12-2023

Development of an Intact Mass Spectrometry Method for the Detection and Differentiation of Major Bovine Milk Proteins

Emily F. Harley-Dowell
University of Nebraska-Lincoln, eharley2@huskers.unl.edu

DEVELOPMENT OF AN INTACT MASS SPECTROMETRY METHOD FOR THE DETECTION AND DIFFERENTIATION OF MAJOR BOVINE MILK PROTEINS

By

Emily F. Harley-Dowell

A THESIS

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Master of Science

Major: Food Science and Technology

Under the Supervision of Professors Melanie Downs and Phillip Johnson

Lincoln, Nebraska

December, 2023

DEVELOPMENT OF AN INTACT MASS SPECTROMETRY METHOD FOR THE DETECTION AND DIFFERENTIATION OF MAJOR BOVINE MILK PROTEINS

Emily F. Dowell, M.S.

University of Nebraska, 2023

Advisors: Melanie L. Downs and Philip E. Johnson

The authentication of products with claims regarding protein sources or compositions is a challenge for traditional analytical methods, which generally lack the required specificity whole protein analysis can provide. For example, the establishment of milk as “A2” is achieved through genetic testing of cows before milk production, with no methods to authenticate milk products themselves. Establishment of A2 milk is completed through genetic testing of the cows before milk production, but with no methods to authenticate the milk products themselves. Intact protein mass spectrometry (MS) has the potential to directly authenticate protein products, including specific proteoform claims. The development of an intact MS method to detect and differentiate major bovine milk proteins (αS1-, αS2-, β-, K-caseins, β-lactoglobulins, and α-lactalbumins) and their proteoforms is needed for protein profile claims and can be an effective tool to analyze milk products for protein authentication.

This was attained through three major phases: generation of a predicted mass database, optimization of sample preparation and instrument parameters conducive with intact bovine milk proteins, and the selection of deconvolution software for protein identification with a mass error tolerance (10 ppm). Fifteen powdered and HTST liquid milk products with an equal distribution of marked A2 and normal commercial products were selected. Each sample was diluted to 1 mg protein/mL in 50 mM ammonium

bicarbonate and then defatted through centrifugation of 15 minutes at 3,000 x g. The samples were then cleaned up, desalted through 3 kDa spin column filters, and then separated and analyzed by liquid chromatography mass spectrometry (LC-MS). Data was deconvoluted using BioPharma Finder sliding windows algorithm that were compared to the predicted database and mass were identified. A mean of 85.27% (± 6.68%, n = 57) of the total signal of powdered and liquid HTST milk could be assigned to the predicted database proteoforms using the finalized method. The average ratio of selected normal commercial products was 25.86% A1 and 0.74.14% A2.

ACKKNOWLEDGEMENTS

I would first like to thank my advisors Dr. Melanie Downs and Dr. Phillip Johnson for their support, guidance, and valuable input throughout my research. Thank you for introducing me to the field of mass spectrometry, food allergen detection, and challenging me to become a more independent scientist. Melanie, thank you for keeping me on track and making time to answer my questions. I am grateful for the opportunity to pursue this research and a degree at the University of Nebraska. I would additionally like to thank my committee members, Dr. Joe Baumert and Dr. Kaustav Majumder. Thank you for your time, input, guidance, and thought-provoking questions throughout the stages of this project.

I would like to sincerely thank Dr. Justin Marsh and Dr. Shyamali Jayasena for their assistance and guidance throughout my years in the FARRP lab. To Dr. Shyamali Jayasena, thank you for double checking my calculations and training me in day-to-day lab responsibilities. To Dr. Justin Marsh, thank you for training me in mass spectrometry, answering questions, checking calculations, guidance, and reminding me I don't have to be a perfect scientist. I will always remember the humorous office conversations, long history lessons, and advice.

I am grateful to all the staff and scientists who have helped me through this program and my time here in Lincoln, Nebraska: Pat Gergen, Julie Nordlee, Julie McManamey, and Dr. Rossana Villa Rojas. Pat, thank you for being a friendly face to talk to, your support, and reoccurring trips to the store to buy my samples. Julie Nordlee, chatting with you in the hallway or in the office always brought a smile to my face. ON, WISCONSIN! Julie McManamey, thank you for helping me with the logistics of my

program and answering my many questions. Rossana, thank you for your guidance and advice throughout my time as your teaching assistant.

Thank you to all my lab mates, however long you were in my life, who assisted added joy and comradery to my time in the department: Dr. Lee Palmer, Dr. Shimin Chen, Dr. Tengfei Li, Maggie, Ellenor, Francisca, and Zhenhao. I would especially like to thank my lab mates Sara, Jess, Olivia, Liyun, and Natalia. Thank you for welcoming me into the lab as the only new student at the time. Sara and Jess, thank you for your guidance throughout my program, reading over very rough drafts of my proposal and thesis, advice, and your friendship. Olivia and Liyun, sharing an office with you two have been a bright point in my program. Thank you for asking how I'm doing and supporting me throughout the program. I will miss the chats we had and the support we gave each other. Natalia, I am so grateful for your friendship and advice. I will miss turning around and getting to talk with you in the office. A special thanks to my peers who became my friends, Dr. Snigdha Guha, Emerson, Prabhashis, Rachel, Vanessa, and Jerry.

Thank you to my family and especially my parents. You have always been supportive of my pursuit of knowledge and have challenged me to be curious. Thank you for all the phone calls and reminding me who I am when faced with adversity. To my sister, Jennifer, thanks for visiting and taking care of me when I knew no one, all the phone calls, and reassurance that I am where I am supposed to be.

To my friends and chosen family at Sower Church, I cannot adequately express how thankful I am for your community, love, and grace. Thank you for pointing me back towards Him when I get distracted with the things in life. A special thanks to Katie, Grace, Phuong, Jordan, Jessica, Colton, Natalie, Cindy, Kara, and Abby.

Lastly, I must express my deepest thanks to my husband, Bradley. Thank you for your endless love, support, and encouragement through the late nights and stressful weeks. Thank you for reminding me what a work-life balance is and when to give myself grace. You bring joy to each day and have comforted me when I need you most.

Finally, I would like to thank FARRP for providing the funding for this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: LITERATURE REVIEW

I. INTRODUCTION

It is customary practice to trust the label on food. This trust is based on the expectation that the producer has used specified ingredients and processed the food as described in a set production system [1]. This mindset has been reinforced through legislation constructed by government regulatory agencies such as the United States Food and Drug Administration (FDA), Food Standards Australia New Zealand (FSANZ), and European Food Safety Authority (EFSA) to ensure product and label are identical. This is done to protect consumers' health, ensure food quality, and provide transparency in the food industry [2]. In the United States, the Federal Food, Drug, and Cosmetic Act (FD&C Act), first established in 1938, granted the FDA authority to oversee the production of medical devices and cosmetics and define food standards [3]. Section 342 in the FD&C Act gives an extended, detailed definition of adulterated food and section 343 covers misbranded food [4, 5]. Adulteration, generally, happens when a product’s quality is lowered and/or altered either with the addition of or substitution of substance(s) [6]. When food is adulterated, it introduces the possibility of foreign substances that could lead to adverse health effects and consumers are unknowingly deceived.

Because of the possibility of adulteration, the need for food authentication is required. Authentication is the analytical process used to establish the truth and accuracy of a food/food product and its label [7]. There are a multitude of authentication methods specific to the ingredients, food products, and possible adulterant in question that have been developed. Proteins have been a popular choice for these methods as analysis can provide details on food properties which is primarily useful for a systematic assessment of the food chain. However, some proteomic techniques face the challenges of excessive

proteoforms found in products when attempting to distinguish accurately and faulty labeled products. Mass spectrometry (MS)-based proteomics can detect and identify similar proteoforms of different proteins within an efficient amount of time and will be the central focus of this review. This review will focus on U.S. cow's (*Bos tarus*) liquid and powdered milk products and legislation covering adulteration and authentication in the United States.

II. ADULTERATION AND AUTHENTICATION

Legislation surrounding adulteration and authentication has been established in many countries to protect consumer health and ensure food products match what is declared on the label. In the United States, there are several laws and regulations to ensure that food is exactly what is declared on the label and safe is consumed. The Federal Food, Drug, and Cosmetic Act section 342 gives a detailed definition of adulterated food, and section 343 covers misbranded food [4, 5]. In the case that economic gain motivates food adulteration, it is classified as economically motivated adulteration (EMA) [6]. Defined by the FDA, EMA occurs when a person decides to add, omit, or substitute an ingredient of value to create an appearance that the food is of higher value. If EMA leads to safety hazards, the FDA Food Safety Modernization Act 2011(FSMA) specifically protects consumers by requiring preventative controls to be put into place for both human and animal food [8]. EMA is a serious crime as it poses a threat to public health [1]. Examples of EMA include, but not limited to, mixing honey or maple syrup with cheaper sweeteners, substituting high value fish species with lower value fish species, and diluting spices by a non-spice plant material [6, 9].

Milk is a target for EMA because of its high nutritional value, utilization in many different products, and consumption by all ages [7, 9-11]. The most notable past example of milk EMA is the addition of melamine, a chemical used in plastics, in infant formula to boost protein content in 2008. This resulted in over hundreds of thousands of hospitalizations and illness and several infant deaths [9]. With the risk of the intentional or even unintentional adulteration of milk, verifying the contents of milk is needed to assure consumer welfare and prevent fraud [12].

III. MILK CONSUMPTION

Milk has been a staple in many households and a major source of nutrients and energy to be enjoyed by all ages as it contains essential amino acids, fats, vitamins, and minerals [7, 9-11, 13]. One of the most versatile animal-derived commodities, it has been a paramount ingredient in many formulations and products, which include, but are not limited to, butter, creams, infant and baby formulas, cheese, baked goods, some alcoholic beverages, and ice cream. Though there has been a decrease in consumption of cow's milk and changing preferences in types of milk (whole milk to a low-fat milk, animal-derived milk to alternative sources (almond, soy, oat, etc.)), milk production has remained the same or even slightly increased as it did in 2022 from 2021 according to the 2022 USDA Milk Production Report [14-16]. However, that is not the case for all dairy products which, in total, have risen at a rapid pace in recent decades. The most notable dairy product, cheese, continues to grow in demand as it has become a pronounced part of U.S. consumers' diets. This can be attributed to a greater diversity of types of cheese, ethnic cuisines that utilize cheese in the dishes, and eating out. In addition to cheese, infant formula consumption has grown. In 2013-2020, approximately 75% of 6-month-

old infants (the age when infants can begin a diet outside of primarily human milk) exclusively receive formula or receive human milk supplemented with infant formula [17]. Powdered formula, the least expensive form of formula, is quite popular and went from 44% to 62% of infant and baby formula sales from 1994-2000 [18]. In 2022, some estimates indicated that the U.S. formula market was valued at $3.962.7 million and is projected to reach $6.973.7 million by 2032 [19]. Though there has been a shift in the type of dairy products consumed, the need for production of cow's milk has not changed and remains a vital part of the food industry.

IV. BOVINE MILK PROTEINS

a. Classification of Proteins

Bovine liquid milk is composed of approximately 3-3.5% (w/w) protein [20]. The protein content can be classified into two major groups: caseins (80%) and whey (20%). The casein proteins are further subdivided and consist of 39-46% αS1-caseins (α_{S1}-CN), 25-35% β-caseins (β-CN), 8-11% α_{S2}-caseins (α_{S2}-CN), 8-15% κ-caseins (κ-CN). The whey proteins consist of 60% β–lactoglobulins (BLG), 20% α-lactalbumins (α-LA), 10% bovine serum albumin (BSA), and 10% immunoglobulins (Ig) (Figure 1.1) [21, 22].

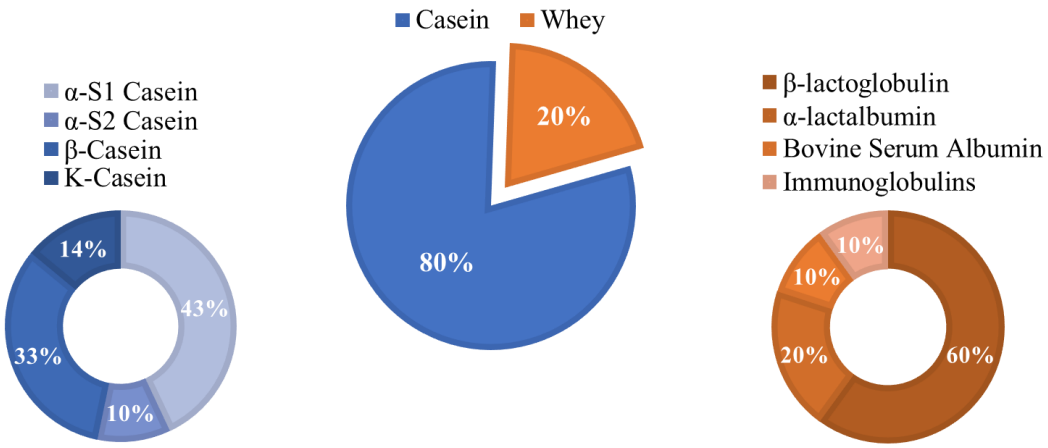


Figure 1.1 Precent Distribution of Major Proteins Found in Bovine milk. Maximum percentage possible is reported. Percent information obtained from Eigel et al., 1984 and Vincent et al., 2016.

Caseins are deposited in large spherical particles called micelles that consist of α-CNs (~23.6 kilodaltons (kDa)) and β-CNs (~24 kDa) surrounded by a layer of κ-CNs (~19 kDa) and α-CNs. κ-CNs are found on the outside due to their amphiphilic properties while the other caseins are more hydrophobic (Figure 1.2) [13, 23].

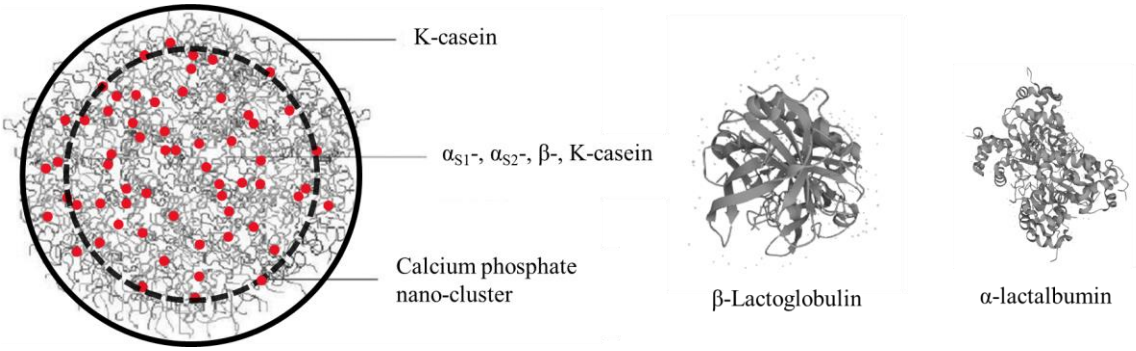


Figure 1.2: Structure of Casein Micelle and Monomers of Whey Proteins. Casein micelle structure is adapted from Hristov et al., 2014 and whey structures were obtained from UniProt P02754 and P00777.

The micelle is stabilized and internally held together by casein hydrophobic interactions and electrostatic interactions between small mass species (phosphoserine clusters and calcium phosphate). The micelles range in size from 50-500 nm in diameter

(120 nm on average) and mass from 10^6 to 3 x 10^9 Daltons (Da) (10^8 on average) [9]. The whey proteins, BLG (~18 kDa) and α-LA (~14 kDa), are naturally independent of this structure. BLG can exist as monomers, dimers, and as tetramers in different pH ranges: between <pH 3.5 and >pH 7.5, pH 5.5-7.5, and pH 3.5-5.5 respectively.

b. Bovine Genetics

Bovine species alleles are either homozygous (producing a single variant) or heterozygous (producing both variants due to co-dominance) [24, 25]. Focusing on β-CNs, the alleles are co-dominant which results in two β-CN types being secreted in milk: A1- and A2-type, which are described in more detail below. A2 milk is produced by cows that possess the genotype A2/A2. This genotype only produces the A2-type variants of β-CN protein. The two other genotypes are A1/A2, which result in the production of both A1- and A2-type protein variants, and A1/A1, which solely produces the A1-type variant of β-CN (Figure 1.3) [26]. Production of the different proteins and variants differ due to gene frequencies which differ from breed to breed [27]. However, some variants do not differ, and others are rarely produced.

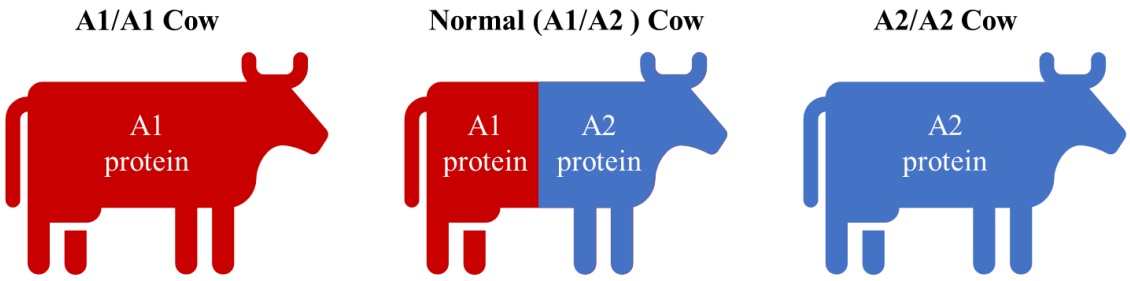


Figure 1.3: Protein Production of A1/A1, Normal A1/A2, and A2/A2 cows. Figure adapted from A2 Milk Company website.

It is believed that the A1-type allele is found primarily in Northern European dairy cattle breeds such as Friesian, Holstein Friesian, Ayrshire, and British Shorthorn

while the A2-type allele is seen in various parts of the world in Guernsey, Jersey, Charolais, Limousin, and Zebu breeds [25]. The change in production of the A2-type to A1-type variants is attributed to a mutation in the European Holstein herd ten millenniums ago [28, 29]. Through genetic testing of herds in Canada, the United States, and Italy, breeds with a higher frequency of the A2/A2 genotype were found to be Jersey, Indian Zebu, Brown Swiss, and Guernsey but seems to vary per country [25, 27, 30, 31]. In the United States, most dairy cows are traditionally Holsteins because they tend to produce more milk per cow, but Jersey and crossbreed cows have gained popularity possibly due to the A2-type protein produced [32]. If farmers want to produce A2 milk, they will genetically test their cows to select for the A2/A2 genotype and isolate cows without the desired genotype from the rest of the herd [29]. When breeding for A2/A2 cows, the genotypes for each parent must be considered to create the desired A2/A2 genotype. Both parents possessing the A2/A2 genotype is preferred as 100% of their offspring will have the genotype A2/A2, but that is not the only option for creating A2/A2 genotype offspring (Table 1.1).

Table 1.1: Possible Genotype Combinations when Breeding Cows to Produce the A2/A2 Genotype[a]

Parent 2	Parent 1		
		A1	A2
A1	A1	A1A1	A1A2
A1	A1	A1A1	A1A2
50% A1/A1 50% A1/A2			

Parent 2	Parent 1		
		A1	A2
A1	A1	A1A1	A1A2
A2	A1	A1A2	A2A2
25% A1/A1 25% A2/A2 50% A1/A2			

Parent 2	Parent 1		
		A1	A2
A2	A1	A1A2	A2A2
A2	A1	A1A2	A2A2
50% A1/A2 50% A2/A2			

Parent 2	Parent 1		
		A1	A1
A2	A1	A1A2	A1A2
A2	A1	A1A2	A1A2
100% A1/A2			

Parent 2	Parent 1		
		A2	A2
A2	A2	A2A2	A2A2
A2	A2	A2A2	A2A2
100% A2/A2			

[a]Adapted from Beavers & Van Doormal, 2016.

The two prominent commercial A2 liquid milk sellers in the United States are The a2 Milk Company originally from Australia and Alexandre Family Farms located in California. The a2 Milk Company has not released details about the breed(s) they use for

their herds but lists their cow breeds as genetically tested cows that only produce the A2 protein[33]. The Alexandre Family Farm does list their breed and uses a crossbreed of New Zealand Kiwi Cross, German Fleckvieh, New Zealand Ayrshire, Dutch Holstein, and Danish Jersey to produce their A2 milk [34]. Because of the crossbreeding of cows and differing production of proteins and variants, the protein profile of a herd of cows from one country or region could be very different from milk in another country or region complicating the comparison of the two.

c. Biological and processing proteoforms

Bovine milk proteins possess many biological and processing modifications in addition to genetic variants, which leads to numerous proteoforms. α_{S1}-CN has four known variants, β-CN has 12 known variants that are distinguished by the 76th amino acid (AA): A1-type with proline and the A2-type with histidine, α_{S2}-CN has no known variants, κ-CN has seven known variants, BLG has five known variants, and α-LA has two known variants (main sequence and Droughtmaster) (Figure 1.4) [35].

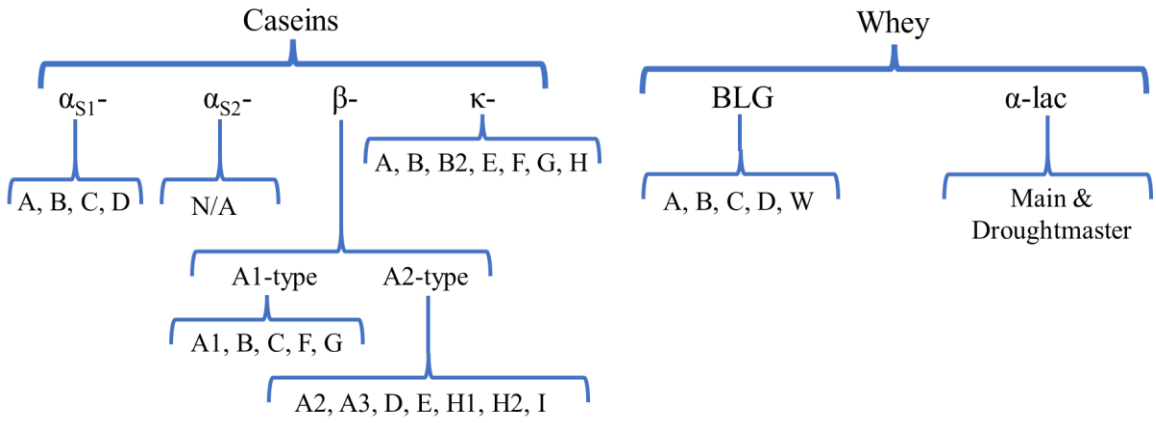


Figure 1.4: Genetic Variants of Bovine Casein and Whey Proteins. Information obtained from UniProt and Farrell et al., 2004.

The main additional biological modifications are the phosphorylation and O-linked glycosylation of caseins and formation of disulfide bonds in caseins and whey

proteins. All caseins undergo phosphorylation at different sites and in different abundances (α_{S1}-CN: up to 8-9 phosphate modifications (Phos), α_{S2}-CN: up to 10-13 Phos, β-CN: up to 4-5 Phos, and κ-CN: up to 1-2 Phos per molecule) [9]. Phosphorylation occurs when caseins are transferred across the endoplasmic reticulum membrane where a casein kinase attaches phosphate groups to two specific amino acids (serine and threonine). This results in in a mass addition of 79.97 Da [36, 37]. O-linked glycosylation primarily affects κ-CN at the serine/threonine residues [38]. Sites of glycosylation are highly variable as are the glycan adducts (eight structural classes) making it hard to monitor and predict on proteins [39-42]. Two caseins and two whey proteins have the potential to form disulfide bonds. Every bond that occurs between two cysteine amino acids results in a mass subtraction of 2.0156 Da because of the hydrogen loss (αS2-CN: up to 2 cysteines, κ-CN: up to 2 cysteines, BLG: up to 4-5 cysteines, and α-lac: up to 8 cysteines) [9].

Milk typically undergoes some form of heat treatment, which can cause additional processing modifications. One example of a heat treatment is pasteurization, usually HTST (high temperature, short time 72°C for 15 seconds), which can lead to the denaturation of whey proteins and increased levels of non-micellar casein in milk due to the change in protein confirmation [43, 44]. During higher temperature treatments such as spray drying, higher degrees of denaturation are often observed, and stimulation of aggregation occurs. As well as confirmation changes, denaturation, and aggregation, the Maillard reaction also occurs [45]. This is a chemical reaction between amino groups and reducing sugars. During heat treatment, Maillard reactions between the free amino group on lysine and lactose (a reducing sugar) cause lactosylation, resulting in a mass addition

of 324.10 Da for impacted proteins. The extent of lactosylation highly depends on the conditions of thermal processes with more intense treatments generally leading to more frequently lactosylated proteins [45, 46]. As such, the milk protein proteoform complexity can be increased by multiple lactosylation events, upon heat treatment. With these modifications, the protein profile of milk becomes very complex.

V. HEALTH CLAIMS AND MARKET VALUE

A2 milk has perceived health benefits and thus increased market value. The call to switch to A2 milk is due to claims about the digestion of certain milk proteins [47]. Normal commercial milk contains a mixture of both A1-type and A2-type β-CNs while A2 milk *only* has the A2-type β-CNs [24, 25]. The major claim is that under gastrointestinal digestion of the A1-type β-CN, β-casomorphin-7 (BCM-7), a bioactive opioid peptide, is released [24]. This peptide could be a cause for concern because of the high affinity for opioid receptors causing opioid-like effects that affect the nervous system and gastrointestinal functions [48]. The enzymes involved in this release are leucine aminopeptidase (LAP) and pancreatic elastase (Figure 1.5). On the other hand, the A2-type is seemingly more resistant to digestion, and the release of BCM-7 is minimal [49]. This is thought to be caused because the bond between the isoleucine (66th AA) and proline (67th AA) is difficult for elastase to cleave while the bond between the isoleucine (66th AA) and histidine (67th AA) can be cleaved more easily. When digesting A2-type protein, the thought is that no or very minimal BCM-7 will be released [24, 49].

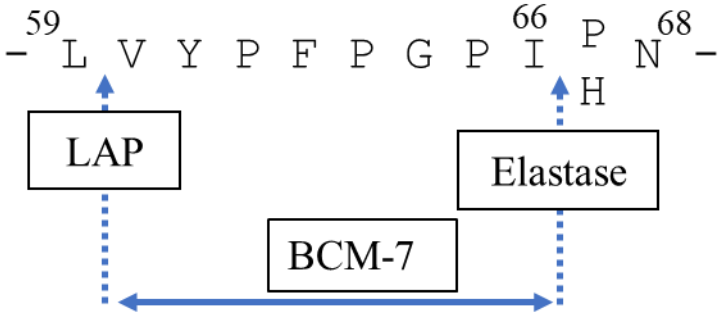


Figure 1.5: Enzymes that Cleave β-casomorphin-7 (BCM-7) from Bovine A1-Type β-Casein.
Arrows show cut sites of enzymes and BCM-7 resulting from the enzymes. Adapted from Jinsmaa & Yoshikawa, 1999. Copyright 1999 by Elsevier Inc.

Opioid receptors can be split into three types of G-protein coupled receptor categories: μ, γ, and κ [50]. Primarily found in the central nervous system, opioid receptors can also be found in the peripheral nervous system, gastrointestinal tract, immune system, and in bone cells. In summary, opioid receptors are distributed throughout the body, allowing opioids to widely affect numerous functions, ranging from pain management to bone metabolism [51]. There are a handful of BCMs that could be released from β-CNs such as BCM-4, -5, -6, -7, -9, -13, -21 (Figure 1.6) [49, 52].

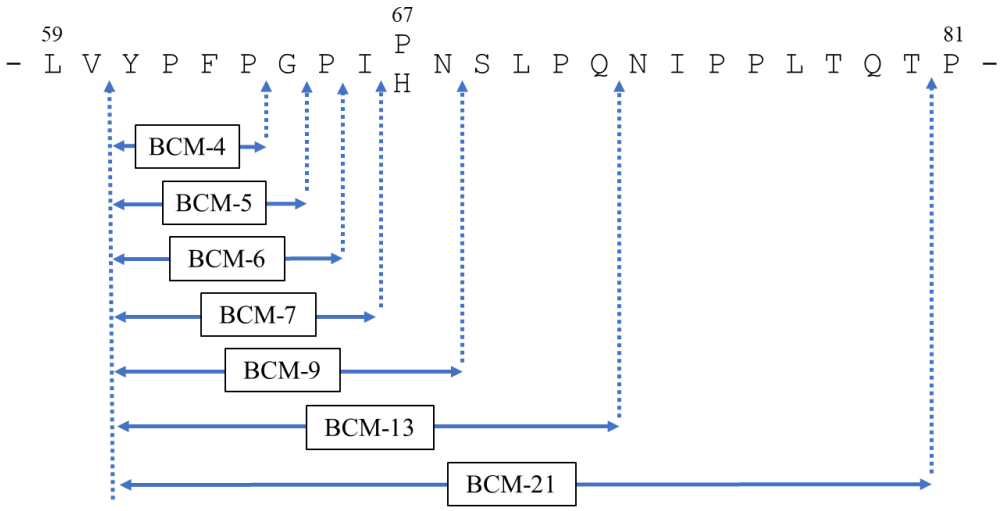


Figure 1.6: β-casomorphin-4, -5, -6, -7, -8, -13, -21 (BCM) from Bovine β-casein.
Arrows show cut sites of enzymes and BCMs resulting from the enzymes. Adapted from Jinsmaa & Yoshikawa, 1999. Copyright 1999 by Elsevier Inc. Information obtained from Jinsmaa & Yoshikawa, 1999 and Arısoy et al., 2019.

In current literature, there have been theories that BCMs could alter aspects of the lower gastrointestinal tract such as the microbiota, epithelial lining, metabolism of acids, gut transit time, and inflammation as well as other health-related attributes including bone growth and weight gain [51]. This thought process is derived from the current understanding that BCMs activate μ-opioid receptors in the gut [49, 50, 53]. To assess the theory and additional health claims, animal models and human studies have been used to investigate this idea.

There have been several animal studies focusing on gastrointestinal function, inflammation, cardiovascular health, type 1 diabetes, gut morphology and histopathological alteration, and gut microbiota [54-61]. These studies were conducted in a variety of animals from various types of mice, rats, pigs, and rabbits. There were many attributes tested in these studies: gastrointestinal transit, C and jejunal dipeptidyl peptidase (to test opioid-dependent functions), effects on parameters of blood (morphology, lipid profiles, liver enzymes, creatin, and urea), gut characteristics (lymphocyte subpopulations, enzymatic activities, cytokine secretion, gut morphology, short-chain fatty acids, and microbiota composition), diabetic specific parameters (blood glucose levels, blood biochemicals), body weight, and pulmonary inflammation.

The results of most of these studies showed some difference between A1- and A2-type β-CNs and respective milks [54, 59-61]. Many of the findings are focused on possible negative effects of A1-type β-CNs, but one study reported increased gut health after consumption of A2 milk. The negative effects of A1-type β-CNs reported in these studies are that it *may* cause progression in type 1 diabetes (takes generations to manifest), is more likely to promote the formation of fatty plaques in the arteries and

could have a proinflammatory effect on the lungs. However, many experiments concluded that there was no difference between being fed A1- or A2-type β-CNs. Furthermore, though animal studies are a good source of information, the models do not directly correlate or reliably predict the outcome in a human [55-58]. That raises a need to conduct human trials.

Focusing on human food clinical trials, studies assess a wide range of claims regarding preferential A2 protein (digestive health effects, muscle soreness, cardiovascular health, type 1 diabetes, glutathione levels) [24, 62-71]. A pattern throughout most of these studies is that they are organized as randomized and double-blind studies that involve some sort of washout period where subjects do not consume any kind of dairy. It should be noted that many of these studies have a small sample size. Six studies focused on testing digestive health effects testing and one study each focused on muscle soreness, cardiovascular health effects, and type 1 diabetes. Of the six studies conducted focusing on digestive health effects there were many different data points collected: bowel movements and respective details surrounding them, abdominal pain/discomfort, bloating, flatulence, diarrhea, PD3, gastrointestinal function (measured by smart pill), Subtle Cognitive Impairment Test (SCIT), serum/fecal laboratory biomarkers, adverse events, serum interleukin-4, immunoglobulins G, E, and G1, and BCM-7 coupled to lower glutathione levels [24, 62-66, 71].

The main results from these studies included no significant difference between the type of milk intake and abdominal pain, stool was softer (via Bristol Stool Scale), and bowel frequency was higher when A1 milk was consumed. Regarding these results, when differences were observed, it was variable between studies if they were statistically

significant or just higher. Studies that focused on t other topics (A2 milk relieving muscle soreness, A1/A2 milk causing cardiovascular health effects, and A1/A2 milk causing type 1 diabetes) all concluded that A2 milk did not cause the benefit or there was no health disadvantage over consumption of A1 casein. It should be noted that one study did conclude that A2 milk caused a higher amount of plasma glutathione levels causing the idea that it promotes greater antioxidant capacity [71]. Unlike the animal studies mentioned before, the results of these studies are variable in support of A2 milk being better than conventional milk, and many do not find correlational between A2 milk and the health positive or A1 milk and the health issue. This highlights the difficulty in applying data derived from animal studies to predicted human outcomes.

The only claim made in support of A2 milk that has multiple scientific studies reporting is that there are some adverse digestive health effects of A1-type β-CNs [47, 72]. Claims that attribute A1 type β-CNs to muscle soreness, negative cardiovascular health effects, and type 1 diabetes are not supported with scientific evidence and should *not* be associated with the A1- and A2-type β-CN debate, although they still can be observed in consumer-facing marketing material. The concept and consumer perception of A2 products is mixed [73]. Consumers do not have a complete understanding of the difference between normal A1/A2 and A2/A2 milk and many only recognize that it is a new product on the market. Though potential health benefits still need to be verified and consumers generally do not fully understand the product, the niche market continues to grow as A2 milk is being sold in 12,000+ stores in the U.S. at two to three times the cost of normal milk [29, 74].

VI. AUTHENTICATION PRACTICES

Foods are complex in makeup, and determining their composition requires specific and suitable analytical tools and methods. When foods are labeled with specific attributes such as genetic, species, or geographical origin (e.g., authentic Champagne can only be produced from the Champagne region in France), method of production (e.g., organic vs traditional), or processing technologies (e.g., freezing or irradiation), the necessity for proof of that claim is instantly created [7, 10]. There are many different reasons to authenticate milk in general. These reasons include but are not limited to high-value milk (sheep's and goat's) adulterated with low-value milk (cow's milk), presence of microbes in pasteurized milk, adulteration of high protein ingredients or chemicals (melamine) instead of milk protein itself, fresh milk with powdered milk, etc. Some popular techniques used specifically are polymerase chain reaction (PCR) for species origin authentication, chromatographic techniques used in tandem with MS (e.g., gas chromatography MS, liquid chromatography MS, and liquid chromatography time-of-flight MS) to identify substitution of high-quality ingredients with low quality ingredients in premium products, spectroscopic methodologies (e.g., mid-infrared (MIR) and near-infrared (NIR) spectroscopy) to identify melamine adulteration, and immunological techniques (e.g. ELISA) to discern food commodities identity such as meat, fish, and dairy [75]. Though other techniques are sufficient, MS has been being used more often in fields of food science, including food authentication, due to its advantages in high sensitivity, selectivity, throughput, and multi-analyte capabilities [76].

In the United States, generally, milk is collected in a pool refrigeration storage tank on a dairy farm until a milk tanker comes to load it up and transport the milk to a dairy processing facility. From there, the cream is separated from the milk and can be

turned into assorted products such as butter, cheese, and cream. The milk is pasteurized and packaged to be sold to supermarkets or to be used as an ingredient in another food product [44, 77-79]. In addition to pooling, the liquid milk can undergo different processes, one of which is spray drying. Spray drying is considered an industrial process that dehydrates liquids that contain solutes via small droplets exposed to hot air leaving a powder [80]. Injected into the spray drying system, milk is first concentrated and then flows though pipes and a series of chambers to be pushed through a nozzle and then dried and cooled to create the powder [43].

As of right now, the only authenticating system set up for A2 milk is genetic testing of cows *before* milk production [29]. With a pooling system set up, milk from a normal A1/A2 cow could cause an undefined amount of A1-type protein in the product. In addition, if a producer of A2 milk was mixing A1/A2 milk with A2/A2 milk to lower costs, there is no system set up to authenticate the product. This calls for a way of authenticating marketed A2 products themselves after genetic testing of cows. Regarding this food authenticity issue, analysis through liquid chromatography mass spectrometry (LC-MS), specifically intact mass spectrometry, may fit well because of the plethora of proteoforms and specificity of differences between A1- and A2-type proteoforms.

