Mass Spectrometry of Lens Fiber Membrane Proteins

by

David Shearer

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements of the degree of

Doctor of Philosophy

Department of Biological Sciences

University of Manitoba

Winnipeg

Abstract

Gap junctions are communicating junctions between cells that allow small molecules to pass from the cytoplasm of one cell to the cytoplasm of an adjacent cell. The pores of gap junctions are comprised of two adjacent connexons on neighboring cells, and each connexon is comprised of six connexin proteins. The eye lens of vertebrates is an avascular tissue that is dependent on gap junctions for the distribution of nutrients as well as the removal of waste products. In addition, as the lens cells develop into fibers, they lose their intracellular organelles including the membrane-bound organelles, and are highly dependent on connexons for movement of metabolites and waste materials. Only two connexins, in *Bos Taurus* Cx44 and Cx49, are highly expressed in lens fiber cells. Thus, the lens offers an excellent system for studying gap junctions. In this study, high-pressure liquid chromatography (HPLC) and mass spectrometry (MS) techniques were used to isolate and characterize connexin proteins from the eye lens of the cow and mouse. Despite over 300 proteins being identified from bovine lens using MS techniques, it was still possible to identify the two connexin proteins following proteolytic digests and MS analysis of the resultant peptides. Several post- translational modifications (PTMs) were identified and characterized in lens fiber connexins, including phosphorylations, acetylations and deamidations and proteolytic cleavages. Changes in phosphorylation of several other lens proteins upon the activation of protein kinase C were also identified.

Detection of the orthologous proteins in mouse lens proved more challenging, but peptides derived from both connexin proteins were also detected from this tissue and PTMs of mouse connexins were also observed. Glutathione-S-transferase fusions to mouse Cx44 and Cx50 were used to identify a number of proteins that may interact with the mouse connexins, and the relevance of those interactions was considered. The utility of mass spectrometry to the identification of specific proteins from complex mixtures was clearly demonstrated, and its application to understanding the functional relevance of PTMs was discussed

Acknowledgements

First I would like to thank my wife Jennifer for agreeing to undertake this long journey with me. Without her support and patience through these past years, I would never have reached this point. I would also like to thank my children for their support and accepting the sacrifices we had to make to achieve this,

I would also like to thank my mother Patti and her husband Maurice for their assistance in helping Jenn with the family while I was busy with my studies, and my father Don for his support.

This project never would have begun without the efforts of Adam Liska who first suggested collaborating with Gunnar Valdimarsson, which ultimately led to this project. He also gave me my introduction into mass spectrometry and interpreting spectra.

Thanks to Oleg Krokhin for his instruction and assistance with the high pressure liquid chromatography.

Thanks to Vic Spicer for his help with analyzing data and assistance with software. And thanks also for his movie collection and gifts through the years.

Thanks to my fellow students Shahong Cheng, Liang Tao, and Tarek Bader for their companionship and help with various things in the lab.

I am grateful to my committee members Werner Ens and Dirk Weihrauch for their encouraging words, thoughtful criticism, and time and attention during busy semesters, and also to Dirk for joining our committee after Gunnar's passing.

I would also like to thank Ross Johnson for his support to Gunnar and for graciously agreeing to be my external examiner and providing invaluable feedback on the thesis.

I would never have been able to continue and finish this work without the assistance and support from Ken Standing. Allowing me to work as an assistant in his lab enabled me to continue with my work.

I would like to thank Dr. Steve Whyard for taking over as my advisor and for all the coffee and donuts. His patience, counsel, and help through the writing of this dissertation have been invaluable.

Finally I will be forever in the debt to Gunnar Valdimarsson, who was willing to take a construction worker into his lab and supported me through the years. His death was a devastating blow, but I am grateful for the few years I had to know him.

Table of Contents

List of Tables

List of Figures

Introduction and research objectives

Connexins are the primary protein component of gap junctions, which are small communication pores between cells that allow passage of small molecules (<1000Da) such as ions and second messengers. This type of communication is important in all multi-cellular organisms because it allows, for example, the coupled cells to provide a coordinated response to a stimulus.

There are twenty different connexin genes in mammals and most cell types express multiple types of connexins. Connexins aggregate in hexameric formations to create a connexon, and two connexons from adjacent cells will align to form the gap junction. Connexons can be homomeric or heteromeric, and as such, there are many possible forms of connexons, and this variation may regulate flow of small molecules by virtue of their diversity. Post-translational modifications of some connexins are also possible mechanisms by which connexons can regulate intercellular communication. Being membrane-bound proteins, the biochemical analysis of connexins can be hampered by the usual problems associated with examining hydrophobic proteins, such as the need to use harsh detergents to dissociate the protein from lipid membranes to adequately characterize them. Given that most cells have many membranous organelles, techniques such as density centrifugation are often used to isolate proteins of the plasma membrane, such as connexins, from internal cell membranes, and generally, there is some contamination between the membranous compartments in these preparations. For these reasons, there is still much to learn of the function and regulation of these important cell adhesion molecules.

The eye lens has several properties that would make it an ideal system for studying connexins. The lens fiber cells lose their organelles as they develop and they only express two types of connexins(White & Paul, 1999). As the lens lacks blood vessels, it is dependent on an internal circulatory system that is made up in part by gap junctions, so the connexins are highly expressed. The importance of connexins to lens homeostasis is highlighted by the fact that mutations of connexin proteins can result in lens cataracts.

Mass spectrometry (MS) is an analytical technique that measures the mass-to-charge ratio of charged particles, and has been used in an increasing number of studies of protein expression and protein structure (reviewed in (Aebersold & Mann, 2003). Membrane proteins however, are still difficult to study with mass spectrometry. Most detergents used for solubilizing membranes proteins are not compatible with mass spectrometry and they tend to overwhelm the spectra. However, because of the lack of organelles in lens cells, it is relatively simple to isolate the plasma membrane using reagents that do not interfere with mass spectrometry analyses.

The main objective of this study was therefore focused on developing methods to purify gap junction-enriched membranes with the aim of using mass spectrometry to characterize connexin proteins and their post-translational modification in lenses derived from two different vertebrates, cow and mouse. Cow lenses were chosen as a study model due to the large size of the eye and the ease by which the cells of the lens could be dissociated from the other tissue. The mouse was chosen as the other model as most of the connexin genes have been analyzed from this animal, and there are a wide variety of genetic techniques that can be used to assess gene functions in mice.

The specific aims of the research study were as follows:

Bovine lens connexins:

1. To develop extraction methods for connexin proteins from cow lens in a suitable form that would facilitate mass spectrometry analysis.

2. To use mass spectrometry analyses to examine the complexity of proteins derived from cow lens and to develop strategies to identify connexin proteins within the complex mixture.

3. To use mass spectrometry to determine whether the bovine connexin proteins were post-translationally modified.

Murine lens connexins:

1. To develop extraction methods that would extract connexin proteins from mouse lens in a suitable form that would facilitate mass spectrometry analysis. Given the structural differences in bovine and murine lens, modifications to the bovine lens proteins extraction protocol would likely be required.

2. To use mass spectrometry analyses to examine the complexity of proteins derived from murine lens, to compare this complexity to that of bovine lens, and to develop strategies to identify connexin proteins within the mixture.

3. To use mass spectrometry to determine whether the murine connexin proteins were post-translationally modified in a manner similar to that of bovine connexins.

4. Using Glutathione-S-transferase (GST)-connexin protein fusion techniques, coupled with mass spectrometry analysis, assess which proteins may be interacting with and possibly regulating connexin function.

Literature Review

Gap Junctions

In order to respond to changes in the environment, all multicellular organisms must be able to coordinate the activity of their cells. Organisms have a few ways to achieve this coordination, such as the secretion of hormones or the control of cellular activity by the nervous system, but the most direct way is by directly coupling the cytoplasm of one cell to an adjacent cell. Animal cells are coupled by gap junctions, which are regions of the cell membranes that contain pores between the cells. In vertebrates, the channels are made up by two hexamers of the protein connexin. Each adjacent cell contributes one hexamer to form a complete channel between the cells. Invertebrate gap junctions are formed by innexin proteins (Phelan et al., 1998), which do not appear to be homologous to the connexins. Vertebrates have pannexins (Baranova et al., 2004), which are homologous to innexins, but it is not clear that they form gap junctions (Sosinsky et al., 2011).

The primary function of connexins is the electrical and metabolic coupling of cells. They allow small molecules <1kDa to move between cells. This permits the electrical coupling by allowing the passage of ions between cells. They also allow second messengers such as cAMP, inositol phosphate or Ca^{2+} to be transferred between cells. There are several different types of connexins and connexons or pores made up of the different types of connexins can have different pore sizes and charge selectivities.

The direct coupling of cells was first observed by Schmidtmann in 1925 (cited in (Harris, 2001), when he observed the transfer of dye between cardiac cells. Twenty-

seven years later, based on electrical conductivity studies of Purkinje cells, it was suggested that the cells making up a fiber must be coupled (Weidmann, 1952). Further electrophysiological work demonstrated coupling between different types of neurons (Furshpan & Potter, 1959; Furakawa & Furshpan, 1963).

In the 1960s, electron microscopic studies identified a junction between cells (Dewey & Barr, 1962) and called the structure the nexus. A similar structure was identified in giant Mauthner cells (Robertson, 1963b) that was associated with the electrical conduction between the cells. (Revel & Karnovsky, 1967a) subsequently identified a structure that had a gap between the membranes of the adjacent cells of about 2nm. They used the word "gap" repeatedly in their paper, but they never used the term gap junction. Nevertheless, after this paper, the junctions were henceforward referred to as gap junctions by all other researchers in the field.

Observations of cell coupling were first described in electrically excitable cells, and at the time, it was thought that gap junctions were only to be found in such cells. This was shown not to be the case, as low resistance electrical connections were shown to be present in epithelial cells (Loewenstein & Kanno, 1964) and glial cells (Kuffler & Potter, 1964). These observations marked the realization that cell to cell coupling is widespread and it is now thought that almost all cells in the body are coupled with gap junctions.

The first researchers to isolate relatively pure gap junctions were Benedetti and Emmelot (1968) (Benedetti & Emmelot, 1968), though they mistakenly thought the hexagonal arrays were tight junctions. (Goodenough & Stoeckenius, 1972) further refined the technique and produced very pure samples of gap junctions. (Goodenough,

1974) partially characterized gap proteins using SDS gel electrophoresis, and although this work misidentified some protein degradation products on the gels, it was the first to define the gap proteins as connexins.

The first connexin gene was cloned by (Paul, 1986). He developed an antibody to the purified protein from liver gap junctions and used it to screen a cDNA expression library. The sequenced cDNA was predicted to produce a protein of a theoretical mass of 32kDa, and hence, the protein was named connexin 32. This sequence was used to produce a probe and he observed expression in many tissues, but not the heart and lens, which at the time were known to have gap junctions. This was evidence for the existence of more than one type of gap junction protein.

At this point, the evidence that connexins were the protein component of gap junctions was only circumstantial, as the proteins were only known to be found on the surface of cells, but their function had not yet been defined. The proof that the connexin expression is both necessary and sufficient to form gap junctions came from further experiments. The first experiments to disrupt gap junctions involved the use of antibodies developed to the purified connexin. (Warner et al., 1984) developed polyclonal antibodies to the gap junctions purified from rat liver and injected them into *Xenopus* embryos. They showed that the injected embryos failed to transfer the dye lucifer yellow between cells and that the embryos showed marked defects in development.

That connexin expression is sufficient for cell to cell communication was shown by (Dahl et al., 1981) by injecting connexin mRNA into communication-defective cells, which subsequently induced cell to cell coupling. Later studies (Werner et al., 1985)

showed that injected connexin mRNA can induce the formation of gap junctions between paired oocytes, which was then used for measuring the electrophysiological properties of the channels.

Structure of gap junctions

The development of gluteraldehyde fixation and appropriate staining procedures (Revel & Karnovsky, 1967b) enabled gap and tight junctions to be distinguished, but because of their low electron opacity, the individual particles were difficult to resolve. The gap junctions of the crayfish (Peracchia, 1973a) are larger and electron micrographs of crayfish neurons gave the first indication that the particles in one membrane line up with the particles in the adjacent membrane. En face views of the membranes had shown that the particles are hexagonally packed (Robertson, 1963a) and with the better resolution of (Peracchia, 1973b), six subunits could be seen.

With the first connexin sequence identified (Paul, 1986), the amino acid hydropathy (Kyte & Doolittle, 1982) plots of connexin sequences permitted the prediction of the proteins' structures. Connexins have four hydrophobic domains with three hydrophilic domains linking the hydrophobic regions. The possible topologies were tested by a combination of antibodies produced against synthetic peptides and limited proteolysis to produce a model of connexin structure (Zimmer et al., 1987) (See Figure 1). The four transmembrane segments are referred to as M1 to M4 and there are two extracellular loops E1 and E2; the N and C termini are on the cytoplasmic side of the membrane along with one other loop domain. All known connexins share this structure and show a high degree of similarity in the transmembrane and extracellular domains. Each extracellular loop contains three cysteines and there is a intramolecular disulfide

bridge between a cysteine on E1 and one on E2. Conversely, the cytoplasmic domains, especially the loop and C-terminus of the protein, vary widely both in sequence and length among different connexins.

Three of the transmembrane domains, M1, M2 and M4, are primarily composed of hydrophobic residues; M3 in contrast, contains a number of charged residues in addition to its hydrophobic residues and has been modeled as an amphipathic α-helix in which the hydrophilic residues line the pore (Bennett et al., 1991)). Structural studies have indicated that two membrane regions line the pore (Unger et al., 1999)) and substitutions of various residues with cysteines suggest that M1 and M3 line the pore (Zhou et al., 1997).

The extracellular regions are responsible for the docking of connexons to form the pore. The docking is mediated by non-covalent forces (Ghoshroy et al., 1995). Each loop contains three cysteine residues, but there have only been intramolecular disulfide bridges found. The second loop is important in determining the docking compatibility between connexin types (White et al., 1994).

As different connexins have different permeabilities and gating behaviors, it is thought that the variable regions mediate the differences in physical properties. Swapping the cytoplasmic regions between connexin 26 and 30 to create chimeric proteins caused changes in the conductance and diffusional properties of the connexins (Manthey et al., 2001)

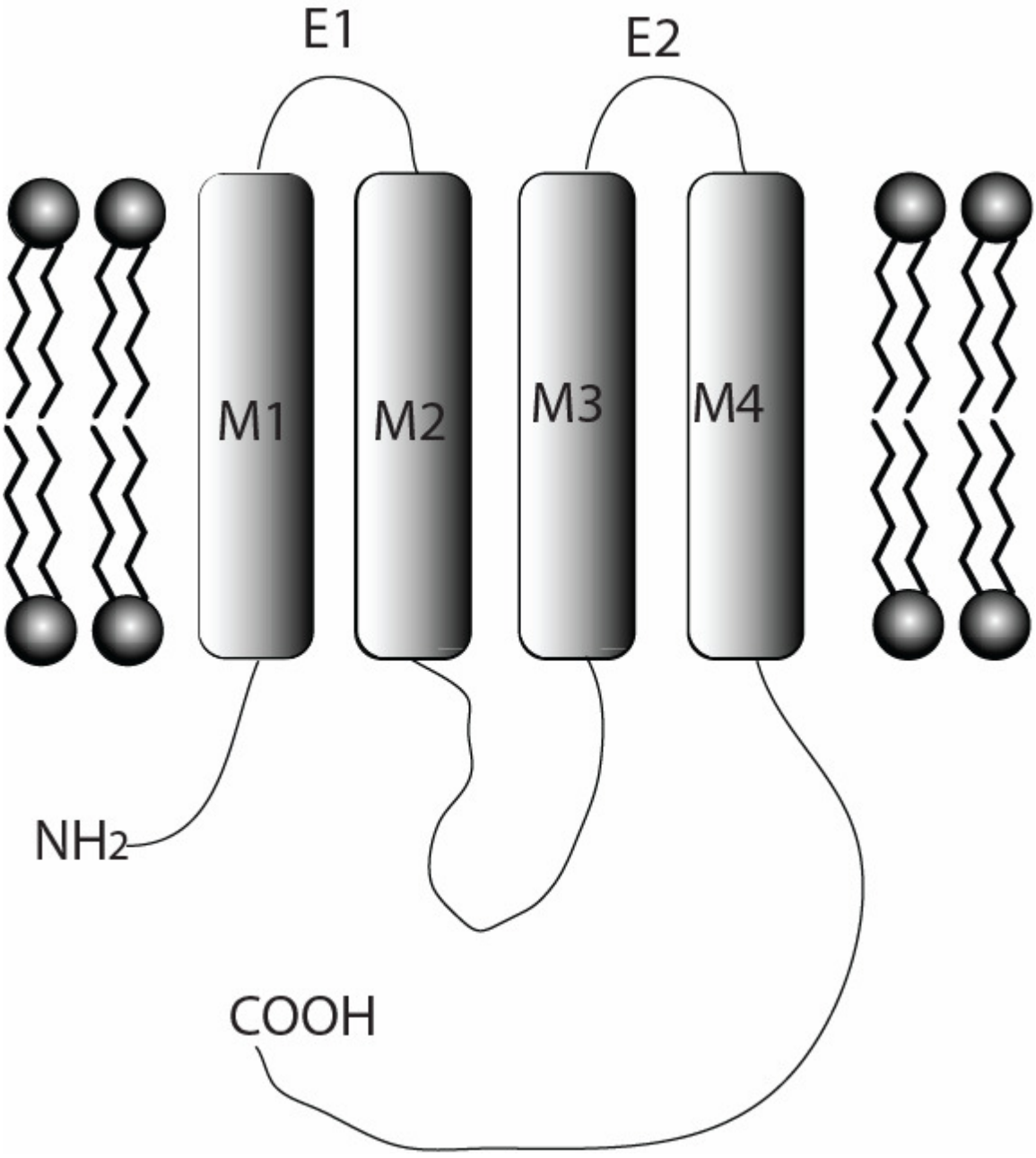


Figure 1. Schematic diagram of a connexin protein. The N and C termini as well as one loop region are on the cytoplasmic side of the cell membrane. There are four transmembrane alpha helices termed M1 to M4 and two extracellular loop domains.

X-ray diffraction studies of 2-D crystals, where the crystals form in doubled membranes, of a truncated form of Cx43 have shown that the four transmembrane domains are alpha helices and confirmed the presence of six subunits in each connexon (Unger et al., 1997). More recent studies (Maeda et al., 2009; Suga et al., 2009) have mapped Cx26 at a 3.5 angstrom resolution. "The hemichannels feature a positively charged cytoplasmic entrance, a funnel, a negatively charged transmembrane pathway, and an extracellular cavity. The pore is narrowed at the funnel, which is formed by the six amino-terminal helices lining the wall of the channel, which thus determines the molecular size restriction at the channel"(Maeda et al., 2009)

Connexin Genes

Even before connexin 32 had been cloned and sequenced, there was evidence of the existence of more than one type of connexin. Antibodies to connexin32 cross-reacted with molecules of different masses that were extracted from other tissues (Dermietzel et al., 1984) and the identified N-terminus of the gap junction proteins purified from rat heart and liver had differing sequences (Nicholson et al., 1985). A second connexin protein derived from heart was cloned using the cDNA from Cx32 as a probe (Beyer et al., 1987). This sequence and its product had a predicted molecular mass of 43 kDa and was named Cx43. Connexin 43 has turned out to be the most widely expressed connexin and also has become the best studied connexin.

With the sequencing of several genomes completed, a comprehensive catalog of connexins has been made. Twenty-one connexin genes have been identified in humans and 20 in mice. Nineteen of the human connexins have orthologs in mice, two are unique

to humans, and one is unique to mice (Willecke et al., 2002). One of the major questions for the field is why so many different connexins are required.

Gap junctions are found in all metazoan life, but only vertebrates have connexins, as invertebrate gap junctions are formed from innexins. These proteins also have four transmembrane segments and the N and C-termini are on the cytoplasmic side, but they have no sequence similarity with connexins. Innexins have been found in several types of invertebrates including the fruit fly *Drosophila melanogaster*, hydra, and the nematode *Caenorhabditis elegans*. Invertebrates can have as great a diversity of innexins as chordates have connexins (e.g. *C. elegans* has 25 different types of innexins, while *Drosophila* has only eight different innexins (Bauer et al., 2005). Chordates have homologous proteins pannexins which may not form gap junctions but act as channels. Interestingly, the primitive tunicate *Ciona intestinalis* has 17 connexin homologs (Sasakura et al., 2003), which suggests that connexins arose in deuterostomes and took over the gap junction-forming role of innexins. Then as chordates evolved, the innexins took on slightly different functions. Early reports showed that the antibody to Cx32 produces the same punctuated staining pattern associated with gap junctions in vertebrates, in hydra, and other invertebrates (Fraser et al., 1987). While the epitope may be common, it is not clear which protein is involved in the immunoreactivity.

The absence of connexin genes in protostomes and their presence in tunicates (White et al., 2004), suggests that they evolved from some ancestral gene prior to the divergence of urochordates. The connexin proteins have similar structures in all mammals, but the relationship between mammalian connexins and connexins in other vertebrates is not always clear (de Boer & van der Heyden, 2005). All connexins share

a similar sequence in the transmembrane domains and almost all have a consistent pattern of three cysteines in each extracellular loop. The N-terminus is also highly conserved but the other two cytoplasmic domains show a great deal of variation. The most widely expressed connexin Cx43 is also highly conserved among mammalian species; Cx43 shows very little sequence variation between species (van der Heyden et al., 2004).

Some connexins genes are also conserved in their untranslated regions (UTRs). Connexin genes usually have two exons, with the second exon containing the entire protein coding region (Willecke et al., 2002). The 5' UTR in most of the connexins shows a great deal of sequence variation. In connexin 43 and connexin 32, the 5' UTR sequences are highly similar and contain an internal ribosome entry site (Hudder & Werner, 2000; Schiavi et al., 1999). The functional significance of the IRES is not clear but has led to some speculation that connexin expression is regulated in part by a terminal differentiation process (Werner, 2000).

Connexin Synthesis and Trafficking

Connexins have a surprisingly short half-life, typically on the order of a few hours (Fallon & Goodenough, 1981; Laird et al., 1991; Beardslee et al., 1998). Even in the lens, the half-life is surprisingly short. In the chick lens, there are two pools of connexin 56, one with a half-life of a few hours and the other with a half-life of a few days (Berthoud et al., 1999). However, given the unique development of the lens, at a certain point, the turnover of connexins must stop and the remaining protein lasts years. The brief half- life of connexins is thought to allow the organism to quickly regulate the degree of coupling between cells, but in the lens of the eye, the tissue shows little change in cellular organization and hence the half-lives of the connexins are greatly extended.

Connexins are generally thought to be inserted in the plasma membrane, as most transmembrane proteins are, co-translationally in the endoplasmic reticulum. Curiously, in a cell-free system, Cx26 was observed to enter microsomes post-translationally (Zhang et al., 1996) and enter the plasma membrane directly (Ahmad & Evans, 2002). However, as the experiments in both of these papers were conducted on cell-free systems, it is not clear that the post-translational translocations of the connexins reflect what happens in cells. In other cell-based studies, fluorescently- tagged Cx26 was found in the Golgi, and brefeldin A treatment caused it to accumulate in the endoplasmic reticulum (Thomas et al., 2005), which suggests that Cx26 is typically inserted into membranes in the usual co-translational manner.

Oligomerization of connexins to form connexons occurs either in the ER-Golgi intermediate compartment or in the trans-Golgi network, depending on the type of connexin and/or cell type (Sarma et al., 2002; Musil & Goodenough, 1993). Connexons are then transported to the plasma membrane where they can act as channels, otherwise known as hemichannels, or dock with connexons on adjacent cells to form intercellular pores. Channels at the membrane form the hexagonal arrays which are characteristic of gap junctions.

When gap junctions oligomerize, they can form either homomeric connexons that are comprised of only one type of connexin protein or they can form heteromeric connexons that are comprised of more than one type of connexin. This interaction can be regulated; for instance, Cx43 and Cx46 form heterotypic connexons in Type I alveolar cells but not in Type II alveolar cells, even though Cx43 and Cx46 are expressed in both cell types (Abraham et al., 2001).

A number of molecules appear to be necessary for the trafficking of connexins from the Golgi to the plasma membrane. A protein termed consortin was identified as an interacting protein with Cx43, and it is required for the transport of at least three types of connexins to the plasma membrane (del Castillo et al., 2010). Cx43 delivery is guided by microtubule tracks to the membrane (Shaw et al., 2007). Vesicles that contain Cx32 utilize kinesin motor proteins and track along microtubules (Fort et al., 2011). However, the transport of connexin proteins to the plasma membrane may not be entirely dependent on these microtubule transport mechanisms, as connexins 43 and 26 were still able to move to the plasma membrane following treatment of the cells with the microtubule disrupting chemical nocodazole (Thomas et al., 2005).

Once at the plasma membrane, connexons freely diffuse in the membrane and dock with connexons from adjacent cells. The docking appears to require the presence of N or E cadherin (Wei et al., 2005; Govindarajan et al., 2010). After they dock, the channels coalesce to form plaques. It is possible that the channels do not open until they become part of plaques, as measurable cell coupling was only observed when large gap junction plaques containing fluorescently-labeled Cx43 were visible (Bukauskas et al., 2000).

Gap junctions can be internalized by a clathrin-dependent process (Piehl et al., 2007). The cell membranes from both cells in a junction are internalized to form a double membrane vesicle termed an annular junction (Marquart, 1977). Annular junctions have not been detected in some cell types, so there may be additional mechanisms for internalization. For example, there is some affinity between Cx43 and caveolin (Schubert et al., 2002), which suggests that this connexin protein may be

internalized by a receptor-mediated endocytosis mechanism involving calveolin. Also, in some cases, gap junction plaques appear to be broken up into smaller aggregates (Fujimoto et al., 1997) and could be internalized as individual connexins or smaller disassociated groups.

Once the gap junctions are internalized, they are delivered to lysozymes and degraded there (Sasaki & Garant, 1986). Later, it was found that degradation of Cx43 requires ubiquitin and an active proteasome (Laing & Beyer, 1995). The current thinking is that mono-ubiquitination of Cx43 is required in the degradation process, but they are ultimately degraded by lysozymes (Berthoud et al., 2004). However, there is some evidence that connexins prior to assembly can be dislocated from the endoplasmic reticulum (VanSlyke & Musil, 2002) and targeted for degradation, which could be mediated by proteasomes.

Post Translational modifications of Connexins

Post-translational modifications (PTMs) are essential in regulating the biological activity of proteins in the cell. PTMs are covalent modifications of the proteins that can modify their activity, state, location or turnover in a cell. More than 200 different types of modifications have been identified (Walsh, 2005); the best studied is phosphorylation or the addition of a phosphate group to a serine, threonine or tyrosine residue. Other common modifications include acetylation, methylation, ubiquitination, acylation, and glycosylation. A protein can also have its activity changed by being proteolytically cleaved. One of the advantages of using mass spectrometry to study proteins is that modified proteins (peptides) can be identified by their shift in mass. Often the modified

residue of a peptide can be identified by collision induced disassociation (Laird et al., 1995).

Almost all connexins have been found to be post-translationally modified. There is evidence that all of the connexins except Cx26 have been found to be phosphorylated (Lampe & Lau, 2000). Several phosphorylation sites of Cx43 have been identified and some of the kinases involved have been identified as well (Lampe & Lau, 2004c). It is thought that phosphorylation of Cx43 is involved in its trafficking from the endoplasmic reticulum to the Golgi or leads to connexin aggregation in the endoplasmic reticulum golgi intermediate compartment (ERGIC). Other phosphorylation sites may also be involved in the gating of the channels and the internalization of gap junctions (Laird, 2005).

Multisite phosphorylation

Cx43 has been shown to be phosphorylated by at least 5 different kinases at serine or tyrosine residues. Multiple phosphorylations on the same protein have been termed multi-site phosphorylations and can greatly increase the regulatory possibilities for the protein. Connexins have the added complexity of being part of a hexamer, so for a connexon comprised of Cx43, there are 2^{12^6} different possible phosphorylation states, assuming that there are no other sites than the 12 that have been identified so far. Theoretical studies of multi-site phosphorylations have suggested that when kinases and phosphatases act in opposition on a multi-site substrate, the system can exhibit stable distributions of different phosphorylation states (Thomson & Gunawardena, 2009).

