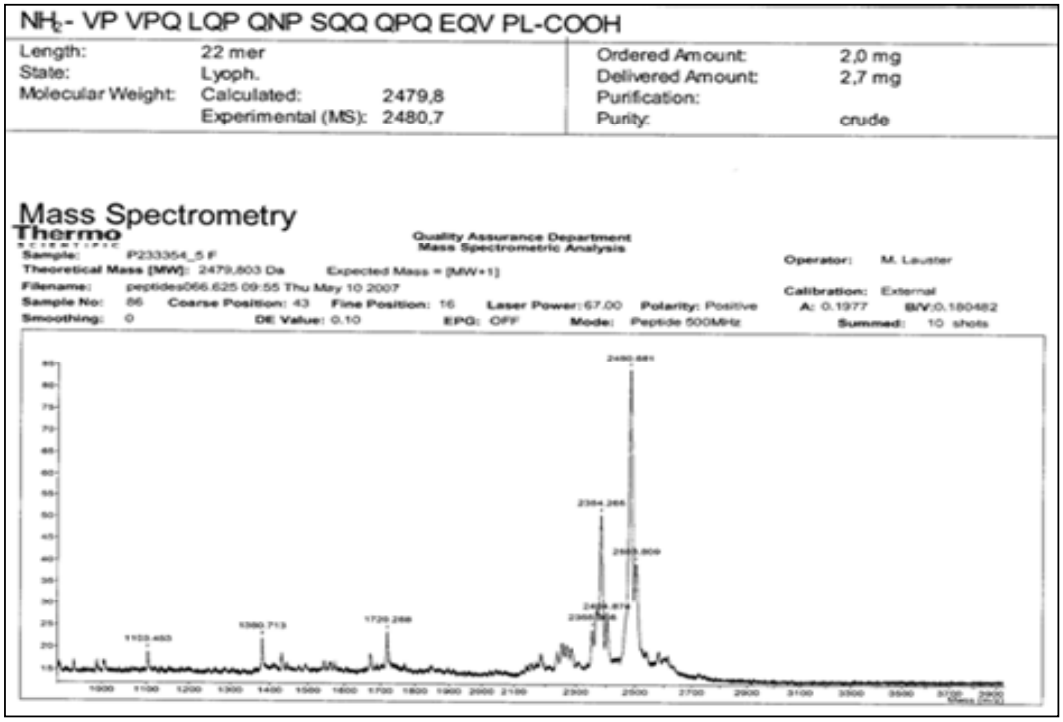


Figure 3.29

MALDI-TOF spectra of synthetic peptide 3 (VPVPQLQPQNPSQQQPQEQVPL)

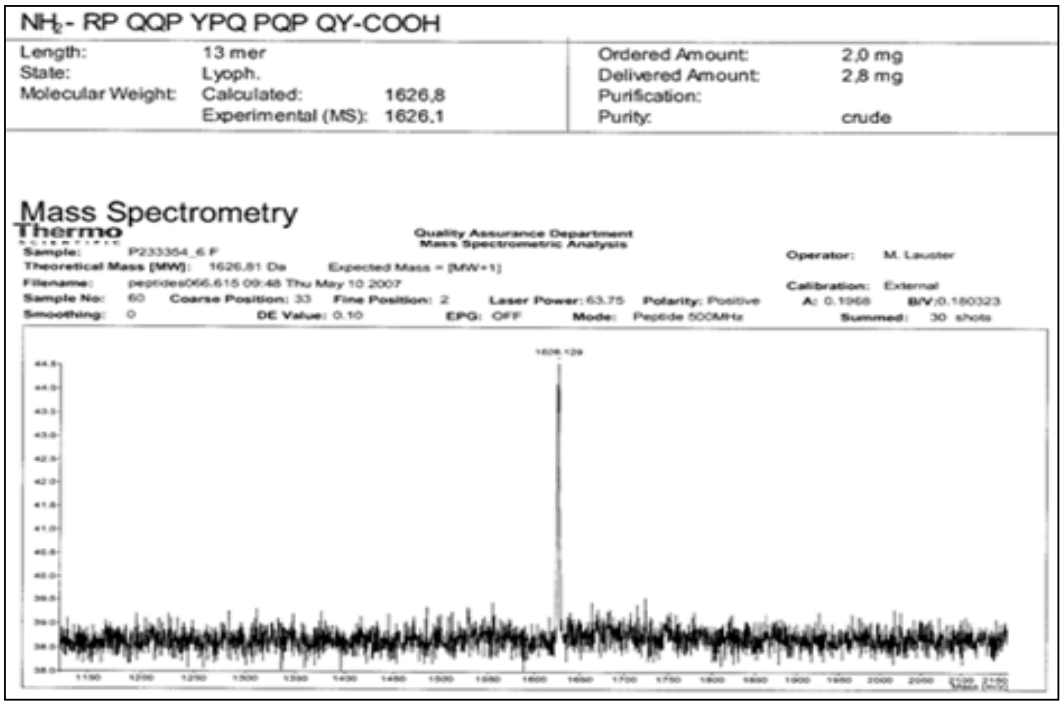


Figure 3.30

MALDI-TOF spectra for synthetic peptide 4 (RPQQPYPQPQPQY)

A

Peptide #	Peptide MW	Peptide sequence
1	1959	LQPQNPSQQQPQEQVPL
2	2150	TQQPQQPFPQQPQQPFPQ
3	2480	VPVPQLQPQNPSQQQPQEQVPL
4	1627	RPQQPYPQPQPQY
5	2150	QPQQPFPQTQQPQQPFPQ
6	703	PQQSPF
7	1532	QPQQPLPQPQQPF

B

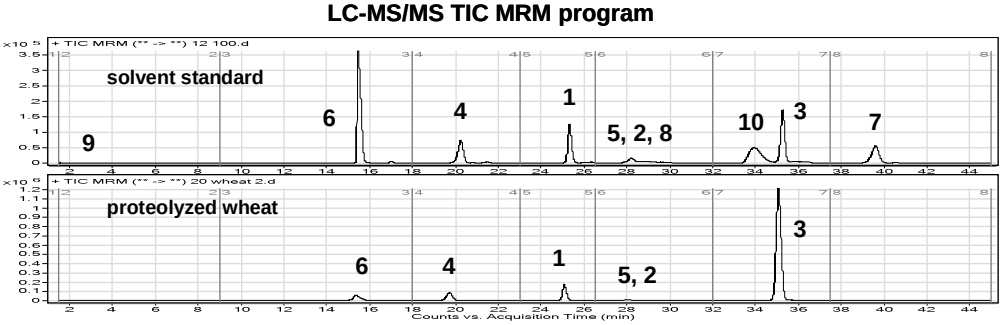


Figure 3.31

Confirmation of identity of potentially immunogenic wheat gluten peptides

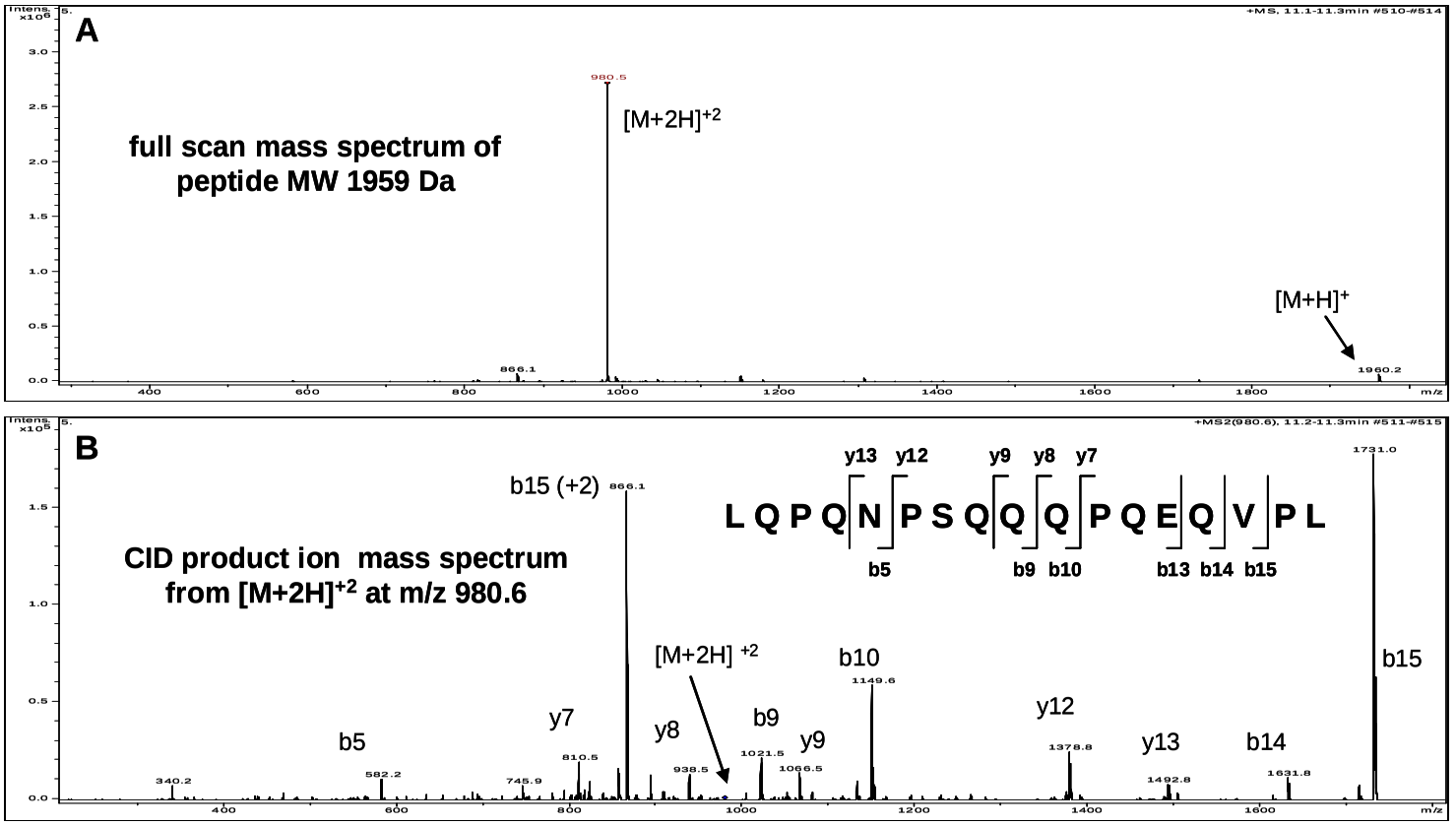


Figure 3.32

Full scan mass spectrum and LC-MS/MS CID product ion mass spectrum for peptide 1

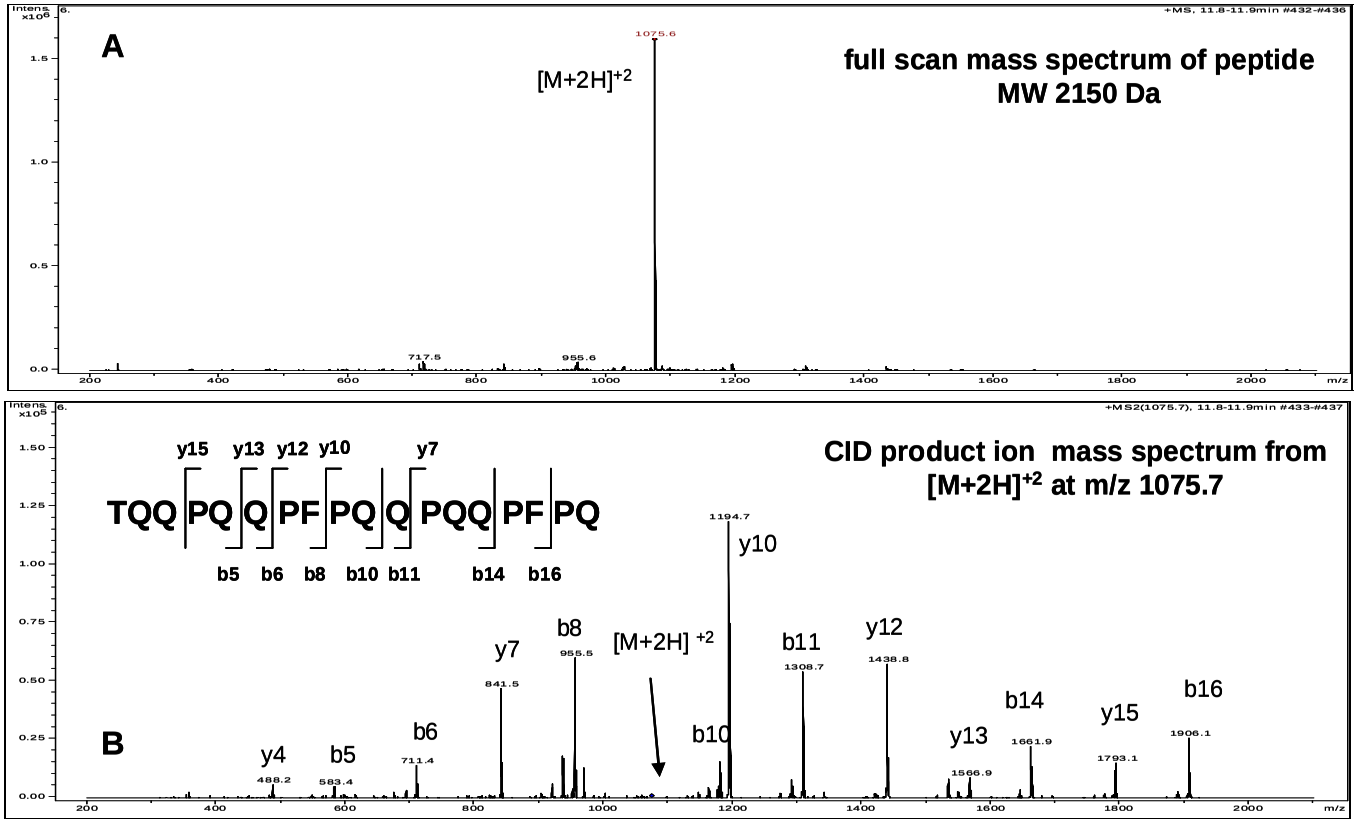


Figure 3.33

Full scan mass spectrum and LC-MS/MS CID product ion mass spectrum for peptide 2

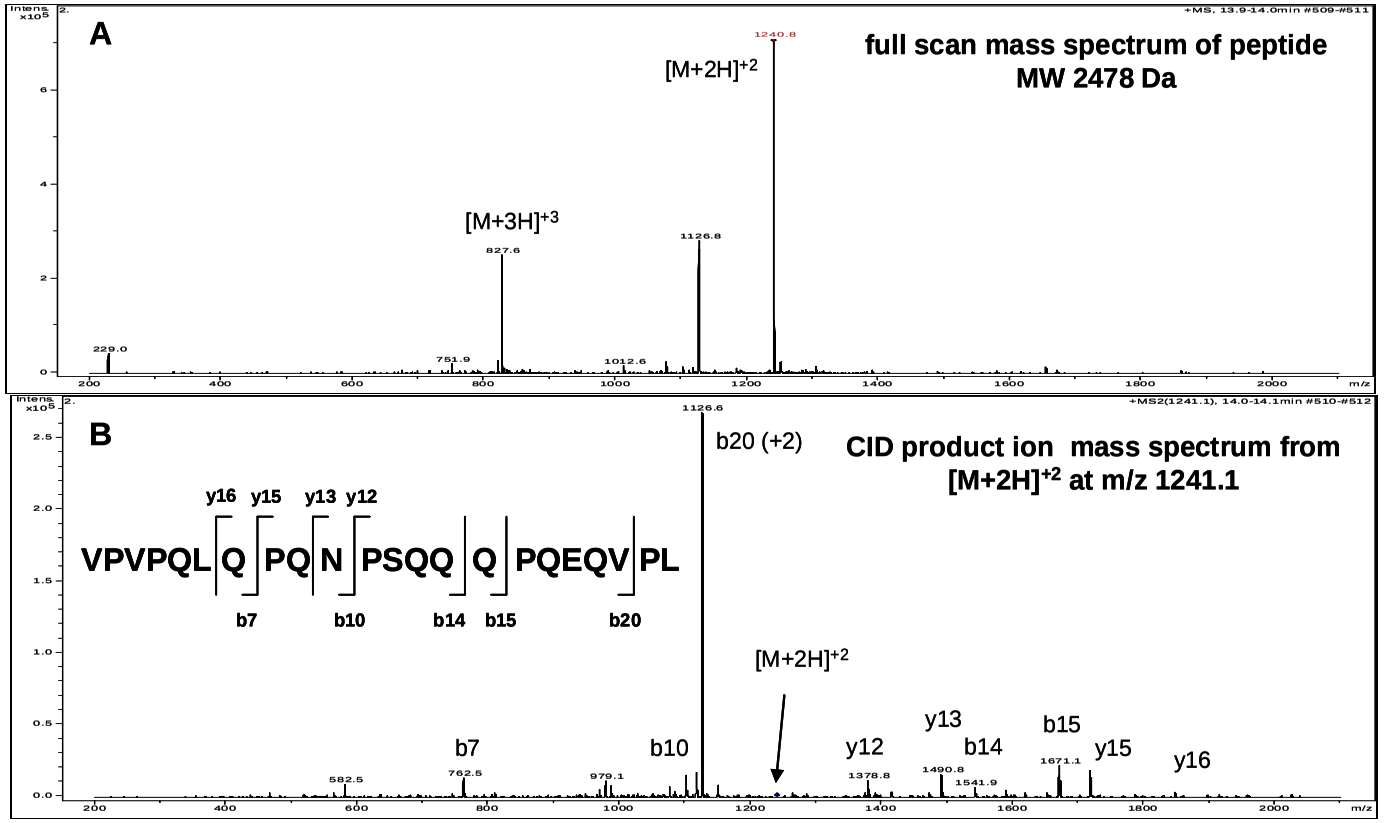


Figure 3.34

Full scan mass spectrum and LC-MS/MS CID product ion mass spectrum for peptide 3

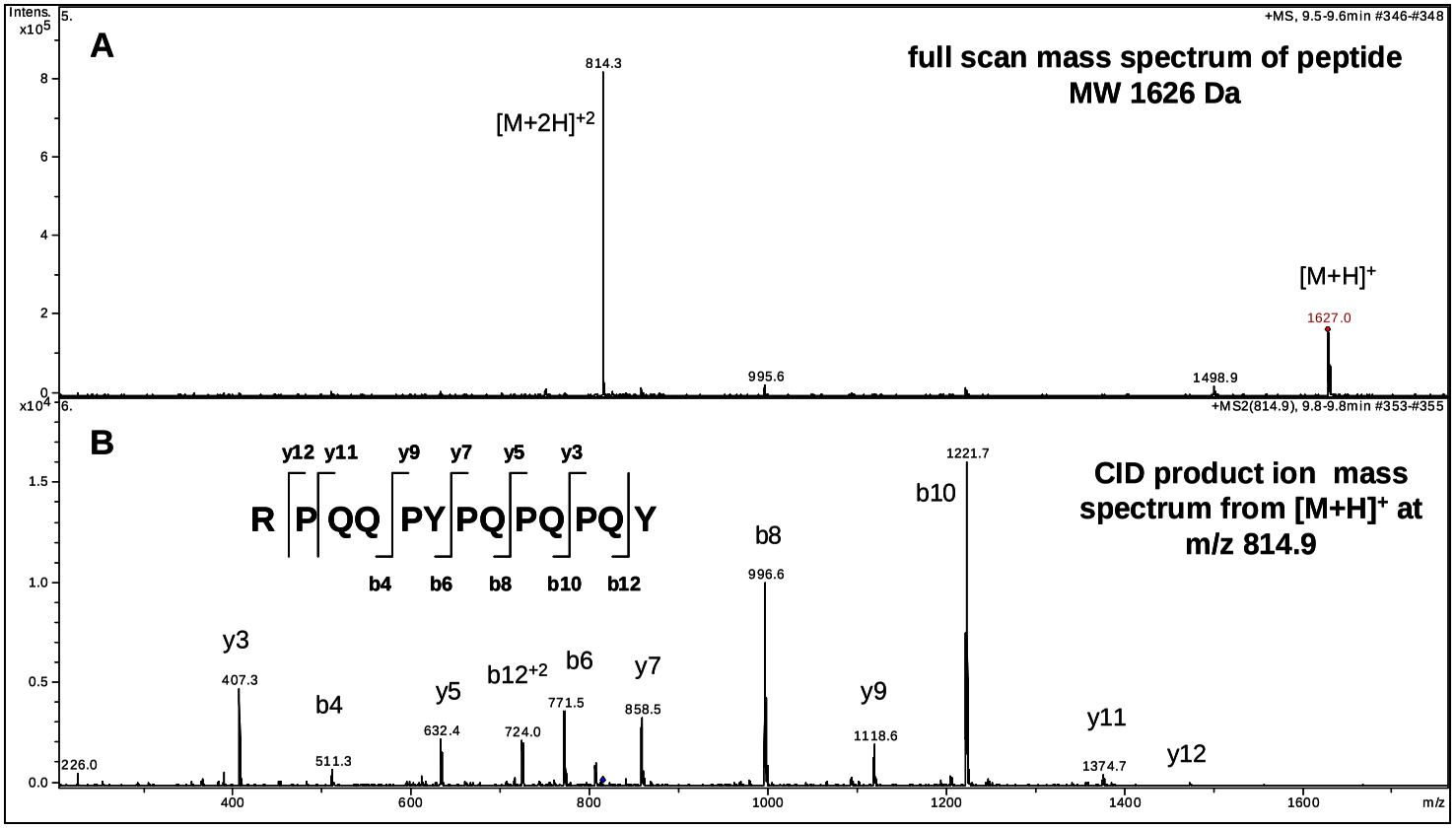


Figure 3.35

Full scan mass spectrum and LC-MS/MS CID product ion mass spectrum for peptide 4

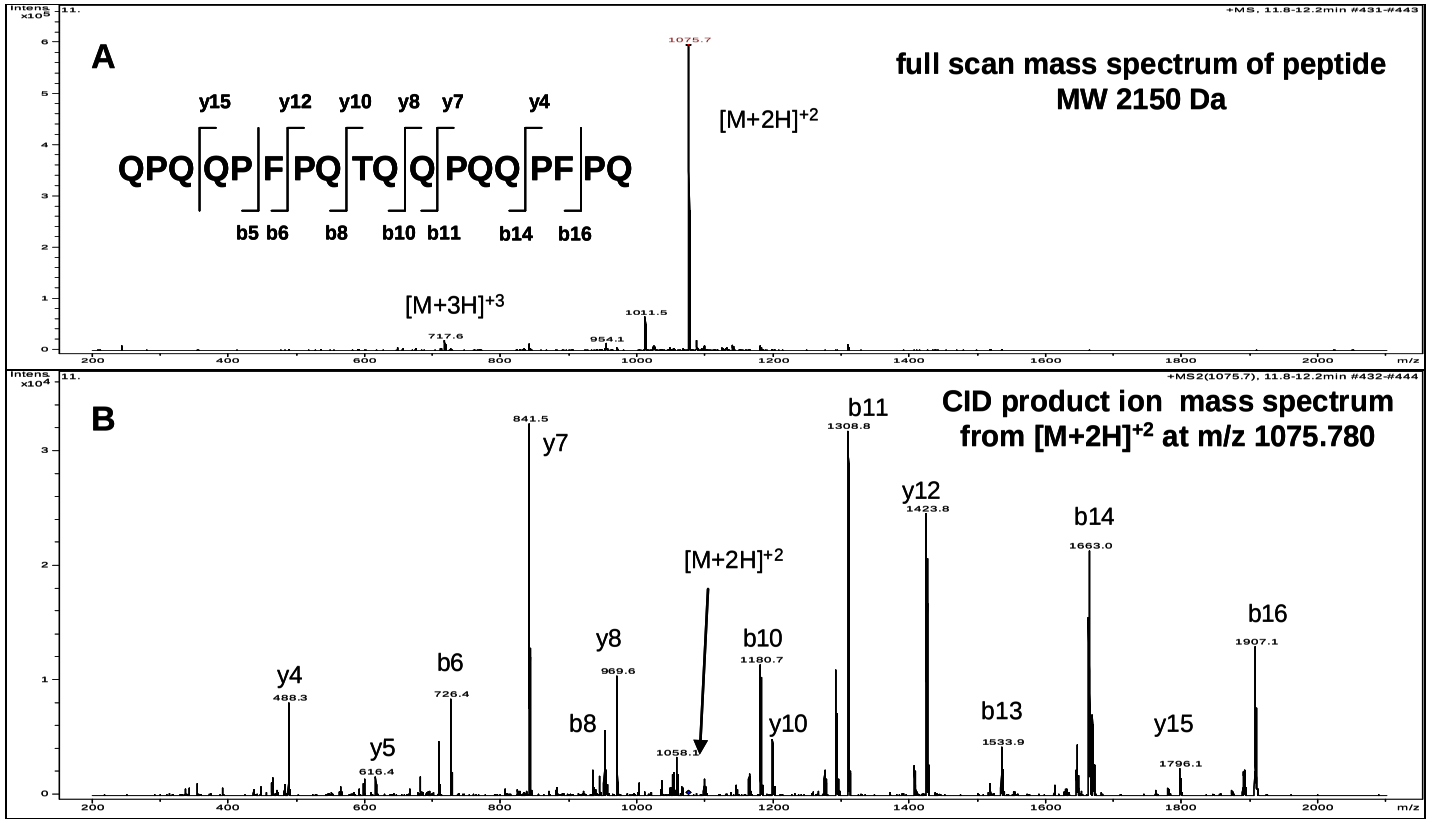


Figure 3.36

Full scan mass spectrum and LC-MS/MS CID product ion mass spectrum for peptide 5

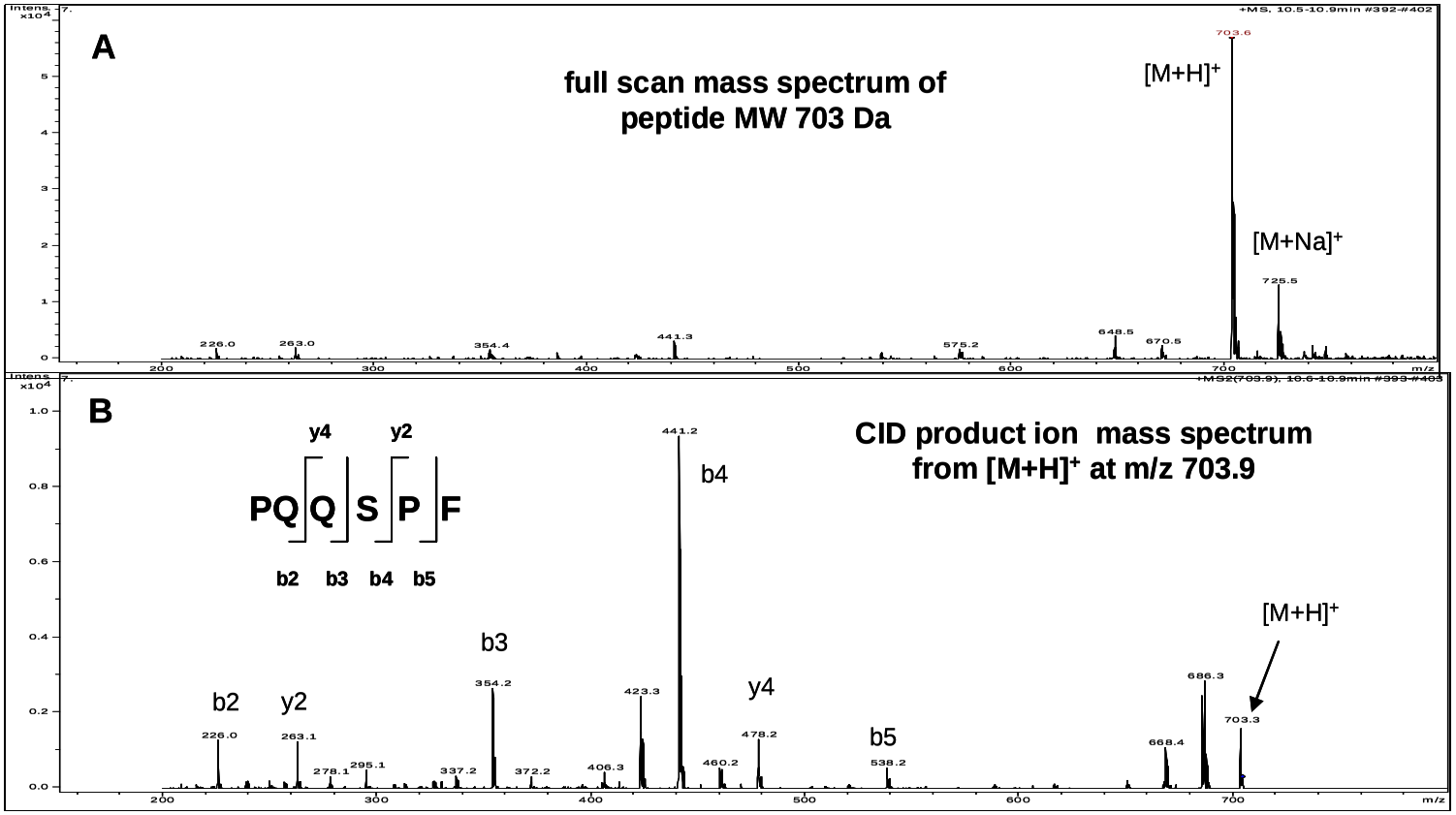


Figure 3.37

Full scan mass spectrum and LC-MS/MS CID product ion mass spectrum for peptide 6

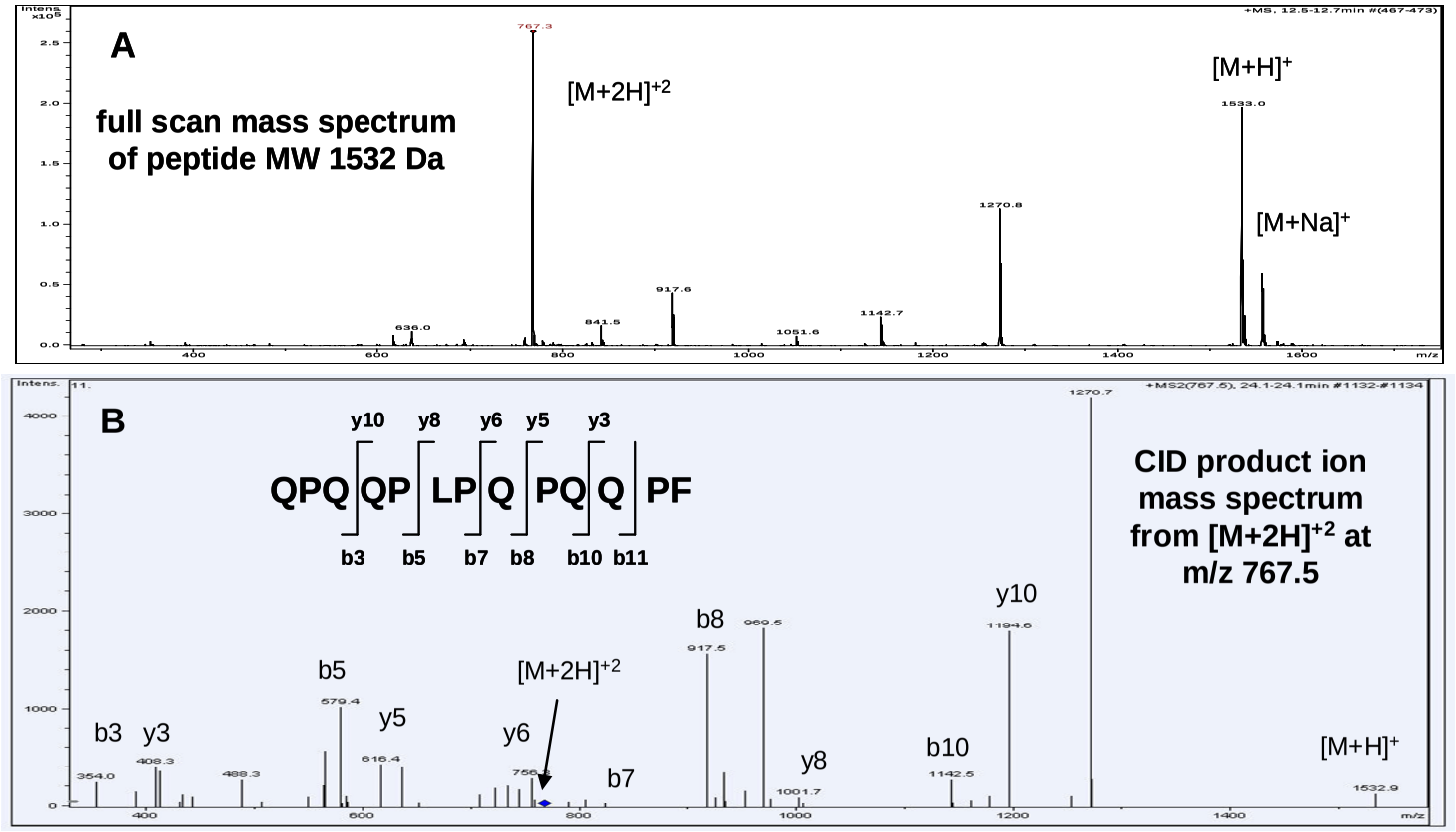


Figure 3.38

Full scan mass spectrum and LC-MS/MS CID product ion mass spectrum for peptide 7

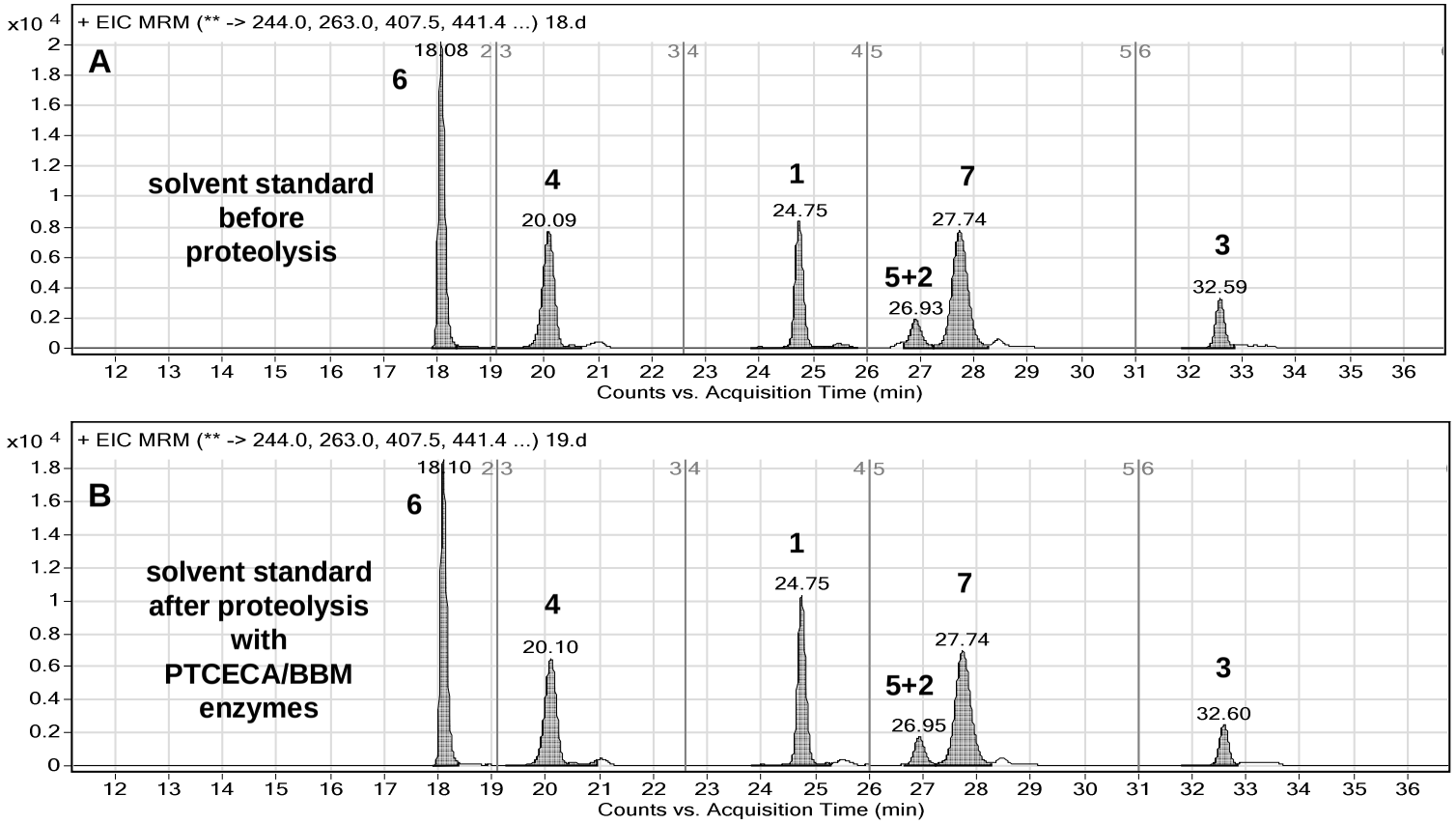


Figure 3.39

Degradation test with PTCECA/BBM proteolytic enzymes on the seven immunogenic wheat gluten peptides