VII. PRINCIPLES OF INTACT MASS SPECTROMETRY

i. Mass Spectrometry (MS) Based Proteomics

The discipline of proteomics focuses on the in-depth analysis of proteins [81]. This includes a range of studies from protein structure and function to discovery and quantification of protein abundance. MS-based proteomics is a versatile and crucial tool used in many proteomic studies which can detect, identify, and quantify (dependent on

the method) from simple to complex protein samples [82]. There are a wide range of settings to which MS proteomic methods could be applied, such as pharmaceuticals, allergen quantification, pesticide residue detection, etc. It is also a great approach for food analysis and has been used for many different aspects of food such as analysis of flavonoids in soft drinks or wines, triacylglycerol and fatty acids in various food matrices, identification of unknown carbohydrates, analysis and quantification of trace ingredients like vitamins, etc. [10].

The major components of MS instruments are the ion source, mass analyzer, ion detector, and data processing electronics [83]. Polar, nonvolatile, and thermally unstable, proteins and peptides need to be softly ionized and transferred into the gas phase for MS-based proteomics [82]. Soft ionization means that through ionization the sample does not extensively degrade. There are two techniques used for ionization in MS-based proteomics: matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) [82, 83]. MALDI requires samples to be mixed with a crystalizing matrix before a laser is used to ionize the solid matrix. On the other hand, ESI produces ions from a solution. This occurs via high voltages while being sprayed and a high temperature to desolvate the sample into the gas phase which then enters the instrument. MALDI primarily generated ions that are singly charged while ESI produces a range of charges.

After ionization, ions flow through the instrument to the mass analyzer. Mass analyzers have the capability to store and separate ions based on the mass-to-charge (m/z) ratios. There are diverse types of mass analyzers, and each differs in properties such as speed, resolution, sensitivity, and mass range. These include ion traps, time of flight

(TOF), quadruple, Orbitrap, and Fourier transform ion cyclotron [82]. In addition, they can be combined to improve capability depending on the objective of analysis. From there, an ion detector reads the number of ions at each m/z value. Then the data processing electronics display the spectra of those m/z values.

Focusing on ESI, even before the sample is run on the instrument, there is usually a separation step to simplify the sample. This is especially beneficial for complex samples that contain a mixture of different proteins to separate before being analyzed on the instrument. In addition to simplifying the sample, it helps detect low-abundance signal that could be overshadowed by a higher abundant signal. High-performance liquid chromatography (HPLC) is commonly used for this separation step and continuously separates (usually paired with a continuous ion source like ESI). The most common HPLC chromatographic materials used are ion exchange, size exclusion, reverse phase, hydrophilic interaction chromatography, and affinity. The column chosen is reliant on the sample chosen and application. From there, after sample separation, it will be injected onto the MS as described previously [82, 83].

ii. Types of Proteomic MS Approaches

There are three major types of proteomic mass spectrometry methods: bottom up, middle down, and top down [82, 84, 85]. Bottom up MS has traditionally been used for complex proteome analysis, but middle and top down MS can be used depending on the sample being analyzed and the goal of the project. Bottom up MS uses proteolytic digestion to cleave proteins into peptides (usually 6 to 20 amino acids) and analyze the peptides from the sample. Top down MS is the opposite of bottom up MS as it analyzes whole proteins without any proteolytic digestion. Middle down MS is the method in-

between bottom up MS and top down MS, as it analyzes partial protein digestion (longer peptides than bottom up MS but not whole proteins) (Figure 1.7).

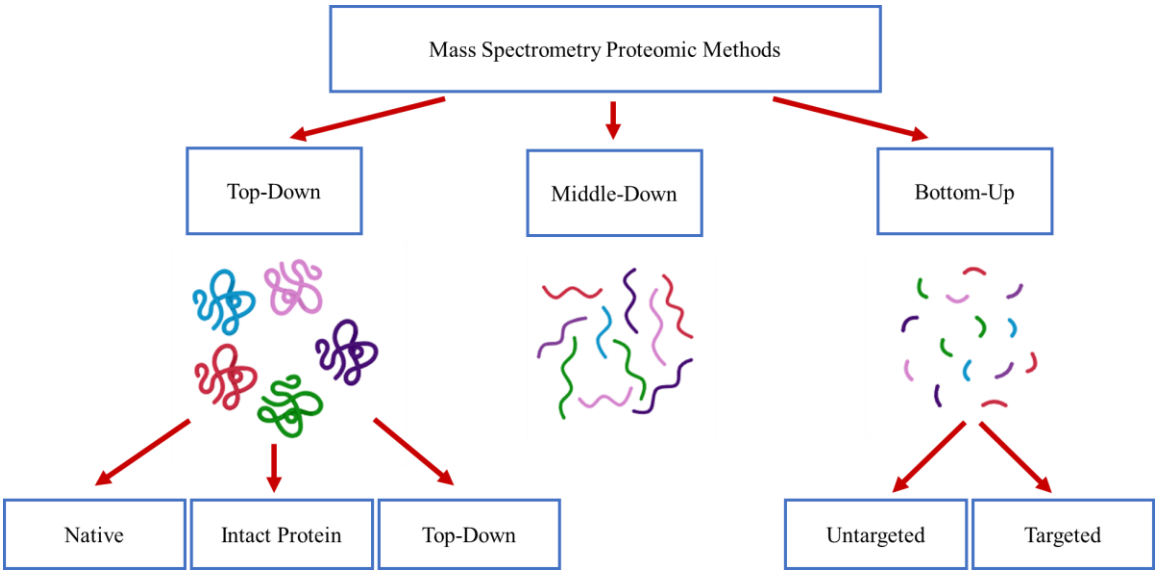


Figure 1.7: Three Main Mass Spectrometry Proteomic Analysis Methods. Top-Down, middle-down, and bottom-up and further subdivided analysis methods. Images provided via BioRender.com. Information obtained from Yates et al., 2009, Moradian et al., 2013, and Schubert et al., 2017.

There are a few subsets of bottom up MS and top down MS. For bottom up MS, the two main groups are untargeted and targeted. Untargeted methods, also known as discovery, look for a wide range of peptides in the sample thus discovering the contents of the sample[86]. For targeted methods, analysis is on predetermined peptides. For top down MS, there are native, intact protein, and top down methods[87]. Native methods require gentle conditions and sample preparation to preserve covalent post translational modifications and protein structures. Intact protein methods use low concentrations of acid, such as formic acid, and heat, which denature proteins dependent on the sample preparation and instrument parameters. Top down methods are very similar to intact protein, but the instrument fragments the ions to be analyzed [82, 85, 87]. Untargeted bottom up MS and intact protein MS will be focused on as middle down MS is not used

frequently, generally for very specific applications that require larger peptides with bottom up MS and smaller than top down MS [88, 89]. Specific examples for middle down MS are chromatin biology and protein deamination characterization.

In terms of sample preparation, bottom up MS and intact MS are very different as well. For bottom up MS, samples are extracted, reduced and alkylated, enzymatically digested, and then cleaned-up and desalted [83, 85]. Trypsin is a popular enzyme because it consistently cleaves at the C-terminal end of arginine and lysine, except after proline. In addition to reliability, the specific amino acids which it cleaves can accept protons, which allows peptides to be easily ionized [83]. For intact MS, there is no enzymatic digestion needed to be done [82, 87]. The main components needed for intact MS are filtering the sample to the protein(s) of interest, dispersing the sample into an MS-friendly solvent, and cleaning up the sample usually by liquid chromatography making it a quality option for sample analysis.

iii. Intact Protein MS

Some aspects to take into consideration for intact protein MS are sample cleanup and storage, method of delivering the sample to the MS (liquid chromatography or direct injection), and a standard that could be used to troubleshoot experimental issues [90]. When deciding sample storage conditions, it is critical to account for the fact that intact proteins could be hard to store. Avoiding freeze thaw cycles, resuspending sample (if a solid) immediately before use, and proper storage temperature dependent on state of sample (-80ºC for solid and 4ºC for liquid) are essential practices. In cleaning up samples, different columns can be used to de-salt, buffer exchange, extract, precipitate, etc. When

working with complex mixtures, chromatographic separation is needed while purified samples do not necessarily need chromatographic separation [82, 87, 90].

As stated before, the main difference between the conventional bottom up MS and intact MS is that bottom up MS requires proteolytic digestion, while intact MS does not [82, 83, 87]. Although the bottom up MS approach provides invaluable information, there are limitations: (1) multi-step preparation (extraction, reduction, alkylation, digestion, and clean up) can introduce modification in the resulting peptides and is time-consuming, (2) a limited number of peptides dependent on sites of protease action with differing ionization characteristics are produced, which may exclude modifications, and (3) proteins with multiple isoforms require unique isoform peptides for quantification, which can be poor targets[87, 91, 92]. Due to these limitations, a proteoform could be missed, regions of a protein could be missed leading to lack of information on PTMs, sequence variants, and modifications of the protein. This causes a possible gap in knowledge of a sample that creates a use for intact MS.

iv. Benefits of Intact Protein MS

Some examples of use of intact MS in other fields currently are pharmaceutical biologic drugs, i.e., immunoglobulin production, purified proteins, and now with complicated matrices such as food. Intact MS offers a potential solution to the limitations of bottom up methods: (1) minimal sample preparation is required, reducing potential modifications and time of assay, (2) the entire protein is being analyzed instead of reconstructing enzymatically derived peptides, (3) though milk proteoforms could lead to an abundance of similar masses, those multiple differences are easily seen in changes of

mass[91, 93]. With the use of intact MS information regarding PTMs, sequence variants, and modifications of the whole protein are not missed.

v. Limitations of Intact Protein MS

One difficulty with intact MS is that numerous proteoforms can coelute and cause the chromatographic protein peaks to broaden and split. This creates peak overlap, making it difficult to analyze single proteoforms. However, sliding window deconvolution analysis and addition of LC improves this distinguishing step [94, 95]. The sliding windows algorithm works by analyzing small windows across the peak instead of the entire averaging across the spectrum (Figure 1.8). This allows lower abundant peaks to be identified instead of being overwhelmed but the most abundant peaks. The main limitation of intact MS is the quality of the database on the protein(s) of interest including the information about sequence, genetic variants, post translational modifications, etc. To utilize intact MS, proteins and their proteoforms must be well characterized to correctly identify the mass deconvoluted from the raw spectra.

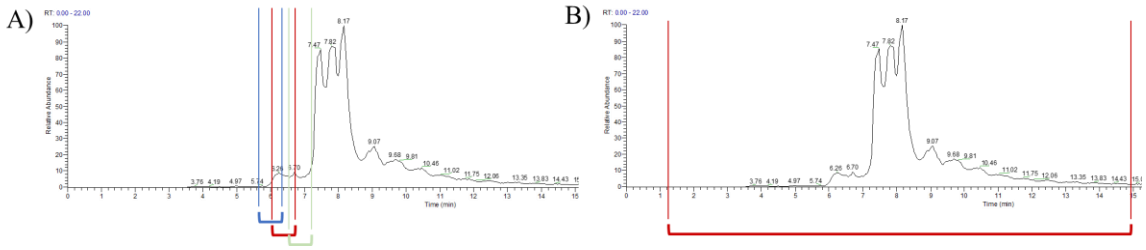


Figure 1.8: Comparison of Two Deconvolution Algorithms Available in BioPharma Finder Software.
A) Sliding window deconvolution algorithm **B)** Average spectrum deconvolution. The sliding window algorithm deconvolutes across the entire peak and only three windows are shown. Information obtained from ThermoFischer Scientific.

VIII. SUMMARY

Cow's milk is a paramount animal-derived product that takes many forms either as a product itself (e.g., liquid milk) or an ingredient in a food matrix (e.g., butter). It is important to regulate due to its popularity and versatile use in many foods. Specifically, the new premium labeled A2 milk and products should be authenticated due to health claims attached to the protein profile. The current establishment of A2 milk and products is conducted through genetic testing of cows *before* milk production, with no validated methods to authenticate the A2 milk products themselves. This gap in authentication approaches opens a need for a robust method to identify and detect the levels of A1- and A2-type β-CNs of the final product itself. To identify and distinguish the many proteoforms included in bovine milk and β-CNs especially, intact MS has the potential to fill the regulatory need though distinguishing a diverse of proteoforms in bovine milk products with varied processing.

IX. REFERENCES

1. Spink, J. and D.C. Moyer, *Defining the public health threat of food fraud.* Food Science, 2011. **76**(9): p. R157-R163.
2. Gökmen, V., *Importance of Food Authentication and Origin Testing. Food chemistry: X.* Food Chemistry: X, 2023. **18**(60): p. 100708.
3. Association, U.S.F.D. *How did the Federal Food, Drug, and Cosmetic Act come about?* 2018 03/28/2018; Available from: https://www.fda.gov/about-fda/fda-basics/how-did-federal-food-drug-and-cosmetic-act-come-about#:~:text=The%20enactment%20of%20the%201938,is%20still%20in%20force%20today.
4. Congress, U.S., *Federal Food, Drug, and Cosmetic Act*, in *21 USC 402*. 1938.
5. Congress, U.S., *Federal Food, Drug, and Cosmetic Act*, in *21 USC 403*. 1938.
6. Administration, U.S.F.D. *Taking time: Economically Motivated Adulteration* 2023 3/22/23; Available from: https://www.fda.gov/food/compliance-enforcement-food/economically-motivated-adulteration-food-fraud.
7. Danezis, G.P., A.S. Tsagkaris, F. Camin, V. Brusic, and C.A. Georgiou, *Food authentication: Techniques, trends & emerging approaches.* TrAC Trends in Analytical Chemistry, 2016. **85**: p. 123-132.
8. U.S. Congress, *Food Safety Modernization Act*. 2011.
9. Fox, P., *Milk: An overview*. 2008. p. 1-54.
10. Di Stefano, V., G. Avellone, D. Bongiorno, V. Cunsolo, V. Muccilli, S. Sforza, A. Dossena, L. Drahos, and K. Vékey, *Applications of liquid chromatography–mass spectrometry for food analysis.* Journal of Chromatography A, 2012. **1259**(2012): p. 74-85.
11. Poonia, A., A. Jha, R. Sharma, H.B. Singh, A.K. Rai, and N. Sharma, *Detection of adulteration in milk: A review.* International Journal of Dairy Technology, 2016. **70**(1): p. 23-42.
12. Ashurst, P.R. and M.J. Dennis, *Food Authentication*. An introduction to food authentication, ed. P.R. Ashurst, Dennis, M.J. 1996, Springer, Boston, MA.

13. Varnam, A.H. and J.P. Sutherland, *MILK and Milk Products Technology, chemistry and microbiology*. 1 ed. Food Product Vol. 1. 1994: Chapman and Hall. 451.
14. Wolf, C.A., T. Trey Malone, and B.R. McFadden, *Beverage milk consumption patterns in the United States: Who is substituting from dairy to plant-based beverages?* Journal of Dairy Science, 2020. **103**(12): p. 11209-11217.
15. Courtney Bir, N.O.W., Christopher Wolf, Michael S. Delgado, *Traditional attributes moo-ve over for some consumer segments: Relative ranking of fluid milk attributes.* Appetite, 2019. **134**(1).
16. National Agricultural Statistics Service, Agricultural Statistics Board, and U.S.D.o. Agriculture, *Milk Production (December 2022)*, N.A.S.S. USDA, Editor. 2022.
17. Prevention, C.f.D.C.a. *Breastfeeding among U.S. children born 2009–2016*. 2019 [cited 2020; CDC National Immunization Survey]. Available from: https://www.cdc.gov/breastfeeding/data/nis_data/results.html#print.
18. Victor Oliveira, M.P., David Smallwood, and Elizabeth Frazão, *WIC and the Retail Price of Infant Formula*. 2004. p. 106.
19. B. Jaya, C. Indranil, and D. Roshan, *U.S. Baby Infant Formula Market by Product Type (Infant Milk, Follow-On Milk, Specialty Baby Milk, and Growing-Up Milk), Ingredients (Carbohydrate, Fat, Protein, Minerals, Vitamins, and Others), Distribution Channel (Hypermarkets, Supermarkets, Pharmacy/Medical Stores, Specialty Stores, Hard Discounter Stores and Others) and Source (Regular, Organic, Milk with HMO, Goat Milk, Plant Based Formula, and Others), Opportunity Analysis and Industry Forecast, 2023-2032*. 2023, Allied Market Research. p. 176.
20. Bobe, G., D.C. Beitz, A.E. Freeman, and G.L. Lindberg, *Separation and Quantification of Bovine Milk Proteins by Reversed-Phase High-Performance Liquid Chromatography*. Journal of Agricultural and Food Chemistry, 1998. **46**(2): p. 458-463.

21. Eigel, W.N., J.E. Butler, C.A. Ernstrom, H.M. Farrell, V.R. Harwalkar, R. Jenness, and R.M. Whitney, *Nomenclature of Proteins of Cow's Milk: Fifth Revision1*. Journal of Dairy Science, 1984. **67**(8): p. 1599-1631.
22. Vincent, D., A. Elkins, M.R. Condina, V. Ezernieks, and S. Rochfort, *Quantitation and Identification of Intact Major Milk Proteins for High-Throughput LC-ESI-Q-TOF MS Analyses.* PLOS ONE, 2016. **11**(10): p. e0163471.
23. Hristov, P., I. Mitkov, D. Sirakova, I. Mehandgiiski, and G. Radoslavov, *Measurement of Casein Micelle Size in Raw Dairy Cattle Milk by Dynamic Light Scattering*. Milk Proteins, ed. I. Gigli. 2016.
24. Jianqin, S., X. Leiming, X. Lu, G.W. Yelland, J. Ni, and A.J. Clarke, *Effects of milk containing only A2 beta casein versus milk containing both A1 and A2 beta casein proteins on gastrointestinal physiology, symptoms of discomfort, and cognitive behavior of people with self-reported intolerance to traditional cows' milk.* Nutrition Journal, 2016. **15**(1): p. 35.
25. Beavers, L. and B. Van Doormaal. *Beta Casein, A2 Milk and Genetics*. 2016; Available from: https://www.cdn.ca/images/uploaded/file/A2%20Milk%20Article%20-%20December%202016.pdf.
26. comapny, a.m. *The A2 protein difference*. 2023 [cited 2023; Available from: https://thea2milkcompany.com/thea2milkdifference.
27. Aschaffenburg, R., *Section G. Genetics. Genetic variants of milk proteins: Their breed distribution.* 3 ed. Journal of Dairy Research. Vol. 35. 1968, Cambridge, UK: Cambridge University Press.
28. Martin, P., L. Bianchi, C. Cebo, and G. Miranda *Genetic Polymorphism of Milk Proteins*. Advanced Dairy Chemistry, ed. P. McSweeney, Fox, P. 2013, Boston, MA: Springer.
29. Dantas, A., H. Kumar, E.S. Prudencio, L.B. de Avila, P. Orellana-Palma, N.S. Dosoky, E. Nepovimova, K. Kuča, N. Cruz-Martins, R. Verma, S. Manickam, M. Valko, and D. Kumar, *An approach on detection, quantification, technological*

properties, and trends market of A2 cow milk. Food Research International, 2023. **167**: p. 112690.

30. Adido, S.M., M.G. Gicheha, C. Adoligbe, and K. Atchou, *Genetic Diversity of Three Indigenous Cattle Breeds Reared in Benin.* American Journal of Animal and Veterinary Sciences, 2019. **14**(1).
31. Sebastiani, C., C. Arcangeli, M. Torricelli, C. M., N. D'avino, G. Cinti, S. Fisichella, and M. Biagetti, *Frequencies Evaluation of β-Casein Gene Polymorphisms in Dairy Cows Reared in Central Italy.* MDPI Animals, 2020. **10**(252): p. 2-7.
32. Agriculture, E.R.S.U.S.D.o. *Background*. Dairy 2023 [cited 2023; Available from: https://www.ers.usda.gov/topics/animal-products/dairy/background/.
33. a2 Milk Company. *Our Cows and Farms*. 2022; Available from: https://a2milk.com/our-cows-and-farms.
34. Farm, A.F. *Crossbreeding for Grazing and A2/A2 Genetics*. 2023; Available from: https://alexandrefamilyfarm.com/pages/a2-a2-cows.
35. Farrell, H.M., R. Jimenez-Flores, G.T. Bleck, E.M. Brown, J.E. Butler, L.K. Creamer, C.L. Hicks, C.M. Hollar, K.F. Ng-Kwai-Hang, and H.E. Swaisgood, *Nomenclature of the Proteins of Cows' Milk—Sixth Revision.* Journal of Dairy Science, 2004. **87**(6): p. 1641-1674.
36. Mercier, J.C. and P. Gaye, *Study of Secretory Lactoproteins: Primary Structures of the Signals and Ezymatic Processing.* Annals of the New York Academy of Sciences, 1980. **343**(1): p. 232-251.
37. Dupont, D., T. Croguennec, and S. Pochet, *Milk Proteins - Analytical Methods.* Reference Module in Food Science, 2018: p. 1-15.
38. Raghav, A. and J. Ahmad, *Fundamentals of Glycosylation*, ed. A.R.a.J. Ahmad. 2022, London, United Kingdom: IntechOpe. 87.
39. Huang, L.-J., J.-H. Lin, J.-H. Tsai, Y.-Y. Chu, Y.-W. Chen, S.-L. Chen, and S.-H. Chen, *Identification of protein O-glycosylation site and corresponding glycans using liquid chromatography–tandem mass spectrometry via mapping accurate mass and retention time shift.* Journal of Chromatography A, 2014. **1371**: p. 136-145.

40. Pisano, A., N.H. Packer, J.W. Redmond, K.L. Williams, and A.A. Gooley, *Characterization of O-linked glycosylation motifs in the glycopeptide domain of bovine κ-casein.* Glycobiology, 1994. **4**(6): p. 837-844.
41. Raghav, A. and J. Ahmad, *Introductory Chapter: Glimpses of Glycosylation*. Fundamentals of Glycosylation, ed. A.R.a.J. Ahmad. 2022.
42. Minkiewicz, P., C.J. Slangen, F.M. Lagerwerf, J. Haverkamp, H.S. Rollema, and S. Visser, *Reversed-phase high-performance liquid chromatographic separation of bovine κ-casein macropeptide and characterization of isolated fractions.* Journal of Chromatography A, 1996. **743**(1): p. 123-135.
43. Lewis, M.J. and H.C. Deeth, *Heat Treatment of Milk*, in *Milk Processing and Quality Management*, A.Y. Tamine, Editor. 2008, Wiley-Blackwell. p. 168-204.
44. Schuck, P., *Effects of Drying on Milk Proteins*. 2 ed. Milk Proteins: From Expression to Food, ed. M.B. Harjinder Singh, Abby Thompson. 2014: Food Science and Technology.
45. Arena, S., G. Renzone, G. Novi, A. Paffetti, G. Bernardini, A. Santucci, and A. Scaloni, *Modern proteomic methodologies for the characterization of lactosylation protein targets in milk.* PROTEOMICS, 2010. **10**(19): p. 3414-3434.
46. Van Boekel, M.A.J.S., *Effect of heating on Maillard reactions in milk.* Food Chemistry, 1998. **62**(4): p. 403-414.
47. Küllenberg de Gaudry, D., S. Lohner, C. Schmucker, P. Kapp, E. Motschall, S. Hörrlein, C. Röger, and J.J. Meerpohl, *Milk A1 β-casein and health-related outcomes in humans: a systematic review.* Nutrition Reviews, 2019. **77**(5): p. 278-306.
48. Mora, L., M.C. Aristoy, and F. Toldrá, *Bioactive Peptides in Foods*, in *Encyclopedia of Food and Health*, B. Caballero, P.M. Finglas, and F. Toldrá, Editors. 2016, Academic Press: Oxford. p. 395-400.
49. Jinsmaa, Y. and M. Yoshikawa, *Enzymatic release of neocasomorphin and β-casomorphin from bovine β-casein.* Peptides, 1999. **20**(8): p. 957-962.
50. Pleurvy, B.J., *Opiod Recpetors and their ligands: natural and unnatural.* British Journal of Anaesthesia, 1991. **66**(3): p. 370-380.

51. Aslam, H., A. Ruusunen, B. Berk, A. Loughman, L. Rivera, J.A. Pasco, and F.N. Jacka, *Unravelled facets of milk derived opioid peptides: a focus on gut physiology, fractures and obesity.* International journal of food sciences and nutrition, 2020. **71**(1): p. 36–49.

52. Arısoy, S., I. Çoban, and Ö. Üstün-Aytekin, *Food-Derived Opioids: Production and the Effects of Opioids on Human Health*, in *From Conventional to Innovative Approaches for Pain Treatment*, C. Marco, Editor. 2019, IntechOpen: Rijeka. p. Ch. 11.

53. Teschemacher, H., *Opioid receptor ligands derived from food proteins.* Current pharmaceutical design, 2003. **9**(16): p. 1331–1344.

54. Barnett, M.P.G., W.C. McNabb, N.C. Roy, K.B. Woodford, and A.J. Clarke, *Dietary A1 β-casein affects gastrointestinal transit time, dipeptidyl peptidase-4 activity, and inflammatory status relative to A2 β-casein in Wistar rats.* International Journal of Food sciences and Nutrition, 2014. **65**(6): p. 720-727.

55. Kaminski, S., E. Kostyra, A. Cieslinska, and E. Fiedorowicz, *Consumption of bovine (3-casein variants (A1 or A2) does not af-fect basic hematological and biochemical indices.* Milchwissenschaft, 2011. **67**(3): p. 238-241.

56. Guantario, B., M. Giribaldi, C. Devirgiliis, A. Finamore, E.C.M.T. Capucchio, R. Evangelista, V. Motta, P. Zinno, S. Cirrincione, S. Antoniazzi, L. Cavallarin, and M. Roselli, *A Comprehensive Evaluation of the Impact of Bovine Milk Containing Different Beta-Casein Profiles on Gut Health of Ageing Mice.* Nutrients, 2020. **12**(7).

57. Beales, P.E., R.B. Elliott, S. Flohé, J.P. Hill, K. Kolb, P. Pozzilli, G.S. Wang, H. Wasmuth, and F.W. Scott, *A multi-centre, blinded international trial of the effect of A(1) and A(2) beta-casein variants on diabetes incidence in two rodent models of spontaneous Type I diabetes.* Diabetologia, 2002. **45**(9).

58. Thakur, N., G. Chauhan, B.P. Mishra, S.K. Mendiratta, A.K. Pattanaik, T.U. Singh, M. Karikalan, S.K. Meshram, and L. Garg, *Comparative evaluation of feeding effects of A1 and A2 cow milk derived casein hydrolysates in diabetic model of rats.* Journal of Functional Foods, 2020. **75**.

59. Chia, J.S.J., J.L. McRae, A.K. Enjapoori, C.M. Lefèvre, S. Kukuljan, and K.M. Dwyer, *Dietary Cows' Milk Protein A1 Beta-Casein Increases the Incidence of T1D in NOD Mice.* Nutrients, 2018. **10**(9): p. 1291.

60. Tailford, K.A., C.L. Berry, A.C. Thomas, and J.H. Campbell, *A casein variant in cow's milk is atherogenic.* Atherosclerosis, 2003. **170**(1): p. 13-19.

61. Shikha Yadav, N.D.S.Y., Atish Gheware , Ankur Kulshreshtha , Pankaj Sharma, V P Singh, *Oral Feeding of Cow Milk Containing A1 Variant of β Casein Induces Pulmonary Inflammation in Male Balb/c Mice.* Scientific Reports, 2020. **10**(1): p. 8053.

62. Crowley, E.T., L.T. Williams, T.K. Roberts, R.H. Dunstan, and P.D. Jones, *Does Milk Cause Constipation? A Crossover Dietary Trial.* Nutrients, 2013. **1**: p. 253-266.

63. Ho, S., K. Woodford, S. Kukuljan, and S. Pal, *Comparative effects of A1 versus A2 beta-casein on gastrointestinal measures: a blinded randomised cross-over pilot study.* European Clinical Journal of Nutrition, 2014. **68**(9): p. 994-1000.

64. He, M., J. Sun, Z.Q. Jiang, and Y.X. Yang, *Effects of cow's milk beta-casein variants on symptoms of milk intolerance in Chinese adults: a multicentre, randomised controlled study.* Nutrition Journal, 2017. **16**(1): p. 72.

65. Sheng, X., Z. Li, J. Ni, and G. Yelland, *Effects of Conventional Milk Versus Milk Containing Only A2 β-Casein on Digestion in Chinese Children: A Randomized Study.* J Pediatr Gastroenterol Nutr, 2019. **69**(3).

66. Ramakrishnan, M., T.K. Eaton, O.M. Sermet, and D.A. Savaiano, *Milk Containing A2 β-Casein ONLY, as a Single Meal, Causes Fewer Symptoms of Lactose Intolerance than Milk Containing A1 and A2 β-Caseins in Subjects with Lactose Maldigestion and Intolerance: A Randomized, Double-Blind, Crossover Trial.* Nutrients, 2020. **12**(12): p. 1-15.

67. Kirk, B., J. Mitchell, M. Jackson, F. Amirabdollahian, O. Alizadehkhaiyat, and T. Clifford, *A2 Milk Enhances Dynamic Muscle Function Following Repeated Sprint Exercise, a Possible Ergogenic Aid for A1-Protein Intolerant Athletes?* Nutrients, 2017. **9**(2): p. 94.

68. Chin-Dusting, J., J. Shennan, E. Jones, C. Williams, B. Kingwell, and A. Dart, *Effect of dietary supplementation with beta-casein A1 or A2 on markers of disease development in individuals at high risk of cardiovascular disease.* Britain Journal of Nutrition, 2006. **1**: p. 136-144.

69. Venn, B.J., C.M. Skeaff, R. Brown, J.I. Mann, and T.J. Green, *A comparison of the effects of A1 and A2 beta-casein protein variants on blood cholesterol concentrations in New Zealand adults.* Atherosclerosis, 2006. **1**: p. 175-178.

70. Thorsdottir , I., B.E. Birgisdottir, I.M. Johannsdottir, D.P. Harris, J. Hill, L. Steingrimsdottir, and A.V. Thorsson, *Different beta-casein fractions in Icelandic versus Scandinavian cow's milk may influence diabetogenicity of cow's milk in infancy and explain low incidence of insulin-dependent diabetes mellitus in Iceland.* Pediatrics, 2000. **106**(4): p. 719-724.

71. Deth, R., A. Clarke, J. Ni, and M. Trivedi, *Clinical evaluation of glutathione concentrations after consumption of milk containing different subtypes of β-casein: results from a randomized, cross-over clinical trial.* Nutrition Journal, 2015. **15**(82).

72. Brooke-Taylor, S., K. Dwyer, K. Woodford, and N. Kost, *Systematic Review of the Gastrointestinal Effects of A1 Compared with A2 β-Casein.* Advances in Nutrition, 2017. **8**(5): p. 739-748.

73. Nogueira, A.C.R., K.A. Amorim, S.C. Bastos, A.C.M. Pinheiro, R.B. Teixeira, A.G. da Cruz, and J.F. Rodrigues, *Information influenced consumers' attitudes toward A2 milk: An approach using text highlighting technique.* Journal of Sensory Studies. **n/a**(n/a): p. e12863.

74. Detz, J. *What Is a2 Milk?* 2019 03/22/23]; Available from: https://www.consumerreports.org/milk/what-is-a2-milk/.

75. Balabin, R.M. and S.V. Smirnow, *Melamine detection by mid- and near-infrared (MIR/NIR) spectroscopy: a quick and sensitive method for dairy products analysis including liquid milk, infant formula, and milk powder.* Talanta, 2011. **85**(1): p. 562-568.

76. Cubero-Leon, E., R. Peñalver, and A. Maquet, *Review on metabolomics for food authentication.* Food Research Journal, 2014. **60**: p. 95-107.

77. Administration, U.S.F.D. *DAIRY PRODUCT MANUFACTURERS-ATTACHMENT 1-4*. 2014 [cited 2023 9/11/23]; Available from: https://www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/inspection-guides/dairy-product-manufacturers-attachment-1-4.

78. Park, Y., . *Park, Y.W. 2014. Safety of goat milk products: Extension education delivery tools for dairy goat producers: A web-based e-book: Chapter 10. Pages 243-263*. 2016. p. 243-263.

79. Niamh, B., A.Z. Krzysztof, S. Mark, H. Paul, P.R. Michael, and C.A. Catherine, *The Dairy Industry: Process, Monitoring, Standards, and Quality*, in *Descriptive Food Science*, D. Antonio Valero and G.-G. Rosa María, Editors. 2018, IntechOpen: Rijeka. p. Ch. 1.

80. Pisecky, J., *Handbook of Milk Powder Manufacture*. 1997, Gladsaxevej 305, DK-2860 Soeborg, Copenhagen, Denmark: GEA Process Engineering A/S.

81. Carrera, M., *Proteomics and Food Analysis: Principles, Techniques, and Applications.* MPDI Foods, 2021. **10**.

82. Yates, J.R., C.I. Ruse, and A. Nakorchevsky, *Proteomics by Mass Spectrometry: Approaches, Advances, and Applications.* Annual Review of Biomedical Engineering, 2009. **11**: p. 49-79.

83. Aebersold, R. and M. Mann, *Mass spectrometry-based proteomics.* Nature, 2003. **422**: p. 198-207.

84. Moradian, A., A. Kalli, M.J. Sweredoski, and S. Hess, *The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications.* Proteomics, 2013. **14**(4-5): p. 489-497.

85. Schubert, O.T., H.L. Röst, B.C. Collins, G. Rosenberger, and R. Aebersold, *Quantitative proteomics: challenges and opportunities in basic and applied research.* Nature Protocols, 2017. **12**: p. 1289–1294.

86. Manes, N.P. and A. Nita-Lazar, *Application of targeted mass spectrometry in bottom-up proteomics for systems biology research.* Proteomics, 2018. **189**: p. 75-90.

87. Catherman, A.D., O.S. Skinner, and N.L. Kelleher, *Top Down proteomics: Facts and perspectives.* Biochemical and Biophysical Research Communnications, 2014. **445**(4): p. 683-693.

88. Sidoli, S. and B.A. Garcia, *Middle-down proteomics: a still unexploited resource for chromatin biology.* Expert review of proteomics, 2017. **14**(7): p. 617–626.

89. Jin, Y., Y. Yi, and B. Yeung, *Mass spectrometric analysis of protein deamidation – A focus on top-down and middle-down mass spectrometry.* Methods, 2022. **200**: p. 58-66.

90. Scientific, T., *Protein Clean Up - Sample Handling*. p. 1-6.

91. Bush, D.R., L. Zang, A.M. Belov, A.R. Ivanov, and B.L. Karger, *High Resolution CZE-MS Quantitative Characterization of Intact Biopharmaceutical Proteins: Proteoforms of Interferon-β1.* Analytical Chemistry, 2016. **88**(2): p. 1138-1146.

92. Chen, J., P. Shiyanov, and K.B. Green, *Top-down mass spectrometry of intact phosphorylated β-casein: Correlation between the precursor charge state and internal fragments.* Journal of Mass Spectrometry, 2019. **54**(6): p. 527-539.

93. Johnson, P., M. Philo, A. Watson, and E.N. Clare Mills, *Rapid Fingerprinting of Milk Thermal Processing History by Intact Protein Mass Spectrometry with Nondenaturing Chromatography*. Journal of Agricultural and Food Chemistry, 2011. **59**(23): p. 12420-12427.

94. Feng, P., C. Fuerer, and A. McMahon, *Quantification of Whey Protein Content in Infant Formulas by Sodium Dodecyl Sulfate-Capillary Gel Electrophoresis (SDS-CGE): Single-Laboratory Validation, First Action 2016.15.* Journal of AOAC INTERNATIONAL, 2017. **100**(2): p. 510-521.