Multiple phosphorylations could also explain the different effects phosphorylation can have on gap junction communication. Activation of mitogen activate protein kinase (MAPK) has been shown to reduce gap junctional intercellular communication (GJIC) in some cell types (Cameron et al., 2003), while activating protein kinase A (PKA) by increasing cAMP concentrations increased GJIC in mouse mammary tumor cells(Bodenstine et al., 2010). The reduction of GJIC can occur via different mechanisms; MAPK activation seems to lead to an increased internalization while the activity of other kinases inhibit or promote the movement of connexins from the endoplasmic reticulum to the plasma membrane (Solan & Lampe, 2007).

Role of Connexins in Disease

In humans, mutations in connexins are associated with many different congenital diseases. Cx43 mutations can cause occulodentaldigital dysplasia (Kjaer et al., 2004), syndactyly, deafness and atrialventricular septal heart defects. Cx32 mutations can cause Charcott-Marie-Tooth disease (Bergoffen et al., 1993), and mutations in Cx31 and Cx26 are both involved in several different types of congenital hearing impairments (Kelsell et al., 1997). In addition, there are several connexins involved in hyperproliferative skin disorders (Richard, 2005). Mutations in Cx46 are associated with lens cataracts (Mackay et al., 1999) and Cx50 is also associated with nuclear pulverulent cataracts (Berry et al., 1999; Arora et al., 2008).

The Lens

The first morphological indication of the development of the eye is the evagination of the diencephalon. The optic primordial continues to expand and eventually forms the optic vesicles. The vesicles displace the mesenchyme and come in close proximity to the surface ectoderm, which will give rise to the lens and cornea. Inductive signals from the optic vesicles cause cells of the ectoderm to thicken and form the lens placode. Lens placode formation coincides with the expression of crystallin genes. The lens placode invaginates to form a hollow ball or lens vesicle.

The hollow ball is filled as the posterior cells elongate toward the anterior surface forming the primary fiber cells. The mass of primary fiber cells forms the embryonic nucleus. The anterior cells proliferate to the equator; epithelial cells are pushed posteriorly to the equator, elongate, and eventually extend from the anterior to the posterior sutures. The lens grows then by new cells being laid down on top of the older cells, so the cells in the center of the lens remain from the original genesis of the lens (for a review of lens development see (McAvoy et al., 1999).

Lens cell morphology

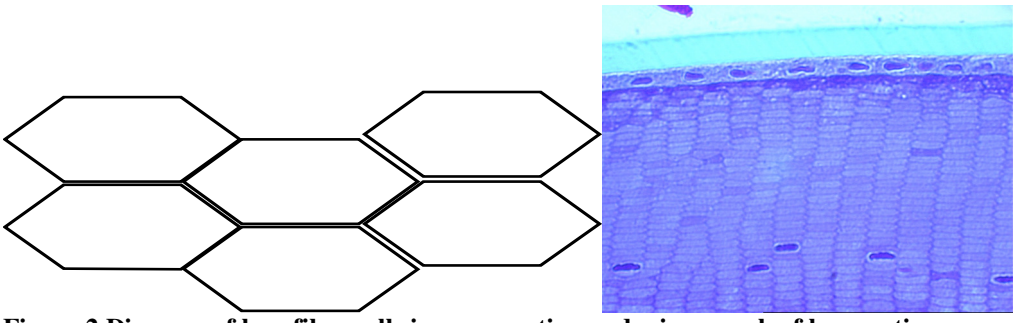


Figure 2 Diagram of lens fiber cells in cross section and micrograph of lens section
The diagram shows the idealized shape of lens fibers as flattened hexagons. They have two broad sides and 4 short sides. The broad sides are perpendicular to the radius of the lens. The micrograph on the right shows a section of mouse lens fixed with neutral buffered formalin and stained with trachoma stain; the light blue layer on top is the lens capsule followed by the single anterior epithelia layer then fiber cells beneath. Magnification 1000X

The primary fiber cells are large and irregularly-shaped compared to the fiber cells that grow later. Lens fibers assume the shape of elongated and flattened hexagons (see Figure 2); the broad sides of the flattened hexagon are perpendicular to the radial direction. The lens fibers meet on the anterior and posterior portions of the lens to form sutures. The membranes between lens fibers form interdigitating processes termed ball and socket joints. Studies of lens structure using electron microscopy have identified three types of gap junctions in lens(Lo & Reese, 1993). The first type is between epithelial cells and has the standard gap junction morphology; the primary protein in these junctions is Cx43. A second type of junction is located on the broad sides of the lens fibers as well as a narrow strip along the middle of the narrow sides both Cx46 and Cx50 are present(Lo et al., 1996). Ultrastructural studies indicate a third type of gap junction is especially enriched in the ball and socket domains. Although morphologically distinct the junctions also contain both Cx46 and Cx50 (Biswas et al., 2010), these regions of the lens are also highly enriched in MIP (i.e. Aquaporin 0).

Cataracts

A cataract can be defined as any opacity of the crystalline lens. Lens transparency results from the ordered structure of lens fiber cells and the tight packing of the proteins within the fibers. The lens consists of an anterior layer of cuboidal cells that are metabolically active and cover the mass of fiber cells. The lens is contained in a connective tissue capsule suspended from the choroid plexus. Cell division occurs at the posterior boundary of the cuboidal cells. The cells that are pushed posteriorly develop into fiber cells and as they elongate they lose their intercellular organelles.

Cataracts can arise from disruption of the cellular architecture or by aggregation of proteins within the cells. Many of the mutant connexins in mice that cause congenital cataracts appear to disrupt the ordered arrangement of the cells themselves. In the double knockout mice Cx50 -/- and Cx46 -/-, the nuclear fiber cells express a reduced amount of gamma crystallin and the cells swell and have a disruption of suture formation (Xia et al., 2006). Cx50 -/- mice develop microphthalmia with normal transparency or mild nuclear cataracts (White et al., 1998), while Cx46 -/- mice develop normal sized eyes with nuclear cataracts (Gong et al., 1998).

In humans, mutation in both Cx50 and Cx46 are associated with cataracts. Mutations in both connexins tend to produce nuclear cataracts or zonular pulverulent cataracts. A missense mutation in Cx46 causing a change from asparagine to serine at residue 63 (N63S) in the first extracellular domain results in a mutant protein that fails to produce functional gap junctions in paired oocytes and does not participate in gap junction formation at all (Pal et al., 2000). Heterozygous individuals with a normal copy of the gene will therefore produce functional connexins. Many of the other mutations are dominant negative because the protein oligomerizes with normal connexins but prevents the formation of working channels (Banks et al., 2009).

It is possible that connexins are a factor in age-related cataracts as well. Cataracts are generally classified on the basis of the location (nuclear, cortical, or subscapular) and appearance (solid, pulverulent). Age is the biggest risk factor for cataracts, but for age-related nuclear cataracts, the cataract is not just a result of the aging process. According to Truscott (2005), oxidation is the key. In lenses with cataracts, there is a progressive oxidation of sulfhydryl and methionine residues, while in a clear lens there is no

significant oxidation. The extent of oxidation may be inversely related to the availability of glutathione (GSH); as long as GSH levels remain above 2mM, significant oxidation of proteins is prevented. Adequate coupling of the lens nucleus to the metabolically active cortex is essential in maintaining GSH levels, and fully functional connexons may therefore play an important role in preventing cataract formation (Sweeney & Truscott, 1998).

The lens is a solid mass of living cells that maintain a high resting voltage, with high concentrations of potassium and low sodium concentrations. Nutrients and antioxidants are required in the nucleus and waste products must be removed. The lens has a microcirculatory system in order to move these small molecules around (Mathias et al., 1997). The metabolically active cells of the epithelium have sodium potassium ATPases that pump out sodium, these channels are especially abundant around the equatorial regions of the lens (Delamere & Tamiya, 2004). The cells have potassium channels, so the potassium that is pumped in leaks out. This results in an electrochemical gradient between the cells at the epithelium and the mature fiber cells underneath. Sodium ions travel from the core of the lens out to the epithelium through gap junctions. By moving these ions, they promote osmosis and so set up a flow of fluid traveling from the center of the lens to the surface. Water and sodium can enter the cells through the sutures and by traveling between cells in a paracellular pathway.

As the lens ages, a permeability barrier develops between the nucleus and the cortex around middle age. Truscott's lab (2005) has shown a reduction in the diffusion of water and also of glutathione in lenses from older individuals. There are several changes that correlate with this reduction in transport. One is an increase in deamidations and

oxidations of susceptible amino acids as well as proteolysis of proteins as time advances. As there is no protein turnover, these post-translational modifications can build up and affect all of the proteins in the nucleus. Another phenomenon is the increase in the amount of insoluble crystallins present in lenses. Much of this insoluble protein becomes associated with the plasma membrane of the lens fibers. These observations have led to several suggested mechanisms for the etiology of age-related cataract. It could be that the loss of GSH in the nucleus causes increased damage from oxidation, which in turn results in the crosslinking of proteins (especially crystallins) in the nucleus which then precipitate out to from the cataract (Truscott, 2005). Alternatively, due to the reduction of ion transport, the amount of calcium in the nucleus increases, which then activates calpains to cleave proteins and can cause them to precipitate (Gao et al., 2004). Or even more simply, just the increase in the level of calcium could be responsible for the precipitation or the loss of water from the nucleus could be enough to cause precipitation of the high levels of crystallins present (Heys et al., 2008).

The one commonality among the models of age-related nuclear cataracts is the loss of transport of substances to and from the nucleus. As the lens transport system is dependent upon gap junctions, one possibility is that the function of gap junctions is inhibited in older lenses. Since there is no protein turnover in differentiated fibers, the loss of function could be due to changes to the connexin protein or the binding of aggregates to the membrane that block the function of the intracellular pores.

Mass Spectrometry

Mass spectrometry technologies have been in use for over a hundred years, but only relatively recently have biologically important molecules been analyzed by mass spectrometry. For any molecule to be analyzed by a mass spectrometer, it must be capable of entering the gas phase and carrying an electrical charge. The original methods developed to ionize an analyte were hard ionization methods such as electron ionization, which would cause the molecules to fragment into various sized ions. A key advance that allowed proteins and peptides to be analyzed by mass spectrometry was the development of soft ionization techniques.

Electrospray ionization (ESI) is one such ionization technique (Fenn et al., 1989). In ESI, gaseous ions are produced by the nebulization of a solution of the analytes such that the droplets carry an electrical charge. There are two models for the formation of gaseous ions from the droplets - the ion evaporation model and the charged residue model(Iribarne & Thomson, 1976; Dole et al., 1968). The main difference between the models is that in the ion evaporation model the charged analyte eventually desorbs from the droplet to form the gaseous ion, while in the charged residue model the charge is transferred as the solvent evaporates away leaving the analyte. Regardless of the ionization mechanism, ESI produces multiply charged ions. The ions can be negative or positive, depending on the polarity of voltages applied to form the spray and accelerate the ions. Peptides are generally analyzed in the positive ion mode. The positive charge is added by the addition of one or more protons, which results in the ion having a mass of the peptide M plus n protons and H is the mass of one proton [M+nH]. The mass spectrometer measures the mass to charge ratio m/z, which for a peptide in positive mode is (M+nH)/n. ESI has an advantage over MALDI (described below) in that it can

be directly coupled with a liquid chromatography system which makes it easy to automate the data acquisition process.

The second of the new soft ionization techniques is matrix assisted laser desorption/ionization (MALDI) (Karas & Hillenkamp, 1988). In this technique, the analyte is suspended in a matrix, which protects the analyte from damage by the laser and promotes vaporization and ionization. The commonly used matrices are 3,5-dimethoxy-4-hydroxycinnamic acid (Sinapinic acid), α-cyano-4-hydroxycinnamic acid (alpha-cyano or alpha-matrix), and the one generally used for the work described here, 2,5-dihydroxybenzoic acid (DHB). All three matrices form crystals, with the analytes suspended in the crystals.

In a typical MALDI experiment, the molecules to be analyzed are mixed with the matrix and then spotted onto a metal plate. Liquid chromatography can be coupled to the deposition and individual fractions deposited directly onto the target. Unlike in ESI,

where formic acid is used, in MALDI trifluoroacetic acid can be used as an ion pairing agent in the chromatographic runs, which can improve the separation and peak characteristics. Ionization is initiated by firing a laser (typically a 337nm UV laser) at the matrix spot. The matrix absorbs the laser's energy causing the sample to vaporize at that spot. The actual ionization is believed to occur secondarily to the formation of the gas pulse or plume. The mechanism of formation of ions in MALDI is still controversial; in one proposed model, the matrix facilitates the formation of ions, while the other model posits that ions are already present in the crystal and just need to be released (Knochenmuss, 2006).

The gaseous ions that are formed from the MALDI spot enter into the mass analyzer. In this study, a quadrupole time of flight (QqTOF) instrument was used (see Chapter 2). In this type of mass spectrometer (Figure 3), a series of quadrupole mass analyzers are coupled to a time- of- flight analyzer. A quadrupole analyzer, as the name implies, consists of four circular metal rods arranged in a parallel array between which an electric field is established. The quadrupole may act as a mass filter for the analytes to pass through or it can be used to pass a wide range of ions, and depending on the pressure of residual gas and their energy, can bused to promote collisional cooling in MS experiments, or it can function as a collision chamber to allow collision-induced disassociation of the analytes for so-called tandem MS experiments. Because peptide ions tend to fragment at the peptide bond, tandem MS spectra can give information on the sequences of the peptides.

In a TOF analyzer, ions enter a flight tube maintained in a high vacuum. The ions are accelerated by an electric field so that they pass through a multichannel plate detector.

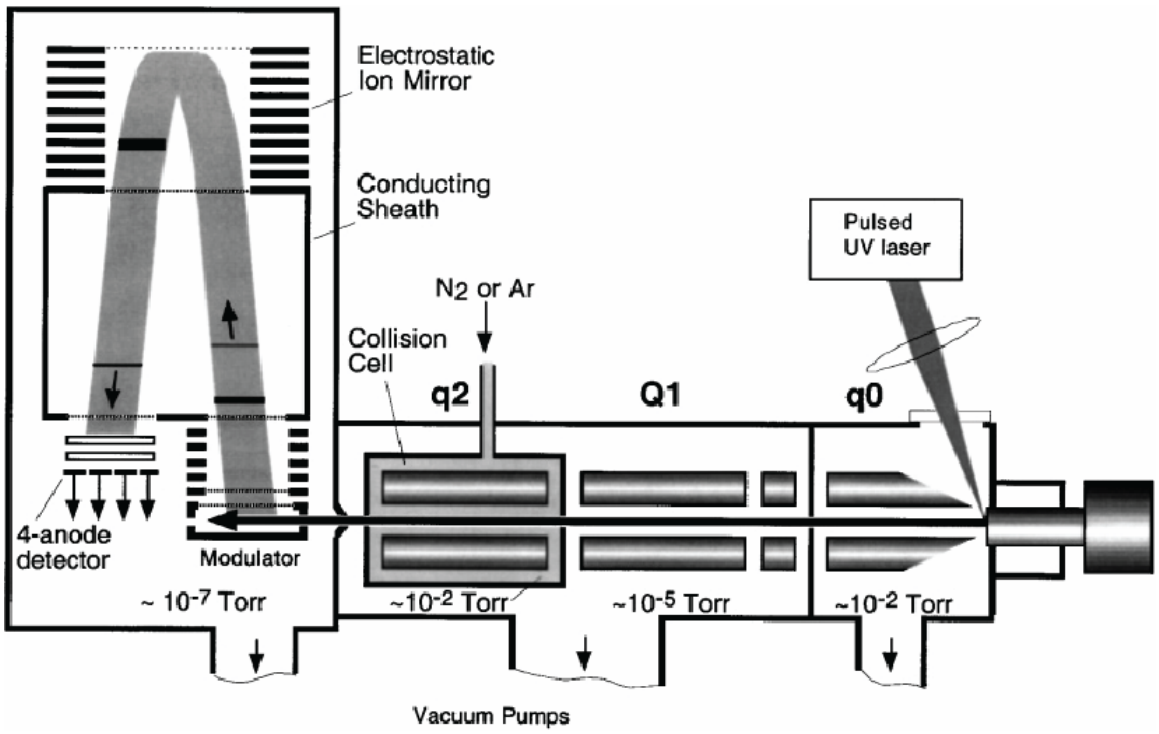


Figure 3. Schematic diagram of a quadrupole time of flight mass spectrometer.

By measuring the time required to traverse the flight tube, m/z values can be determined. The advantage of a TOF analyzer is that it offers a full scan sensitivity over a wide mass to charge (m/z) range. There is limit on the dynamic ranges of such detectors; on the high end, they are limited because of the detector's dead time following the detection of an ion. Also they are limited by the duty cycles (the percentage of time the instrument is able to detect ions during a run) of the mass spectrometer. Typically, in orthogonal TOF instruments, the duty cycle is between 5% and 30%, which means that a large amount of ions produced are lost.

Tandem mass spectrometry refers to a process where multiple steps of mass selection occur with fragmentation occurring between the steps. On a QqTOF instrument, the first quadrupole is set to allow only ions from a limited window of m/z

values to travel to the next quadrupole. The second quadrupole functions as a collision chamber, where ions will collide with a non-reactive gas such as argon. When ionized molecules collide, they begin to vibrate and the vibrations cause the molecule to fragment in a process known as collision-induced disassociation (CID). Current LC-ESI instruments are able to perform multiple tandem mass spectrometry experiments every second by automatically choosing the precursor masses.

In low energy CID analyses, the ionized peptides can fragment along the backbone of the peptide, which can produce several types of ions: a, b, or c ions contain the n-terminus while x, y, or z ions contain the c-terminus (See Figure 4). The most common point of cleavage is at the peptide bond itself, producing y and b ions; a type ions are also common, and it is possible to identify a series of b ions by looking for the corresponding a ions. In low energy CID spectra, the a, b, and y ions tend to be the predominant ions in the spectra. Other types of ions can form, including internal fragments formed by multiple cleavages of the peptide backbone, the loss of side chains, or the formation of immonium ions.

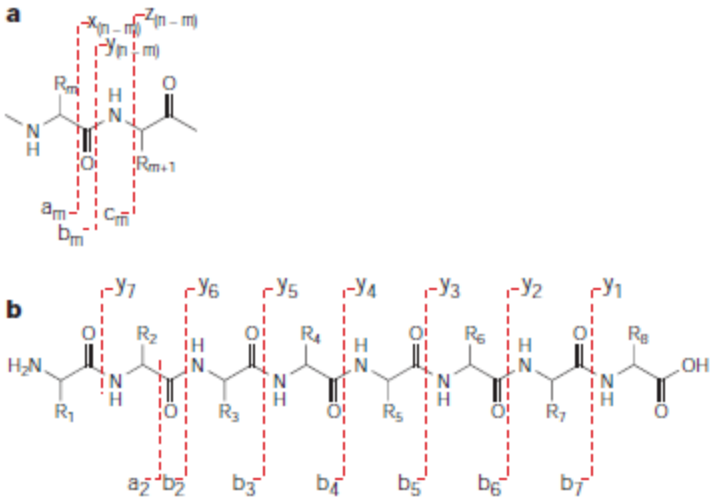


Figure 4 Ion nomenclature for peptide fragmentation

Peptides usually break at the peptide bond; the resulting fragments on the C terminal side are termed y ions and the N-terminal fragments are b ions. The cleavage can also occur on the other side of the carbonyl group resulting in a and x ions, but usually only a ions are seen in the spectrum.

A typical spectrum from a peptide is shown in Figure 5. The x- axis indicates m/z values while the y- axis indicates the number of ions produced. To accurately interpret a spectrum, it is good practice to account for every peak and to identify the corresponding fragment from a proposed chemical structure. This particular spectrum had a number of internal cleavage ions from a cleavage between threonine and proline as well as a number of the expected a, b and y ions.

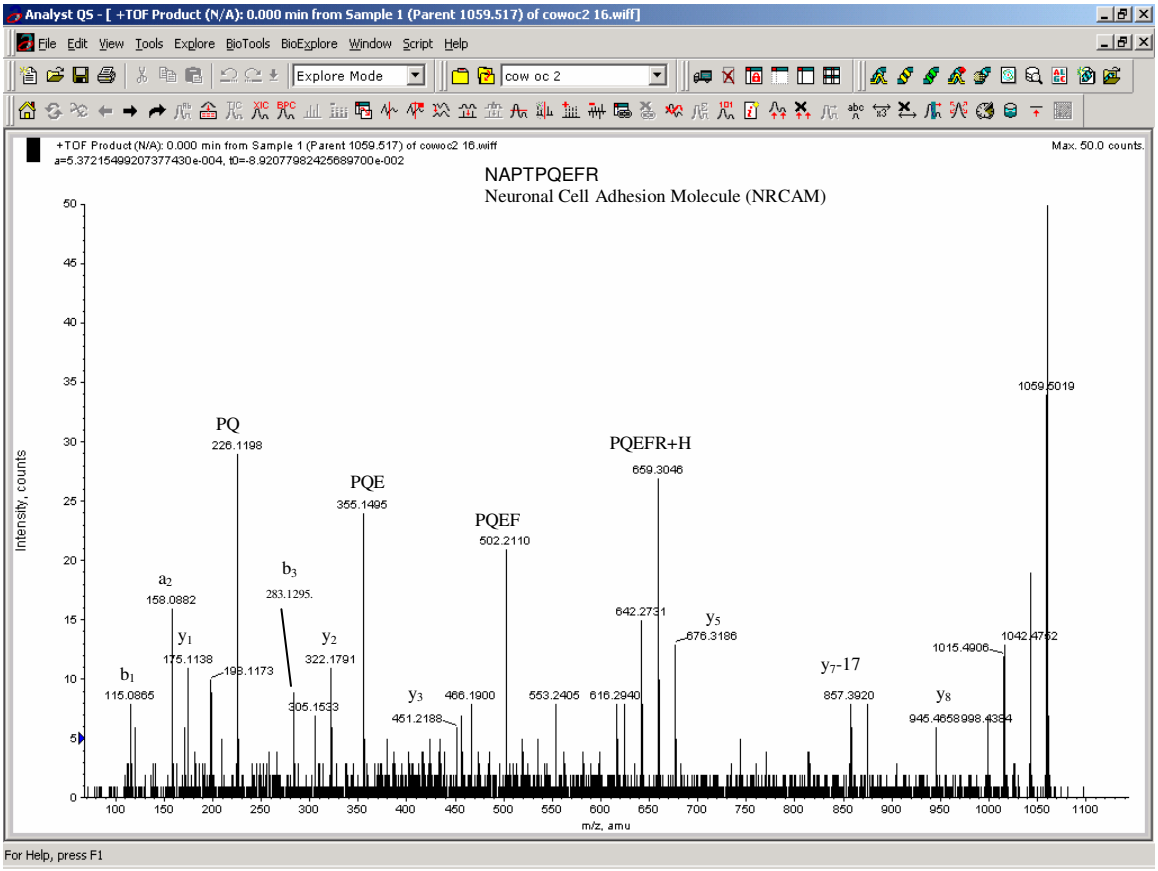


Figure 5. A sample tandem mass spectrum

Certain post-translational modifications, such as acetylation, create easily discernable changes to a protein's mass spectrum. Phosphorylations of serine and threonine are a little less straight-forward to identify on mass spectra. A phosphorylation increases the mass of the parent peptide by 80 Da, but in the collision chamber, the phosphate group is extremely labile and readily disassociates from the peptide in the form of phosphoric acid, thereby reducing the mass by 98 Da. It is possible to identify the phosphorylated residue by looking for a shift in the serine or threonine peaks by 18 Da. Unfortunately, these residues will often spontaneously release water, thereby reducing the mass of the peptide by 18 Da and shifting these peaks as well which can hinder identification of the modified residue, especially if it contains multiple serines and/or threonines.

In this study, mass spectra were obtained using a matrix-assisted laser desorption/ionization (MALDI)-quadrupole time-of-flight (QqTOF) mass spectrometer built at the University of Manitoba (Loboda et al., 2000b). A mass spectrum of each of the 40 fractions was acquired and combined in a set for each HPLC separation/target plate. The MS and chromatographic data were analyzed with software developed in-house (described later), to identify ion peaks of interest. Lists of precursor masses were then generated and tandem mass spectra (MS/MS) were manually acquired according to the list using the same MALDI-QqTOF instrument.

Sequencing/ identification of peptides

Peptide mass fingerprinting is a well established method to identify proteins. It was developed as a method to identify proteins from a band or spot in a gel. Typically,

the protein is digested either chemically or with an enzyme that cleaves at specific residues. The masses of the peptides can be measured with a mass spectrometer and the masses are then used to interrogate a database and identify the protein or proteins that would produce cleavage products of the measured masses. There may be more than one protein present in the band and as long as it is a small number of proteins, they can still be identified by the standard peptide mass fingerprint.

If the sample contains several proteins, the PMF will not work as the there are several factors that cause a reduction in the confidence of any identification. When there are many different peptides present in a spot, some peptides will ionize more readily and suppress the signal from less ionizable peptides. Also, the presence of several peaks from many different proteins increases the likelihood of a false positive match. By adding reverse phase HPLC separation to the process, we can deal with much more complex samples. This is the basis for the sMART (sequence specific Mass and Retention Time) software program, where a protein database is used to create a database of tryptic peptides, the expected molecular weight and retention time is calculated for each peptide, and then the database can be queried with data of masses and retention times to identify what is present in a given sample.

Mass Spectrometry as a tool to study lens proteins

Shortly after the development of the soft ionization techniques, researchers began using mass spectrometry to examine lens proteins. They began looking at crystallins to confirm sequence data and identify post-translational modifications. The first studies used ESI-MS and fast atom bombardment (Smith et al., 1991) to sequence alpha

crystallins. They confirmed the sequence and identified a phosphorylated residue as well. Several years later, MALDI was applied to study lens proteins (Kilby et al., 1996). Bovine aquaporin was sequenced by cyanogen bromide cleavage followed by HPLC and MALDI, which again identified a phosphorylation site as well as a truncation (Schey et al., 1997c).

Mass spectrometry has not been used frequently to study connexins. Cooper and colleagues (Cooper et al., 2000) were the first to use mass spectrometry to conduct a preliminary identification of phosphorylation sites in Cx43. Mass spectrometry has subsequently been used to identify proteins that bind to Cx43 (Singh & Lampe, 2003) and to examine post- translational modifications of connexins 26 and 32 (Locke et al., 2006b). In this latter study, they identified a hydroxylation and/or a phosphorylation near the amino terminus of both connexins, gamma-carboxyglutamate residues in the cytoplasmic loop of both connexins, a phosphorylation in the carboxyl-terminal domain of Cx32, and palmitoylation at the carboxyl-terminus of Cx32.

Mass spectrometry is a technology that offers considerable promise in the analysis of proteins. It offers considerable improvements over Edmund degradation in speed, sensitivity, and relative ease. Although sequencing of mRNA is easier and often gives more complete coverage of the sequence, mass spectrometry provides more direct protein sequence information and more importantly gives information about post- translational changes to the protein. In this study, mass spectrometry techniques were used to examine connexins in both the bovine and murine lens, which had not previously been studied.

METHODS AND MATERIALS

Obtaining lens material

Bovine eyes were obtained from a local abattoir and transported on ice to the lab. The lenses were dissected out and the lens capsule removed. While chilled, the lens material was easy to identify as it remained cloudy, but if warmed above 4ºC, the lenses became clear. The lens tissue was very soft and gelatinous, making it difficult to distinguish between the lens cortex and the nucleus. The lenses were placed in 5mM Tris-HCl, pH 8.0 and frozen at -20ºC until required.

Mouse eyes were collected from mice obtained from the University of Manitoba's Animal Holding Facility. Eyes were enucleated from the mice and the lenses dissected out and frozen in 5 mM Tris-HCl, pH 8.0 at -20 ºC.

Isolation of lens fiber membranes:

Lens fiber membranes were isolated using a previously described protocol (Kistler et al., 1994). Approximately 500 mg lens fibers, derived from 1 bovine lens or 25-50 mg lens fibers derived from 5-10 mouse lenses, were scraped off the lens and placed in 5 mL of 5 mM Tris, pH 8.0, 5 mM EDTA/EGTA and homogenized in a glass homogenizer with 7 strokes. The mixture was transferred to a centrifuge tube and 20 mL of additional buffer was added. The mixture was sonicated in a Microson XL2000 probe sonicator for thirty seconds at a setting of 5. The membrane preparations were then pelleted from the lysates by centrifuging for 20 min at 25,900 rpm in a SW41 Ti rotor (Beckman Instruments).