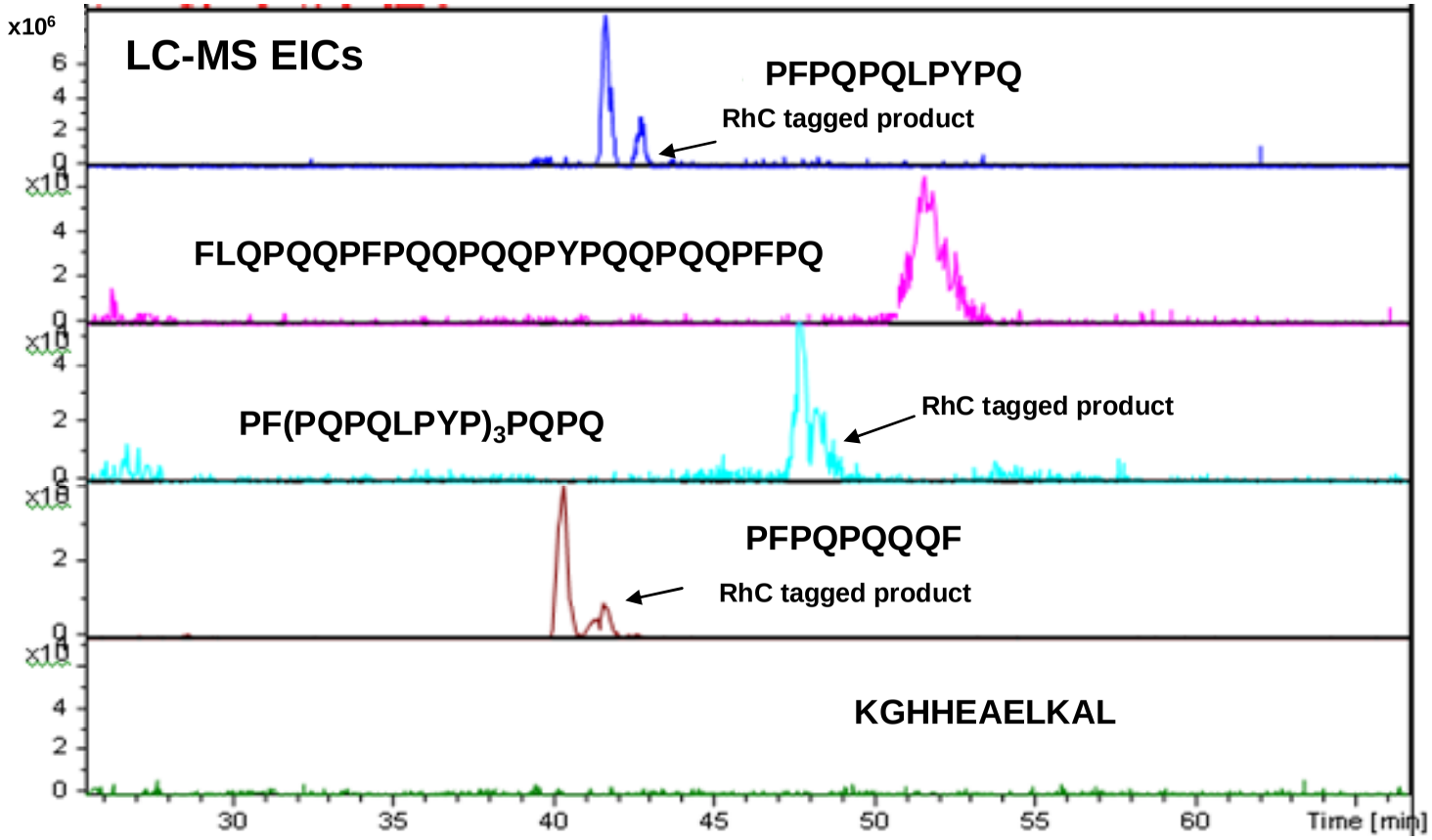


Figure 3.40

Specificity of TG2 with reference synthetic Q-containing peptides

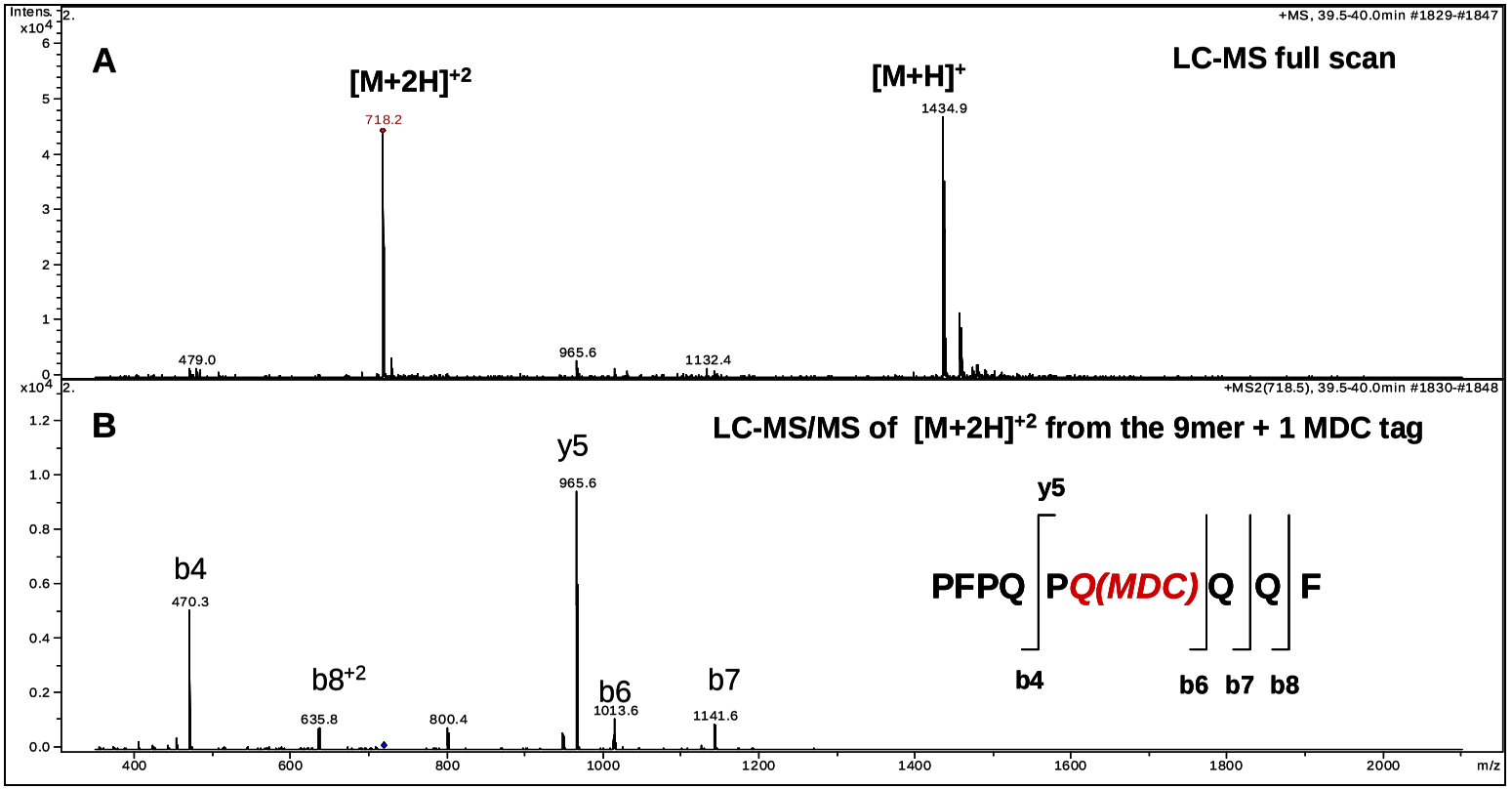


Figure 3.41

Specificity of TG2 within PFPQPQQQF using a MDC tag

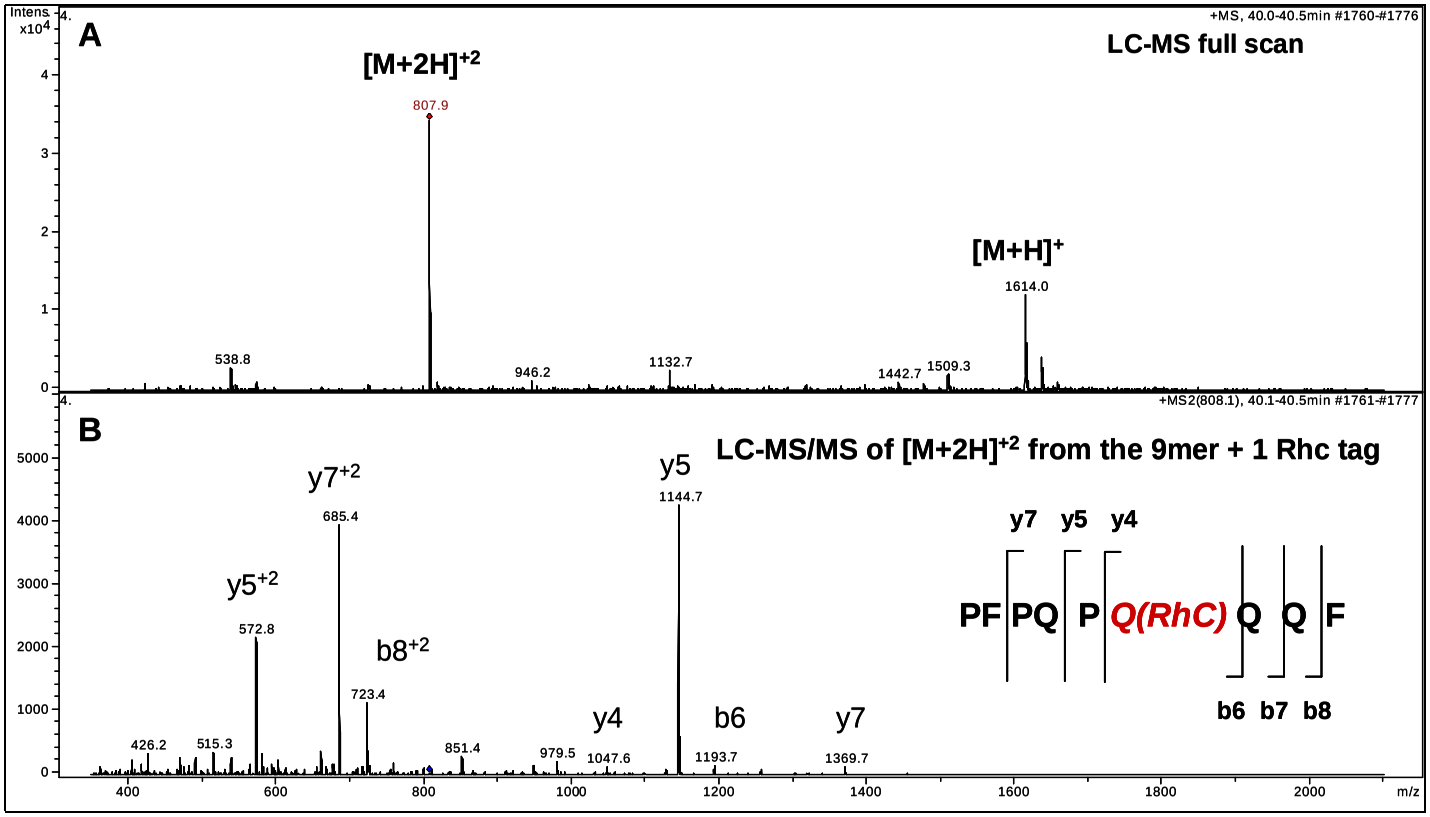
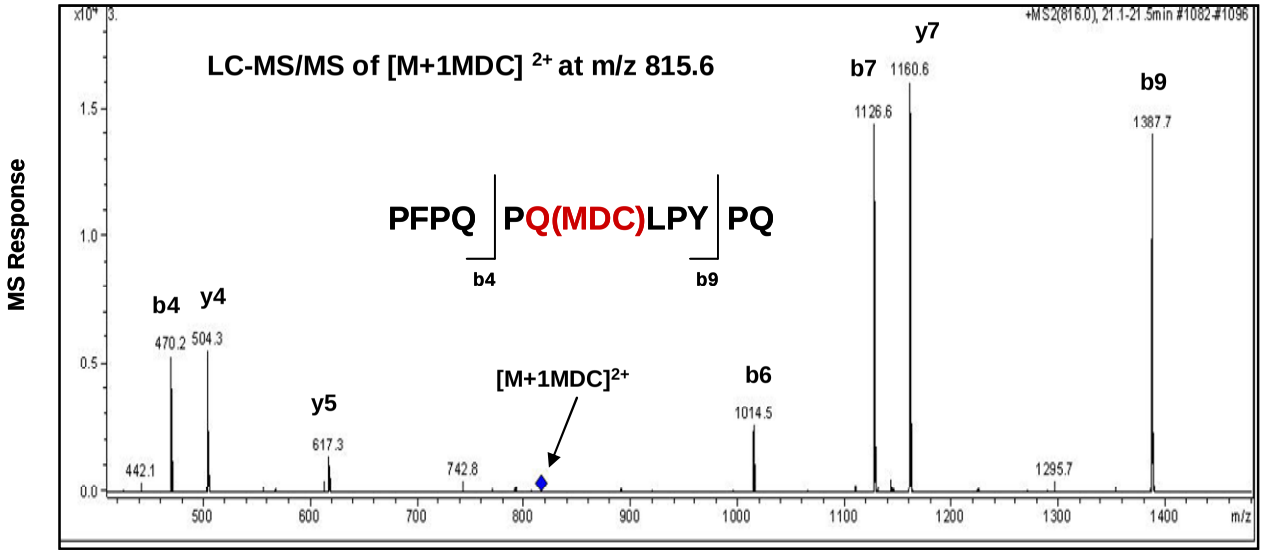


Figure 3.42

Specificity of TG2 within PFPQPQQQF using a RhC tag



	b4 (m/z)	b9 (m/z)	% observed
Q4	788	1387	<1
Q6	470	1387	99
Q11	470	1068	<0.1

Figure 3.43

Specificity of TG2 within PFPQPQLPYPQ using a MDC tag

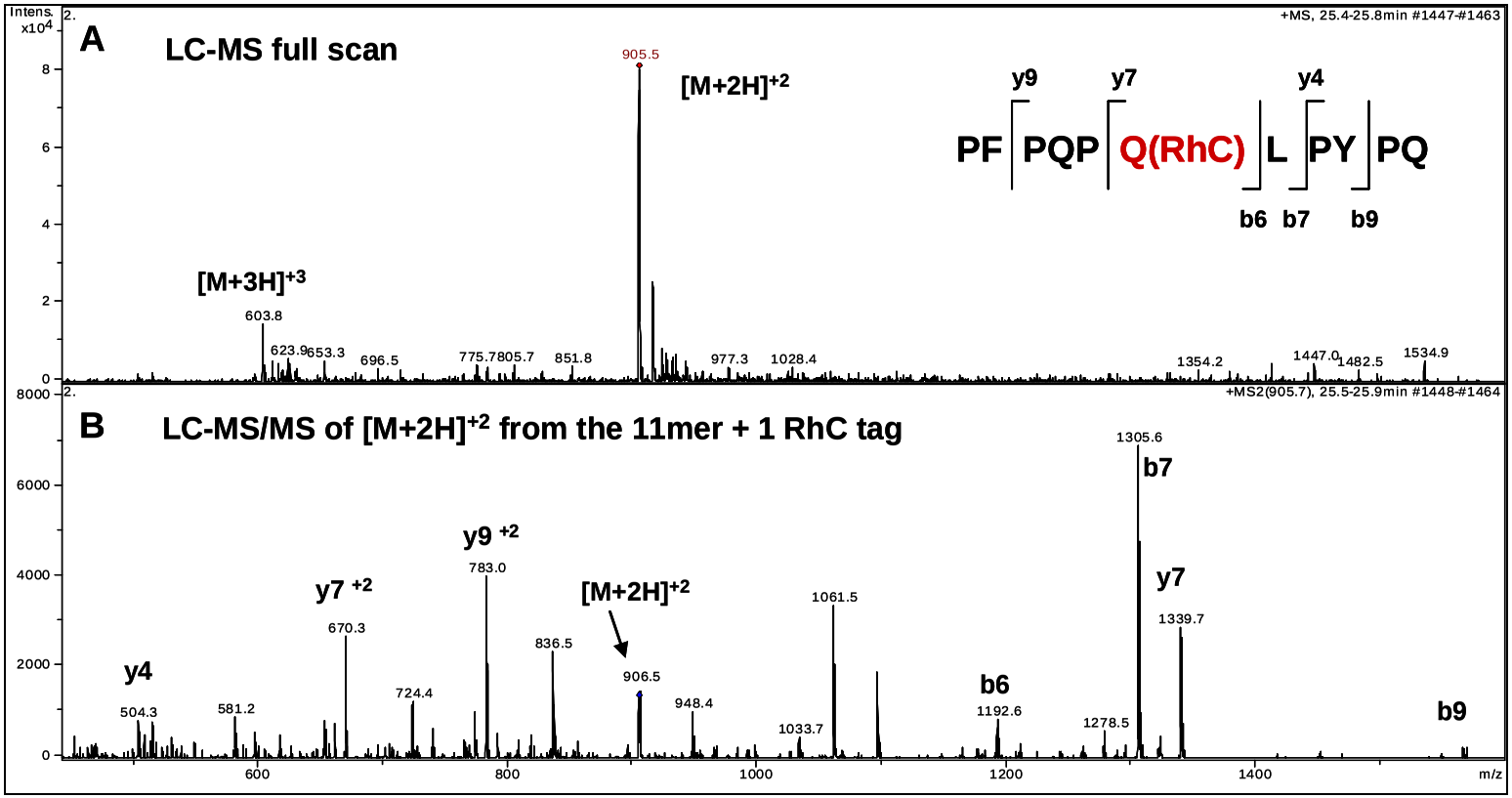


Figure 3.44

Specificity of TG2 within PFPQPQLPYPQ using a RhC tag

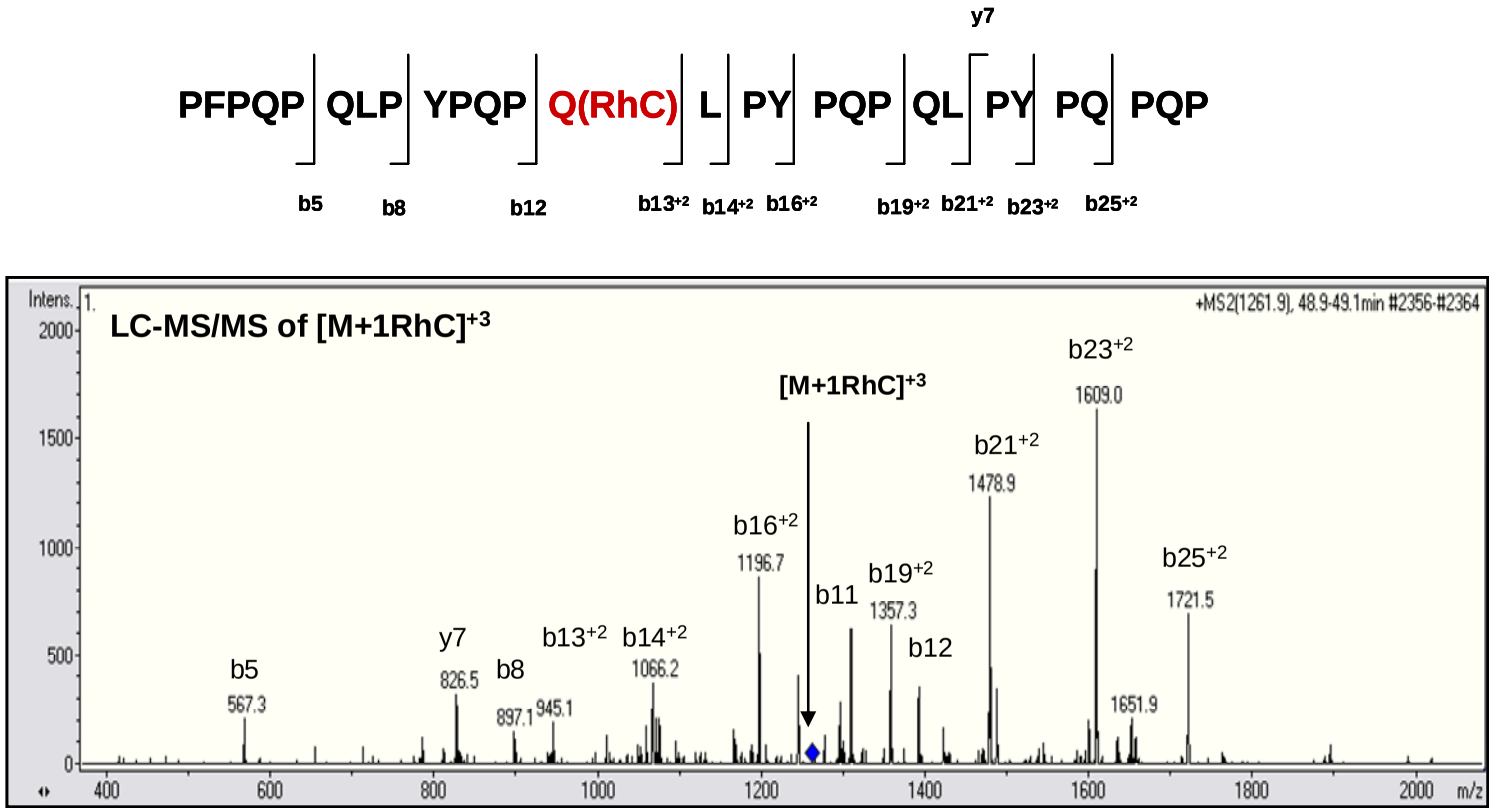


Figure 3.45

Specificity of TG2 within PF(PQPQLPY)3PQPQP using a RhC tag

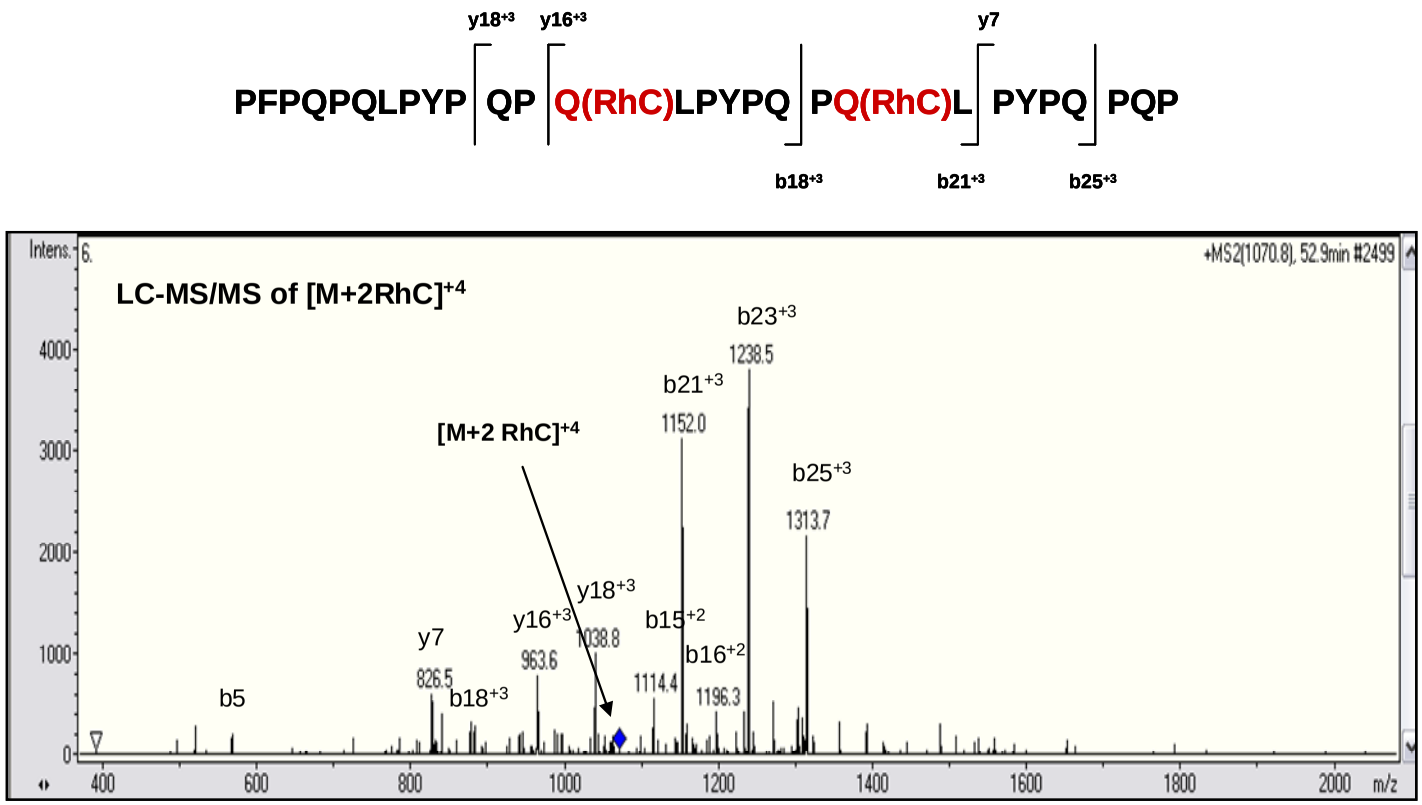


Figure 3.46

Specificity of TG2 within PF(PQPQLPY)3PQPQP using a RhC tag

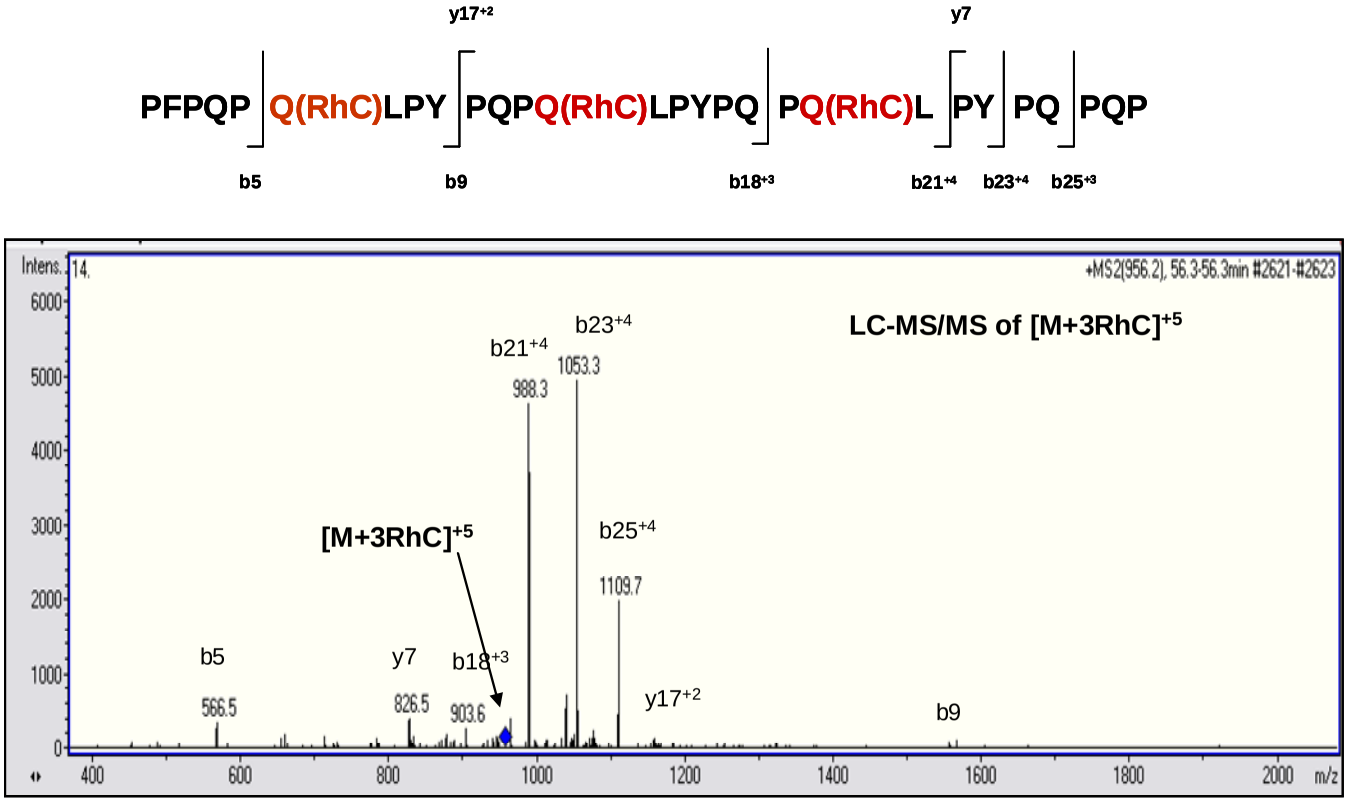
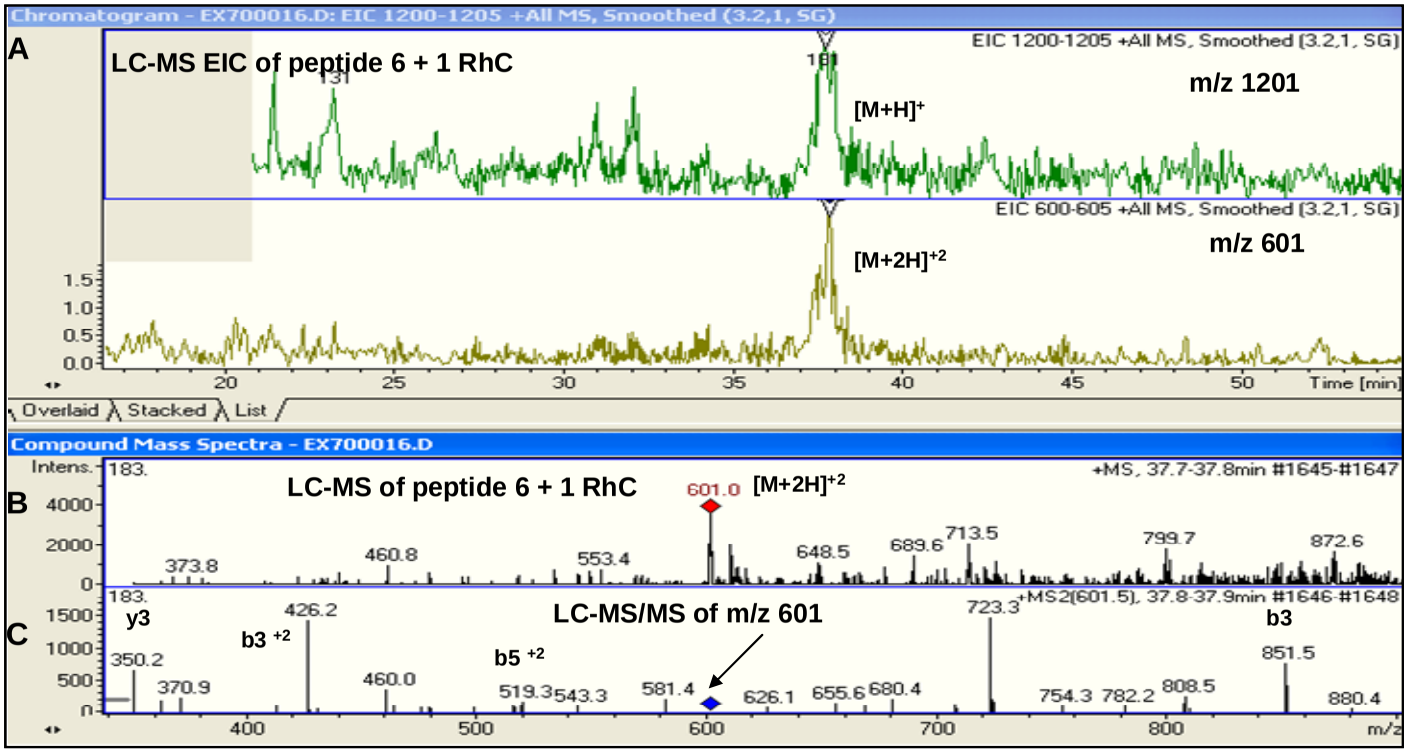


Figure 3.47

Specificity of TG2 within PF(PQPQLPY)3PQPQP using a RhC tag



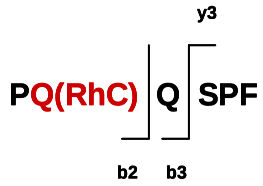
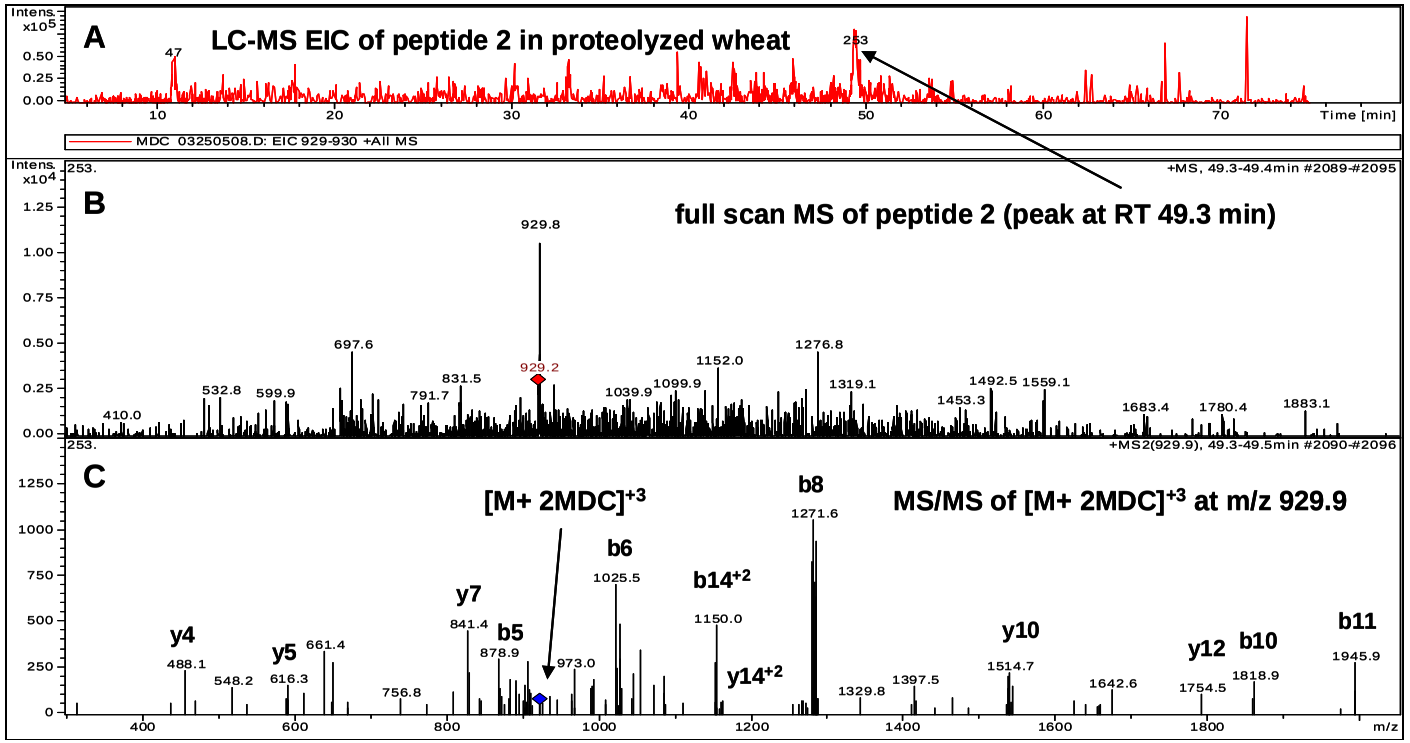


Figure 3.48

Specificity of TG2 within peptide 6 using a RhC tag



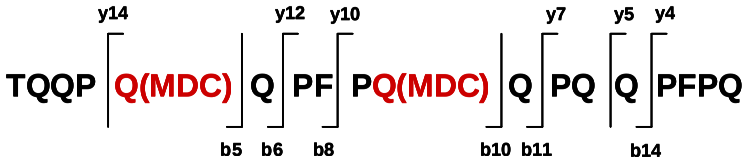


Figure 3.49

Specificity of TG2 within peptide 2 using a MDC tag

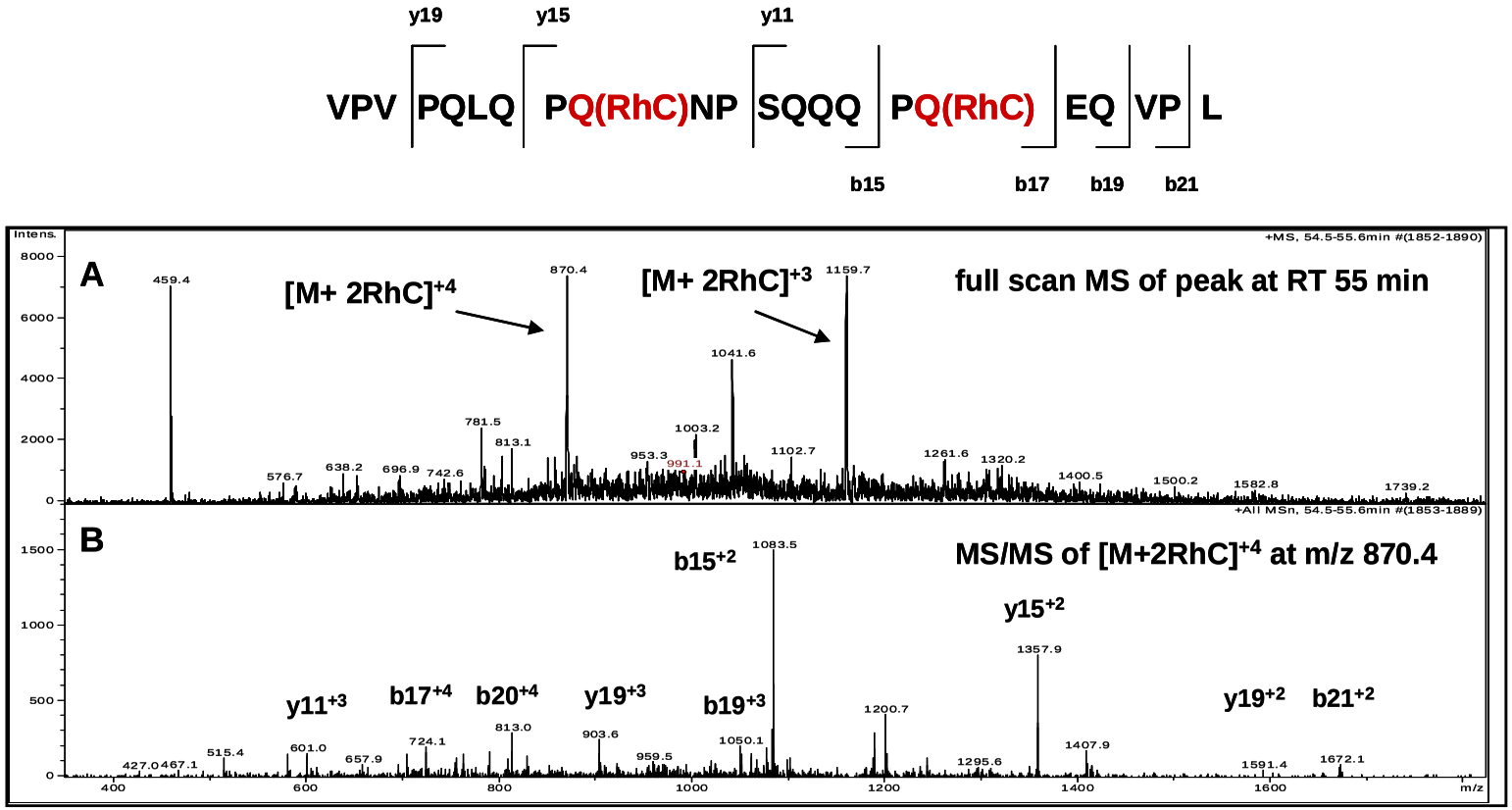
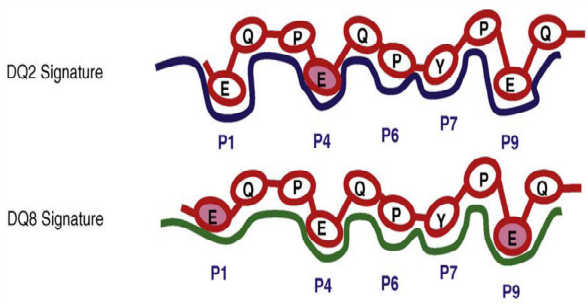


Figure 3.50

Specificity of TG2 within peptide 3 using a RhC tag



preferred binding characteristics:

prolines @ P1, 3, 5 and 8

glutamines @ P1, 4, 6, 7 and 9

glutamic acids @ P1, 4 and 9

peptide 1: **LQPQNPSQQQPQEQVPL**

1 2 3 4 5 6 7 8 9

peptide 2: **TQQPQQPFPQQPQQPFPQ**

1 2 3 4 5 6 7 8 9

peptide 4: **RPQQPYPQPQPQY**

peptide 5: **QPQQPFPQTQQPQQPFPQ**
QPQQPFPQTQQPQQPFPQ

Figure 3.51

Relationship between the specificity of TG2 and the preferred binding register of the core 9-residue immunogenic epitope sequence to DQ2/8

(adapted from [29])

Figure 3.52

Database search of the Triticeae tribe for peptide 1 (LQPQNPSQQQPQEQVPL)

3.5 References

[1] Shan, L.; Molberg, O.; Parrot, I.; Hausch, F.; Filliz, F.; Gray, G., M.; Sollid, L.; Khosla, C. *Science* **2002**, *297*, 2275-2279.