95. Scientific, T., *BioPharma Finder User Guide, software version 3.0, XCALI-97898 Revision A*. 2017.

CHAPTER II: DEVELOPMENT OF CONDITIONS OPTIMAL FOR INTACT MASS SPECTROMETRY

I. ABSTRACT

Products with a particular protein source or composition declared on label pose a challenge to authenticate when using traditional analytical methods that generally do not have the required specificity. For example, the establishment of milk as “A2” is achieved through genetic testing of cows before milk production, with no methods to authenticate milk products themselves. This work aimed to develop an intact protein mass spectrometry (MS) method to analyze major bovine milk proteins in powdered and high-temperature, short-time (HTST) milk for direct authentication of bovine milk protein products, including specific proteoform claims. This was attained through three major phases: generation of a predicted mass database, optimization of sample preparation and instrument parameters conducive with intact bovine milk proteins, and the selection of deconvolution software for protein identification with a mass error tolerance (10 ppm). Within the selected 15 commercial products (nonfat dry milk, whole milk powder, infant formulas, and HTST milk), a mean of 85.27% (± 6.68%, n = 57) of the total signal of could be assigned to the predicted monoisotopic mass database proteoforms using the finalized method.

II. INTRODUCTION

Currently, the authentication of products with claims regarding protein sources or compositions is a challenge for traditional analytical methods, which generally lack the required specificity whole protein analysis can provide. For example, A2 milk, produced by cows with the genotype A2/A2, only contains the A2-type variant of β-casein instead of the conventional mix of A1- and A2-type variants. This protein profile has reported health benefits and thus increased market value [1-3]. A2 milk is established through genetic testing of the cows before milk production, but with no methods to authenticate the milk products themselves [4]. Intact protein mass spectrometry (MS) has the potential to directly authenticate protein products, including specific proteoform claims. Presently, there is no established regulatory definition for the exact protein profile of A2 milk because there are no methods to specifically distinguish the two β-casein variants (A1- and A2-type). Bovine milk proteins have many genetic variants and are variably modified, resulting in a multitude of proteoforms. The development of an intact protein MS method to detect and differentiate major bovine milk proteins (α_{S1}-, α_{S2}-, β-, κ-caseins, β-lactoglobulins, and α-lactalbumins) and their proteoforms is needed to confirm A2 protein profile claims and help regulators define A2 milk. This work will demonstrate that intact protein MS can effectively analyze milk products for protein authentication.

III. MATERIALS AND METHODS

To develop the finalized intact protein MS method, the generation of a predicted mass database for intact bovine proteoforms was first completed using UniProt and NCBI databases. From there, the development and optimization of the method was completed and could be grouped into two steps starting with direct infusion MS and then liquid

chromatography (LC) MS. Subsets of direct infusion MS focused on sample preparation followed by instrument parameter optimization before the transition to use of LC (Figure 2.1). Acetonitrile (ACN), water, and formic acid (FA) used were Optima™ LC/MS Grade (Fisher Chemical, Thermo Scientific).

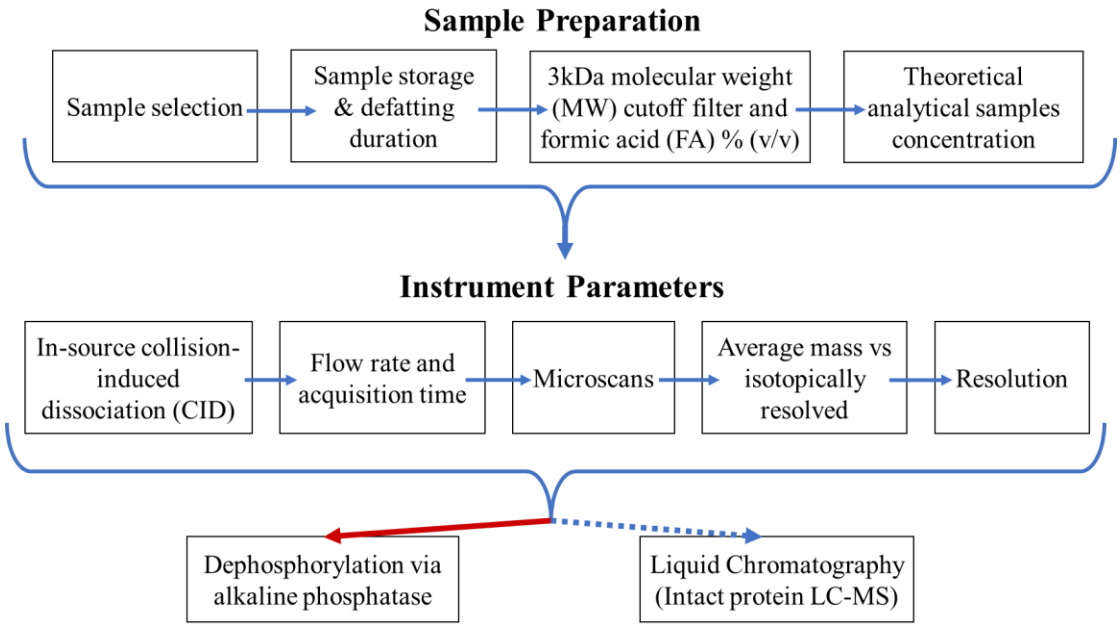


Figure 2.1: Flow Chart of Pilot Experiments to Pptimize Direct Infusion Intact Protein MS Method.
Each box was completed before moving on to the tier below. Blue arrows shows the steps that were successful in determining a parameter while red arrows show a path that was unsuccessful.

a. Sequence Database Selection

The major proteins found in milk are α_{S1}-, α_{S2}-, β-, κ-caseins (CN), β-lactoglobulins (BLG), α-lactalbumins (α-lac), bovine serum albumin (BSA), and immunoglobulins (Ig). First, information on variability, abundance, and size were gathered to select the target proteins for the method. This was cross referenced with the capabilities of the instrument to narrow down, if necessary, the list of proteins. From there, sequence databases, UniProt and NCBI were used to compile all known sequences

of selected proteins. Beginning with the UniProt database, sequences were collated including those of genetic variants for each protein. Then, a UniProt provided service, “Retrieve/ID mapping”, was used to produce a list of NCBI sequence accession numbers that were similar to the sequences collected from UniProt [5]. This served to expand the database for additional variants and proteoforms. Alignments were created using Clustal Omega to distinguish unique proteoforms between the sequences found within UniProt and NCBI. Sequences that did not have the entire sequence characterized (noted with an “X” in the sequence) were removed. From there, sequences from UniProt and NCBI were combined to create the primary sequences in the database. Monoisotopic and average masses were assigned to each primary sequence using the program mMass [6]. Information regarding modifications for each selected protein was considered using UniProt and previous research and literature [5, 7-45]. Each feasible modification was reviewed for each proteoform, and monoisotopic and average masses were generated and noted to construct the final database of predicted masses and the proteoform attributed to.

b. Sample Sourcing

Commercial products both containing a normal profile (both A1- and A2-type β-CNs) and marketed A2-type β-CN profiles were considered. The major products found using A2 labeling were HTST liquid milks, powdered milks (nonfat dry milk (NFDM) and whole milk powder (WMP), and infant formulas (Table 2.1). A general search of HTST liquid milks that could be locally sourced and from where was collected as those samples would need to be continually available to allow repeated purchase as necessary. Powdered milks and infant formulas could be sourced non-locally due to their longer shelf stability. Where possible, both normal and A2 products from the same

manufacturers were obtained. In addition, lyophilized purified bovine proteins from Sigma-Aldrich® (α-, β-, κ-CN, BLG, and α-lac) were used as single-protein controls. Samples were chosen based on consistent availability locally or through reputable online sources, with an equal distribution of samples within the categories in mind (product types and mixed and non-mixed profiles).

Table 2.1: Selected Commercial Samples of Normal and A2-type β-CN Protein Profiles

Product Type	Mixed A1- & A2-type β-CNs	Abbreviation	Marketed as " Only A2-type β-CNs"	Abbreviation
Powdered Milk	Nestlé Carnation Instant Nonfat Dry Milk	Nestlé NFDM	a2 Instant Milk Powder Full Cream 1 kg (Australia Import)	a2 company A2 WMP
	Hy-Vee Instant Nonfat Dry Milk	Hy-Vee NFDM	Azure Market A2/A2 Organic Dry Milk Powder	Azure Market A2 WMP
Powdered Infant Formula	Pure Bliss® by Similac® Infant Formula	Similac Formula	Similac Organic A2 Infant Formula Powder Toddler	SimilacA2 Formula
	Gerber Good Start Baby Formula Powder, GentlePro	Gerber Formula	Gerber Good Start Infant Formula with Iron, A2, 0 to 12 Months - 20 oz	Gerber A2 Formula
	Enfamil Infant formula	Enfamil Formula	Enfamil A2 Premium Infant Formula, Milk-Based Powder with Iron - 19.5 oz	Enfamil A2 Formula
			Serenity 12+ months A2 whole milk toddler formula	Serenity A2 Formula
HTST Milk	Hiland Vitamin D Milk	Hiland HTST milk	Whole milk by "The a2 milk company"	a2 company A2 HTST milk
	Hy-Vee Vitamin D Milk	Hy-Vee HTST milk	A2/a2 Organic 4% Whole Milk by Alexandre ECO DAIRY	Alexandre A2 HTST milk

c. Direct Infusion MS

vi. Direct Infusion MS Sample Preparation Optimization

To develop conditions suitable for intact protein MS, the direct infusion method was chosen initially for its simplicity and minimal amount of sample preparation before analysis. This method required a sample preparation in which the sample was cleaned up and resuspended in a solvent compatible with the instrument as there is no separation step before analysis. This sample preparation optimization was split into three main steps: sample storage and defatting conditions, desalting and clean up conditions, and protein loading optimization.

1. Sample Storage and Defatting Conditions

In duplicate, 1 mL of normal samples, Nestlé NFDM and Horizon ultra-high-temperature processing (UHT) milk, were stored at -20°C, room temperature (RT), and 37°C for 10 minutes, 30 minutes, and overnight (o/n). Samples were then centrifuged at 3,000 x g for 0, 15, and 30 minutes. Pictures of sample fat layer thickness produced were documented and compared to decide which combinations to test further. After completion, a new set of 1 mL Nestlé NFDM and Horizon UHT milk were stored at room temperature, frozen at -20ºC for 15 minutes, or frozen at -20°C o/n and were centrifuged for 10 and 15 minutes. A total of six combinations were evaluated using NuPAGE™ 4-12% Bis-Tris Gels (1.00 mm x 12 well) (Invitrogen™, Thermo Scientific™) to determine the optimal combination that ensured protein remained in solution and enough fat was removed from the sample.

The sample loaded into each well consisted of 4 µL defatted sample, 5 µL 4X NuPAGE LDS Buffer, 10 µL water, and 1 µL of ß-mercaptoethanol (BME) (Sigma-Aldrich) to reduce the sample (95°C for 5 minutes). Out of the 20 µL of sample prepared, 15 µL of each sample was loaded into each well. NuPAGE SDS Running Buffer was used in a Mini-Cell Electrophoresis Chamber (Invitrogen, Thermo Scientific). The gel was run for 35 minutes at a constant voltage of 200 V. For each gel, the Precision Plus Protein™ Dual Xtra Standard was run (Bio-Rad Laboratories). Gels were stained with Coomassie Brilliant Blue R-250 Staining Solution for one hour and destained using Coomassie Brilliant Blue R-250 Destaining Solution for two to three hours (Bio-Rad Laboratories) while placed on a Rotomix 50800 Orbital Shaker (Thermolyme). The gel was rinsed with water and then imaged under a bright light table. Examining and

identifying bands of protein and possible smearing determined the combination of storage temperature and centrifugation time.

2. Desalting and Clean Up Conditions

To clean up and desalt the samples, Amicon® Ultra 0.5 mL Centrifugal Filters with a 3 kDa molecular weight (MW) cutoff were used. A cutoff of 3 kDa was selected to allow small molecules and salts to pass through while the larger proteins are stopped by the small filter pores. The use of the filters limited the amount of ACN the proteins could be resuspended into (up to 20%) and FA needed to be added after the filtering process as it was not compatible with the filters.

To determine if the initial resuspension buffer, use of MW spins columns, and/or buffer exchange in sample preparation could cause protein precipitation, Nestlé NFDM was selected for evaluation (Figure 2.2).

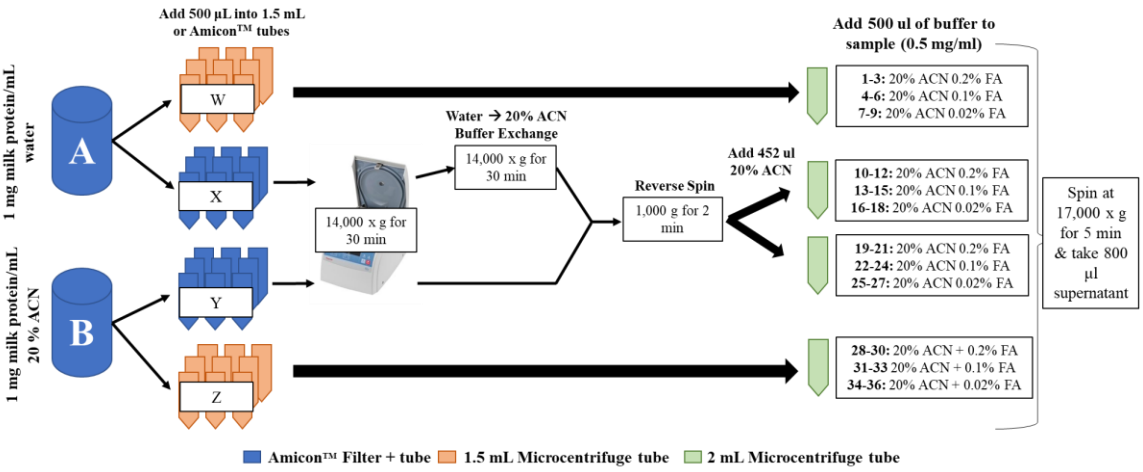


Figure 2.2: Part One Assessment of Sample Preparation for Protein Precipitation.

Nestlé NFDM was resolubilized separately in water and in 20% ACN (1 mg/mL). From there, the resolubilized samples were split into two groups: one group filtered through the Amicon® Ultra 0.5 mL Spin Columns while the other is directly diluted into

20% ACN. Run in triplicate, concentrations of 0.2%, 0.1%, and 0.02% (v/v) FA to achieve an end concentration of 0.1%, 0.05%, and 0.01% (v/v) FA were added to each sample. The resulting tubes were spun at 17,000 x g for five minutes to quickly sediment and remove protein not soluble in solution. From there, a total protein quantification of samples was completed using the Pierce™ BCA Protein Assay Kit according to the manufacturer's instruction. Each combination was run in duplicate and a standard curve of each solution was made. Evaluating the theoretical protein content of each solution and method, a finalized protocol was established. A second round of Nestlé NFDM (1 mg milk protein/mL water) was made without the defatting step and diluted using water, 40% ACN, 40% ACN 0.2% (v/v) FA, 40% ACN 0.1% (v/v) FA, 40% ACN 0.02% (v/v) FA to further assess buffer effects (Figure 2.3).

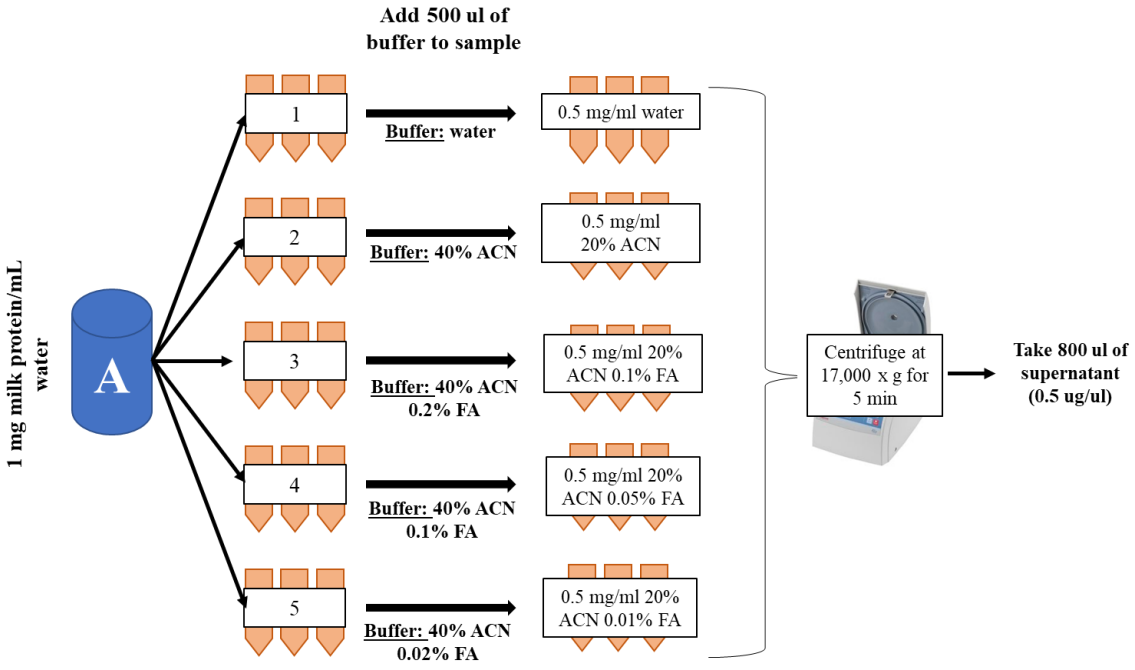


Figure 2.3: Part Two Assessment of Sample Preparation for Protein Precipitation.

To determine if end solution pH containing the final sample was near the isoelectric point of caseins (pH 4.6), BLG (pH 5.1), or α-lac (pH 4.2), pH strips were

used [46]. Each of the three different pH strip brands, Fishers, Baxter, and Color-pHast, were used to assess each of the solutions: 0.5 mg milk protein/ml water, 0.5 mg milk protein/ml 20% ACN 0.01% (v/v) FA, 0.5 mg milk protein/ml 20% ACN 0.05% (v/v) FA, and 0.5 mg milk protein/ml 20% ACN 0.1% (v/v) FA.

3. Evaluation of Finalized Sample Preparation Protocol

To assess the chosen defatting and clean up steps in the sample preparation method, each stage of preparation (before sample preparation, after defatting step, after Amicon® Spin Filter) was examined using Nestlé NFDM, Serenity A2 Formula, Alexandre A2 HTST milk, and Hiland HTST milk NuPAGE™ 4-12% Bis-Tris Gels (1.00 mm x 12 well) (Figure 2.4).

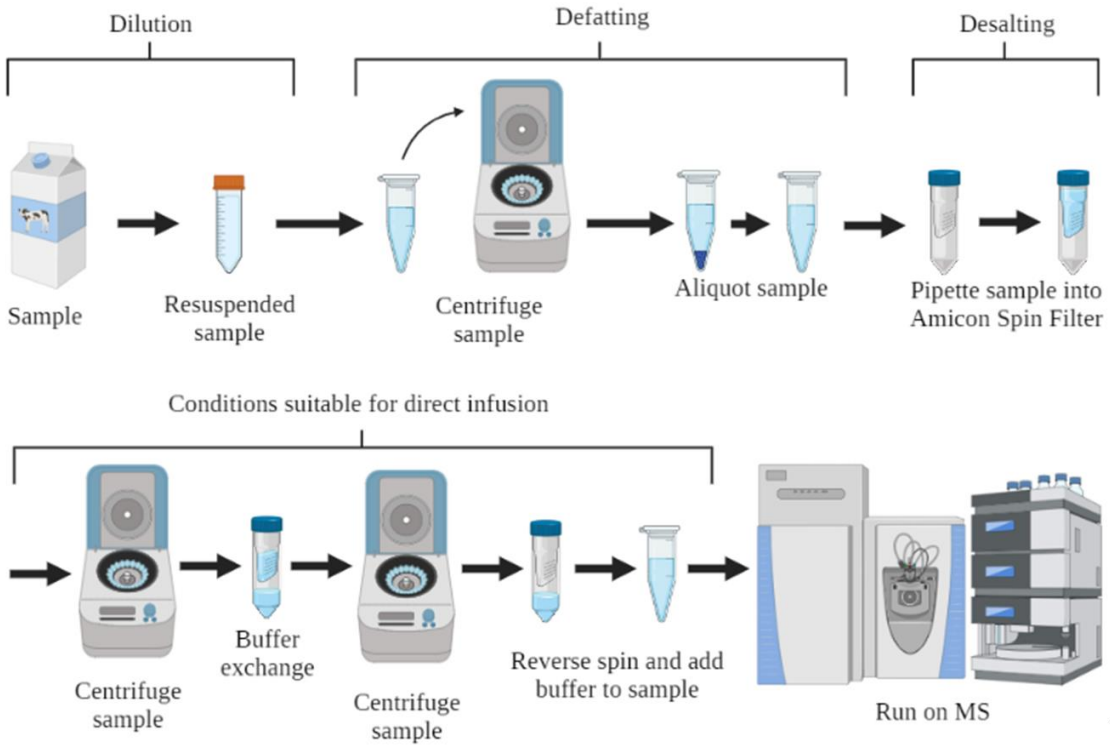


Figure 2.4: Direct Infusion Intact Protein MS Analysis Sample Preparation Summary. Created in BioRender.com.

NuPAGE™ 4-12% Bis-Tris Gels were used to ensure there was no significant loss of protein at each step, especially the loss of β-CNs. The sample loaded into each well consisted of 7.5 µL MS ready sample, 7.5 µL 4X NuPAGE LDS Buffer, 14 µL water, and 1 µL of BME to reduce the sample (95°C for 5 minutes). Out of the 30 µL of sample prepared, 20 µL of each sample was loaded into each well. For each gel, the Precision Plus Protein™ Dual Xtra Standard was run. The gel was run, stained and destained as previously stated. Examining and identifying bands of protein determined the success of the sample preparation method.

4. Protein Loading Optimization

To evaluate the optimal protein concentration of the samples for MS analysis, Nestlé powder and Horizon UHT milk were run in triplicate through sample preparation and then diluted into a series of concentrations: 0.0005, 0.005, 0.05, and 0.5 µg/µL. A Thermo Scientific™ Q Exactive™ Plus hybrid quadrupole-Orbitrap™ mass spectrometer was used for analysis. A 500 µL syringe (Thermo Scientific™) and compatible tubing attached to a Rheodyne needle port used for HPLC (Thermo Scientific™) were used to directly infuse sample into the instrument. Instrument parameters were set as follows: resolution = 70,000 m/z, scan range = 600.0 to 3,000 m/z, fragmentation (in-source collision induced-dissociation (IS-CID)) = 20.0 eV, polarity = positive, sheath gas flow rate = 7, aux gas = 1, sweep gas flow rate = 5, spray voltage = 4.00, capillary temperature = 320°C, AGC target = 1e6, maximum inject time = 200 milliseconds and microscans = 1. Deconvolution was completed using Freestyle Xtract deconvolution with the averaging across window algorithm. The parameters are set to output mass = M, adduct element = H+ (1.00727663), charge range = 3 – 50, analyzer type = OT, relative abundance

threshold (%) = 0, isotope table = protein, and minimum number of detected charges = 3. Total intensity of sample signal, average fractional abundance of total signal of identified proteins, and number of identified proteins were used to evaluate the optimal protein loading optimization.

vii. Direct Infusion MS Instrument Parameter Optimization

A Thermo Scientific™ Q Exactive™ Plus hybrid quadrupole-Orbitrap™ mass spectrometer was used for all direct infusion analysis. A 500 µL syringe and compatible tubing attached to a Rheodyne needle port used for HPLC were used to directly infuse sample into the instrument (same as the protein loading optimization equipment as described above). Before samples were directly infused into the instrument, a cycle of washing the syringe and tubing using 100% methanol, 20% ACN 0.1% (v/v) FA, and air was implemented before each sample. The full scan MS method was run in positive ion mode with the following instrument parameters: resolution = 70,000 m/z, scan range = 600.0 to 3,000 m/z, fragmentation (IS-CID) = 20.0 eV, polarity = positive, sheath gas flow rate = 7, aux gas = 1, sweep gas flow rate = 5, spray voltage = 4.00, capillary temperature = 320˚C, AGC target = 1e6, maximum inject time = 200 milliseconds and microscans = 1. Specific instrument parameters, IS-CID, microscans, resolution, and flow rate and acquisition time, were later optimized and changed subsequently as described below.

5. IS-CID

Purified proteins (α_{S1}-, α_{S2}-, β-, κ-CN, BLG, and α-lac) at the concentration of 0.5 mg protein/mL 20% ACN 0.1% (v/v) FA were directly infused into the instrument. The samples were not prepared though the established sample preparation due to the purified

nature of the samples. The same instrument parameters were used as stated before with the change in IS-CID starting at 0, 20, 30, 50, and 100 eV for each purified protein. The sample spectra were acquired in one minute and compared to evaluate the optimal IS-CID.

6. Microscans

After IS-CID was optimized, microscan was the next instrument parameter to be optimized. Nestlé NFDM at a concentration of 0.5 mg protein/mL 20% ACN 0.1% (v/v) FA was prepared as described before. The evaluation of microscans was completed through the collection of one minute spectra acquisitions. The same instrument parameters used as stated above with the change in number of microscans (1, 5, and 10). The sample spectra were acquired in one minute and compared. Following spectral comparison, extracted ion chromatograms (XIC) were created using BLG peaks to note peak width (minutes) and number of points across the peak for each selected microscan setting was calculated to evaluate the optimal number of microscans.

7. Resolution

After IS-CID and microscans were optimized, resolution was the next instrument parameter to be optimized. Nestlé NFDM at a concentration of 0.5 mg protein/mL 20% ACN 0.1% (v/v) FA was prepared as described before. The evaluation of resolution at 70,000 or 140,00 was completed through the collection of one minute acquisitions for both resolutions. The same instrument parameters were used as described above with the change in resolution settings and deconvoluted using FreeStyle Xtract averaging across window algorithm and the parameters set as stated previously. Resolution was later assessed again with Nestlé NFDM in duplicate at a concentration of 0.5 mg/mL in 20%

ACN 0.1% (v/v) FA and deconvoluted using FreeStyle Xtract averaging across window algorithm and the parameters set as stated previously.

8. Flow Rate and Acquisition Time

After IS-CID, microscans, and resolution, were optimized, flow rate and acquisition time was the next instrument parameter to be optimized. The flow rates of 10 μL/min and 20 μL/min with three different acquisition times of 1, 3, and 5 minutes were evaluated using 0.5 μg protein/μL 20% ACN 0.1% (v/v) FA of four different samples: Nestlé powder, Serenity A2 Formula, Alexandre A2 HTST milk, and Hiland HTST milk. Each sample was prepared through established sample preparation protocol and analyzed with the previous set instrument parameters with the difference of flow rate and acquisition time. Serenity A2 Formula with the flow rate at 10 μL/min with a five minute acquisition and 20 μL/min with a one minute acquisition time was deconvoluted using FreeStyle Xtract averaging across window algorithm and the parameters set as stated previously.

viii. Direct Infusion Deconvolution Software Optimization

To compare the two available types of software for intact protein MS, FreeStyle Xtract and Biopharma Finder Xtract, Nestlé NFDM, Serenity A2 Formula, Hiland HTST milk, and Alexandre A2 HTST milk were prepared at 0.5 mg protein/mL with the established sample preparation protocol and deconvoluted with both software programs. The settings for FreeStyle Xtract were set as follows: output mass = M, adduct element = H+ (1.00727663), charge range = 3 – 50, analyzer type = OT, relative abundance threshold (%) = 0, isotope table = protein, minimum number of detected charges = 3.

The settings for Biopharma Finder Xtract deconvolution were set as follows: output mass range = 3,000 – 30,000, output mass = M, S/N Threshold = 3, relative abundance threshold (%) = 5%, charge range = 3 – 50, minimum number detected charge = 3, fit factor = 80%, consider overlaps = true, minimum intensity = 1, expected intensity error = 3, source spectra method = sliding windows or average spectrum, target average spectrum width = 0.5 minutes, target average spectrum offset scan = 1, merge tolerance = 30 ppm, merge scheme = legacy merge scheme, number of detected intervals = 3. The mass outputs (the number of masses deconvoluted from the spectra), total identified proteins (masses that matched masses within the predicted monoisotopic mass database), and proteins not identified in each software were compared.

Proteins were identified through comparison of the predicted monoisotopic mass proteoform database. Identification was only considered if experimental monoisotopic mass was within ten parts per million (ppm) of the predicted monoisotopic mass. This gave high confidence that the deconvoluted mass was the predicted proteoform within the generated database. From there, signal intensity attached to the proteoform provided by BiopharmaFinder deconvolution software was recorded.

Furthermore, the same set of data was evaluated for isotopically resolved (Xtract) and unresolved data (ReSpect^{TM}) via BioPharma Finder software average spectra algorithm. To assess the programs' ability to detect low abundance proteins, all purified proteins were prepped to 1 mg milk protein/mL 20% ACN 0.1% (v/v) FA and analyzed on the instrument. The amount of mass outputs and total identified proteins between Xtract and ReSpect^{TM} were evaluated. The settings for Biopharma Finder Xtract deconvolution were set as follows: output mass range = 6,000 – 60,000, output mass = M,

S/N Threshold = 3, relative abundance threshold (%) = 1%, charge range = 3 – 30, minimum number detected charge = 3, fit factor = 80%, remainder threshold (%) = 25, negative charge = false, charge carrier = H, minimum intensity = 1, expected intensity error = 3, and source spectra method = average over selected retention time.

The settings for Biopharma Finder ReSpect^{TM} deconvolution were set as follows: output mass range = 6,000 – 60,000, deconvoluted spectra display mode = Isotopic Profile (new), deconvolution mass tolerance = 20 ppm, choice of peak model = intact protein, model mass range = 8,000 – 70,000, charge state range = 7 – 100, minimum adjacent charges (low and high model mass) = 4 – 4, relative abundance threshold (%) = 0, quality score threshold = 0%, target mass = 70,000 Da, number of peak models = 1, left/right peak shape = 2:2, peak detection quality measure = 95%, peak model width factor = 1, intensity threshold scale = 0.01, noise comparison = true, negative charge = false, and source spectra method = average over selected retention time. The mass outputs (the number of masses deconvoluted from the spectra) and total identified proteins (masses that matched masses within the predicted average mass database were used to determine the optimal deconvolution program.

ix. Final Optimization of Direct Infusion MS method

To assess the possibility of simplifying the complicated milk product matrix, alkaline phosphatase from bovine intestinal mucosa (Sigma-Aldrich®) was introduced into the sample preparation workflow to dephosphorylate the proteins withing the sample, specifically β-CNs (Figure 2.5). The enzyme was added at different concentrations of 1, 5, and 10 mg/mL in triplicate and incubated for one hour. Spectra was deconvoluted

using Biopharma Finder Xtract software average over the spectrum algorithm as described above.

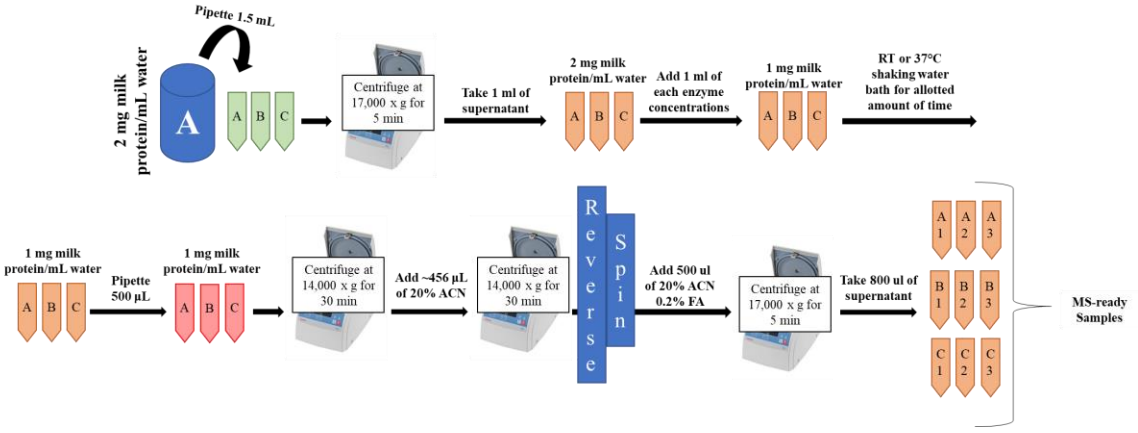


Figure 2.5: Sample Preparation Method with Alkaline Phosphatse Addition.

d. Liquid Chromatography Mass Spectrometry

After optimization of sample preparation, instrument parameters, deconvolution software, and final optimization attempt with alkaline phosphatase, the number of masses reliably identified through direct infusion were still lower than 50% protein identification of the total sample signal. This low percentage of identification led to the addition of LC to the method. The addition of LC further separated the proteins and enhanced the detection of lower abundant target proteins compared to the direct infusion method that analyzed all proteoforms at once.

The column selected for LC was UltiMate 3000 RSL® liquid chromatography (UPLC) system (Thermo Scientific™), equipped with a Acquity UPLC Protein BEH C4, 300 Å, 1.7 µm, 2.1 mm × 150 mm and a VanGuard™ pre-column, 2.1 mm × 150 mm (Waters Corporation, Milford, MA) due to previous studies completed in this area [47-51]. Reverse phase columns come in a variety of different hydrophobic alkyl chains that make up the stationary phase [52]. The C4 chains are usually used for large proteins

molecules because larger proteins are more likely to possess more hydrophobic moieties (part of a molecule) to interact with the column. If higher chains (C8, C18) were used, the large protein molecules would take much longer to elute off the column, if at all.

Though there was a change in how the sample was introduced into the instrument, sample preparation and instrument parameters optimized for direct infusion MS were carried over to the intact protein LC method as described below.

x. Finalized Liquid Chromatography MS Instrument Parameters

Instrument parameters optimized for direct infusion were kept for LC-MS as described below in Table 2.2.

Table 2.2: Established Parameters for Intact Protein LC-MS Carried Over from Direct Infusion

HESI Parameters	
Capillary temperature	320°C
Polarity	Positive
Sheath gas flow rate	50
Auxiliary gas	10
Sweep gas flow rate	0
Spray voltage	3.5kv
MS Parameters	
Scan range	600.0 to 2,500.0 m/z
Microscans	5
AGC target	1e6
Maximum injection time	200 ms
Resolution	140,000

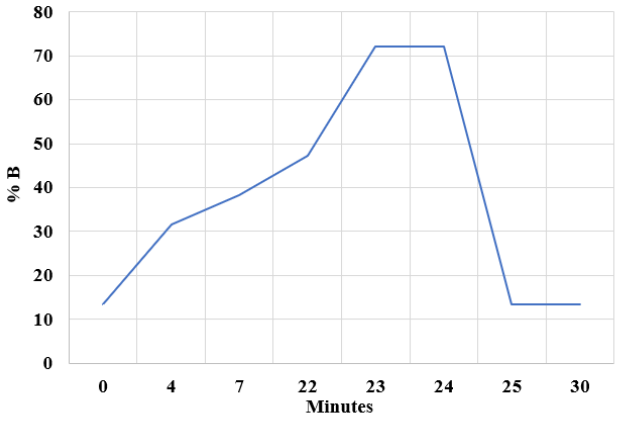
xi. Finalized Liquid Chromatography MS Deconvolution Settings

With the addition of LC, the Xtract sliding windows algorithm, which BioPharma Finder provides, was chosen. The sliding windows analyzes selected intervals of the chromatography (the window) with a chosen amount of overlap to deconvolute in a step-wise fashion until it reaches across the entire spectra [53]. This approach to

deconvolution helps identify low abundance, large molecules in a complicated sample with poorly defined chromatographic peaks and can identify components that co-elute at overlapping retention time ranges. The deconvolution parameters were set as follows for intact protein LC-MS deconvolution: output mass range = 3,000 – 30,000, output mass = M, S/N Threshold = 3, relative abundance threshold (%) = 5%, charge range = 3 – 50, minimum number detected charge = 3, fit factor = 80%, consider overlaps = true, minimum intensity = 1, expected intensity error = 3, source spectra method = sliding windows or average spectrum, target average spectrum width = 0.5 minutes, target average spectrum offset scan = 1, merge tolerance = 30 ppm, merge scheme = legacy merge scheme, number of detected intervals = 3.

xii. Finalized Liquid Chromatography MS Gradient

With the addition of LC, the gradient of solvents needed to be optimized. Buffer A consisted of water containing 0.1% (v/v) FA and buffer B consisted of 100% (v/v) ACN containing 0.1% (v/v) FA. Samples were injected at 15 µL on-column and proteins separated and eluted from the column using a gradient of 13.5–72 % mobile phase B over 31 min at a flow rate of 250 µL.min-1 (Figure 2.6). Analysis of sample on MS was scheduled between 2 and 23 minutes while the other was directed into waste to equilibrize the column and wash. Optimal sample protein concentration and load on column would later be evaluated.