To remove non-integral proteins adhering to the membranes, the membrane pellet was resuspended in 4 M urea, 5mM Tris-HCl, pH 9.5, 5 mM EDTA/EGTA and then the membranes were precipitated by centrifuging for 20 minutes under the same conditions. To remove additional contaminating proteins, the resulting pellet was resuspended in 20 mM NaOH and centrifuged as described above for 45 min. The pellet was then resuspended in 5 mM Tris-HCl, pH 8.0, and centrifuged again for 45 minutes under the same conditions. The pellet, now enriched with integral proteins, including connexins, was resuspended in 800 µL water and stored in 4 aliquots of 200 µL each at -20 ºC.

In solution digests

The membrane proteins were briefly centrifuged at 18,000 x g in a benchtop microfuge to precipitate the proteins and were then resuspended in 100mM ammonium bicarbonate and 10% acetonitrile. The proteins were then subjected to proteolysis by adding trypsin to a final concentration of 5 µg/ml and incubating the mixture overnight at 37 ºC. The sample was then centrifuged at 18,000 x g to pellet any undigested, insoluble proteins and the supernatant was recovered for analysis. The sample was subjected to an identical second trypsin digest as just described, centrifuged briefly to pellet any further undigested material, and the supernatant was recovered for subsequent analysis. This second digest generally produced the greatest yield of unique peptides.

While most of the digests were performed using trypsin, digests with endoproteinase GluC (5 µg) and chymotrypsin (5 µg) were also used to produce additional peptides to obtain more complete coverage of the proteins in subsequent analyses.

HPLC

Separations of the peptides derived from the in-solution digests were performed using an Agilent 1100 HPLC System (Agilent Technologies, Wilmington, DE). Deionized water, ACN, and trifluoroacetic acid (TFA) were used for the preparation of the eluents. Dried peptides were dissolved in 20 µL of water, and 5 µL was injected into a 5.0-µL loop, and then pumped onto a 150-µm x 150-mm Vydac 218TP C15 Reverse Phase column (5.0 µm; Grace Vydac, Hesperia, CA), which was maintained at 30°C. The sample was eluted from the column with a linear gradient of 1% to 46% ACN (0.1% TFA) in 35 minutes (1.32% ACN per minute at 4 µL/min flow rate), followed by an increase in ACN concentration from 47% to 99% over 5 minutes. PEEK 65 µm inner diameter (ID) and fused silica 50 µm ID tubings were used for pre- and post-column connections, respectively. The column effluent (4 µL/min) was mixed online with MALDI matrix solution (9 mg 2,5-dihydroxybenzoic acid [DHB], Aldrich, Milwaukee, WI, dissolved in 500 µL ACN and 500 µL methanol) at a flow rate of 2.0 µL/min. Forty fractions were deposited onto a gold target at 1-minute intervals (4 µL eluent + 2 µL matrix solution per fraction) by a computer-controlled robot built in-house. Online mixing was performed (Microtee P775; Upchurch Scientific, Oak Harbor, WA), and the fractions were air-dried and subjected to MALDI-MS analysis.

Mass Spectrometry

Mass spectra (MS) were obtained using a matrix assisted laser desorption/ionization (MALDI)- quadrupole time-of-flight (QqTOF) mass spectrometer

QSTAR prototype built at the University of Manitoba (Loboda et al., 2000a). A nitrogen laser (model VSL-337ND, Laser Science Inc.), was used to ionize the sample. The laser was coupled to a 200 μm diameter optical fiber and its output was focused by two lenses to form an image area of 0.3 mm^2 on the target peptides. The ions that traveled in the flight tube were injected into the flight tube pulses triggered at 7500 Hz, which results in a maximum flight time of 125 μs and corresponds to a m/z value of 4500. Spectra were acquired over a m/z range of 500-4500 Da. On occasion, to scan for larger molecules, the pulser frequency was reduced and the mass range was increased.

Tandem mass spectra were acquired for selected precursor ions by targeting the spot containing the precursor and setting the first quadrupole as a mass filter to allow entry of ions of the measured mass. The second quadrupole functions as the collision chamber and the voltage range for the collision was initially set to 50 V/ 1000 Da. The voltage was manipulated up or down during aquisition usually in 5 volt increments to produce spectra with a balanced appearance. Occasionally, the voltage was greatly reduced to allow the intact parent ion through to enhance the parent ion peaks. Spectra were acquired until the operator judged that the intensity of the peaks were high enough for the spectrum to be interpreted .

A mass spectrum of each of the 40 fractions was acquired and combined in a set for each HPLC separation/target plate. The mass spectra were recorded using TOMFA software that was developed in-house. The instrument was calibrated with two peptides, delargin (monoisotopic mass 726.394 Da) and mellitin (monoisotopic mass 2845.762 Da) and the mass of a third peptide substance p (monoisotopic mass 1347.736) was used to check the calibration. Each mass spectrum was generally acquired using a laser

frequency of 10 shots per second over a period of 1 min, although the frequency and time could be adjusted by the operator, depending upon the quality of the spectrum produced.

The MS and chromatographic data were analyzed with software developed in-house to identify ion peaks of interest. Lists of precursor masses were then generated and tandem mass spectra (MS/MS) were manually acquired for the precursor ions listed on the same MALDI-QqTOF instrument.

Data formats and software

The data files in the tofma format were converted to wiff files, which is the format used by the Bioanalyst software from Applied Biosystems MDS Sciex. The conversion process is a two-step process where the tofma files are first converted to a matlab file and then a conversion utility provided by Sciex translates the file into a wiff file. The tofma format is one spectrum per file while the wiff format allows for several spectra to be stored in one file. The tofma to matlab utility can combine several spectra into one file.

Neither wiff files nor tofma files are compatible with X!Tandem software. The Bioanalyst software was therefore used to convert wiff files into a compatible mascot generic format (MGF). Alternatively, we also used software written in house to convert a set of tofma files into MGF files.

Peptide identification

Peptides were initially identified based on tandem MS spectra which were then submitted to X!Tandem (Craig & Beavis, 2004) to be identified, or they were sequenced

manually using bioanlalyst software. During the early phase of this project, software was developed at the University of Manitoba, which identifies tryptic peptides based on sequence specific Mass and Retention Time (sMART) (Krokhin et al., 2006). sMART was used to identify proteins in many of the later chromatograms from the mouse samples as well as providing a check on the identifications derived from the cow samples sMART does not require tandem MS, but is based only on the mass chromatagram. The software has a database of tryptic peptides derived from the Swiss-Prot database (http://www.expasy.org ; Swiss Institute of Bioinformatics, Geneva, Switzerland), where each peptide has a mass and a hydrophobicity rating. The hydrophobicity was calculated based on the sequence using algorithms developed from the sequence-specific retention calculator (Krokhin et al., 2004). The software was then able to identify proteins present in the sample and assign peptides to peaks in the chromatogram. A tolerance of ±3 minutes for retention time and ±10 ppm for the mass was used. This method is similar to a mass fingerprint, but the added dimension of retention time allowed proteins from a complex mixture to be identified. MS/MS was performed to confirm the sequences of peptides identified by sMART, as well as to sequence peptides that sMART did not identify. The tandem mass spectra were processed by peak-selection software developed in-house, which produced peak lists in the Mascot Generic Format. The peak lists were then submitted to the GPM Website (http://thegpm.org/ The Global Proteome Machine Organization Proteomics Database) for identification. GPM searches were performed using the ENSEMBL *Bos taurus* database (ensembl http://www.ensembl.org). The measurement error allowed was 0.4 Da, and the cutoff for the expectation value was log e < -1. The complete modifications field applies a shift in mass to every peptide that

contains the modified residue. Many of the samples were not alkylated so this field was set to none. However, some of the samples were reduced and then alkylated with iodoacetamide, which attaches a carbamidomethyl group to cysteines, resulting in a mass shift of 57 Da; as this is a standard procedure, it is the default setting of X!tandem. Potential modifications that were considered for this study included oxidation, deamidation, and phosphorylation. GPM automatically includes pyroglutamine formation as well as N-terminal acetylation, and it also scans for peptides with non-tryptic cleavages in its refinement step. The device and parent ion method was set to TOF (100 ppm). The spectra were also converted into the wiff format to be manually sequenced with Analyst software to pinpoint modified residues (MDS Sciex, Concord ON, Canada).

Cloning of the Cx46 and Cx50 C terminal domains:

The C-terminal domains of Cx46 and Cx50 were PCR-amplified from the mouse DNA using the primers listed in Table 1. These primers were designed with either EcoRI or XhoI linkers to enable the PCR products to be eventually ligated into the pGEX6p-1 plasmid (purchased from G.E. Healthcare and Life Science).

Table 1. Primer sequences used to PCR-amplify the C-termini of connexin genes Cx46 and Cx50 from mouse

Cx46CTF1	5'GTAATGAATTCTACCACCTGGGCTGGAAGAAG
Cx46CTR1	5'GATGACTCGAGCCACTGAAGGGCTGCTAACACC
Cx50CTF1	5'GTAATGAATTCGAGATGAGCCACCTGGG
Cx50CTR1	5'GATGACTCGAGCCTAGTTGTACAGCGTG

The PCR was conducted using the proofreading DNA polymerase PWO (Roche) in standard 25 µl PCR reactions containing 1µl cDNA template, 17.3 µl H_2O, and the following reagents from Invitrogen: 1 µl 10 mM dNTPs, 0.5 µl 50 mM $MgCl_2$ (note that this is 50% of standard conditions), 2.5 µl 10x PCR reaction buffer, and 0.2 µl PWO DNA polymerase. The following program was used to amplify the DNA sequences: 94°C for 3 min, followed by 30 cycles of 94°C for 30 sec, 55°C for 30 sec, 72°C for 1 min, followed by a final extension of 72°C for 10 min.

The PCR products were purified using a Qiagen QIAquick spin kit and the DNA was then restriction enzyme-digested with EcoRI and XhoI restriction enzymes. The pGEX6p-1 plasmid was similarly digested and the restriction digest products were resolved on 1% agarose gels and purified using a Genclean kit (Bio 101). The connexin gene fragments were ligated into the pGEX6p-1 plasmid by incubating the digested fragments at 15°C with T4 DNA ligase for 3 hours. 2.5 uL of the ligation reaction was used to transform Stratagene Gold Competent *E. coli* cells. The Stratagene cells were used to amplify and store the plasmid but BL21 cells (supplied along with pGex plasmid from G.E. Healthcare) are recommended for protein expression, so liquid cultures of the Gold cells were grown overnight and the plasmid purified using the QIAquick Plasmid DNA kit.

Preparing Competent Cells

Lyophilized BL21 cells were thawed and resuspended and grown overnight in LB media. The culture was then used to streak a plate which was incubated for 8 hours and then a single colony was used to inoculate a 5 mL liquid culture and grown overnight. One ml

of the culture was transferred to 100mL LB and incubated until the OD600 value reached 0.5, which took ~2 hours. The flask was chilled on ice for 20 minutes then centrifuged at 1500 rpm for 5 min at 4 °C and resuspended in TSS buffer (10% polyethylene glycol, 5% DMSO, 20 mM $MgCl_2$). Aliquots of 150 uL were dispensed in 1.5 mL microfuge tubes and frozen at -80°C until needed for transformation

Transformation of bacterial cells

To transform cells, 1 uL of plasmid was added to the competent cells. They were then heated to 42 °C for 2 minutes, then transferred to ice for 3 minutes. SOC media (800 uL) was added to the tubes and they were incubated on a shaker at 37 °C for one hour. The cells were plated and grown overnight.

Growth and production of protein

For small-scale production of protein, a single colony was used to inoculate a flask with 12 ml of LB medium. The cells were allowed to grow for 3 hours, when the A600 reached ~0.6 – 0.8. 10uL of IPTG was added and the culture was incubated for another 2 hours. The culture was transferred to centrifuge tubes and the cells pelleted and then resuspended in 600uL of ice cold PBS. Cells were lysed by adding 6uL of 10mg/mL lysozyme and then freeze-thawed 10 times. The lysate was centrifuged at full speed (14,000 rpm) in a microcentrifuge to remove insoluble material.

Purification of protein

Proteins were purified by adding the bacterial lysate to GSH-agarose beads. The lysate was added to the beads and the proteins were allowed to bind for one hour at 4 °C. The beads were then washed four times with Tris-buffered saline to remove unbound proteins. To confirm the sequence of the affinity-purified fusion proteins, a peptide mass fingerprint was performed on each of the samples. 5 µL of the beads were transferred to a fresh spin column, and the proteins were eluted with 10 µL GST. An equal amount of 100 mM ammonium bicarbonate was added and 1 µg of trypsin and the proteins were digested overnight. The digest was directly spotted on a MALDI target and the resulting MS spectra were compared with expected spectra from a theoretical digest of the predicted sequence.

Covalent binding of proteins to beads

The initial pull-down experiments were performed without covalently binding the bait to the beads, but the bait protein then tended to dominate the MS spectra. Hence, the GST-fusion proteins as well as the GST control were cross-linked to the GSH-agarose beads with bis-(sulfosuccinimidyl) suberate (BS3), which links the primary amine of GSH to lysines in the GST portion of the protein. Beads were transferred to a spin column and washed 4 times with PBS. The bottom of the spin column was plugged with glass wool and 0.5 mL of PBS was added to the column, followed by BS3 in 5mM sodium citrate. The column end was capped and the cross-linking reaction was allowed to continue for 45 minutes with shaking. The column was spun at 4000 rpm for 2 min and the quenching

buffer (1 M Tris-HCl, pH 8.2) was added to stop the reaction. The beads were then washed with glutathione elution buffer followed by wash buffer.

Protein Pull-down Experiments

Lens fiber lysate was added to the recombinant proteins Gst-Cx46Ct or Gst-Cx50Ct bound to glutathione-conjugated beads. GST alone was used as a negative control. The experiments were also performed with the GST-fusion proteins not covalently bound. Each mixture was incubated at 37° C for 1 hour and then the beads were washed with TBS six times. The bound proteins were eluted with reduced glutathione and the eluted proteins were digested with trypsin as described above. The digested peptides were separated by RP-HPLC (as previously described) and 40 fractions were deposited on a MALDI target. The spots were analyzed by mass spectrometry and then selected peaks were analyzed by tandem mass spectrometry

In vitro phosphorylation assay.

The GST fusion proteins were also used for *in vitro* phosphorylation assays. Samples of the fusion proteins were incubated with 0.5 ng/uL Protein Kinase C (Upstate) in 40 mM HEPES (pH 7.5), 0.04% Triton X-100, 100 mM ATP 1 mM $CaCl_2$, 10mM $MgCl_2$ and the PKC lipid activator (Invitrogen). A histone H1 peptide was used as a control reaction. The proteins were digested with trypsin and analyzed by MALDI time-of- flight mass spectrometry.

Results

Identification of bovine lens proteins

As a first step to characterize the proteins present int the lens membrane, the extracts were analyzed by SDS polyacrylamide gel electrophoresis. A typical gel is shown in Figure 6. A total of 17 bands were excised from lane 5 and were subjected to in-gel digestion with trypsin to identify the proteins in each band the labeled bands were subjected to further LC MS/MS analysis. For the initial characterization, peptide mass finger print analyses were performed with MALDI. Most of the samples had several prominent peaks in the mass spectra. The peak lists were submitted to the Mascot Peptide Mass Fingerprint database via the matrix science web portal. Although there were several peaks in the spectra, the mass fingerprinting was inconclusive, because most of them had scores well below the significance threshold. For instace beta b1 crystallin was identified in some samples, but the mascot score was a low 67 with a resulting expectation value of 0.17which is well above the value of 0.05 for statisctical significance. To be considered significant, the score should be above 72 (Pappin et al., 1993). MALDI Tandem MS was performed on a few peaks from each band and confirmed that beta b1 crystallin was present and also identified some peptides from other known lens proteins such as phakinin, filensin and aquaporin 0.

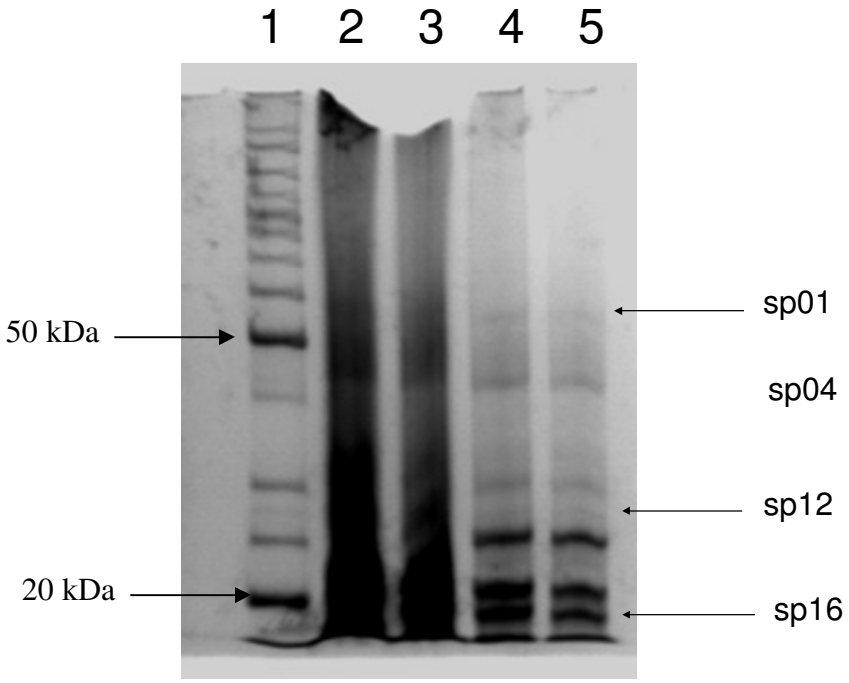


Figure 6 SDS PAGE of the urea- and NaOH-insoluble fraction of the bovine lens. Lane 1 contains molecular weight markers, lane 2 contains the initial preparation and subsequent lanes 3, 4, and 5 are 1:2, 1:10 and 1:100 dilutions of the initial preparation, respectively. Proteins were stained with Coomassie Blue-R for visualization. Numbered bands in lane 5 were subsequently subjected to LC-MS/MS.

Samples from four of the bands were also subjected to ESI LC-MS/MS performed on a commercial qstar instrument. Table 1 lists the proteins identified from each of the digests. Each protein band contained multiple proteins, between 13 to 30.The expectation value is the probability that a spectrum would match the peptide identified by chance, and hence, a lower value indicates a better match. The table also indicates the number of peptides matched from each protein. The results are all derived from the assembled *Bos taurus* database and the description is based on the results of matching a protein sequence to annotated proteins from other studied genomes. The results are similar to the mass fingerprinting results, in that crystallins and aquaporin 0 are present in all samples excised from the gel. One of the primary proteins of interest, Cx44, was identified in one of the bands (sp04), but only one spectrum was matched to Cx44.

Several other proteins were also identified in this band, including the ubiquitous crystalline beta b1, as well as phakinin and glutathione-S-transferase, all of which may have interfered with identifying more peptides from the Cx44 band.

Table 1a. Proteins identified by in-gel digestion of band 1.

log(e)[a]	unique[b]	total[c]	Mr[d]	Accession[e]	Description
-151.1	17	17	65.8	sp|K2C1_HUMAN|	Cytokeratin 1
-108.2	11	11	62.1	sp|K1C9_HUMAN|	Cytokeratin 9
-90.1	10	10	59.5	sp|K1C10_HUMAN|	Cytokeratin 10
-87.2	9	9	65.8	sp|K22E_HUMAN|	cytokeratin 2e
-79.4	10	11	83.1	ENSBTAP00000029150	Filensin
-49.9	5	5	28	ENSBTAP00000025671	Crystallin beta B1
-42.9	6	6	19.8	ENSBTAP00000004073	Crystallin alpha A
-27.7	3	3	24.4	sp|TRYP_PIG|	Trypsin
-20.4	2	2	23.7	ENSBTAP00000025674	Crystallin beta A4
-17.4	3	3	56.3	ENSBTAP00000007860	Leucine amino peptidase
-17.1	2	2	16.7	ENSBTAP00000032372	Crystallin beta A4
-10.7	2	2	45.7	ENSBTAP00000024638	Phakinin
-9.9	2	2	49.9	ENSBTAP00000001950	similar to alpha 1 Tubulin
-8.1	1	1	24.3	ENSBTAP00000001656	Crystallin beta B3
-6.4	1	1	49.2	ENSBTAP00000001988	Keratin type I
-5.5	1	1	29.4	ENSBTAP00000020243	NA
-5.3	1	1	55.3	ENSBTAP00000010661	Retinal dehydrogenase
-4.5	1	1	19.5	ENSBTAP00000049180	NA
-2.7	1	1	20	ENSBTAP00000000556	Crystallin Alpha B

-1.8	1	1	68.5	ENSBTAP00000033759	NA
-1.5	1	1	303.3	ENSBTAP00000010295	Thyroglobulin precursor

a- Expectation value
b- Number of unique peptides matching the identified protein
c- Total number of peptides (unique or shared) matching to the identified protein
d- Relative mass in daltons
e- Accession number of closest matching protein in the *Bos taurus* ensemble database

Table 1b. Proteins from band 4

log(e)[a]	unique[b]	total[c]	Mr[d]	Accession[e]	Description
-157.6	16	17	65.8	sp|K2C1_HUMAN|	Keratin, type II
-88.3	10	10	45.7	ENSBTAP00000024638	Phakinin
-81.8	7	7	62.1	sp|K1C9_HUMAN|	Keratin, type I
-71.1	7	7	65.8	sp|K22E_HUMAN|	Keratin Type II
-61.2	6	7	83.1	ENSBTAP00000029150	Filensin
-52.6	6	6	28	ENSBTAP00000025671	Crystallin beta B1
-50.1	5	6	24.3	ENSBTAP00000001656	Crystallin beta B3
-49.2	6	7	19.8	ENSBTAP00000004073	Crystallin alpha A
-29	4	5	59.5	sp|K1C10_HUMAN|	Keratin type I
-27.8	4	5	24.4	sp|TRYP_PIG|	Trypsin
-26.9	4	4	20	ENSBTAP00000000556	Crystallin alpha B
-26.4	3	3	23.7	ENSBTAP00000025674	Crystallin beta A4
-19.4	1	1	62.6	ENSBTAP00000021456	Keratin type II
-14.3	2	2	45.7	ENSBTAP00000036739	Actin
-10	2	2	54.1	ENSBTAP00000040563	glucose transporter type 1
-9.7	1	1	16.7	ENSBTAP00000032372	Crystallin beta A4

-8.7	2	2	44.9	ENSBTAP00000001187	Phosphoglycerate kinase 1
-6.4	1	1	29.4	ENSBTAP00000020243	Carbonic anhydrase 3
-4.7	1	1	55.3	ENSBTAP00000010661	Retinal dehydrogenase 1
-3.6	1	1	28.3	ENSBTAP00000013360	Aquaporin-0
-3.5	1	1	25.1	ENSBTAP00000007037	Crystallin beta A3
-2.9	1	1	68.5	ENSBTAP00000033759	Radixin
-2.4	1	1	73.7	ENSBTAP00000025207	na
-2	1	1	49.2	ENSBTAP00000001988	KRT18 protein
-1.9	1	1	53.5	ENSBTAP00000020622	NA
-1.9	1	1	54.1	ENSBTAP00000040493	glucose transporter type 3
-1.8	1	1	24.5	ENSBTAP00000010119	Alpha S1 casein
-1.6	1	1	69.3	sp|ALBU_HUMAN|	Albumin
-1.4	1	1	50.5	ENSBTAP00000006992	Rab GDP dissociation inhibitor beta
-1.2	1	1	22.6	ENSBTAP00000005642	regulator of G-protein signaling g16

a- Expectation value
b- Number of unique peptides matching the identified protein
c- Total number of peptides (unique or shared) matching to the identified protein
d- Relative mass in daltons
e- Accession number of closest matching protein in the *Bos taurus* ensemble database

Table 1c. Proteins identified from band 12

log(e)[a]	unique[b]	total[c]	Mr[d]	Accession[e]	Description
-111.2	10	10	28	ENSBTAP00000025671	Crystallin beta B1
-79.2	8	8	65.8	sp|K2C1_HUMAN|	Keratin type II cytoskeletal 1
-50.5	6	6	65.8	sp|K22E_HUMAN|	Keratin type II cytoskeletal 2

-50	5	5	24.3	ENSBTAP00000001656	Crystallin beta B3
-34.2	4	4	62.1	sp|K1C9_HUMAN|	Keratin type I cytoskeletal 9
-31	4	5	19.8	ENSBTAP00000004073	Crystallin Alpha A
-25.5	4	4	59.5	sp|K1C10_HUMAN|	Keratin type 1 cytoskeletal 10
-19.7	3	3	45.7	ENSBTAP00000024638	Phakinin
-16.4	2	2	24.4	sp|TRYP_PIG|	Trypsin
-14.7	3	3	25.6	ENSBTAP00000023627	Glutathione S transferase
-9.1	1	1	23.7	ENSBTAP00000025674	Crystallin beta A4
-8.4	2	2	23.3	ENSBTAP00000005341	Crystallin beta B2
-7.6	1	1	26.8	ENSBTAP00000042488	GSTM2 protein
-7.3	1	1	29.4	ENSBTAP00000020243	Carbonic anhydrase 3
-5.3	1	1	28.8	ENSBTAP00000032864	phosphoglycerate mutase 1
-4.5	1	1	26.7	ENSBTAP00000026358	Triosephosphoate isomerase
-2.6	1	1	16.7	ENSBTAP00000032372	Crystallin beta A4
-2.6	1	2	28.3	ENSBTAP00000013360	Aquaporin-0
-2.3	1	1	541.9	ENSBTAP00000013316	FAT tumor suppressor homolog 4
-2.2	1	1	43.6	ENSBTAP00000027188	Connexin 44
-1.3	1	1	192.6	ENSBTAP00000012423	NA
-1.2	1	1	313.1	ENSBTAP00000011898	centromere protein E
-1.1	1	1	106.2	ENSBTAP00000006967	zinc finger CW with cc domain 2
-1.1	1	1	530.6	ENSBTAP00000015828	NA

a- Expectation value
b- Number of unique peptides matching the identified protein
c- Total number of peptides (unique or shared) matching to the identified protein

d- Relative mass in daltons
e- Accession number of closest matching protein in the *Bos taurus* ensemble database

Table 1d. Proteins identified from band 16.

log(e)[a]	unique[b]	total[c]	Mr[d]	Accession[e]	Description
-132.7	13	16	20	ENSBTAP00000000556	Crystallin alpha B
-129.2	14	24	19.8	ENSBTAP00000004073	Crystallin alpha A
-55.7	6	6	65.8	sp|K22E_HUMAN|	Cytokeratin 2e
-52.2	6	6	65.8	sp|K2C1_HUMAN|	Cytokeratin 1
-21	3	3	24.4	sp|TRYP_PIG|	trypsin
-15.4	2	2	28	ENSBTAP00000025671	Crystallin beta B1
-14.4	2	2	62.1	sp|K1C9_HUMAN|	Cytokeratin 9
-14.2	3	3	20.9	ENSBTAP00000019664	Crystallin gamma C
-7.3	1	1	21.1	ENSBTAP00000009553	Crystallin gamma B
-4.6	1	1	59.5	sp|K1C10_HUMAN|	Cytokeratin 10
-3.2	1	1	28.3	ENSBTAP00000013360	Aquaporin 0
-3	1	1	24.5	ENSBTAP00000010119	Alpha S1 Casein
-1.2	1	1	34.7	ENSBTAP00000040267	NA

a- Expectation value
b- Number of unique peptides matching the identified protein
c- Total number of peptides (unique or shared) matching to the identified protein
d- Relative mass in daltons
e- Accession number of closest matching protein in the *Bos taurus* ensemble database

The protein alpha crystallin was one of the most abundant proteins detected, found in virtually all bands subjected to mass spectrometry analysis. A variety of peptides were derived from this protein, and a sample of the peptides derived from one band (band

4) is provided in Table 1b. The overall protein expectation value is calculated from each of the peptide expectation values, which are listed in the tables. The formula for calculating the protein expectation from the peptide values is described below.

Table 2 Peptides identified from alpha crystallin in band 4 of the gel.