[2] Koning, F.; Roepstorff, P.; Sollid, L.; McAdam, S. *Journal of Experimental Medicine* **2000**, *191*, 603-612.

[3] Heap, G.; van Heel, D. *Seminars in Immunology* **2009** (in press).

[4] Jabri, B.; Kasarda, D.; Green, P. *Immunological Reviews* **2005**, *206*, 219-231.

[5] Green, P.; Cellier, C. *The New England Journal of Medicine* **2007**, *357*, 1731-1743.

[6] Cassinotti, A.; Birindelli, A.; Clerici, M.; Trabattoni, D.; Lazzaroni, M.; Ardizzone, S.; Colombo, R.; Rossi, E.; Bianchi-Porro, G. *American Journal of Gastroenterology* **2009**, *104*, 195-217.

[7] Qiao, S.; Sollid, L.; Blumberg, R. *Current Opinion in Immunology* **2009**, *21*, 111-117.

[8] Mowat, A. *Lancet* **2003**, *361*, 1290-1292.

[9] Clark, D.; Mycek, M.; Neidle, A.; Waelsch, H. *Archives of.Biochemistry and Biophysics* **1957**, *79*, 338-354.

[10] Pisano, J.; Finlayson, J.; Peyton, M.; *Science* **1968**, *160*, 892-893.

[11] Griffin, M.; Casadio, R.; Bergamini, C. *The Biochemical Journal* **2002**, *368*, 377-396.

[12] Fleckenstein, B.; Molberg, O.; Qiao, S.; Schmid, D.; von der Mulbe, F.; Elgstoen, K.; Jung, G.; Sollid, L. *The Journal of Biological Chemistry* **2002**, *277*, 34109-34116.

[13] Molberg, O.; McAdam, S.; Lorner, R.; Quarsten, H.; Kristiansen, C.; Madsen, L.; Fugger, L.; Scott, H.; Noren, O.; Roepstorf, P.; Lundin, K.; Sjostrom, H.; Sollid, L. *Nature Medicine* **1998**, *4*, 713-717.

[14] Aeschlimann, D.; Paulsson, M.; *Thrombosis and Haemostasis* **1994**, *71*, 402-415.

[15] Elli, L.; Bergamini, C.; Bardella, M.; Schuppan, D. *Digestive and Liver Disease* **2009** (in press).

[16] Vader, L.; de Ru, A.; van der Wal, Y.; Kooy, Y.; Benckhuijsen, W.; Mearin, L.; Drijfhout, J.; van Veelen, P.; Koning, F. *Journal of Experimental Medicine* **2002**, *195*, 643-649.

[17] Arentz-Hansen, H.; Korner, R.; Molberg, O.; Quarston, H.; Vader, W.; Kooy, Y.; Lundin, K.; Loning, F.; Roepstorff, P.; Sollid, L.; McAdam, S. *Journal of Experimental Medicine* **2000**, *191*, 603-612.

[18] Marzari, R.; Sblattero, D.; Florian, F.; Tongiorgi, E.; Not, T.; Tommasini, A.; Venture, A.; Bradbury, A.; *Journal of Immunology* **2001**, *166*, 4170- 4176.

[19] Caputo, I.; D'Amato, A.; Troncone, R.; Auricchio, S.; Esposito, C. *Amino Acids* **2004**, *26*, 381-386.

[20] Sjostrom, H.; Lundin, K.; Molberg, O. *Scandinavian Journal of Immunology* **1998**, *48*, 111-115.

[21] Marsh, M. N. *Nature Medicine* **1997**, *3*, 725.

[22] Piper, J.; Gray, G.; Khosla, C. *Biochemistry* **2002**, *41*, 386-393.

[23] Mamone, G.; Ferranti, P.; Melck, D.; Tafuro, F.; Longobardo, L.; Chianese, L.; Addeo, F. *FEBS Letters* **2004**, *562*, 177-182.

[24] Shan, L.; Qiao, S.; Arentz-Hansen, H.; Molberg, O.; Gray, G.; Sollid, L.; Khosla, C. *Journal of Proteome Research* **2005**, *4*, 1732-1741.

[25] Lundin, K.; Nilsen, E.; Scott, H.; Loberg, E.; Bratlie, J.; Skar, V.; Mendez, E.; Lovik, A.; Kett, K. *Gut* **2003**, *52*, 1649-1652.

[26] Picarelli, A.; Di Tola, M.; Sabbatella, L.; Gabrielli, F.; Di Cello, T.; Anania, M.; Mastracchio, A.; Silano, M.; De Vincenzi, M. *American Journal of Clinical Nutrition* **2001**, *74*, 137-140.

[27] Marti, T.; Molberg, O.; Li, Q.; Gray, G.; Khosla, C.; Sollid, L. *The Journal of Pharmacology and Experimental Therapeutics* **2005**, *312*, 19-26.

[28] Kilmartin, C.; Wieser, H.; Abuzakouk, M.; Kelly, J.; Jackson, J.; Feighery, C. *Digestive Diseases and Sciences* **2006**, *51*, 202-209.

[29] Dorum, A.; Qiao, S.; Sollid, L.; Fleckenstein, B. *Journal of Proteome Research* **2009**, *8*, 1748-1755.

CHAPTER 4

Development of an HPLC-MS/MS Assay for the Quantitative Detection of Novel Immunogenic Gluten Peptides

4.1 Introduction

The ability to be able to detect and quantify trace levels of dietary gluten from wheat, barley and rye in commercially available food and consumer products is important because of the established relationship between exposure to gluten and human health. Dietary gluten is a principal trigger of a variety of immune diseases including food allergies and intolerances. An analytical method that has the capability to detect and quantify trace levels of gluten, as accurately and sensitively as possible, is clearly needed in order to address worldwide product safety issues.

For the purposes of gluten-free labeling and certification, there currently exists some controversy surrounding the type of analytical methods (and their respective gluten standards used), considered acceptable to determine and quantify the amount of gluten present in food and consumer products. Several organizations (such as the FDA, the Codex Alimentarius Commission (a joint committee with delegates from both the Food and Agriculture Organization of the United Nations [FAO-UN] and the World Health Organization [WHO]) and the Association of Analytical Communities (AOAC)) all endorse different methods [1, 2]. These

methods are variations of an antibody-based enzyme-linked immunosorbent assay (ELISA), which is based on a specific monoclonal antibody(s) that cross-reacts with a specific sequence in a gluten prolamin protein. In the USA, the FDA (Food and Drug Association) supports all ELISA methodologies, but currently has a mandate to establish the best analytical method possible, in its revised food allergen labeling proposal [3].

The “ω-gliadin ELISA” assay is one method that is officially endorsed by the AOAC. This is a sandwich assay that is based on monoclonal antibodies to ω-gliadin [4]. Sandwich assays require two epitopes (antibody binding sites). Should only one epitope exist in the sample, then that gluten protein would not be positively identified by this method. One advantage to using this method for establishing wheat gluten content in a sample is that the ω-gliadin fraction of wheat gliadin does not denature when heated. Other gluten fractions of wheat do denature when heated. Therefore, this method can be used to assess the content of wheat ω-gliadins (along with a small number of similar types of proteins found in rye and barley) in foods that are native (uncooked) and in foods that have been heated (due to cooking or processing). This method cannot be used to quantify hydrolyzed gluten proteins in foods. Partially hydrolyzed gluten is found in many food products that contain components that originate from wheat starch hydrolysates.

Another method endorsed for gluten determination is the “R5 ELISA”. This is also a sandwich ELISA assay and is based on the R5 monoclonal antibody to the pentapeptide “QQPFP” epitope and other similar epitopes that are found in all wheat, barley and rye prolamin proteins. Because QQPFP is found in all three

grains, this assay detects more gluten proteins than its counterpart (ω-gliadin ELISA). This QQPFP epitope has also been found to be resistant to heat. Therefore, one advantage to this assay is that it can detect more of the gluten content present in heated and uncooked foods than the ω-gliadin ELISA assay. This method cannot be used to quantify glutelin proteins (the polymeric fraction of gluten proteins) or hydrolyzed proteins.

Overall, applications of both of these two ELISA assays have proven successful in gluten analysis, but each has its limitations (as noted in the previous paragraphs). Therefore, in order to ensure product safety, it is imperative to know which ELISA method to use in each analytical application [5, 6 and 7]. Variations of both endorsed ELISA methods are under development (as described in the work by Sousa [8] and Dekking [9]) in order to compensate for these limitations and provide a more comprehensive assay. It is important to note, however, that it is practically impossible to provide a complete representation of gluten in food products due the complexity and heterogeneity of the proteins. The immunogenicity of gluten in a product varies also, depending on the variety and source of the grains that were used as ingredients and the procedures that were used for product processing. Notwithstanding, many gluten peptide sequences have been found to be homologous within one species of grain and others across several species. For example, certain sequences of wheat α-gliadins can also be found in barley and rye. Although these ELISA assays do recognize some specific sequences, such as “QQPFP”, which are present in various types of wheat gluten proteins, they may not be present in other celiac-active grains (barley and/or rye). In other words, these

ELISA assays are not comprehensive; nor to they yield specific information on the presence of the actual disease-causing peptides, those which are actually released by digestive processes in the gut.

Several researchers have presented work focused on gluten protein and peptide quantification using mass spectrometric methods. HPLC-MS has been applied successfully in the quantification of gluten exorphin peptides in biological fluids [10, 11]. Although this group of five small peptides studied here are in fact gluten peptides, they are of a different type than the immunostimulatory peptides that are involved in the pathogeneses of the destruction of intestinal tissue in celiac disease. Quantitative analysis of gluten proteins using mass spectrometry has also been done using HPLC-MS [12] and MALDI-TOF-MS [13, 14 and 15]. These methods rely on the visualization of gluten proteins in food by observation of characteristic patterns of wheat gliadins and glutenins and do not focus on the physiologically relevant peptides, discussed in this research.

Although these methods have proven to be valuable in the characterization of gluten and in the assessment of total gluten content, each has limited application in trace gluten analysis in food. Small amounts of gluten can easily find their way into food and consumer products from cross contamination in ingredient handling, by improper transporting or processing equipment maintenance procedures, from hidden or unlabeled ingredients in a product, or possibly, by the incomplete removal of gluten during processing. By establishing the presence of any of these gluten peptides in a product, it would affirm that the product is unsafe for use by those with gluten sensitivities. Because prolonged exposure to even the smallest amounts of

gluten can result in severe damage to the small intestine, there exists a clear need to implement a methodology that can establish the basic presence or absence of gluten as accurately and sensitively as possible.

4.1.1 Quantitative LC-MS/MS assay development

HPLC-MS has been used successfully to study trace contamination of a variety of analytes in food [16 - 20]. Therefore, it is quite feasible that such a method could be developed to target trace levels of immunogenic gluten proteins and peptides in food. My initial efforts to this end utilized an enzymatic/chemical labeling approach which afforded the identification of seven physiologically relevant immunogenic gluten peptides. Verification of the structure of these seven peptides was followed by the experimental determination of their presence / absence in native celiac active and other grains (Chapter 3). The ability to detect and ultimately quantify trace levels of any of these immunogenic gluten peptides in complex native and processed food samples, would provide an excellent means to render products free from gluten.

Utilizing the excellent analytical capabilities of HPLC-QQQ-MS, a method was developed and optimized in order to provide the best possible accuracy, and sensitivity for the quantitative detection of trace levels of the seven immunogenic wheat gluten peptides.

4.1.2 HPLC-MS for gluten quantification

To achieve the best sensitivity, accuracy and specificity possible in the quantitative detection of the seven gluten peptides, the HPLC-ion trap-MS method

(developed in Chapter 3) was transferred to the HPLC-triple quadrupole (QQQ) mass spectrometry system. This system was better able to meet the requirements for trace quantitative detection.

The established technology of HPLC-QQQ-MS/MS is well suited for the comprehensive task of detection and quantification of immunogenic gluten related peptides in complex matrices. The anticipated sensitivity for LC-MS operation in such peptide analysis was expected to permit the detection and quantification well below the parts per million (ppm) level, corresponding to detection limits reported for the most sensitive ELISA methods currently available [5, 21].

Tandem mass spectrometry, as performed on a scanning instrument such as a triple quadrupole mass spectrometer (QQQ-MS) operated under multiple reaction monitoring (MRM) conditions, provides better accuracy and sensitivity for the quantitative detection of target or unknown peptides, than if performed on an ion trap. Under MRM conditions, there is less dependence on space charge issues that occur within an ion trap (i. e. matrix ions that fill up the trap) that result in fluctuation and/or decreased target analyte ion signal and/or reduced quantitative accuracy.

Limitations to the applicability of this tandem mass spectrometric technique become important to consider, however, once the size of the target peptides increase much above 3000 m/z (assuming singly charged ions). The level of achievable fragmentation and overall sensitivity of collision-induced dissociation (CID) detection decreases, due to the reduced capability of internal energy transfer in gas phase collisions (as noted by the center of mass collision theory [22]). Electrospray ionization offers the ability to create multiply charged parent ions, which

typically result in CID product ions ranging in mass 500-1500 m/z. This allows sequencing information to become available for peptides of MW greater than 3000. However, a caveat exists in this situation, because the charge states of product ions resulting from CID of multiply charged (z>2) parent ions are sometimes difficult to determine without some prior knowledge of their sequence. This creates challenges in analyzing the data obtained for large peptides.

4.2 Experimental

4.2.1 Materials and Reagents

4.2.1.1 Chemicals

Pepsin, trypsin, chymotrypsin, $KHNaPO_4$ and NaOH were obtained from Sigma-Aldrich (St. Louis, MO, USA). HPLC-grade acetonitrile, dithiothreitol (DTT), ethanol, isopropanol, tris-(hydroxymethyl) aminomethane and HCl were obtained from Fischer Scientific. Water was obtained from an in-house Milli-Q water purification system (Millipore, Billerica, MA, USA).

4.2.1.2 Peptide standards

Synthetic standards of peptides (LQPQNPSQQQPQEQVPL, PQQSGQGVSQSQQQSQQQ, TQQPQQPFPQQPQQPFPQ, FPLQPQQSF, VPVPQLQPQNPSQQQPQEQVPL, RPQQPYPQPQPQY, QPQQPFPQTQQPQQPFPQ, PQQQFPQTQQPQQPFPQP, QPQQPLPQPQQPF and PQQSPF were obtained from Thermo Electron, Ulm, Germany and were analyzed using HPLC-MALDI-TOF mass spectrometry in order to determine their exact molecular weights and purity (>90% pure).

4.2.1.3 Calibration and quality control standards, procedure and analytical blanks

Calibration standards were prepared as cocktails of 6 and 7 of the synthetically prepared target peptides [(1) LQPQNPSQQQPQEQVPL, (2) TQQPQQPFPQQPQQPFPQ, (3) VPVPQLQPQNPSQQQPQEQVPL, (4) RPQQPYPQPQPQY, (5) QPQQPFPQTQQPQQPFPQ, (6) PQQSPF and (7) QPQQPLPQPQQPF]. Standards were prepared using appropriate aliquots taken from individual stock peptide solutions of each peptide, prepared in 80% / 20% water / acetonitrile. Calibration standard concentration levels were as follows: (0.1, 0.3, 1, 3, 10, 30, 100 and 300) pg/µL and (1, 3, 10, 30 and 100) ng/µL. This corresponded to a range of about 3 ppb to 100 ppm (or ng/mg).

Two types of quality control standards were prepared. The first type was prepared by spiking stock peptide solutions at concentration levels that matched the solvent calibration standards into aliquots of corn flour that had already been enzymatically digested. The second type of quality control standard was prepared by spiking the stock peptide solutions at concentration levels that matched the solvent calibration standards into aliquots of water that had also been enzymatically treated.

Two types of procedure blanks were prepared. The first type used corn flour as a representative gluten-free matrix and was enzymatically digested with no target peptides spiked in. The second type of procedure blank used water as the matrix and was also enzymatically treated with no target peptide spiked in.

Analytical blanks consisted of aliquots of the HPLC mobile phase A (95% water / 5% acetonitrile (ACN) + 0.025% trifluoroacetic acid (TFA)).

4.2.1.4 Samples

Stone ground whole grain corn flour, vital wheat gluten flour, stone ground whole wheat flour (brand 1) and whole wheat flour (brand 2) were obtained from a local supermarket (Raleigh, NC, USA).

4.2.2 In-vitro proteolytic digestion procedure

Samples of native cereal grains were proteolyzed with pepsin, trypsin, chymotrypsin (PTC) according to the following protocol:

30 mg of homogenized sample was dissolved into 1mL pepsin solution (0.01M HCl adjusted to pH 2), to establish a 1:100 pepsin to protein ratio; heat (at 38˚C) and shake for 2 hours. To each sample, add 50 µL of a 50 mM phosphate buffer and 35 µL of a 0.1 M NaOH solution, to establish the pH between 7-7.5. To each sample, add 25 µL of a 1:100 trypsin / chymotrypsin to protein solution (in 50 mM phosphate). Heat (at 38˚C) and gently shake for 30 minutes. For each sample, take 200 µL of the supernatant and add 200 µL of Mobile phase A (95% water / 5% acetonitrile (ACN) + 0.025% trifluoroacetic acid (TFA)). Spin down all samples and aliquot 200 µL for HPLC-MS/MS analysis.

4.2.3 Instrumentation and analysis conditions

4.2.3.1 Assay development

HPLC-ESI-QQQ-MS/MS was used in the development of this quantitative method for the analysis of the 7 target peptides. High performance liquid chromatography (HPLC) was performed using an Agilent 1200 Rapid Resolution LC system (Agilent Technologies, Santa Clara, CA, USA). The HPLC was coupled to an

Agilent 6410 QQQ mass spectrometer (Agilent Technologies, Santa Clara, CA, USA), operated in positive atmospheric pressure electrospray ionization (AP-ESI) mode.

4.2.3.2 Optimized assay

The assay was performed on the same HPLC was coupled to the Agilent 6410 QQQ mass spectrometer, operated in positive atmospheric pressure electrospray ionization mode. Mass calibration of the mass spectrometer was conducted according to the manufacturer's documented procedures, using the Agilent tune compound. Daily mass calibration checks were performed in order to assure instrument response was accurate and consistent. Sample analyses were performed in multiple reaction monitoring mode (MRM) employing time programming to obtain maximum possible sensitivity and specificity for each peptide.

The atmospheric pressure ionization (API) source was operated at a capillary voltage of 3800 V (half moon electrode in the API chamber), using a nitrogen drying gas at a flow rate of 9.5 L/minute and heated to 350˚C. The nebulizer was operated at 45 PSI. The fragmentation voltage (capillary exit voltage) used for all target peptides was 150V, except for peptide 6, whereby 120V was used. Each parent to product transition was monitored for 100 ms, except for time segment 5, whereby each transition was monitored for 50 ms. Table 4.1 lists the seven target peptide sequences, their respective monoisotopic molecular weights, time program segments, parent to product ion transitions and respective collision energies that were used for the analyses.

Typically, 20 µL of sample was injected into the HPLC-MS system. For ultra-trace level analyses, a sample volume of 100 µL was used. Sample aliquots were injected onto a C18 reversed phase Ascentis Express (Sigma-Aldrich/Supelco) chromatography column of dimensions 2.1 mm i.d. X 150 mm with 2.7 µm superficially-porous silica particles. For the separation of peptides, a gradient of 0-22% B over 35 minutes, followed by 22-60% B over 10 minutes was used at 25 ˚C and a flow rate of 300 µL/minute. Mobile phase A was 95% water / 5% acetonitrile (ACN) + 0.025% trifluoroacetic acid (TFA). Mobile phase B was 5% water / 95% acetonitrile (ACN) + 0.025% trifluoroacetic acid (TFA).

4.3 Results and Discussion

4.3.1 In-vitro proteolytic digestion

Prior to analysis by HPLC-MS, wheat gluten samples were treated with various proteases *in-vitro*, using conditions and enzymes that model the gastric and duodenal protein digestion in humans. This enzymatic digestion procedure was designed to release some or all of the seven target peptides from a gluten-containing sample. This procedure had been developed and tested using a variety of other native gluten-containing and gluten-free grain samples and found to work successfully and reproducibly, thus releasing the target peptides when gluten proteins were present in the grain (Chapter 3).

In brief the procedure was as follows: 30 mg samples of native grains were treated with pepsin, trypsin, chymotrypsin, elastase and carboxypeptidase A (PTCECA) according to the following protocol. 30 mg of homogenized sample was

dissolved into 1mL pepsin solution (0.01M HCl adjusted to pH 2), to establish a 1:100 pepsin to protein ratio; heat (at 38˚C) and shake for 2 hours. To each sample, add 50 µL of a 50 mM phosphate buffer and 35 µL of a 0.1 M NaOH solution, to establish the pH between 7-7.5. To each sample, add 25 µL of a 1:100 trypsin/chymotrypsin to protein solution (in 50 mM phosphate); heat (at 38˚C) and shake for 2 hours. Heat to 95˚C for 15 minutes and cool back down to room temperature. To each sample, add 20 µL of a 1:500 elastase solution (in 50 mM Tris); adjust pH to 7.5 with HCl; heat (at 38˚C) and shake for 2 hours. Heat to 95˚C for 15 minutes and cool back down to room temperature. To each sample, add 25 µL of a 1:100 carboxypeptidase A solution (in 50 mM phosphate); heat (at 38˚C) and shake for 2 hours. Heat to 95˚C for 15 minutes and cool back down to room temperature. Spin down samples and aliquot 200 µL of the supernatant for HPLC-MS analysis.

It was found that this digestion procedure took approximately 10 to 11 hours to complete and was very labor intensive. A procedure was desired that was more streamlined, efficient and cost effective. Therefore, in an effort to reduce the time involved and to maximize the yield of the gluten peptides produced, each step of the procedure was re-evaluated. A variety of conditions and analytical parameters were explored as to their effect on the yield of the peptides. HPLC-ESI-QQQ MS/MS detection was used in all experiments.

4.3.1.1 Evaluation of alternative cocktails of digestive (gastric and pancreatic) enzymes

The enzymes used for the working *in-vitro* digestion procedure were pepsin, trypsin, chymotrypsin, elastase and carboxypeptidase A (PTCECA). The procedure was re-evaluated using various combinations of these enzymes to determine if all enzymes were actually required in order to release the target gluten peptides. Figure 4.1 displays the differences in yield of target peptides 1-7 from the digestion of wheat flour using enzyme cocktails: (A) all five enzymes (pepsin, trypsin, chymotrypsin, elastase and carboxypeptidase A), (B) all but carboxypeptidase A (pepsin, trypsin, chymotrypsin, elastase) and (C) only pepsin, trypsin and chymotrypsin. From these experiments, it was determined that pepsin, trypsin and chymotrypsin were essentially all that were required to produce a reasonable yield of the 7 target peptides. The absence of pepsin, trypsin or chymotrypsin resulted in the absence of the target peptides. Removing the steps in the procedure which utilized elastase and carboxypeptidase A would then efficiently reduce the sample preparation time by several hours.

4.3.1.2 Evaluation of proteolytic digestion (reaction) times

The amount of time that each enzyme is allowed to participate in the digestion of food corresponds directly to the amount of proteolyzed products that result. The existing procedure used a time period of 2 hours, which represented an estimate of the approximate amount of time that it would take to digest an average meal. In actuality, it is difficult to deduce the exact time that each enzyme spends working *in-vivo*, therefore an estimated time of 2 hours was used. In an effort to investigate how

long it took to produce the highest yield of the seven target gluten peptides, a series of *in-vitro* experiments were performed that monitored peptide formation during a digest of wheat reacting with pepsin, trypsin and chymotrypsin. These experiments were designed to hold the digestion time for pepsin constant at four different time periods, while varying the amount of time that trypsin and chymotrypsin were allowed to react.

Figure 4.2 illustrates results obtained by this set of experiments. The graph was normalized to the longest digestion time of 120 minutes for pepsin and 120 minutes for trypsin and chymotrypsin. Figure 4.2(A) shows the yield of the peptides with the reaction time for pepsin set to 15 minutes, while trypsin and chymotrypsin reaction times range from 15, 30, 60 to 120 minutes. Parts (B), (C) and (D) of this figure show how the yield of peptides changes while holding the pepsin reaction time constant at 30, 60 and 120 minutes respectively, and again, varying the trypsin and chymotrypsin reaction times. The results shown represent peptides 1, 2, 3, 4 and 6. From these results it was concluded that the highest yield of peptides was obtained with a pepsin reaction time of 2 hours and a trypsin / chymotrypsin reaction time of 30 minutes.

4.3.1.3 Evaluation of various protein to enzyme ratios

The ratio of protein to enzyme also affects the yield of proteolyzed products. Wheat was proteolyzed with pepsin, trypsin and chymotrypsin at protein to enzyme ratios 20:1, 100:1 and 500:1 and allowed to digest over the following four time periods: (1) pepsin 30 minutes + trypsin and chymotrypsin 30 minutes, (2) pepsin 30 minutes + trypsin and chymotrypsin 120 minutes, (3) pepsin 120 minutes + trypsin

and chymotrypsin 30 minutes and (4) pepsin 120 minutes + trypsin and chymotrypsin 120 minutes.

Figure 4.3 shows how the formation of peptides 1, 2, 3, 4 and 6 varies with digestion time and protein to enzyme ratio. From these results it appeared that the protein to enzyme ratio that formed the highest peptide yield was 100 to 1.

4.3.1.4 Evaluation of various procedures to quench proteolytic activity

One mechanism that is commonly used to terminate an enzyme's activity is to heat the reaction mixture in order to destroy the enzyme. It was postulated that this mechanism may also alter or hydrolyze some of the target peptides. Consequently, three alternative procedures to quench enzymatic reactivity were investigated: (1) addition of a solution of 50% water / 50% acetonitrile (ACN), (2) addition of HPLC mobile phase A (95% water / 5% ACN + 0.025% trifluoroacetic acid (TFA)) and (3) chill on ice for 30 minutes.

Figure 4.4 shows what happens to the yield of peptides after incorporating each of these quenching mechanisms into the procedure. Clearly, the highest yield of all peptides is seen from the experiment which utilized the addition of HPLC mobile phase A to quench the enzymatic reaction.

4.3.1.5 Evaluation of the need for an alcohol pre-extraction procedure

The inclusion of an aqueous alcohol (50% ethanol) pre-extraction step is commonly reported in literature where ELISA methodologies are utilized in the analysis of wheat gluten [7, 23]. Its purpose is to pre-extract the alcohol soluble

monomeric gliadin proteins from the remaining mix of the other proteins in the sample, thus improving the detection capabilities of the assay.

Therefore, to compare the efficiency of the direct proteolysis approach with this alternate methodology (which employs the aqueous alcohol pre-extraction step), the yield of peptides from the proteolysis of wheat gluten with PTC was compared. Figure 4.5 shows that the yield of peptides 3, 4 and 6 is actually higher without the alcohol pre-extraction step, while the yield of peptides 1 and 2 has actually more than doubled using with the alcohol pre-extraction. Based on the similarities in structure of all seven target peptides, an explanation for this was not forthcoming.

Repeating the experiment with wheat flour (versus wheat gluten), the opposite results were obtained for peptides 1 and 2. Figure 4.6 compares data from the digestion of wheat flour with pepsin, trypsin and chymotrypsin (PTC): (1) directly (no alcohol extraction), (2) after an alcohol (60% ethanol) pre-extraction without reducing agent (DTT), (3) after an alcohol (60% ethanol) pre-extraction with DTT, (4) after an alcohol (50% ethanol) pre-extraction without DTT and (5) after an alcohol (50% isopropanol) pre-extraction with DTT.

Interestingly, even using isopropanol (recent literature recommends this [21]) and with the addition of a reducing agent DTT (used to reduce disulphide bonds thus allowing polymeric glutenin subunits to become soluble in alcohol), the yield was still higher without the alcohol pre-extraction step. A possible explanation for this may be the contribution of target peptides that have formed from the direct digestion of the alcohol-insoluble polymeric glutenin proteins. These would not be extracted by the alcohol pre-extraction step. It has only recently been determined that glutenin

proteins are also immunostimulatory [24]. Glutenin proteins basically consist of multiple units of monomeric gliadins held together by an extensive network of cross-linked disulphide bridges [25]. The traditional understanding that gliadins and glutenins could be cleanly fractionated from each other using an aqueous alcohol extraction, based on their solubility characteristics, has given way to the more recent observations that some glutenin subunits are in fact alcohol soluble and some gliadins are actually alcohol insoluble [24, 25 and 26]. Therefore, the currently endorsed ELISA methodologies could conceivably miss detecting many gluten proteins because their alcohol extraction step did not extract them from the sample. Even with the addition of a reducing agent, designed to catch the alcohol soluble glutenin monomers, this method does not yield as many peptides as the direct proteolysis approach.

4.3.1.6 Improving detection limits

Efforts to improve the overall sensitivity of this assay involved experiments that evaluated the pre-concentration of samples, prior to HPLC-MS analysis. Following the completed proteolysis procedure, the supernatant was removed, taken down to dryness and reconstituted to a smaller volume with an appropriate solvent. Various solvents and solutions were evaluated, but results showed very little increase in the final yield of peptides. It appeared that significant losses in peptide yield occurred during the various stages of this pre-concentration procedure.

Further experiments involved increasing the sampling size of the initial wheat sample from 30 mg up to 300 mg. The enzyme concentrations were increased accordingly, in order to maintain the 100 to 1 (protein to enzyme) ratio. Results from

these experiments are shown in Figure 4.7. In theory, increasing the sample size ten times should have increased the yield of peptides by 10. However, losses in peptide yield from incomplete proteolysis, sample handling and an increased number of sample dry down/reconstitution steps needed in this procedure outweighed the benefits of using such a large sample size.

An increase in the amount of sample injected (from 10 µL to 100 µL) into the HPLC-MS did allow for a significant increase in sensitivity, which would be helpful for samples requiring ultra-trace component analysis.

4.3.1.7 Final optimized in-vitro digestion procedure

Upon evaluation of the aforementioned parameters and conditions, the best conditions from each were combined into a final condensed procedure, which is as follows:

30 mg of homogenized sample was dissolved into 1mL pepsin solution (0.01M HCl adjusted to pH 2) to establish a 1:100 pepsin to protein ratio; heat (at 38˚C) and gently shake for 2 hours. To each sample, add 50 µL of a 50 mM phosphate buffer and 35 µL of a 0.1 M NaOH solution, keeping the pH between 7-7.5. To each sample, add 25 µL of a 1:100 trypsin + chymotrypsin to protein solution (in 50 mM phosphate); heat (at 38˚C) and shake for 30 minutes. To quench the enzymatic reaction, take 200 µL of the supernatant and add 200 µL of Mobile phase A (95% water / 5% acetonitrile (ACN) + 0.025% trifluoroacetic acid (TFA)). Spin down samples and aliquot 200 µL for HP LC-MS/MS analysis.

4.3.2 High resolution chromatographic separation of proteolyzed wheat gluten peptides

The separation of proteolyzed wheat gluten peptides by HPLC resulted in complicated chromatograms, representing hundreds peptides. Previous experimental results (Chapter 3) showed that further optimization was required in order to provide better resolution of these eluting peptide components. Therefore, in order to achieve adequate separation and resolution of the seven target peptides from other peptides as well as other digest components, a variety of chromatographic columns, stationary phases, particles and pore sizes were evaluated.

The HPLC conditions and column chosen for a particular analysis should be done so as to provide adequate specificity and sensitivity for that analysis. For this type of HPLC-MS/MS analysis, chromatography conditions needed to be chosen to reduce the complexity of the sample entering the MS. This was necessary to reduce matrix ion suppression and interferences from common product ions that could hinder the specificity or sensitivity achieved by MS/MS. To that extent, a short 1-10 minute HPLC separation, typical of a quantitative analysis of one target analyte, would in no way be sufficient to reduce the number of co-eluting components that could interfere with the peptide analyses. Therefore, it was observed that HPLC separation times of approximately 60 minutes were needed, in order to minimize these matrix effects.

The choice of chromatographic column also plays a major role in the specificity and sensitivity for the analysis. Improving the separation of the components reduces the number of co-eluting components that can suppress

ionization or interfere in the MS/MS analysis. Improved resolution also results in sharper, more concentrated peaks, which would result in a better response in a concentration detector, such as UV and electrospray ionization mass spectrometry. Various columns and particle chemistries were evaluated. They consisted of the following: (1) SB C18 column of dimensions 2.1 x 50 mm using 1.8 µm porous particles, (2) Poroshell C18 column of dimensions 2.1 x 75 mm using 5-7 µm coated particles and 300A pores, (3) SB C18 column of dimensions 2.1 x 30 mm using 1.8 µm porous particles and (4) Asecntis Express C18 column of dimensions 2.1 x 150 mm using 2.7 µm superficially porous particles (1.7 µm particle core with a 0.5 µm porous layer and a C18 coating).

The column packed with the superficially porous fused core particles appeared to result in the best performance. The peaks were noticeably sharper than those seen with the smaller 1.8 µm particles. This improved chromatographic performance using these particles could be a result of the nature of the particles in this column and/or the increased length of this column. By placing a solid core at the center of the particle, the potential diffusion path length for the peptides through the 0.5 µm porous layer was shortened considerably, compared to that in a totally porous particle. This shorter diffusion path should reduce dispersion from resistance to mass transfer, thus reducing peak broadening. This type of particle, although larger in overall size, should theoretically, generate similar efficiency to a smaller totally porous particle without generating high backpressures. It was observed that the backpressures from this column were lower, as compared with those observed with the other columns evaluated.

For this application of gluten peptide analysis, the 2.7 µm superficially-porous silica particles in the Ascentis Express C18 column provided much better chromatographic peak shape and resolution for the separation of the seven target peptides.

4.3.3 Mass spectrometric detection of the seven target gluten peptides

Various mass spectrometric conditions and parameters were optimized in order to achieve the best possible sensitivity and specificity for the seven target peptides. In order to set up a quantitative analytical procedure for target analytes, the MS system first needed to be optimized for each analyte. Most instrument parameters were mass dependent and were optimized using a tuning mix (provided by the instrument manufacturer), in order to obtain the best possible ion transmission yet maintain mass resolution. This was initially done via an auto-tune function, which is available on the Agilent 6410 triple quadrupole system. Following the auto-tune, a fine tune was performed manually, in order to tweek any settings deemed necessary to improve the signal. Two key MS instrument parameters, however, were compound dependent, requiring individual optimization for each peptide. These parameters were fragmentor voltage and collision energy.

4.3.3.1 Fragmentor optimization

The “fragmentor” is the term Agilent Technologies uses to describe the variable voltage at the capillary exit in the electrospray interface (Figure 4.8). The capillary samples the ions formed by electrospray ionization (ESI) in the atmospheric chamber and brings them to the first stage of pumping in the mass spectrometer.

The potential drop between the capillary exit and first skimmer aids in the focusing and transmission of the ions through the skimmer region. In this region, the pressure is in the mtorr range. As such, the mean free path of the ions becomes long enough to allow them to gain kinetic energy. As collisions occur between ions in this region, the internal energy of the ions can increase and this can result in fragmentation (sometimes termed “up front CID”). This type of fragmentation is not desired in this particular type of assay, because it could reduce the molecular ion precursors specifically chosen for each peptide in the MS/MS process. Therefore, this fragmentation voltage is compound dependent as the bond strengths of the atoms in the compound and the transfer of kinetic to internal energy in a compound are dependent on the structure and molecular weight of each peptide.

In order to optimize the fragmentor voltage setting for each of peptides, flow injection analysis (FIA) was performed on each peptide individually. With each injection of a peptide standard, the fragmentor voltage was incremented by 30-40 volts. Figure 4.9 shows the full scan MS data for peptide 3, with the fragmentor voltage optimized to 150V. Figure 4.10 illustrates how the voltage setting of 150V would maximize both the $[M+2H]^{2+}$ (m/z 1240.9) and $[M+3H]^{+3}$ (m/z 827.6) precursor ions for peptide 3.

4.3.3.2 Collision energy optimization

Analogous to the fragmentation voltage, the collision energy (CE) voltage is also compound dependent. CE depends on the bond strength, molecular weight of the compound and the pathway in which the ion is formed (i. e. directly from the parent ion or through a series of sequential intermediates). The collision energy

needs to be optimized in order determine the optimal voltage to produce the most intense characteristic product ions that will be monitored for each peptide in the MS/MS analysis. The collision cell is actually a hexapole in the Agilent 6410 QQQ mass spectrometer. It has nitrogen introduced into it, which serves as the collision gas used to transfer kinetic energy into internal energy. The voltage on the collision cell sets the kinetic energy for the ion. However, unlike the fragmentor voltage, which is chosen to maximize the precursor ion intensity, the CE voltage is chosen to produce the most abundant product ions from the chosen precursor ion.

The CE optimization was also performed by FIA, whereby the CE voltage was incremented by 5 volts with each subsequent injection of peptide standard. Figure 4.11 shows the MS/MS spectra of all products ions formed from (A) the parent ion at m/z 1240.9 ($[M+2H]^{+2}$) and from (B) the parent ion at m/z 827.5 ($[M+3H]^{+3}$) from peptide 3, with the collision energies optimized to 30V and 15V respectively. The product ions chosen for the optimization are highlighted by the blue arrows. The voltages 30V and 15V were chosen as the optimal voltages based on the best intensity for these product ions, as shown in Figures 4.12 and 4.13.