Time	Flow (mL/min)	%B (Acetonitrile 0.1% Formic Acid)
Equilibration		
-2.000	0.250	13.5
Run		
0.000	0.250	13.5
4.000	0.250	31.5
7.000	0.250	38.2
22.000	0.250	47.3
23.000	0.250	72.0
24.000	0.250	72.0
25.000	0.250	13.5
30.000	0.250	13.5
31.000	*Stop Run*	

Figure 2.6: Intact Protein LC C4 Column Gradient.
Solvent A: water containing 0.1% (v/v) FA. Solvent B: 100% (v/v) ACN containing 0.1% (v/v) FA.

xiii. Liquid Chromatography MS Sample Preparation Optimization

A similar sample preparation used for direct infusion was transferred over to LC-MS with the omission of two steps. The first step omitted was the buffer exchange on the Amicon® Ultra 0.5 mL Centrifugal Filters since the sample flows through the column with MS solvents (Figure 2.7). The second step omitted was the centrifugation step, 17,000 x g for five minutes, because the guard column would prevent particulates and chemical contaminants not suitable for the instrument [54].

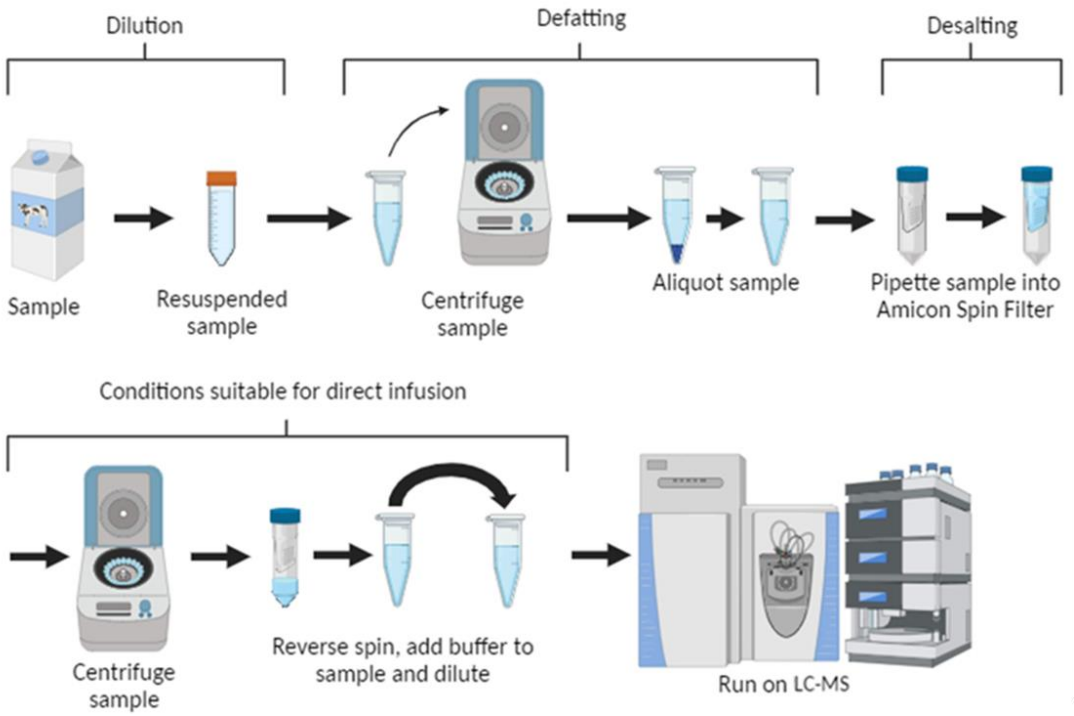


Figure 2.7: Intact Protein LC-MS Sample Preparation Summary. Created with BioRender.com.

1. Protein Loading and End Buffer Optimization

For the intact protein LC-MS method, protein loading and end buffer (buffer that the sample was diluted into before analysis) were optimized simultaneously as described below. In addition, information regarding the protein loading optimization for direct infusion (0.5 µg milk protein/µL) was considered at the beginning of optimization as described below.

Buffer types, water containing 0.1% (v/v) FA, 6.35 mM trisodium citrate containing 0.1% (v/v) FA, and 6.25 mM trisodium citrate 23 mM dithiothreitol (DTT) (DL-1,4-Dithiothreitol, BioUltra Sigma-Aldrich®) containing 0.1% (v/v) FA, were tested with sample protein concentrations of 0.8 mg/mL, 0.625 mg/mL, and 0.5 mg/mL as shown in Figure 2.8. Comparing the spectra and total intensity of the different protein loads and end buffers, combinations were chosen to be assessed more.

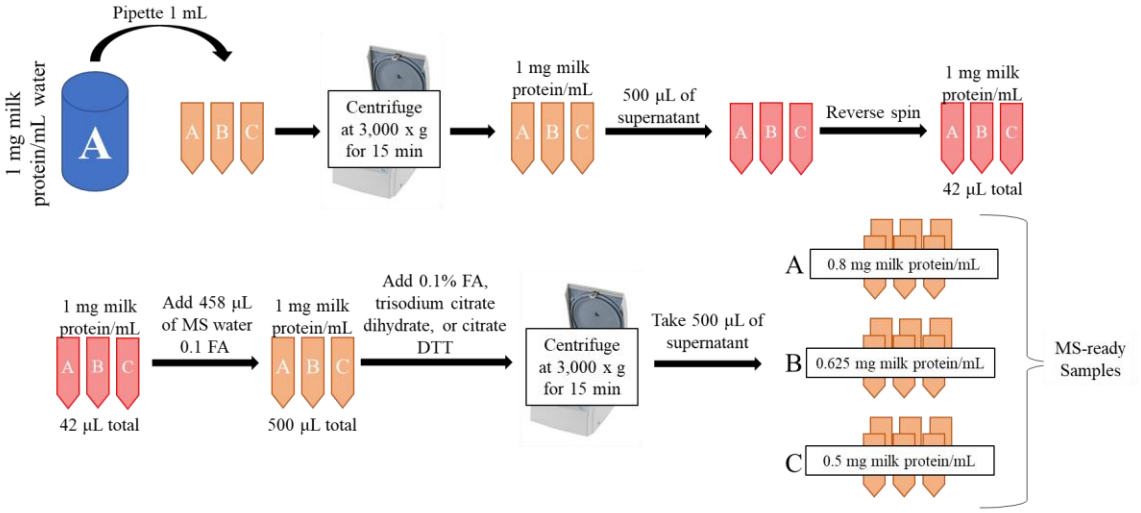


Figure 2.8: Protein Loading and Sample Buffer Optimization using Nestlé NFDM Samples and 6.25 mM Trisodium Citrate Buffers.
Protein loading concentrations were 0.8, 0.625, and 0.5 mg milk protein/mL. Buffers tested were water containing 0.1% (v/v) FA, 6.25 mM trisodium citrate containing 0.1% (v/v) FA, and 6.25 mM trisodium citrate 23 mM DTT containing 0.1% (v/v) FA.

After the first investigation into protein loading and end buffer optimization condition was completed, 6.25 mM trisodium citrate and 23 mM DTT was chosen to be further explored with the addition of a heat step after the addition of DTT to reduce the proteins. A new set of Nestlé NFDM samples prepared to 0.05 mg milk protein/mL and 0.005 mg milk protein/mL were analyzed (Figure 2.9). In addition, purified BLG at concentrations of 0.075 mg/mL and 0.25 mg/mL with three different buffers of 6.25 mM trisodium citrate, 6.25 mM trisodium citrate and 23 mM DTT, and 6.25 mM trisodium citrate and 23 mM DTT with a heated step (37°C for 5 minutes) was performed. Comparing the spectra, total intensity of the different protein loads and end buffers, combinations were chosen to be assessed furthermore.

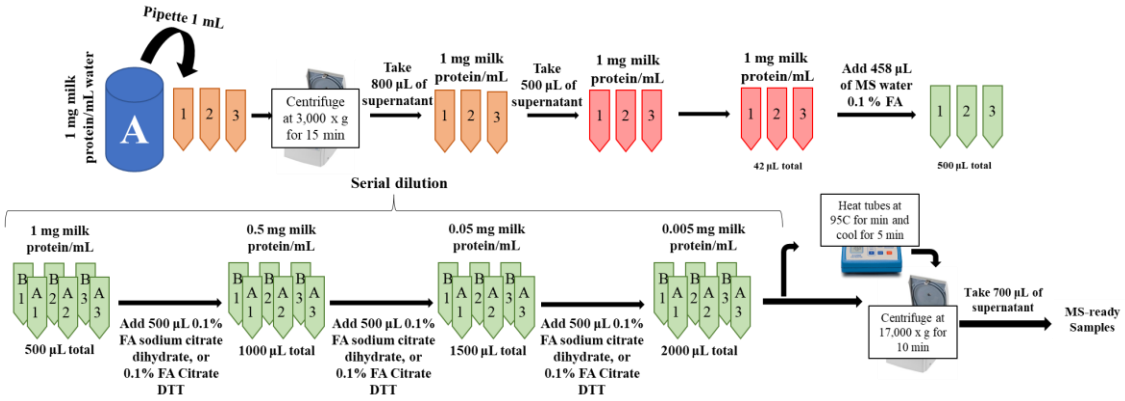
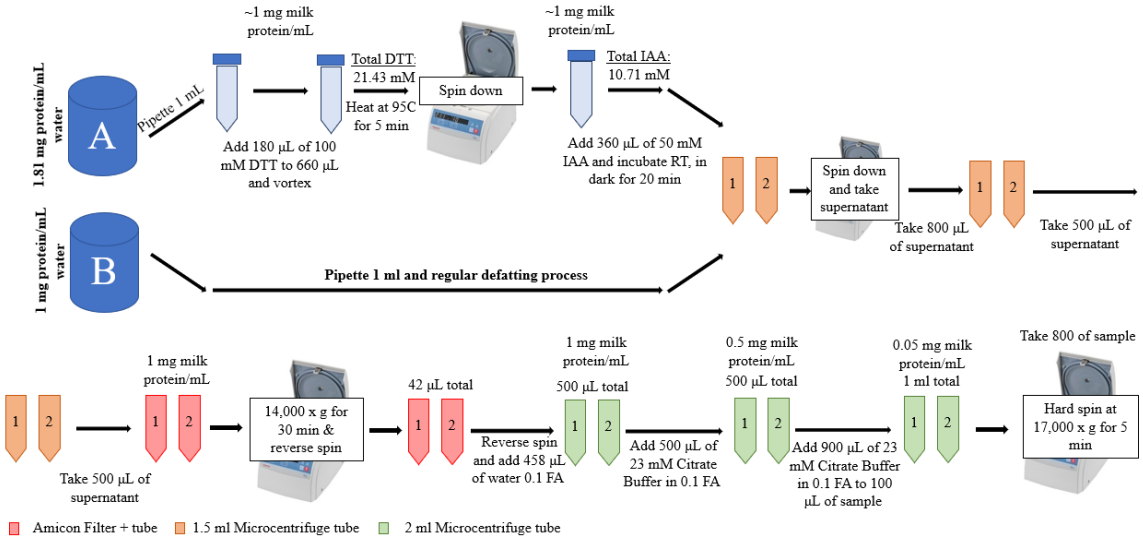


Figure 2.9: Protein Loading and Sample Buffer Optimization using Nestlé NFDM Samples and 6.25 mM Trisodium Citrate Buffers and Heating.
Protein loading concentrations were 0.5, 0.05mg milk protein/mL. Buffers tested were 6.25 mM trisodium citrate containing 0.1% (v/v) FA, and 6.25 mM trisodium citrate 23 mM DTT containing 0.1% (v/v) FA with a 37ºC for five minutes heating step.

After the second investigation into protein loading and sample buffer optimization condition was completed, 6.25 mM trisodium citrate and 23 mM DTT was chosen again to be further explored with the addition of a heat step after the addition of DTT and iodoacetamide (IAA) (BioUltra Sigma-Aldrich®) to reduce and alkylate the proteins. A new set of Nestlé NFDM samples were prepared to 0.5 mg/mL and 0.05 mg/mL 6.25 mM sodium citrate buffer 21.43 mM total DTT 10.71 mM total IAA containing 0.1% (v/v) FA. The new sample preparation method was assessed through MS analysis, deconvoluted using Biopharma Finder Xtract sliding windows algorithm with parameters and described above (Figure 2.10).



Figrue 2.10: Protein Loading and Sample Buffer Optimization using Nestlé NFDM and Purified Protein Samples and 6.25 mM Trisodium Citrate Buffers and Heating. Protein loading concentrations were 0.5, 0.05mg milk protein/mL. Buffers tested were 6.25 mM trisodium citrate containing 0.1% (v/v) FA, and 6.25 mM trisodium citrate 23 mM DTT 21.43 mM total DTT 10.71 mM total IAA containing 0.1% (v/v) FA with a 37ºC for five minutes heating step.

In addition to the spectral comparison and deconvolution results of Nestlé NFDM and purified proteins (α_{S1}-, α_{S2}-, β-, κ-CN, BLG, and α-lac), the reduction and alkylation with DTT and IAA in the sample preparation was evaluated using reduced and non-reduced NuPAGE™ 4-12% Bis-Tris Gels (1.00 mm x 12 well). For the non-reduced gels, the sample loaded into each well consisted of 4 µL the MS ready sample, 5 µL 4X NuPAGE LDS Buffer, and 11 µL water. The samples were heated at 95°C for 5 minutes. Out of the 20 µL of sample prepared, 15 µL of each sample was loaded into each well. For the reduced gels, the sample loaded into each well consisted of 4 µL the MS ready sample, 5 µL 4X NuPAGE LDS Buffer, and 10 µL water, and 1 µL of BME to reduce the sample and heated at 95°C for 5 minutes. NuPAGE SDS Running Buffer was used in a Mini-Cell Electrophoresis Chamber. For each gel, the Precision Plus Protein™ Dual Xtra Standard was run. The gel was run, stained, and destained as previously stated.

Examining and identifying bands of protein determined the success of reduction and alkylation.

A new set of the samples, Nestlé NFDM and purified proteins, were evaluated on reducing NuPAGE™ 4-12% Bis-Tris Gels (1.00 mm x 12 well). These gels served as a protein reduction control to compare to the samples prepared using the sample preparation with 6.25 mM sodium citrate buffer 21.43 mM total DTT 10.71 mM total IAA containing 0.1% (v/v) FA as described above. The sample loaded into each well consisted of 7 µL MS ready sample, 5 µL 4X NuPAGE LDS Buffer, and 8 µL water. Out of the 20 µL of sample prepared, 15 µL of each sample was loaded into each well. For each gel, the Precision Plus Protein™ Dual Xtra Standard was run. The gel was run, stained, and destained as previously stated. Examining and identifying bands of protein determined the success of reduction and alkylation.

To further assess the sample preparation, the heat step (37°C for 30 minutes) was evaluated using the same samples as before (Nestlé NFDM and purified proteins of αS1-, αS2-, β-, κ-CN, BLG, and α-lac) with reduced and non-reduced NuPAGE™ 4-12% Bis-Tris Gels (1.00 mm x 12 well). The sample loaded into each well consisted of 4 µL the MS ready sample, 5 µL 4X NuPAGE LDS Buffer, and 10 µL water, and 1 µL of BME (Sigma-Aldrich®) to reduce the sample (heated at 95°C for 5 minutes). Out of the 20 µL of sample prepared, 15 µL of each sample was loaded into each well. For each gel, the Precision Plus Protein™ Dual Xtra Standard was run. The gel was run, stained, and destained as previously stated.

Ultimately, after 6.25 mM trisodium citrate was assessed with the combinations of reduction, heat, alkylation steps, the method produced low identification percentages

of the total signal. This led to the optimization using 50 mM ammonium bicarbonate. This buffer was chosen for its volatile nature and handles up to pH 6.9 to ensure the proteins, especially the caseins, are not near the isoelectric point [55, 56]. The new buffer was evaluated using Nestlé NFDM and purified proteins (α-, β-, κ-CN, BLG, and α-lac) without the use of DTT, heat, or IAA at protein loading concentrations of 50 and 500 µg/mL.

xiv. Final Liquid Chromatography Optimization of All Commercial Samples

After the shift to 50 mM ammonium bicarbonate containing 0.1% (v/v) FA as the final sample buffer, ten commercial samples, Nestlé NFDM, a2 Company WMP, Azure Market A2 WMP, Similac A2 Formula, Gerber A2 Formula, Enfamil A2 Formula, Serenity A2 Formula, Hiland HTST milk, a2 Company HTST milk, and Alexandre A2 milk were prepared in triplicate with the established sample preparation method but resolubilized in both water and 50 mM ammonium bicarbonate. Subsequently, with the addition of more samples to the set later, fifteen commercial samples, Nestlé NFDM, Hy-Vee NFDM, a2 Company WMP, Azure Market A2 WMP, Similac Formula, Similac A2 Formula, Gerber Formula, Gerber A2 Formula, Enfamil Formula, Enfamil A2 Formula, Serenity A2 Formula, Hiland HTST milk, Hy-Vee HTST milk, a2 Company HTST milk, and Alexandre A2 milk, were prepared in triplicate with the established sample preparation method but resolubilized in both water and 50 mM ammonium bicarbonate. Samples were analyzed on the MS to be deconvoluted with BioPharma Finder sliding windows algorithm (parameters described previously).

IV. RESULTS AND DISCUSSION

a. Sequence Database Selection

i. Primary Sequences

The major proteins were evaluated, and six of the eight proteins were chosen. The six proteins selected were α_{S1}-, α_{S2}-, β-, κ-CN, BLG, and α-lac based on low variability, high abundance (>10 % of total milk protein), and within 11-26 kDa in size. BSA (P02769) was excluded because of its low abundance and large protein size (69,293 Da) [5, 46, 57-62]. Ig was excluded because of its high level of variance and large mass [5]. Recommended resolutions for full scan MS methods and past intact protein work found in literature did not have the resolution lower than 70,000. From there, two resolutions were chosen to assess: 70,000 and 140,000. Resolution relates to the amount of time spent on each scan and the general rule that higher resolution works better with smaller masses while lower resolution works better with larger masses [53, 63]. This means that larger proteins cannot be adequately detected at that resolution as lower resolution is needed, so a smaller range was required.

Beginning with the UniProt database, 43 sequences for the selected proteins and their genetic variants were collated. Using the Retrieve/ID mapping function, a list of 174 sequences from the NCBI protein database in the form of GI IDs were generated. These IDs within the NCBI database were used for potential unique sequences not listed in UniProt. Out of the 174 matches from NCBI, 27 sequences were found to be unique. Combining the 66 sequences from UniProt and 27 sequences from NCBI, 93 primary sequences were included in the database.

ii. Modifications and Proteoform Consideration

The additional modifications that were initially considered were disulfide bonds, a free oxidized sulfhydryl, glycosylation, phosphorylation, and multiple lactose adducts. Disulfide bonds mainly affect α_{S2}-, κ-CN, BLG and α-lac because of the cysteines in the sequence that could create the disulfide bonds between each other. This bond would result in the loss of -1.0079 Da for each bond present (one hydrogen molecule). The possibility of one free oxidized sulfhydryl only applies to this group as well. Glycosylation of κ-CN and potentially other caseins were not considered as O-linked glycosylation complicates the search due to the high variance thus challenging to predict the final mass [64-67]. κ-CN and β-CN can undergo different types of proteolysis [46, 68]. κ-CN are split into a smaller and larger molecule by chymosin: para kappa casein (~12 kDa) and glycomacropeptide (~6.5 kDa) (GMP) and β-CN can undergo proteolysis by plasmin to generate γ-CNs (~11 kDa).

Phosphorylation primarily affects α_{S1}-, α_{S2}-, and β-CNs. Each protein has a variable amount of phosphorylation sites (β-CN with 5, α_{S1}-CN with 9, and α_{S2}-CN with 13) [46]. Three sequences from the NCBI database are labeled as partial sequences that have N-terminus cleavage (CAC37028.1, ABR10906.1, ABL74247.1). This removes the possible phosphorylation sites that have been documented, so those options of phosphorylation were removed. Based on previous research on bovine milk samples that have been spray dried (NFDM, WMP, and infant formula), the mass addition of 324.1 Da (lactose molecule without water molecules) can be seen and attributed to masses detected and identified [7]. The high temperature processing can lead to the addition of multiple lactose additions to lysine residues. Most of the proteins selected are usually found in

monomer form, but BLG can be found in a dimer form resulting in a mass of ~36 kDa. With the resolution limitation, however, the mass would not be within range of the instrument. Considering all modifications mentioned, excluding glycosylation, the final predicted mass database consists of 1,000+ monoisotopic masses with each mass attributed to a specific proteoform of one of six proteins (see supplemental file attachment). Note that one database contained monoisotopic masses and a separate identical database was created with average masses.

b. Direct Infusion Mass Spectrometry

i. Direct Infusion MS Sample Preparation Optimization

1. Sample Storage and Defatting Conditions

Evaluation with NuPAGE™ 4-12% Bis-Tris Gels (1.00 mm x 12 well) of 1 mL Nestlé NFDM and Horizon UHT milk samples at differing storage temperatures and centrifugation durations showed that no proteins were removed (e.g., precipitation out of solution) from the sample under the selected storage temperatures or centrifugal durations (Figure 2.11).

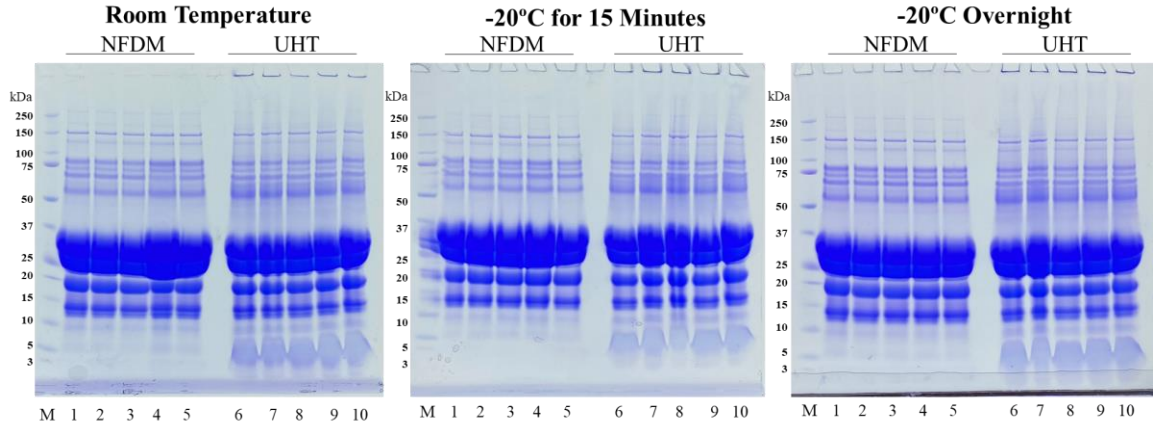


Figure 2.11: SDS-PAGE for Analysis of Centrifugation Parameters in Nestlé NFDM and Horizon UHT Milk.
M: Molecular weight markers; lane 1: control (no change in temperature or centrifugation); lanes 2-3: centrifuge for 10 minutes ; lanes 4-5: centrifuge for 15 minutes. Each lane contained 101.7 µg of protein.

Since there was not a discerning difference in band intensity between the treatments visually, room temperature storage was chosen. Centrifugation duration of fifteen minutes was chosen since the fat layer was thicker than samples centrifuged at ten minutes (Table 2.3).

Table 2.3: Comparison of Centrifugation Times at 3,000 x g at Room Temperature for Nestlé NFDM and Horizon UHT milk

Sample	Centrifugation Time (minutes)			
	5	10	15	30
Horizon UHT Milk				
Nestlé NFDM				

2. Desalting and Clean Up Protocol

Amicon® Ultra 0.5 mL Centrifugal Filters and buffers (water and 20% ACN containing 0.1, 0.05, and 0.01% (v/v) FA) were evaluated using the Pierce™ BCA Protein Assay Kit established sample preparation for direct infusion. It should be noted that the Pierce™ BCA Protein Assay Kit detects Cu^{+1}, not protein itself [69]. The total protein quantification is based upon the understanding that protein can reduce Cu^{+2} to Cu^{+1} in an alkaline solution and results in a purple color with bicinchoninic acid which is then read by a plate reader (562 nm).

As shown in Figure 2.12, there is no discernable difference in the use of Amicon® spin filters, but a major loss in protein when sample is initially resolubilized in 20% ACN. The loss resulted in the choice to resolubilize the samples initially in water followed by a buffer exchange with Amicon® spin filters.

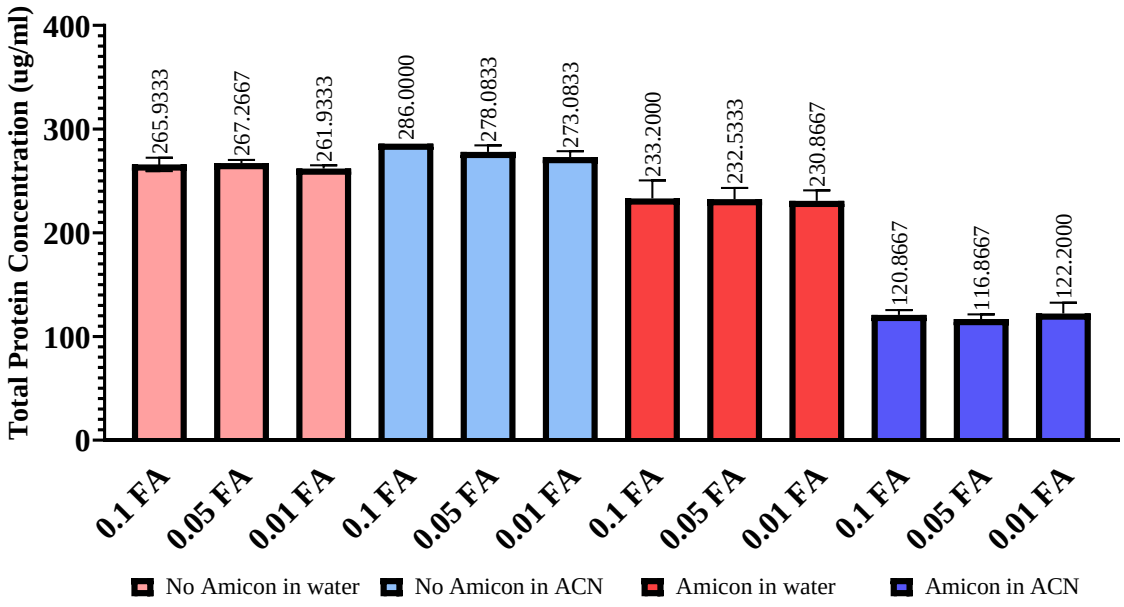


Figure 2.12: Evaluation of Amicon® Spin Filter and Buffers for Sample Preparation Optimization using Nestlé NFDM.
Protein concentrations were 500 ug/µL.

Samples solubilized in water and 20% ACN as the final buffer resulted in a lower amount of protein detected within the sample (152.80 ± 1.00 μg/mL and 157.80 ± 1.73 μg/mL respectively) compared to samples containing water and FA (Figure 2.13). Regarding the samples with the final buffer containing 0.1%, 0.05%, and 0.01% (v/v) FA, there is no observable difference (246.47 ± 2.31 μg/mL, 247.80 ± 13.08 μg/mL, and 247.80 ± 1.53 μg/mL respectively) between the percent FA used. Since there was no discernable difference, the traditional concentration of 0.1% (v/v) FA used for MS methods was chosen.

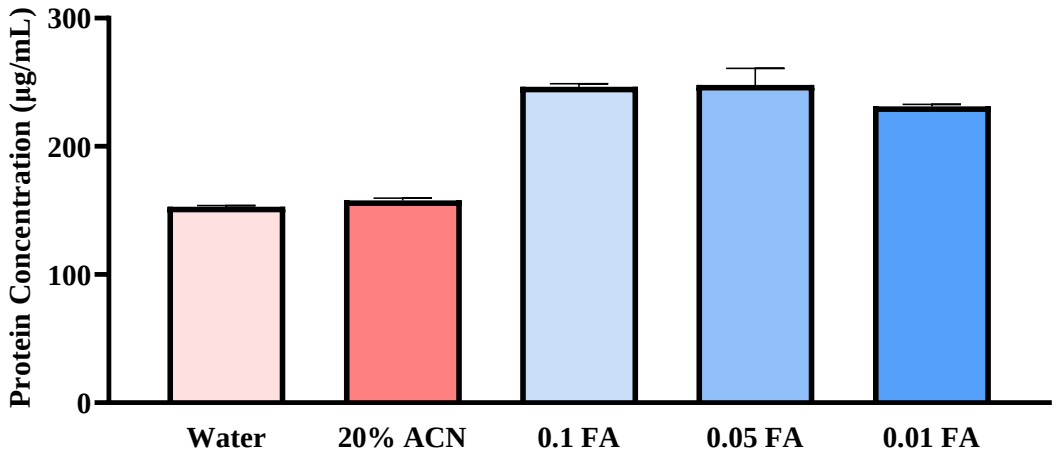


Figure 2.13: Evaluation of Buffers for Sample Preparation Optimization using Nestlé NFDM.
Protein concentrations were 500 ug/μL.

The pH of the buffers tested are all above or below the isoelectric point of caseins (pH 4.6) and thus supports the decision of 0.5 mg milk protein/ml 20% ACN 0.1% (v/v) FA (Table 2.4). There are consistent results that percent concentration FA does not affect the protein concentration when coupled with 20% ACN and makes protein more soluble. In addition, 20% ACN does not cause the protein to precipitate after being resuspended in water first.

Table 2.4: Evaluation of Final Buffer for Sample Preparation Optimization using 0.5 mg Milk Protein/mL using Nestlé NFDM.

Solution	Fisher brand	Baxter	Color-pHast
0.5 mg milk protein/mL water	6	6	5-5.5
0.5 mg milk protein/mL 20% ACN 0.1% FA	6	5.5	5-5.5
0.5 mg milk protein/mL 20% ACN 0.05% FA	3	3	2.5
0.5 mg milk protein/mL 20% ACN 0.01% FA	5	3	3
0.5 mg milk protein/mL 20% ACN 0.005% FA	5	3.5	3

3. Evaluation of Finalized Sample Preparation Protocol

SDS-PAGE gels were completed to evaluate the established protein loss between steps of the sample preparation (Figure 2.14). Though there are minor concentration losses between lanes of before sample preparation, after the defatting, and after Amicon® spin filter steps, all the bands of expected proteins remain. This confirms that neither defatting nor Amicon® spin filter step causes a significant amount of protein loss. With the established parameters of sample preparation and the confirmation that the protocol does not reduce the expected proteins extensively, a sample preparation protocol was established as shown in figure. With the established parameters of sample preparation and the confirmation that the protocol does not noticeably reduce the expected proteins, a sample preparation protocol was established.

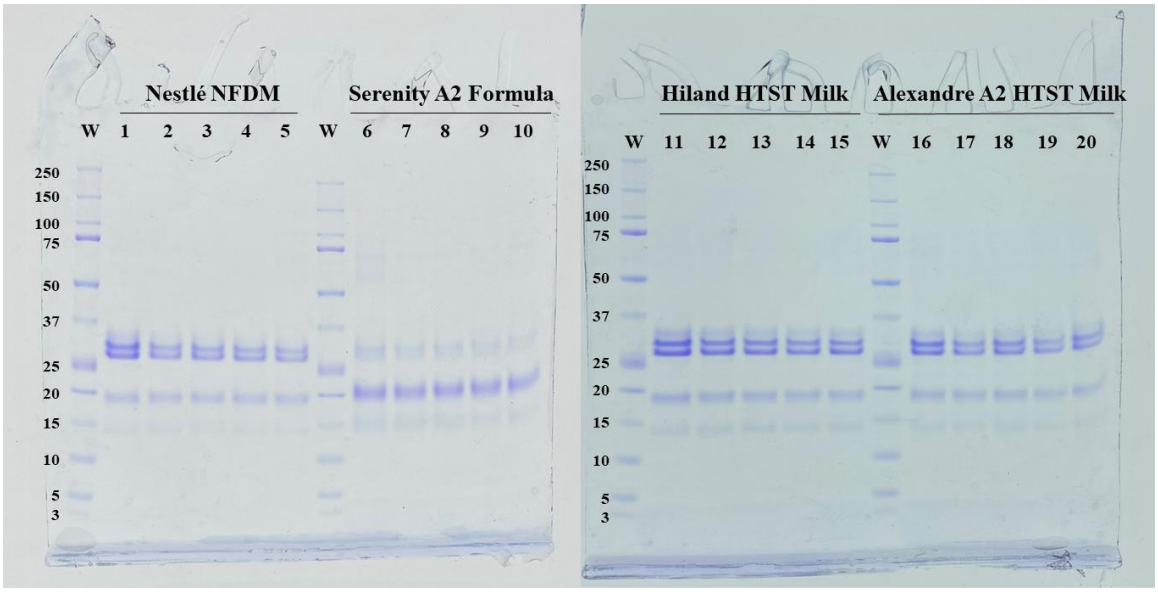


Figure 2.14: Evaluation Finalized Sample Preparation Protocol using Nestlé NFDM, Serenity A2 Formula, Hiland HTST Milk, and Alexandre A2 HTST Milk with SDS-PAGE Gels.
Lanes are as follows: before sample preparation = 1, 6, 11, 16; after defatting step = 2, 3, 7, 8, 12, 13, 17, 18; after Amicon step = 4, 5, 9, 10, 14, 15, 19, 20.

4. Protein Loading Optimization

The protein concentration of 0.5 µg protein/µL was selected based on highest signal intensity and greatest number of masses deconvoluted from the spectra for Nestlé NFDM (Figure 2.15) and Horizon UHT milk (Figure 2.16). The average fractional abundance of total signal of identified proteins in the sample for Nestlé powder and Horizon UHT at 0.0005, 0.005, 0.05, and 0.5 µg protein/µL were highly variable and generally lower than 50% identification as detailed in Table 2.5. Samples that had higher than 50% identification had few masses deconvoluted from raw data and thus should not be considered a success in signal identification of the sample. Combining the greater signal intensity, average fractional abundance of total signal of identified proteins, and number of identified proteins, 0.5 ug protein/uL was chosen as the optimal protein concentration.

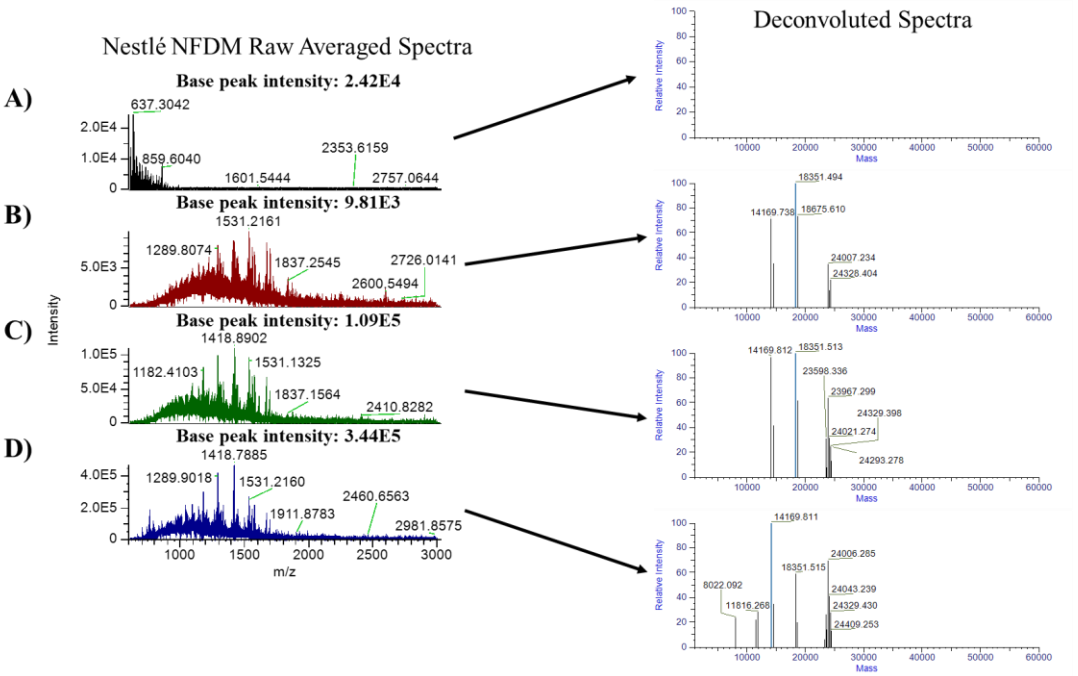


Figure 2.15: Initial Direct Infusion Spectra of Different Concentrations of Nestlé NFDM and Deconvoluted Spectra.
A) 0.0005 µg/µL protein, **B)** 0.005 µg/µL milk protein, **C)** 0.05 µg/µL milk protein, and **D)** 0.5 µg/µL milk protein. This figure shows one replication out of three.