Spectrum[a]	log(e)[b]	log(I)[c]	M+H[d]	Delta[e]	Z[f]	Sequence[g]	mods[h]
242.1	-10.1	2.21	1443.709	-0.021	2	MDIAIQHPWFK	M[1]15.9949 M[1] 42.0106
340.1	-8.5	2.61	1427.714	-0.023	2	MDIAIQHPWFK	M [1] 42.0106
175.1	-2.5	2.64	1037.541	-0.022	2	TLGPFYPSR	
154.1	-3	2.64	1175.627	-0.029	2	TVLDSGISEVR	
349.1	-1.9	2.23	980.5815	-0.0026	2	FVIFLDVK	
172.1	-4.6	2.49	1172.595	-0.007	2	HFSPEDLTVK	
112.1	-6.2	1.68	1224.597	0.0036	2	IPSGVDAGHSER	

a - spectrum identification number that gpm assigns to each spectrum
b - Expectation value
c - Fragment ion intensities
d - Parent ion mass
e - difference between measured and expected parent ion masses
f - charge state
g - peptide sequence
h - modified residues: 42Da is acetylation and 16 Da Oxidation

It is worth noting in Table 2 that two N-terminal peptides of alpha crystallin (MDIAIQHPWFK) were detected by the tandem mass spectrum analysis; one represented an acetylated N-terminus, while the other represented an oxidized methionine residue. An ESI spectrum analysis also produced a doubly- charged precursor ion at 722.3 (Figure 5). The dominant fragment ions are singly charged, though occasionally doubly charged ions are seen.

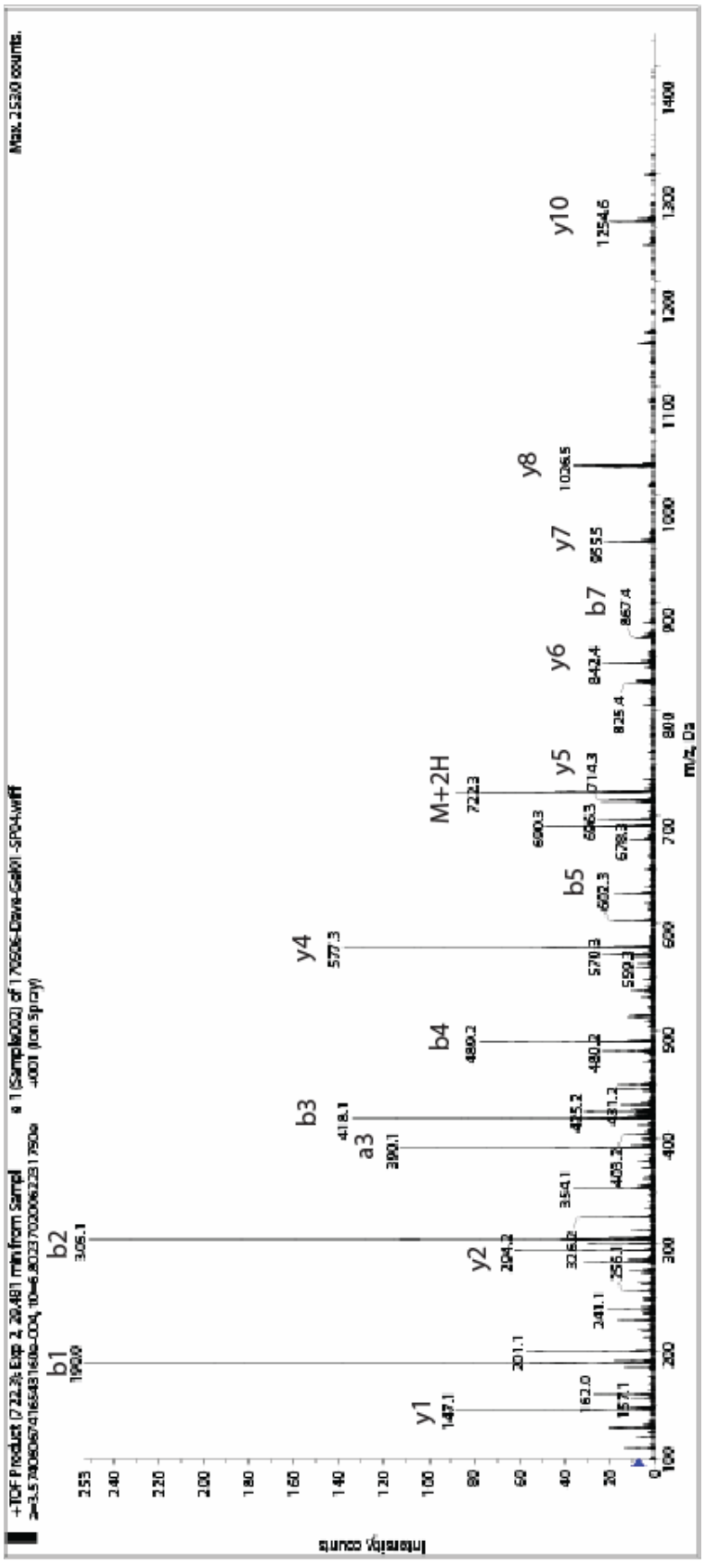


Figure 7 MS spectra of the alpha crystalin peptide Automated peptide matching depends on matching b and y ions to theoretical spectra.

As the X!Tandem program (via GPM website portal) was used for many of the protein identifications, it is worthwhile examining how the data from this program was interpreted. The scoring method is decribed in (Fenyo & Beavis, 2003) which is the source of the following description and formulas. Table 2 shows the peptides identified from the putative alpha crystallin band from the gel digest of sample 4. Note that the term "peptides" refers to the products of a proteolytic digest (typically with trypsin), whereas "protein" refers to the polypeptide or protein present in the original sample. The spectrum number is only a label that the GPM software has used to identify each of the tandem mass spectra that have been submitted. The expectation value is the chance that the matching of the mass spectrum to that particular sequence happened by chance. The protein expectation values reported in Table 1 (e.g. log -49.2 for alpha crystallin in sample 4) are based on the peptide expectation values from each set of peptides identified from a particular protein. The protein expectation value is given by:

$$E_{pro} = \left(\frac{\beta^n (1-\beta)^{s-n}}{sN^{n-1}} \right) \times \left(\prod_{j=1}^{n} e_j \right) \times \left(\prod_{i=0}^{n-1} \frac{(s-i)}{(n-i)} \right)$$

Where N = number of peptide sequences scored to find the n unique peptides

e is the peptide expectation value

β = N/total number of peptides in the proteome considered

s = number of spectra in the data set

if only one peptide match is made from the protein the $E_{pro} = e_1$.

The peptide expectation values e are estimated by X!Tandem software. X! Tandem matches spectra to sequences in a multistep process. In the first step, the spectra

are matched against every theoretical tryptic peptide present in the chosen database that matches the mass of the parent ion. The matching of the theoretical spectra depends on matching peaks actual spectra with the y and b ions in the theoretical spectra. The preliminary score is calculated by:

$$\sum_{i=1}^{n} I_i \times P_i$$

Where I_i is the measured intensity and Pi is given a value of 1 if a y or b ion is predicted and 0 if no ion is predicted at the given mass. Spectra are stored as multidimensional vectors, so a score is the dot product between the measured spectrum and the theoretical spectrum. The hyper score is calculated from the preliminary score by multiplying the preliminary score by the factorial of the number of b ions matched and the number of y ions matched:

$$\left(\sum_{i=1}^{n} I_i \times P_i \right) \times N_b ! \times N_y !$$

X! Tandem plots a histogram of all of the scores from the theoretical spectra that might match the actual spectra and assumes the highest scoring match is correct. The values plotted in the histogram are used to estimate the survival function. The survival function is defined as:

$$s_{j(x)} = \Pr(X > x) = \sum_{i=j(x)}^{\infty} p_i$$

x is the score of a particular spectrum to sequence match and $\Pr(X > x)$ is the probability of getting a score greater than x by random matching and $p_{j(x)}$ is the discrete probability of getting the score x. As presumably all but one of the matches is random, the $p_{j(x)}$ values are estimated, based on the frequencies of particular scores plotted in the

first histogram. A sample survival function is plotted in Figure 8. Finally, the expectation value p of a sequence-to-spectral match is $e_{j(x)} = ns_{j(x)}$ where n is the number of sequences scored.

In the initial spectrum- to- sequence matching process, the peptides are assumed to be unmodified tryptic peptides or to contain complete modification of a residue such as alkylation of cysteine residues. It then uses a reduced protein database only of proteins that had a tryptic match and looks for modified peptides as specified by the user as well as semi-tryptic peptides, i.e. peptides that may be produced by digest with trypsin and some other enzyme so they would have the same N or C terminal as the tryptic peptides, but would be truncated at a residue other than lysine or arginine.

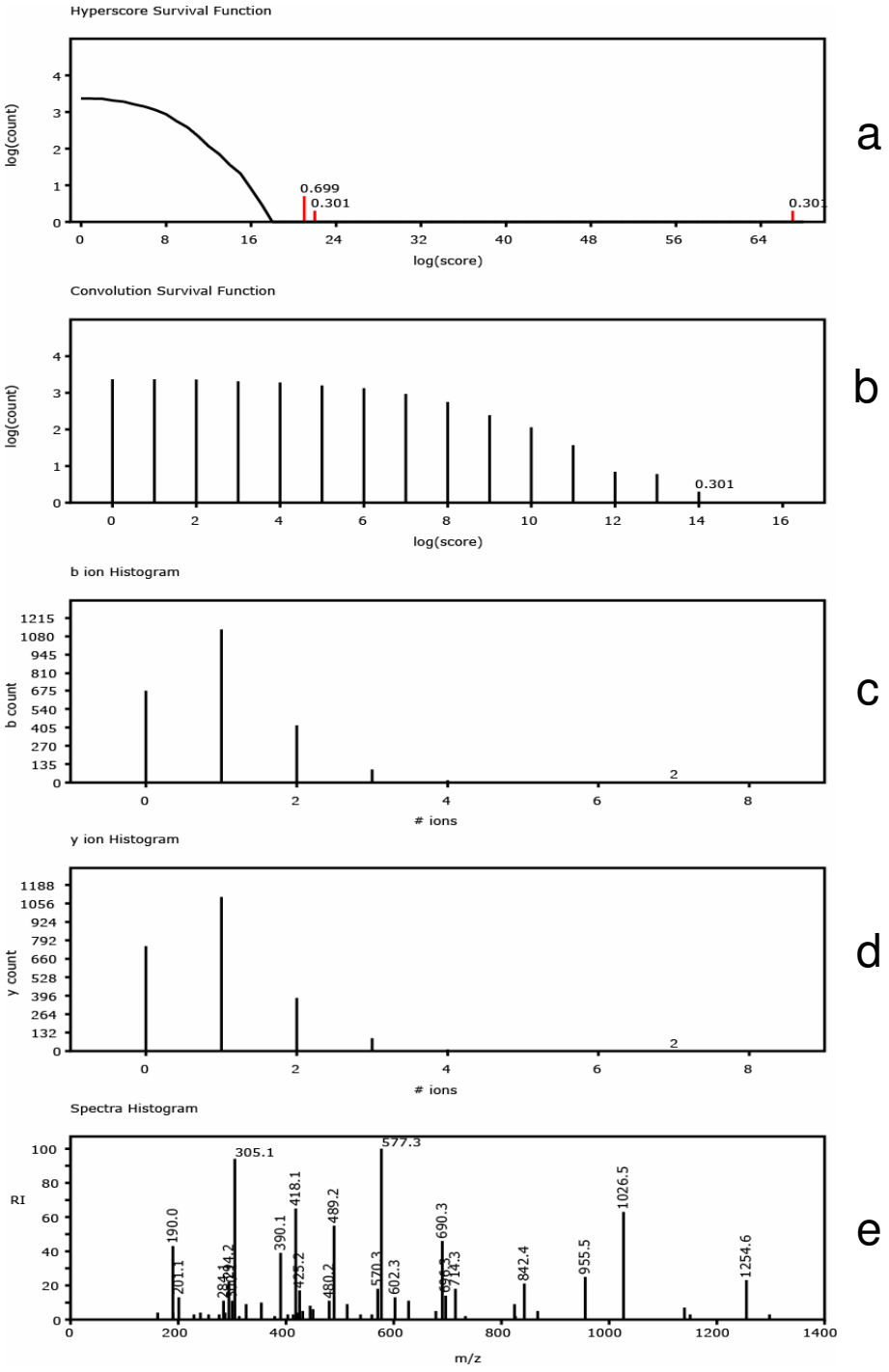


Figure 8. Graphs produced by X!tandem showing supporting data for the GPM assignment of the N-terminal peptide of alpha crystallin to the submitted spectrum.

a The graphs to the left show the evidence for the match made by GPM. The first graph shows, the number of peptides that have a particular score, the three red lines indicate the highest scoring matches with the closest matching peptide situated far to the left.

b The second graph shows the survival function of the value for the match. The graph only shows the regions with high counts. According to the graph above the matching peptide has a log score ~80. By extending the curve, essentially extending a line along the declining region of the histogram the expectation value can be determined log(e)=-10 in this case.

c and d. The next set of histograms show the number of peptides that have the same parent mass and match the given number of ions. So in the first graph, about 1080 peptides have a match of one y ion.

e. The last graph is a recreation of the original spectrum

In Solution digests

As a few groups have reported success analyzing membrane proteomw using LC-MS/MS strategies(Blonder et al., 2004; Wu & Yates, III, 2003), in solution digests of the lens membrane samples were performed concurrently with the gel digests, and the initial results were encouraging as several peptides from both Cx44 and Cx49 were identified. Based on these preliminary results, in-gel digests were subsequently replaced with an in-solution shotgun approach. The initial chromatograms produced hundreds of peaks but were dominated again by crystallins and aquaporin 0, and only a few peptides of other proteins were present. However, subsequent tryptic digests of the same samples produced much more diverse chromatograms with many more strong peaks visible. The rnumber of peaks in the spectra ranged from 800 to 1500 peaks. A number of peaks were derived from both Cx44 and Cx49. Table 3 lists the proteins that GPM identified from the tandem MS spectra taken from the second tryptic digest of the lens sample. Contamination from keratins was present, but was reduced relative to the gel samples because all of the sample processing occurs in closed microfuge tubes.

Table 3 Proteins identified by X!tandem of tandem ms spectra from in-solution digest of bovine lens proteins

log(e)[a]	% [b]	unique[c]	total[d]	Mr[e]	Accession[f]	Description
-276.6	57	30	33	54	ENSBTAP00000024572	Vimentin
-111.3	48	10	10	23.3	ENSBTAP00000005341	Crystallin beta B2
-95.6	18	12	12	43.6	ENSBTAP00000027188	Connexin 44
-69.4	47	9	10	19.8	ENSBTAP00000004073	Crystallin alpha A
-67.1	22	7	8	39.9	ENSBTAP00000029102	Connexin 50
-65.1	19	7	7	68.5	ENSBTAP00000049952	NCAM
-57.4	17	7	7	99.8	ENSBTAP00000028238	N-cadherin
-53.5	15	6	6	65.8	sp|K2C1_HUMAN|	Cytokeratin 1
-51.1	15	7	7	64.8	ENSBTAP00000006868	NA
-42.6	26	5	6	25.1	ENSBTAP00000007037	Crystallin beta A3
-39.7	36	6	8	20	ENSBTAP00000000556	Crystallin alpha B
-35.5	30	4	6	24.3	ENSBTAP00000001656	Crystallin Beta B3
-28.6	21	3	3	28	ENSBTAP00000025671	Crystallin Beta B1
-27.6	11	4	4	28.3	ENSBTAP00000013360	Aquaporin 0
-26.4	11	3	3	34.3	ENSBTAP00000022919	Ubiquitin
-23.6	19	3	4	23.7	ENSBTAP00000025674	Crystallin beta A4
-22.8	8.3	3	3	62.1	sp|K1C9_HUMAN|	Cytokeratin 9
-19.9	21	3	3	22.2	ENSBTAP00000019735	Crystallin Beta A2
-15.6	6.2	2	2	65.8	sp|K22E_HUMAN|	Cytokeratin 2e
-12.8	5.6	3	3	59.5	sp|K1C10_HUMAN|	Cytokeratin 10
-11.7	7.4	2	3	50.1	ENSBTAP00000019318	Elongation factor 1-alpha
-11.7	5.7	1	1	16.7	ENSBTAP00000032372	Crystallin beta A4

-11.1	7.7	2	2	29.4	ENSBTAP00000020243	Carbonic anhydrase 3
-9.8	1.3	2	2	274.2	ENSBTAP00000009208	SPTBN1
-9.5	9.8	1	1	20.9	ENSBTAP00000019664	Crystallin gamma C
-8.6	6	1	1	35.8	ENSBTAP00000037577	GAPDH
-7.9	3	2	2	132.4	ENSBTAP00000008864	NCAM
-7.6	14	2	2	20.9	ENSBTAP00000004084	Crystallin gamma S
-7.2	7.7	1	1	26.4	ENSBTAP00000044393	NCAM homologue
-6.9	14	1	1	12.7	ENSBTAP00000005078	Synaptobrevin-2
-6.4	9.8	2	2	45.7	ENSBTAP00000024638	Phakinin
-6.3	1.5	1	1	103.7	ENSBTAP00000053343	armadillo repeat protein
-6.2	6.5	2	2	47	ENSBTAP00000013288	flotillin-2
-6.2	4.3	1	1	33.3	ENSBTAP00000004664	CD47 antigen
-5.7	2	1	1	157.5	ENSBTAP00000053549	NA
-5.7	8	1	1	20.8	ENSBTAP00000001253	RAB2A
-5.6	4.4	1	1	40.1	ENSBTAP00000006400	Coxsackie virus receptor
-5.2	5.2	1	1	30.7	ENSBTAP00000038149	NA
-5.2	9.7	1	1	17.1	sp|MYG_HORSE|	Myoglobin
-3.9	1.5	1	1	68.5	ENSBTAP00000033759	Radixin
-3.8	4	1	1	47.6	ENSBTAP00000013135	Flotillin-1
-3.6	9.8	1	1	21.1	ENSBTAP00000018740	Crystallin gamma F
-2.6	3	1	1	45.7	ENSBTAP00000036739	Crystallin
-1.2	1.2	1	1	92.5	ENSBTAP00000053623	NA
-1.1	5.6	1	1	15.9	ENSBTAP00000027839	NA

a - expectation value

b - % sequence coverage

c - unique spectra peptide matches
d - total spectra peptide matches
e- molecular weight of protein
f- ensemble *Bos taurus* database id of protein

Sequence coverage of specific proteins.

Aquaporin 0

The most abundant membrane protein in a lens cell is Aquaporin 0 and it produced very prominent spectra in all samples. Because this protein has been studied extensively by mass spectrometry in Kevin Schey's laboratory (Schey et al., 1997b; Swamy-Mruthinti & Schey, 1997a; Han et al., 2004; Schey et al., 2000; Han & Schey, 2004b), it was relatively easy to identify many of its peptides from the bovine lens. In the initial lens digest, 6 peptides were identified as Aquaporin 0 tryptic products, and subsequent digests and analyses identified additional peptides from the sequence as well as some modified and truncated peptides. Two phosphorylation sites were identified. Figure 9 shows the spectra of peptide at m/z 1131 and its phosphorylated version 1211. The b5 ion 559.2 is prominent in both the 1131 and 1211 spectra, which indicates that the third serine (ser 235) in the peptide SVSERLSILK is the primarily phosphorylated residue. The other phosphorylated pairs were peptides that were phosphorylated on either serine 243 or 245.

Table 4 Identification of aquaporin peptides by tandem MS

log(e)[a]	m+h[b]	delta[c]	sequence	mods[d]
-2.4	975.562	0.068	SFAPAILTR	
-1.1	1131.673	0.061	SVSERLSILK	
-7.2	3165.664	0.106	LSILKGSRPSESNGQPEVTGEPVELKTQAL	N [247] 0.984016

-4.1	2752.436	0.094	LSILKGSRPSESNGQPEVTGEPVELK	N [247] 0.984016
-5.3	2611.285	0.079	GSRPSESNGQPEVTGEPVELKTQAL	N [247] 0.984016
-3.2	2198.057	0.073	GSRPSESNGQPEVTGEPVELK	N [247] 0.984016

a- expectation value
b- measured m/z
c - difference between measured mass and expected mass
d- post-translational modification in the match

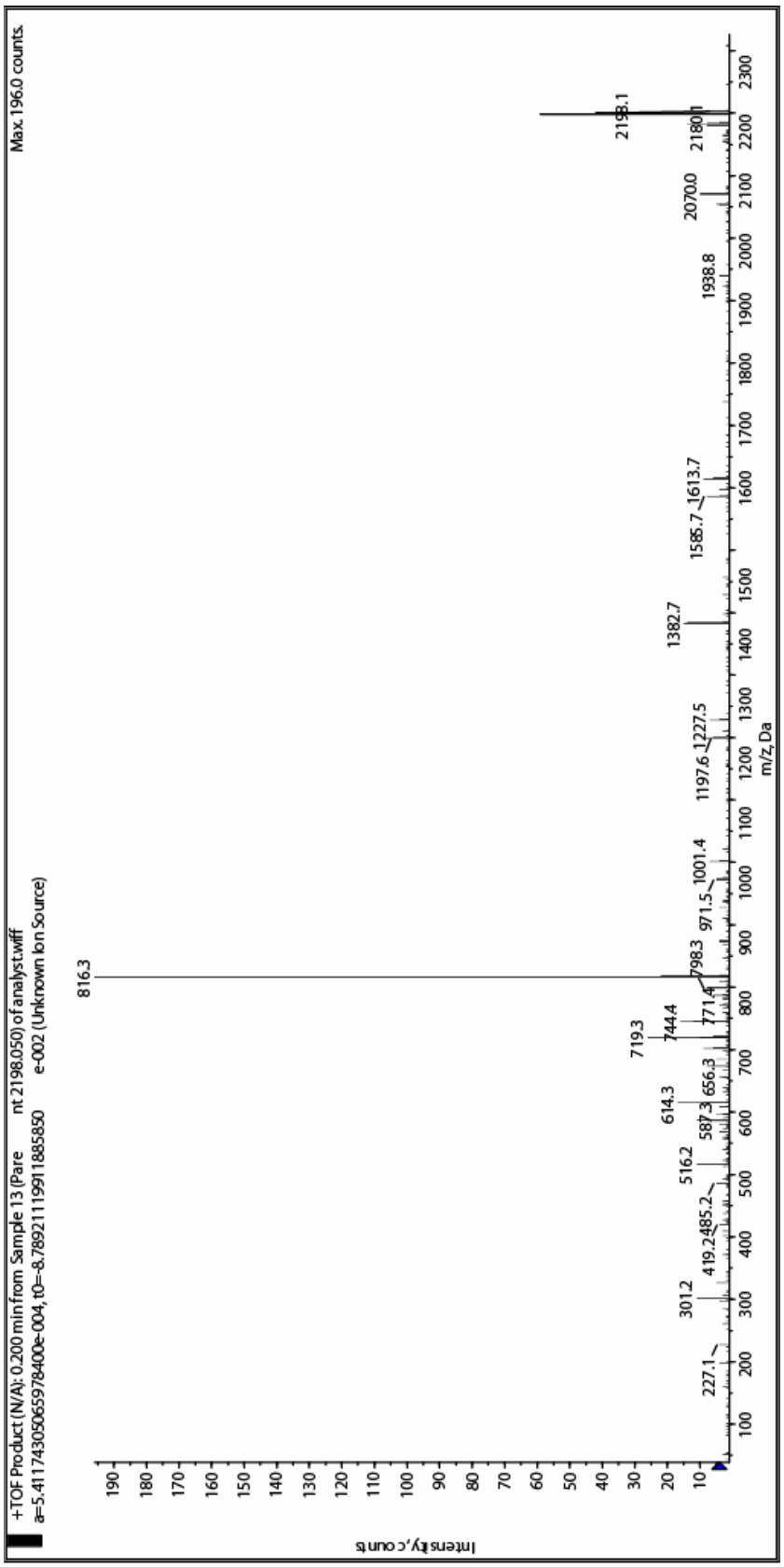


Figure 9 Tandem mass spectrum of aquaporin peptide GSRPSESNGQPEVTGEPVELK

Table 5. Masses of identified peptides derived from aquaporin 0 in bovine lens

Mr	Sequence
2198.13	GSRPSESN(1)GQPEVTGEPVELK
2611.27	GSRPSESN(1)GQPEVTGEPVELKTQAL
3164.49	SASFWRAICAEFFASLFYVFFGLGASLR
975.5	SFAPAILTR
2691.37	GSRPSESNGQPEVTGEPVELKTQAL + H2PO3
2299.2	GSRPSESN(1)GQPEVTGEPVELKT
2212.16	GSRPSESN(1)GQPE(14)VTGEPVELK
2313.22	GSRPSESN(1)GQPEVTGEPVELKT (+14)
2625.41	GSRPSESN(1)GQPEVTGEPVELKTQAL (14)
2827.41	GSRPSESN(1)GQPEVTGEPVELKTQAL (+216)
2278.13	GSRPSESN(1)GQPEVTGEPVELK + H2PO3
2000.13	LSILKGSRPSESDGQPEVT
1131.734	SVSERLSILK
1211.7	SVSERSILK + H2PO3
3723.942	SVSERLSILKGSRPSESN(1)GQPEVTGEPVELKTQAL
3165.765	LSILKGSRPSESN(1)GQPEVTGEPVELKTQAL
1794.947	GQPEVTGEPVELKTQAL
3245.73	LSILKGSRPSESN(1)GQPEVTGEPVELKTQAL + H2PO3
1827.847	GSRPSESN(1)GQPEVTGEPV
1897.893	PSESN(1)GQPEVTGEPVELK
1129.653	(154.095)SFAPAILTR
1445.635	GSRPSESN(1)GQ(1)PEVT
2334.064	GSRPSESN(1)GQPEVTGEPVELK(136)
2070.01	GSRPSESN(1)GQPEVTGEPVEL
2364.051	GSRPSESN(1)GQPEVTGEPVELK(166)
2363.069	GSRPSESN(1)GQPEVTGEPVELK(135)
2347.025	GSRPSESN(1)GQPEVTGEPVELK(149)
997.541	(21.9849)SFAPAILTR

3451.8	NFTNHWVYWVGPVIGAGLGSLLYDFLLFPR
2752.438	LSILKGSRPSESN(1)GQPEVTGEPVELK
3165.672	LSILKGSRPSESN(1)GQPEVTGEPVELKTQAL
750.359	M(16)WELR
1956.889	GSRPSESN(1)GQPEVTGEPVE
1349.683	GQPEVTGEPVELK
1212.631	ISGAHVNPAVTF (chymotrypsin digest)

A total of 35 peptides derived from aquaporin 0 were identified by manual examination of the spectra from the bovine lens digests. There are a number of truncated peptides from the C-terminal peptide as well as two phosphorylated peptides. There are also a number of spectral peaks that have adducts, with a mass shift of 21.9 that is caused by a sodium adduct and the shift of 154 is probably a DHB adduct. The other mass shifts of 136, 216, 149 are of uncertain origin, however the spectral peaks show a shift.

Connexin Peptides

Initially, attempts were made to purify the connexins by SDS-PAGE, followed by in-gel digestion. Unfortunately, this method did not obtain significant coverage of the protein sequence. In contrast, by digesting the whole membrane sample and then separating the peptides with RP-HPLC, it was possible to obtain significant coverage of the sequence of Cx44 and Cx49. A set of mass spectra from a 40 spot MALDI target contained 600 to 900 monoisotopic peaks (Figure 10). The number of peaks obtained was dependent on the number of hours the sample was permitted to digest; a 12 h digest of the proteins produced about 600 peaks, whereas a 24 h digest produced the greater number (~900) of peaks. Some samples were separated with a shallower gradient onto

100 spot targets, and while the number of distinct peaks doubled, there were no peaks from Cx44 or Cx49 that were not already detected in the 40 spot target.