4.3.3.3 Mass spectrometric mode of operation

A time-programmed multiple reaction monitoring (MRM) MS acquisition profile was developed. Time programming was used to maximize the MS-response and achieve the best sensitivity possible, through use of multiple segments based on the LC retention times of the peptides. Each segment was set so as only to monitor the MRM transitions for the target peptides that eluted during that particular time segment. Each segment monitored one peptide, except segment 5, whereby three

peptides (2, 5 and 7) were monitored. Dividing the chromatogram into segments allowed for longer dwell times to monitor the target ions, compared to the monitoring of all ions (14 ions) over the entire course of the analysis. It was observed that the longer dwell time improved the signal/noise ratio for the measurement of MS response by 2-3 times.

The MRM transitions were chosen to be both specific to the target peptide sequence and for sensitivity. All peptides except peptide 6 used the doubly charged parent ion ($[M+2H]^{+2}$) as the precursor ion for the MRM experiment. Peptide 6 being of low molecular weight (702 Da), only formed a singly charged ion ($[M+H]^{+}$). Product ions from each peptide were chosen based on sensitivity and specificity of the respective “b-” or “y-” type fragments that formed. The choice of ions was not always the most intensely responding fragments, but those that did not interfere with signals from other peptides or matrix components that were present in that particular time segment. Peptides 2 and 5 proved to be challenging, since they are isomers and were not well chromatographically resolved. They were resolved by MS/MS, however, since the product ions chosen for these peptides were different in MW and were specific for their respective sequences. It appeared the product ions for these two peptides favored cleavage between the amino acids F (phenylalanine) and P (proline), resulting in a “b8” fragment ion at m/z 956.6 for peptide 5 and a “b6” ion at m/z 726.3 for peptide 2. Figures 4.14 – 4.20 represent peptides 1 through 7, showing their respective HPLC-MS-MS product ion mass spectra, sequences of the product ions, types of fragment ions and quantification/confirmation parent to product transition ions for each.

An interesting observation was noted for parent to product MRM transitions where the product ion was at a higher m/z value than the precursor ion. In these instances it was observed that noise in the spectra was significantly reduced. This has important connotations when performing trace component analysis, such as contamination in food, where the target product ions are low in intensity. This condition happened to be applicable for MRM transitions for three of the seven target peptides. The m/z of the singly charged confirming product ion for peptides 1, 2 and 4 was above that of the doubly charged precursor ion. While the MS-intensity of these product ions was lower than that of other product ions (those whose m/z were below the precursor m/z), the level of noise for these transitions was seen to be 5-7 times lower. Therefore, the signal to noise (S/N) ratio was ultimately higher, which allowed better sensitivity. This indicated a real advantage in choosing product ions above the multiply charged precursor ion, whenever possible. An example of this situation is illustrated in the Figure 4.21, where the EIC of the confirming ion transition for peptide 4 (m/z 814.3 - 1221.8) shows basically no noise as compared with the level of noise that is seen in the quantification ion transition (m/z 814.3 – 407.2).

A representative time programmed MS/MS-extracted ion chromatogram (EIC) of a cocktail of all seven synthetic wheat gluten peptide calibration standards [(1) LQPQNPSQQQPQEQVPL, (2) TQQPQQPFPQQPQQPFPQ, (3) VPVPQLQPQNPSQQQPQEQVPL, (4) RPQQPYPQPQPQY, (5) QPQQPFPQTQQPQQPFPQ, (6) PQQSPF and (7) QPQQPLPQPQQPF] at a concentration 3 ng/mg, is shown in Figure 4.22. Figure 4.22(A) shows the primary

MRM parent to product transitions for each peptide. These are the most intensely MS-responding ions and were chosen as the transitions used for quantification purposes. The secondary MRM transitions are shown in Figure 4.22(B). These are the confirmation transition ions used for positive identification of the presence of the peptides in samples. Table 4.1 summarizes the target peptide sequences, the established time program segments, parent and product ion transitions and respective collision energies used for all analyses.

4.3.4 HPLC-MS/MS method performance

The overall performance of this HPLC-MS/MS assay was evaluated according to a variety of parameters, commonly used to assess the capabilities of such an assay and verify that it can meet the requirements deemed necessary to be used efficiently for the purposes of quantitative detection of target compounds in real samples.

4.3.4.1 Linearity, sensitivity, accuracy and precision

Linearity and sensitivity of the assay is demonstrated from data obtained from the evaluation of calibration data. Table 4.2 summarizes data from a representative regression analysis, from an analysis performed using a corn flour matrix (proven free of any gluten peptides) spiked with a standard cocktail of peptides 1-7. The respective limits of detection (LOD) and quantification (LOQ), correlation coefficient (R^2), as well as residuals for each peptide are all shown. These data demonstrate that this assay can quantify select immunogenic wheat gluten peptides over a range of about 10 to 1000 pg/mg of food. This corresponds to being able to measure the

peptides over a range of about 10 – 1000 ppb (parts per billion) in food. Detection and quantification at higher levels was also done, using a second calibration curve, covering a range from 1 ng/mg to 100 ng/mg (ppm). Representative calibration curves for peptides 1-6 over this range are shown in Figure 4.23.

4.3.4.2 Specificity

Specificity of the assay is demonstrated by monitoring two different MRM MS transitions for each of the seven target peptides (refer to Table 4.1).The most intense MS-responding ions constituted the primary MRM transition, which was used for quantification purposes. The less intense ions constituted the secondary MRM transition, which were used for confirmation purposes.

4.3.4.3 Spike recovery

Spike recovery experiments were conducted to test the assay's absolute ability to detect the seven peptides, free from interferences, when spiked in both an analyte-free and analyte-containing matrix (i. e. corn versus wheat). Two spike recovery experiments were performed, whereby a standard cocktail of peptides 1-7 was spiked: (1) into corn flour matrix at two levels: (a) representing 0.06 ng/mg peptide in food and (b) representing 30 ng/mg peptide in food; and (2) into wheat flour matrix representing 60 ng/mg peptide in food. Figure 4.24 shows representative MS/MS-extracted ion chromatograms of: (A) a 60 ppb solvent standard cocktail of peptides 1-7, (B) an analytical blank and (C) a 60 ppb standard cocktail of peptides 1-7 spiked into corn flour. A calibration curve was prepared for each peptide which bracketed the concentration of peptide spiked into each matrix (proteolyzed corn and

wheat). The areas of each of the peptide peaks in the spiked samples were measured and the corresponding quantity of that peptide in the samples was determined using the respective calibration curve. The calculated results of these recovery experiments are tabulated in Table 4.3. Percent (%) accuracy was calculated as [(measured value/theoretical value) x 100]. The “measured value” refers to the peak area of each peptide measured in matrix and the “theoretical value” refers to the corresponding peak area measured in the solvent standard. These data show that the range of accuracy of detection of the 0.06 ng/mg and 30 ng/mg standards spiked into corn were 69-104% and 92-108% respectively. The lower level (0.06 ng/mg) spike represented a level near the LOQ of the assay. This concentration is equivalent to about 60 ppb. The 30 ng/mg level represents a higher level, perhaps more representative of wheat contamination that could be encountered in food. An even higher concentration (60 ng/mg) was required to determine spike recovery in proteolyzed wheat, due to the presence of native wheat gluten peptides in the matrix. Recovery from wheat was based on subtracting endogenous peptide levels from a wheat sample that had not been spiked. The peak area for each peptide in the spiked wheat sample was determined by subtracting the corresponding background from the blank wheat sample.

4.3.4.4 Matrix effects and robustness

Examination of procedure blanks and analytical blanks showed no matrix effects or carryover. Solvent standards, quality controls and procedure blanks were all found to be stable for well over eight months, kept at 0˚C.

The use of isotopically labeled internal standards was considered, but not implemented, because of cost and also because the results from the qualification experiments demonstrated the method to be acceptable. Furthermore, because a target analyte extraction step was not required and because isotopically labeled standards could not be used to gage the formation of the target peptides from the direct enzymatic digestion, their incorporation would only serve as an injection standard, of sorts. Actually, an injection standard is not needed either, considering that the injection to injection precision, demonstrated over the course of these experiments, was found to be more that adequate (<12% overall).

These experiments were carried out over the course of many months. Over this timeframe normal variations in laboratory conditions (such as temperature, mobile phase preparation, age and status of the chromatography columns used) caused retention times of the peptides and segment times in the MS-time programmed acquisition to shift around slightly. This was not seen as a problem because the peptides were always analyzed in combination with standards and quality controls, thus allowing the proper retention times to be verified on a run to run basis.

4.3.4.5 Analysis of wheat and corn flours for the seven target gluten peptides

An overall demonstration of the performance of the optimized HPLC-MS/MS method involved the analyses of samples of native wheat and corn flour, for the quantitative determination of the seven physiologically relevant gluten peptides. Confirmation of the presence (in wheat) and absence (in corn) of peptides 1 and 5 are shown in Figures 4.25 and Figure 4.26, respectively.

Figures 4.27 shows representative MS/MS-extracted ion chromatograms (EICs) of proteolyzed native wheat and corn flours evaluated by this HPLC-ESI-MS/MS assay in order to confirm the presence of all seven target peptides in wheat flour and their absence in corn flour. Traces shown include (A) a solvent standard of the seven target peptides at a concentration of 20 ng/mg, (B) corn flour and (C) wheat flour. The concentration of the standard (20 ng/mg) is representative of the proposed FDA and Codex Alimentarius guideline of 20 ppm total gluten. This is considered an acceptable quantity of gluten in any product tested. The intensity of peptide 7 was found to be quite low in native wheat flour. Therefore, it may not be practical to use this particular peptide as a marker peptide for the presence of trace amounts of gluten in food samples. The concentration of each of these seven immunogenic peptides was calculated to be: (1) 239, (2) 237, (3) 255, (4) 42.5, (5) 246, (6) 13.5 and (7) 1.98 ng/mg in wheat flour. Relating these concentrations of individual gluten peptides to a total gluten content value is difficult, if not impossible, because this assay is not designed to do that. The purpose of this assay is to detect the presence of any of these seven peptides, which would infer that immunogenic wheat gluten is present in the product tested, thus rendering that the product would not be safe for consumption by anyone with a type of gluten sensitivity. The ability to quantify the peptides down to the low ppb level offers the unique capability for ultra-trace detection of wheat gluten in products that may be contaminated with gluten.

4.4 Conclusions

A sensitive, specific and accurate analytical LC-MS/MS method has been developed for the quantitative detection of seven uniquely identified immunogenic wheat gluten marker peptides in foods.

Experimental results from the analysis of native cereal grains has proven that this method can accurately, sensitively and reproducibly quantify seven immunogenic wheat gluten peptides over the range 10 pg/mg to 100 ng/mg. The final stage of development for this methodology will be to test its usefulness in a practical application. Analysis of common commercially available food and consumer products will be described next, in Chapter 5.

Peptide #	Sequence	Monoisotopic MW (Da)	Time Segment	MRM Transition *	Product Ion Type	Collision Engergy (V)
1	LQPQNPSQQQPQEQVPL	1958.0	4	**980.7(+2) - 866.6(+2)** *980.7(+2) – 1150.7*	b15 b10	20 30
2	TQQPQQPFPQQPQQPFPQ	2148.0	5	*1075.9(+2) – 1195.6* **1075.9(+2) - 956.6**	y10 b8	30 25
3	VPVPQLQPQNPSQQQPQEQVPL	2478.3	6	**1240.9(+2) – 1126.7(+2)** *1240.9(+2) – 762.4*	b20 b7	25 30
4	RPQQPYPQPQPQY	1627.8	3	*814.6(+2) – 1221.8* **814.6(+2) – 407.3**	b10 y3	25 30
5	QPQQPFPQTQQPQQPFPQ	2148.0	5	*1075.9(+2) - 726.3* **717.6(+3) – 244.1**	b6 y2	27.5 17.5
6	PQQSPF	702.3	2	**703.4 - 441.4** *703.4 - 263.3*	b4 y2	25 35
7	QPQQPLPQPQQPF	1530.8	5	**767.1-263.2** 767.1 - 917.6	Y2 b8	15 15

* charge states other than 1 is shown in paranthees
MRM transitions in **BOLD** were used for quantification
MRM transitions in *italics* were used for confirmation

Table 4.1

MS/MS conditions

Peptide #	Sequence	Correlation Coefficient (R2)	Residual Range (% accuracy)	LOD (pg/mg)	LOQ (pg/mg)
1	LQPQNPSQQQPQEQVPL	0.9996	78 - 104	3.5	20
2	TQQPQQPFPQQPQQPFPQ	0.9960	91 - 102	25	100
3	VPVPQLQPQNPSQQQPQEQVPL	0.9992	93 - 103	14	50
4	RPQQPYPQPQPQY	0.9916	85 - 103	3	20
5	QPQQPFPQTQQPQQPFPQ	0.9963	77 - 109	30	100
6	PQQSPF	0.9977	83 - 112	1	10
7	QPQQPLPQPQQPG	0.9996	90 - 111	12	50
		ave: 0.9971			

Table 4.2

Sensitivity and linearity

Peptide #	Sequence	%Accuracy in corn (0.06 pg/mg)	%Accuracy in corn (30 pg/mg)	%Accuracy in wheat (60 pg/mg)
1	LQPQNPSQQQPQEQVPL	96.2	92.8	99.4
2	TQQPQQPFPQQPQQPFPQ	85.6	95.8	94.1
3	VPVPQLQPQNPSQQQPQEQVPL	103.5	97.1	97.4
4	RPQQPYPQPQPQY	90.4	92.3	97.5
5	QPQQPFPQTQQPQQPFPQ	90.1	99.1	96.8
6	PQQSPF	68.7	108.3	102.2
7	QPQQPLPQPQQPF	95.6	97.8	96.1
	Ave.	90.0	97.6	97.6
	Rel. Std. Dev (%)	11.0	5.3	2.6

Table 4.3

Spike recovery

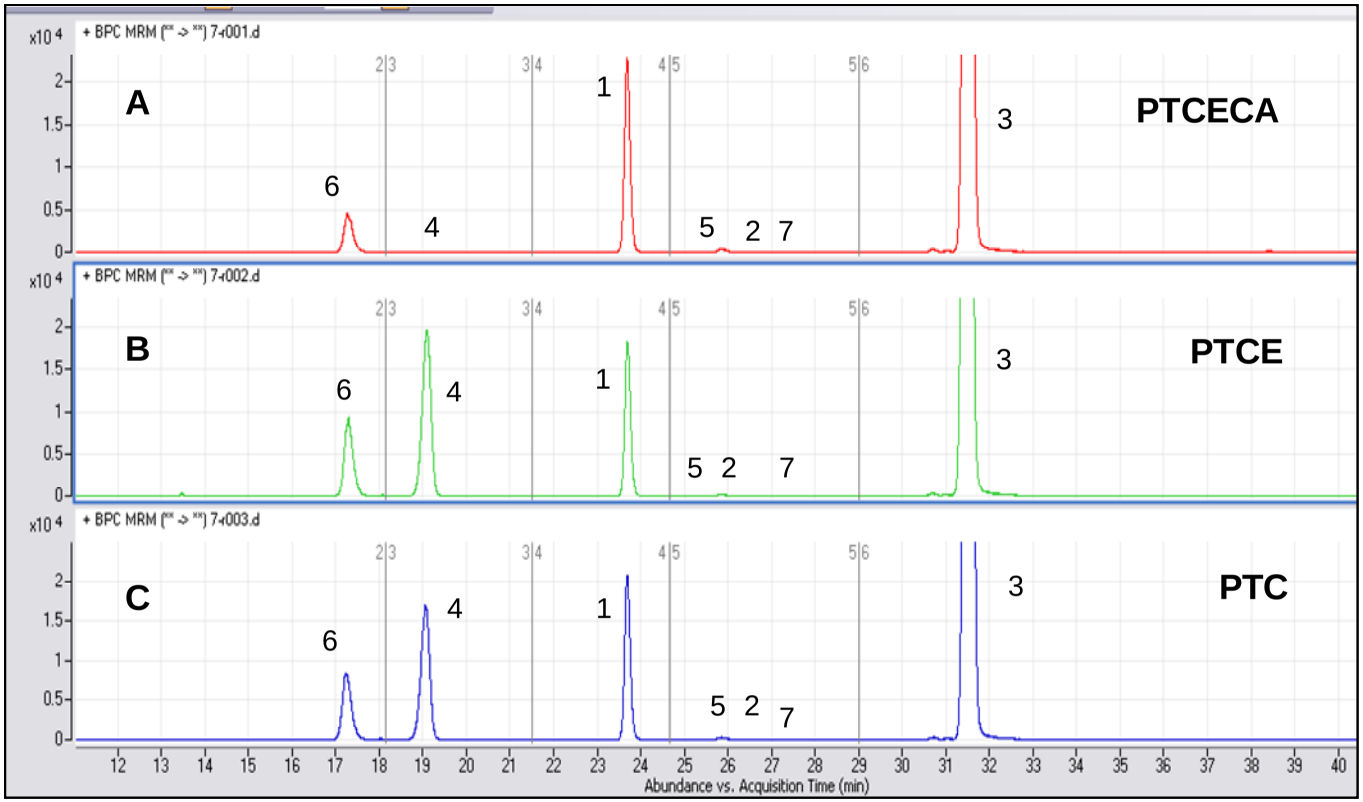


Figure 4.1

Evaluation of alternative cocktails of digestive enzymes: pepsin (P), trypsin (T), chymotrypsin (C), elastase (E) and carboxypeptidase A (CA)

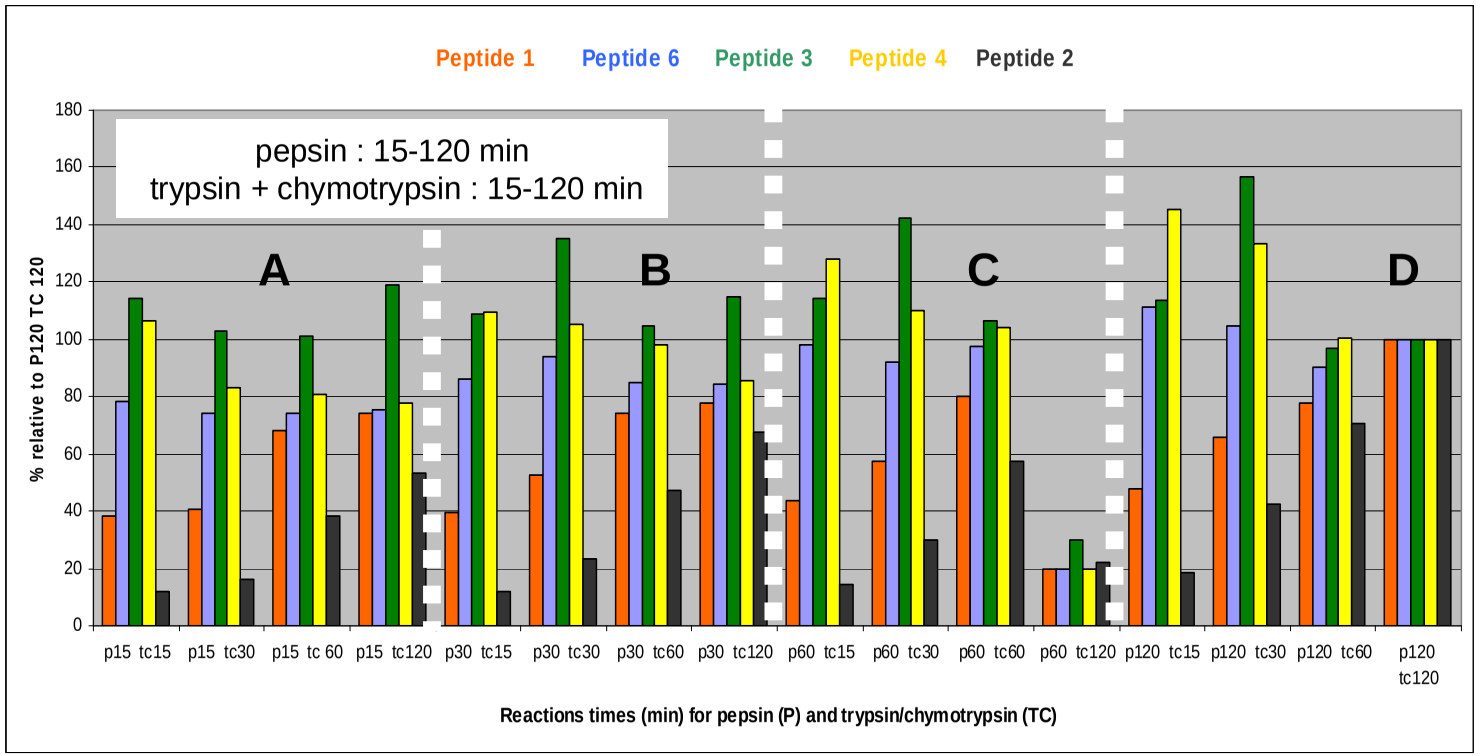


Figure 4.2

Optimization of proteolytic digestion (reaction) times

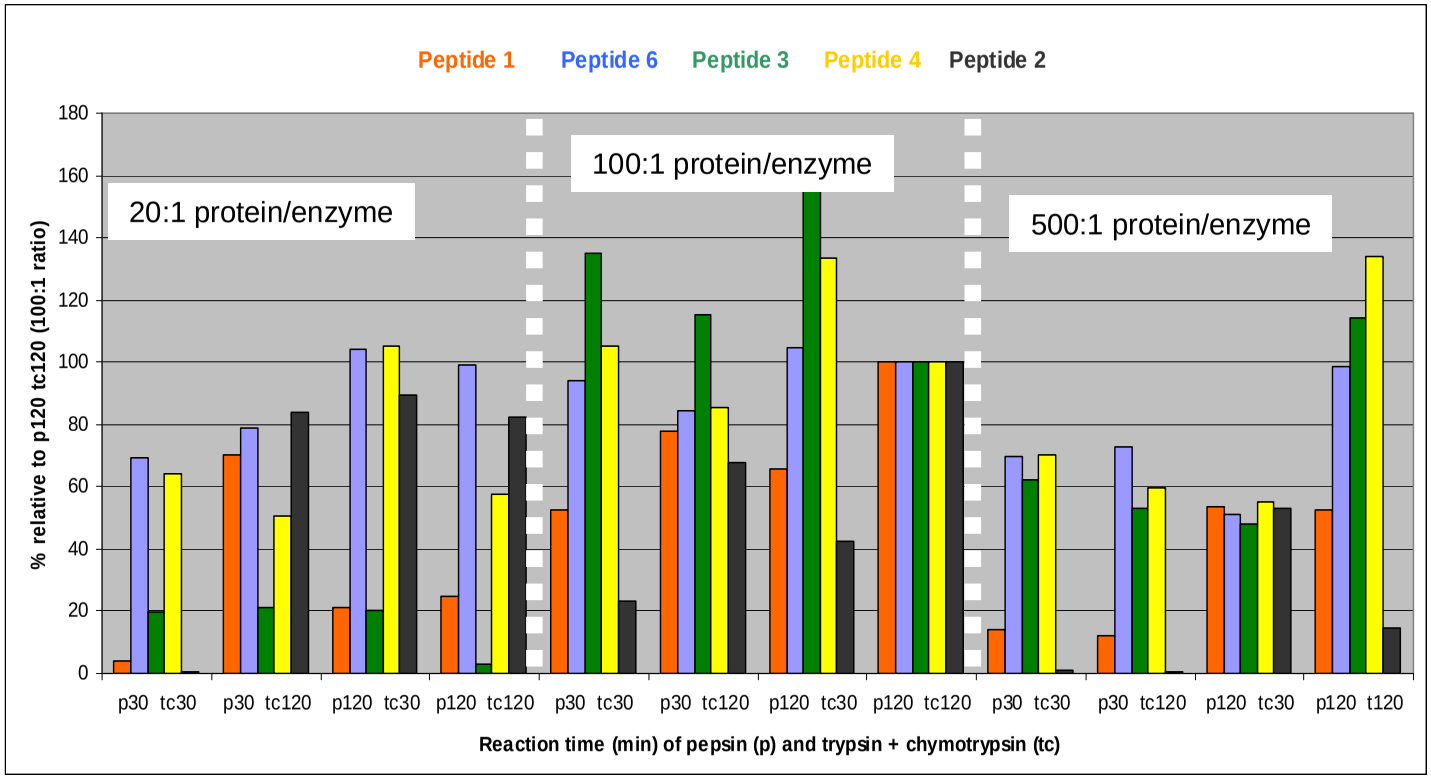


Figure 4.3

Optimization of the protein to enzyme ratio at different digestion times in the proteolysis of wheat with pepsin, trypsin and chymotrypsin