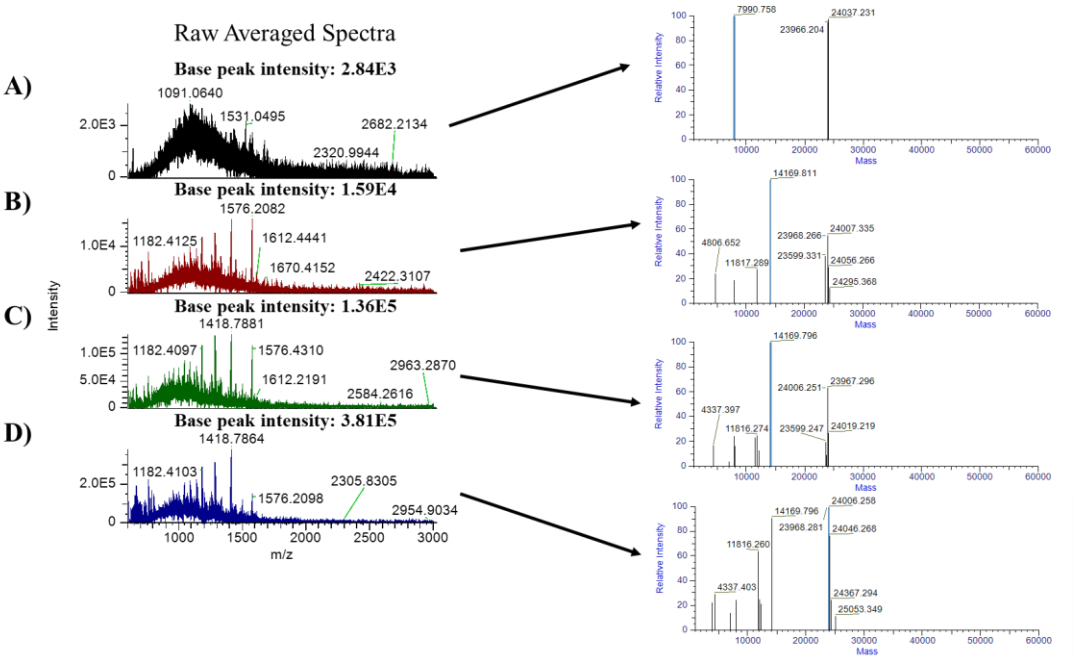


Figure 2.16: Initial Direct Infusion Spectra of Different Concentrations of Horizon UHT Milk and Dwconvoluted Spectra.
0.0005 µg/µL protein, **B)** 0.005 µg/µL milk protein, **C)** 0.05 µg/µL milk protein, and **D)** 0.5 µg/µL milk protein. This figure shows one replication out of three.

Table 2.5: Initial Comparison of Deconvoluted Masses in Different Protein Loading Concentrations for Nestle NFDM and Horizon UHT Milk Resolubilized in Water.

Sample	Concentration (µg protein/µL)	Identification (%)	STD	Number of identified proteins	STD	Total Masses	STD
Nestle Powder	0.0005	70.63	41.54	3	0	5.5	5.5
	0.005	53.45	12.23	6.67	1.53	15.67	2.31
	0.05	48.25	10.32	9.33	2.52	26.33	3.21
	0.5	32.24	17.30	11	3.61	38.67	0.58
Horizon UHT Milk	0.0005	36.46	41.33	1.33	1.15	6	6
	0.005	47.89	17.72	5.33	1.53	14	2.65
	0.05	34.84	7.36	7	3	23.33	3.06
	0.5	27.88	6.45	8.33	2.08	30	2.65

ii. Direct Infusion MS Instrument Parameter Optimization

After the optimization of sample preparation conditions, specific instrument parameters were optimized (IS-CID, microscans, resolution, and flow rate and acquisition time) and changed subsequently as described below.

1. IS-CID

After analysis of purified proteins (α_{S1}-, α_{S2}-, β-, κ-CN, BLG, and α-lac) at the concentration of 0.5 mg protein/mL 20% ACN 0.1% (v/v) FA, fragmentation through IS-CID was chosen. For all five purified proteins, 50 and 100 eV for IS-CID was too high as shown in Figure 2.17. To help break apart any small molecules that might still be in the sample, IS-CID is preferred to be used with direct infusion intact mass spectrometry. Since most proteins did not start breaking apart between 20 to 30 eV IS-CID, 25 eV was chosen.

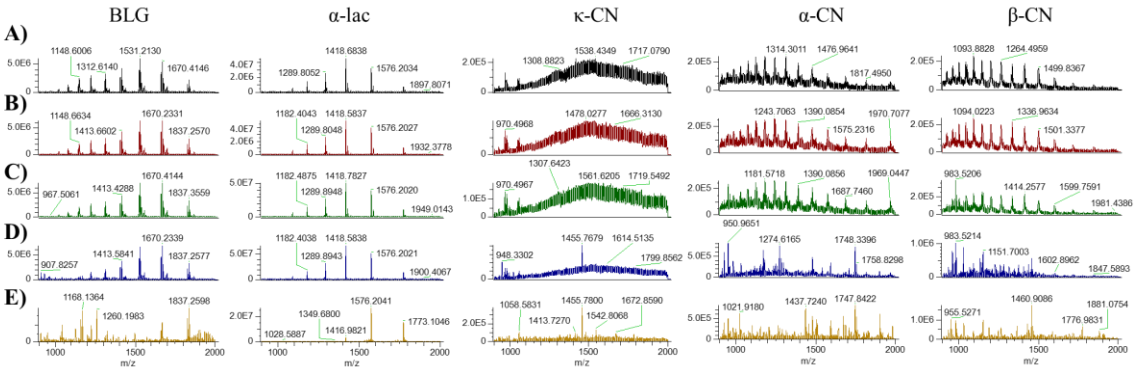


Figure 2.17: Direct Infusion Intact MS Comparison of Six Purified Major Milk Proteins to Optimize IS-CID.

2. Microscans

Acquisitions using microscan settings of 1, 5, and 10 were performed with Nestlé NFDM at a concentration of 0.5 mg protein/mL 20% ACN 0.1% (v/v) FA. The spectra of the three runs does not give a definite answer to a microscan decision besides the baseline using 1 microscan drifts (Figure 2.18) [70]. Using extracted ion chromatograms to determine the peak width is ~30 seconds, the following calculation from 1 to 10 microscans was done (Table 2.6). This determined that 5 microscans should be chosen as that allowed 11 points across the peak with the assumption that the peak is 30 seconds.

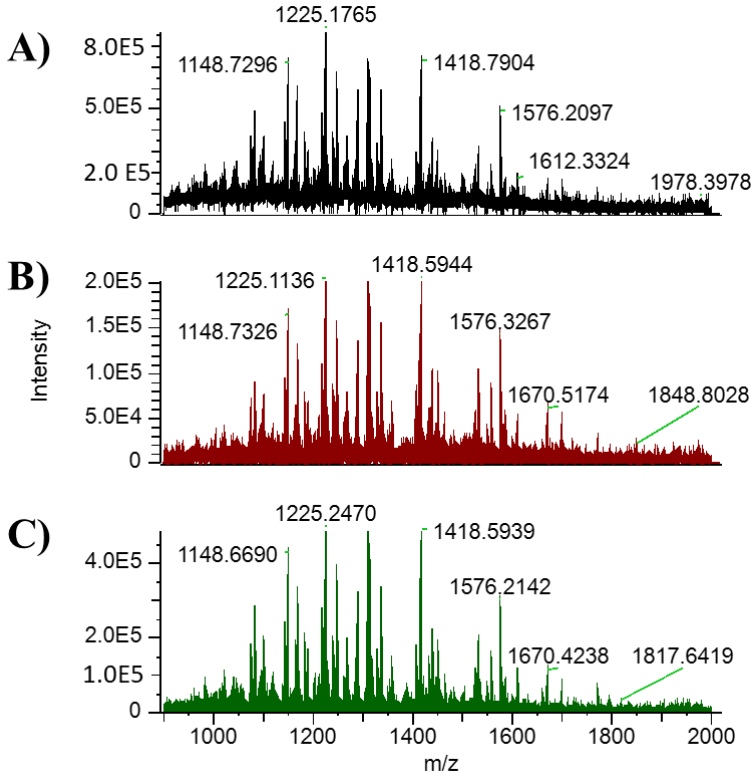


Figure 2.18: Spectral Comparison of 0.5 mg protein/mL Nestlé NFDM to Evaluate Microscan Optimization.

Table 2.6: Direct Infusion Intact Protein MS Microscan Calculation.

Microscan(s)	Difference between data points (seconds)	XIC assumed peak duration	Points across the peak
1	0.518785921	30	57.82732098
2	1.037571843	30	28.91366049
3	1.556357764	30	19.27577366
4	2.075143686	30	14.45683024
5	2.593929607	30	11.5654642
6	3.112715529	30	9.63788683
7	3.63150145	30	8.261045854
8	4.150287372	30	7.228415122
9	4.669073293	30	6.425257887
10	5.187859215	30	5.782732098

3. Resolution

The evaluation of resolution with the Nestlé NFDM sample was collected to evaluate which to use with the expected proteins in the sample. The spectra of the two runs shows a baseline drift with 70,000 resolution (Figure 2.19).

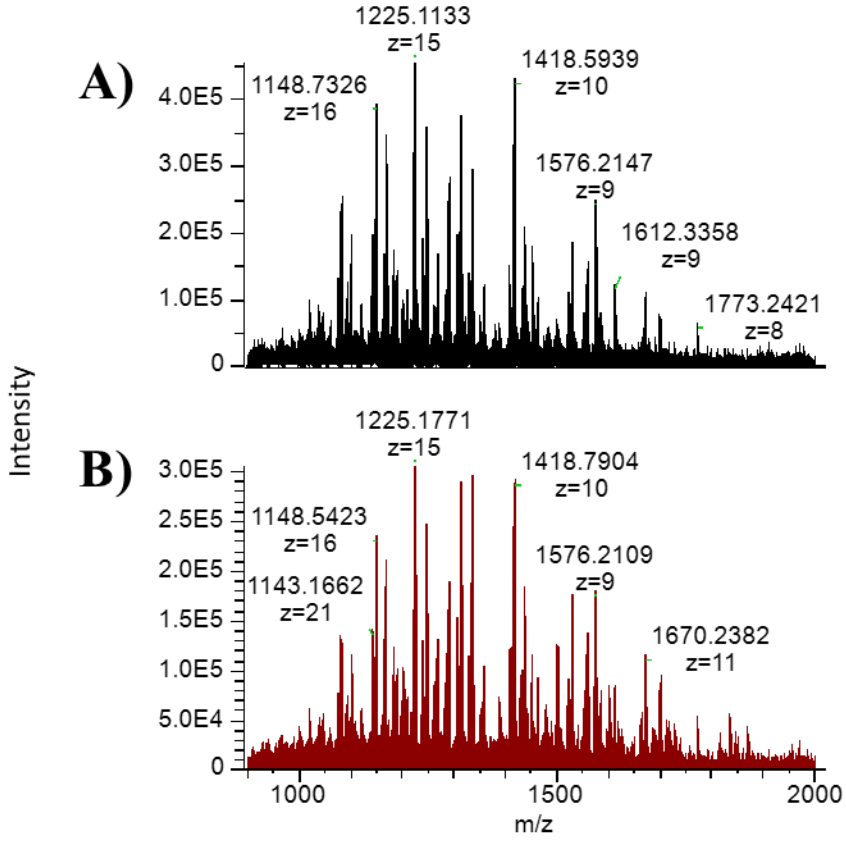


Figure 2.19: Direct Infusion Intact Protein MS Spectral Comparison for Optimization of Resolution.
A) 70,000 **B)** 140,000.

Upon deconvolution using FreeStyle Xtract, both 70,000 and 140,000 produced a similar number identified proteins (14 and 13 respectively) with A1-type β-CNs and κ-CNs not being identified. There was a higher signal to noise ratio when using 70,000 resolution, meaning there's a greater sample signal to noise signal, and thus 70,000 was chosen at the time. However, when investigating multiple replicates with three minute captures of Nestlé NFDM and Serenity A2 Formula, the greater number of masses detected for both samples using 140,000 and lack of masses while using 70,000 (Figure 2.20). Considering that the higher the resolution, the greater the ability to distinguish between ions differing in m/z and proteoforms can be a few Daltons away from one another, the higher resolution would be the better choice between the two as well [70].

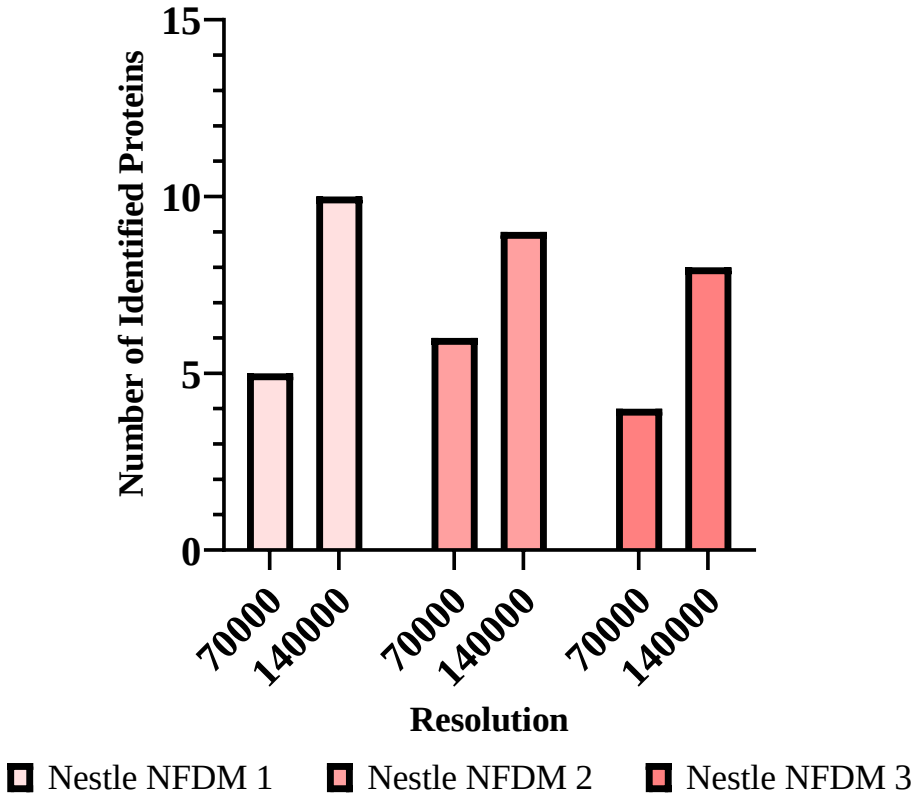


Figure 2.20: Comparison of the Number of Identified Masses using Direct Infusion Intact Protein MS for the Optimization of Resolution.

4. Flow Rate and Acquisition Time

The spectra of each combination of flow rate and capture time was compared as shown in Figure 2.21. Serenity A2 Formula was analyzed first with the main two combinations compared were 10 µL/min with a capture time of 5 minutes and 20 µL/min with a capture time of 1 minute. There was no significant difference looking at the spectra and comparing the intensities for 10 µL/min with a capture time of 5 minutes and 20 µL/min were 1.71E8 and 3.39E8. When deconvoluted using Freestyle Xtract and parameters described above, the 10 µL/min with a capture time of 5 minutes and 20 µL/min with a capture time of one minute had a similar number of identified proteins (5

of 26 and 5 of 25 respectively) and similar percent identification (27.90 ± 4.08% and 27.9 ± 2.96% respectively).

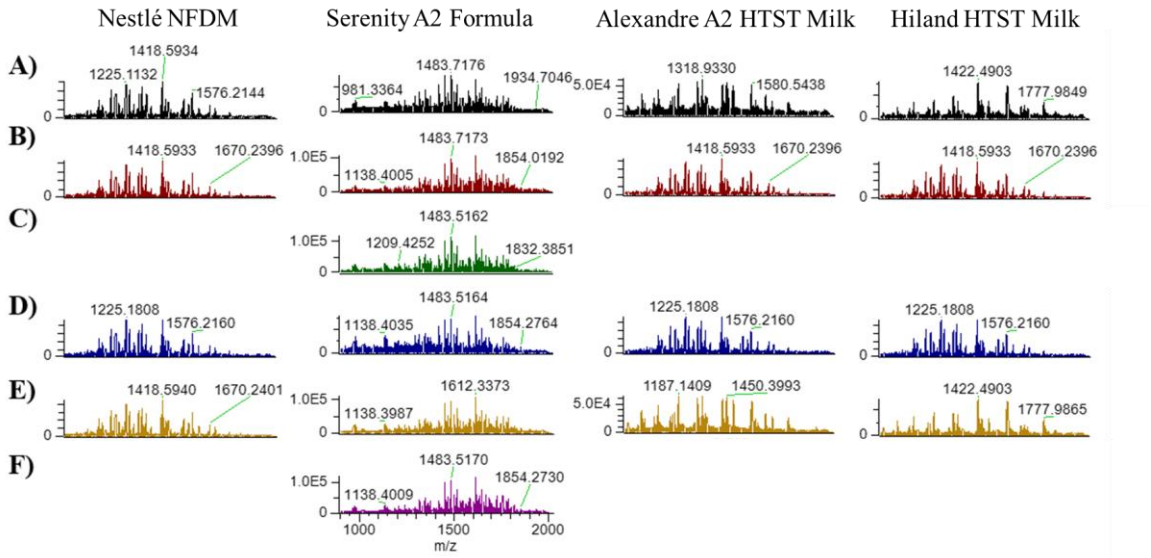


Figure 2.21: Flow rate and acquisition time comparison of Nestlé NFDM, Serenity A2 Formula, Alexandre A2 HTST Milk, and Hiland HTST Milk.
A) flow rate: 10 µl/min, acquisition time: 1 minute **B)** flow rate: 10 µl/min, acquisition time: 3 minutes **C)** flow rate: 10 µl/min, acquisition time: 5 minutes **D)** flow rate: 20 µl/min, acquisition time: 1 minute **E)** flow rate: 20 µl/min, acquisition time: 3 minutes **F)** flow rate: 20 µl/min, acquisition time: 5 min.

Due to the similar results of both flow rates and acquisition time, the compromise between the two, 20 uL/min with a three minute acquisition time, was selected. This was the focus while analyzing the other samples. Analysis after deconvolution of those samples of powdered and liquid milks at 0.5 µg/µL with the selected flow rate and capture time resulted in 74.19% (±6.54%), 19.15% (±2.31), 18.29% (±3.50), and 19.15% (±2.30) identification through fraction abundance of total signal for Nestlé NFDM, Serenity A2 Formula, Alexandre A2 HTST milk, and Hiland HTST milk. The number of proteins identified for Nestlé NFDM, Serenity A2 Formula, Alexandre A2 HTST milk, and Hiland HTST milk were 12 of 35, 6 of 32, 4 of 31, and 8 of 38 respectively.

iii. Direct Infusion MS Deconvolution Software Optimization

Evaluation of the two available deconvolution software options, FreeStyle Xtract and Biopharma Finder, with the average across the spectrum algorithms produced very similar mass outputs and protein identification within the sample signal (Table 2.7). FreeStyle Xtract deconvolution required manual averaging while BioPharma Finder averaged across the spectra systematically and produced excel files and reports. There were more deconvolution parameters that could be modified and the ability to use different algorithms. Because of this, averaging across the peaks using BioPharama Finder was chosen. The main difference between the two was the ability for sliding windows algorithm (explained in later section).

Table 2.7: Comparison of Deconvolution Programs FreeStyle Xtract and Biopharma Finder.

	A2 (Nestle NFDM)	**B2** (A2 Serenity NFDM)	**C3** (Hiland Non-A2 Milk)	**D2** (A2 Alexandre Milk)
BioPharma Outputs[a]	31	26	35	39
Xtract (3C) Outputs[a]	32	26	35	39
Total Matched Output Masses	30	26	35	39
Close matches (off by 1 Da)	0	N/A	N/A	N/A
Outputs not in BioPharma	2	N/A	N/A	N/A
Outputs not in Xtract	1	N/A	N/A	N/A
Total Identified Proteins[b]	13	3	3	8
Not identified in BioPharma	0	0	0	0
Not identified in Xtract (3C)	1	0	0	0

[a]Output masses are the list of masses as the result of deconvolution
[b]Identified masses are masses compared to the predicted masses database

Within the Biopharma Finder deconvolution software, the deconvoluted masses can be generated either as isotopically resolved (Xtract) or unresolved (ReSpect). When compared with Nestlé NFDM, Serenity A2 Formula, Hiland HTST milk, Alexandre A2 HTST Milk and purified proteins (α_{S1}-, α_{S2}-, β-, κ-CN, BLG, and α-lac), using averaging across the spectrum, ReSpect produced more deconvoluted masses from the data. However, the number of identifications were not greater and created more noise in data to

analyze compared to Xtract for both commercial samples (Figure 2.22) and purified proteins (Figure 2.23). Of the protein identifications, ReSpect identified 11 proteins, while Xtract identified 22 proteins. Because of this, the Xtract program was chosen. With the effort of changes to the sample preparation and deconvolution software, the highest fractional abundance percent identification of the total signal only reached around 70% with the lack of consistent identification of the A1-type β-CN.

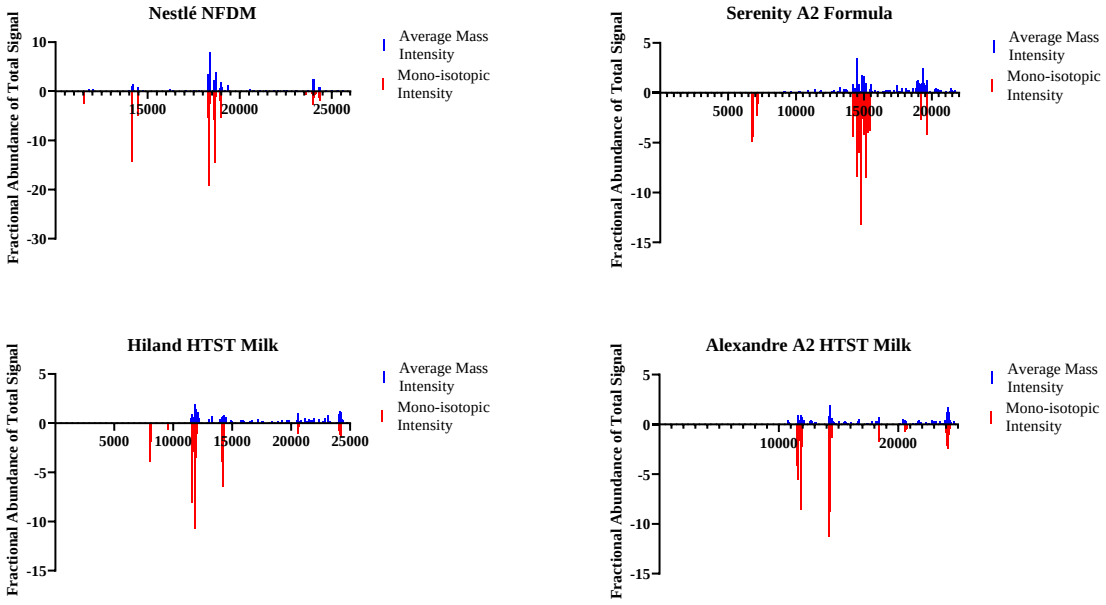


Figure 2.22: Number of Deconvoluted Masses of Major Bovine Milk Proteins in Selected Commercial Producrs to Evaluate BioPharma Finder Xtract (Monoisotopic Mass) ReSpect.

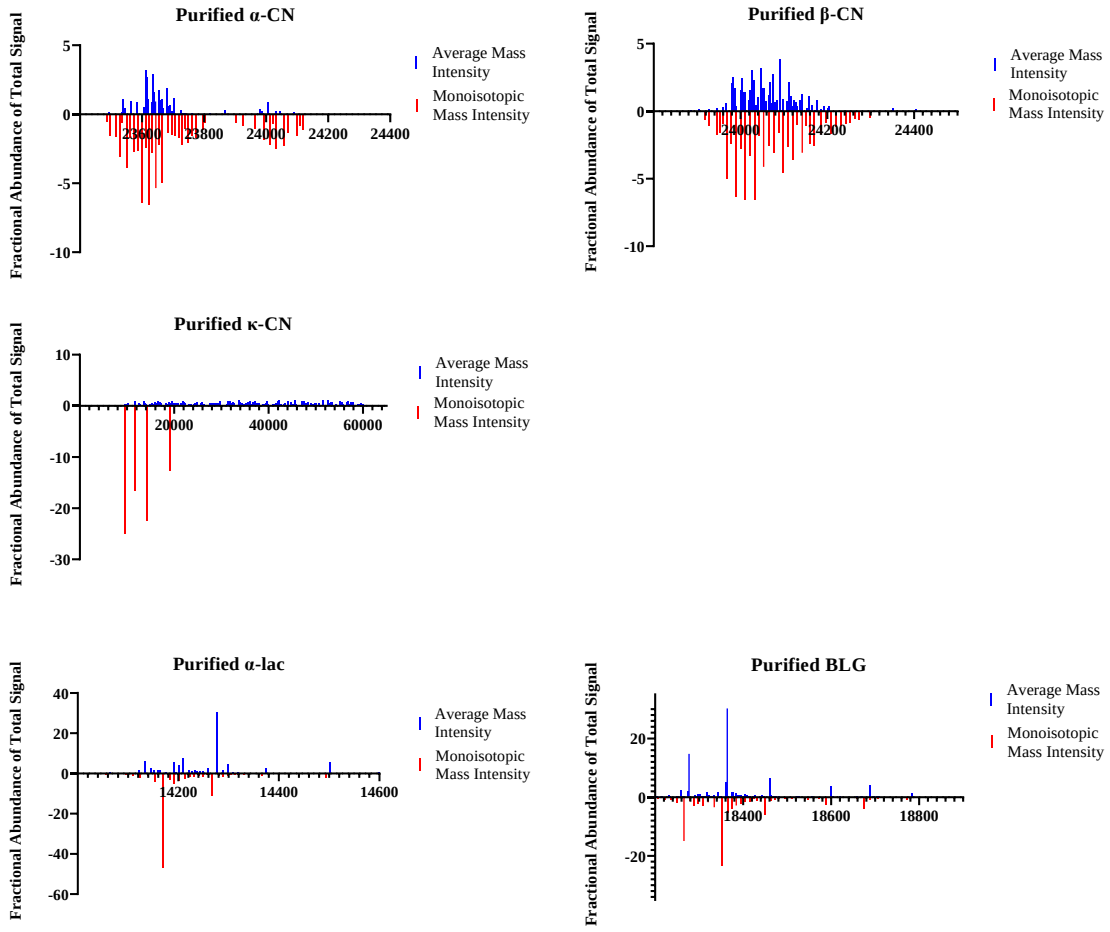


Figure 2.23: Number of Deconvoluted Masses of Major Bovine Milk Protein in Purified Major Bovine Milk Protein to Evaluate BioPharma Finder Xtract (Monoisotopic Mass) ReSpect.

iv. Final Optimization of Direct Infusion MS method

With the development of the sample preparation and selecting deconvolution software, the highest fractional abundance percent identification of the total signal only reached around 70% with the lack of consistent identification of the A1-type β-CN. The addition of alkaline phosphatase caused the proteins in sample to be digested by other enzymes not listed in the purified enzyme due to the origin, bovine intestine. This was concluded due to the increase of low masses as the enzyme concentration that was added increased (Figure 2.24). Masses below 13,952 Da increased from 37 proteins at 1 mg/mL

alkaline phosphatase to 63 proteins at 5 mg/mL alkaline phosphatase to 113 proteins at 10 mg/mL. Because of this, further use of alkaline phosphatase was removed from the method, and a new method type was pursued.

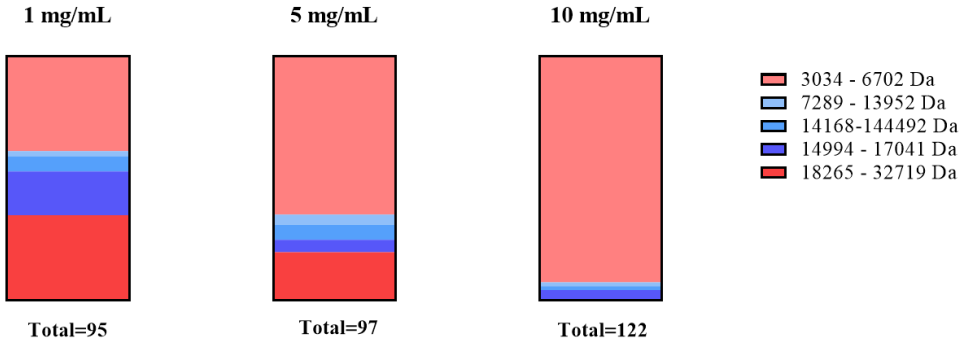


Figure 2.24: Deconvoluted Mass Ranges of Three Concentrations of Alkaline Phosphatase Addition.
Ranges were determined by major protein type analyzed (i.e. α-lac = 14168-144492).

c. Liquid Chromatography Mass Spectrometry

When comparing the deconvoluted data of 500 mg/mL Nestlé NFDM with the difference of direct infusion method and intact LC-MS method, the amount of deconvoluted masses and proteins identified are very different (Figure 2.25). The improvement which using LC is very apparent visually and supports the decision of movement from direct infusion to the addition of liquid chromatography.

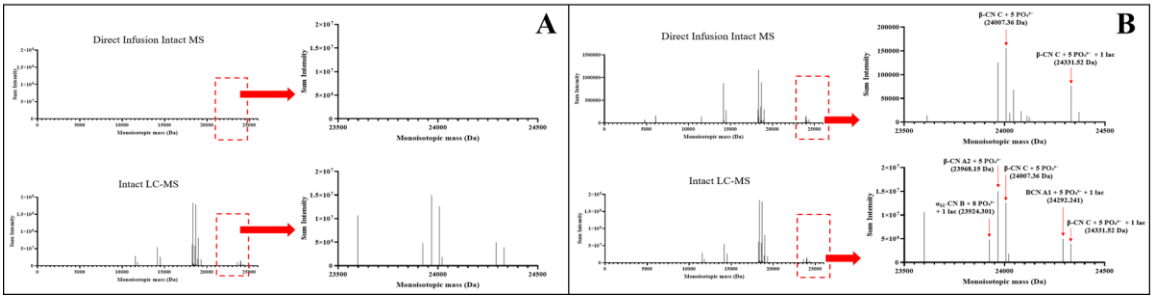


Figure 2.25: Comparison of Direct Infusion Intact Protein MS and Intact Protein LC-MS Spectra.
(A) shows the same y axis range while **(B)** changes the y axis to show spectra.

i. Finalized Liquid Chromatography MS Instrument Parameters

Liquid Chromatography MS Instrument Parameters was finalized as described above (Materials and Methods section dx).

ii. Finalized Liquid Chromatography MS Deconvolution Settings

Liquid Chromatography MS Deconvolution Settings was finalized as described above (Materials and Methods section dxi).

iii. Liquid Chromatography MS Sample Preparation Optimization

The selection of 0.500 mg/mL protein concentration of Nestlé NFDM was chosen compared to 0.625 mg/mL, and 0.800 mg/mL because the chromatography did not show a difference in peak shape, but all concentrations were at the highest limit to what the MS could analyze with base peak intensities as high as 1e9 (Figure 2.26). A dilution gradient starting at 0.500 mg/mL and lower was needed for further analysis. The selection 6.25 mM trisodium citrate and 6.25 mM trisodium citrate 23 mM DTT to resolubilize the sample needed to be further analyzed.

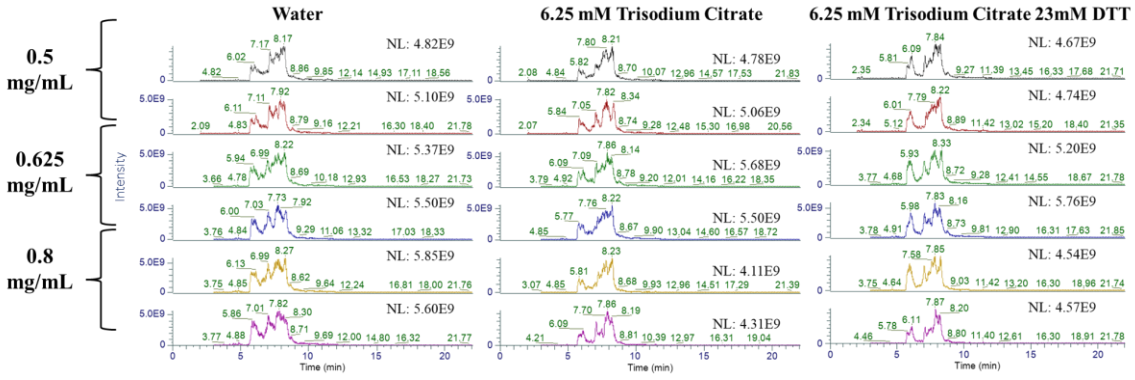


Figure 2.26: Chromatography Comparison to Evaluate Protein Loading and Buffers. Buffers include water and 0.1% (v/v) FA, 6.25 mM trisodium citrate 0.1% (v/v) FA, and 6.25 mM trisodium citrate 23 mM DTT 0.1 FA. Protein loading concentrations include 0.500 mg milk protein/mL, 0.625 µg milk protein/mL, and 0.800 mg milk protein /mL.

The next dilution gradient created using Nestlé powder at 5 µg protein/mL and 50 µg protein /mL and the latter was chosen based on peak shape and intensity as 5 µg protein /mL produced too low intensity as reflected in the chromatography (Figure 2.27). The selection of buffer to resolubilize the sample in was 6.25 mM trisodium citrate and 6.25 mM trisodium citrate 23 mM DTT to be further analyzed. Purified BLG went through the same type of analysis and produced similar results (data not shown). The addition of the 37ºC heat step for 5 minutes did have an effect as two peaks became visible around the 6 and 11.5 minute mark on chromatography.

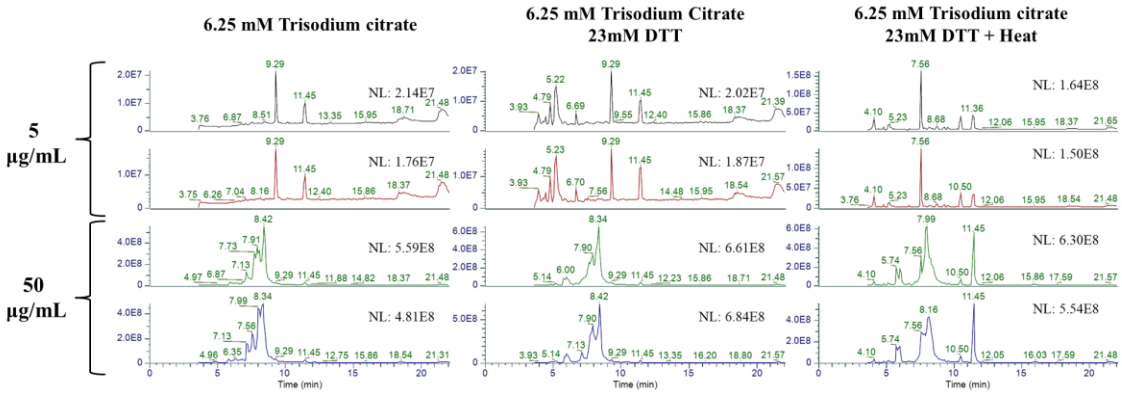


Figure 2.27: Chromatography Comparison to Evaluate Protein Loading and Buffers with a Heat Step.
Buffers include 6.25 mM trisodium citrate 0.1% (v/v) FA, 6.25 mM trisodium citrate 23 mM DTT 0.1% (v/v) FA, and 6.25 mM trisodium citrate 23 mM DTT 0.1% (v/v) FA with a heating step. Protein loading concentrations include 5 µg milk protein/mL and 50 µg milk protein/mL.