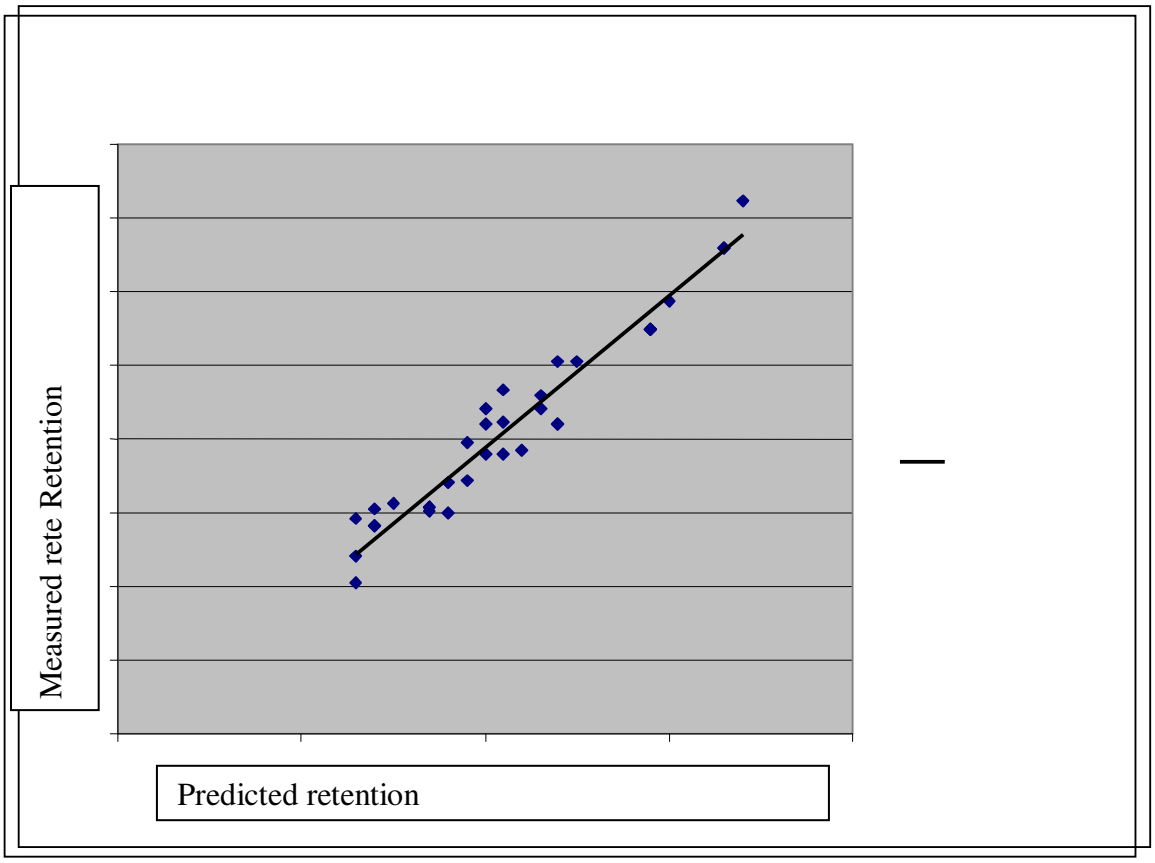


Figure 10 Number of peaks identified by sMART as having the mass and retention time to match peptides from connexin 49. The hydrophobicity retention model predicts that the peptides will elute on the line but there is some variation in the actual elution.

Peptides were identified by sMART based on retention time and mass (Table 6). Most of the Cx49 peptides were accurately identified by this analysis. However, one mis-identified peptide, 2202.9, was identified by sMART as WPCPNVVDCFVSRPTEK, but was subsequently identified by tandem MS to be a peptide from vimentin (EMEENFSVEAANYQDTIGR).

Table 6 . Peptides derived from Cx49 identified by sMART

Meas Mr	Exp Mr	Frac	exp	Sequence	Mods
773.48	773.48	20	19	ILPLYR	
800.481	800.475	18	17.1	LEGTLLR	
826.407	826.407	14	15.2	QELTPEK	Q/C-17
1066.584	1066.577	20	21	VGPGPLGDLSR	
1103.643	1103.645	24	25.2	FRLEGTLLR	
1107.544	1107.545	19	19.8	GYQLLEEEK	
1263.626	1263.622	13	10.3	APLAADQGSVKK	S+80
1348.673	1348.675	18	15	SLHSIAVSSIQK	S+80
1306.67	1306.677	21	23.3	AKGYQLLEEEK	
1309.637	1309.64	13	14.6	EVEKEEPPPEK	
1490.802	1490.809	19	17.2	RPVEQPLGEIPEK	
1490.803	1490.809	21	21.2	RPVEQPLGEIPEK	
1686.753	1686.744	14	14.1	EAEELSQQSPGNGGER	
1687.732	1687.728	14	14.1	EAEELSQQSPGNGGER	Q+0.98
1767.697	1767.695	13	12.1	EAEELSQQSPGNGGER	S+80 Q+0.98
1847.677	1847.661	13	12.1	EAEELSQQSPGNGGER	S+80 S+80 Q+0.98
1730.883	1730.894	23	22.9	VSTEGQETLAVLEVEK	
1924.047	1924.042	21	19	SAFKRPVEQPLGEIPEK	
1972.896	1972.872	15	15.7	EREAEELSQQSPGNGGER	Q+0.98
2201.951	2201.927	22	19.2	WPCPNVVDCFVSRPTEK	W+16*2 S+80
2217.105	2217.105	20	22.1	VEPPEVEKEVEKEEPPPEK	
2471.205	2471.208	34	36.2	GDWSFLGNILEEVNEHSTVIGR	
2503.205	2503.198	34	36.2	GDWSFLGNILEEVNEHSTVIGR	W+16*2
2725.297	2725.279	18	17.1	EAEELSQQSPGNGGERAPLAADQGSV K	Q+0.98
2805.257	2805.245	17	15.1	EAEELSQQSPGNGGERAPLAADQGSV K	S+80 Q+0.98
2741.506	2741.507	24	21	RPVEQPLGEIPEKSLHSIAVSSIQK	
2821.473	2821.474	24	21	RPVEQPLGEIPEKSLHSIAVSSIQK	S+80

2853.377	2853.374	17	15.4	EAEELSQQSPGNGGERAPLAADQGSV KK	Q+0.98
2933.335	2933.34	17	15.4	EAEELSQQSPGNGGERAPLAADQGSV KK	S+80 Q+0.98
3013.321	3013.307	17	15.4	EAEELSQQSPGNGGERAPLAADQGSV KK	S+80 S+80 Q+0.98
3418.635	3418.639	25	25.3	SQEAERVSTEGQETLAVLEVEKVEPPE VEK	S+80
3366.694	3366.684	30	29.4	IVSHYFPLTEVGMVEASPLSAKPFSQFE EK	
3382.674	3382.679	29	27.4	IVSHYFPLTEVGMVEASPLSAKPFSQFE EK	M+16
3446.668	3446.651	29	27.4	IVSHYFPLTEVGMVEASPLSAKPFSQFE EK	S+80
3526.648	3526.617	29	27.4	IVSHYFPLTEVGMVEASPLSAKPFSQFE EK	S+80 S+80
3731.736	3731.732	23	22.1	AYQETLPSYAQVGAQEGVEEEQPVEAA AEPEVGEK	
4456.239	4456.219	33	33	GYQLLEEEKIVSHYFPLTEVGMVEASPL SAKPFSQFEEK	
4472.224	4472.214	33	33	GYQLLEEEKIVSHYFPLTEVGMVEASPL SAKPFSQFEEK	M+16

Table 7 lists all of the peptides identified for Cx49 by tandem MS. The GPM expectation values are given for spectra that GPM software was able to match; however a few of the spectral peaks were only matched by sequencing manually. Most of the unidentified spectral peaks were ones that had a non-tryptic cleavage such as 1205 IVSHYFPLTE, which GPM could not be expected to identify. Other spectral peaks that were not identified by GPM represent multiple missed cleavages, which GPM would have identified if the settings were changed. A few peaks were not identified because the intensity of the peak had a low signal to noise ratio, such as was seen in the acetylated N-terminal peptide and some of the multiply- phosphorylated peptides.

Table 7 Cx49 peptides detected by tandem MS and identified by GPM

Location	Sequence[a]	Mass[b]	Measured masses[c,d]	GPM[e]
2 – 22	GDWSFLGNILEEVEHSTVIGR	2472.278	2472.278, 2514.2	-6.1
108 – 125	EREAEELSQQSPGNGGER	1973.881	1973.887, 2053.846[2]	-4.1
110 – 125	EAEELSQQSPGNGGER	1688.737	1688.745, 1768.702[2], 1848.683[3]	-5.7
110 – 137	EAEELSQQSPGNGGERAPLAADQGSVKK	2853.397	2854.383, 2934.34[2], 3014.341[3]	
126 – 137	APLAADQGSVKK	1184.664	1184.665, 1264.627[2]	-5.9
148 – 156	FRLEGTLLR	1104.653	1104.658	-1.1
241 – 257	SAFKRPVEQPLGEIPEK	1925.049	1925.071	
241 – 269	SAFKRPVEQPLGEIPEKSLHSIAVSSIQK	3175.748	3175.75, 3255.624[2]	-4.8
245 – 157	RPVEQPLGEIPEK	1491.817	1491.9, 1492.8	-1.9
245 – 269	RPVEQPLGEIPEKSLHSIAVSSIQK	2742.817	2742.607, 2822.502[2], 2902[3], 2982[6], 3062[7]	-3.0
258 – 269	SLHSIAVSSIQK	1269.287	1269.721	-6.5
258 – 271	SLHSIAVSSIQKAK	1468.849	1468.910	-3.8
258 – 276	SLHSIAVSSIQKAKGYQLL	2043.160	2043.16	
258 – 284	SLHSIAVSSIQKAKGYQLLEEEKIVSH	2994.626	2994.626	
258 – 290	SLHSIAVSSIQKAKGYQLLEEEKIVSHYFPLTE	3744.985	3744.98	
270 – 290	AKGYQLLEEEKIVSHYFPLTE	2494.287	2494.278	-7.3
272 – 280	GYQLLEEEK	1108.552	1108.553	-3.5
272 – 284	GYQLLEEEKIVSH	1544.796	1544.794	-5.6
272 – 286	GYQLLEEEKIVSHYF	1854.927	1854.93	
272 – 290	GYQLLEEEKIVSHYFPLTE	2295.154	2295.16	

281 – 290	IVSHYFPLTE	1205.621	1205.619	-1.3
281 – 310	IVSHYFPLTEVGMVEASPL SAKPFSQFEEK	3367.692	3367.685, 3383.701[4], 3447.667[2], 3527.651[3]	-8.5
311 – 321	VGPGPLGDLSR	1067.585	1067.552	-3.5
322 – 256	AYQETLPSYAQVGAQEGV EEEQPVEAAAEPEVGEK	3732.740	3732.715	-8.2
363 – 378	VSTEGQETAVLEVEK	1731.902	1731.883	-2.5
379 – 397	VEPPEVEKEVEKEEPPPE K	2218.113	2218.113	-8.8
387-397	EVEKEEPPPEK	1310.648	1310.646	-2.7

a - list of the peptides detected in the sample that could be assigned to Cx50 from the tryptic and chymotryptic digests.
b- the theoretical parent (M+H) mass
c - lists the measured mass and the masses of peptides with modifications.
d- the superscripts represent the shift in masses due to modifications 1) Acetylated 2) singly phosphorylated 3) Doubly phosphorylated 4) Oxidized Methionine 5) glutamine to pyroglutamine 6) triply phosphorylated 7) quadruply phosphorylated
e- the log of the expectation value reported by the GPM.

Table 8 Connexin 44 peptides identified by tandem MS/MS and GPM analysis (see Table 7 for description of column values)

Location	Sequence	Expect	Measured Masses	GPM
2 – 9	GDWSFLGR	937.453	937.453, 979.504[1]	-2.0
106 – 115	RKEREEEPPK	1297.686	1297.674	-1.2
106 – 132	RKEREEEPPKAAGPEGHQDPAPVRDDR	3066.510	3066.663	-2.7
107 – 115	KEREEEPPK	1141.585	1141.568	-1.1
108 – 115	EREEEPPK	1013.490	1013.486	-2.3
108 – 124	EREEEPPKAAGPEGHQD	875.847	1875.837	-9.2
108 – 129	EREEEPPKAAGPEGHQDPAPVR	2396.159	2396.171	-8.1

108 – 132	EREEEPPKAAGPEGHQDPAPVRDDR	2782.314	2782.375	-1.9
110 – 115	EEEPPK	728.346	728.339	
110 – 129	EEEPPKAAGPEGHQDPAPVR	2111.016	2111.013	-7.0
110 – 132	EEEPPKAAGPEGHQDPAPVRDDR	2497.171	2497.182	
116 – 129	AAGPEGHQDPAPVR	1401.687	1401.694	-3.6
116 – 132	AAGPEGHQDPAPVRDDR	1787.842	1787.835	
116 – 134	AAGPEGHQDPAPVRDDRGK	1972.959	1972.948	-3.6
144 – 152	TYVFNIIFK	1144.641	1144.634	-3.0
226 – 242	LKQGMTSPFRPDTPGSR	1874.955	1874.943, 1954.912^{2}, 2034.887^{3}, 1890.939^{4}, 1970.915^{2,4}, 2050.885^{3,4}	-2.3
228 – 242	QGMTSPFRPDTPGSR	1633.775	1633.79, 1713.761^{2}, 1793.736^{3}, 1649.775^{4}, 1729.74^{2,4}, 1809.705^{3,4}, 1616.742^{5}, 1776.682^{2,5}	
243 – 255	AGSVKPVGGSPLL	1181.689	1181.702, 1261.660	-4.4
316 – 331	AQNWANREAEPQTSSR	1845.847	1845.851, 1925.845	-4.3

The raw mass spectra were processed by peak picking software that selects monoisotopic peaks with a signal to noise ratio greater than 3.5. The software also eliminates duplicate peaks in adjacent fractions and peaks that are due to sodium adducts, loss of water, etc. These peak lists are then analyzed by software based on the sequence specific retention calculator (Krokhin, 2006). Allowing a mass error of 10 ppm and a

retention error of +/- 3 minutes, 27 possible matches were made to peptides derived from Cx49 including 12 peptides that contained post-translational modifications (Table 7). The software also identified 14 possible matches from Cx44 including 5 post-translationally modified peptides (Table 8).

Confirmation of peptide identity by tandem ms.

The identity of peptides that were chosen on the basis of retention time and mass was confirmed by sequencing the peptides with Tandem MS/MS. The retention time constraint was also eliminated and any possible peaks from either Cx44 or Cx49 that may have eluted outside the predicted time were sequenced as well. In general, these turned out to be peptides from other proteins; one exception was the N-terminal peptide from Cx49. In some chromatographic runs, it eluted later than predicted, while in others, it eluted in the predicted fraction. The tandem mass spectra were sequenced by hand and also submitted to the GPM to get an objective measure of confidence in the identity of a particular peptide. In some cases, the level of confidence is quite low e.g. FRLEGTLLR from Cx 49 has a log e value of 1.1. Manual inspection of all of the matching spectra was done to confirm the sequences.

Identification of post translational modifications

Several common modified peptides were observed including the oxidation of methionine residues (+15.99 Da) and the shift of glutamine to pyroglutamine (-17.03) in peptides where the initial N-terminal residue was glutamine. The Asn121 of connexin 49 had a loss of 0.98 Da, which is characteristic of a deamidation.

Both Cx44 and Cx49 were modified at the N-terminus. The initial Met residue was cleaved from both connexins. Both N-termini were acetylated as well. The acetylation increased the retention time of the peptides and shifted the mass by 42.01 Da. In the MS/MS spectra of the acetylated peptides, the Y ions remain identical to the non-acetylated version but all of the b-ions are shifted up by 42 Da.

The initial analysis of the MS spectra was not be able to identify any peptides that showed non-tryptic cleavage, such as would be present in protein that had been proteolytically cleaved within the cells. A list of peaks that could not be assigned to any proteins on the basis of mass and retention time was generated and tandem MS spectra taken. A few of these turned out to be from Cx44 and Cx49. Two non-tryptic peptides were found from Cx44; one from the cytoplasmic loop region ending at Asp124 and another from the C-terminus regions at Leu255. Non-tryptic peptides from Cx49 terminated at His284 and Glu290.

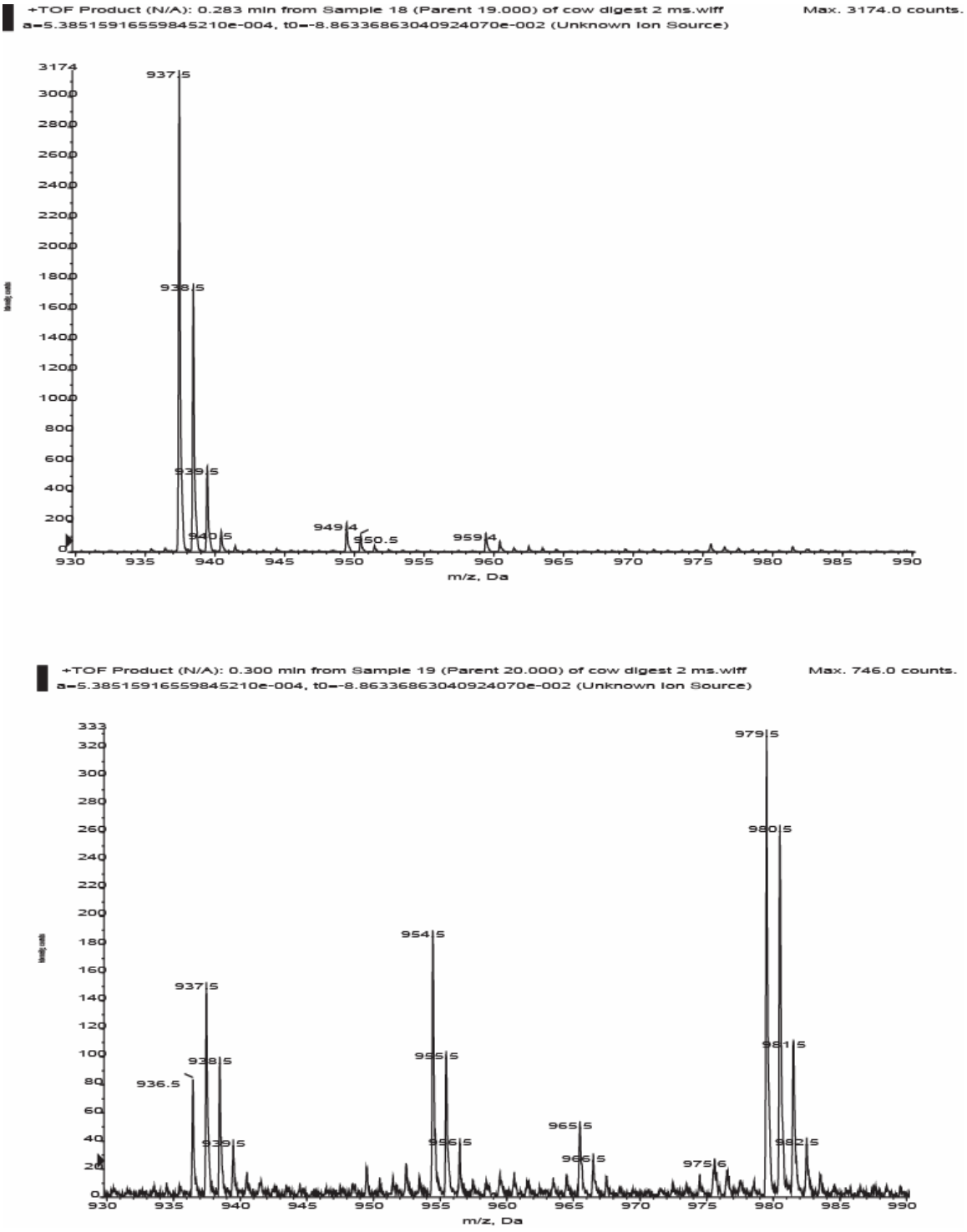


Figure 11. MS of N terminal peaks of Cx44. The top spectrum shows the unacetylated version peak 937.5 from fraction 18; the bottom spectrum is from the subsequent fraction and shows the acetylated version 42 daltons higher at 979.5. There is also a small amount of non acetylated peptide in this fraction.

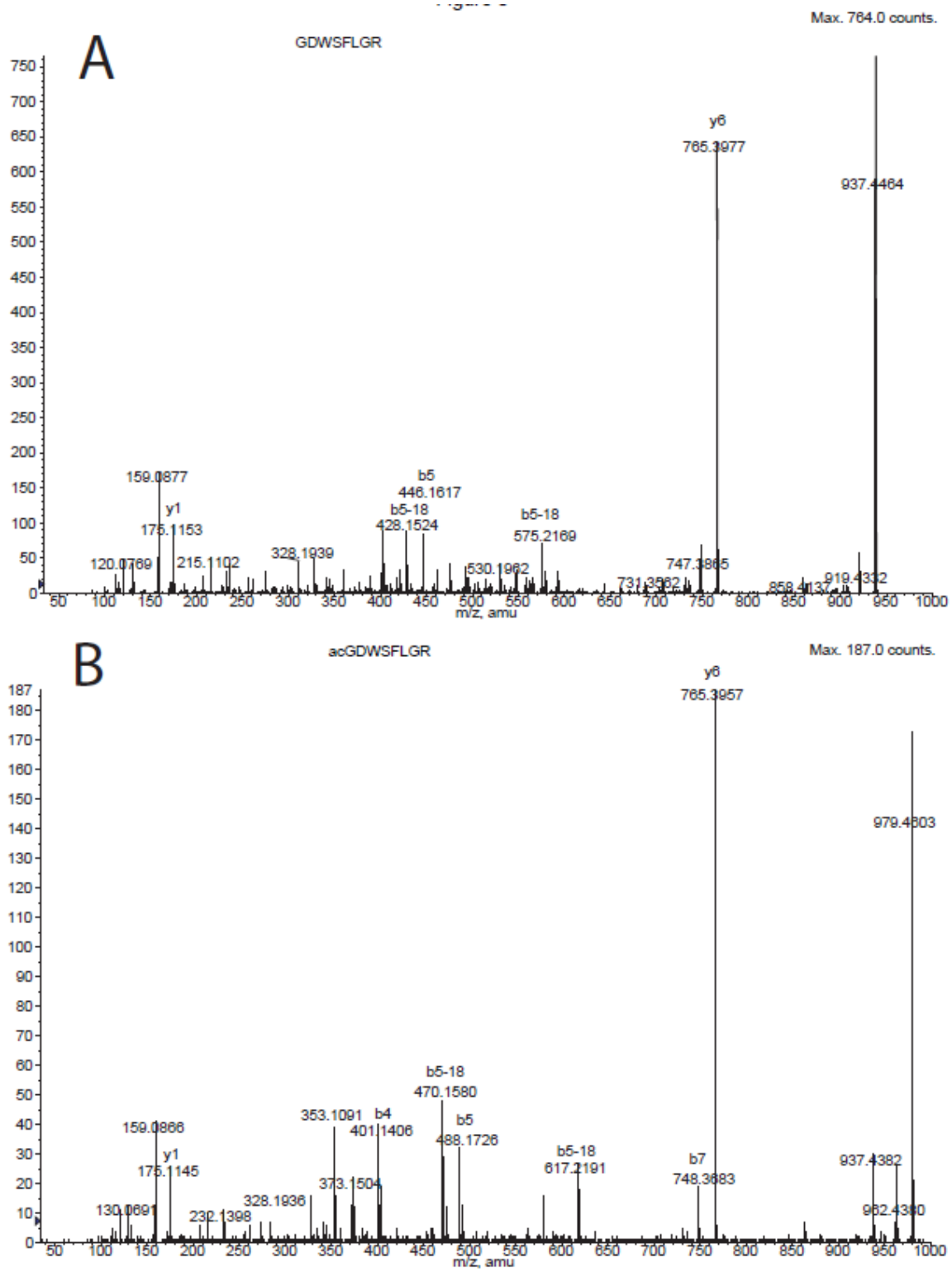


Figure 12. Tandem MS of acetylated and non-acetylated N-terminus of Cx44. The spectra have the same y ions which is to be expected because they lacked the N terminal residues. The b ions and their derivatives are all shifted 42 Da higher.

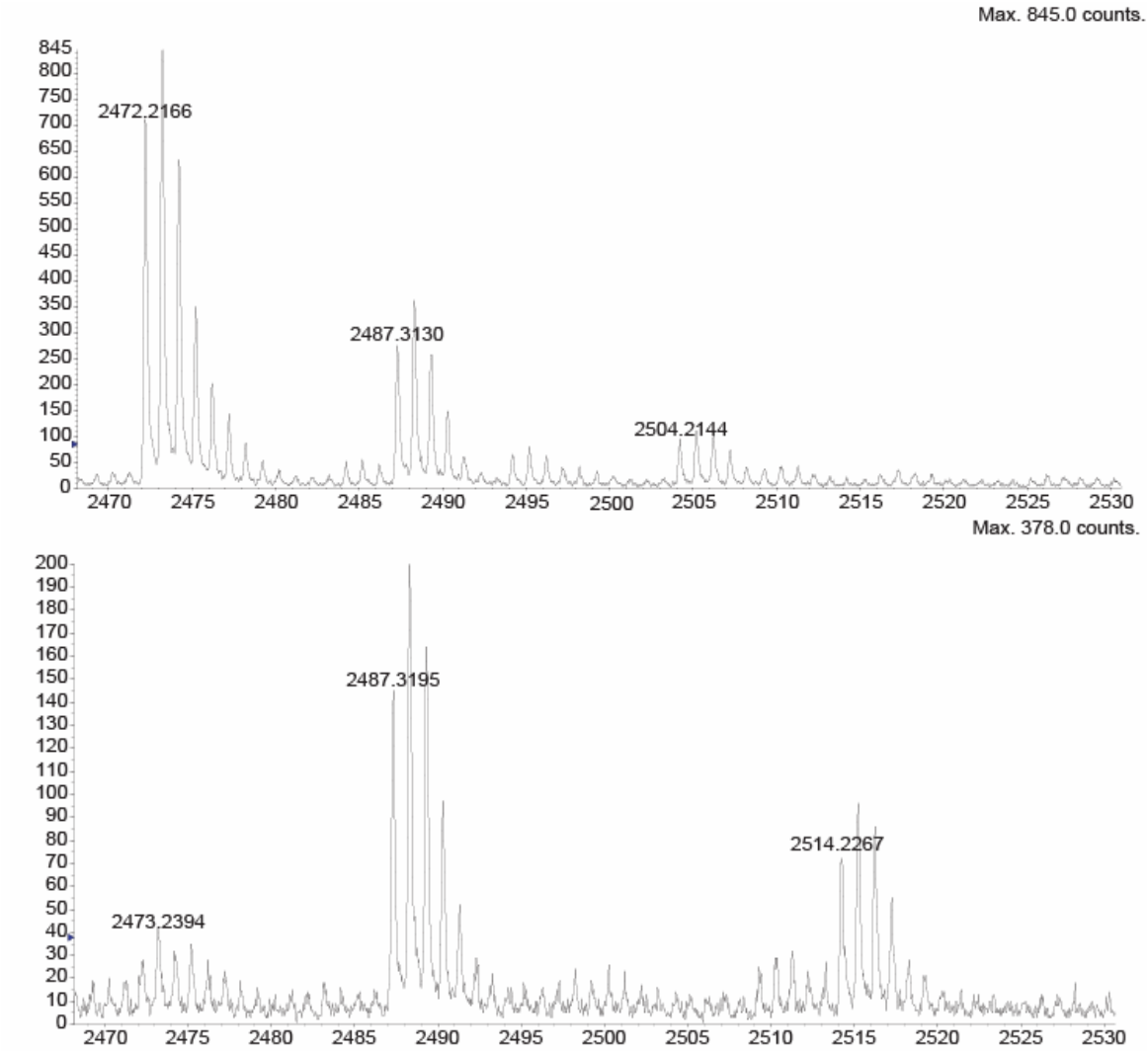


Figure 13. MS showing non- acetylated and acetylated N-terminus of Cx49. The non- acetylated version of the peptide is in the first spectrum with a mass of 2472.2. The acetylated version is present in the subsequent fraction with a mass of 2414.2