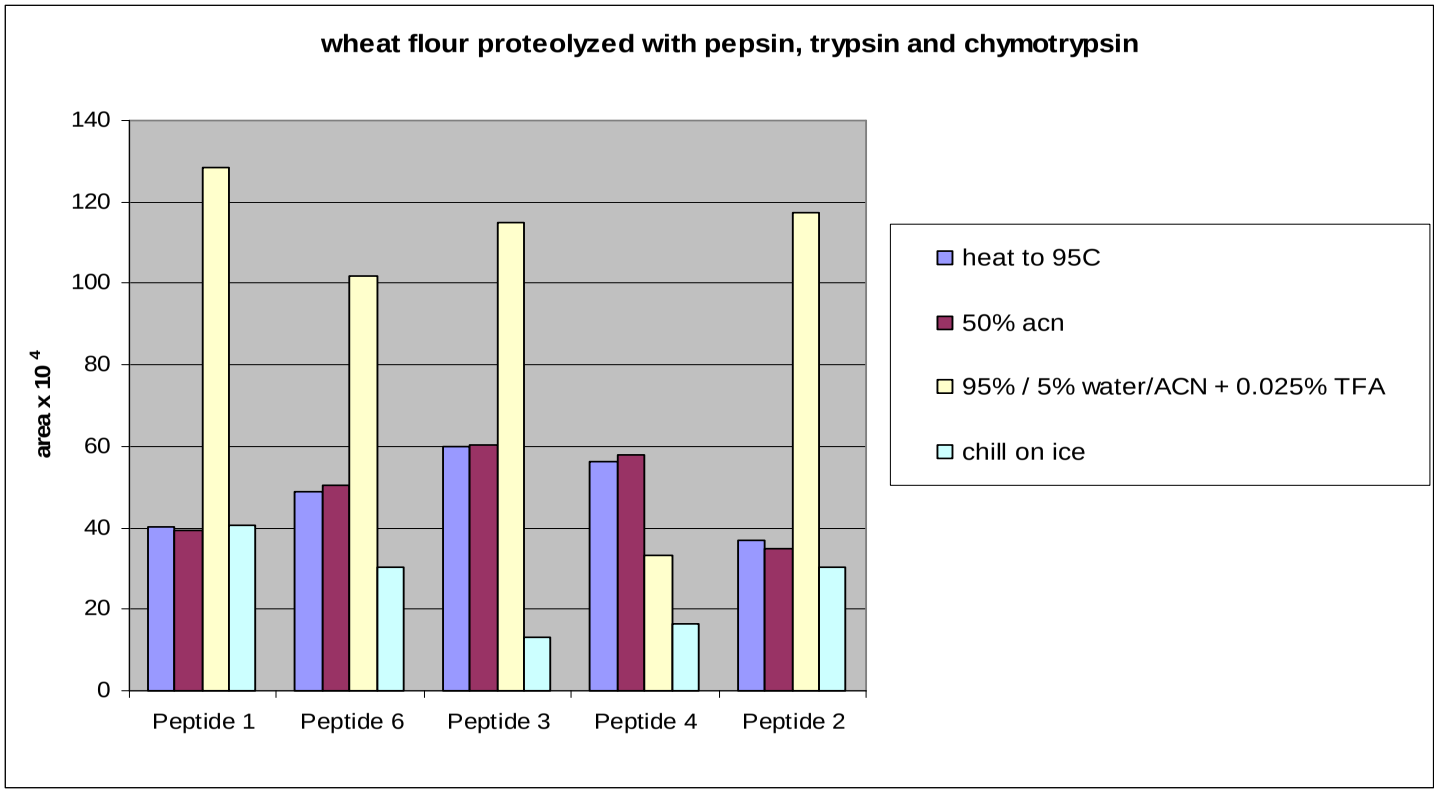


Figure 4.4

Evaluation of various procedures to quench proteolytic activity

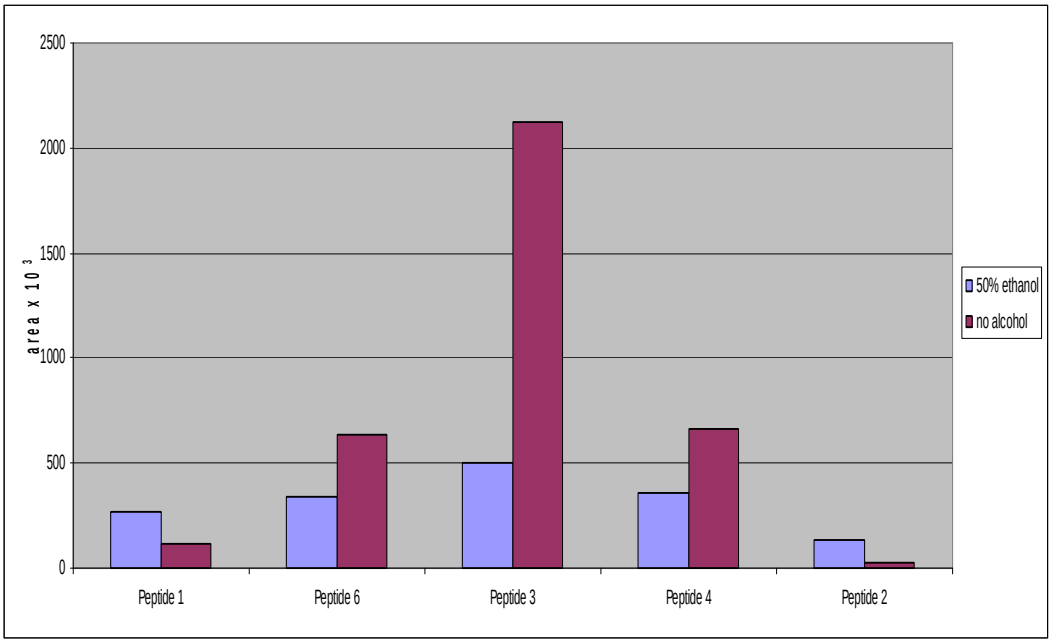


Figure 4.5

50% ethanol pre-extraction versus direct proteolysis of wheat gluten

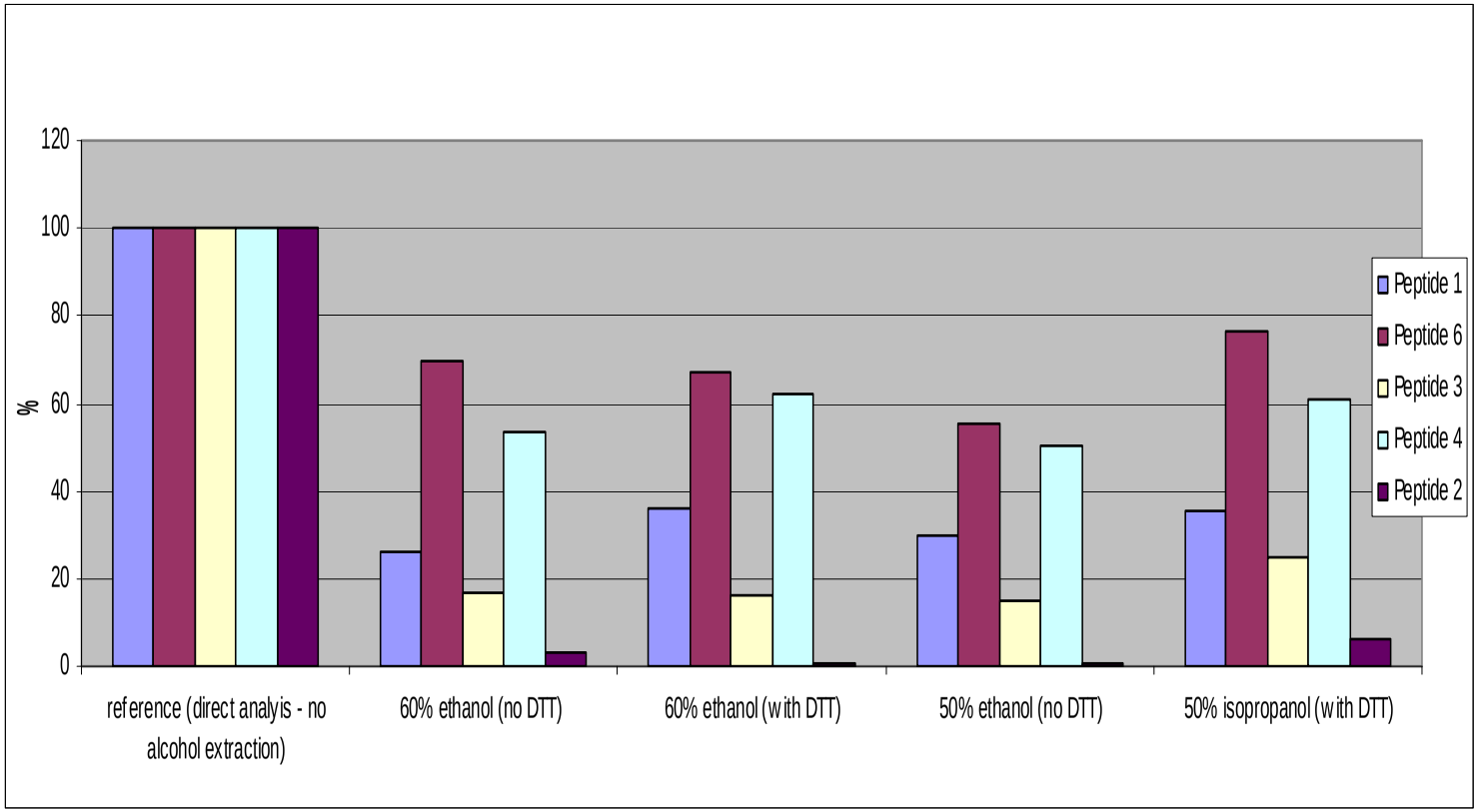


Figure 4.6

Aqueous alcohol pre-extraction versus direct proteolysis of wheat flour

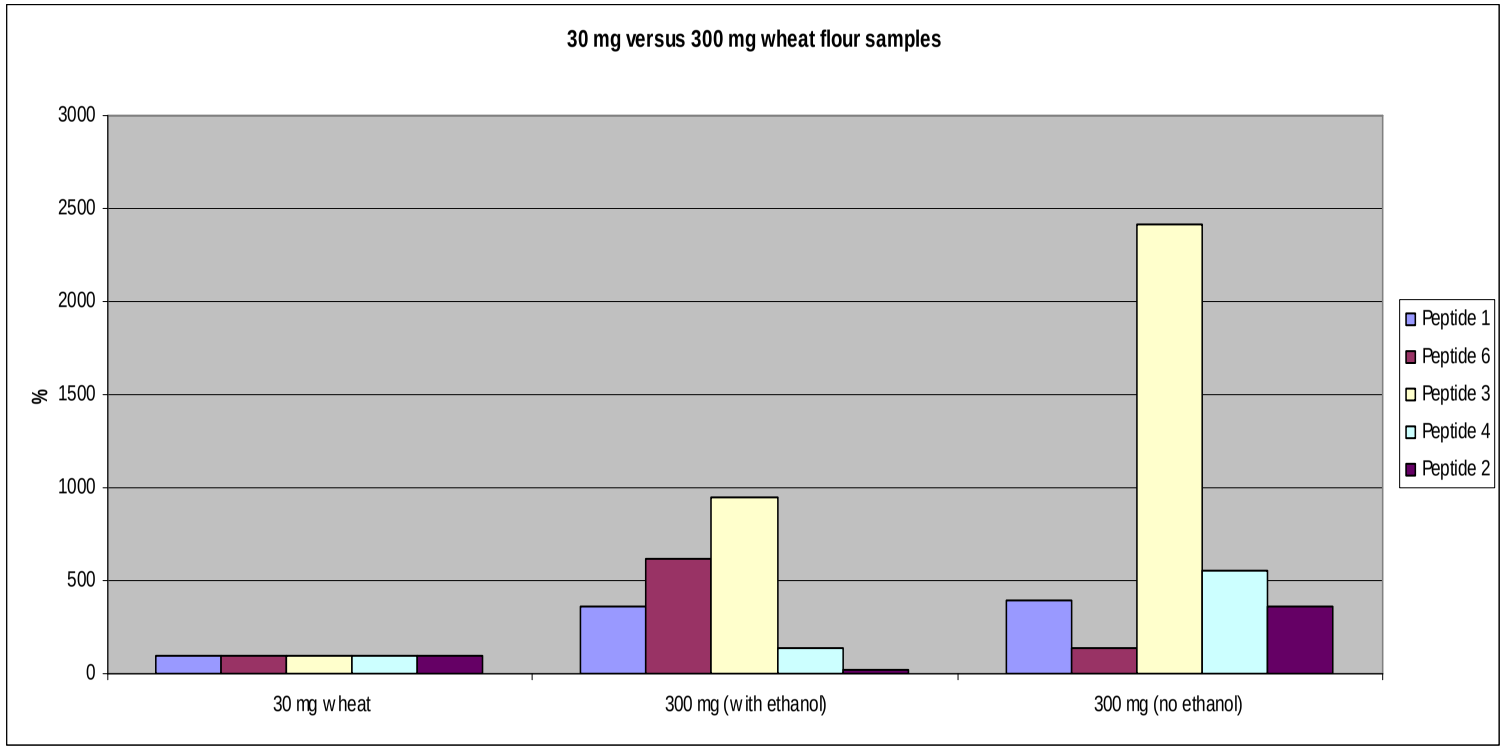


Figure 4.7

Evaluating detection limits by increasing the sampling size

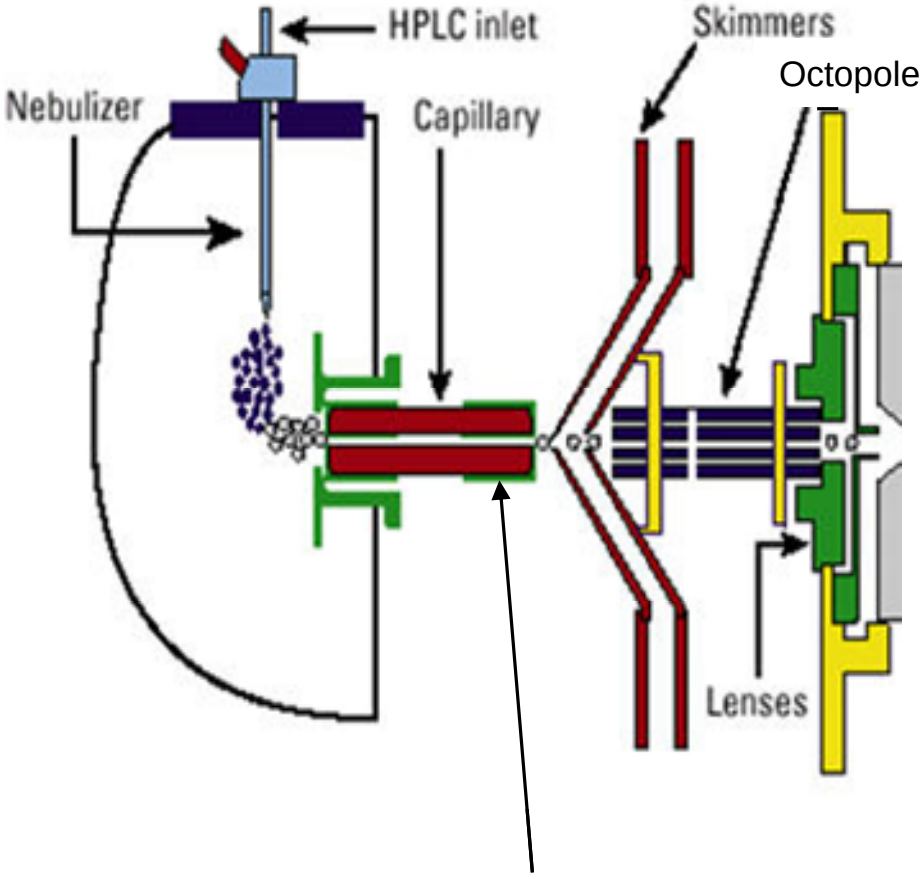


Figure 4.8

Ion transport region of the electrospray ionization interface

VPVPQLQPQNPSQQQPQEQVPL

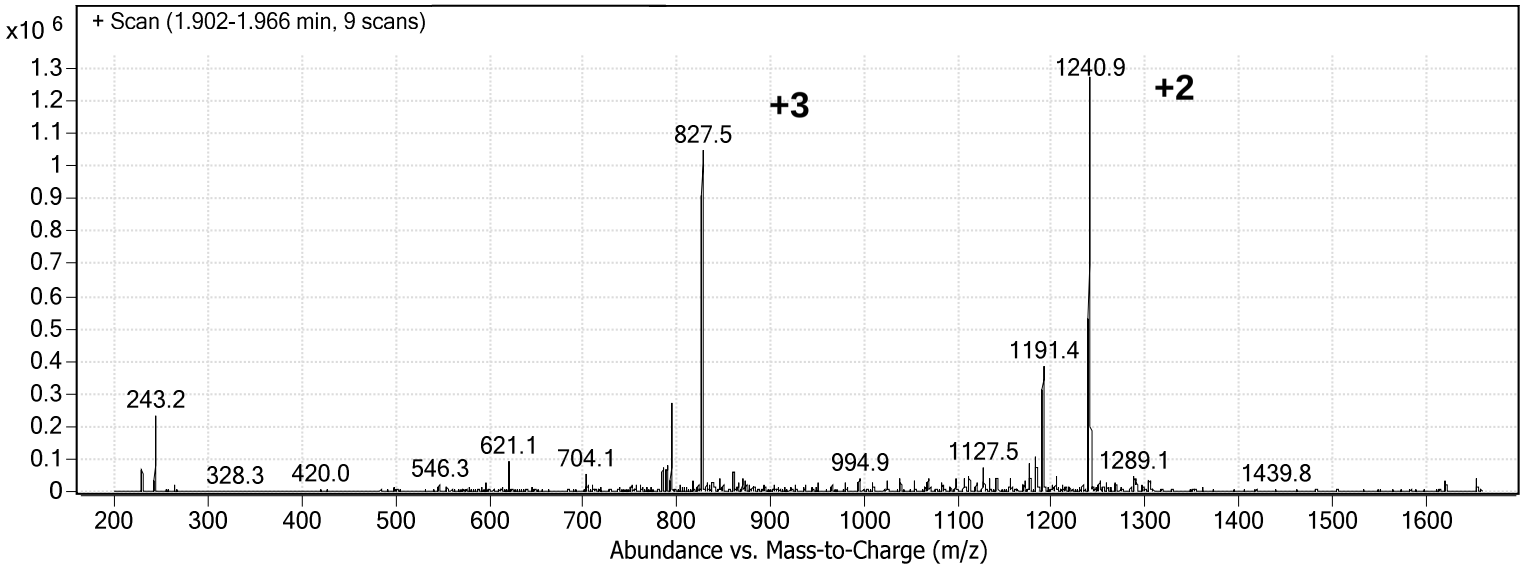


Figure 4.9

Full scan MS spectra of target peptide 3 at a fragmentor setting of 150V

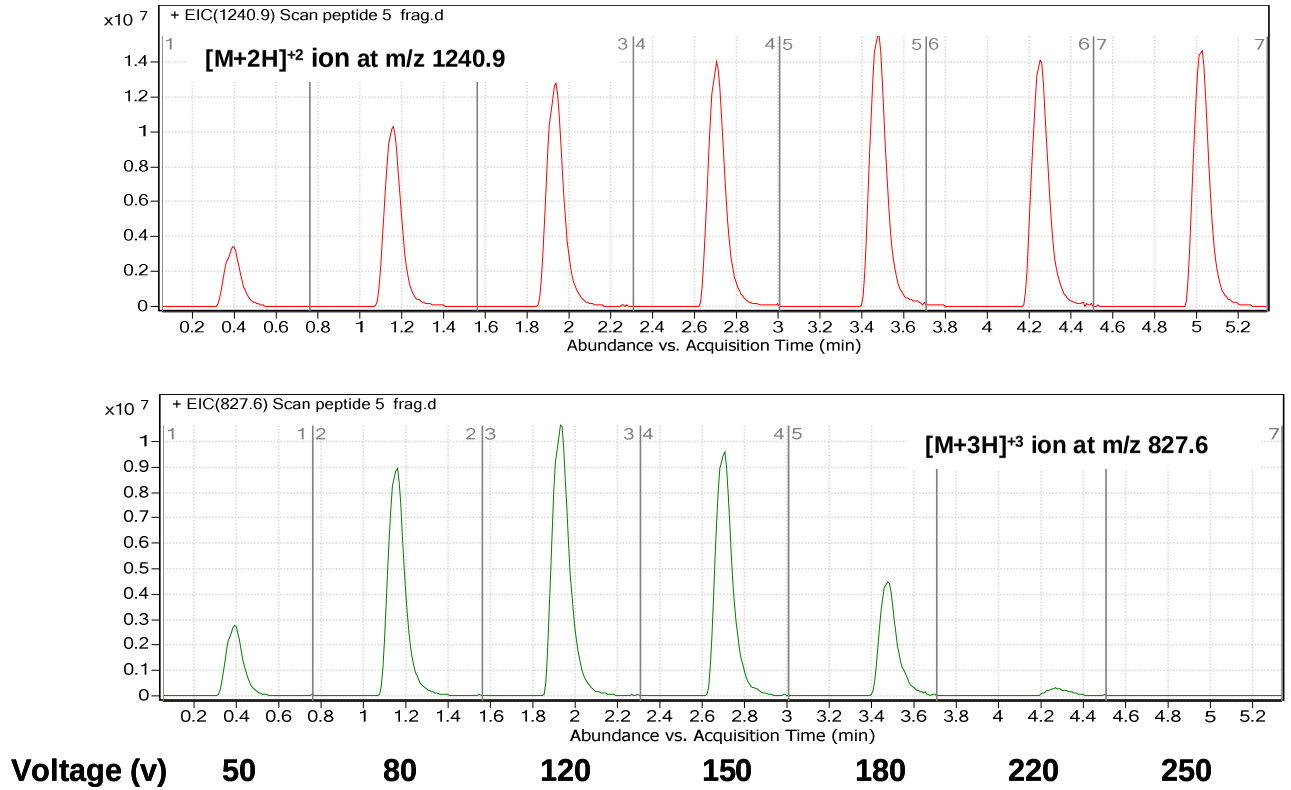


Figure 4.10

Fragmentor optimization of two precursor ions for peptide 3

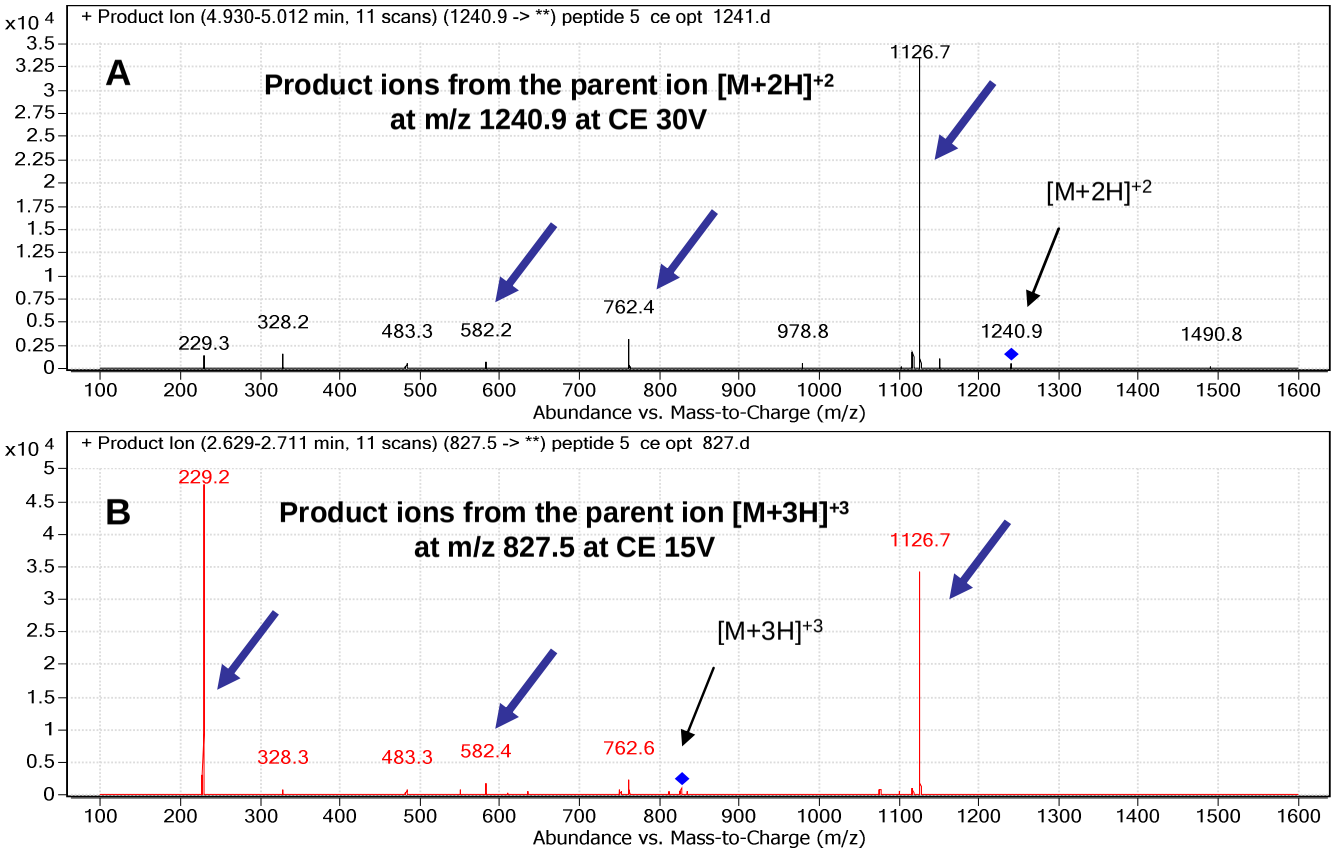


Figure 4.11

Collision energy optimization of the product ions from peptide 3

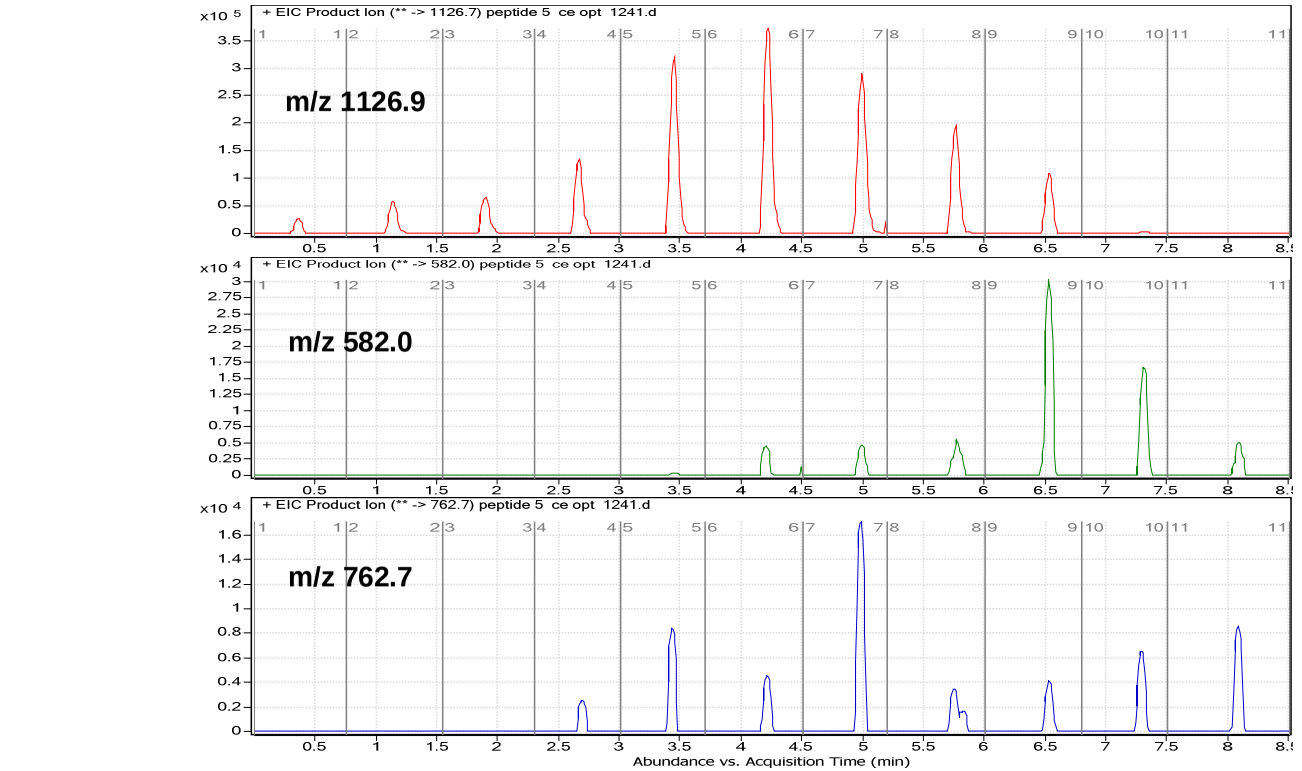
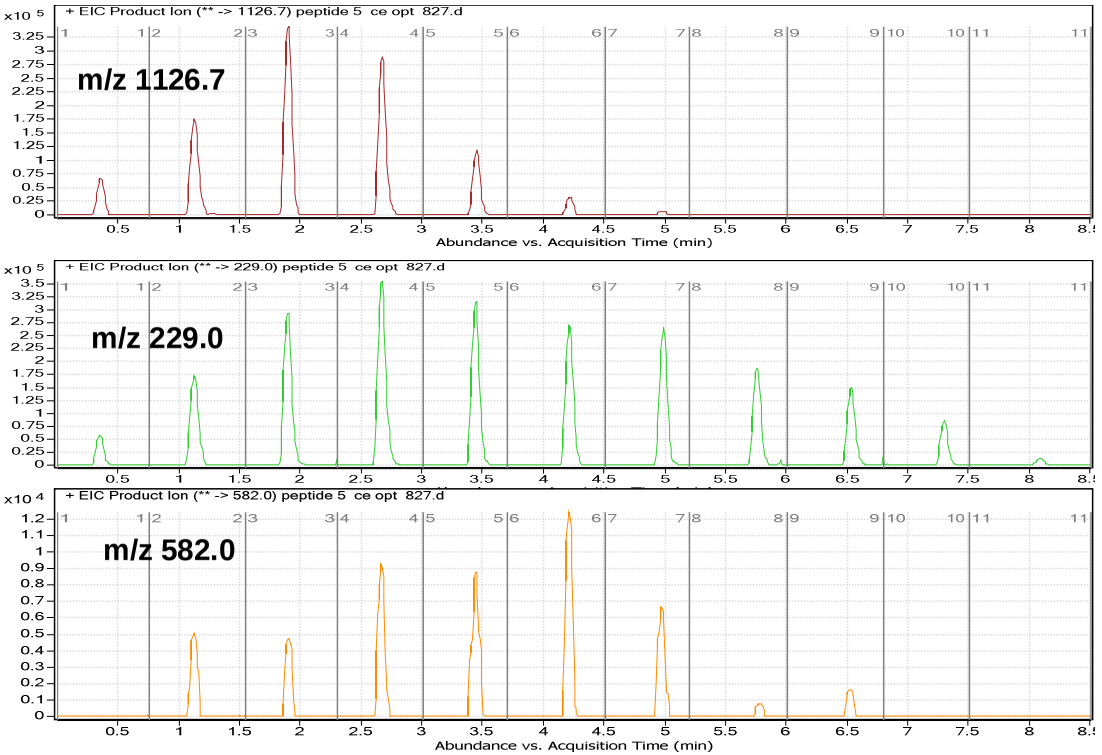


Figure 4.12

Collision energy optimization of 3 product ions from the precursor $[M+2H]^{+2}$ (m/z 1240.9) ion of peptide 3



Collision voltage: 0 5 10 15 20 25 30 35 40 50 60

Figure 4.13

Collision energy optimization of the 3 product ions from the precursor $[M+3H]^{+3}$ (m/z 827.5) ion of peptide 3

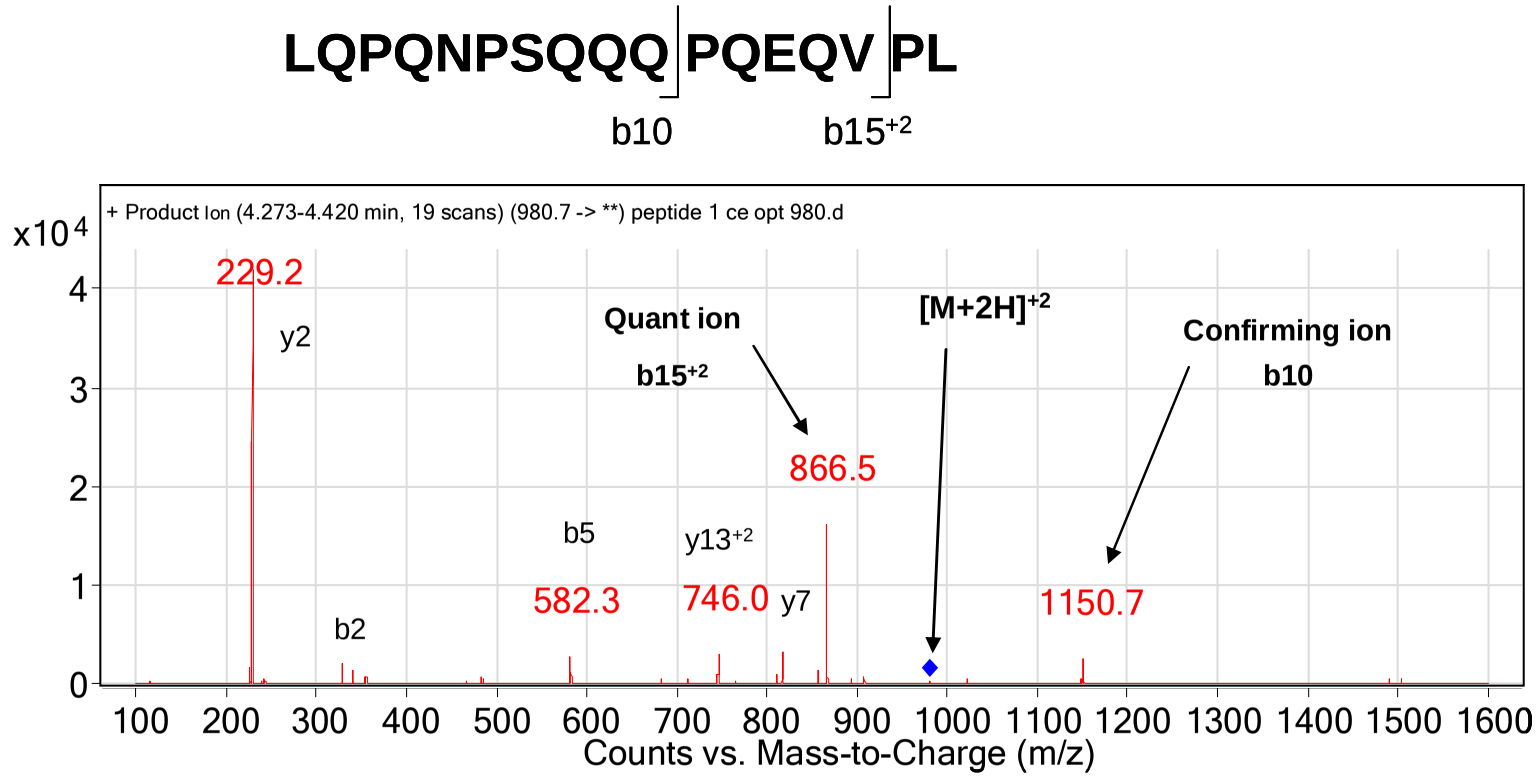


Figure 4.14

LC-MS/MS product ion mass spectrum generated from the $[M+2H]^{+2}$ precursor ion for peptide 1

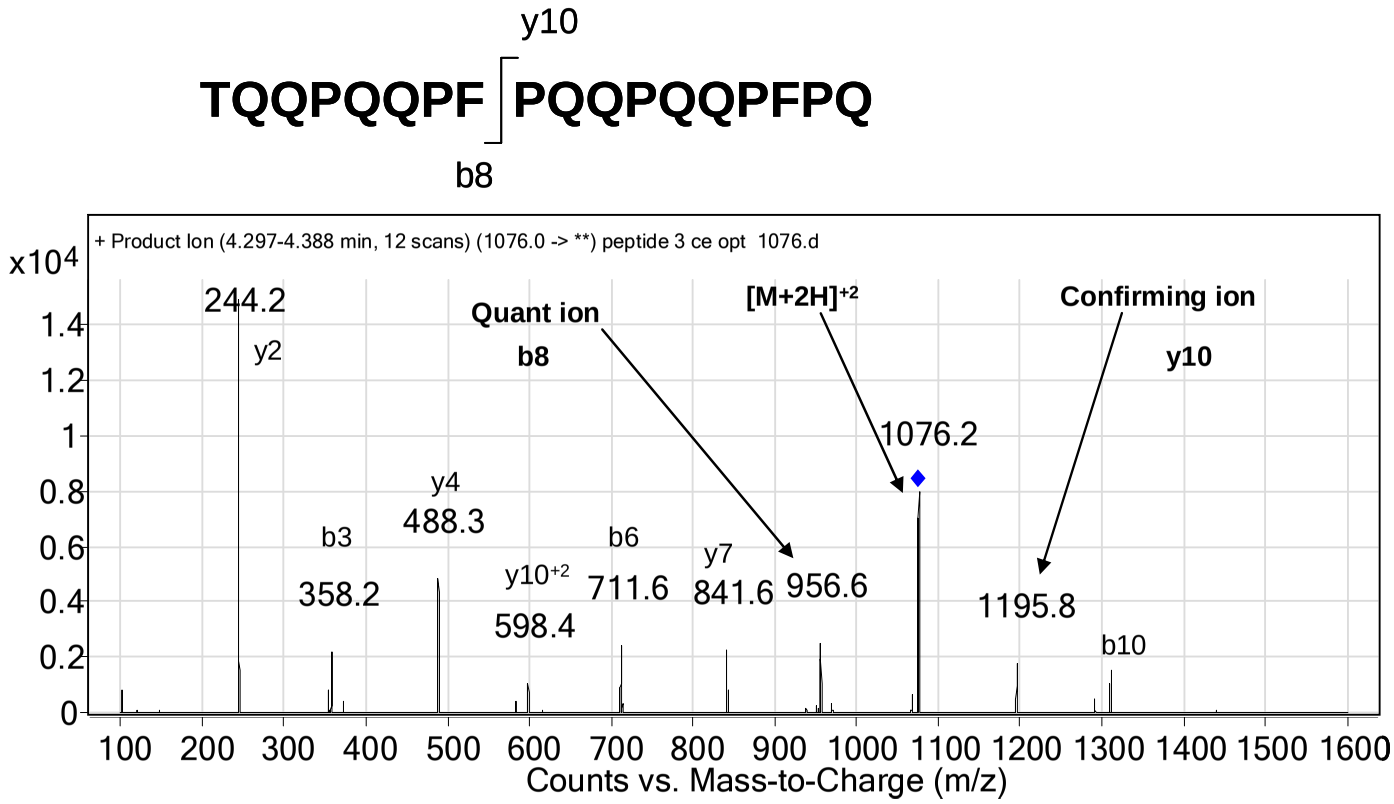


Figure 4.15

LC-MS/MS product ion mass spectrum generated from the $[M+2H]^{+2}$ precursor ion for peptide 2

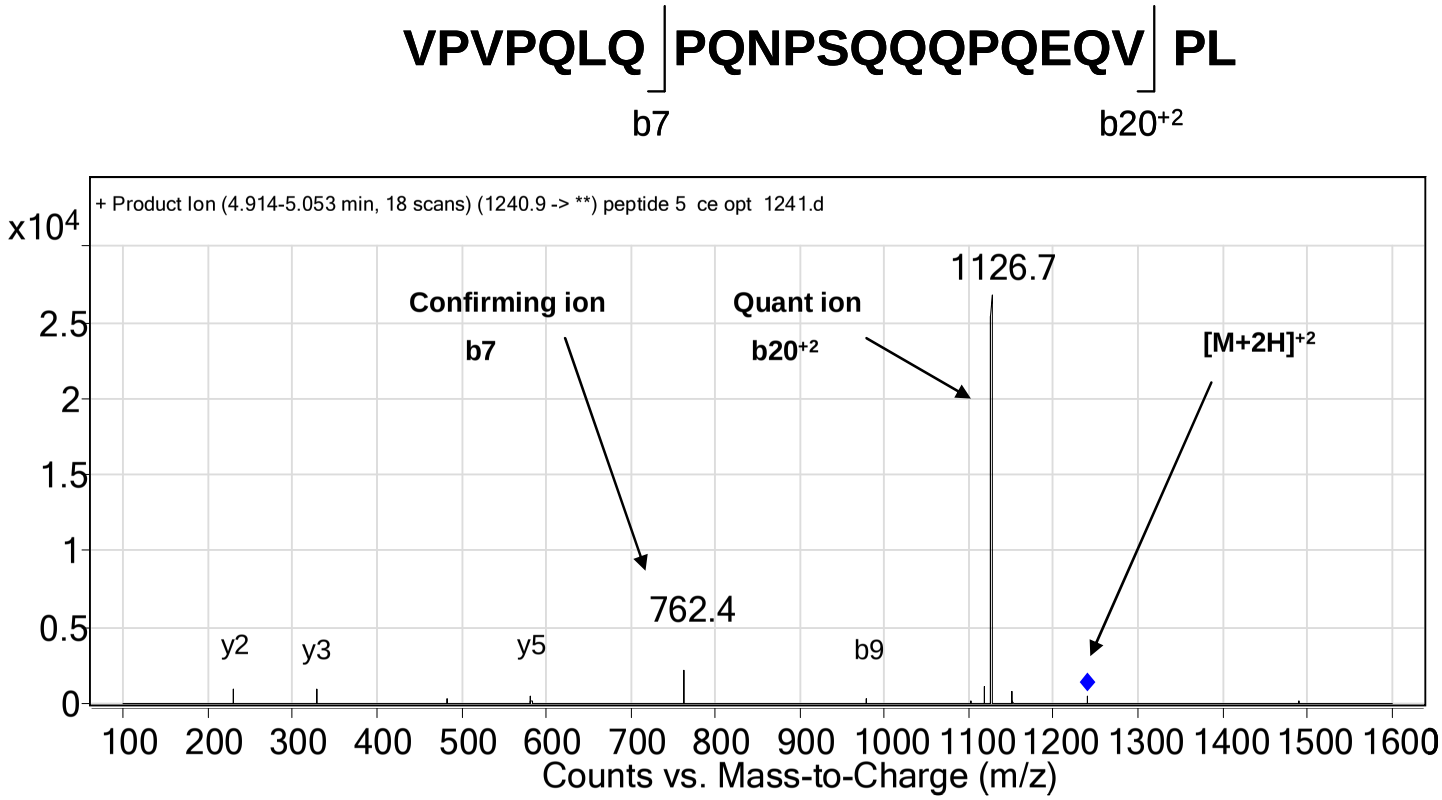


Figure 4.16

LC-MS/MS product ion mass spectrum generated from the $[M+2H]^{+2}$ precursor ion for peptide 3

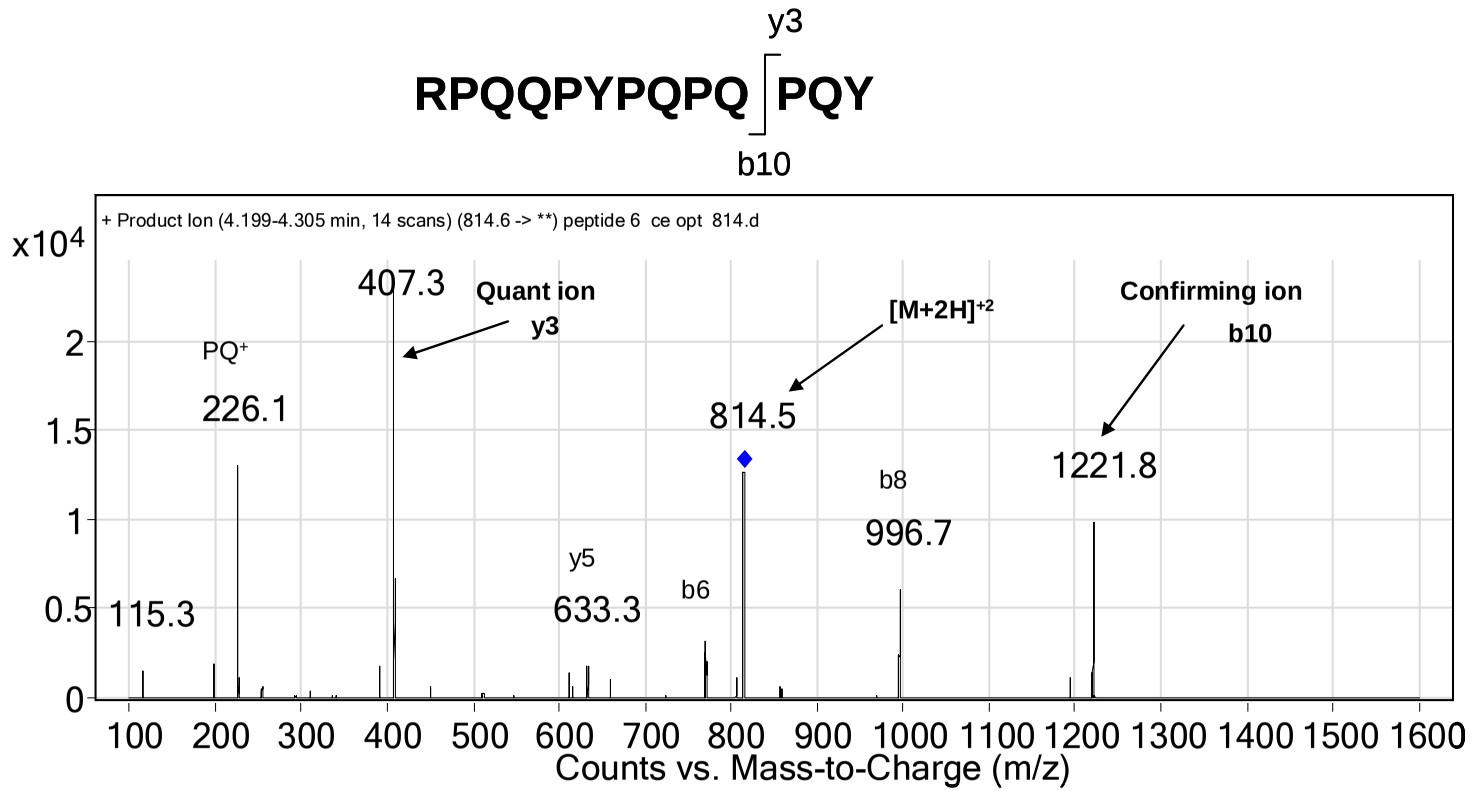


Figure 4.17

LC-MS/MS product ion mass spectrum generated from the $[M+2H]^{+2}$ precursor ion for peptide 4

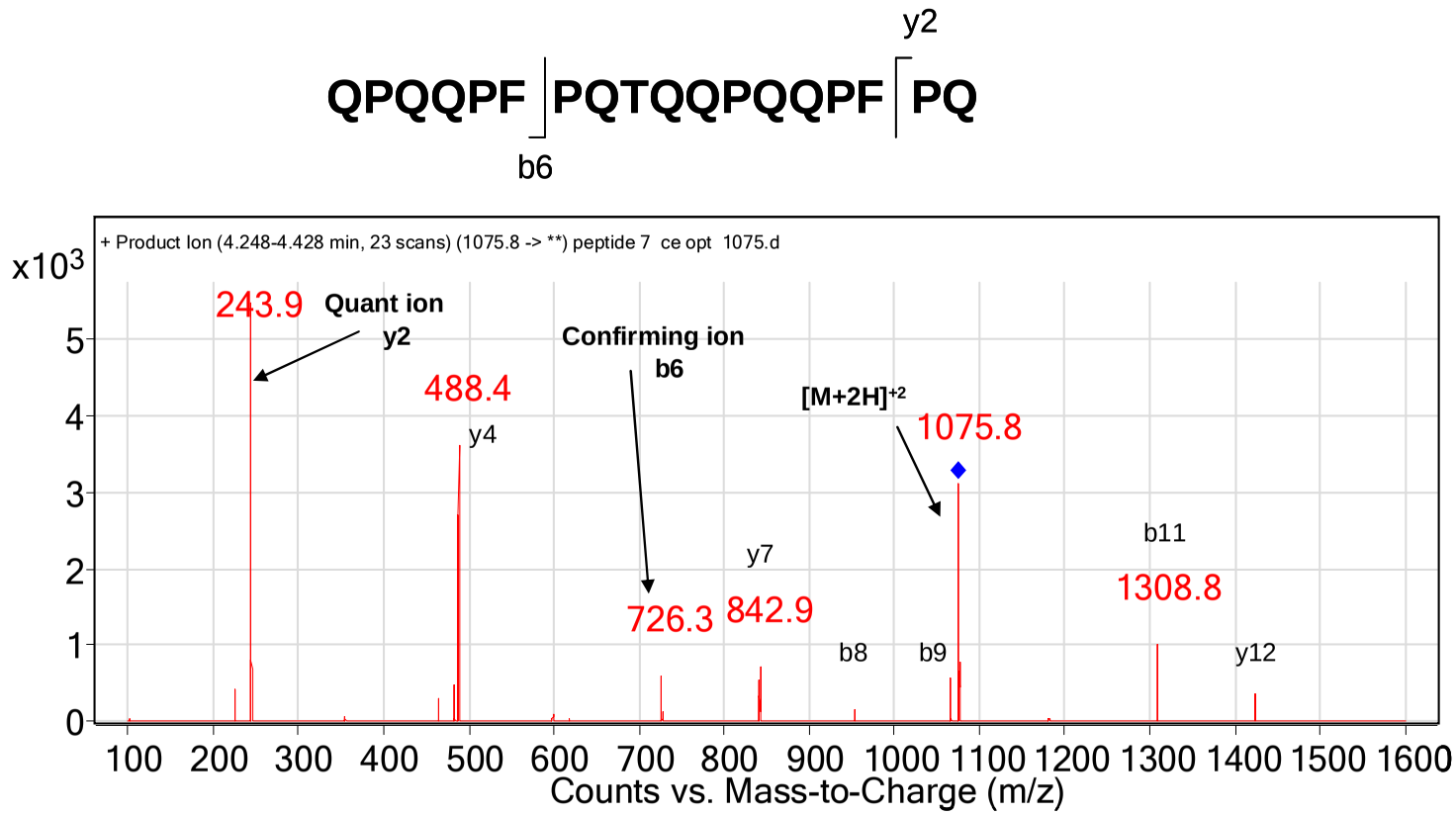


Figure 4.18

LC-MS/MS product ion mass spectrum generated from the $[M+2H]^{+2}$ precursor ion for peptide 5

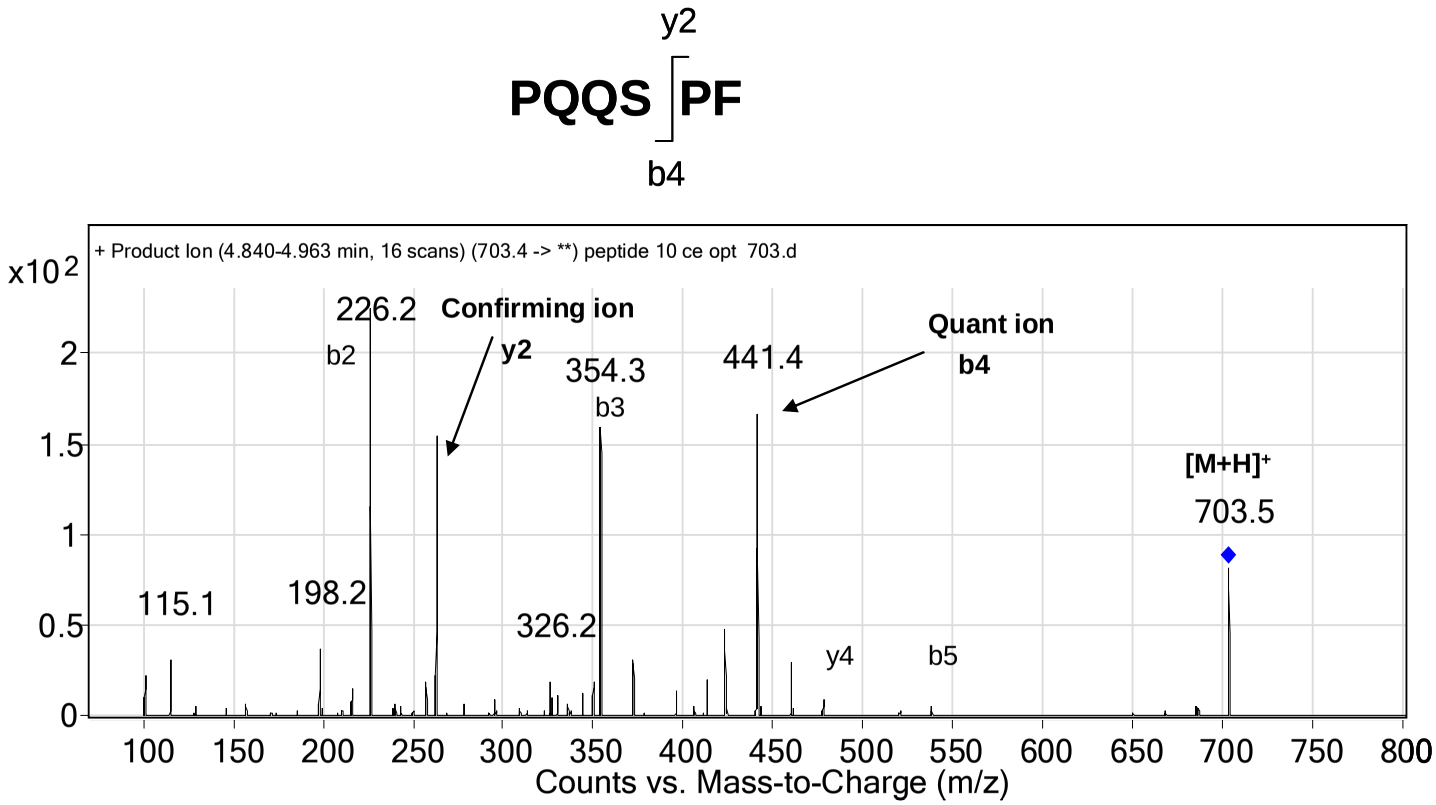


Figure 4.19

LC-MS/MS product ion mass spectrum generated from the $[M+H]^+$ precursor ion for peptide 6