When deconvoluted, all three samples of Nestlé NFDM in the three different buffers produced an average of 9 protein identifications of various BLG and α-lac proteoforms with one A2-type β-CN proteoform and sometimes one A1-type β-CN proteoform. The fractional abundance of proteins identified within the signal for 6.25 mM trisodium citrate, 6.25 mM trisodium citrate 23mM DTT, and 6.25 mM trisodium citrate 23 mM DTT was 95.70% (±1.72%), 87.08% (±3.07%), and 84.21% (±12.01%)

respectively. Since an alkylation step was not used, the low rate of identity could be attributed to disulfide bonds coming back together over time, so that step was added.

Using the control reduced and non-reduced gels containing the same samples, Nestlé NFDM and purified proteins (α_{S1}-, α_{S2}-, β-, κ-CN, BLG, and α-lac), were completed through SDS-PAGE reduced and non-reduced gels (Figure 2.28). This gave the reference for which protein bands should look non-reduced and reduced.

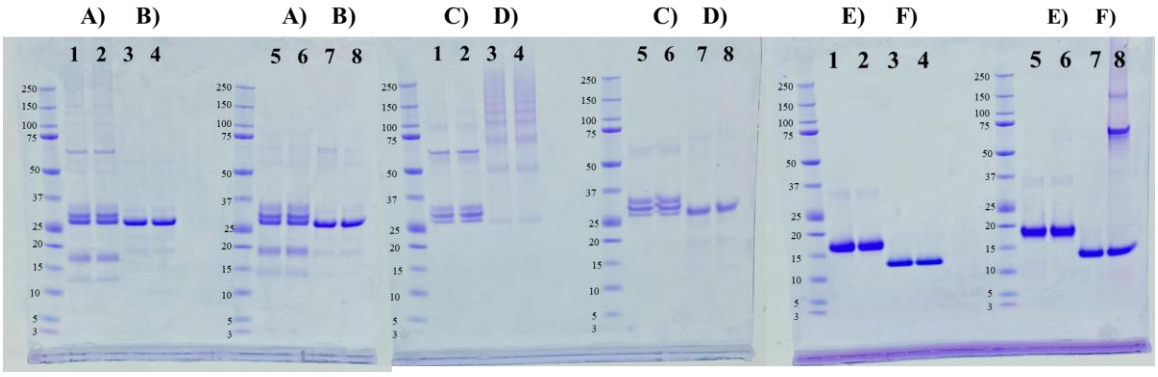


Figure 2.28: Reduced and non-reduced SDS-Page Gels to Evaluate the Reduction and Alkylation of Nestlé NFDM and Purified Proteins (αS1-, αS2-, β-, κ-CN, BLG, and α-lac).
Lanes 1-4 are non-reduced, and lanes 5-8 are reduced with β-mercaptoethanol. A) Nestlé NFDM, B) β-CN, C) α-CN, D) κ-CN, E) BLG, F) α-lac. Proteins with cystines: αs2-CN (2), κ-CN (2-3), BLG (4-5), α-lac (8).

With the reduction and alkylation of proteins in sample, further analysis of the samples, Nestlé NFDM and purified proteins (α_{S1}-, α_{S2}-, β-, κ-CN, BLG, and α-lac), through SDS-PAGE gels, the addition of DTT and IAA. Evaluating the before and after reduction and alkylation gels, proteins precipitated out of solution reflected in the SDS-PAGE gels shown in Figure 2.29. In addition to the SDS-PAGE gels, sequential chromatograms and spectra of sample shown in Figure 2.30, Figure 2.31, Figure 2.32, Figure 2.33, Figure 2.34, and Figure 2.35.

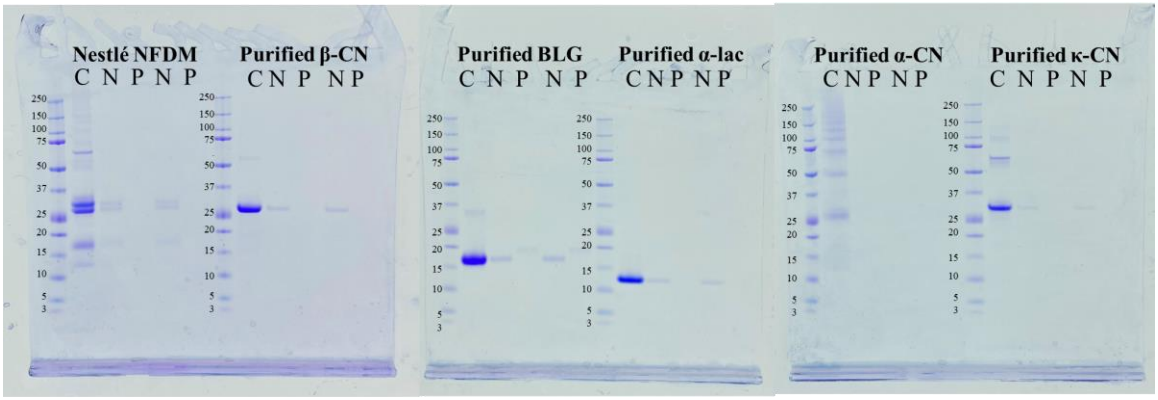


Figure 2.29: SDS-PAGE of Nestlé NFDM and Purified Proteins (αS1-, αS2-, β-, κ-CN, BLG, and α-lac) to Evaluate the Reduction and Alkylation with DTT and IAA. C = control (sample made up to 1 mg protein/ml water) N = "Negative", same sample prep as before P = protein reduced and alkylated.

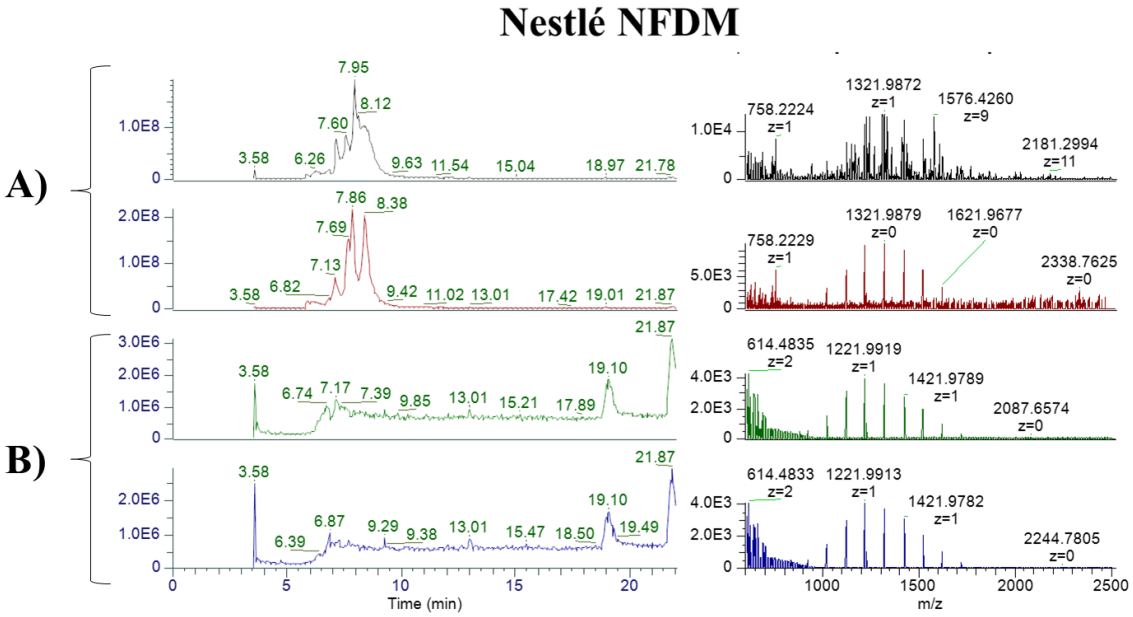


Figure 2.30: Chromatography and Averaged Spectra of 50 µg/mL Nestlé NFDM to Evaluate Reduction and Alkylation Protein Loss of Sample.
A) 23mM Trisodium Citrate 0.1%FA **B)** 23mM Trisodium Citrate 21.43 mM total DTT 10.71 mM total IAA 0.1%FA in duplicate.

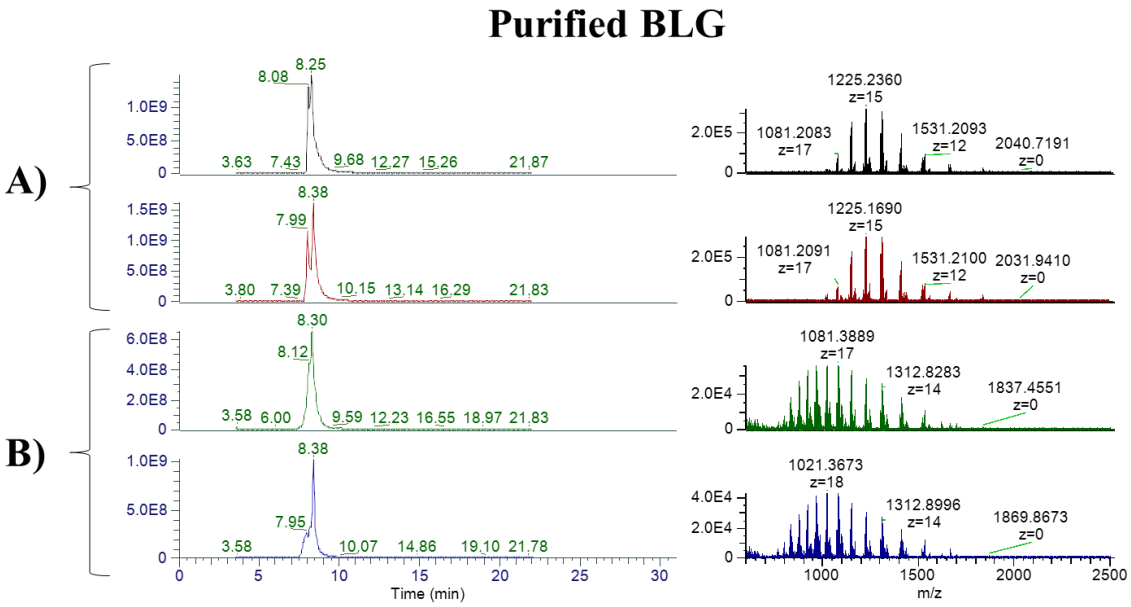


Figure 2.31: Chromatography and Averaged Spectra of 50 µg/mL Purified BLG to Evaluate Reduction and Alkylation Protein Loss of Sample.
A) 23mM Trisodium Citrate 0.1%FA B) 23mM Trisodium Citrate 21.43 mM total DTT 10.71 mM total IAA 0.1%FA in duplicate.

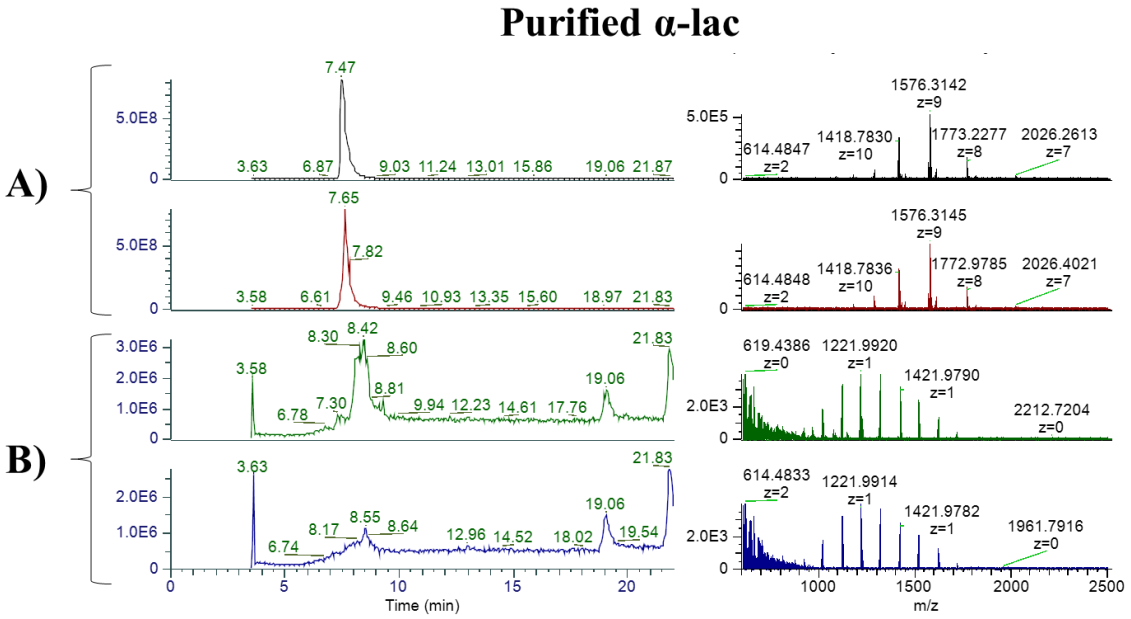


Figure 2.32: Chromatography and Averaged Spectra of 50 µg/mL Purified α-lac to Evaluate Reduction and Alkylation Protein Loss of Sample.
A) 23mM Trisodium Citrate 0.1%FA **B)** 23mM Trisodium Citrate 21.43 mM total DTT 10.71 mM total IAA 0.1%FA in duplicate.

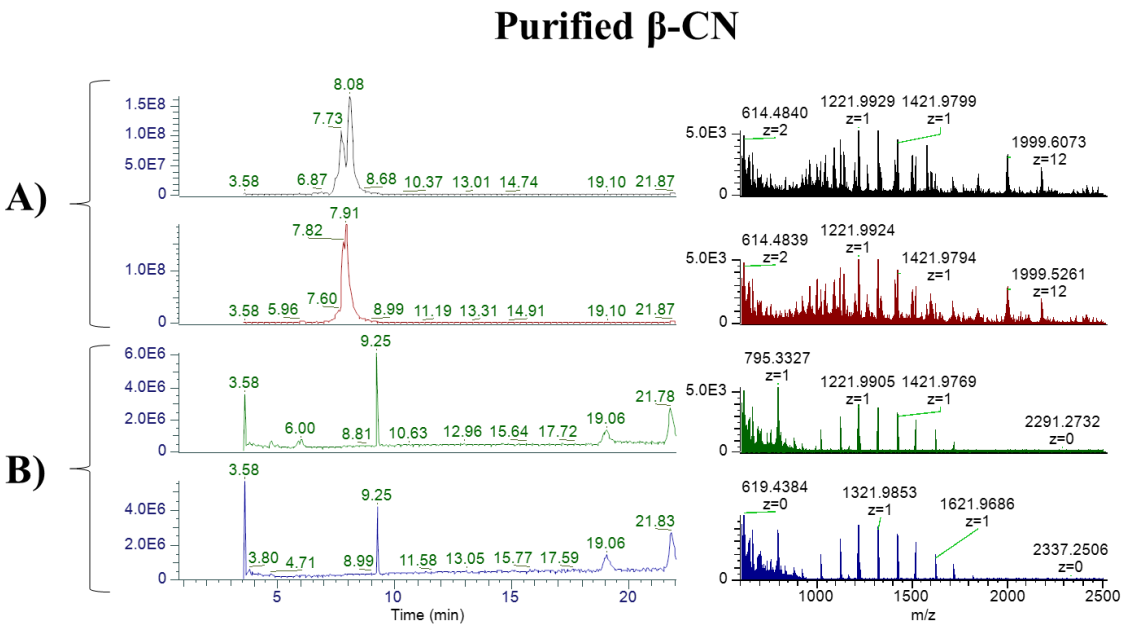


Figure 2.33: Chromatography and Averaged Spectra of 50 µg/mL Purified β-CN to Evaluate Reduction and Alkylation Protein Loss of Sample.
A) 23mM Trisodium Citrate 0.1%FA **B)** 23mM Trisodium Citrate 21.43 mM total DTT 10.71 mM total IAA 0.1%FA in duplicate.

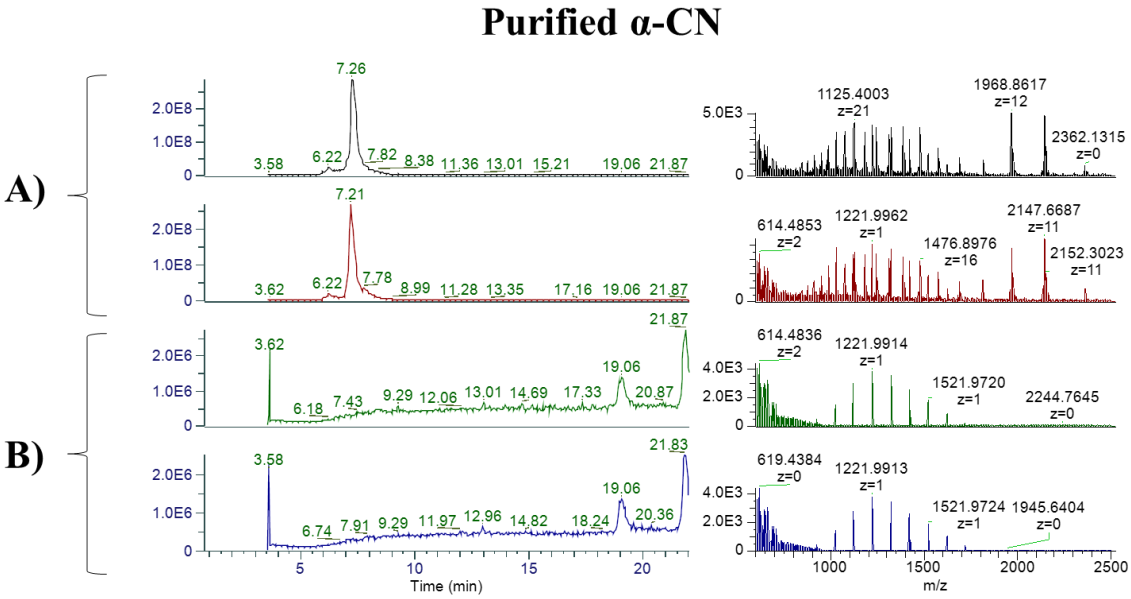


Figure 2.34: Chromatography and Averaged Spectra of 50 µg/mL Purified α-CN to Evaluate Reduction and Alkylation Protein Loss of Sample.
A) 23mM Trisodium Citrate 0.1%FA **B)** 23mM Trisodium Citrate 21.43 mM total DTT 10.71 mM total IAA 0.1%FA in duplicate.

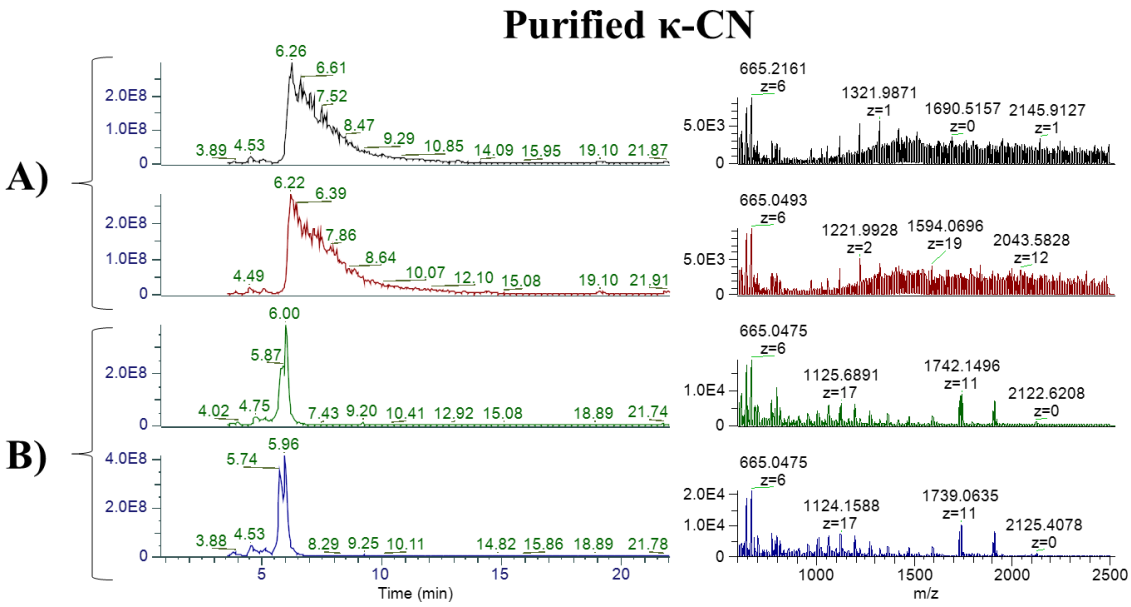


Figure 2.35: Chromatography and Averaged Spectra of 50 µg/mL Purified κ-CN to Evaluate Reduction and Alkylation Protein Loss of Sample.
A) 23mM Trisodium Citrate 0.1%FA **B)** 23mM Trisodium Citrate 21.43 mM total DTT 10.71 mM total IAA 0.1%FA in duplicate.

To investigate the heating step as the culprit for lack of protein found in the reduced and alkylated samples, completed through SDS-PAGE gels, verified that the heat step was not the cause of precipitation and protein loss (Figure 2.36). To note, there was a slight loss of protein when heated, but bands were very comparable. Due to the noticeable protein loss with use of reduction and alkylation steps and that the heating step was not the sole cause of protein loss, were removed from the sample preparation method.

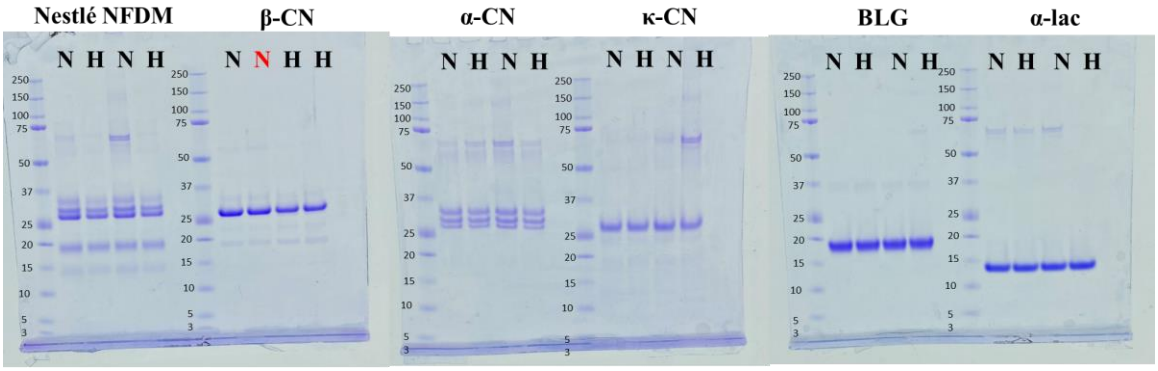


Figure 2.36: Reduced and Non-Redcued SDS-Page Gels of Nestlé NFDM and Purified Proteins (αS1-, αS2-, β-, κ-CN, BLG, and α-lac) to Evaluate the Heating Step in Sample Preparation for Protein Loss.
Lanes 1-4 are non-reduced, and lanes 5-8 are reduced with β-mercaptoethanol. Proteins with cystines: α_{s2}-CN (2), κ-CN (2-3), BLG (4-5), α-lac (8).

iv. Final Liquid Chromatography Optimization of All Commercial Samples

With the switch to 50 mM ammonium bicarbonate (Puriss, Fluka® Analytical), the concentration of 50 µg/mL was chosen between 50 µg/mL and 500 µg/mL. Further investigations with selected commercial samples warranted both water and 50 mM ammonium bicarbonate to be used. The use of 50 mM ammonium bicarbonate caused the peak between five to ten minutes shown in the water samples chromatography to be split into multiple peaks (50 µg/mL protein loading concentration shown) (Figure 2.37). With the addition of deconvoluting the spectra, the evaluation of the two final buffers resulted in a similar protein identification profile for water and 50 mM ammonium bicarbonate samples were (Table 2.8).

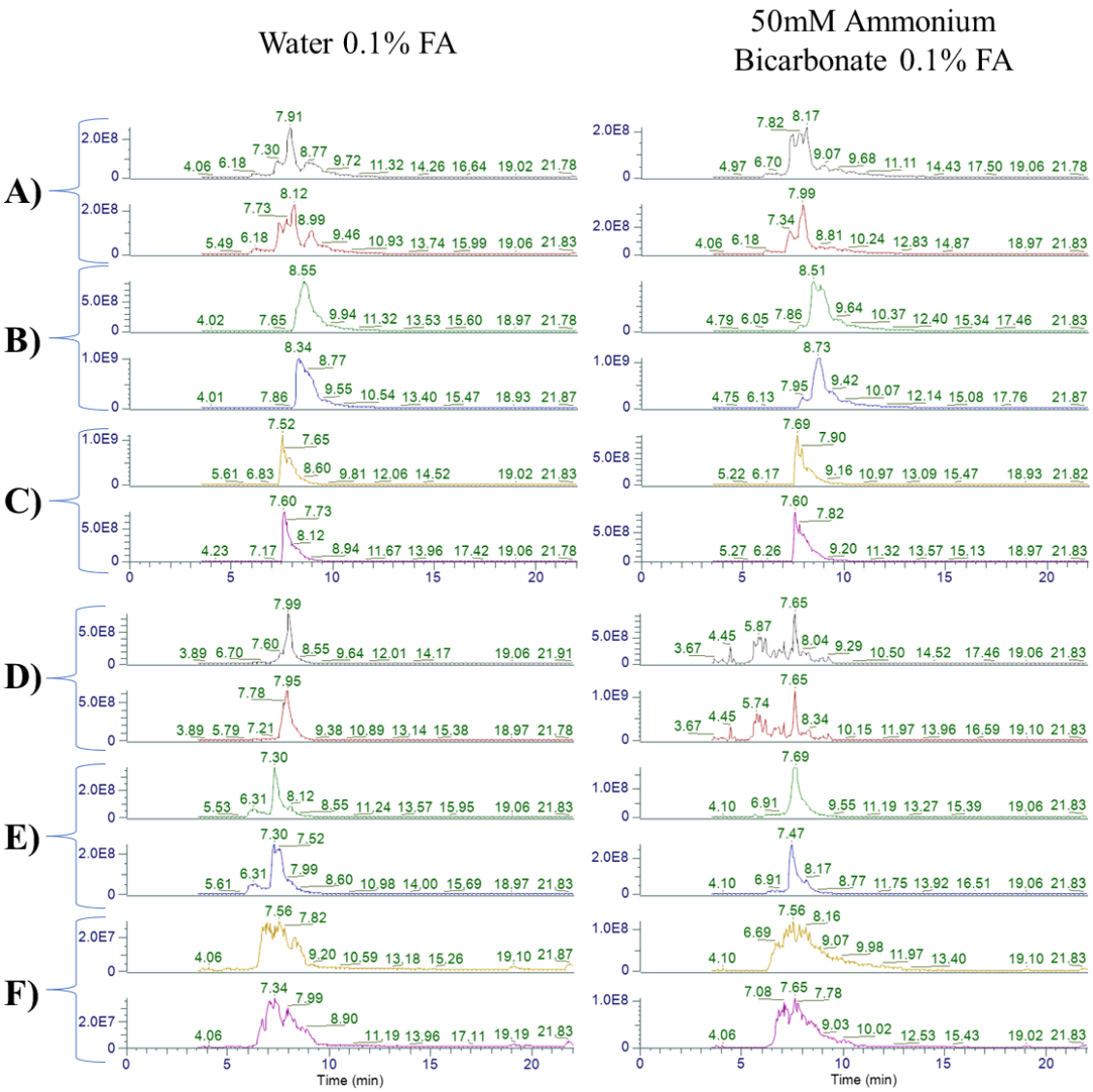


Figure 2.37: Chromatography Comparison of 50 µg/mL Nestlé NFDM and Purified Proteins (αS1-, αS2-, β-, κ-CN, BLG, and α-lac) to Evaluate Optimal Buffer. **A)** Nestlé NFDM **B)** BLG **C)** α-lac **D)** β-CN **E)** α-CN **F)** κ -CN.

Table 2.8: Comparison of 50 and 500 µg/mL in Nestlé NFDM and Purified Proteins in water or 50 mM Ammonium Bicarbonate.

Sample	50 µg/mL Water	50 µg/mL 50 mM Ammonium Bicarbonate	500 µg/mL Water	500 µg/mL 50 mM Ammonium Bicarbonate
Nestlé NFDM 1	20	25	20	17
Nestlé NFDM 2	20	23	16	18
Purified BLG 1	9	9	10	7
Purified BLG 2	8	10	9	10
Purified α-lac 1	4	6	5	5
Purified α-lac 2	2	4	4	5
Purified β-CN 1	5	5*	9	6*
Purified β-CN 2	4	5*	4	4*
Purified α-CN 1	6	4	6	4
Purified α-CN 2	4	5	5	6
Purified κ-CN 1	1*	1*	2*	1*
Purified κ-CN 2	1*	1*	1*	2*

*Could not identified and proteins amongst the deconvoluted masses in that specific protein category

The completed analysis of powdered and HTST samples resulted in the overall average percent fractional abundance of samples resolubilized in water was 84.48% and samples resolubilized in 50 mM ammonium bicarbonate was 85.16% with the exclusion of Serenity A2 Formula and Enfamil A2 Formula (Table 2.9). The high percent of fractional abundance gives confidence not only in the predicted monoisotopic mass database that it is extensive enough to be able to identify the abundant proteins and their proteoforms, but percent fractional abundance of total signal became a marker of whether a sample was suitable for intact LC-MS. In addition to this, both A1-type and A2-type β-CNs were able to be identified consistently between samples and replicates.

Table 2.9: Average Percent Fractional Abundance of the Total Signal of Identified Proteins in Commercial Samples in a Final Buffer of Water or 50 mM Ammonium Bicarbonate

	Resolubilized in Water					Resolubilized in 50mM Ammonium Bicarbonate				
Sample	**a**	**b**	**c**	**AVG**	**STD**	**a**	**b**	**c**	**AVG**	**STD**
Nestlé NFDM	97.10	95.45	94.70	95.75	1.23	94.64	96.07	94.71	95.14	0.81
Serenity A2 Formula	32.18	32.36	31.57	32.04	0.41	24.13	23.98	23.29	23.8	0.45
a2 company WMP	93.72	94.66	90.23	92.87	2.33	85.53	90.6	94	90.04	4.26
Similac A2 Formula	77.85	80.73	78.43	79.00	1.52	80.75	79.55	76.52	78.94	2.18
Enfamil A2 Formula	34.18	40.56	42.89	39.21	4.51	25.66	28.15	30.82	28.21	2.58
Gerber A2 Formula	87.68	81.35	78.85	82.63	4.55	85.92	85.86	80.81	84.20	2.93
Azure A2 WMP	93.13	89.59	84.28	89.00	4.45	84.34	86.67	86.78	85.93	1.38
Hiland HTST Milk	75.79	78.9	79.93	78.21	2.16	84.71	80.3	83.31	82.77	2.25
Alexandre A2 HTST Milk	80.26	84.94	85.65	83.62	2.93	85.72	79.7	83.83	83.08	3.08
a2 company A2 HTST milk	81.27	70.72	72.21	74.73	5.71	74.53	72.16	87.88	81.21	9.44

The statistical analysis between the use of water or 50 mM ammonium bicarbonate solvent resulted in the latter chosen due to the difference between the two averages of A1- and A2-type β-CNs. Both percent fractional abundance and sum intensities for each sample were analyzed with the exclusion of Serenity A2 Formula and Enfamil A2 Formula. The use of fractional abundance of A2-type β-CNs for statistical analysis, all samples excluding Serenity A2 Formula and Enfamil A2 Formula were found to be significantly different statistically (Table 2.10).

Table 2.10: Sum Intensity p-values for Selected Commercial Samples of A1- and A2-Type β-CNs to Evaluate Optimal Sample Buffer.

	% ID of A1	% ID of A2	% ID of A1 & A2
Nestle NFDM	0.5821	0.01485	0.0731
A2 Serenity Infant Formula	0.0001517	0.2902	0.0007935
A2 company whole milk powder	N/A	0.01399	0.01399
A2 Similac Infant Formula	N/A	0.02279	0.02279
A2 Enfamil Infant Formula	0.102	0.7702	0.6486
A2 Gerber Infant Formula	N/A	0.04488	0.04488
A2/A2 Azure whole milk powder	N/A	0.02472	0.02472
Hiland Liquid Milk	0.2536	0.006454	0.0172
A2 Alexandre Liquid Milk	N/A	0.01289	0.01289
A2 company liquid milk	N/A	0.03923	0.03923

Using p-value information in combination with the knowledge that the percent identification (fractional abundance) of A2-type β-CNs resuspended in 50 mM ammonium bicarbonate was greater than samples resuspended in water, 50 mM ammonium bicarbonate was chosen to be the buffer (Table 2.11).

Table 2.11: Average Identified A2-Type Percent Fractional Abundance of the Total Signal of a Sample to Evaluate Water or 50 mM Ammonium Bicarbonate Buffers.

	Resolubilized in Water					Resolubilized in 50mM Ammonium Bicarbonate				
Sample	**a**	**b**	**c**	**AVG**	**STD**	**a**	**b**	**c**	**AVG**	**STD**
Nestlé NFDM	6.23	7.18	5.75	6.39	0.73	9.14	9.86	11.47	10.16	1.19
a2 company A2 WMP	16.84	16.93	17.68	17.15	0.46	24.23	28.43	24.45	25.70	2.36
Similac A2 Formula	8.08	24.8	24.43	19.10	9.55	51.93	37.72	45.19	44.95	7.10
Gerber A2 Formula	22.14	15.92	22.83	20.30	3.80	27.06	31.66	27.10	28.61	2.64
Azure A2 WMP	27.85	29.86	28.04	28.58	1.11	38.61	33.36	36.74	36.24	2.66
Hiland HTST Milk	8.39	7.38	8.35	8.04	0.57	10.39	10.82	11.84	11.02	0.75
Alexandre A2 HTST Milk	25.92	28.42	24.82	26.39	1.85	33.05	31.35	28.65	31.02	2.22
a2 company A2 HTST milk	26.36	31.16	28.65	28.72	2.40	34.93	34.93	32.5	34.12	1.40

With the addition of more normal commercial products prepped with a similar same sample preparation method, a clearer picture of method success was confirmed. An average of 85.34% (±7.83%) fractional abundance of proteins identified was the result across many different types of samples and protein profiles (Table 2.12).

Table 2.12: Evaluation of Average Identified A2-type Percent Fractional Abundance of the Total Signal of a Sample in 50 mM Ammonium Bicarbonate.

	a	b	c	AVG	STD
Nestlé NFDM	96.97	96.03	97.20	96.73	0.62
HyVee NFDM	88.48	90.26	92.54	90.43	2.04
A2 Company A2 WMP	92.73	91.94	92.58	92.42	0.42
Azure A2 WMP	91.30	89.34	91.15	90.60	1.09
Serenity A2 Formula	27.14	28.44	24.23	26.60	2.16
Similac A2 Formula	80.15	85.5	78.53	81.39	3.65
Similac Formula	60.01	81.27	64.45	68.58	11.21
Gerber A2 Formula	83.3	88.08	81.52	84.30	3.39
Gerber Formula	0	0	0	0	0
Enfamil A2 Formula	32.74	31.79	31.60	32.04	0.61
Enfamil Formula	32.70	31.02	29.54	31.09	1.58
Hiland HTST Milk	88.59	87.31	88.19	88.03	0.65
HyVee HTST Milk	85.52	85.87	83.32	84.90	1.38
Alexandre A2 HTST Milk	82.05	86.41	85.54	84.67	2.31
A2 Company A2 HTST Milk	74.72	71.8	83.54	76.69	6.11

When sum intensities of total signal were used to determine statistically significance between use of water and 50 mM ammonium bicarbonate, there were high p-values, as shown in Table 2.13, meaning there was no difference between the two, however, the intensities of samples resolubilized in 50 mM ammonium bicarbonate were higher (Table 2.14).

Table 2.13: Sum Intensity p-values of A1- and A2-Type β-CNs in Selected Commercial Products.

	% ID of A1	% ID of A2
Nestlé NFDM	0.72610	0.13900
a2 company A2 WMP	N/A	0.03295
Similac A2 Formula	N/A	0.13360
Gerber A2 Formula	N/A	0.51470
Azure A2 WMP	N/A	0.95500
Hiland HTST Milk	0.24130	0.07212
Alexandre A2 HTST Milk	N/A	0.20590
a2 Company HTST milk	N/A	0.55990

Table 2.14: Average Identified A1- and A2-Type Sum Intensities of the Total Signal of Selected Commercial Products in Water or 50 mM Ammonium Bicarbonate.