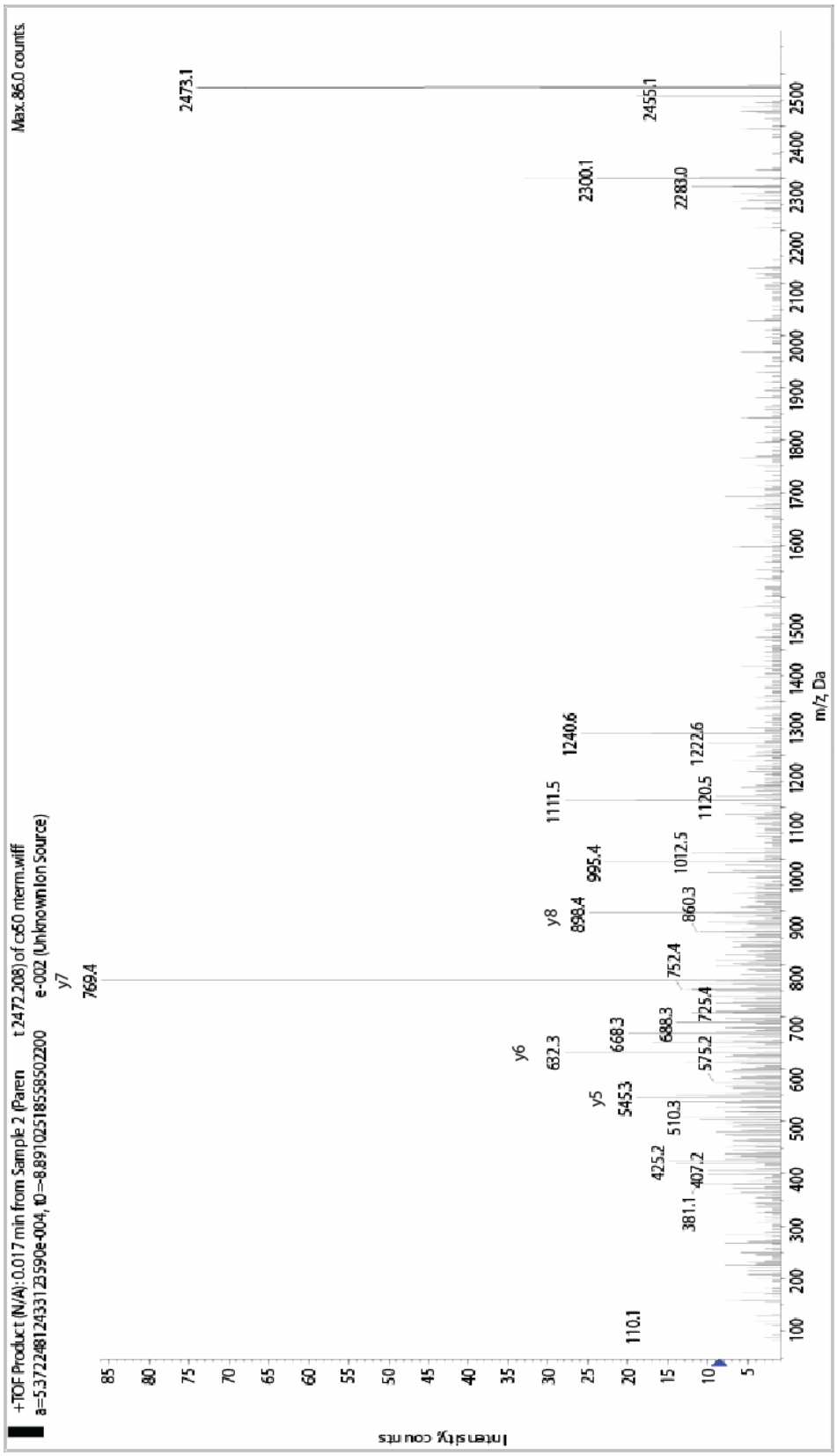


Figure 14. Tandem MS spectrum of N-terminal peptide of connexin 50

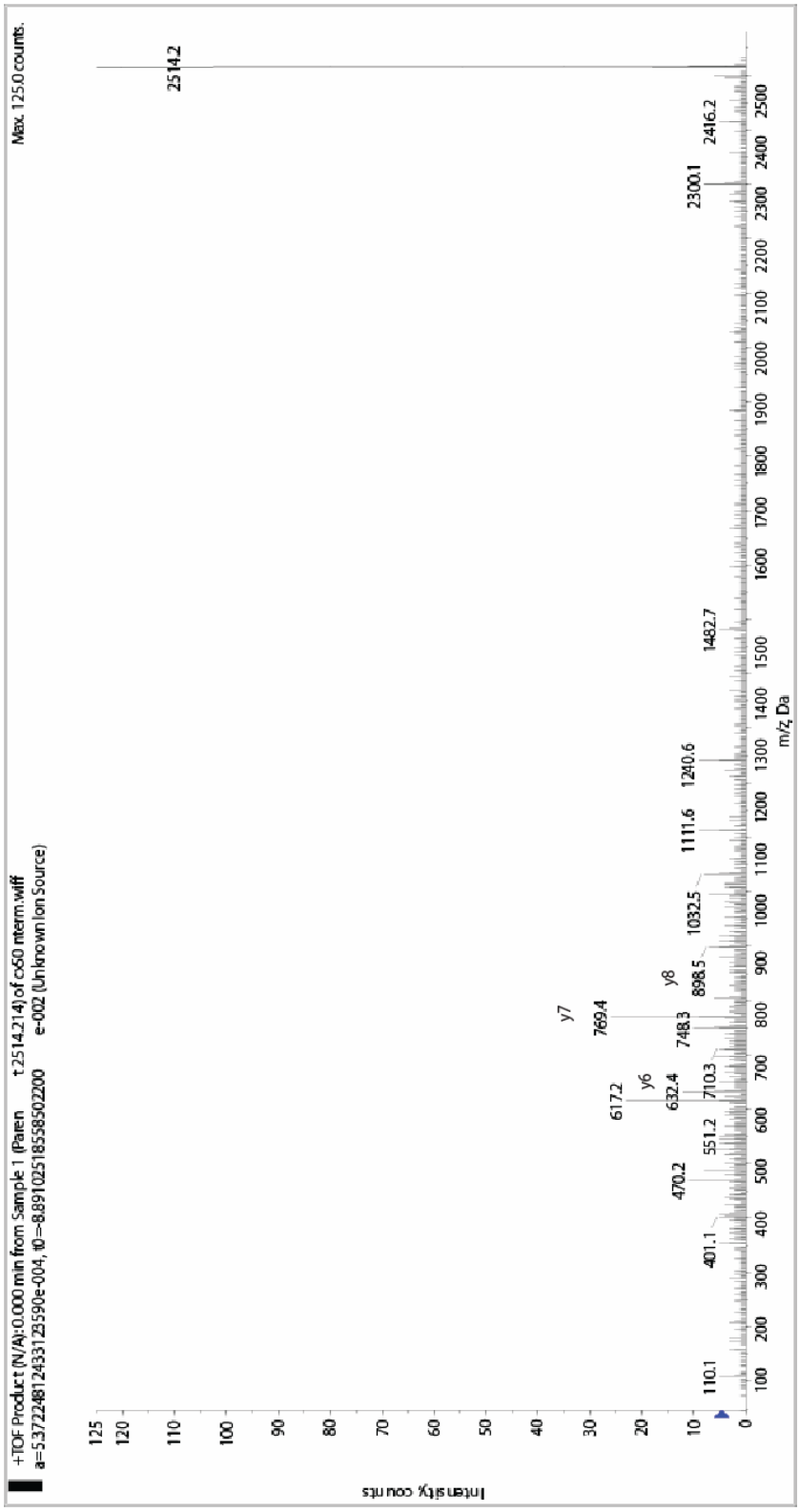


Figure 15. Tandem ms spectra of acetylated n-terminus connexin 50. This spectrum has relatively low counts and only the y ion predominate. They match the y ions from the prevoius spectrum.

The masses of phosphorylated peptides are shifted from the non-phosphorylated version by a multiple of 79.996 Da (Figure 13). In the 40 spot analyses, the phosphorylated peptides generally eluted in the same or the preceding fraction. The presence of a phosphorylated residue was confirmed by MS/MS and in most cases the specific residue that is phosphorylated was identified. Peptides that are phosphorylated on a serine or threonine residue usually show a strong loss of phosphoric acid (98) in their fragmentation pattern. The resulting shift in the ions of –18 Da compared with the non-phosphorylated peptide can indicate which residue is phosphorylated (Figure 14). Nine phosphorylated serines were identified in Cx49. Three of the phosphorylated serines were from the cytoplasmic loop region Ser115, Ser118, and Ser134; the other sites were from the C-terminus, Ser258, 261, 265, 266, 297, and 300. In Cx44, two serines (Ser241 and Ser 245) and three threonines (Thr238, 300, and 303) were phosphorylated. There was one additional phosphorylated peptide in Cx44, which was phosphorylated on Thr328 ,Ser 329 or Ser330, but it was not possible to determine which residue was phosphorylated.

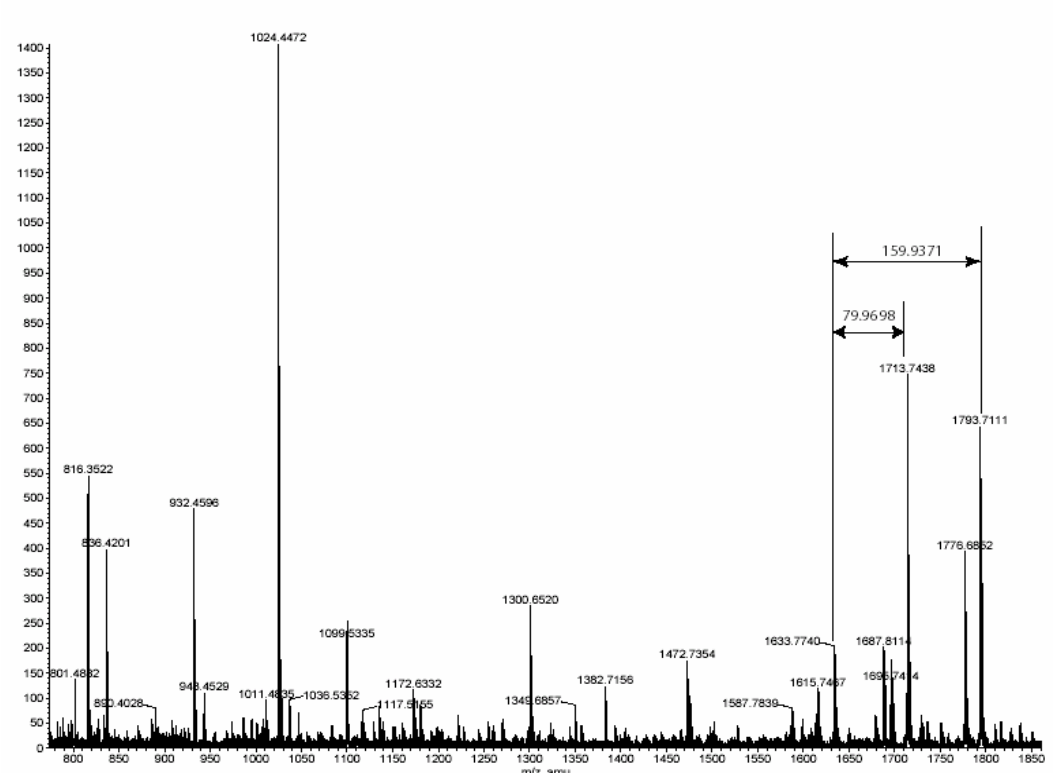


Figure 16. Mass spectrum showing peaks representing singly and doubly phosphorylated peptides. The non- phosphorylated peptide has a peak at 1633.7, the single phosphorylated peak is ~80 Da higher at 1713.7 and the double phosphorylated peak is at 1793.7

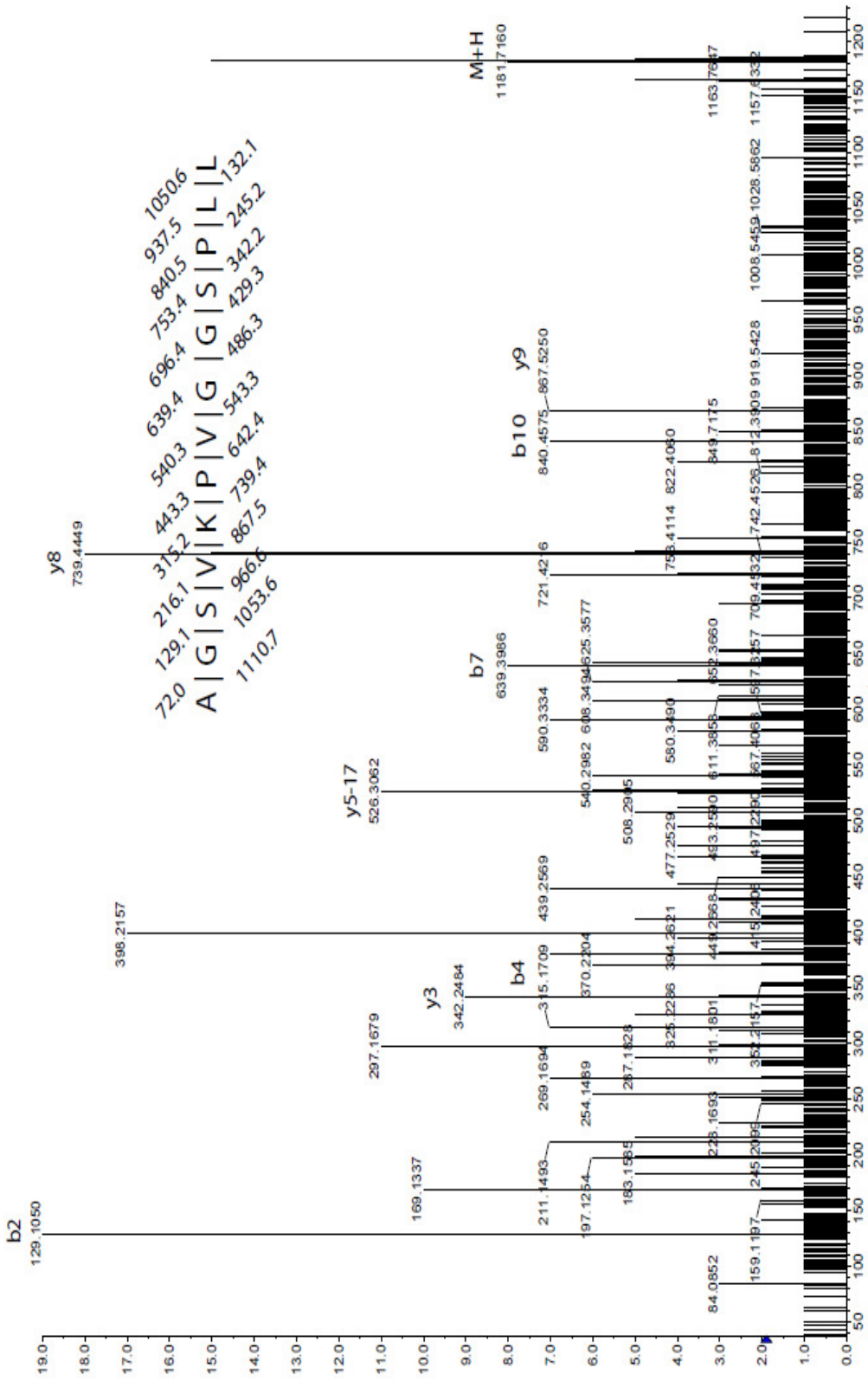


Figure 17. Tandem MS of phosphorylated and truncated peptides from connexin 44

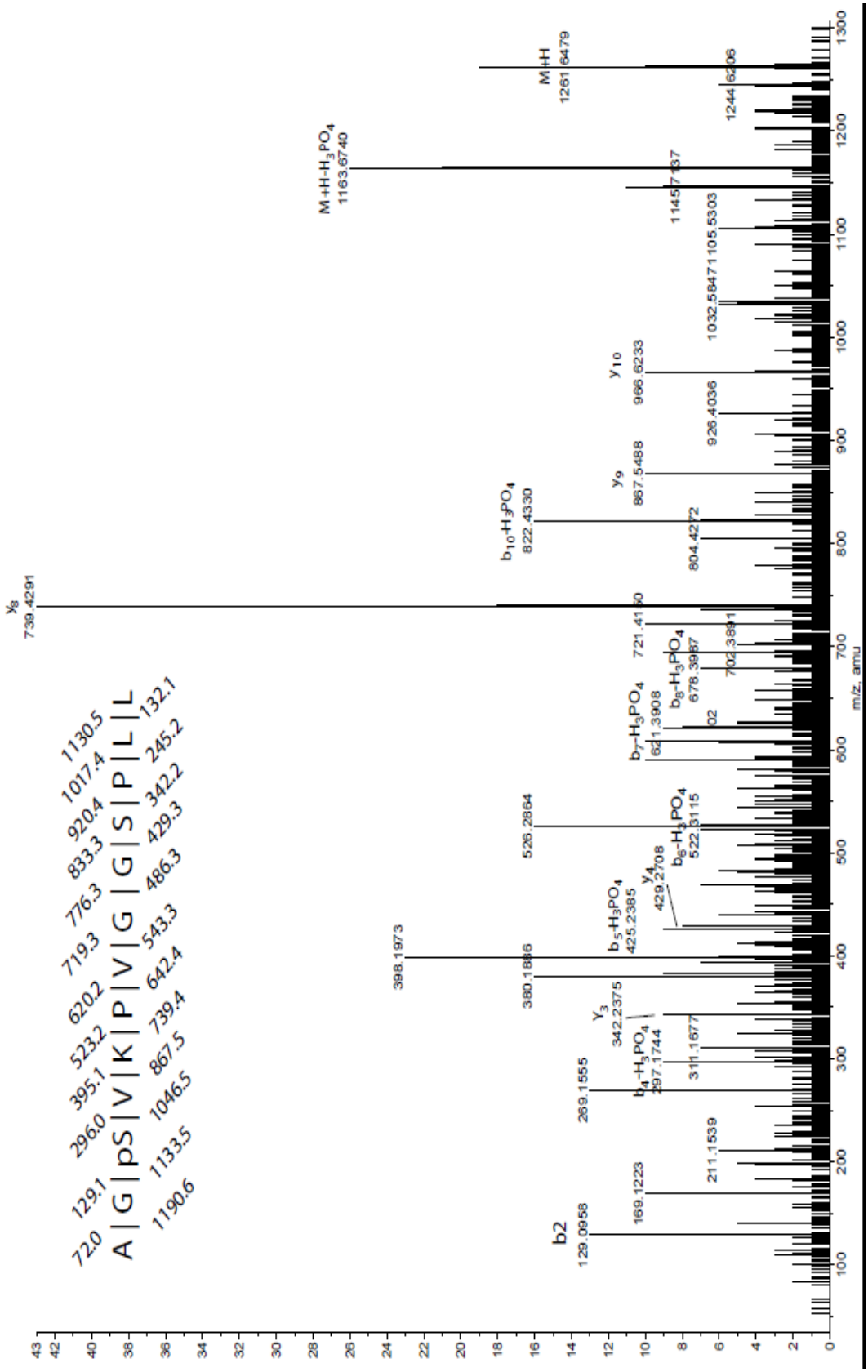


Figure 18. Tandem MS of phosphorylated and truncated peptides from connexin 44

Chymotryptic digest

Cx44 has large regions in its C-terminus that do not contain Lys or Arg residues. To get some coverage of this region of the protein, a digest with chymotrypsin was performed. Two peptides from this region were identified in the chymotryptic digest. One fragment, spanning residues 377-388 (HAPPEPPADPGR), was probably cut on the C-terminal by contaminating trypsin within the chymotrypsin reagents. The other peptide 281-312 (GQASAPGYPEPPPPAALPGTPGTPGGGGNQGL) was found both singly and doubly phosphorylated (Thr 300 and Thr303). All the spectra are shown below because they have some interesting properties. Normally, the phosphate is extremely labile in collision-induced disassociation and is lost immediately along with water for a total loss of 98 Da. However, these spectra show fragments with the phosphate group still attached. The consistent b ions show that the phosphates are attached in the C-terminal regions of the peptide which can only be the two threonine residues.

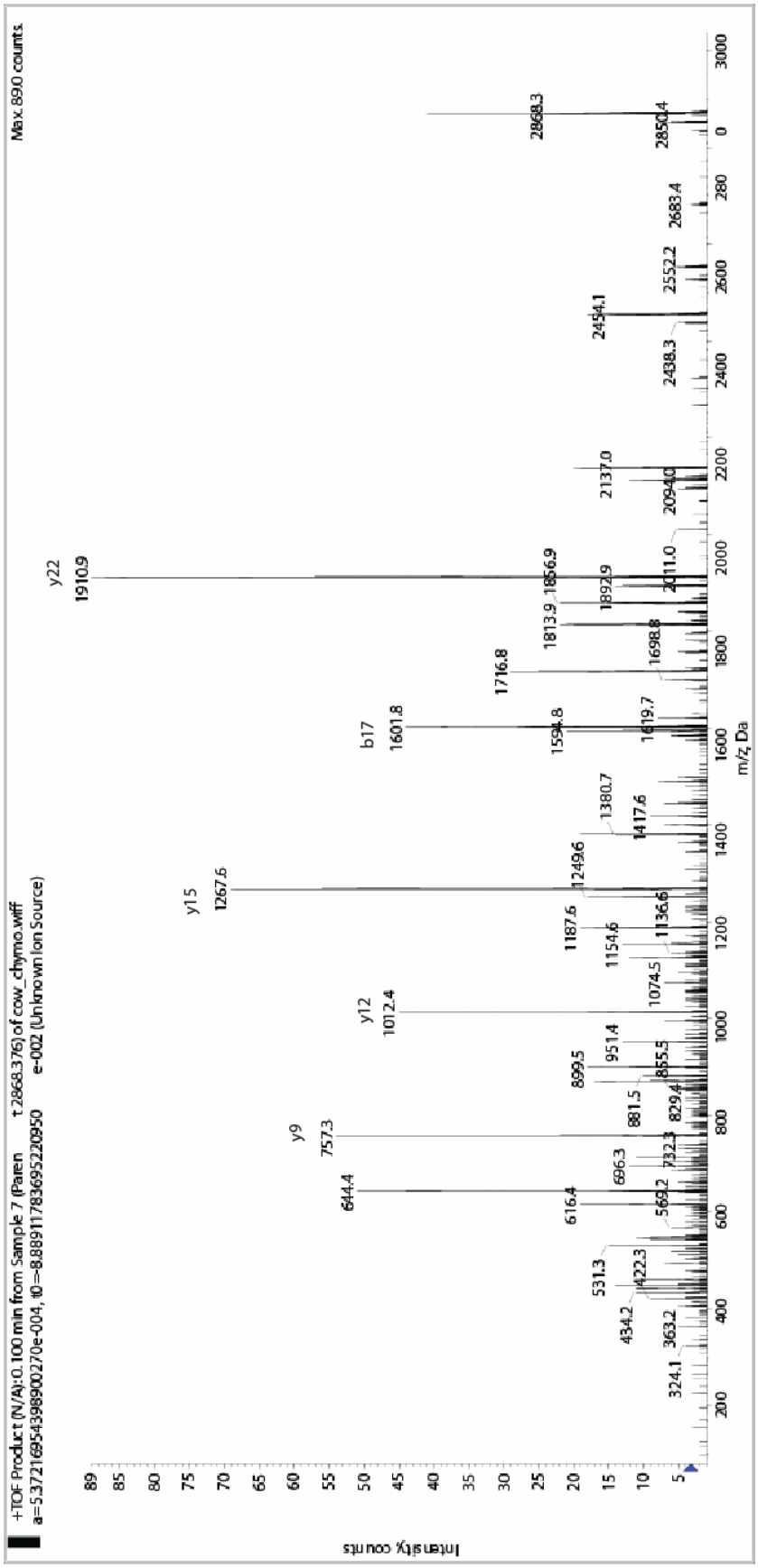


Figure 19. Tandem MS of GQASAPGYPEPPPPAALPGTPGTPGGGGNQGL

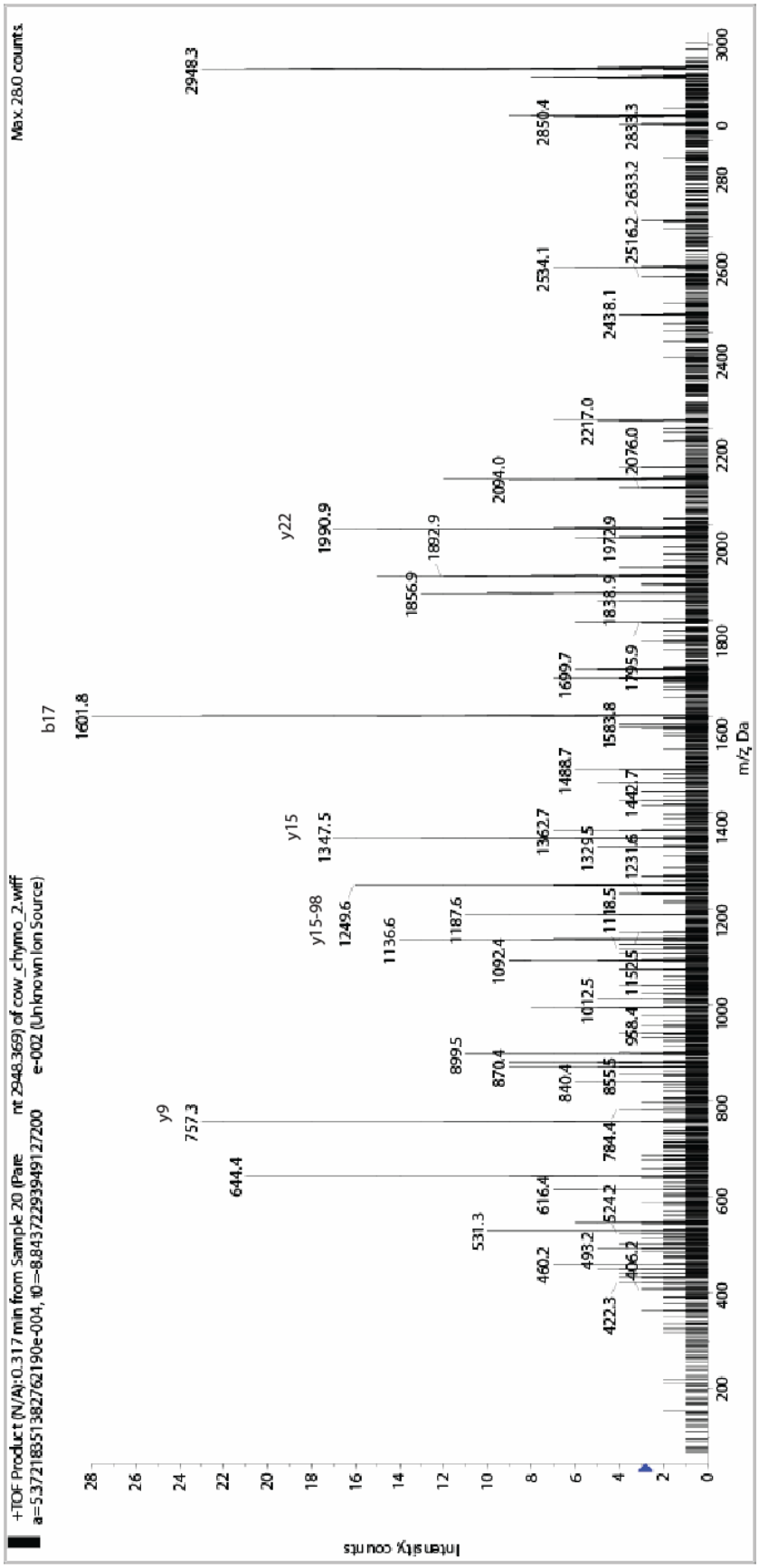


Figure 20. Tandem MS of singly phosphorylated peptide
Several of the ions still contain the phosphate group e.g. y22 is 1990.9 80 more than the y22 ion in the previous spectrum

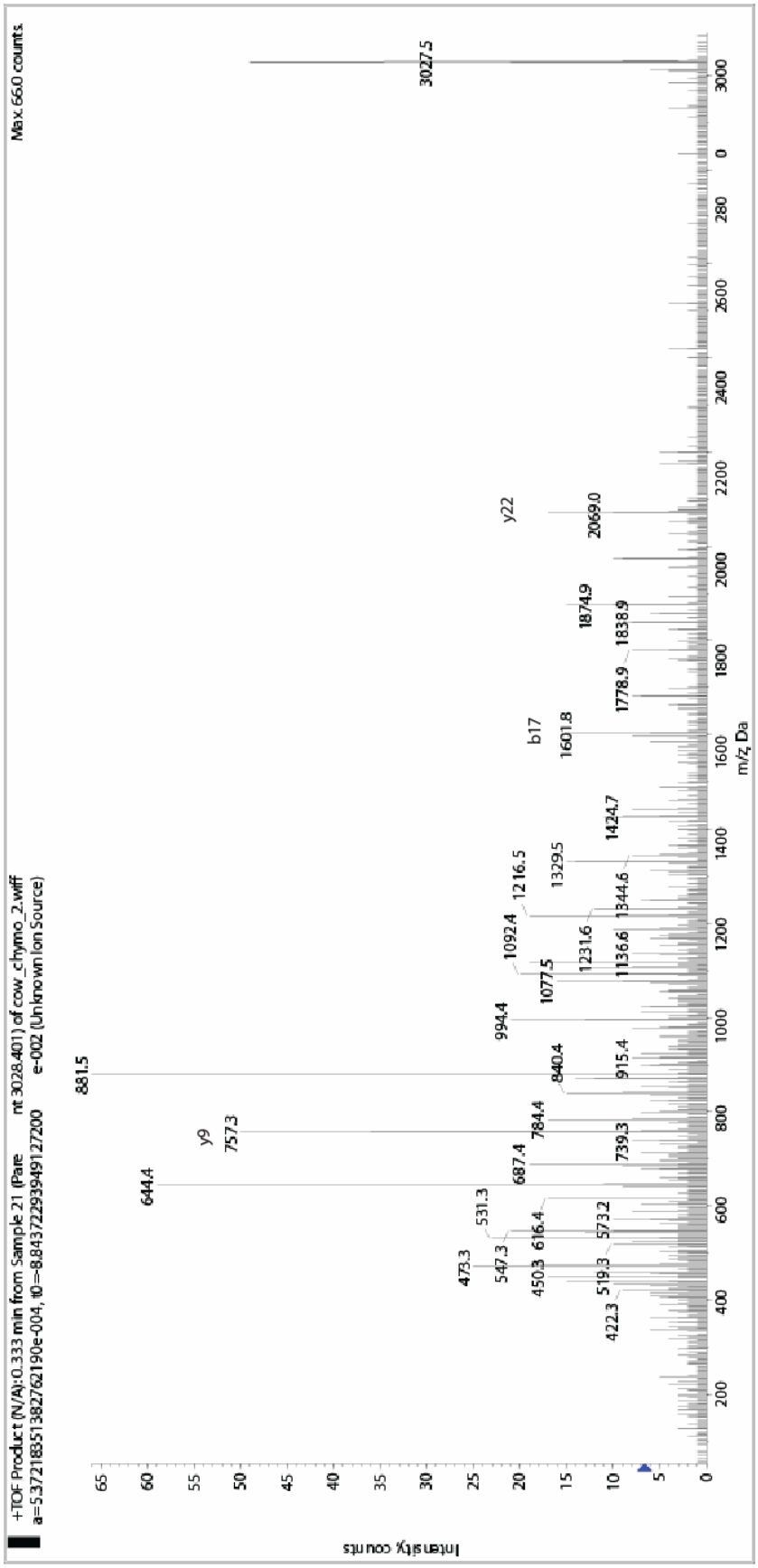


Figure 21. Tandem MS of doubly phosphorylated peptide

Figures 22-23 summarize which regions of the protein were sequenced by mass spectrometry and which residues were modified. .