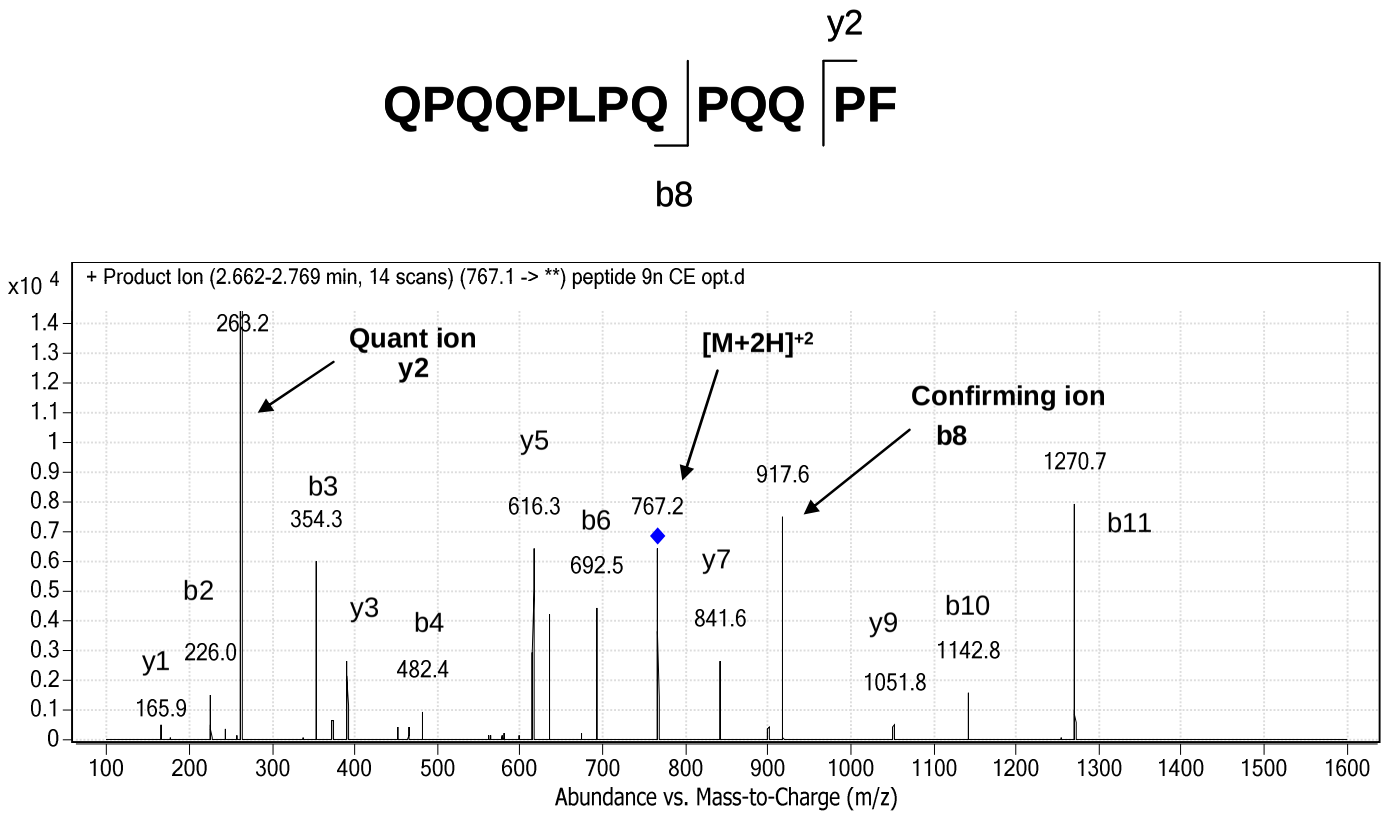
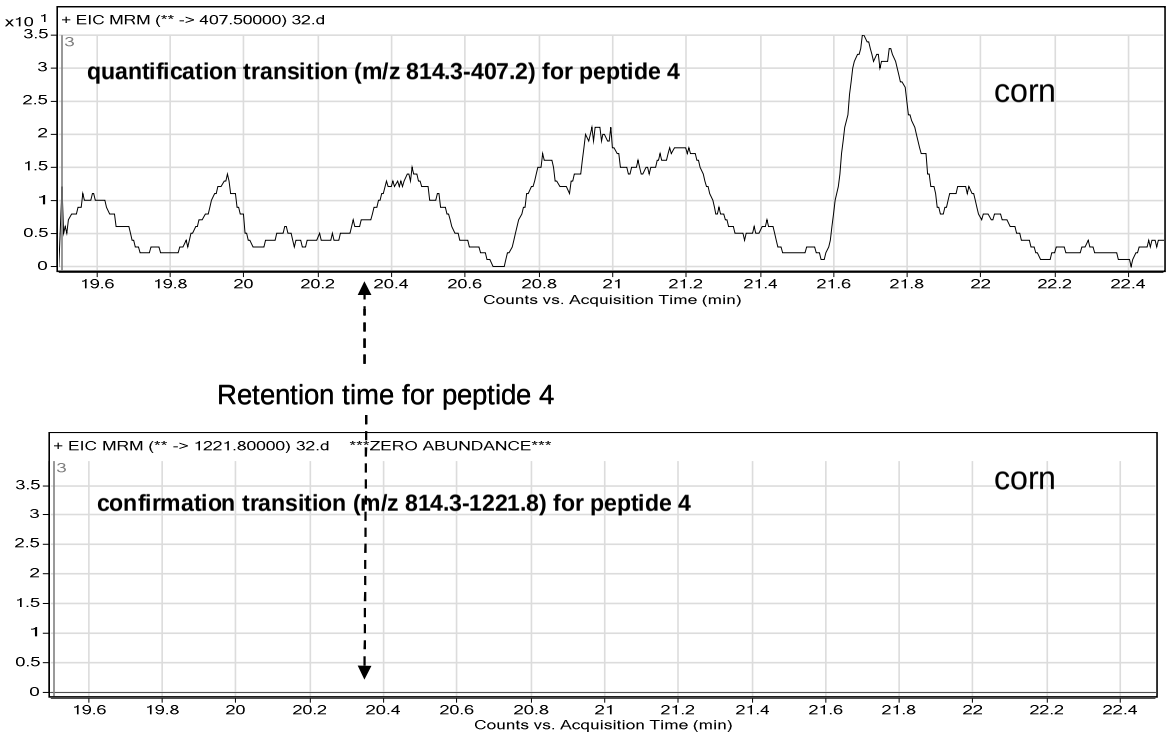


Figure 4.20

LC-MS/MS product ion mass spectrum generated from the $[M+2H]^{+2}$ precursor ion for peptide 7

MS/MS MRM EIC for a product ion below the m/z of the parent ion



MS/MS MRM EIC for a product ion above the m/z of the parent ion

Figure 4.21

Reducing chemical noise in HPLC-MS/MS analyses

LC-MS/MS EIC of a solvent standard cocktail of peptides 1-7 at 3 ng/mg

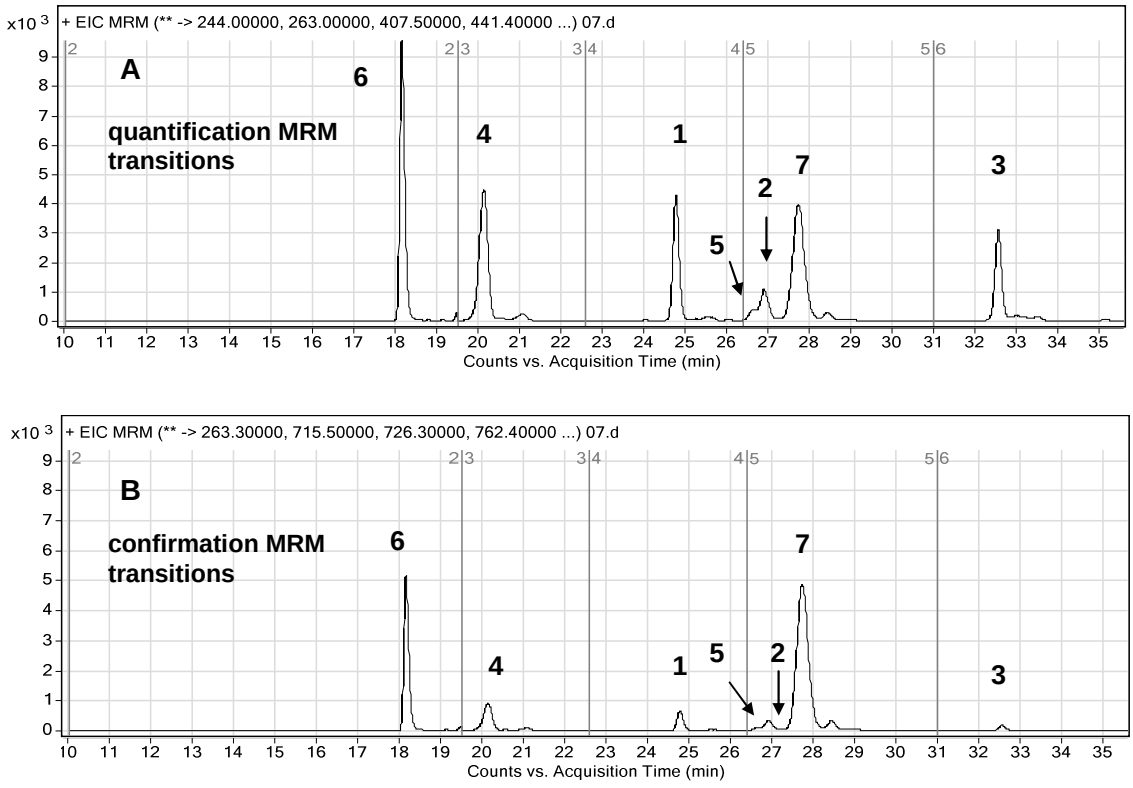


Figure 4.22

Time programmed multiple reaction monitoring acquisition profile

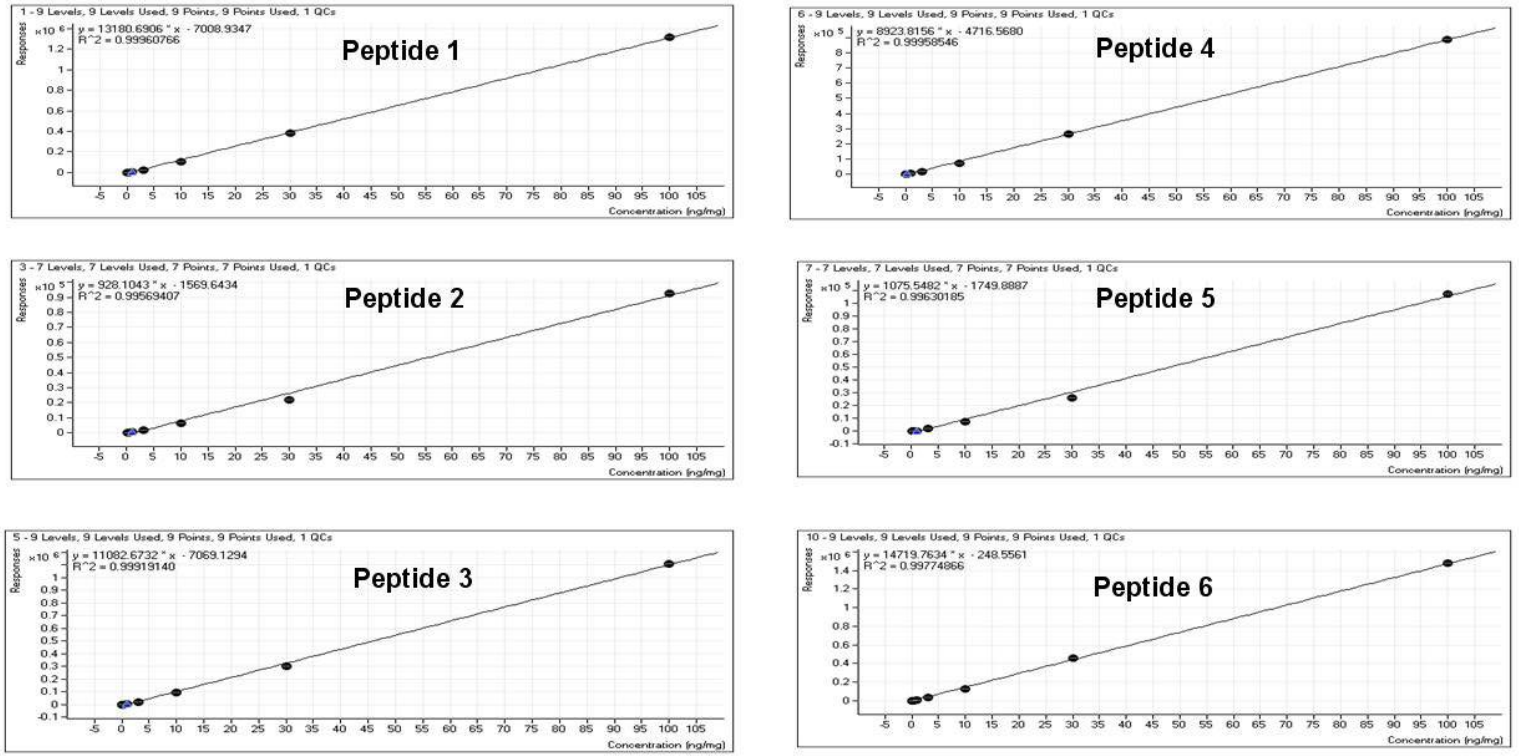


Figure 4.23

Calibration curves for peptides 1 to 6 over the range 0.1 to 100 ng/mg

LC-MS/MS EIC peptides 1-7 at 60 ppb

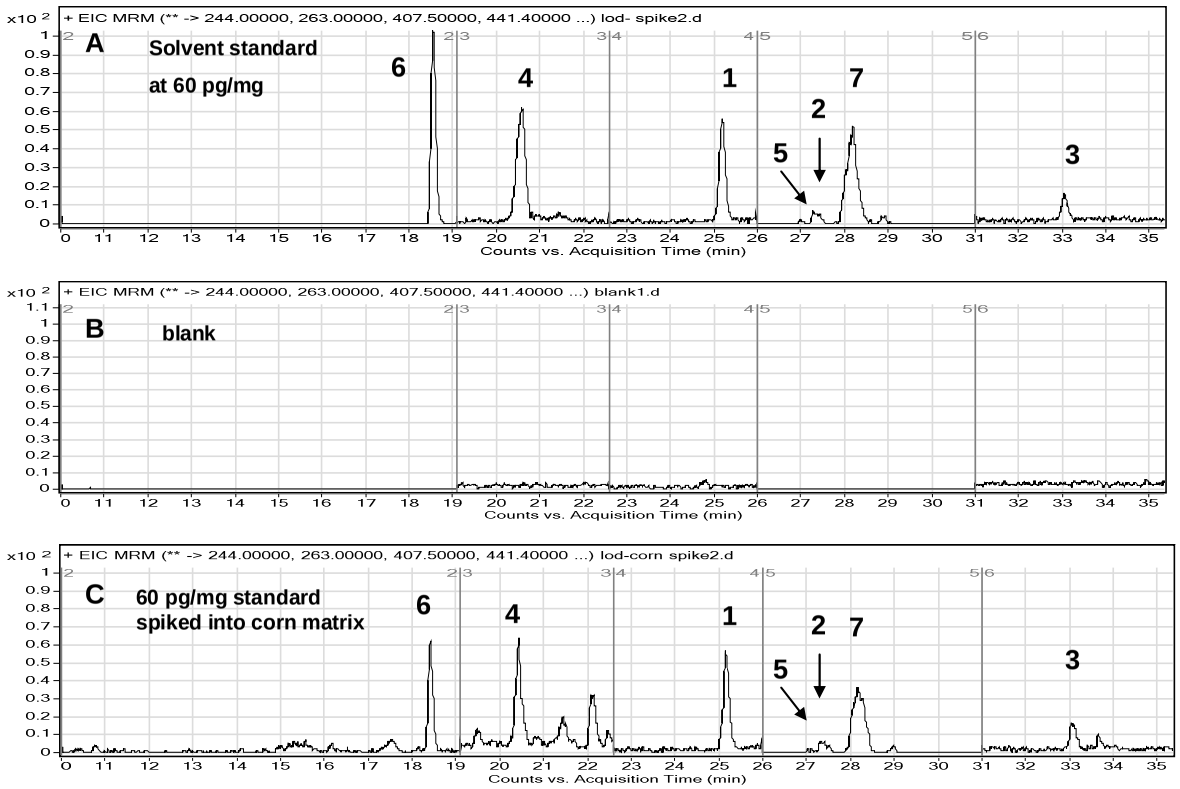


Figure 4.24

Spike recovery study of peptides 1-7 in corn

LC-MS/MS EIC of peptide 1 at a retention time of 25.4 minutes

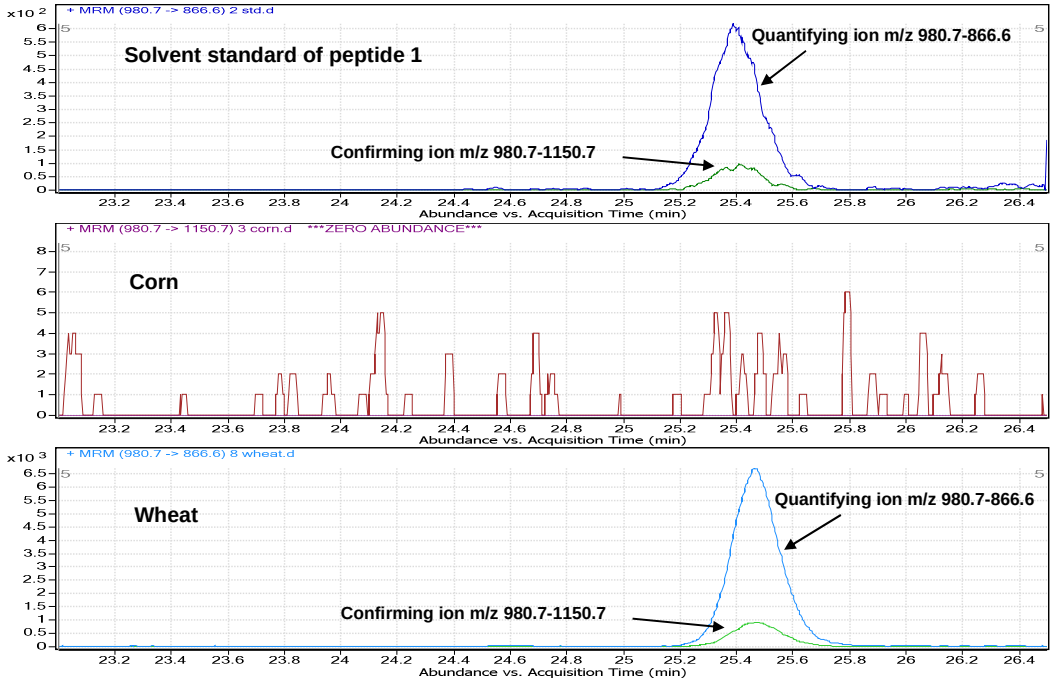


Figure 4.25

Confirmation of the presence of peptide 1 (LQPQNPSQQQPQEQVPL) in wheat

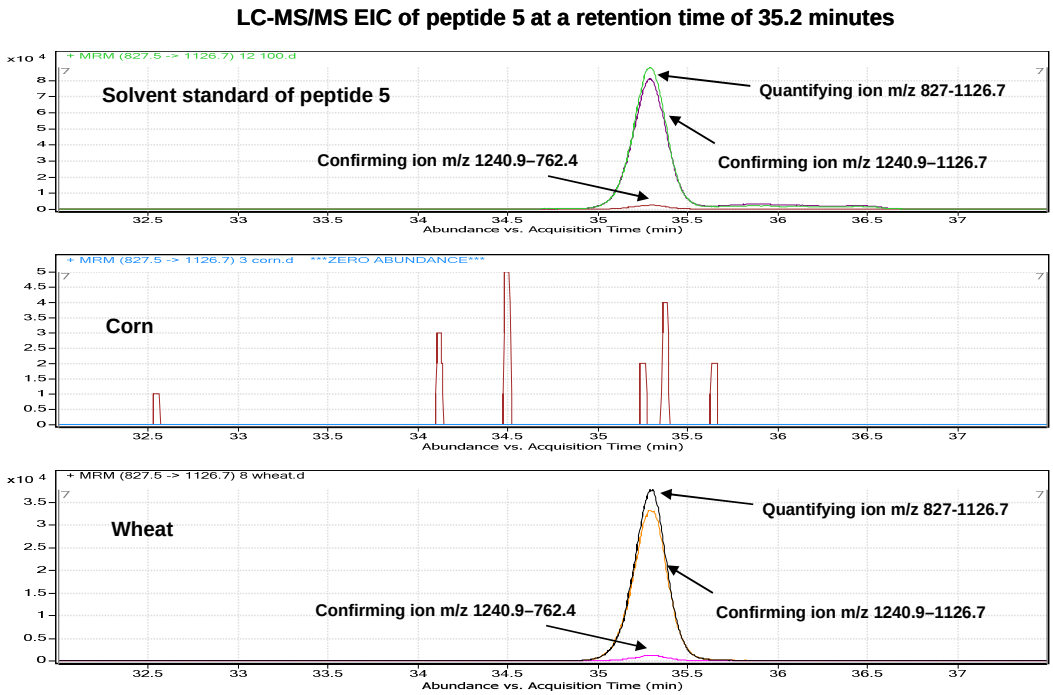


Figure 4.26

Confirmation of the presence of peptide 5 (VPVPQLQPQNPSQQQPQEQVPL) in wheat

LC-MS/MS EIC of peptides 1-7

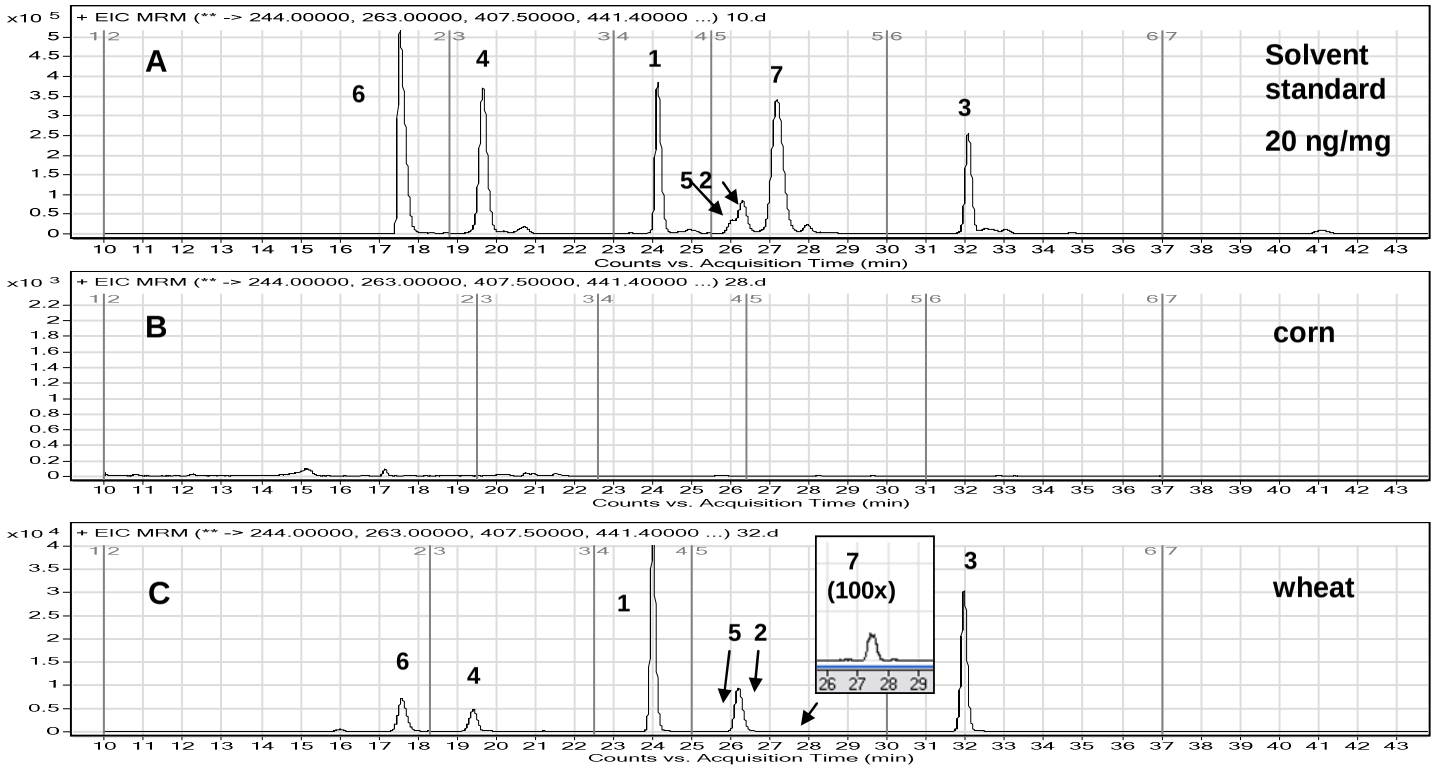


Figure 4.27

HPLC-MS/MS analysis of native proteolyzed wheat and corn flours

4.5 References

[1] AOAC (Association of Analytical Communities) Official Method 9991.19, AOAC International conference, Gaithersburg, MD **2005**.

[2] Joint Food and Agriculture Organization of the United Nations/World Health Organization Food Standards Program, Rome, Italy **2006**.

[3] FDA (US Food and Drug Administration, Center for Food Safety and Applied Nutrition), Food Allergen Labeling and Consumer Protection Act of **2004**; (http://www.cfsan.fda.gov/~lrd/fr07013.html).

[4] Skerritt, J.; Hill, A. *Journal of the Association of Analytical Communities International* **1991**, *74*, 257-264.

[5] Thompson, T.; Mendez, E. *Journal of the American Dietetic Association* **2008**, *108*, 1682-1687.

[6] Denery-Papini, S.; Nicolas, U.; Popineau, Y. *Journal of Cereal Science* **1999**, *30*, 121-131.

[7] Garcia, E.; Lorente, M.; Hernando, A.; Kieffer, R.; Wieser, V.; Mendez, E. *European Journal of Gastroenenterology & Hepatology* **2005**, *17*, 529.

[8] Moron, B.; Cebolla, A.; Manyani, H.; Slvarez-Maqueda, A.; Megias, M.; del Carmen, T.; Lopez, M.; Sousa, C. *American Journal of Clinical Nutrition* **2008**, *87*, 405-414.

[9] Mitea, C.; Looy-Winkelaar, Y.; van Veelan, P.; de Ru, A.; Drijfhout, J.; Koning, F.; Dekking, L. *American Journal of Clinical Nutrition* **2008**, *88*, 1057-1066.

[10] Fanciulli, G.; Azara, E.; Wood, T.; Dettori, A.; Delitala, G.; Marchetti, M. *Journal of Chromatography B* **2006**, *833*, 204-209.

[11] Pennington, C.; Dugresne, C.; Fanciulli, G.; Wood, T. *The Open Spectroscopy Journal* **2007**, *1*, 9-16.

[12] Mamone, G.; Ferranti, P.; Chianese, L.; Scafure, L.; Addeo, F. *Rapid Communications in Mass Spectrometry* **2000**, *14*, 897.

[13] Camafeita, E.; Alfonso, P.; Mothes, T.; Mendez, E. *Journal of Mass Spectrometry* **1997**, *32*, 940-947.

[14] Dworschak, R.; Ens, W.; Standing, K.; Preston, K.; Marchylo, B.; Nightingale, M.; Stevenson, S.; Hatcher, D. *Journal of Mass Spectrometry* **1998**, *33*, 429.

[15] Hernando, A.; Valdes, I.; Mendez, E. *Journal of Mass Spectrometry* **2003**, *38*, 862.

[16] Andersen, W.; Roybal, J.; Gonzales, S.; Turnipseed, S.; Pfenning, A.; Kuck, L. *Analytica Chimica. Acta* **2005**, *529*, 145-150.

[17] Bogusz, M.; Hassan, H.; Al-Enazi, E.; Ibrahim, Z.; Al-Tufail, M. *Journal of Chromatography B* **2004**, *807*, 343-356.

[18] Zhou, K.; Su, L.; Yu, L. *Journal of Agriculture and Food Chemistry* **2004**, *526*, 6114.

[19] Freitas, L.; Gotz, C.; Ruff, M.; Singer, H.; Muller, S. *Journal of Chromatography A* **2004**, *1028*, 277-286.

[20] Sealey-Voyksner, J.; Khosla, C.; Voyksner, R.; Jorgenson, J. *Journal of Chromatography A* **2009** (in press)

[21] van den Broeck, H.; America, A.; Smulders, M.; Bosch, D.; Hamer, R.; Gilissen, L.; van der Meer, I. *Journal of Chromatography B* **2009**, *877*, 975-982.

[22] Sleno, L.; Volmer, D. A.; *Journal of Mass Spectrometry* **2004**, *39*, 1091 – 1230.

[23] Stern, M.; Ciclitira, P.; van Eckert, R.; Feighery, C.; Janssen, F.; Mendez, E.; Mothes, T.; Troncone, R.; Weiser, H. *European Journal of Gastroenterology & Hepatology* **2001**, *13*, 741.

[24] Howdle, P. *European Journal of Gastroenterology & Hepatology* **2006**, *18*, 703.

[25] Shewry, P.; Tatham, A.; Kasarda, D. *Coeliac Disease*, Oxford: Blackwell Scientific Publications **1992**, 305.

[26] Wieser, H. *Balliere's Clinical Gastroenterolgy* **1995**, *9*, 191.

CHAPTER 5

Practical Applications of the HPLC-MS/MS Assay for the Quantitative Determination of the Presence of Immunogenic Gluten Peptides.

5.1 Introduction

5.1.1 Current guidelines for measuring gluten in food and consumer products

Currently there is no concise, universally accepted definition of “gluten-free” [1 - 3]. The United States Food and Drug Administration (FDA) proposed a definition for this term, which was posted in the Federal Register on January 3, 2007. The final form of the definition was expected to have been published in August, 2008; but it still remains in a draft format [4].

Conversely, the Codex Alimentarius Commission (a joint committee with delegates from both the Food and Agriculture Organization of the United Nations [FAO-UN] and the World Health Organization [WHO]) has revised its “Standard for Foods for Special Dietary Use for Persons Intolerant to Gluten”, in July 2008 [5]. This Codex definition is now considered to be the operational standard for the amount of gluten allowed in a commercially produced food product and as such, these products would be considered “gluten-free” for the purposes of international trade. In short, the guideline states that for a product to be labeled “gluten-free”, it

must not contain more than 20 ppm (parts per million) of wheat, barley or rye gluten. This correlates to about 1 mg of total gluten (prolamins and glutelins) in 50 g of food product.

From the perspective of a patient who suffers with any kind of gluten sensitivity, such as celiac disease, a “safe” level of exposure to gluten has not yet been established [6, 7]. Therefore, the only means to maintain a healthful existence is to carefully limit the exposure to gluten as effectively as possible. This requirement entails a detailed review of all product packaging and ingredient labels, in order to determine if any ingredient is present that may contain gluten, or if the product was manufactured in a facility that also manufactures products that contain wheat, barley or rye cereal grains. From other perspectives, such as those responsible for manufacturing processes, quality control and product labeling, there poses an obvious concern as to the means to ensure that products are analytically proven to be free of gluten, as sensitively and accurately as possible.

The work described in this thesis report has attempted to address this subject area by the development of an analytical methodology whose function is to determine the presence of immunogenic wheat gluten in a complex food matrix. Two practical examples of the application of this methodology are described in this chapter. (1) One practical application of the methodology involved its utilization in the evaluation of the quality of commonly available consumer foods and products. (2) A second application involved the assessment of the capabilities of a new therapeutic strategy under development, which attempts to address the challenges

faced from the continuous daily exposure to trace levels of gluten, in the management of celiac disease.

5.1.2 HPLC-MS/MS analysis of food and consumer products

The HPLC-MS/MS assay that has been developed and optimized is capable of detecting and quantifying select physiologically relevant wheat gluten peptides in complex food samples. Experimental results from the analysis of several varieties of commonly found native dietary grains have demonstrated that this method can successfully detect and quantify these peptides over the range 10 pg/mg -100 ng/mg of grain sample. The concluding stage of this work assessed the usefulness of this methodology in a practical application to evaluate the composition of common commercially available (native and processed) food and consumer products.

This HPLC-MS/MS method can be considered a sensitive and comprehensive alternative to the commonly used immunological antibody-based ELISA, which at best, is able to quantify gluten proteins down to the low ppm (part per million) level in specific applications [8]. Various ELISA assays have been successfully applied to the analysis of gluten, but do not have the capabilities to effectively determine if food products have inadvertently been contaminated with trace levels of gluten. Therefore, the ability to do this would give merit to this HPLC-MS method as a means of providing a much needed means of ensuring food safety.

5.1.3 Assessment of the effectiveness of an orally based combination enzyme therapy

Unless a facility is totally committed to manufacturing only gluten-free products, cross contamination can easily occur from improper handling of gluten-

containing ingredients or careless transporting or processing equipment maintenance procedures. Such instances could easily allow small amounts of gluten to erroneously become part of the finished product. Unfortunately, a strict gluten-free lifestyle can easily become disrupted by this type of lingering exposure to gluten, resulting in ongoing tissue damage and failure to completely heal.

One strategy to address the issue of how to surmount the effects of a continuous and unavoidable exposure to gluten involves the development of an oral enzyme that can detoxify small amounts of gluten in the gut before it has the chance to mediate the inflammatory response. This orally based therapy consists of a combination of food-grade proteases and has been under development in recent years by several researchers [9 - 12]. The design of this enzyme therapy is based on targeting the small proteolytically resistant proline and glutamine rich gluten peptides that seem to persist in the gut and are responsible for triggering the immune reaction in the body. The goal is to target these peptides and digest them into smaller fragments that can then go on to be completely proteolyzed by the other gastric and pancreatic proteases present in the gut. Although research toward the identification of T-cell stimulatory peptides from gluten is on-going, it may never be possible to completely characterize all physiologically relevant gluten peptides in every species of dietary grain. Therefore, an oral enzyme that has the ability to target and aid with the digestion of any such peptides that share a typical representative structure would greatly enhance the quality of life of anyone who suffers from T-cell mediated inflammatory reactions to gluten.

Two products that are currently commercially available consist of combinations of common food grade glutamine and proline specific endopeptidases. Both products contain a proprietary blend of dipeptidyl peptidase IV (DPPIV) and other proteases that are designed to work together with pepsin at the gastric stage of digestion. HPLC-MS/MS was used to evaluate the proteolytic capability of these two gluten detoxifying protease cocktails as compared with one general digestive enzyme cocktail, specifically with regard to their effect on degrading the seven immunogenic peptides identified in this research project.

5.2 Experimental

5.2.1 Materials and Reagents

5.2.1.1 Chemicals

Pepsin, trypsin, chymotrypsin, KHNaPO4 and NaOH were obtained from Sigma-Aldrich (St. Louis, MO, USA). HPLC-grade acetonitrile and HCl were obtained from Fischer Scientific. Water was obtained from an in-house Milli-Q water purification system (Millipore, Billerica, MA, USA).

5.2.1.2 Peptide standards

Synthetic standards of peptides (LQPQNPSQQQPQEQVPL, TQQPQQPFPQQPQQPFPQ, VPVPQLQPQNPSQQQPQEQVPL, RPQQPYPQPQPQY, QPQQPFPQTQQPQQPFPQ, QPQQPLPQPQQPF and PQQSPF) were synthesized (Thermo Electron, Ulm, Germany) and were analyzed using HPLC-MALDI-TOF mass spectrometry in order to determine their exact molecular weights and purity (>90% pure).

5.2.1.3 Calibration and quality control standards, procedure and analytical blanks

Calibration standards were prepared as various cocktails of the synthetically prepared target peptides [(1) LQPQNPSQQQPQEQVPL, (2) TQQPQQPFPQQPQQPFPQ, (3) VPVPQLQPQNPSQQQPQEQVPL, (4) RPQQPYPQPQPQY, (5) QPQQPFPQTQQPQQPFPQ, (6) PQQSPF and (7) QPQQPLPQPQQPF] as well as the αG-33 peptide (LQLQPFPQPQLPYPQPQLPYPQPQLPQPQPQPF). Standards were prepared using appropriate aliquots taken from stock peptide solutions, prepared in 80%/20% water/acetonitrile, over the range 10 pg/mg - 100 ng/mg.

Two types of quality control standards were prepared: (1) the first type was prepared by spiking the peptide stock peptide solutions into aliquots of corn flour that had already been enzymatically digested with pepsin, trypsin and chymotrypsin (per the procedure outlined in section 5.2.2.1), at concentration levels that matched the solvent calibration standards; (2) the second type was prepared by spiking the stock peptide solutions into aliquots of water that had also been enzymatically treated, at concentration levels that matched the solvent calibration standards.

Two types of procedure blanks were also prepared: (1) the first type used corn flour as a representative gluten-free matrix and was systematically enzymatically digested; (2) the second type used water as the matrix and was also enzymatically treated.

Analytical blanks consisted of aliquots of the HPLC mobile phase A (95% water / 5% acetonitrile (ACN) + 0.025% trifluoroacetic acid (TFA)).