	Resolubilized in Water					Resolubilized in 50mM Ammonium Bicarbonate				
Sample	**a**	**b**	**c**	**AVG**	**STD**	**a**	**b**	**c**	**AVG**	**STD**
Nestlé NFDM	5281284.25	6828358.52	5172619.88	5760754.22	926167.478	343712.7	1921524.9	2099548.7	1454928.8	966449.148
a2 company A2 WMP	7866955.63	9132852.38	10181679.6	9060495.87	1159057.11	11759992.2	12741765	11006825.9	11836194	869976.14
Similac A2 Formula	834473.1	3647906.42	4009343.72	2830574.41	1738095.1	7055628.9	3614871.3	5829265.02	5499921.7	1743861.65
Gerber A2 Formula	8940165.44	5712368.62	9852142.55	8168225.54	2175166.77	8348534.07	10532967	8762157.57	9214553	1160359.18
Azure A2 WMP	10203034.7	12264865.9	10175287.2	10881062.6	1198489.14	12886670.8	9407524.3	10573386	10955860	1770827.51
Hiland HTST Milk	27125325.7	25773024	27519658.3	26806002.7	916055.803	26459672.6	30830911	31300130.2	29530238	2669516.8
Alexandre A2 HTST Milk	14691985	14099442.2	12647287.4	13812904.9	1052033.67	16102172.4	13780232	15975959.5	15286121	1305664.2
a2 company A2 HTST milk	16308171.1	21831352.6	20120474.8	19419999.5	2827433.9	21129641.1	19458219	21143662.1	20577174	969068.906

Statistical analysis of percent identified, and sum intensities confirmed the success of the method to be used to analyze data and calculate A1- and A2-type abundances within product labeled normally and products labeled as A2 products. Overall, between the two finalized sample preparations with 50 mM ammonium bicarbonate experiments, a mean of 85.27% (± 6.68%, n = 57) of the total signal of powdered and HTST milk could be assigned to the predicted monoisotopic database proteoforms.

V. SUMMARY

An informatic database search for a collection of all known major milk protein proteoforms was completed to create a predicted monoisotopic mass database of over 1000 intact masses. The development for conditions of intact mass spectrometry was developed using fifteen commercial products. This resulted in an established sample preparation method involving the centrifugation to defat diluted samples (1 mg protein/mL), use of molecular weight spin filters to clean up and desalt, and remobilization and dilution into 50 mM ammonium bicarbonate 0.1% (v/v) FA. Sample protein was separated and analyzed by ultra-high performance liquid chromatography mass spectrometry (UHPLC-MS). Raw data was deconvoluted to generate a list of monoisotopic masses. Mass events of 6-27 kDa were analyzed against the predicted monoisotopic mass database (mass error tolerance = 10 ppm). A mean of 85.27% (± 6.68%, n = 57) of the total signal of powdered and liquid HTST milk could be assigned to the predicted monoisotopic mass database proteoforms. Of most of these samples, BLG, α-lac, β-CN, and α-CNs were continually detected and identified. The method could not detect κ-CN and samples that used hydrolyzed proteins due to the intact protein focus and resolution limitations of the method.

VI. REFERENCES

1. Jinsmaa, Y. and M. Yoshikawa, *Enzymatic release of neocasomorphin and β-casomorphin from bovine β-casein.* Peptides, 1999. **20**(8): p. 957-962.
2. Küllenberg de Gaudry, D., S. Lohner, C. Schmucker, P. Kapp, E. Motschall, S. Hörrlein, C. Röger, and J.J. Meerpohl, *Milk A1 β-casein and health-related outcomes in humans: a systematic review.* Nutrition Reviews, 2019. **77**(5): p. 278-306.
3. Detz, J. *What Is a2 Milk?* 2019 03/22/23]; Available from: https://www.consumerreports.org/milk/what-is-a2-milk/.
4. Jianqin, S., X. Leiming, X. Lu, G.W. Yelland, J. Ni, and A.J. Clarke, *Effects of milk containing only A2 beta casein versus milk containing both A1 and A2 beta casein proteins on gastrointestinal physiology, symptoms of discomfort, and cognitive behavior of people with self-reported intolerance to traditional cows' milk.* Nutrition Journal, 2016. **15**(1): p. 35.
5. consortium, U., *UniProt: the Universal Protein Knowledgebase in 2023*. 2023.
6. Strohalm, M., *mMass - Open Source Mass Spectrometry Tool*. 2005 - 2013.
7. Johnson, P., M. Philo, A. Watson, and E.N. Clare Mills, *Rapid Fingerprinting of Milk Thermal Processing History by Intact Protein Mass Spectrometry with Nondenaturing Chromatography*. Journal of Agricultural and Food Chemistry, 2011. **59**(23): p. 12420-12427.
8. Farrell, H.M., R. Jimenez-Flores, G.T. Bleck, E.M. Brown, J.E. Butler, L.K. Creamer, C.L. Hicks, C.M. Hollar, K.F. Ng-Kwai-Hang, and H.E. Swaisgood, *Nomenclature of the Proteins of Cows' Milk—Sixth Revision.* Journal of Dairy Science, 2004. **87**(6): p. 1641-1674.
9. Han, S.K., Y.C. Shin, and H.D. Byun, *Biochemical, molecular and physiological characterization of a new beta-casein variant detected in Korean cattle.* Animal genetics, 2000. **31**(1): p. 49-51.
10. Grosclaude, F., M.F. Mahe, and G.F. Voglino, *Le variant beta E et le code de phosphorylation des caséines bovines [The beta E variant and the phosphorylation code of bovine caseins].* FEBS letters, 1974. **45**(1): p. 3-5.

11. Bonsing, J., J.M. Ring, A.F. Stewart, and A.G. Mackinlay, *Complete nucleotide sequence of the bovine beta-casein gene.* Australian journal of biological sciences, 1988. **41**(4): p. 527–537.
12. Schmelzer, C.E., R. Schöps, L. Reynell, R. Ulbrich-Hofmann, R.H. Neubert, and K. Raith, *Peptic digestion of beta-casein. Time course and fate of possible bioactive peptides.* Journal of chromatography, 2007. **1166**(1-2): p. 108-115.
13. Jimenez-Flores, R., Y.C. Kang, and T. Richardson, *Cloning and sequence analysis of bovine beta-casein cDNA.* Biochemical and biophysical research communications, 1987. **142**(2): p. 617–621.
14. Ribadeau Dumas, B., G. Brignon, F. Grosclaude, and J.C. Mercier, *Structure primaire de la caséine beta bovine. Séquence complète [Primary structure of bovine beta casein. Complete sequence].* European journal of biochemistry, 1972. **25**(3): p. 505–514.
15. Ribadeau Dumas, B., F. Grosclaude, and J.C. Mercier, *Localisation dans la chaîne peptidique de la caséine bovine de la substitution His=Gln différenciant les variants génétiques A 2 et A 3 [Localization in the peptide chain of bovine beta casein of the His-Gln substitution differentiating the A2 and A3 genetic variants].* Comptes rendus hebdomadaires des seances de l'Academie des sciences. Serie D: Sciences naturelles, 1970. **270**(19): p. 2369–2372.
16. Jones, D.S., W. Heerma, P.D. van Wassenaar, and J. Haverkamp, *Analysis of bovine beta-casein tryptic digest by continuous-flow fast-atom bombardment mass spectrometry.* Rapid communications in mass spectrometry : RCM, 1991. **5**(4): p. 92–195.
17. Carles, C., J.C. Huet, and B. Ribadeau-Dumas, *A new strategy for primary structure determination of proteins: application to bovine beta-casein.* FEBS letters, 1988. **229**(2): p. 265–272.
18. Wu, S.L., J. Kim, W.S. Hancock, and B. Karger, *Extended Range Proteomic Analysis (ERPA): a new and sensitive LC-MS platform for high sequence coverage of complex proteins with extensive post-translational modifications-comprehensive analysis of beta-casein and epidermal growth factor receptor (EGFR).* Journal of proteome research, 2005. **4**(4): p. 1155–1170.

19. Imanishi, S.Y., V. Kochin, Ferraris, S. E., , A. de Thonel, H.M. Pallari, G.L. Corthals, and J.E. Eriksson, *Reference-facilitated phosphoproteomics: fast and reliable phosphopeptide validation by microLC-ESI-Q-TOF MS/MS.* Molecular & cellular proteomics : MCP, 2007. **6**(8): p. 1380–1391.
20. Brew, K., F.J. Castellino, T.C. Vanaman, and R.L. Hill, *The complete amino acid sequence of bovine alpha-lactalbumin.* The Journal of biological chemistry, 1970. **245**(17): p. 4570–4582.
21. Vanaman, T.C., K. Brew, and R.L. Hill, *The disulfide bonds of bovine alpha-lactalbumin.* The Journal of biological chemistry, 1970. **245**(17): p. 4583–4590.
22. Gordon, W.G., R. Aschaffenburg, A. Sen, and S.K. Ghosh, *Amino acid composition of several alpha-lactalbumins.* Journal of Dairy Science, 1968. **51**: p. 947.
23. Shewale, J.G., S.K. Sinha, and K. Brew, *volution of alpha-lactalbumins. The complete amino acid sequence of the alpha-lactalbumin from a marsupial (Macropus rufogriseus) and corrections to regions of sequence in bovine and goat alpha-lactalbumins.* The Journal of biological chemistry, 1984. **259**(8): p. 4947-4956.
24. Brignon, C. and B.R. Dumas, *Localisation dans la chaine peptidique de la beta lactoglobuline bovine de la substitution Glu-Gln differenciant les variants genetiques B et [Localization of the Glu-Gln substitution differentiating B and D genetic variants in the peptide chain of bovine beta lactoglobulin].* FEBS letters, 1973. **33**(1): p. 73-76.
25. Godovac-Zimmermann, J., I. Krause, J. Buchberger, G. Weiss, and H. Klostermeyer, *Genetic variants of bovine beta-lactoglobulin. A novel wild-type beta-lactoglobulin W and its primary sequence.* Biological chemistry Hoppe-Seyler, 1990. **371**(3): p. 255-260.
26. Alexander, L.J., G. Hayes, M.J. Pearse, C.W. Beattie, A.F. Stewart, I.M. Willis, and A.G. & Mackinlay, *Complete sequence of the bovine beta-lactoglobulin cDNA.* Nucleic acids research, 1989. **17**(16): p. 6739.

27. Jamieson, A.C., M.A. Vandeyar, Y.C. Kang, J.E. Kinsella, and C.A. Batt, *Cloning and nucleotide sequence of the bovine beta-lactoglobulin gene.* Gene, 1987. **61**(1): p. 85-90.

28. Braunitzer, G., R. Chen, B. Schrank, and A. Stangl, *Die Sequenzanalyse des beta-Lactoglobulins [The sequence of beta-lactoglobulin (author's transl)].* Hoppe-Seyler's Zeitschrift fur physiologische Chemie, 1973. **354**(8): p. 867–878.

29. Willis, I.M., A.F. Stewart, A. Caputo, A.R. Thompson, and A.G. Mackinlay, *Construction and identification by partial nucleotide sequence analysis of bovine casein and beta-lactoglobulin cDNA clones.* DNA (Mary Ann Liebert, Inc.), 1982. **1**(4): p. 375-386.

30. Kuwata, K., M. Hoshino, V. Forge, S. Era, C.A. Batt, and Y. Goto, *olution structure and dynamics of bovine beta-lactoglobulin A.* Protein science : a publication of the Protein Society, 1999. **8**(11): p. 2541–2545.

31. Grosclaude, F., M.F. Mahe, M. J.C., and B. Ribadeau-Dumas, *Localization of amino-acid substitutions that differenciate bovine kappa-casein variants A and B.* Ann. Genet. Sel. Anim., 1972. **4**: p. 515-521.

32. Rasmussen, L.K., P. Højrup, and T.E. Petersen, *The multimeric structure and disulfide-bonding pattern of bovine kappa-casein. .* European journal of biochemistry, 1992. **207**(1): p. 215-222.

33. Mercier, J.C., G. Brignon, and B. Ribadeau-Dumas, *Structure primaire de la caséine kappa B bovine. Séquence complète [Primary structure of bovine kappa B casein. Complete sequence].* European journal of biochemistry, 1973. **35**(2): p. 222-235.

34. Reid, J.R., T. Coolbear, C.J. Pillidge, and G.G. Pritchard, *Specificity of hydrolysis of bovine kappa-casein by cell envelope-associated proteinases from Lactococcus lactis strains.* Applied and environmental microbiology, 1994. **60**(3): p. 801–806.

35. Jollès, J., F. Schoentgen, C. Alais, A.M. Fiat, and P. Jollès, *Studies on the primary structure of cow kappa-casein. Structural features of para-kappa-casein; N-terminal sequence of kappa-caseinoglycopeptide studied with a sequencer. .* Helvetica chimica acta, 1972. **55**(8): p. 2872–2883.

36. Guérin, J., C. Alais, J. Jollès, and P. Jollès, *Kappa-casein from bovine colostrum.* Biochimica et biophysica acta, 1974. **351**(2): p. 325–332.

37. Mercier, J.C., F. Grosclaude, and B. Ribadeau-Dumas, *Structure primaire de la caséine s1 -bovine. Séquence complète [Primary structure of bovine s1 casein. Complete sequence].* European journal of biochemistry, 1971. **23**(1): p. 41-51.

38. Grosclaude, F., M.F. Mahé, J.C. Mercier, and B. Ribadeau-Dumas, *Caractérisation des variants génétiques des caséines sl et bovines [Characterization of genetic variants of a S1 and bovine caseins].* European journal of biochemistry, 1972. **26**(3): p. 328–337.

39. Larsen, M.R., T.E. Thingholm, O.N. Jensen, P. Roepstorff, and T.J. Jørgensen, *Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns.* Molecular & cellular proteomics, 2005. **4**(7): p. 873–886.

40. Grosclaude, F., M.F. Mahé, and B. Ribadeau-Dumas, *Structure primaire de la caséine alphasl et de la caséine beta bovines. Correctif [Primary structure of alpha casein and of bovine beta casein. Correction].* European journal of biochemistry, 1973. **40**(1): p. 323-324.

41. Grosclaude, F., M.F. Mahé, J.C. Mercier, and B. Ribadeau-Dumas, *ocalisation, dans la partie NH(2)-terminale de la caséine alpha(s1) bovine, d'une délétion de 13 acides aminés différenciant le variant a des variants B et C.* FEBS letters, 1970. **11**(2): p. 109–112.

42. Neuteboom, B., M.G. Giuffrida, and A. Conti, *Isolation of a new ligand-carrying casein fragment from bovine mammary gland microsomes.* FEBS letters, 1992. **305**(3): p. 189–191.

43. Ramalho-Santos, M., P. Veríssimo, C. Faro, and E. Pires, *Action on bovine alpha s1-casein of cardosins A and B, aspartic proteinases from the flowers of the cardoon Cynara cardunculus L.* Biochimica et biophysica acta, 1996. **1297**(1): p. 83-89.

44. Brignon, G., B. Ribadeau Dumas, J.C. Mercier, J.P. Pelissier, and B.C. Das, *Complete amino acid sequence of bovine alphaS2-casein.* FEBS letters, 1977. **76**(2): p. 274–279.

45. Zucht, H.D., M. Raida, K. Adermann, H.J. Mägert, and W.G. Forssmann, *Casocidin-I: a casein-alpha s2 derived peptide exhibits antibacterial activity.* FEBS letters, 1995. **372**(2-3): p. 185-188.
46. Fox, P., *Milk: An overview*. 2008. p. 1-54.
47. Miranda, G., L. Bianchi, Z. Krupova, P. Trossat, and P. Martin, *An improved LC–MS method to profile molecular diversity and quantify the six main bovine milk proteins, including genetic and splicing variants as well as post-translationally modified isoforms.* Food Chemistry: X, 2020. **5**: p. 100080.
48. Vincent, D., A. Elkins, M.R. Condina, V. Ezernieks, and S. Rochfort, *Quantitation and Identification of Intact Major Milk Proteins for High-Throughput LC-ESI-Q-TOF MS Analyses.* PLOS ONE, 2016. **11**(10): p. e0163471.
49. Bordin, G., F. Cordeiro Raposo, B. de la Calle, and A.R. Rodriguez, *Identification and quantification of major bovine milk proteins by liquid chromatography.* Journal of chromatography. A, 2001. **928**(1): p. 63-76.
50. Frederiksen, P.D., K.K. Andersen, M. Hammershøj, H.D. Poulsen, J. Sørensen, M. Bakman, K.B. Qvist, and L.B. Larsen, *Composition and effect of blending of noncoagulating, poorly coagulating, and well-coagulating bovine milk from individual Danish Holstein cows.* Journal fo Dairy Science, 2011. **94**(10): p. 4787-4799.
51. Bonfatti, V., L. Grigoletto, A. Cecchinato, L. Gallo, and P. Carnier, *Validation of a new reversed-phase high-performance liquid chromatography method for separation and quantification of bovine milk protein genetic variants.* Journal of Chromatography A, 2008. **1195**(1-2): p. 101-106.
52. Scientific, T. *Reversed Phase HPLC and UHPLC Columns*. 2006-2023 [cited 2023; Available from: https://www.thermofisher.com/us/en/home/industrial/chromatography/liquid-chromatography-lc/hplc-uhplc-columns/reversed-phase-hplc-columns.html#menu1.
53. Scientific, T., *BioPharma Finder User Guide, software version 3.0, XCALI-97898 Revision A.* 2017.

54. Scientific, T. *HPLC Guard Columns, Guard Cartridges, and Accessories*. 2006-2023; Available from: https://www.thermofisher.com/us/en/home/industrial/chromatography/liquid-chromatography-lc/hplc-uhplc-columns/hplc-guard-columns-cartridges-accessories.html#:~:text=HPLC%20guard%20columns%20are%20short,affinity%20for%20the%20stationary%20phase.

55. Hedges, J.B., S. Vahidi, X. Yue, and L. Konermann, *Effects of Ammonium Bicarbonate on the Electrospray Mass Spectraof Proteins: Evidence for Bubble-Induced Unfolding.* Analytical chemistry, 2013. **85**(13): p. 6469–6476.

56. Information, N.C.f.B. *PubChem Compound Summary for CID 14013, Ammonium Bicarbonate*. 2023 [cited 2023 October 26]; Available from: https://pubchem.ncbi.nlm.nih.gov/compound/Ammonium-Bicarbonate.

57. McGillivray, R.T., D.W. Chung, and E.W. Davie, *Biosynthesis of bovine plasma proteins in a cell-free system. Amino-terminal sequence of preproalbumin.* European journal of biochemistry, 1979. **98**(2): p. 477–485.

58. Patterson, J.E. and D.M. Geller, *Bovine microsomal albumin: amino terminal sequence of bovine proalbumin.* Biochemical and biophysical research communications, 1977. **74**(3): p. 1220–1226.

59. Hirayama, K., S. Akashi, M. Furuya, and K. Fukuhara, *Rapid confirmation and revision of the primary structure of bovine serum albumin by ESIMS and Frit-FAB LC/MS.* Biochemical and biophysical research communications, 1990. **173**(2): p. 639–646.

60. Strawich, E. and M.J. Glimcher, *Tooth 'enamelins' identified mainly as serum proteins. Major 'enamelin' is albumin.* European journal of biochemistry, 1990. **191**(1): p. 47-56.

61. Carraway, R.E., D.E. Cochrane, W. Boucher, and S.P. Mitra, *Structures of histamine-releasing peptides formed by the action of acid proteases on mammalian albumin(s).* Journal of immunology, 1989. **143**(5): p. 1680–1684.

62. Reed, R.G., F.W. Putnam, and T. Peters, Jr *Sequence of residues 400--403 of bovine serum albumin.* The Biochemical journal, 1980. **191**(3): p. 867–868.

63. Scientific, T., *Protein Clean Up - Sample Handling*. p. 1-6.

64. Huang, L.-J., J.-H. Lin, J.-H. Tsai, Y.-Y. Chu, Y.-W. Chen, S.-L. Chen, and S.-H. Chen, *Identification of protein O-glycosylation site and corresponding glycans using liquid chromatography–tandem mass spectrometry via mapping accurate mass and retention time shift.* Journal of Chromatography A, 2014. **1371**: p. 136-145.

65. Pisano, A., N.H. Packer, J.W. Redmond, K.L. Williams, and A.A. Gooley, *Characterization of O-linked glycosylation motifs in the glycopeptide domain of bovine κ-casein.* Glycobiology, 1994. **4**(6): p. 837-844.

66. Raghav, A. and J. Ahmad, *Introductory Chapter: Glimpses of Glycosylation.* Fundamentals of Glycosylation, ed. A.R.a.J. Ahmad. 2022.

67. Minkiewicz, P., C.J. Slangen, F.M. Lagerwerf, J. Haverkamp, H.S. Rollema, and S. Visser, *Reversed-phase high-performance liquid chromatographic separation of bovine κ-casein macropeptide and characterization of isolated fractions.* Journal of Chromatography A, 1996. **743**(1): p. 123-135.

68. Fuerer, C., R. Jenni, L. Cardinaux, F. Andetsion, S. Wagnière, J. Moulin, and M. Affolter, *Protein fingerprinting and quantification of β-casein variants by ultra-performance liquid chromatography–high-resolution mass spectrometry.* Journal of Dairy Science, 2020. **103**(2): p. 1193-1207.

69. Scientific, T., *Pierce™ BCA Protein Assay Kit User Guide*. 2020.

70. Murray, K.K., *Resolution and Resolving Power in Mass Spectrometry.* Journal of the American Society for Mass Spectrometry, 2022. **33**(12): p. 2342-2347.

CHAPTER III: EVALUATION OF INTACT MASS SPECTROMETRY METHOD

I. ABSTRACT

With the lack of an authentication method for A2 milk products themselves, a repeatable and robust analytical method is needed. This method will check that the marketed protein profile claims will match the product. Intact mass spectrometry has been previously used for mainly pharmaceutical purposes of testing purified monoclonal antibodies and in addition, can be used to monitor proteoforms within a complex mixture such as milk. Fifteen samples of powdered and HTST liquid milks with an equal distribution of marked A2 and non-A2 commercial products were selected. Each sample was diluted to 1 mg protein/mL in 50 mM ammonium bicarbonate and then defatted through centrifugation of 15 min at 3,000 x g. The samples were then cleaned up and desalted through 3 kDa spin column filters and then separated and analyzed by ultra-high performance liquid chromatography mass spectrometry (UHPLC-MS). Biopharma Finder deconvolution software was used on raw files to produce mass outputs that were compared to the predicted database and mass were identified. A mean of 85.27% (± 6.68%, n = 57) of the total signal of powdered and liquid HTST milk could be assigned to the predicted database proteoforms. Fifteen selected commercial products were evaluated, and the average ratio of selected normal commercial products was 25.86% A1 and 0.74.14% A2.

.

II. INTRODUCTION

Authentication of products marketed to possess specific protein sources or compositions can be a challenge for traditional methods that generally lack required protein specificity. An example of this situation is A2 milk products. A2 products exist in both powdered and liquid forms and are mainly targeted at parents who buy powdered infant formula. Presently, the establishment of the claim is established through the genetic testing of cows before milk production. This drives the need for further authentication of the product itself. Intact protein mass spectrometry (MS) has the potential to directly authenticate protein products, including specific proteoform claims. The development of an intact MS method to detect and differentiate major bovine milk proteins (α_{S1}-, α_{S2}-, β-, κ -caseins (CNs), β-lactoglobulins (BLG), and α-lactalbumins (α-lac)) and their proteoforms is needed to confirm A2 protein profile claims and help regulators define A2 milk. This work will demonstrate that intact protein MS can be an effective tool to analyze milk products for protein authentication in general.

As described in Chapter 2, an informatic approach to create a sequence database of the major milk proteins was first completed. Then, information on genetic and processing modifications were gathered to produce a predicted monoisotopic mass database of over a thousand monoisotopic masses attributed to a specific proteoform. The development for conditions of intact mass spectrometry resulted in an established sample preparation method involving the centrifugation of 1 mg/mL samples, clean up and desalting with molecular weight spin filters, and dilution into 50 mM ammonium bicarbonate and 0.1 FA. This resulted in BLG, α-lac, β-CN, and α-CNs continually detected and identified within most of the selected samples. The method could not detect

κ-CN and samples that used hydrolyzed proteins due to the intact protein focus and resolution limitations of the method.

III. MATERIALS AND METHODS

a. Summary of Intact Protein LC-MS Method Optimization

Commercial products both containing a normal profile (both A1- and A2-type β-CNs) and marketed A2-type β-CN profiles were selected as described in Chapter 2 (Table 2.1). Where possible, both normal and A2 products from the same manufacturers were obtained. In addition, lyophilized purified bovine proteins from Sigma-Aldrich® (α-, β-, κ-CN, BLG, and α-lac) were used as single-protein controls.

Table 2.1: Selected Commercial Samples of Normal and A2-type β-CN Protein Profiles

Product Type	Mixed A1- & A2-type β-CNs	Abbreviation	Marketed as " Only A2-type β-CNs"	Abbreviation
Powdered Milk	Nestlé Carnation Instant Nonfat Dry Milk	Nestlé NFDM	a2 Instant Milk Powder Full Cream 1 kg (Australia Import)	a2 company A2 WMP
	Hy-Vee Instant Nonfat Dry Milk	Hy-Vee NFDM	Azure Market A2/A2 Organic Dry Milk Powder	Azure Market A2 WMP
Powdered Infant Formula	Pure Bliss® by Similac® Infant Formula	Similac Formula	Similac Organic A2 Infant Formula Powder Toddler	SimilacA2 Formula
	Gerber Good Start Baby Formula Powder, GentlePro	Gerber Formula	Gerber Good Start Infant Formula with Iron, A2, 0 to 12 Months - 20 oz	Gerber A2 Formula
	Enfamil Infant formula	Enfamil Formula	Enfamil A2 Premium Infant Formula, Milk-Based Powder with Iron - 19.5 oz	Enfamil A2 Formula
			Serenity 12+ months A2 whole milk toddler formula	Serenity A2 Formula
HTST Milk	Hiland Vitamin D Milk	Hiland HTST milk	Whole milk by "The a2 milk company"	a2 company A2 HTST milk
	Hy-Vee Vitamin D Milk	Hy-Vee HTST milk	A2/a2 Organic 4% Whole Milk by Alexandre ECO DAIRY	Alexandre A2 HTST milk

A Thermo Scientific™ Q Exactive™ Plus hybrid quadrupole-Orbitrap™ mass spectrometer was used for all direct infusion analysis. The instrument parameters optimized for the developed intact protein LC-MS method (as described in Chapter 2) were as follows: Full scan MS method, resolution = 140,000 m/z, scan range = 600.0 to

2,300 m/z, fragmentation (IS-CID) = 25.0 eV, polarity = positive, sheath gas flow rate = 7, aux gas = 1, sweep gas flow rate = 5, spray voltage = 4.00, capillary temperature = 320◦C, AGC target = 1e6, maximum inject time = 200 milliseconds and microscans = 5.

The LC gradient was optimized prior resulting in a gradient of 13.5–72 % mobile phase B over 31 min at a flow rate of 250 µL.min-1 (See Figure 2.6). Analysis of sample on MS was scheduled between 2 and 23 minutes while the other was directed into waste to equilibrize the column and wash. Buffer A consisted of water containing 0.1% (v/v) FA and buffer B consisted of 100% (v/v) ACN containing 0.1% (v/v) FA. Samples were injected at 15 µL on-column.

Sample preparation was optimized to include a 15 min centrifugation at 3,000 x g defatting step, cleaning up and desalting the sample with a 3kDa Amicon® molecular weight (MW) cutoff spin column, and diluting into 50 mM ammonium bicarbonate 0.1% (v/v) FA (Figure 2.7). The protein loading chosen was 50 ug/mL in a final buffer of 50 mM ammonium bicarbonate containing 0.1% (v/v) FA. The raw data would then be deconvoluted with BioPharma Finder deconvolution software.

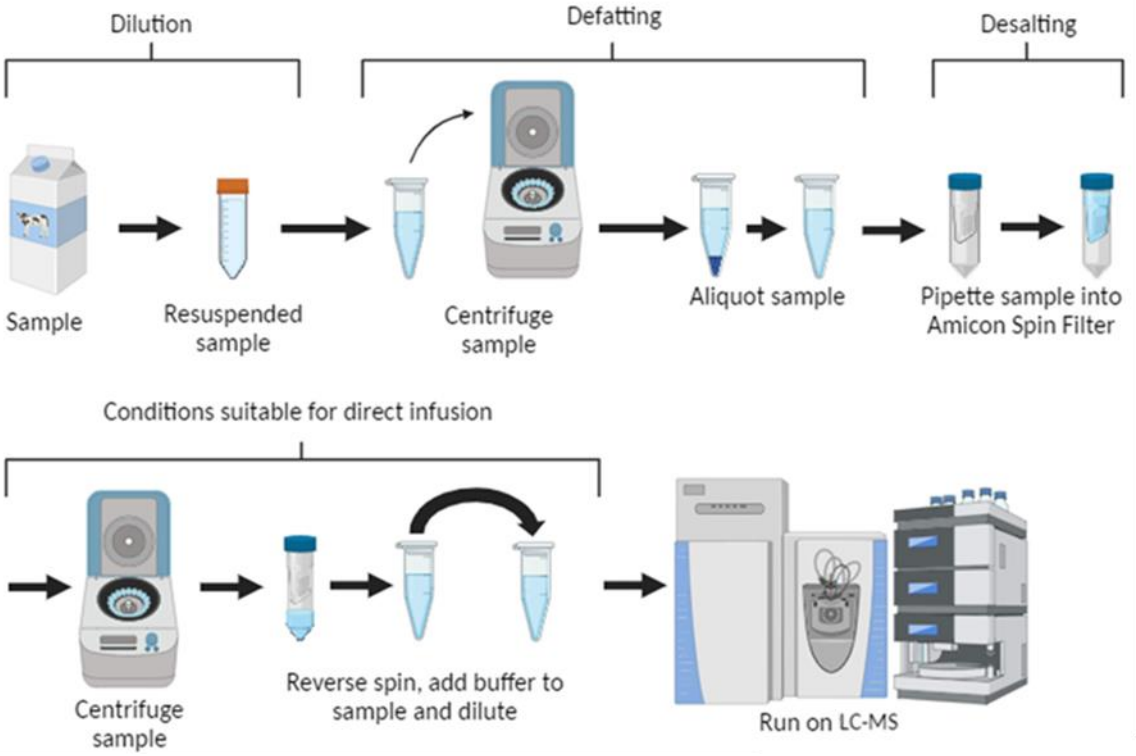


Figure 2.7: Intact Protein LC-MS Sample Preparation Summary. Created in BioRender.com.

The optimized deconvolution software chosen was BioPharma Finder Xtract sliding windows algorithm. The deconvolution parameters were set as follows for intact protein LC-MS deconvolution: output mass range = 3,000 – 30,000, output mass = M, S/N Threshold = 3, relative abundance threshold (%) = 5%, charge range = 3 – 50, minimum number detected charge = 3, fit factor = 80%, consider overlaps = true, minimum intensity = 1, expected intensity error = 3, source spectra method = sliding windows or average spectrum, target average spectrum width = 0.5 minutes, target average spectrum offset scan = 1, merge tolerance = 30 ppm, merge scheme = legacy merge scheme, number of detected intervals = 3. Deconvoluted masses would then be identified through the predicted monoisotopic database created for bovine milk proteoforms within a mass tolerance of 10 ppm. Signal intensity and fractional abundance

of total signal of identified proteins were recorded. The summarized intact protein LC-MS method is detailed in Figure 3.1.

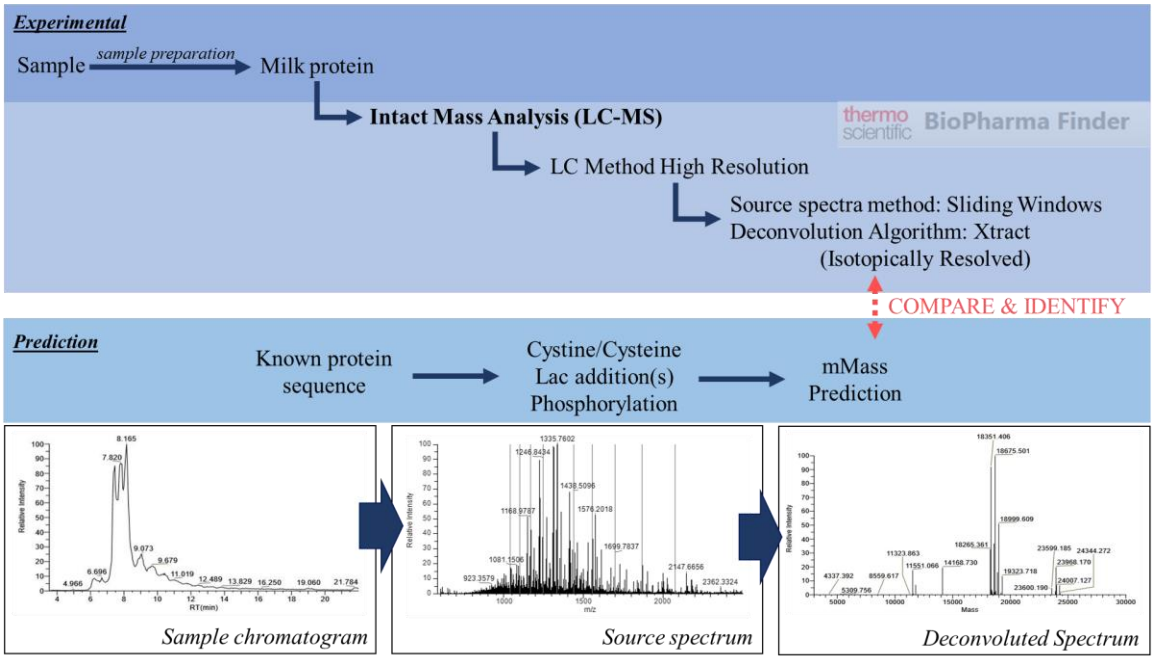


Figure 3.1: Summary Data Analysis Flowchart for Developed Intact Protein LC-MS Method.

b. Identified A1 to A2 Fraction of Total Signal Intensity of Selected Commercial Samples

The first set of commercial samples, as described above, were processed through the same sample preparation method and then deconvoluted with Biopharma Finder computer software as described above and in Chapter 2. These samples included Nestlé NFDM, a2 Company A2 WMP, Azure Market A2 WMP, Similac A2 Formula, Gerber A2 Formula, Enfamil A2 Formula, Serenity A2 Formula, Hiland HTST milk, a2 Company A2 HTST milk, and Alexandre A2 HTST milk. A new second set of commercial samples, including samples from the initial set and additional samples added, we prepared and analyzed in the same way (sample preparation, instrument settings, etc.).

These additional samples included Hy-Vee NFDM, Similac Formula, Gerber Formula, Enfamil Formula, and Hy-Vee HTST milk.

Proteins were identified through comparison of the predicted monoisotopic mass proteoform database. Identification was only considered if experimental monoisotopic mass was within ten parts per million (ppm) of the predicted monoisotopic mass. This gave high confidence that the deconvoluted mass was the predicted proteoform within the generated database. From there, signal intensity attached to the proteoform provided by BiopharmaFinder deconvolution software was recorded. To calculate the amount of A1- to A2-type β-CNs of the sample's total signal intensity, the sums of identified A1-type β-CNs, A2-type β-CN, and sum of both A1- and A2-type β-CNs were individually recorded. Taking the sum of the A1-type β-CNs over the total sum of A1- *and* A2-type β-CNs, a fraction, in the form of a decimal was calculated. The same was done for A2-type β-CNs. This was done for both the first and second selected commercial sample sets.

c. Dilution Series of Conventional Product in A2 Product

Nestlé NFDM and a2 Company A2 WMP were resolubilized in 50 mM ammonium bicarbonate at the concentration of 1 mg/mL. Then, a two-fold factor serial dilution of Nestlé NFDM into a2 Company powder was completed. This series started at 50% Nestlé NFDM diluted down to 0.390625% Nestlé NFDM. After the dilutions were finished, controls of both un-diluted products were run in tandem. Three replicate sets of the dilution series were prepared, and each dilution was injected three times. It should be noted that the controls, the un-diluted products, of each dilution set were run singularly (three injections of each product total).

d. Infant Formula Investigation

In the initial set of commercial samples analysis, two products, Enfamil A2 Formula and Serenity A2 Formula, were separated out from the other samples due to their low percent protein identity (via fractional abundance of total signal) of ~30%. This warranted further analysis of the cause behind the low protein identity.

i. SDS-PAGE Gels

In addition to the two powdered infant formulas in question, two control samples, Nestlé NFDM and Similac A2 Formula were analyzed as well. Nestlé NFDM was chosen due to its use throughout the method development process and Similac A2 Formula was chosen for its highest signal intensity A1- and A2-type β-CNs of the powdered formulas. Samples were analyzed using reduced and non-reduced NuPAGE™ 4-12% Bis-Tris Gels (1.00 mm x 12 well) (Invitrogen™, Thermo Scientific™) to assess the protein profiles between selected samples. The sample loaded into each well differed for each sample and detailed in Table 3.1.