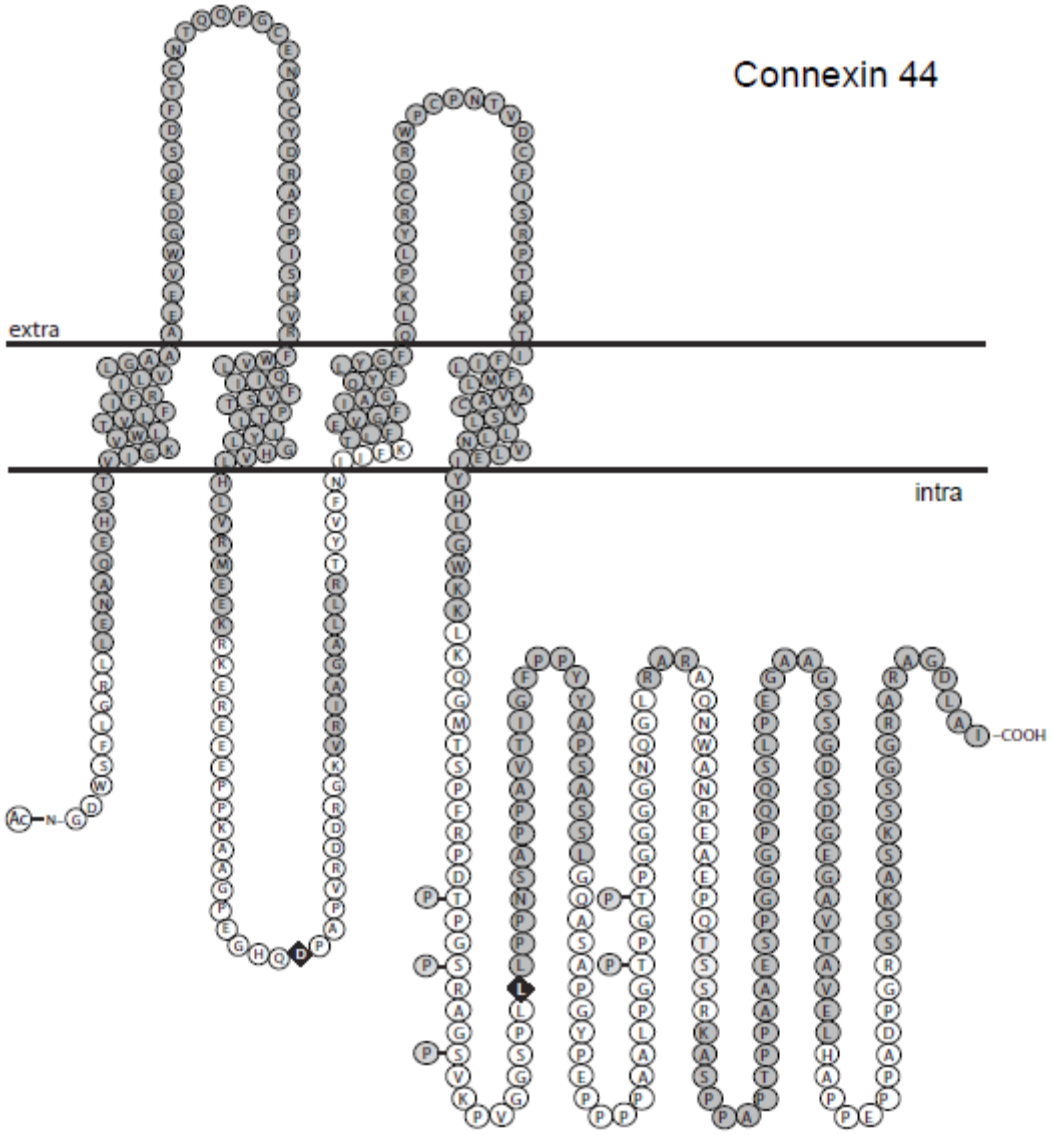


Figure 22. Diagram of connexin 44 showing the post-translational modifications identified in this study. Grey residues indicate portions of the protein that were not identified in the mass spec analysis. P indicates phosphorylated residues, Ac indicates acetylated residues. The D and L residues in the solid diamonds were from truncated peptides and could represent *in vivo* cleavage sites.

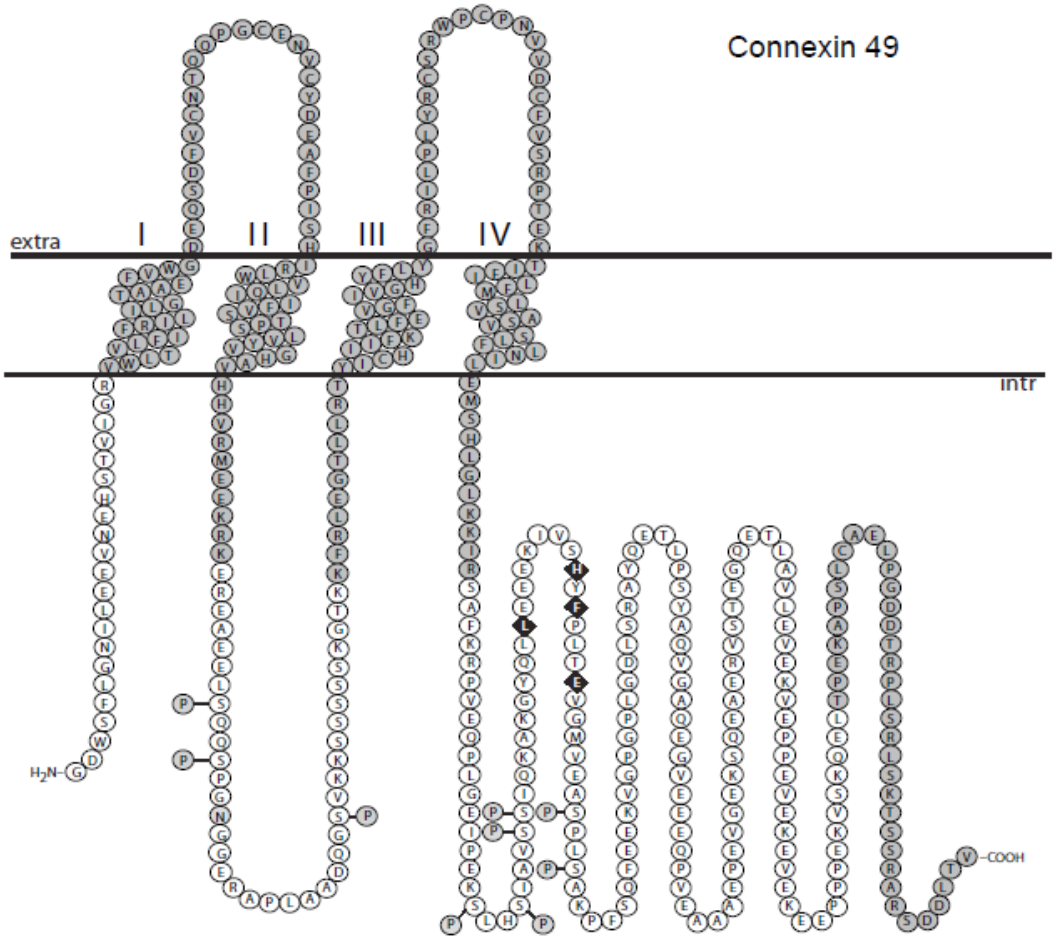


Figure 23 Diagram of connexin 49 showing coverage and PTMs identified. Gray residues were not detected and the black diamonds were from truncated peptides. In addition, there was a deamidation of the asparagine residue in the cytoplasmic loop.

Potential Kinases

To determine which kinases may be involved in the phosphorylation of Cx44 and Cx49, the obtained sequences were submitted to the NetPhosK 1.0 web server (Blom et al., 1999) (Tables 9 and 10). Several of the phosphorylation sites identified from Cx49 are consistent with sites phosphorylated by PKC. Potential Map kinase (MapK) sites were also identified in both connexins along with a Casein Kinase I (CKI) site. Both

MapK and CKI are known to phosphorylate connexins (Lampe & Lau, 2004b). Several sites in Cx44 are potential Glycogen Synthetase Kinase 3 sites, an enzyme which has not been shown to phosphorylate connexins previously.

Table 9 Predicted kinases for sites detected in Cx44.

Residue	Putative kinase
Thr 238	GSK3, P38MAPK
Ser241	GSK3
Ser242	PKA, PKC
Thr 300	GSK3, P38MAPK
Thr303	GSK, P30MAPK, CK1
Ser (330,331)	PKC

Table 10 Predicted kinases for sites detected in Cx49

Residue	Putative kinase
Ser 115	CK1, ATM, DNAPK
Ser 118	CK1
Ser 134	PKC

Mouse lens proteins

Based on the previous success of identifying post-translational modifications of the connexins from the small amounts of bovine lens material, a similar approach with mouse lens was attempted. The first experiments with mouse lens were performed with the lens capsule attached, as the mouse lenses tended to slough off a considerable portion of the outer lens layers when the capsule was removed. The initial analyses of mouse

lens proteins yielded considerably fewer protein peaks on the MS spectra compared to bovine lens, and the two corresponding mouse connexin proteins, Cx46 and Cx50, were particularly difficult to detect compared to their bovine counterparts. Ions from heparin sulfate proteoglycan protein (HPSG) were extremely prominent in the mass chromatogram. Laeverin was also an abundant protein within the mouse lens. Sequence coverage of Cx46 and Cx50 was not as good as in the cow samples, accounting for only 20 and 22 % of the protein sequences, respectively. Despite the low yields of the connexins, a phosphorylated site was detected from both connexins, Ser249 in Cx46 and Ser134 in Cx50. There was no N-terminal acetylation detected on either connexin in the mouse lens. Table 11 lists peptide identifications derived from Cx46 that were first identified by mass and retention time values and then confirmed by tandem MS analysis.

Table 11 Identification of peptides derived from mouse Cx46

Meas Mr	Intens	Frac	Exp Mr	error	Frac Error	Sequence	
936.452	19400	59	936.445	7	2.1	GDWSFLGR	Y
943.502	2651	34	943.509	-6	0.7	EEELLRR	Y
982.495	1864	24	982.483	12	-1	DPPLRDDR	Y
1167.596	1119	24	1167.599	-2	1.3	DPPLRDDRGK	Y
1641.769	588	36	1641.749	11	1	QGVTNHFNPDASEAR	Y
1642.731	435	37	1642.733	-1	1.5	QGVTNHFNPDASEAR (+0.98)	Y
1722.72	1289	37	1722.	3	1.5	QGVTNHFNPDASEAR (+80)	Y
1865.801	529	37	1865.786	7	0.5	HCNGHHLTTEQNWTR +0.98 3+2	N
2181.022	268	31	2181.03	-3	3.7	QVAEQQTPASKPSSAASSPDGR	Y
2309.098	355	30	2309.125	-11	4.3	QVAEQQTPASKPSSAASSPDGRK (-17)	N

2326.134	928	27	2326.151	-7	2.8	QVAEQQTPASKPSSAASSPDGRK	N
2398.031	26864	61	2398.05	-8	0.4	CDRWPCPNTVDCFISRPTEK +32	N
2398.073	1628	65	2398.05	9	2.4	CDRWPCPNTVDCFISRPTEK +32	N
2514.284	1825	79	2514.25	13	1.4	GDWSFLGRLLENAQEHSTVIGK +16 +42	N
2514.284	1825	79	2514.275	3	2.9	AIGFPGAPLSPADFTVVTLNDAQGR	Y
2593.259	970	78	2593.258	0	2.4	AIGFPGAPLSPADFTVVTLNDAQGR +80	N

Several peptides from this analysis were mis- identified by sMART. As was observed for the bovine lens proteins, mis-identifications are particularly common for peptides that arise from several modifications. While MS/MS cannot always identify a peptide with certainty, it can often provide possible matches, which can then be reasonably interpreted by a researcher who is familiar with how multiple modifications can affect the masses of peptides.

Table 12 lists the peptides derived from Cx50 that were identified from the same 100 chromatographic fractions as the Cx46 analysis. Once again, a number of peptides are misidentified.

.

Table 12. Identification of peptides derived from mouse Cx50

Measured Mass	Intens	Frac	Exp Mr	error	Frac Error	Sequence	
707.317	3906	12	707.327	-14	-1.3	MEEKR [M+16]	Y
1219.507	2620	25	1219.525	-14	0.6	VTKQGLSAEK [T+80 S+80]	N
1107.535	43789	45	1107.545	-8	2	GYQLLEEEK	Y

1196.528	1775	48	1196.526	1	2.5	IGTGPLADMSR [T+80]	N
1165.652	1975	40	1165.645	5	-1.7	VPIAPDQASIR	Y
1253.578	1361	36	1253.59	-9	3	EELQAEKVTK [T+80]	N
1253.606	11645	33	1253.59	12	1.5	EELQAEKVTK [T+80]	N
1268.715	1640	46	1268.709	4	-1.9	SLHSIAVSSIQK	Y
1428.644	1594	49	1428.641	1	-0.4	SLHSIAVSSIQK [S+80 S+80]	N
1293.735	1413	36	1293.74	-4	-2.8	VPIAPDQASIRK	y
1464.791	59366	45	1464.793	-1	-2.9	RPVEQPLGEIAEK	Y
1751.875	1100	35	1751.88	-2	-0.9	VAPEGQETVAVPDRER	Y
1766.893	16920	44	1766.891	1	-0.7	SNGGERVPIAPDQASIR [N+0.98]	Y
1845.88	537	40	1845.873	3	-2.7	SNGGERVPIAPDQASIR [S+80]	Y
1846.857	9348	42	1846.857	0	-1.7	SNGGERVPIAPDQASIR [S+80 N+0.98]	Y
1967.928	433	38	1967.945	-8	-2.5	EELQAEKVTKQGLSAEK [T+80 Q+0.98]	N
1894.986	14467	41	1894.986	0	-1.3	SNGGERVPIAPDQASIRK [N+0.98]	Y
1973.982	880	39	1973.968	6	-2.3	SNGGERVPIAPDQASIRK [S+80]	Y
1974.957	76935	40	1974.952	2	-1.8	SNGGERVPIAPDQASIRK [S+80 N+0.98]	Y
2840.249	668	61	2840.245	1	2	QGLSAEKAPSLCPELTTDDNRPLSR	N

A number of tandem MS spectra were submitted to the GPM website and Table 13 lists the proteins identified from the sample.

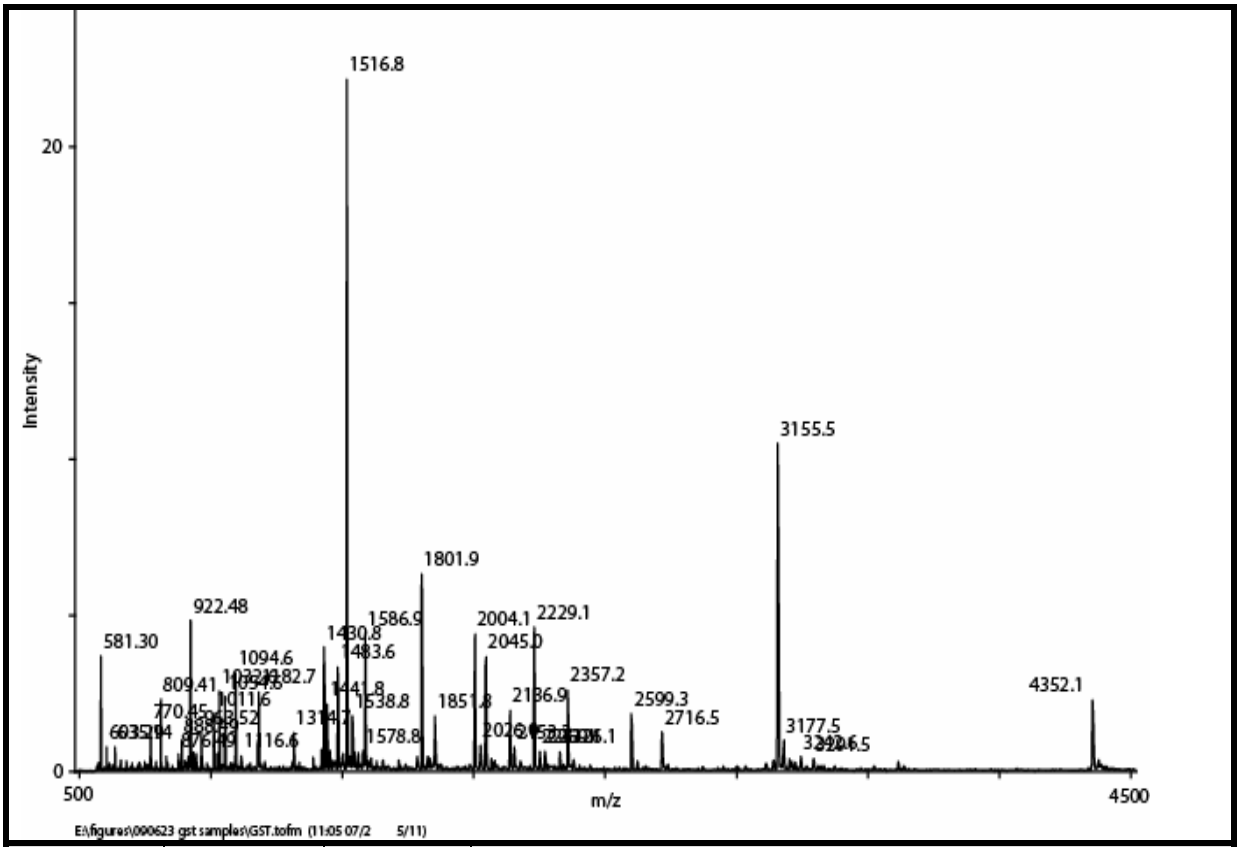
Table 13 Proteins identified from mouse lens based on tandem MS and X!Tandem

log(e)	unique	total	Mr	Accession	Description
-192.6	21	24	468.7	ENSMUSP00000030547	Perlecan
-108.8	10	12	65.8	sp|K2C1_HUMAN|	Cytokeratin-1
-101.4	12	13	24.3	ENSMUSP00000113572	Crystallin beta B3

-72.3	5	6	65.8	sp|K22E_HUMAN|	Cytokeratin-2e
-43.1	6	6	53.7	ENSMUSP00000028062	Vimentin
-41.4	6	9	21.2	ENSMUSP00000084617	Crystallin gamma E
-38.2	5	7	160.6	ENSMUSP00000033898	Collagen type IV alpha1
-37.1	5	9	28.2	ENSMUSP00000026455	Aquaporin 0
-35.6	4	4	25.2	ENSMUSP00000077693	Crystallin Beta A1
-32.6	4	5	19.8	ENSMUSP00000110146	Crystallin alpha A
-32.3	4	4	62.1	sp|K1C9_HUMAN|	Cytokeratin 9
-28.2	1	4	21.1	ENSMUSP00000058548	Crystallin gamma A
-23.9	3	3	20.1	ENSMUSP00000091371	Crystallin alpha B
-22.1	3	5	167.2	ENSMUSP00000033899	Collagen type IV alpha2
-19.5	3	4	28	ENSMUSP00000031286	Crystallin beta B1
-18.2	3	3	20.9	ENSMUSP00000027089	Crystallin gamma C
-17.7	2	2	21.1	ENSMUSP00000027090	Crystallin gamma B
-16.9	3	3	59.5	sp|K1C10_HUMAN|	Cytokeratin 10
-12.1	3	3	49.6	ENSMUSP00000049532	Connexin 50
-9.4	2	2	20.8	ENSMUSP00000043588	Crystallin gamma S
-8.2	1	2	23.4	ENSMUSP00000031295	Crystallin beta B2
-6.4	2	2	99.7	ENSMUSP00000025166	Cadherin 2
-6.2	2	2	45.7	ENSMUSP00000116249	Phakinin
-5.5	2	2	54.6	ENSMUSP00000080613	Keratin 83
-4.3	1	1	44.5	ENSMUSP00000056622	Keratin 34
-4.1	1	2	46.3	ENSMUSP00000059587	Connexin 46
-3	1	2	73	ENSMUSP00000028907	Filensin
-2.9	1	1	28.3	ENSMUSP00000048739	Aquaporin 5
-2.9	1	1	92	ENSMUSP00000081048	Adam9
-2.6	1	1	6.7	ENSMUSP00000038352	ribosomal protein S29
-1.9	1	1	46.6	ENSMUSP00000029866	cyclin E2
-1.2	1	1	59.1	ENSMUSP00000046877	TBC1
-1.2	1	1	6.9	ENSMUSP00000052968	Lens epithelial protein

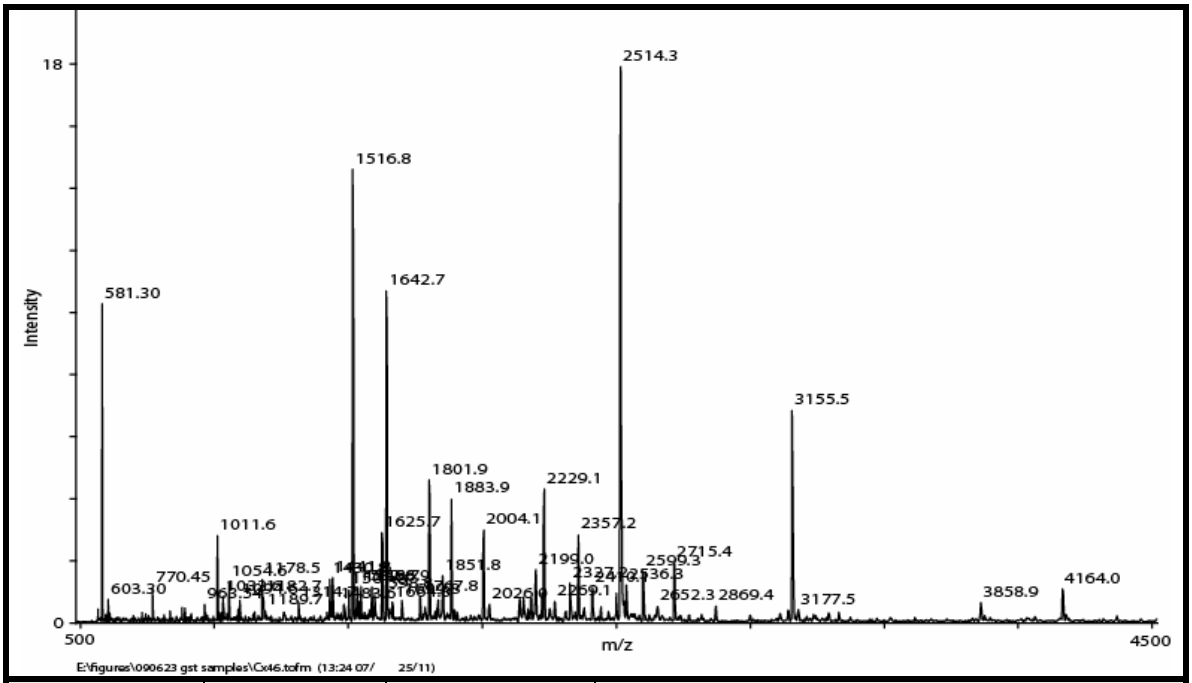
GST- connexin fusion proteins and pull-down experiments

Figures 24, 25, and 26 show mass spectra and expected masses from GST-connexin fusion proteins expressed in *E. coli* cells. Interestingly, even after overnight digests of the cell lysates, there were still a number of missed predicted cleavages. The C-terminus of both Cx46 and Cx50 is thought to be important in interactions with Zona Occludens-1 and also subject to phosphorylation. In this study, it was not possible to get coverage of that region of the connexin proteins derived from the cow lens. However, by over-expressing the mouse Cx50 fusion proteins within *E. coli* cells, a spectral peak corresponding to the peptide ARSDDLTI, matched the extreme C-terminal peptide of Cx50. The intensity of this peak was very low compared with other peptides from the protein, and even after HPLC purification to resolve the peptides, the intensity of the peak corresponding to that region of the protein was very low.