5.2.1.4 Samples

Quinoa flour, stone ground whole grain corn flour, stone ground whole grain soy flour, vital wheat gluten flour, stone ground whole wheat flour (brand 1), whole wheat flour (brand 2), rye flour, barley flour, rice flour, oat flour, powdered ice tea mix, pasta, orzo, cheerios, hot sauce, bread, goldfish crackers, white vinegar, toothpaste, body lotion and body wash were all obtained at a local supermarket (Raleigh, NC, USA). Beer, gin, vodka, rum, red and white wine were obtained from a local specialty store (Durham, NC, USA). Gluten-free items (beer, rice, potato pasta, rice pasta, quinoa pasta, soy pasta, bread, pretzels, protein bar, quinoa cereal, rice seasoning mix, pad thai seasoning mix and crackers), as well as the two brands of glutenase enzymes (glutenase x and glutenase y) and one general digestive enzyme (GDE) product were obtained from a local specialty store (Chapel Hill, NC, USA).

5.2.2 Sample preparation

5.2.2.1 In-vitro proteolytic digestion procedure

Prior to analysis by HPLC-MS, samples were proteolyticallly digested *in-vitro*, using the optimized conditions determined in Chapter 4. This procedure was designed to release some or all of the seven target peptides, should gluten from wheat, be present in the sample.

In brief, 30 mg of homogenized sample was dissolved into 1 mL pepsin solution (0.01M HCl adjusted to pH 2), to establish a 1:100 pepsin to protein ratio; heat (at 38˚C) and shake for 2 hours. To each sample, add 50 µL of a 50 mM phosphate buffer and 35 µL of a 0.1 M NaOH solution, allowing the pH to rise above

6 and keeping it between 7 – 7.5. To each sample, add 25 µL of a 1:100 trypsin / chymotrypsin to protein solution (in 50 mM phosphate); heat (at 38˚C) and shake for 30 minutes. From each sample, take 200 µL of the supernatant and add 200 µL of Mobile phase A (95% water / 5% acetonitrile (ACN) + 0.025% trifluoroacetic acid (TFA)). Spin down all samples and aliquot 200 µL for HPLC-MS/MS analysis.

5.2.2.2 In-vitro oral enzyme cocktail digestion procedure

Prior to analysis by HPLC-MS, food samples were proteolytically digested *in-vitro*, using the following procedure.

30 mg of homogenized sample was dissolved into 1 mL pepsin solution (0.01M HCl adjusted to pH 2), to establish a 1:100 pepsin to protein ratio and a 1:10 solution of glutenase (x or y) to protein or digestive enzyme to protein solution (in 50 mM phosphate); heat (at 38˚C) and shake for 2 hours. To each sample, add 50 µL of a 50 mM phosphate buffer and 35 µL of a 0.1 M NaOH solution, allowing the pH to rise above 6, then keeping the pH between 7 – 7.5. To each sample, add 25 µL of a 1:100 trypsin / chymotrypsin to protein solution (in 50 mM phosphate); heat (at 38˚C) and shake for 30 minutes. From each sample, take 200 µL of the supernatant and add 200 µL of Mobile phase A (95% water / 5% acetonitrile (ACN) + 0.025% trifluoroacetic acid (TFA)). Spin down all samples and aliquot 200 µL for LC-MS analysis.

5.2.3 HPLC-ESI-MS Instrumentation and analytical conditions

HPLC-ESI-MS/MS was used to both detect and quantify each of the 7 target peptides. High performance liquid chromatography (HPLC) was performed using an

Agilent 1200 Rapid Resolution LC system (Agilent Technologies, Santa Clara, CA, USA). The HPLC was coupled to an Agilent 6410 QQQ mass spectrometer (Agilent Technologies, Santa Clara, CA, USA), operated in positive atmospheric pressure electrospray ionization (AP-ESI) mode. Mass calibration of the mass spectrometer was conducted using the Agilent tune compound, according to the manufacturer's documented procedures. Daily mass calibration checks were performed in order to assure that instrument response was accurate and consistent. Sample analyses were performed in multiple reaction monitoring mode (MRM) using time programming to obtain maximum possible sensitivity for each compound.

The atmospheric pressure ionization (API) source was operated at a capillary voltage of 3800 V (half moon electrode in the API chamber), with the nitrogen drying gas flow rate of 9.5 L/minute and temperature heated to 350˚C. The nebulizer was operated at 45 PSI. The fragmentation voltage (capillary exit voltage) was 150V, except for peptide 6, which was 120V. Each transition was monitored for 100ms, except for time segment 5, whereby each transition was monitored for 50ms. Table 5.1 lists the seven target peptide sequences, their respective time program segments, parent and product ion transitions and respective collision energies used for the analyses.

Typically, 20 µL of sample was injected into the LC-MS system. For ultra-trace level analyses, a sample volume of 100 µL was used. Sample aliquots were injected onto a C18 reversed-phase Ascentis Express (Sigma-Aldrich/Supelco) column of dimensions 2.1 mm i.d. X 150 mm packed with 2.7 µm superficially-porous silica particles. The gradient used for separation was 0-22% B over 35

minutes, then 22-60% B over 10 minutes at 25 ˚C, with a flow rate of 300 µL/minute. Mobile phase A was 95% water / 5% acetonitrile (ACN) + 0.025% trifluoroacetic acid (TFA). Mobile phase B was 5% water / 95% acetonitrile (ACN) + 0.025% trifluoroacetic acid (TFA).

5.3 Results and Discussion

5.3.1 HPLC-MS/MS analysis of food and consumer products

The HPLC-MS/MS assay, developed over the course of this research project, was used to both detect and quantify the seven immunogenic marker gluten peptides in a variety of commercially available food and consumer products. Products were manufactured in various countries around the world. Some products contained gluten, others were naturally gluten-free or labeled as “gluten-free”. The detection of any of the seven peptides was intended to represent a positive indication of the presence of wheat gluten in the product.

Following the completion of the *in-vivo* digestion procedure (section 5.2.2.1), HPLC-MS/MS analysis was performed on all sample extracts. Confirmation of the presence of the target peptides in the samples was based on the following criteria, adapted from existing FDA guidelines on qualitative target compound identification: (1) target compound retention time (within 2% of the representative calibration standard); (2) detection of all MRM product ions (all MRM parent to product transitions must be seen); (3) MS/MS parent to product ion ratios (within 30% of those in the representative calibration standard) and (4) absence of target peptides in all procedure blanks and analytical blanks.

Figures 5.1, 5.2, 5.3 and 5.4 show representative HPLC-MS/MS extracted ion chromatograms (EICs) of the target peptides that were detected in various food products. Tables 5.2(A) – 5.2(D) present the quantitative results from all analyses. Interestingly, no particular peptide profile was noted in any of the samples. All seven peptides were detected in native wheat and rye flours. Peptides 1, 3, 4, 6 and 7 were also detected in native barley flour. Oat and rice flour was shown to contain very low levels of two of the peptides. This could be a result of gluten contamination that occurred during the milling of the grains because most manufacturers handle many different types of grains. Oats has not been completely ruled out as a celiac active grain. As expected, all of the wheat containing processed foods contained various concentrations of all seven peptides. Interestingly, the powdered ice tea mix contained four of the seven peptides. None of the seven target peptides were detected in most of the gluten-free products, with the exception of one brand of gluten-free pasta and a pad thai seasoning mix. Potato vodka, rum and wine did not contain any of the peptides. As expected, beer was found to contain five of the seven peptides. Distilled products, such as gin are thought to be free of gluten, owing to the nature of the distillation process. However, this study found gin to contain a very low level of peptide 3. Vinegar and products that contain vinegar are expected to contain gluten, unless the vinegar has been distilled. The vinegar analyzed in this study was not distilled and did show the presence of four of the peptides. Hot sauce was found to contain a very low level of peptide 1. The label listed vinegar as an ingredient, but did not list the vinegar as “distilled”. The only consumer product that was found to contain any of the peptides was the body wash.

Several peptides were found to have concentrations that were above the upper limit of the calibration range (100 ng/mg). These samples were diluted and re-analyzed to ensure that their results were within the calibration range.

This novel analytical methodology is a combination enzymatic digestion-LC-MS procedure that was developed specifically to determine if wheat gluten is present, by releasing, detecting and quantifying any of seven immunogenic wheat gluten peptides. In the samples analyzed in these experiments, the exact variety(s) of wheat in each sample is unknown. Therefore, from these data it can be concluded that the presence of any of these peptides would render the product unsafe for use by people with gluten sensitivities. The data shown demonstrates that this assay can accurately detect and quantify immunogenic wheat gluten peptides, in a variety of food and consumer products, down to the low ppb (part per billion) level. In comparison, ELISA methods (currently endorsed by the FDA and Codex for the quantification of trace levels of wheat gluten), can at best and with limited applications, quantify down to the low ppm range [8, 13 and 14]. Results from a recent review of current literature in this area indicate that this LC-MS assay may possibly be the most comprehensive, sensitive and versatile assay currently available for the detection and quantification of trace amounts of wheat gluten. Given that the Codex standard for gluten content in a food product is 20 ppm (or 1 mg/50g food), this assay can confirm the presence of wheat gluten well below this, down to 0.5 µg/50g (corresponding to about 10 ppb) of food.

A unique ability of this novel methodology, allows for the detection of wheat peptides that originate from both monomeric gliadins and the polymeric glutenins,

because the sample is digested directly prior to LC-MS analysis. This type of proteoysis does not need any additional sample preparation step and occurs in a similar fashion as it would be *in-vivo.* This is a major advantage over any of the existing ELISA methodologies. ELISA methods use an alcohol pre-extraction step in order to separate the alcohol soluble monomeric prolamins [14, 15], then detect them using immunochemistry. Intact glutenins are polymeric proteins and do not dissolve in aqueous alcohol. Neither do hydrolyzed nor denatured gluten proteins that come about from heat processing. This is one of the limitations encountered with some ELISA methods and can lead to inaccurate quantitative results and incorrect calculated total gluten content in food [1, 8 and13].

The ability to detect any of the seven immunogenic wheat peptides provides an indication of the presence of wheat gluten (from gliadins and/or glutenins). Exposure to any product that contains unexpected or unlabeled gluten contributes to prolonged inflammatory tissue damage. The results presented here demonstrate that these seven target peptides can also be found in other grains (rye and barley), which are also implicated in the pathogenesis of T-cell inflammatory disease.

Figure 5.5 shows LC-MS/MS-extracted ion chromatograms of various gluten containing and gluten-free native grain flours, to illustrate the presence/absence of the target peptides in immunogenic and non-immunogenic grains. Traces shown include: (A) a solvent standard of the seven target peptides, (B) corn flour, (C) rice flour, (D) wheat flour and (E) wheat gluten. Database searches were also conducted using the National Center for Biotechnology Information (NCBI) protein database to confirm that each of the target peptide sequences were present in wheat proteins

and absent from corn and rice proteins. The presence of peptide 4 in rice may be due to contamination.

Figure 5.6 shows LC-MS/MS extracted ion chromatograms of peptide 7 in native wheat gluten as well as wheat, rye and barley flours. Interestingly, the response for peptide 7 is the smallest in wheat gluten and wheat flour, but the largest in barley and rye flours.

By examining Figure 5.7, it can be seen from the LC-MS/MS extracted ion chromatogram of a popular powered ice tea mix (manufactured in the USA), that trace levels of wheat gluten peptides 1, 3, 4 and 6 are present. The last peak seen in the trace was identified as the αG-33 (33mer) peptide. This 33mer is one of the most potent T-cell stimulatory peptides identified to date [16]. “Wheat gluten” (or gluten referenced by any number of alternate names) is absent from the ingredient list for this product. Fortunately, this is an example of how this HPLC-MS assay can detect instances of hidden or unlabeled gluten in a food product.

An example of a potentially contaminated food product is shown in Figure 5.8. Figure 5.8(A) – (C) are LC-MS/MS-extracted ion chromatograms of various pasta products: (A) wheat-containing pasta; (B) and (C) are two different brands of gluten-free pasta. The upper trace in Figure 5.8(B) shows the primary MRM parent to product ions, used for quantification. By examining the quantification ions that are present in the top trace, it appears that a trace level of peptide 4, along with a very small detected peak corresponding to peptide 3, can be seen in the gluten-free pasta (brand 2; manufactured in Italy). By looking at the lower part of figure 5.8(B), it can be seen that the confirming ion for peptide 4 is present, but is absent for peptide

3. Therefore, a positive identification of peptide 3 in this product cannot be made. This product is labeled “gluten-free”, however, its package also displays a disclaimer that the product was manufactured in a facility that also manufactures products containing wheat. The gluten-free pasta, shown in Figure 5.8(C) shows no hint of any wheat gluten contamination. As noted from this product’s package labeling, this product was manufactured in a facility dedicated only to gluten-free processing.

Exposure to the deleterious nature of gluten is not just limited to ingestion. Products such as lotions and toothpaste can also facilitate progression of disease. Figure 5.9 shows the profile of peptides detected in a dermatologist recommended daily moisturizing body wash. Peptides 1, 3, 4 and 6 can all be seen clearly, with a small contribution from peptides 2 and 5. Although hydrolyzed wheat protein is listed as an ingredient, this example is one where exposure to wheat may not be obvious.

These examples provide much needed evidence as to how important it is to ascertain the presence of trace levels of gluten in commercially available food and consumer products. From the consumer’s point of view, it is not always evident that gluten is present in products. Therefore, this LC-MS assay provides evidence that it could be potentially unhealthy for a person with celiac disease, or any other form of gluten sensitivity, to be exposed to some products even though they are labeled “gluten-free”. From a manufacturing perspective, the capabilities of this LC-MS assay provide an effective tool to ensure food safety and permit compliance with worldwide food labeling guidelines.

5.3.2 Comparison of the proteolytic abilities of glutenase and digestive enzymes by HPLC-MS/MS

The evaluation of the proteolytic abilities of two commercially available oral gluten-based protease products (glutenase x and glutenase y) and one general digestive enzyme (GDE) cocktail, was conducted with emphasis toward the digestion and detoxification of gluten peptides 1-7 present in various food samples.

Initial investigation evaluated the effectiveness of the proteolysis of the seven gluten peptides at three protein to enzyme ratios (10:1, 100:1 and 500:1). A wheat flour sample was digested following the protocol outlined section 5.2.2.2. The glutenase enzymes and general digestive enzymes were added at the same time as pepsin, in order to simulate gastric digestion. To simulate duodenal digestion, the pancreatic enzymes (trypsin and chymotrypsin) were added accordingly. Total digestion time involving the glutenases (and GDE) and pepsin was 2 hours, with an additional 30 minutes for digestion with trypsin + chymotrypsin. Data from these experiments determined that a protein to enzyme ration of 10 to 1 provided the best conditions for maximum possible detoxification of the seven peptides.

Three experiments were performed which evaluated three types of samples: (1) solvent standard (20 ng/mg) of peptides 1-7, (2) 100% wheat and (3) 1% wheat in corn. Samples (1) - (3) were all digested per the procedure detailed in above. The glutenases are designed to be the most effective where trace levels of gluten are present, which was theoretically represented by sample (3). Following the *in-vivo* digestion procedure (using a protein to enzyme ratio of 10:1), HPLC-MS/MS analysis was performed on all samples to evaluate which peptides resisted proteolysis.

Figures 5.10, 5.11 and 5.12 show representative LC-MS/MS EICs of the effect each enzyme cocktail had on peptides 1-7. Figure 5.10 displays results from the digested solvent standard containing peptides 1-7 as well as the αG-33 peptide. Figure 5.11 shows what happened when peptides 1-7, found naturally in a PTC digested native wheat flour sample, were proteolyzed and Figure 5.12 presents the remaining peptides 1-7 found in a sample of digested 1% wheat in corn, respectively. Table 5.3 summarizes the quantification results of the peptides that remain after these three digestion experiments. The values shown represent the percentage (+/- <15%) of each of the seven peptides that remain following treatment with pepsin, trypsin and chymotrypsin (controls for each experiment), glutenase x, general digestive enzyme (GDE) and glutenase y.

Several interesting observations can be made from these results. In general, the glutenases appeared to be more effective at degrading the seven peptides than the general digestive enzyme. With respect to the specific proteolytic abilities of glutenase x versus glutenase y, it appeared that brand y was more effective than brand x, under these experimental conditions. Looking more closely at the results of each of the experiments, the more concentrated levels of the seven peptides, as found in the solvent standard, were proteolyzed by the GDE by approximately 40 to 70%, with the exception of peptides 4 and 6, where <10% remain. The concentration of peptides found in the 100% wheat and 1% wheat samples were significantly lower than in the solvent standard and both showed almost complete digestion of peptides 4 and 6 by the GDE, while the other peptides remained pretty much intact. In all three cases, very small amounts of peptide remained following treatment with the

glutenases. The peptides that were released from the digestion of the 1% wheat in corn sample were digested the most. Therefore, it appears that the glutenase products do what they are designed to do, which is to effectively digest trace levels of gluten, as would be found in products contaminated with gluten.

Based on what current literature reports, potent T-cell stimulatory peptides are characteristically proteolytically resistant [9, 16, 17 and 18]. Therefore, from these data, one could conclude that peptides 1-7 do show resistance to further degradation. Peptides 4 and 6 do show moderate degradation and therefore, may not be highly immunostimulatory. Interestingly, the peptide that has been reported as the immunodominant (most potent) T-cell stimulatory peptide [11, 16], the αG-33 (33mer), was also almost totally digested by the GDE and completely digested by the glutenases. This 33mer peptide was spiked into the solvent standard along with peptides 1-7 in order to compare its resistance to digestion. The presence of the native 33mer was detected in wheat using this HPLC-MS assay; however, it was detected at very low levels as compared to the other seven immunostimulatory peptides.

5.4 Conclusions

Two practical examples of the application of this methodology are described in this chapter. (1) One practical application of the methodology involved its utilization in the evaluation of the quality of commonly available consumer foods and products. (2) A second application involved the assessment of the capabilities of a new therapeutic strategy under development, which attempts to address the

challenges faced from the continuous daily exposure to trace levels of gluten, in the management of celiac disease.

High performance liquid chromatography-mass spectrometry is a powerful analytical technique for use in the analysis of food proteins. The goal of this research was to identify a subset of potentially immunogenic wheat gluten peptides and develop an analytical methodology for their quantitative detection in complex samples. This novel analytical methodology is a combination enzymatic digestion-LC-MS procedure that was developed specifically to determine if wheat gluten is present, by releasing, detecting and quantifying any of seven immunogenic wheat gluten peptides in food (native or processed). The data shown demonstrate that this assay can detect and quantify these immunogenic gluten peptides, in a variety of grains, food and consumer products, down to approximately 0.5 µg/50g (corresponding to about 10 ppb) of product.

Peptide #	Sequence	Time Segment	MRM Transition *	Collision Engergy (V)
1	LQPQNPSQQQPQEQVPL	4	**980.7(+2) - 866.6(+2)** *980.7(+2) – 1150.7*	20 30
2	TQQPQQPFPQQPQQPFPQ	5	*1075.9(+2) – 1195.6* **1075.9(+2) - 956.6**	30 25
3	VPVPQLQPQNPSQQQPQEQVPL	6	**1240.9(+2) – 1126.7(+2)** *1240.9(+2) – 762.4*	25 30
4	RPQQPYPQPQPQY	3	*814.6(+2) – 1221.8* **814.6(+2) – 407.3**	25 30
5	QPQQPFPQTQQPQQPFPQ	5	*1075.9(+2) - 726.3* **717.6(+3) – 244.1**	27.5 17.5
6	PQQSPF	2	**703.4 - 441.4** *703.4 - 263.3*	25 35
7	QPQQPLPQPQQPF	5	**767.1-263.2** 767.1 - 917.6	15 15

Note:
* charge states other than 1 is shown in paranthees
MRM transitions in **BOLD** were used for quantification
MRM transitions in *italics* were used for confirmation

Table 5.1

MS/MS Conditions

product description	product	Measured amount of peptide (ng/mg) peptide 1	peptide 2	peptide 3	peptide 4
native flours	* corn flour	ND	ND	ND	ND
	* soy flour	ND	ND	ND	ND
	wheat flour	240	240	250	43
	wheat gluten	510	730	700	220
	rye flour	0.22	0.36	0.16	0.12
	barley flour	0.02	ND	1.1	D
	* quinoa flour	ND	ND	ND	ND
	* oat flour	D	ND	D	ND
	* rice flour	ND	ND	ND	D
processed foods	bread	91	140	130	67
	crackers	13	28	9.7	11
	goldfish crackers	190	180	140	14
	cheerios	3.3	2.3	3.1	0.63
	pasta	4.4	52	29	26
	orzo	12	90	160	150
	* rice crackers	ND	ND	ND	ND
	* powdered ice tea mix	0.046	ND	0.25	0.16
	LOD (ng/mg)	**0.004**	**0.025**	**0.014**	**0.003**
	LOQ (ng/mg)	**0.020**	**0.100**	**0.050**	**0.020**

NOTE:
LOD = limit of detection
LOQ = limit of quantification
ND = not detected (below LOD)
D = between the LOD and LOQ
*** = expected to be gluten free**

Table 5.2(A)

Immunogenic gluten peptides 1-4 quantified in native flours and processed foods

Measured amount of peptide (ng/mg)

product description	product	peptide 5	peptide 6	peptide 7
native flours	* corn flour	ND	ND	ND
	* soy flour	ND	ND	ND
	wheat flour	250	14	2
	wheat gluten	600	65	5.3
	rye flour	D	D	3.5
	barley flour	ND	0.015	2.6
	* quinoa flour	ND	ND	ND
	* oat flour	ND	ND	ND
	* rice flour	ND	ND	ND
processed foods	bread	150	14	0.11
	crackers	28	2.8	0.33
	goldfish crackers	140	11	0.09
	cheerios	3.7	0.06	0.14
	pasta	D	4.4	0.07
	orzo	2.9	12	0.12
	* rice crackers	ND	ND	ND
	* powdered ice tea mix	ND	0.025	ND
	LOD (ng/mg)	**0.030**	**0.001**	**0.012**
	LOQ (ng/mg)	**0.100**	**0.010**	**0.050**

NOTE:
LOD = limit of detection
LOQ = limit of quantification
ND = not detected (below LOD)
D = between the LOD and LOQ
*** = expected to be gluten free**

Table 5.2(B)

Immunogenic gluten peptides 5-7 quantified in native flours and processed foods

product description	product	Measured amount of peptide (ng/mg) peptide 1	peptide 2	peptide 3	peptide 4
gluten-free labeled products	* gluten free pasta (1)	ND	ND	ND	ND
	* gluten free pasta (2)	ND	ND	D	0.11
	* gluten free bread	ND	ND	ND	ND
	* gluten free pretzels	ND	ND	ND	ND
	* gluten free rice seasoning mix	ND	ND	ND	ND
	* gluten free crackers	ND	ND	ND	ND
	* gluten free beer	ND	ND	ND	ND
	* gluten free pad thai seasoning	ND	ND	ND	ND
	* quinoa pasta	ND	ND	ND	ND
	* soy pasta	ND	ND	ND	ND
	* protein bar	ND	ND	ND	ND
	* quinoa cereal	ND	ND	ND	ND
beverages and sauces	beer	D	D	1.57	0.088
	gin	ND	ND	0.06	ND
	* potato vodka	ND	ND	ND	ND
	hot sauce	0.15	ND	ND	ND
	vinegar	D	D	2.61	0.12
	* rum	ND	ND	ND	ND
	* red wine	ND	ND	ND	ND
	* white wine	ND	ND	ND	ND
consumer products	body wash	0.49	0.23	0.7	0.83
	* toothpaste	ND	ND	ND	ND
	* body lotion	ND	ND	ND	ND
	LOD (ng/mg)	**0.004**	**0.025**	**0.014**	**0.003**
	LOQ (ng/mg)	**0.020**	**0.100**	**0.050**	**0.020**

NOTE:
LOQ = limit of quantification
LOD = limit of detection
D = between the LOD and LOQ
ND = not detected (below LOD)
*** = expected to be gluten free**

Table 5.2(C)

Immunogenic gluten peptides 1-4 quantified in gluten-free foods, beverages, sauces and consumer products

		Measured amount of peptide (ng/mg)		
product description	**product**	**peptide 5**	**peptide 6**	**peptide 7**
gluten-free labeled products	* gluten free pasta (1)	ND	ND	ND
	* gluten free pasta (2)	ND	ND	ND
	* gluten free bread	ND	ND	ND
	* gluten free pretzels	ND	ND	ND
	* gluten free rice seasoning mix	ND	ND	ND
	* gluten free crackers	ND	ND	ND
	* gluten free beer	ND	ND	ND
	* gluten free pad thai seasoning	ND	ND	D
	* quinoa pasta	ND	ND	ND
	* soy pasta	ND	ND	ND
	* protein bar	ND	ND	ND
	* quinoa cereal	ND	ND	ND
beverages and sauces	beer	D	0.091	D
	gin	ND	ND	ND
	* potato vodka	ND	ND	ND
	hot sauce	ND	ND	ND
	vinegar	1.5	0.026	D
	* rum	ND	ND	ND
	* red wine	ND	ND	ND
	* white wine	ND	ND	ND
consumer products	body wash	0.32	0.12	D
	* toothpaste	ND	ND	ND
	* body lotion	ND	ND	ND
	LOD (ng/mg)	**0.030**	**0.001**	**0.012**
	LOQ (ng/mg)	**0.100**	**0.010**	**0.050**

NOTE:
LOD = limit of detection
LOQ = limit of quantification
ND = not detected (below LOD)
D = between the LOD and LOQ
*** = expected to be gluten free**

Table 5.2(D)

Immunogenic gluten peptides 5-7 quantified in gluten-free foods, beverages, sauces and consumer products

experiment	%peptide1	%peptide 2	%peptide 3	%peptide 4	%peptide 5	%peptide 6	%peptide 7	33 mer
digestion of solvent standard								
Control (PTC digest of solvent standard)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
peptides 1-7: gluten enzume x	0.2	ND	0.7	0.1	ND	ND	ND	ND
peptides 1-7: general digestive enzyme	33.0	52.8	56.1	0.3	40.5	8.9	36.0	4.3
peptides 1-7: gluten enzume y	ND	ND	ND	0.2	ND	0.1	ND	ND
digestion of 100% wheat								
Control (PTC digest of wheat)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
100% wheat: gluten enzume x	0.8	1.5	2.2	0.6	1.4	ND	ND	ND
100% wheat: general digestive enzyme	50.6	112.6	112.5	1.5	113.2	1.6	71.4	ND
100% wheat: gluten enzume y	0.7	0.2	0.9	0.7	0.2	ND	ND	ND
digestion of 1% wheat in corn								
Control (PTC digest of 1% wheat in corn)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	ND
1% wheat: gluten enzume x	1.2	ND	0.1	ND	ND	ND	ND	ND
1% wheat: general digestive enzyme	39.6	101.6	104.5	0.4	105.4	0.4	102.5	ND
1% wheat: gluten enzume y	0.2	ND	0.3	0.7	ND	ND	ND	ND