Table 3.1. Reduced and Non-Reduced NuPAGE™ 4-12% Bis-Tris Gel Sample Load Specification

Resuspended directly in SDS-PAGE Buffer							
Sample	Reduced?	Sample (mg)	Water (µL)	BME (µL)	LDS (µL)	Total Volume (µL)	Loaded into well (µL)
Nestle	Y	1	299	1	100	400	10
	N	1	200	0	100	400	10
Similac	Y	5	299	1	100	400	10
	N	5	300	0	100	400	10
Enfamil	Y	2.3	299	1	100	400	10
	N	2.3	300	0	100	400	10
Serenity	Y	3	299	1	100	400	10
	N	3	300	0	100	400	10
Resuspended in water (1 mg/mL) before							
Sample	Sample (µL)	Water (µL)	BME (µL)	LDS (µL)	Total Volume (µL)	Loaded into well (µL)	
Reduced samples*	14	0	1	5	20	12.5	
Non-reduced samples*	14	1	0	5	20	12.5	

*Note: These samples included all four selected commercial sample products

Both reduced and non-reduced samples were heated at 95°C for 5 minutes and spun down before loaded into the well. NuPAGE™ MES SDS Running Buffer was used in a Mini-Cell Electrophoresis Chamber (Invitrogen, Thermo Scientific). The gel was run for 35 minutes at a constant voltage of 200 V. For each gel, Precision Plus Protein™ Dual Xtra Standard was run in tangent (Bio-Rad Laboratories). Gels were stained with Coomassie Brilliant Blue R-250 Staining Solution for one hour and destained using Coomassie Brilliant Blue R-250 Destaining Solution for two to three hours (Bio-Rad Laboratories) while placed on a Rotomix 50800 Orbital Shaker (Thermolyme). The gel was rinsed with water and imaged under a table of bright light.

The same selected samples were then analyzed using reduced and non-reduced Novex™ 16% Tricine Gels (1.00 mm x 12 well) to further asses the protein profile focusing on the lower separation range of 4 to 30 kDa. The sample loaded into each well differed for each sample and detailed in Table 3.2. Both reduced and non-reduced samples were heated at 95°C for 5 minutes and spun down before loaded into the well. Novex™ Tricine SDS Running Buffer was used in a Mini-Cell Electrophoresis Chamber. For each gel, the Precision Plus Protein™ Dual Xtra Standard was run. The gel was run, stained and destained as described above.

Table 3.2. Reduced and Non-Reduced Novex™ Tricine 16% Gel Sample Load Specification

Resuspended directly in SDS-PAGE Buffer							
Sample	Reduced?	Sample (mg)	Water (µL)	BME (µL)	LDS (µL)	Total Volume (µL)	Loaded into well (µL)
Nestle	Y	1	199	1	200	400	10
	N	1	200	0	200	400	10
Similac	Y	5	199	1	200	400	10
	N	5	200	0	200	400	10
Enfamil	Y	2.3	199	1	200	400	10
	N	2.3	200	0	200	400	10
Serenity	Y	3	199	1	200	400	10
	N	3	200	0	200	400	10

Resuspended in water (1.5 mg/mL) before						
Sample	Sample (µL)	Water (µL)	BME (µL)	LDS (µL)	Total Volume (µL)	Loaded into well (µL)
Reduced samples*	56	3	1	60	120	12.5
Non-reduced samples*	56	4	0	60	120	12.5

*Note: These samples included all four selected commercial sample products

IV. RESULTS AND DISCUSION

a. Identified A1 to A2 Fraction of Signal

Completed in triplicate, the initial analysis of selected commercial products resulted in two of the ten samples that have a normal (both A1- and A2-type β-CNs) protein profile detected (Figure 3.2).

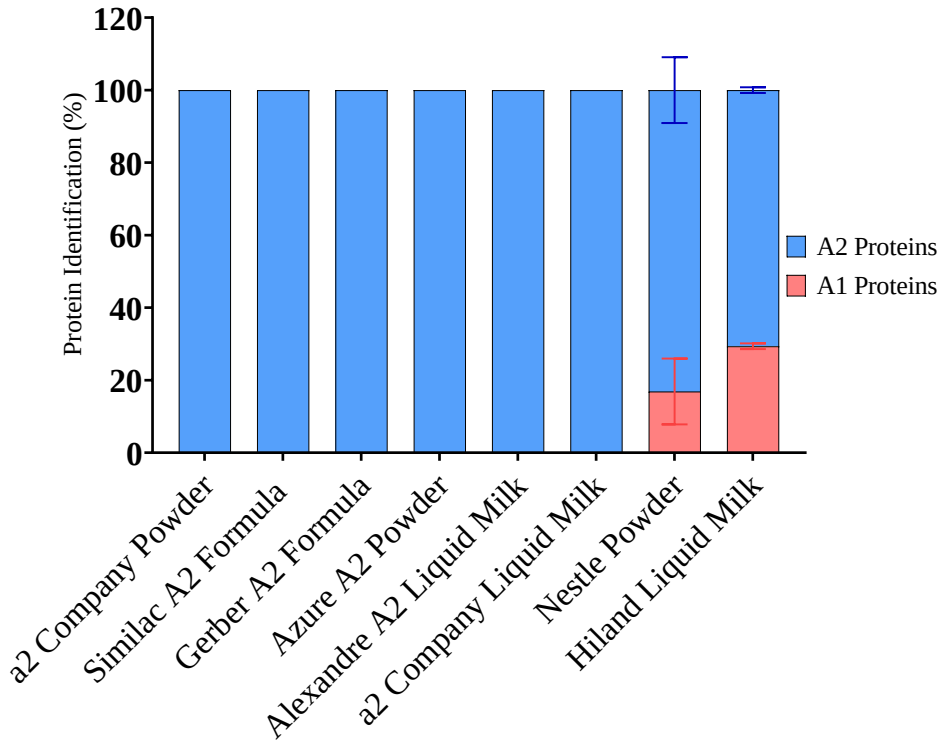


Figure 3.2: A1- and A2-type β-Casein Distribution in Initial Selected Commercial Products.
Percentage is calculated from total sum intensity of β-Caseins and each sample was run in triplicate.

Two samples, Serenity A2 Formula and Enfamil A2 Formula, were omitted from the percent distribution of β-CNs because there were no β-CN proteoforms identified within the sample. The Nestle powder distribution of A1- and A2-type β-CNs was 16.90% (± 9.09%) and 83.10% (± 9.09%) correspondingly. The Hiland HTST Liquid Milk distribution of A1- and A2-type β-CNs was 29.37% (± 0.79%) and 70.63% (± 0.79%) respectively. The rest of the samples did not identify any A1-type β-CNs. Additional samples were later added to collect additional data from sample types with normal and marketed A2 β-CN profiles.

Using the newly selected sample set, a second set of samples were processed through the same method and A1- and A2-type β-CN distributions were calculated as

described before (Figure 3.3). There were no intact A1-type β-CNs within the predicted monoisotopic mass database that were identified in the selected marketed A2 commercial products. There is a stark difference between the normal products compared to the marketed A2 products. This does not confirm an absence of A1-type β-CNs within the selected samples, but that there are no detectable masses that match the predicted monoisotopic mass database created. With the current lack of authentication of products themselves, the absence of detectable A1-type β-CNs is reassuring.

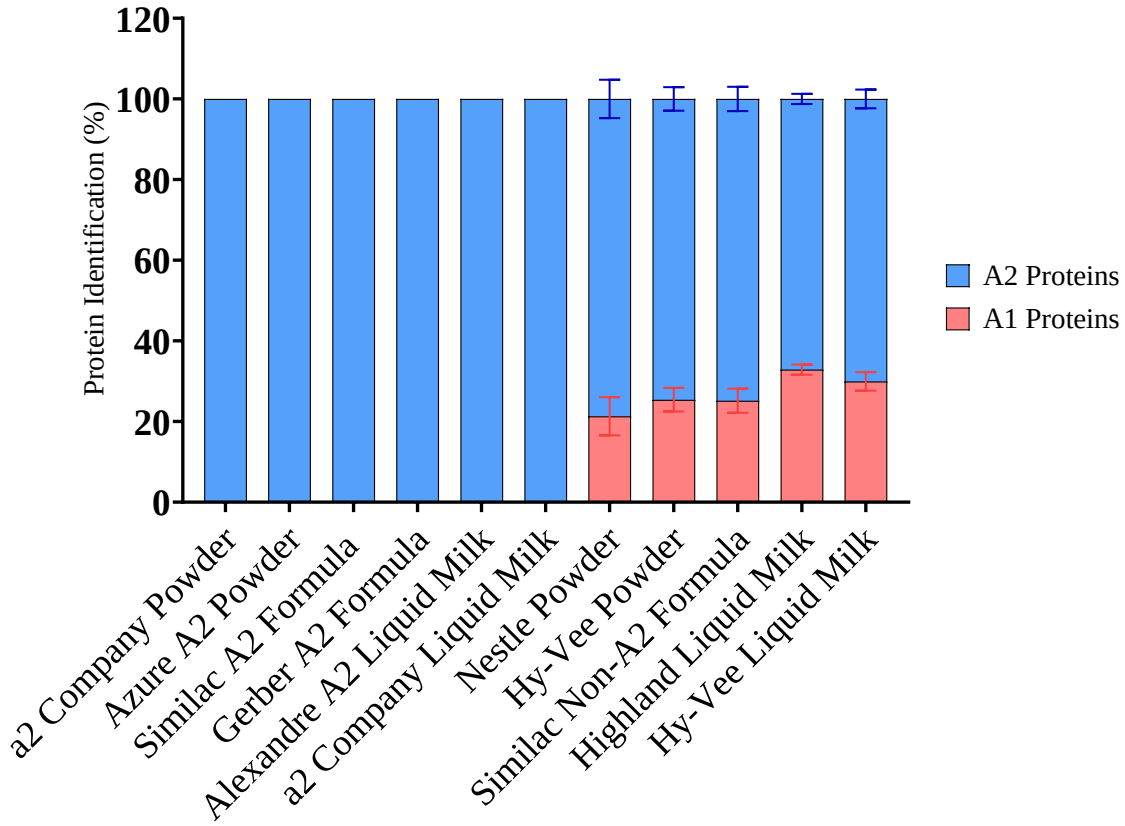


Figure 3.3: A1- and A2-type β-Casein distribution in second selected commercial products.
Percentage is calculated from total intensity of β-Caseins and each sample was run in triplicate.

The A1- and A2-type distribution of each normal sample is detailed in Table 3.3. Using the A1- and A2-type distribution average percentage (25.86% A1 SD = 6.32% *n* = 21 and 74.14% A2 SD = 6.32% *n* = 21) with the knowledge that 80% of total milk protein are caseins and 25-35% of caseins are β-CNs, the approximate percentage of A1- and A2-type β-CNs of total proteins in milk are 5.17 – 7.24% and 14.83 – 20.76% respectively [1, 2].

Table 3.3: Normal Commercialized Products' Percent β-CN Distribution of A1- and A2-Types

Sample	A1-type %	A2-type %
Nestle powder	21.33% (± 4.74%)	78.67% (± 4.74%)
Hy-Vee powder	25.42% (± 2.91%)	74.58% (± 2.91%)
Similac non-A2 powder	25.18% (± 2.99%)	74.82% (± 2.99%)
Hiland liquid milk	32.88% (± 1.25%)	67.12% (± 1.25%)
Hy-Vee Liquid Milk	29.96% (± 2.32%)	70.04% (± 2.32%)

b. Dilution Series of Conventional Product in A2 Product

To test the amount of A1-type β-CNs that could be detected in relation to the amount of A2-type β-CNs, Nestlé NFDM was chosen to be serially diluted in a2 Company A2 WMP evaluate the detection. This essentially evaluates the method's ability to detect adulteration of a marketed A2 milk product with A1-type β-CNs from traditional milk sources. First, the total percentage of identified proteins of the samples and dilution was compared to ensure there were no inconsistencies of protein detection and identification between them. This was calculated through the summing of the fractional abundance of the proteins that could be identified in the spectra from each sample. Fractional abundance is an output from the deconvolution software indicating the

percent abundance of the deconvoluted mass within the total sample spectra. The average overall total percentage of identified proteins of all samples was 89.88% ± 2.86% and ranged from 86.03 – 98.97% which was determined to not distinctly vary for further analysis of the detection and identification A1- and A2-type β-CN distribution of dilutions (Figure 3.4).

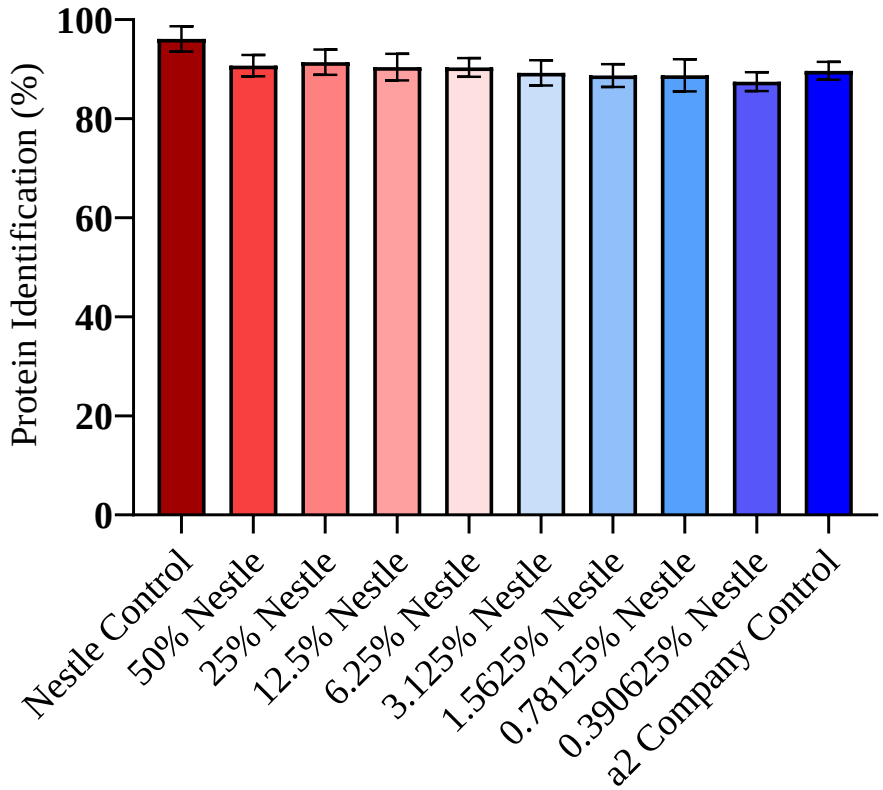


Figure 3.4: Percent Protein Identification Based on the Fractional Abundance of Total Signal Intensity.
Controls averaged from three replicates and dilutions were averaged from nine replicates.

After the detection and identification of proteins within the samples was established as not be discernably different, an evaluation of a dilution set of Nestlé NFDM in a2 Company A2 WMP was completed. Starting at 100% Nestlé NFDM, the sample would be serially diluted into the same protein concentration of a2 Company A2 WMP until reaching 0.390625% Nestlé NFDM in a2 Company A2 WMP. After more than a two-fold dilution (50% Nestlé NFDM), the A1-type β-CNs were not able to be

identified with the methods utilized (Figure 3.5). The total sum intensities of identified A1- and A2-type β-CN proteoforms across the dilution series were recorded from the deconvolution and found to be similar but somewhat inconsistent. The pattern of the intensities increases with increased proportion of A2 milk powder until plateauing at 6.25% Nestlé NFDM. This could be attributed to diluting one product into a different product with different processing methods. For example, if Nestlé NFDM was processed more than the a2 Company A2 WMP, this would lead to different processing proteoforms. The source of milk between the two products and milk sources resulting in a different profile of β-CNs detected. That could lead to the increase of the accumulation of A2-type β-CNs in the a2 Company A2 WMP until it plateaus as shown in Figure 3.5.

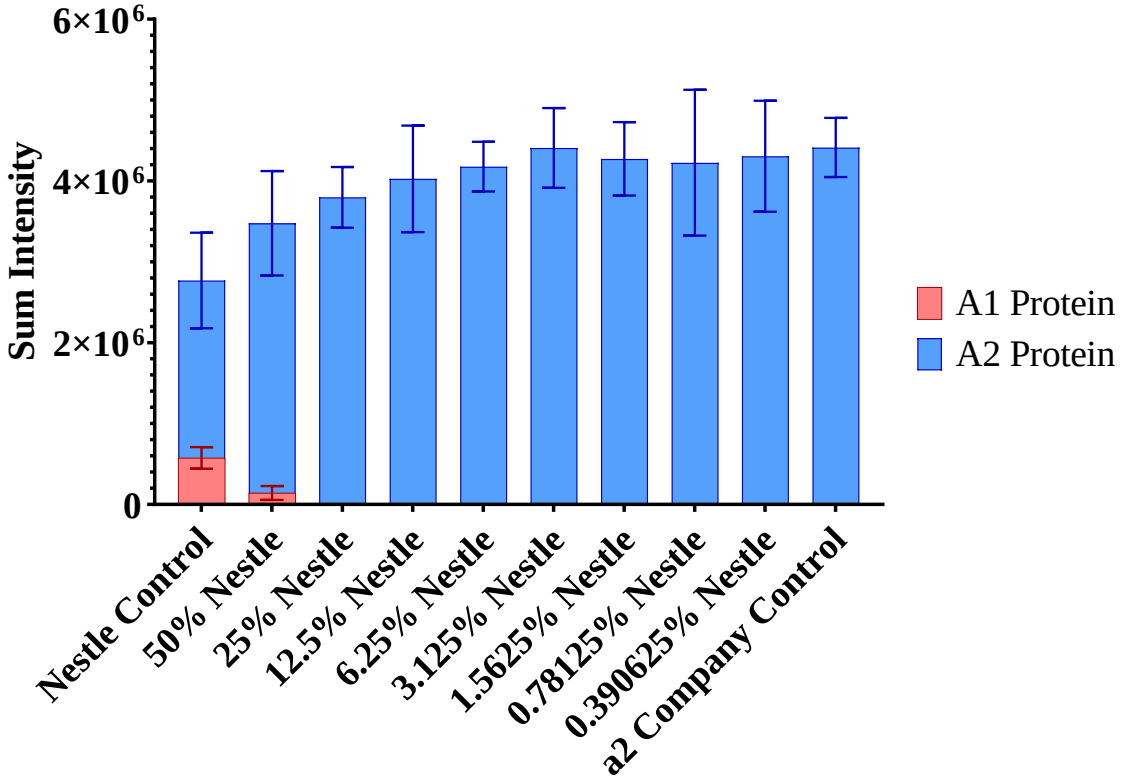


Figure 3.5: The Sum Intensity of A1- and A2-type β-CNs in Controls and Dilutions. Controls averaged from three replicates and dilutions were averaged from nine replicates.

c. Infant Formulas Investigation

Infant formulas, Enfamil A2 and Serenity A2 Formula, produced a very low protein identification percentage (per fractional abundance of total sample signal). Due to this low protein identification, the evaluation of the protein profiles of Nestle powder, Similac A2 formula, Enfamil A2 formula, and Serenity A2 formula was completed through SDS-PAGE 4-12% Bis-Tris and 16% Tricine gels. The 4-12% Bis-Tris gels showed that Similac, Enfamil, and Serenity formulas have smaller masses (around 3-5 kDa) and smear towards the end of the gel while the Nestlé NFDM does not (Figure 3.6). There are minor protein losses between samples directly resuspended in SDS-PAGE buffer compared to samples resuspended in water before SDS-PAGE buffer.

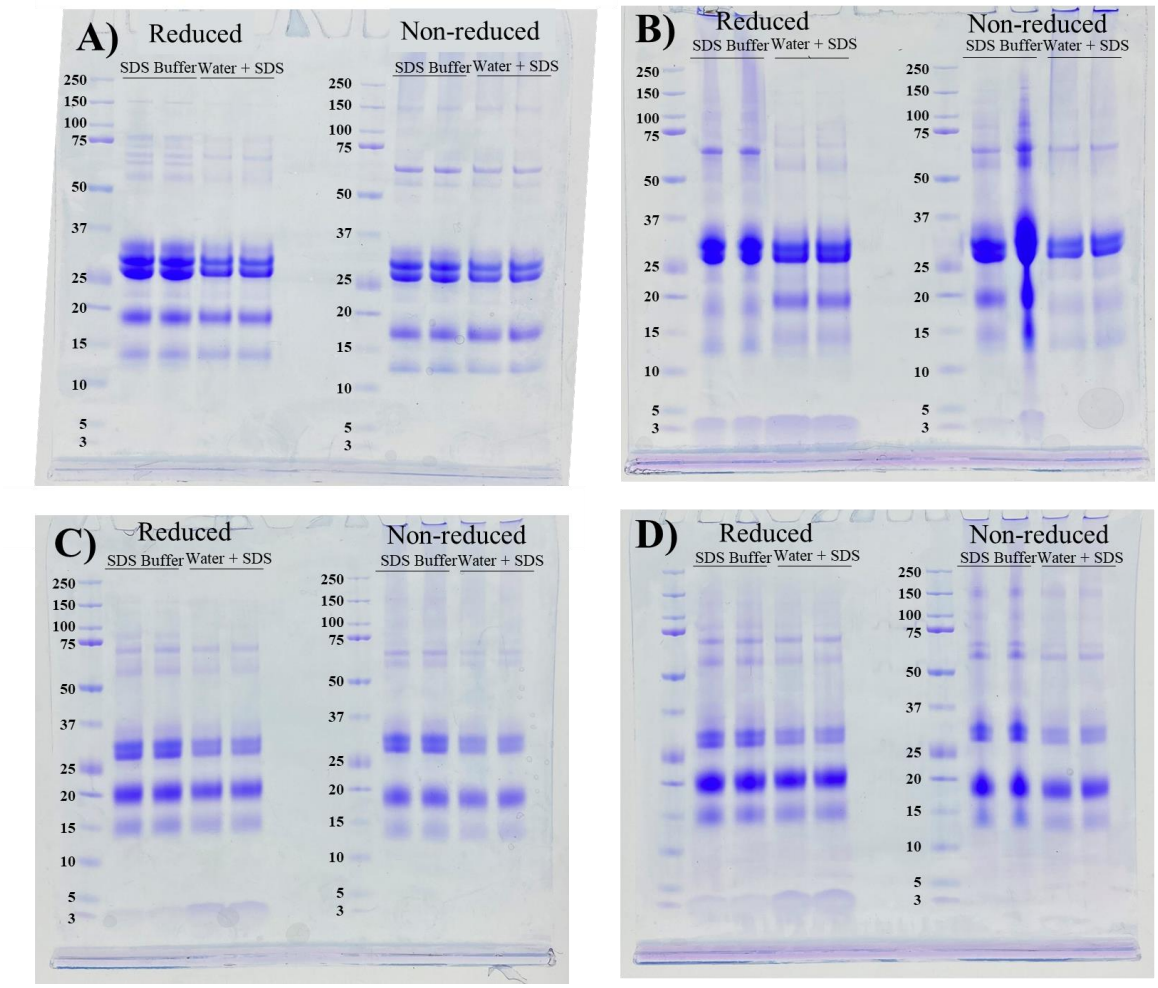


Figure 3.6: Reduced and non-reduced SDS-PAGE Bis-Tris Gel analysis on Nestlé NFDM, Similac A2 Formula, Enfamil A2 Formula, and Serenity A2 Formula. A) Nestle powder B) Similac A2 formula C) Enfamil A2 formula D) Serenity formula resuspended directly into SDS buffer or in water at 1 mg/mL before loaded onto gel.

Further investigation with the aim of evaluating the smaller masses (4-30 kDa) within the sample via 16% Tricine gels produced a similar profile as seen on the 4-12% Bis-Tris gels (Figure 3.7). There is a smear of smaller masses on all the formulas but not Nestle powder. However, the smear pattern is more similar between the Enfamil and Serenity formulas. Though the formulas have similar profiles, Similac has a higher protein identification percentage of ~80% while Enfamil and Serenity have ~30%

identification. The band around 25kDa, where expected α- and β-CN masses should be, is still present but no identifications in that mass range have been made for both Enfamil A2 and Serenity A2 Formulas.

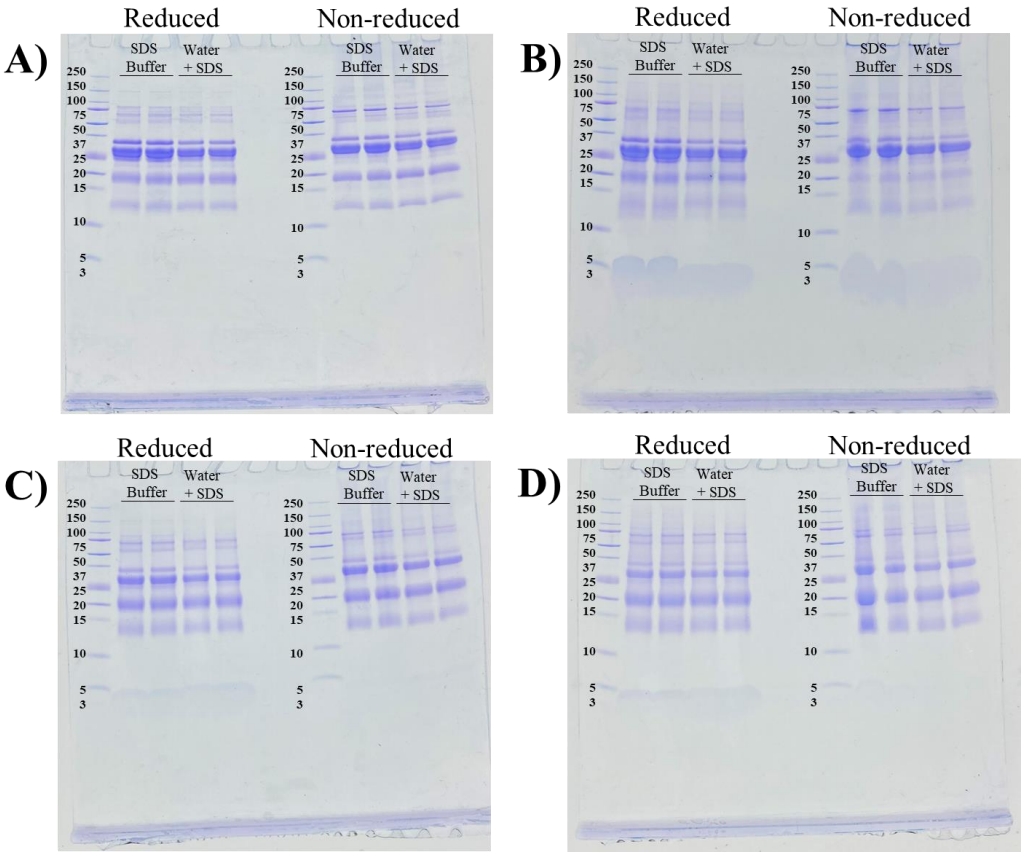


Figure 3.7: Reduced and non-Reduced SDS-PAGE Tricine Gel analysis on Nestlé NFDM, Similac A2 Formula, Enfamil A2 Formula, and Serenity A2 Formula. A) Nestle powder B) Similac A2 formula C) Enfamil A2 formula D) Serenity A2 formula resuspended directly into SDS buffer or in water at 1.5 mg/mL before loaded onto gel.

Further investigations of Serenity A2 formula and Enfamil A2 formula samples found that the low percent identity could possibly be attributed to processing and ingredient formulation. The top 5 ingredients for Serenity A2 formula are organic lactose, organic A2 whole milk powder, organic galactooligosaccharides, organic whey protein concentrate, and organic extra virgin olive oil. The top five ingredients for Enfamil A2 formula are skim milk, lactose, vegetable oil (palm olein, coconut, soy, and high oleic

sunflower oils), whey protein concentrate, and less than 2% galactooligosaccharides. The major difference between these samples and others was that these samples included whey protein concentrate in their ingredient lists while other samples did not.

With the later addition of samples, two more samples could be added to the list of products that included whey protein concentrate in their ingredient list: Gerber non-A2 formula and Enfamil non-A2 formula. The top five ingredients in the Gerber non-A2 formula is whey protein concentrate (from milk, enzymatically hydrolyzed, reduced in minerals), vegetable oil (palm olein, soy, coconut, high oleic safflower, or high oleic sunflower), lactose, corn maltodextrin, and less than 2% of potassium hydroxide. The top five ingredients for Enfamil non-A2 formula were nonfat milk, lactose, vegetable oil (contains one or more of the following: palm olein oil, coconut oil, soy oil, high oleic sunflower oil), whey protein concentrate, and less than 2% galactooligosaccharides. Gerber non-A2 formula did not have any identifications and could be attributed to the whey protein concentrate and no addition of milk powder itself.

The three of the four low percent identification formulas (Serenity A2 Formula, Gerber Formula, Enfamil Formula is marketed with a similar phase along the lines of "easy to digest." This could indicate products marketed as "easy to digest" are not compatible with the intact MS method due to possible partial hydrolysis of the proteins. However, Enfamil Formula was not marketed with this phrase but still has a low percentage identification. In addition, other products (Gerber A2 and Similac A2 formula) that did not have an issue identifying over 60% of the signal also had a similar phrase to "easy to digest." This method was produced for intact protein masses, and the addition of whey protein concentrate (WPC) to products, though similar in protein

amounts, creates an issue where proteins within the predicted monoisotopic mass database cannot be matched with the experimental mass. Caseins represent a relatively low portion of the signal because total casein signal is not a few proteoforms. The signal is split into the different genetic variables and processing proteoforms. With the addition of WPC, the casein signal might be overwhelmed by the WPC signal as whey proteins do not have as many proteoforms. In addition to this possibility, WPC could be hydrolyzed itself. This poses a problem for detection because the method was designed for whole proteins. Partial protein products were generally not considered to be included in the predicted monoisotopic mass database (i.e., γ2, γ3, para κ-CN, and GMP). For these reasons, products that include WPC pose challenges for the method in its current form.

V. SUMMARY

An intact protein LC-MS method was developed to detect and differentiate major bovine milk proteins in powdered and HTST milk products. Eleven normal and A2 milk products were evaluated and ratios of A1- and A2-type β-CN of the total signal were successfully created. Products containing WPC were further investigated using this method resulting in the recommendation that those products are not compatible with this intact protein LC-MS method. The ability to detect, identify, and calculate the fraction of A1- and A2-type β-CNs opens the ability for regulators and industry to begin to understand the β-CN protein profile in products and ingredients and future authentication. Though A2 milk products were focused on for this work, the method could be used for other adulteration/authentication purposes such as adulteration with bovine milk or a specific protein such as BLG.

VI. FUTURE WORK

The developed intact protein mass spectrometry method for the authentication of A2 products have a couple of aspects that could be addressed through future work. These aspects include processing modification assessment, quantification, chromatography optimization, multi-instrumental methods that target different groups of proteins, and a method transfer.

Commercial products with various degrees of heat treatment were the primary samples used for method development. To expand information on process-dependent modifications and how well they are detected on MS, raw milk from genetically tested A1A2 and A2A2 cows could be collected and analyzed using this developed method. This collection of liquid milk would be pasteurized at high temperature low time and then spray dried at known parameters. There was no information regarding spray drying parameters with the selected commercial products and set parameters will help discover patterns of processing. Then raw, pasteurized, and spray dried samples of each milk type would be processed through the sample preparation to be analyzed on the MS. This would test the comprehensiveness of the predicted monoisotopic mass database based on total fractional abundance of identified proteins and the compatibility of proteins to be detected and analyzed by the instrument. In summary, this addition would extend the information about A1 and A2 protein characteristics.

While this method can produce an approximate protein distribution, this method cannot quantify A1- and A2-type β-CNs presently. To verify the A1- and A2-type β-CN content, purified native forms of both groups of β-CNs would be needed. There are a couple of companies selling claimed purified samples of both protein groups but have not

been assessed for proteoform distribution for this method. Biosensis is one of those companies selling purified A1 and A2 β-CNs and could be evaluated for purity by intact LC-MS. If sufficiently pure, they could be spiked into the analytical samples after labelling (e.g., via dimethyl labeling) or used for standard addition [3]. In addition, recombinant proteins could be manufactured, however, the similarity to the native proteins may not be able to be achieved.

To further improve the method, the chromatography gradient optimization could produce further separation of the proteins within sample. As of right now, there are two main large peaks coming off the column around 4-8 minutes. The two peaks are somewhat variable from sample type to sample type but remain similar. Deconvolution software with the sliding windows algorithm was included to detect and differentiate the proteoforms. The default sliding windows algorithm itself was coded with the goal in mind to be able to detect low abundance proteins. An attempt via shallowing the window where the two major peaks come off the column with no difference between the original and shallowed gradient was the only attempt to optimize the gradient and could have additional work added. The substitution of FA with trifluoroacetic acid (TFA) is also another possibility to sharpen peaks, however, TFA is considered an ion suppressor.

In addition to improving the method and targeting specific proteins, the developed method could provide a template for multiple methods targeting specific proteins and/or proteoforms in cow's milk. β-CNs are in low abundance compared to other proteins, and trying to detect and differentiate is already challenging as the whey proteins can drown out the β-CNs. A group of multiple methods run consecutively with a few parameters (i.e., resolution, microscans, etc.) changed between the methods could produce a better

overall picture of the sample. In this case, κ-CNs and other proteins outside the established 3-25 kDa could not be detected at 140,000 resolution and could theoretically be detected and characterized with the change in instrument parameters.

In addition to the additions or expansions expressed previously, a method transfer to a different MS instrument could improve the method output. The method was developed specifically on the Thermo Scientific™ Q Exactive™ Plus hybrid quadrupole-Orbitrap™ mass spectrometer. There have been upgraded models (e.g., Thermo Scientific™ Orbitrap Exploris™ 240) which could be good options to improve the method with the new enhancement's applications. With the transfer of the method, comparison of the fractional abundance of identified proteins, proteins and specific proteoforms identified, variability between samples, and A1/A2 ratio could be compared to the current developed method and assess improvement. This would be important for future work as the limit of this instrument currently is the number of settings you can change (resolution).

Though authentication of A2 labeled products was the focus of this project, the data presented is only a fraction of what the method has achieved. Data regarding β-CNs have been focused on primarily because of the sample selection and objective. However, there is a wealth of knowledge concerning protein proteoforms detected and identified (BLG, α-lac, and α-CNs). Further analysis could even provide a connection between other proteins and β-CNs. Outside of marketed A2 product authentication, this method could be applied to other authentication purposes as well. This could include the authentication of other species' dairy products (lack of adulteration with cow's milk) or even modifying the method to target a different species' milk such as goat or buffalo milk

to check if the milk has been adulterated with cow's milk (would need the species to have a characterized sequence database).

VII. REFERENCES

1. Eigel, W.N., J.E. Butler, C.A. Ernstrom, H.M. Farrell, V.R. Harwalkar, R. Jenness, and R.M. Whitney, *Nomenclature of Proteins of Cow's Milk: Fifth Revision1*. Journal of Dairy Science, 1984. **67**(8): p. 1599-1631.
2. Vincent, D., A. Elkins, M.R. Condina, V. Ezernieks, and S. Rochfort, *Quantitation and Identification of Intact Major Milk Proteins for High-Throughput LC-ESI-Q-TOF MS Analyses.* PLOS ONE, 2016. **11**(10): p. e0163471.
3. Cao, W., J.L. Baumert, and M.L. Downs, *Evaluation of N-terminal labeling mass spectrometry for characterization of partially hydrolyzed gluten proteins.* Journal of Proteomics, 2020. **210**: p. 103538.