Note:
% peptide = % remaining after digestion
ND = not detected

Table 5.3

HPLC-MS/MS evaluation of glutenase and digestive enzyme activity

Figure 5.1

Target peptides 1-7 detected in gluten-containing food

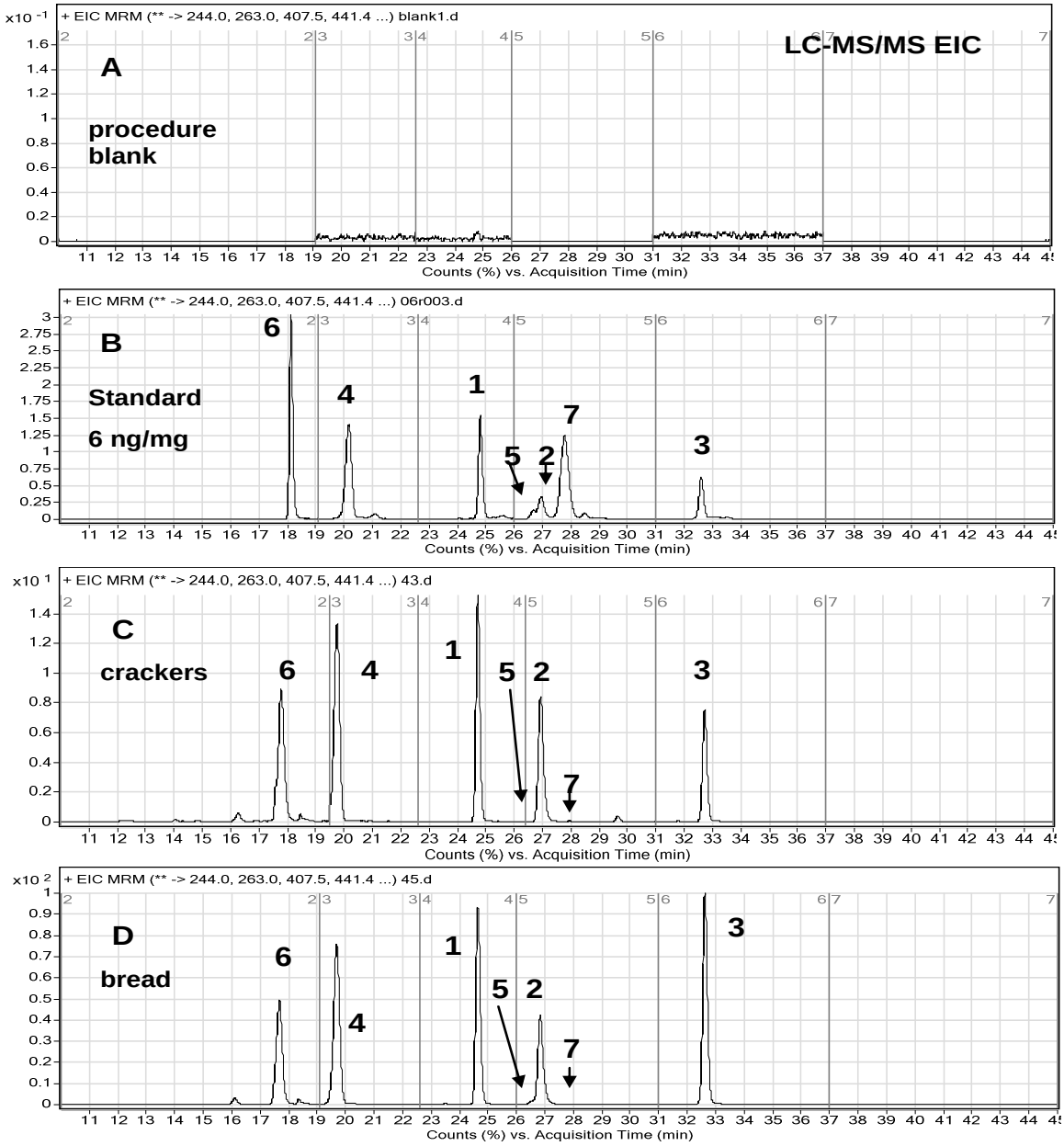


Figure 5.2

Target peptides 1-7 detected in gluten-containing food

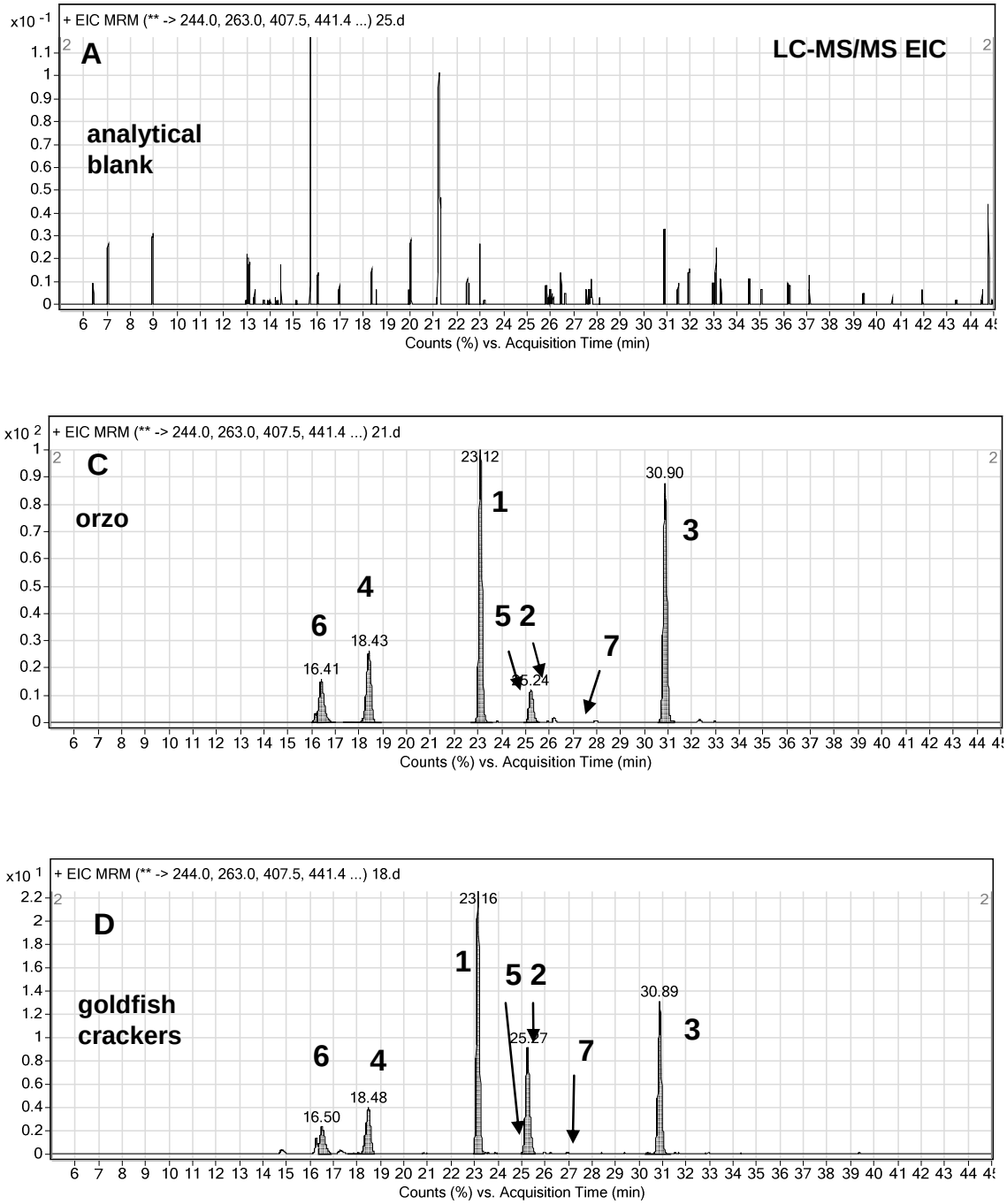


Figure 5.3

Target peptides 1-7 detected in gluten-free food

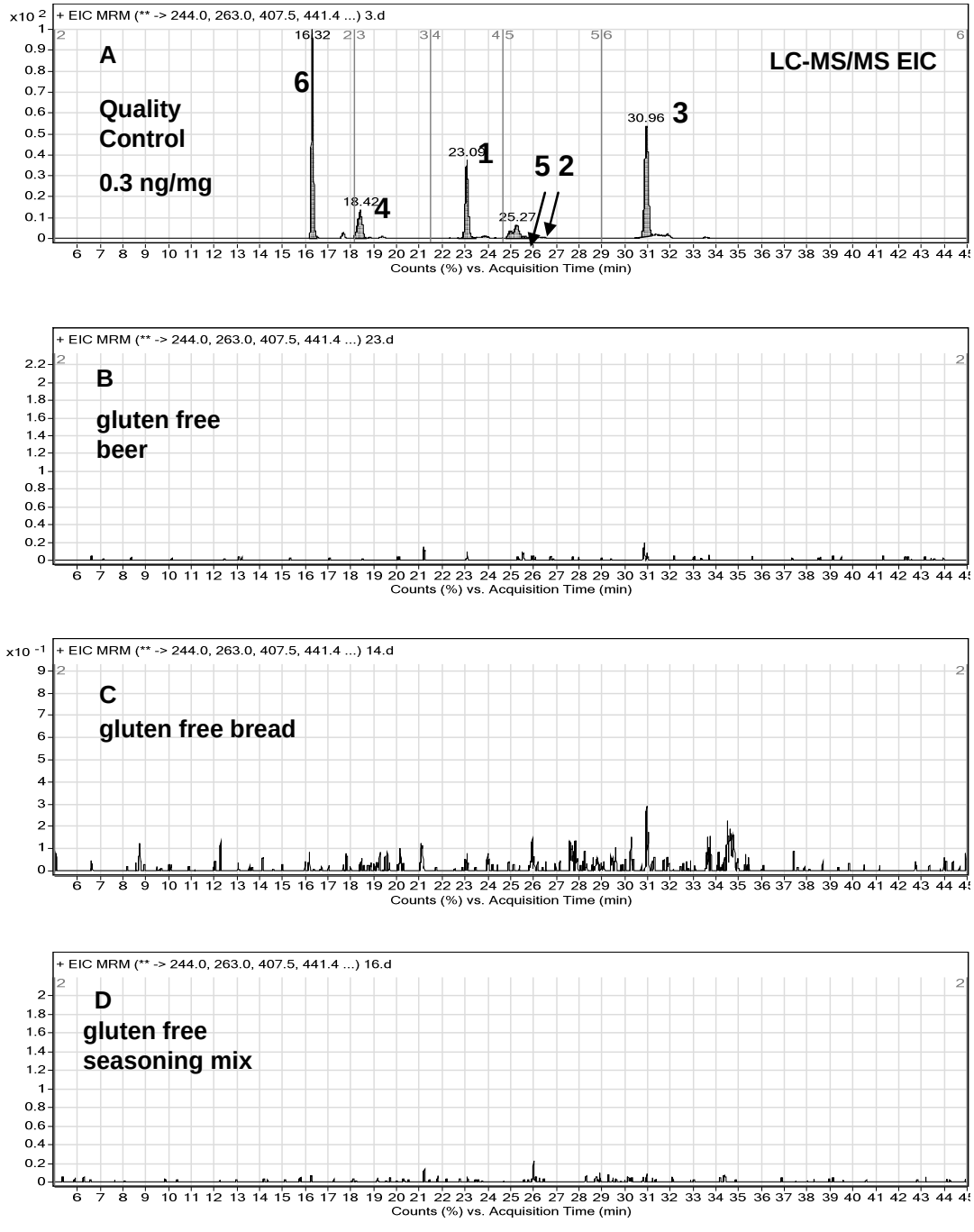


Figure 5.4

Target peptides 1-7 detected in gluten-containing liquids

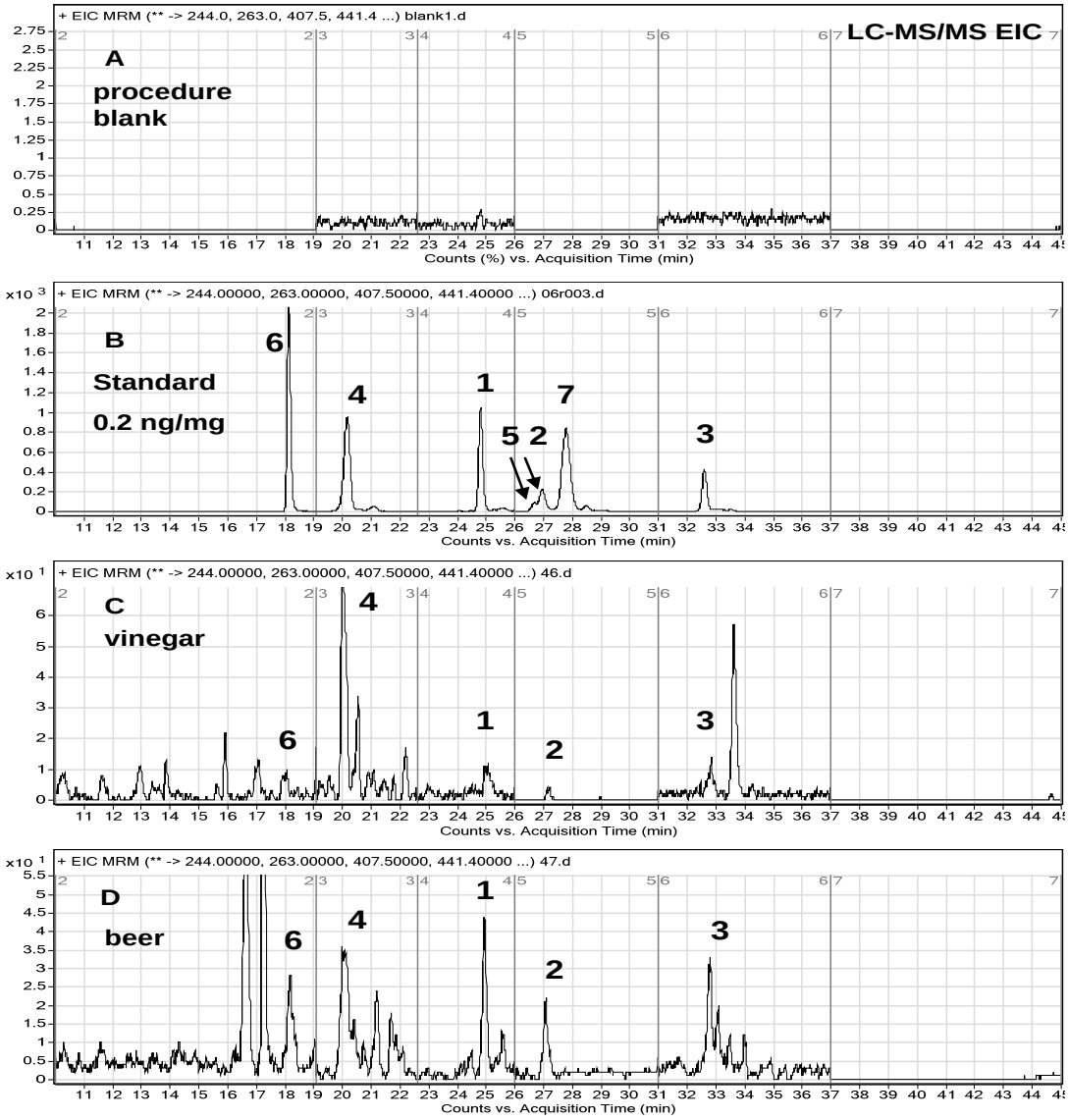


Figure 5.5

Target peptides 1-7 detected in native grains

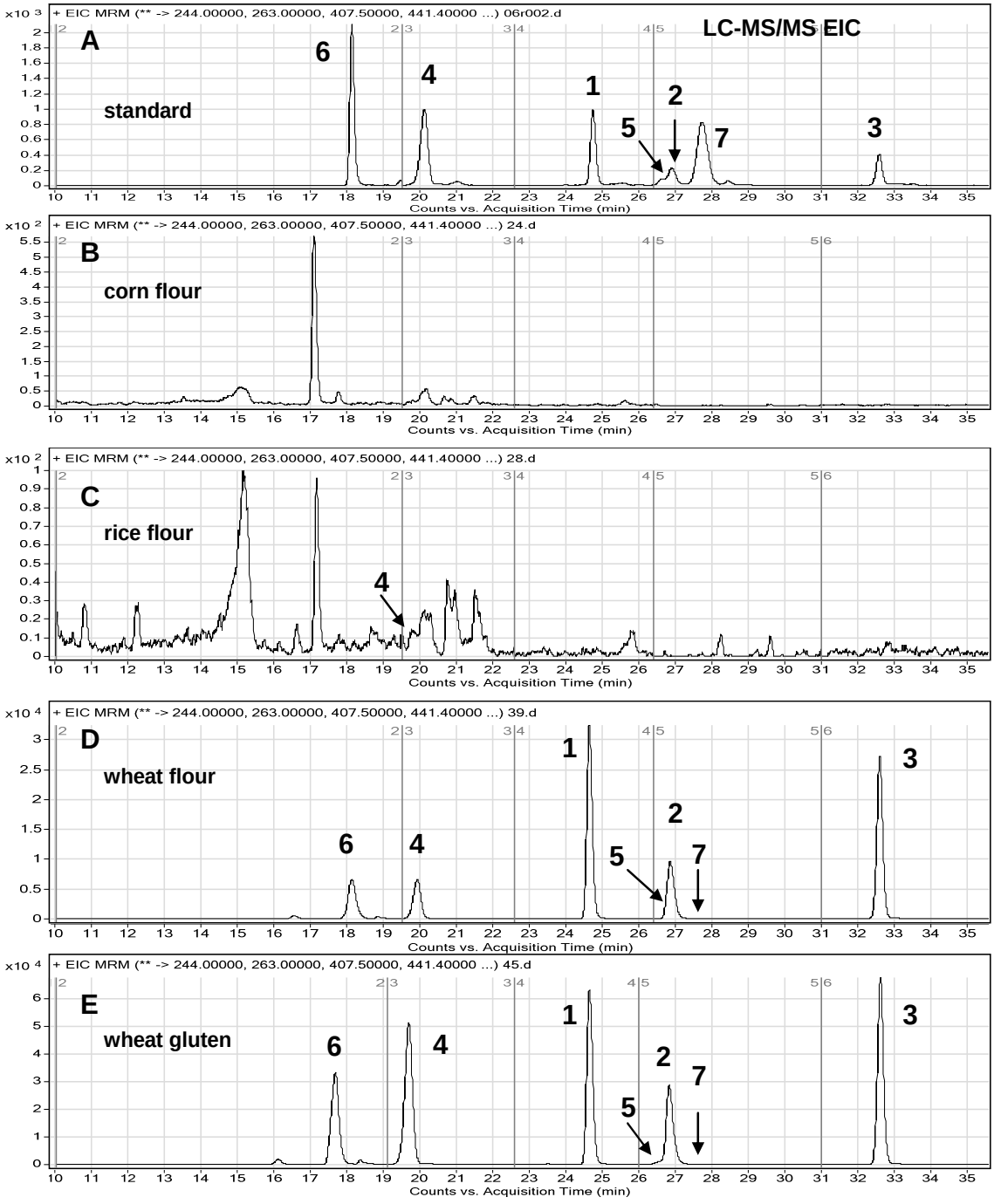


Figure 5.6

Peptide 7 in wheat gluten and native flours

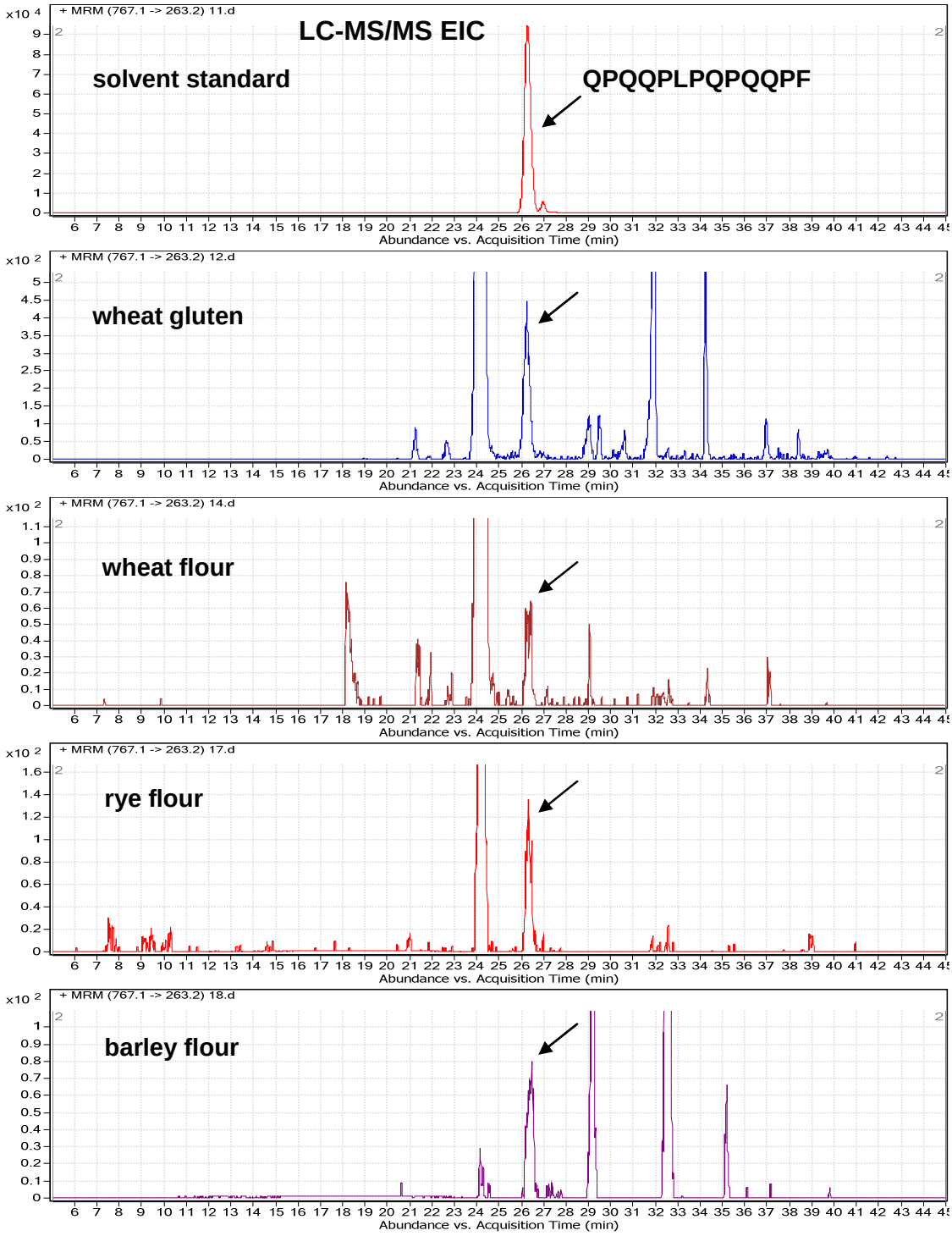


Figure 5.7

Gluten peptides found in a powdered ice tea mix

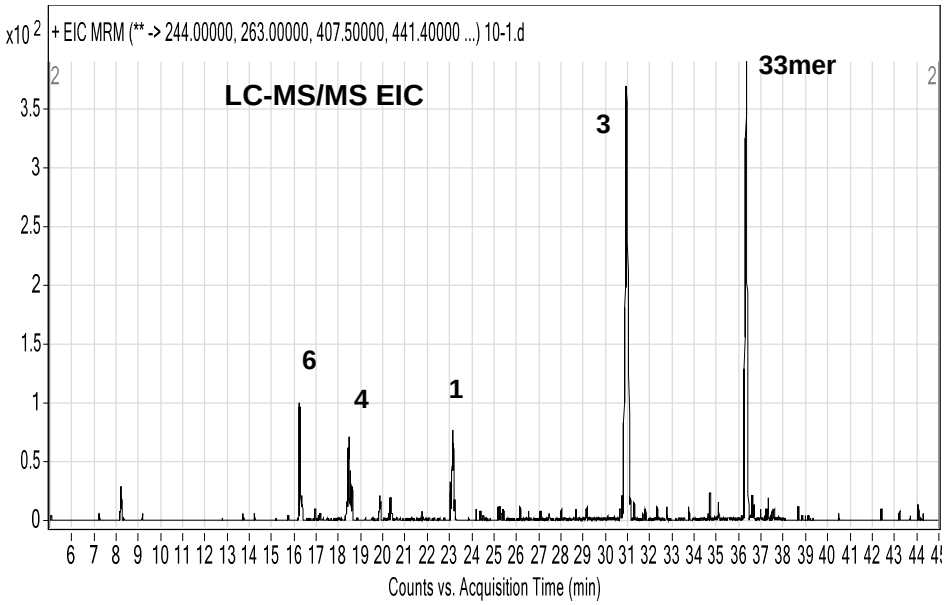


Figure 5.8

Gluten peptides 1-7 detected in wheat-containing and gluten-free pasta

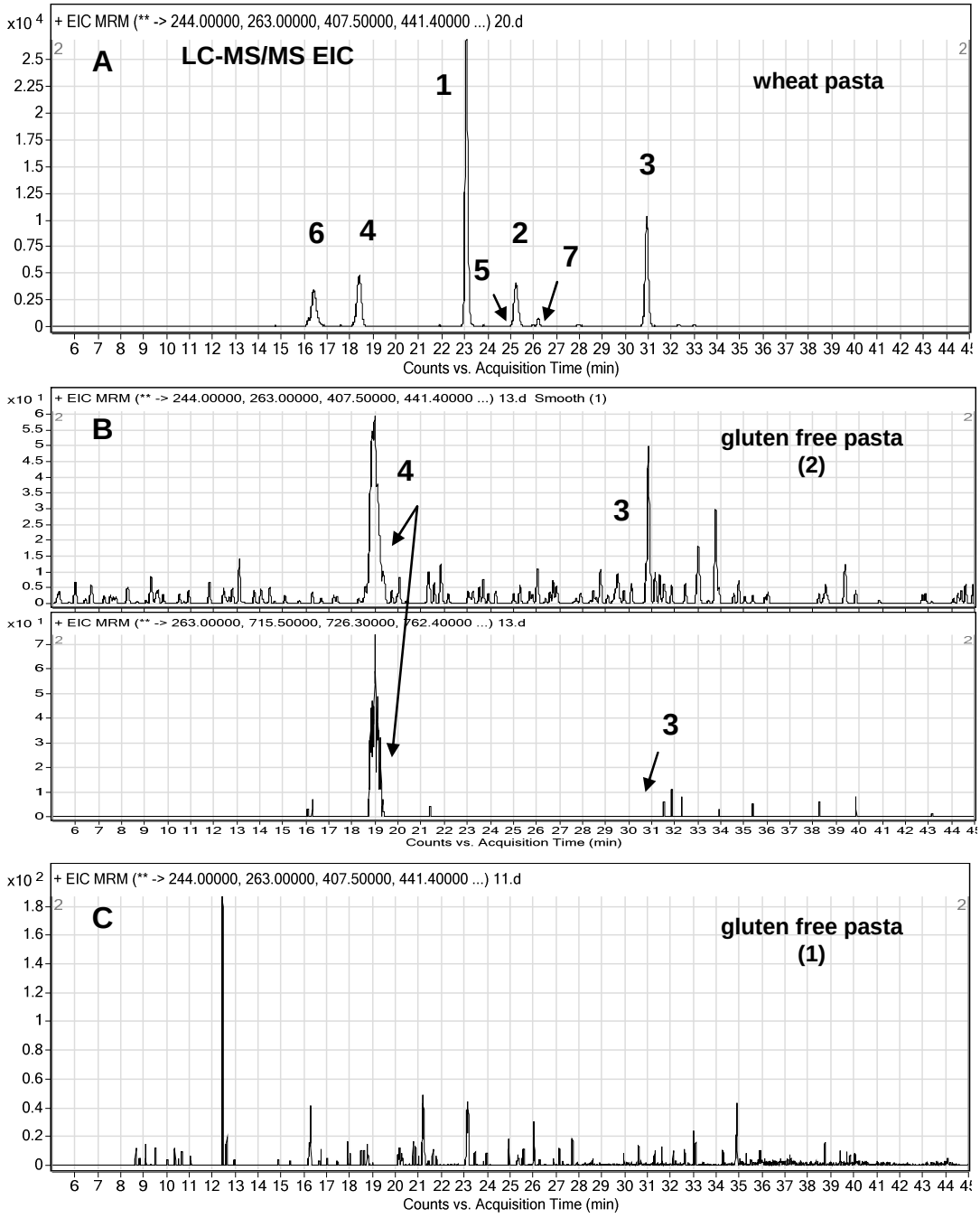


Figure 5.9

Gluten peptides 1-7 detected in a body wash product

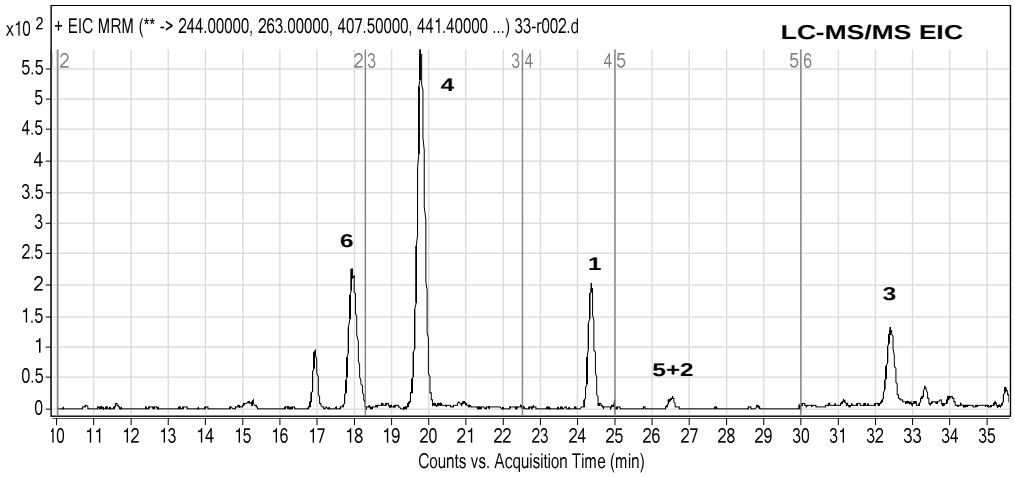


Figure 5.10

Effects of glutenase and digestive enzymes on a solvent standard cocktail of peptides 1-7 + αG-33

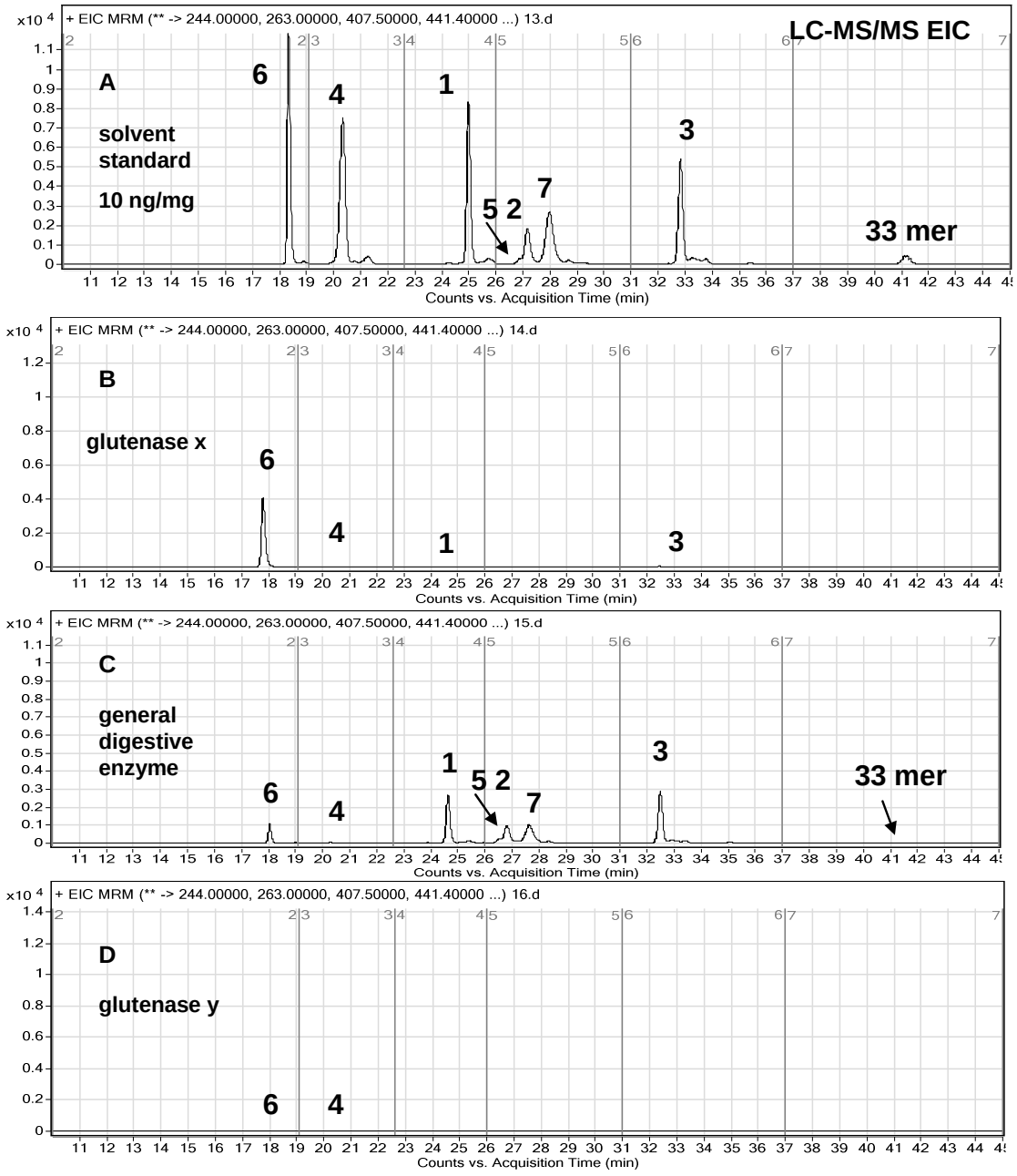


Figure 5.11

Effects of glutenase and digestive enzymes on peptides 1-7 and αG-33 in 100% wheat

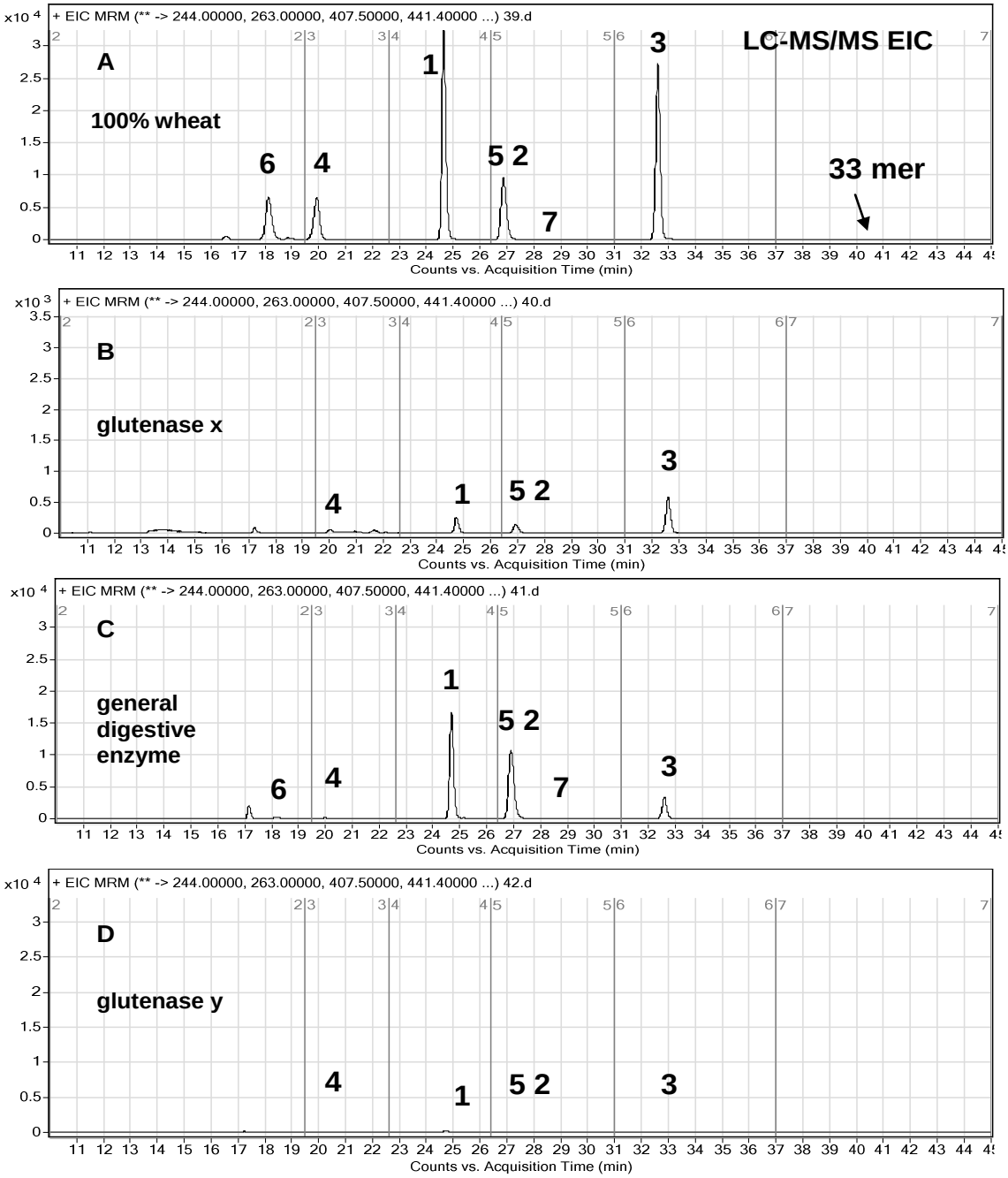
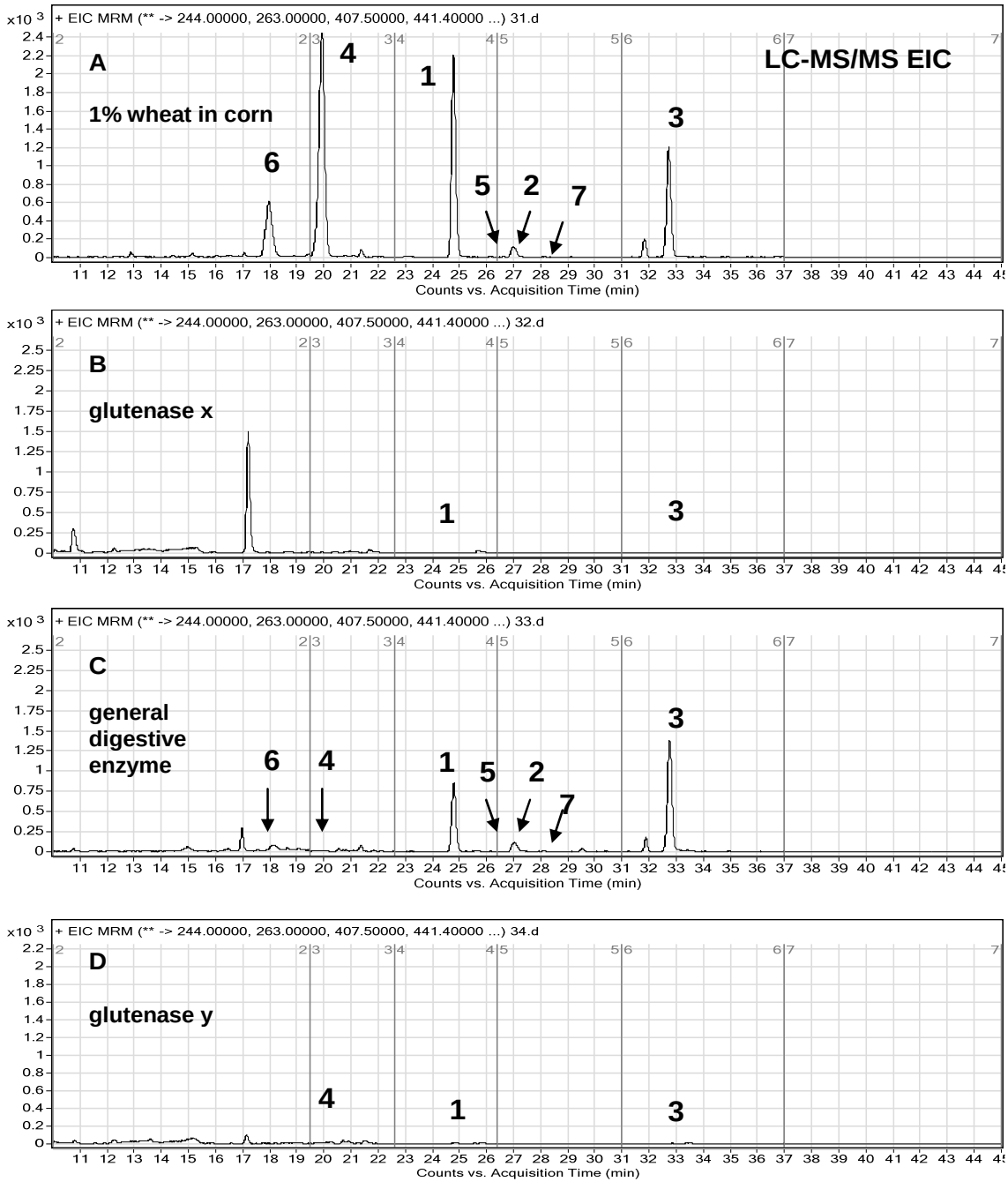


Figure 5.12

Effects of glutenase and digestive enzymes on peptides 1-7 in 1% wheat spiked into corn



5.5 References

[1] Lester, D. *Plant Methods* **2008**, *4*, 1-5.

[2] Troncone, R.; Auricchio, R.; Granata, V. *Current Opinion in Clinical Nutrition and Metabolic Care* **2008**, *11*, 329.

[3] Niewinski, M. *Journal of the American Dietetic Association* **2008**, *108*, 661.

[4] US Department of Health and Human Services; Food and Drug Administration: Food labeling; Gluten-free labeling of foods: Proposed rule (http://www.cfsan.fda.gov/~lrd/fr07013.html).

[5] Codex Alimentarius Commission: Draft revised standard for gluten-free foods, Step 8[http://www.codexalimentarius.net/web/archives.jsp?year=08] ALINORM 08/31/26 **2008**.

[6] Catassi, C.; Fabiani, E.; Iacono, G.; D'Agate, C.; Francavilla, R.; Biagi, F.; Volta, U.; Accomando, S.; Picarelli, A.; DeVitis, I.; Ianelli, G.; Gesuita, R. *American Journal of Clinical Nutrition* **2007**, *85*, 160.

[7] Hischenhuber, C.; Crevel, R.; Jarry, B.; Makis, M.; Moneret-Vautrin, D.; Romano, A.; Troncone, R.; Ward, R. *Alimentary Pharmacology &Therepeutics* **2006**, *23*, 559.

[8] Thompson, T.; Mendez, E. *Journal of the American Dietetic Association* **2008**, *108*, 1682-1687.

[9] Mitea, C.; Havenaar, R.; Drijfhout, J.; Edens, L.; Dekking, L.; Koning, F. *Gut* **2008**, *57*, 25-32.

[10] Gass, J.; Bethune, M.; Siegel, M.; Spencer, A.; Khosla, C. *Gastroenterology* **2007**, *133*, 472-480.

[11] Camarca, A.; Anderson, R.; Mamone, G.; Fierro, O.; Facchiano, A.; Costantini, S.; Zanzi, D.; Sidney, J.; Auricchio, S.; Sette, A.; Troncone, R.; Gianfrani, C. *The Journal of Immunology* **2009**,*182*, 4158-4166.

[12] Ehren, J.; Moron, B.; Martin, E.; Bethune, M.; Gray, G.; Khosla, C. *PLoS ONE* **2009**, *4*, e6313.

[13] Denery-Papini, S.; Nicolas, U.; Popineau, Y. *Journal of Cereal Science* **1999**, *30*, 121-131.

[14] Garcia, E.; Lorente, M.; Hernando, A.; Kieffer, R.; Wieser, V.; Mendez, E. *European Journal of Gastroenenteroogy & Hepatology* **2005**, *17*, 529.

[15] Stern, M.; Ciclitira, P.; van Eckert, R.; Feighery, C.; Janssen, F.; Mendez, E.; Mothes, T.; Troncone, R.; Weiser, H. *European Journal of Gastroenterology & Hepatology* **2001**, *13*, 741.

[16] Shan, L.; Molberg, O.; Parrot, I.; Hausch, F.; Filliz, F.; Gray, G., M.; Sollid, L.; Khosla, C. *Science* **2002**, *297*, 2275-2279.

[17] Shan, L.; Khosla, C.; *Immunology, Endochronology and Metabolism Agents in Medicinal Chemistry* **2007**, *7*, 187-193.

[18] Hausch, F.; Shan, L.; Santiago, N.; Gray, G.; Khosla, C. *American Journal of Physiology: Gastrointestinal and Liver Physiology* **2002**, *283*, 996-1003.

CHAPTER 6

Future Research and Directions

6.1 Summary and final conclusions

This research focused on the study of cereal grain proteins, specifically those in the large gluten protein sub-family. Dietary gluten is a staple for people all over the world. However, dietary gluten is also a trigger of a variety of immune diseases, food intolerances and allergies. One such autoimmune disorder is celiac disease. Those who have the genetic predisposition can develop the disease, whereby exposure to proteins of the Triticeae tribe of the Gramineae (grass) family (including wheat, barley and rye) triggers both innate and adaptive immune responses. For those who must limit or exclude exposure to dietary gluten, it is essential to know if trace levels of gluten are present in commercially available food and consumer products. Because gluten is not always listed on a product label, or may be listed under an unfamiliar name, or may be present as a component due to contamination during processing, it is not always evident that gluten is present.

The objectives of this research were: (1) to identify and characterize potential immunogenic wheat gluten peptides and (2) to develop an analytical methodology using HPLC-MS for use in screening commercially available food and consumer

products for the quantitative detection of trace quantities of the wheat gluten peptides identified.

Initial efforts toward the identification of immunogenic wheat gluten peptides utilized an *in-vitro* procedure that was designed to simulate the *in-vivo* enzymatic digestive process. Gluten peptides, known to be immunogenic, resist complete proteolysis *in-vivo* and have sequences consisting of less than 40 amino acids, of which many are proline and glutamine. The *in-vitro* proteolysis procedure succeeded in its attempt to release gluten peptides from proteolyzed wheat gluten proteins. Many appeared to resist further proteolysis, thus providing the first proof of the principle of the concept. One such peptide was identified as a 17mer with the sequence LQPQNPSQQQPQEQVPL.

Experiments involving metal ion adducts to proline were employed in order to aid in the identification of the peptides released by proteolysis. Several metals did show an affinity for proline and sequence information was obtained about the C-terminal end of some of the peptides released. However, hundreds of gluten peptides were released via proteolytic digestion, and this type of affinity was not sufficient to provide enough specificity to make immunogenic peptide identification possible.

An alternate strategy was then developed which attempted to discriminate between those peptides that are immunogenic and those that are not. The strategy focused on certain aspects of how the body itself determines which peptides it considers as “antigens” and how it targets them during the autoimmune responses to gluten. A novel methodology was developed that involved the chemical tagging of

physiologically relevant wheat gluten peptides using recombinant human tissue transglutaminase 2. This approach used HPLC-ESI ion trap MS^n to identify and characterize seven previously unreported physically relevant immunogenic gluten peptides in proteolyzed wheat flour (LQPQNPSQQQPQEQVPL, TQQPQQPFPQQPQQPFPQ, VPVPQLQPQNPSQQQPQEQVPL, RPQQPYPQPQPQY, QPQQPFPQTQQPQQPFPQ, QPQQPLPQPQQPF and PQQSPF).

An analytical method was then developed for the quantitative detection of trace levels of the seven identified immunogenic gluten peptides using HPLC-MS/MS. Experimental results defined the capabilities of the method with respect to the levels of accuracy, sensitivity and specificity of detection of the peptides possible in a complex food matrix. In order to test the effectiveness of the developed assay, it was used in the evaluation of the quality of commonly available consumer foods and products. The data presented demonstrate how this method can detect trace levels of immunogenic gluten in products down to low ppb level. This is believed to be the most sensitive and comprehensive assay available for the trace detection of immunogenic gluten peptides in food and consumer products.

Another practical application of the developed methodology provides evidence as to how it could be utilized in medical research, in order to further our knowledge about gluten-related diseases and to aid in the development of methods that aim to treat them. The assay was used successfully in the assessment of the proteolytic capabilities of a new orally based food enzyme cocktail. This type of oral

enzyme therapy is designed to address instances of unavoidable everyday exposure to small amounts of gluten.

Overall, it was observed that under these experimental conditions, the seven physically relevant peptides identified do resist general gastric and pancreatic digestion. They all contain known immunostimulatory epitopes and they are all targets for the TG2 enzyme, which collectively, would infer that they belong in the class of gluten peptides that participate in the pathogenesis of the T-cell mediated inflammatory mechanism. The analysis, characterization and quantification of cereal grain gluten peptides and proteins are all complicated processes. However, the methodology developed through this research has the capabilities to successfully screen a wide variety of complex samples and report the presence of trace levels of gluten down the low ppb level.

6.2 Objectives for future research

This methodology is a refreshing new approach, complimentary to the traditional immunochemical methods currently endorsed for the detection of gluten in food. This unique non-immunochemistry based methodology will be able to contribute to future research capabilities toward development of drugs and therapies targeted at the destruction of immunogenic peptides *in vivo* and the detection of such peptides in food and consumer products for gluten-free labeling and certification purposes.

Future efforts with respect to this research will continue and will focus on three objectives: (1) continued efforts in method development of the HPLC-MS

assay for screening food and consumer products will be directed toward a commercial application, as a service to both the celiac community and the food industry, by providing an accurate and economic means to generate much needed data for researchers developing treatments for patients with gluten sensitivities and manufacturers producing and labeling products that are safe; (2) future work will also continue to attempt to identify additional immunogenic wheat peptides (from both gliadins and glutelins) and study rye and barley grains in similar detail. Interest lies in the identification of peptides common to all grains and as well as peptides unique to individual grains, in order to form an even more comprehensive list of marker peptides for use with the HPLC-MS/MS assay; (3) investigation of the *in-vivo* response of the various peptide sequences identified will provide interesting information as to the relative level of cytotoxicity of these peptide sequences (i. e. which peptides are the most immunogenic). To date, no results from such a study have been published. Comparing results of activity with human TG2 and T-cell stimulatory studies of peptide sequences from different grains will also provide information about the relative immunogenicity between grains and between patients.