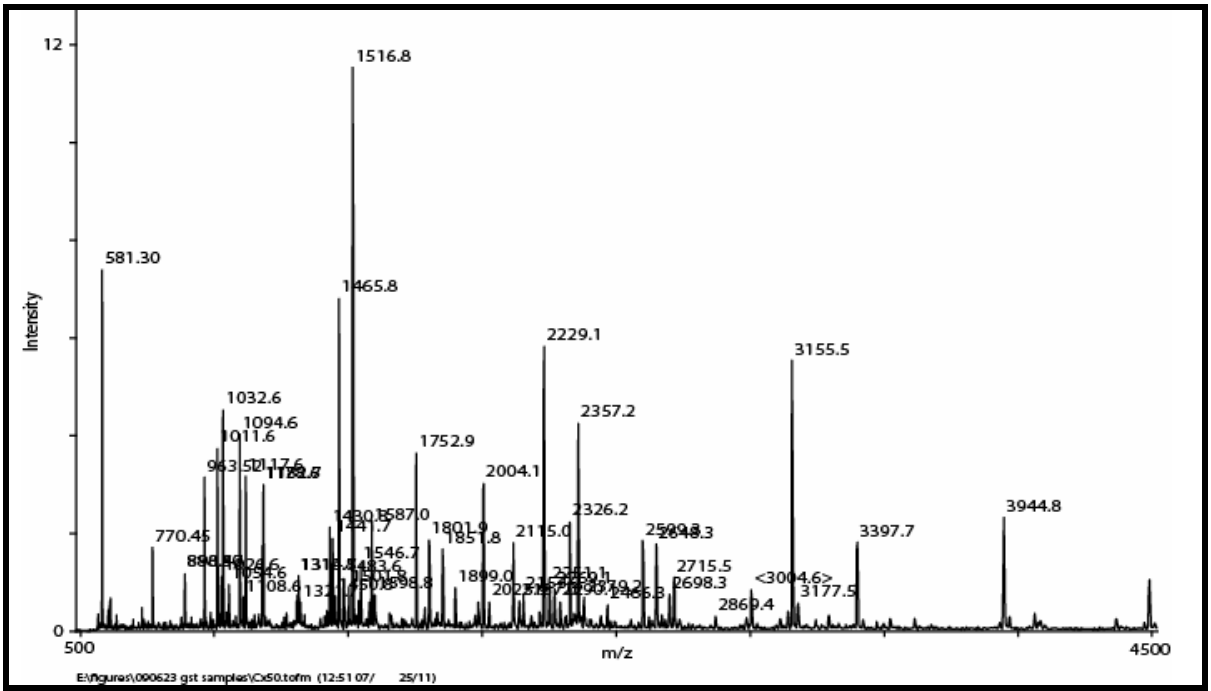
Location	Expected	Measured	Sequence
104 - 108	581.3047	581.303	YGVSR
109 - 113	581.3299	581.303	IAYSK
41 - 44	603.3367	603.292	WRNK
12 - 18	770.4524	770.453	GLVQPTR
10 - 18	1011.6315	1011.629	IKGLVQPTR
65 - 73	1032.5876	1032.593	LTQSMAIIR
1 - 9	1094.5708	1094.573	MSPILGYWK
182 - 191	1182.6846	1182.684	RIEAIPQIDK
109 - 119	1314.6945	1314.681	IAYSKDFETLK
183 - 194	1430.8259	1430.823	IEAIPQIDKYLK
114 - 125	1441.7578	1441.769	DFETLKVDFLSK
90 - 103	1516.8045	1516.808	AEISMLEGAVLDIR
88 - 103	1801.9482	1801.944	ERAEISMLEGAVLDIR
219 - 237	2045.0344	2045.048	SDLEVLFQGPLGSPEFPGR
46 - 64	2229.1119	2229.109	FELGLEFPNLPYYIDGDVK
19 - 35	2269.1392	2269.145	LLLEYLEEKYEEHLYER
198 - 218	2326.1409	2326.141	YIAWPLQGWQATFGGGDHPPK
45 - 64	2357.2069	2357.189	KFELGLEFPNLPYYIDGDVK
19 - 42	3155.533	3155.529	LLLEYLEEKYEEHLYERDEGDKWR
46 - 73	3242.6811	3242.619	FELGLEFPNLPYYIDGDVKLTQSMAIIR
198 - 237	4352.1569	4352.13	YIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSPEFPGR

Figure 24. Mass fingerprint of GST protein expressed in *E. coli* cells and purified on a glutathione spin column. The table shows a number of peaks that correspond to the GST protein sequence.



Location	Expected	Measured	sequence
109 - 113	581.3299	581.3	IAYSK
41 - 44	603.3367	603.305	WRNK
12 - 18	770.4524	770.451	GLVQPTR
10 - 18	1011.632	1011.629	IKGLVQPTR
65 - 73	1032.588	1032.593	LTQSMAIIR
1 - 9	1094.571	1094.572	MSPILGYWK
132 - 140	1178.545	1178.543	MFEDRLCHK
182 - 191	1182.685	1182.683	RIEAIPQIDK
109 - 119	1314.695	1314.683	IAYSKDFETLK
183 - 194	1430.826	1430.825	IEAIPQIDKYLK
114 - 125	1441.758	1441.768	DFETLKVDFLSK
90 - 103	1516.805	1516.808	AEISMLEGAVLDIR
74 - 87	1546.751	1546.744	YIADKHNMLGGCPK
182 - 194	1586.927	1586.929	RIEAIPQIDKYLK
244 - 258	1642.757	1642.747	QGVTNHFNPDASEAR
109 - 125	2004.069	2004.063	IAYSKDFETLKVDFLSK
341 - 362	2199.064	2199.045	QVAEQQTPASKPSSAASSPDGR
46 - 64	2229.112	2229.11	FELGLEFPNLPYYIDGDVK
19 - 35	2269.139	2269.147	LLLEYLEEKYEEHLYER
341 - 363	2327.159	2327.15	QVAEQQTPASKPSSAASSPDGRK
45 - 64	2357.207	2357.188	KFELGLEFPNLPYYIDGDVK
296 - 320	2514.299	2514.276	AIGFPGAPLSPADFTVVTLNDAQGR
43 - 64	2599.345	2599.317	NKKFELGLEFPNLPYYIDGDVK
19 - 42	3155.533	3155.529	LLLEYLEEKYEEHLYERDEGDKWR

Figure 25. Mass fingerprint of GST-Cx44ct fusion protein expressed in *E. coli*. The table shows a number of peaks that correspond to peptides derived from the expected GST-Cx44ct sequence



Location	Expected	Measured	Sequence
104 - 108	581.3047	581.301	YGVSR
12 - 18	770.4524	770.451	GLVQPTR
438 - 445	890.4583	890.471	ARSDDLTI
10 - 18	1011.632	1011.629	IKGLVQPTR
183 - 191	1026.584	1026.586	IEAIPQIDK
65 - 73	1032.588	1032.592	LTQSMAIIR
1 - 9	1094.571	1094.574	MSPILGYWK
277 - 285	1108.553	1108.56	GYQLLEEEK
316 - 326	1117.568	1117.57	IGTGPLADMSR
132 - 140	1178.545	1178.563	MFEDRLCHK
182 - 191	1182.685	1182.686	RIEAIPQIDK
181 - 191	1310.78	1310.776	KRIEAIPQIDK
109 - 119	1314.695	1314.703	IAYSKDFETLK
183 - 194	1430.826	1430.824	IEAIPQIDKYLK
114 - 125	1441.758	1441.745	DFETLKVDFLSK
250 - 262	1465.801	1465.8	RPVEQPLGEIAEK
90 - 103	1516.805	1516.808	AEISMLEGAVLDIR
74 - 87	1546.751	1546.742	YIADKHNMLGGCPK
182 - 194	1586.927	1586.957	RIEAIPQIDKYLK
368 - 383	1752.888	1752.87	VAPEGQETVAVPDRER
88 - 103	1801.948	1801.947	ERAEISMLEGAVLDIR
246 - 262	1899.034	1899.044	SAFKRPVEQPLGEIAEK
413 - 430	1984.976	1984.968	APSLCPELTTDDNRPLSR
109 - 125	2004.069	2004.063	IAYSKDFETLKVDFLSK
406 - 430	2698.347	2698.344	QGLSAEKAPSLCPELTTDDNRPLSR
19 - 42	3155.533	3155.529	LLLEYLEEKYEEHLYERDEGDKWR
286 - 315	3397.703	3397.711	IVSHYFPLTEVGMVETSPLSAKPFSQFEEK

327 - 361	3944.893	3944.838	SYQETLPSYAQVGVQEVEREEPPIEEAVEPEVGEK

Figure 26 Mass fingerprint of GST-Cx50ct fusion protein expressed in *E. coli* cells.

The GST-connexin fusion proteins were used in pull-down experiments. Both the negative (GST alone) control and the two fusion proteins, GST-Cx46ct and GST-Cx50ct, had a number of non-specific binding proteins that were identified by the mass spectrometry analysis, many of which were members of the crystallin family. The GST-Cx50ct fusion protein did pull out some unique proteins, but only two proteins were reproducibly present, actin and galectin-related interfiber protein (Table 15).

In addition, in some of the pull-down experiments, the native proteases present in the mouse lens were activated and some naturally-cleaved fragments of both connexins were identified. The expected calpain-cleaved fragment from Cx50 with a mass of 1204.6 (IVSHYFPLTE) was present as well as two peptides from Cx46. Both a peptide with mass 2117.1 (GLIDSSGSSLQESALVVTPEE) and a shorter peptide 1191.5 (GLIDSSGSSLQE) were present in the sample. These are similar to the Cx50 site and could possibly be due to cleavage by the same calpain.

Table 14 Proteins identified by pulldown with GST alone

Peptides[a]	Coverage[b]	Accession[d]	Description
8	27	P24622	Crystallin Alpha A
5	44	Q9JJV1	Crystallin Beta A2
4	31	Q9JJV0	Crystallin Beta A4
5	30	Q9WVJ5	Crystallin Beta B1

9	43	P62696	Crystallin Beta B2
13	68	Q9JJU9	Crystallin Beta B3
8	33	P10065	Crystallin Gamma A
6	28	P04344	Crystallin Gamma B
8	35	Q61597	Crystallin Gamma C

a - Number of peptides matched from sequence
b- sequence coverage of protein
c- Swiss prot ID

Table 15 Proteins identified in GST-Cx46ct pulldown

Peptides[a]	Coverage[b]	Accession[d]	Description
17	81	P24622	Crystallin Alpha A
22	75	P23927	Crystallin Alpha B
5	28	Q9JJV1	Crystallin Beta A2
4	31	Q9JJV0	Crystallin Beta A4
15	68	Q9WVJ5	Crystallin Beta B1
8	25	P62696	Crystallin Beta B2
9	34	Q9JJU9	Crystallin Beta B3
15	55	P10065	Crystallin Gamma A
19	59	P04344	Crystallin Gamma B
13	56	Q61597	Crystallin Gamma C
11	40	Q03740	Crystallin Gamma E
8	30	Q9cxv3	Crystallin Gamma F
12	46	O35486	Crystallin Gamma S

a - Number of peptides matched from sequence
b- sequence coverage of protein
c- Swiss prot ID

Table 16 Proteins identified in GST-Cx50ct pulldown

Peptides[a]	Coverage[b]	Accession[d]	Description
25	76	P24622	Crystallin Alpha A
20	67	P23927	Crystallin Alpha B
16	60	P02525	Crystallin Beta A1
12	44	Q9JJV1	Crystallin Beta A2
12	53	Q9JJV0	Crystallin Beta A4
21	88	Q9WVJ5	Crystallin Beta B1
25	88	P62696	Crystallin Beta B2
17	68	Q9JJU9	Crystallin Beta B3
21	80	P04344	Crystallin Gamma B
13	56	Q61597	Crystallin Gamma C
14	54	Q03740	Crystallin Gamma E
18	76	O35486	Crystallin Gamma S
10	41	P63260	Actin cytoplasmic 2
5	27	Q9D1U0	Galectin-related interfiber protein

a - Number of peptides matched from sequence
b- sequence coverage of protein
c- Swiss prot ID

Discussion

The initial aim of this project was to study lens fibers connexins using both mass spectrometry and a variety of cell and molecular techniques. At the time the project started, very few studies of connexins with mass spectrometry had been done, and hence, it was anticipated that this approach to the study of these proteins would, at best, serve only as a small component of the research, complementing the primary focus using more traditional cell biological techniques. However, given the time required to develop the mass spectrometry techniques, the emphasis gradually shifted more to an exploration and testing of how mass spectrometry could be applied to address questions of connexin protein identification and functional regulation.

Mass spectrometry is a powerful analytical tool that has revolutionized the field of proteomics, mainly by allowing the rapid identification of proteins and the detection of covalent modifications (Patterson & Aebersold, 2003). The study of membrane-bound proteins however, has always presented unique challenges, as it is often difficult to study their true structure and function using many of the conventional *in vitro*, traditional biochemical methods, as most of these methods strip the protein from its natural lipid-based microenvironment. Mass spectrometry analysis of membrane-bound proteins typically faces similar difficulties of resolving structure-function relationships as the detergents used to solubilize membrane proteins are not compatible with mass spectrometry.

In this study on bovine lens proteins, mass spectrometry techniques proved instrumental in identifying several post-translationally modified sites in the cytoplasmic domains of bovine Cx44 and Cx49. The results show what was identified in a few individual runs. Overall a total of 350 peptides from 110 other co-purified proteins were also identified, some of which were also found to be post- translationally modified. By far the most commonly detected peptides from the MS analysis were derived from crystallins, for instance alpha crystallin and beta b1 crystalin were present in all the gel slices and the intensity of their MS peaks were quite strong in the in solution digest, but peptides from a few other membrane proteins were also identified.

In a parallel study using mouse lens, the MS analysis once again identified an abundance of crystallin proteins, but identified far fewer other peptides relative to the bovine studies. Only a few of the non-crystallin peptides from the mouse lens protein extractions were identified as being derived from Cx46 and Cx50. To identify proteins that interacted with these two mouse connexin proteins, pull-down experiments using the C-terminal tails of Cx46 and Cx50 were performed. Unfortunately, due to the high abundance of crystallins in the mouse lens, crystallins were the most abundant peptides detected in both the control and experimental fractions of the pull-down experiments. Nevertheless, a few unique proteins were identified that appear to interact with Cx50, including actin and galectin-related inter fiber protein, and their relevance to connexins will be considered in a following section.

Aquaporin 0

Aquaporin 0 is the most abundant protein in the lens fiber membranes. It has been studied extensively with mass spectrometry in Kevin Schey's lab (Swamy-Mruthinti & Schey, 1997b; Schey et al., 1997a; Han & Schey, 2004a) and with classical techniqes by others (Lampe & Johnson, 1989; Lampe & Johnson, 1990). Lampe and Johnson (1990) identified serine 243 and 245 as being phosphorylated. Schey et al. 1997 identified serine 235 as the major site of phosphorylation. My data indicate that both of the sites can be phosphorylated, and show no reason to claim one site is the major site of phosphorylation.

Post Translational Modifications of Bovine Connexins

Most connexins that have been examined have been shown to be phosphorylated (Lampe & Lau, 2004a), but with the exception of mammalian Cx43, only a few specific phosphorylation sites have been identified. More recently, however, mass spectrometry has been used to examine modifications in rat Cx26 and Cx32 (Locke et al., 2006a). In their study, the authors used isoelectric focusing to resolve the proteins from rat liver and then MALDI mass spectrometry to identify post- translational modifications of the cytoplasmic domains. They identified hydroxylations and carboxylations as well as the phosphorylation of the N-terminal domains and the phosphorylation and palmitoylation of the C-terminal domain of Cx32. While mass spectrometry is not as sensitive as immunoassays, it offers the ability to acquire more information about phosphorylations over larger areas of the protein.

Phosphorylation of connexins has mainly been studied in Cx43 (Lampe and Lau, 2004a). The bovine Cx44 and Cx49 that were examined in this study are both members

of the alpha family of connexins, which includes Cx43, and so there may be similar residues that are phosphorylated in these related connexins. When the three sequences are aligned, there is very little similarity observed in most of the cytoplasmic loops and carboxy terminus domains, except for the last 20 residues of the carboxy tail. In this portion of Cx43, there are several serine residues that are phosphorylated in a PKA - dependent manner, and one residue, Ser368, which is phosphorylated by PKC and is thought to be the crucial residue for the effects seen in gap junction communication when PKC is activated (Lampe et al., 2000). Unfortunately, it was not possible to obtain MS coverage of this portion of either Cx44 or Cx49. The *in silico* digests of both Cx44 and Cx49 show that the digest would have produced small peptides within this region that would be difficult to identify, and for that reason, digestion with chymotrypsin was used, which should have produced peptides in the mass range suitable enough for detection by MS/MS.

Three serine residues in Cx43 are phosphorylated by MAPK (Warn-Cramer et al., 1996). Cx44 has three threonine residues that Netphos predicted to be MAPK sites. Each of these sites show a PXTP/SP motif that is characteristic of an optimal MAPK phosphorylation site (Gonzalez et al., 1991). Cx49 has only one site that is predicted to be a MAPK target site, but its has a shorter SP motif, which is considered sufficient but not optimal for phosphorylation by MAPK. There were no clear examples of any of the MAPK consensus docking sites (Caldwell et al., 2006).

Cx44 cleavage sites in the lens.

It is impossible to say with certainty that the cleavage sites predicted in Cx44 are the actual sites of cleavage in a normal lens. It has been shown that Cx44 is truncated at the C-terminal as the lens cells mature (Jacobs et al., 2004a). The peptide with a truncation at Leu255 was not detected in samples derived from the outer regions of the lens, but was only detected in samples that contained the core of the lens. If this truncation was an artifact of the sample processing, it would be present in all samples, which suggests that it may be a genuine cleavage site. Conversely, the peptide with the cleavage at Asp131 was seen in samples derived from all regions of the lens. Because of the way the lens develops, the proteins present in the nucleus and inner cortex show no turnover. Hence, the proteins present in the membranes from those regions are old, and in some cases, as old as the organism itself. If the cleavage was a normal developmental event, one would expect the cleavage products to be more abundant in the older regions of the lens. But based on the intensity of the peaks, the Asp131 cleavage product does not appear to be any more abundant in a particular region of the lens. Interestingly, in the chick, it was recently shown that Cx45.6 is cleaved in the cytoplasmic loop as the lens fibers differentiate (Yu et al., 2005); this cleavage interferes with an interaction between aquaporin 0 and the cytoplasmic loop domain of Cx45.6.

Acetylation of the N-terminus.

The acetylation of the N-terminus of some, but not all Cx44 and Cx50 proteins is one of the most interesting results observed in the cow lens study. There are other proteins from the lens that were acetylated on the N-terminus, but acetylation is usually an all-or-nothing phenomena. For instance, the N-terminus of alpha A crystallin is

acetylated, but the non-acetylated N terminus peptide has never been observed; this is also true for several beta crystallins and alpha B crystallin as well (Strous et al., 1974; Driessen et al., 1983). N-terminal acetylation occurs co-translationally and the enzyme complex is physically associated with the ribosome (Strous et al., 1974; Pestana & Pitot, 1975; Gautschi et al., 2003).

It has been over 50 years since the first discovery of an N-terminal acetylated protein (Narita, 1958), but the function of the acetylation has been difficult to pin down. Only for a handful of proteins has a specific function been determined. Actin and tropomyosin require N-terminal acetylation for proper binding with each other (Polevoda et al., 2003) while a couple of GTPases require it for targeting to the Golgi apparatus (Behnia et al., 2004; Setty et al., 2004). Conversely, N-terminal acetylation also has been shown to inhibit targeting to the endoplasmic reticulum of a number of secreted proteins (Forte et al., 2011). Another possible function of N-terminal acetylation is the prevention of protein ubiquitination and subsequent degradation (Hershko et al., 1984), but there are cases where the opposite is true and acetylation seems to increase the rate of protein degradation (Hwang et al., 2010).

The first 10 residues of connexins lie in the channel pore and create a vestibule for the connexon channel and act as a voltage sensor (Purnick et al., 2000). Acetylation of the N-terminus eliminates the charged amino group and could possibly have effects on the gating of a connexon channel. In Cx43, hyper-acetylation of lysines causes a disassociation from gap junction plaques and changes the localization of the connexin protein (Colussi et al., 2011). The histone deacetylase P300/CBP associated factor co-immunoprecipitates with Cx43, and other histone deacetylases co-localize with Cx43 as

well (Colussi *et al.*, 2011), which suggests that the acetylation/deacetylation of Cx43 is regulated. It is possible then that the acetylation of the N-terminus has a functional role in the connexins. Despite all these observations, further research will be required to confirm that acetylation is important in the functioning of the lens fiber connexins.

There are a few possibilities for how the connexins are acetylated. A subset of the connexins could be acetylated co-translationally as they are being synthesized at the ribosome. Alternatively, all of the connexins could be acetylated as they are being synthesized, and then later in their life cycle, the acetyl group could be removed, or the acetyl group could be attached in other stages of the connexin cycle (e.g. once the protein reaches the plasma membrane). One way this could be determined is to produce antibodies to the acetylated and non-acetylated versions of the N-terminal regions and determine where in the lens fibers the acetylated connexins are present.

Another way to investigate the functional relevance of acetylation would be to introduce mutant versions of the connexin into the mouse. The normal N-terminal acetylation depends on the protein sequence and altering the second residue in the sequence to proline could inhibit the acetylation reaction. This approach would work only if the mouse actually produces the acetylated version of the protein. Based on the ambiguous spectra obtained in this study, it is still unclear whether the mouse lens fiber connexins are acetylated.

Mouse Connexins

The study of connexins in the cow lens identified a number of sites with post-translational modifications, but the cow is not a very tractable animal for additional animal modification experiments, as they are very costly, and there are few genetic tools

available for use in cows. While some bovine connexin sites were found to be phosphorylated, it is only possible to speculate about their functional relevance. In contrast, the mouse model is considerably more tractable for further experiments that could uncover some of the functional roles of the post translational modifications of various connexins. Given that it is possible to produce transgenic mice with targeted mutations (Thomas & Capecchi, 1987), site directed mutagenesis could be used to convert a connexin phosphorylation site to either alanine as a loss-of-function experiment or to a glutamate which somewhat mimics a phosphoserine or threonine as a gain-of-function experiment (Maciejewski et al., 1995). The modified genes could be introduced into the mouse genome to replace the wild type genes to examine the effect *in vivo*. This could be done either by introducing the mutated gene into the genome (a knock-in), or given that the Cx50 knock-out mouse has already been generated (White et al., 1998), the gene could be re-introduced using a plasmid expression system. These types of experiments are laborious and require considerable genetic and molecular analyses, and are not worth attempting in cows, which is the main reason for working with a mouse model.

However, the qualities of the mouse lens are quite different from those of the cow lens. Mouse lens is much harder than that of cow because mice do not change the lens shape in the process known as accommodation. In addition, mouse lenses are much smaller than those of cows and therefore they require much less conductance both for the transport of ions and other metabolites, per surface area. Because volume increases to the third power while the surface area only increases by the second power as a cell or tissue grows, the number of gap junctions in a given area of the lens at the outer region of a

mouse lens can be much less than a cow would require. Consequently, the function and regulation of connexins in the two species could differ considerably.

That said, there are some obvious similarities in the processing of the proteins in the organisms. The mouse Cx50 is cleaved at the same site in the C-terminus as the cow connexin homolog. Phosphorylation was detected in the cytoplasmic loop section as well as the C-terminal domain, however the sequence coverage and strength of the detected signals was much lower than that observed with the cow peptides. Cx 46 also had one phosphorylated peptide, but the overall sequence coverage was even lower than Cx50, so one cannot say if the multiple phosphorylations seen in the cow are present in mice. An interesting experiment would be to modify the connexin cleavage site so that it is not recognized by the calpain that cuts it. If it had the same conduction properties as wild type Cx50, it could be used as a knock-in to see what, if any, function the cleavage has in the mouse lens development.

Protein Complexes and Pull-downs

Many processes in the cell are mediated by protein-protein interactions. Some structures such as the nuclear pore are fairly stable structures, formed by many different proteins. Other interactions can be much more transient, such as the interaction between a protein kinase and its substrate protein. In the gap junction field, there has been much speculation about the formation of a supermolecular structure beyond the connexon (Herve et al., 2007; Duffy et al., 2002) and there is some evidence that structural proteins such as actin and tubulin and the scaffold protein zona occludens 1 have an affinity for Cx43 (Giepmans et al., 2001) as well as both Cx44 and 50 (Nielsen et al., 2003).

Techniques such as the yeast two hybrid system as well as immuno-coprecipitation can also be used to identify interacting partners. Affinity pull-down methods using GST fusion proteins have been used to identify proteins that interact with Cx43 (Singh & Lampe, 2003) and were used in this study. Yeast two hybrid techniques are very sensitive and can detect very transient, low affinity interactions, but they also can produce many false positives (Vidal & Legrain, 1999). With GST pull-down methods, the bait protein is typically used at a much higher concentration than is normally found within the cell, and can therefore falsely identify low affinity interactions that do not normally occur in the cell. The immuno-coprecipitation technique can use the naturally-occurring proteins that are present in the cell, and thereby avoid the problem of adding too much of a particular interacting protein, but in practice, the components that are being precipitated are often over-expressed in the cell to facilitate the ease of detection. Clearly then, any detected interaction has to be confirmed in the cell by other techniques such as fluorescent immunohistochemistry, to show that the locations of the two proteins coincide within the cell.

In the pull-downs performed here, many of the crystallins appeared in both the experimental and control samples, and likely obscured the less abundant proteins. There were several MS peaks that were unique to the experimental samples, but subsequent tandem MS often showed that they were from crystallins that had not been fully cleaved by the proteases, due to their high abundance. There is however, some evidence for specific crystallins having affinities for certain proteins, such as gamma e and gamma f crystallins, which interact with aquaporin 0 (Fan et al., 2005). Crystallins also interact with each other to form complexes (Srivastava et al., 2008) so it is also possible that a

binding interaction between a specific crystallin and the Cx50 could result in the pull down of several other crystallins. It is also possible that specific crystallins are indeed interacting with connexins and are therefore present in the experimental sample in greater amounts, but the techniques used in this study were not capable of accurately quantifying the interactions. Ideally, quantitative MS techniques could be used to compare protein affinities in different samples, but quantitative MS methods are still in the developmental stages and were beyond the scope of this current study.

One interesting protein that was found to bind to Cx50 was galectin-related interfiber protein. This protein is similar to galectin binding protein and is expressed in the lens. It is found localized in the regions of the cell membrane on the short sides of lens fibers (Ogden et al., 1998). Cx50 gap junctions are generally thought to be located in the broad sides of the lens fibers (Kistler et al., 1985). However, others have shown both by electron microscopy and work with fluorescent immunostaining for connexins that small gap junction plaques are present on the short sides of fiber cells in the cortex, and deeper in the lens, the larger plaques are broken up and smaller plaques are found in both the broad and short sides of the fiber cells (Jacobs et al., 2004b). Other studies have shown that plaques in the lens fibers are made up of both Cx46 and 50 (Paul et al., 1991), so it seems likely that the Cx50 shows a similar distribution in the broad and short sides, and if so the galectin related protein could interact there.

Actin was also commonly found in the pull-down experiments, and due it is abundance within cells, could be considered a false-positive contaminant. It was however, only occasionally identified in the negative control samples but was always found in the Cx50 pull-down experiments. Resolution of the pull-down proteins on SDS-

PAGE gels confirmed that a protein of 40 kDa, corresponding to actin, was present in samples with GST-Cx50 fusion proteins but not in the GST negative controls (results not shown). Electron microscopy has shown F-actin bundles in close proximity to gap junctions in primate lenses. In rat lens, this actin-gap junction association was not detected as actin was limited to the short sides of the lens fibers (Lo et al., 1994), which is the same distribution as the galectin- related interfiber protein described above.

Systems biology

One of the exciting applications of combining mass spectrometry with proteomics is the potential development of more powerful tools for studies in systems biology research. While the term systems biology is used in many contexts, it is often considered to involve a very broad examination of interacting or interdependent components that form an integrated whole. In a sense, biology has always been concerned with systems, as every object studied in biology is composed of smaller components. Physiology especially has always been concerned with how the function of various components are integrated in the greater whole. For example, the Hodgkin Huxley model of the propagation of the action potential (Hodgkin & Huxley, 1952) is representative of systems level thinking. Likewise, Denis Noble's model of electrical conduction in the heart (Noble, 2002) is another example of systems biology.

The current excitement around systems level thinking applies more to molecular systems biology and refers to attempts to integrate the large amount of data produced by the various –omics techniques. A definition of systems biology from the Institute for Systems Biology in Seattle that is relevant to this molecular biology approach is:

Systems biology is the study of an organism, viewed as an *integrated* and *interacting network* of genes, proteins and biochemical reactions which give rise to life. Instead of analyzing individual components or aspects of the organism, such as sugar metabolism or a cell nucleus, systems biologists focus on all the components and the interactions among them, all as part of one system (http://www.systemsbiology.org/Intro_to_Systems_Biology/Systems_Biology_--_the_21st_Century_Science).

Of course, an organism can be viewed as a network of genes and proteins, but this approach to systems biology may fail to address the complexity and hierarchy of organismal-level biology. The fundamental unit of life is generally thought to be the cell, and most biochemical and genetic networks function in the cell. However, even within the cell, different proteins can have different functions depending on the cellular compartment they are in. For example, beta catenin is a transcription factor in the nucleus but acts as a component of tight junctions at the cell surface.

One requirement of systems biology is that it is quantitative; the aforementioned Hogkin Huxley and Noble models are both examples mathematical models that help define the systems accurately. One major hurdle that proteomics needs to overcome if it is really going to be applied fully to a systems biology approach is the requirement for quantitative standards. Currently, it is possible to do relative quantification with techniques such as iTRAQ (isobaric tags for relative and absolute quantification) or differential gel analysis, but currently, absolute quantification can only be achieved by examining a few targets, coupled with a synthetic standard that is constructed for comparison. The lack of universal standards makes it very difficult to compare results from one laboratory to the next. When trying to resolve protein-protein interactions *in vivo*, for example, it currently is not possible to measure a multitude of parameters, such as binding coefficients. The lack of such information will continue to hinder the future

development of realistic systemic models. I can see the development of cell culture lines with metabolically-labeled proteins as offering at least a partial solution to this problem.

Many researchers who use MS to detect protein phosphorylations would like to know how much of the protein is phosphorylated. Currently, HPLC and MALDI in the methods used here are not amenable to quantitative analysis. As a peptide and modified peptide elute, some of the peaks of each peptide may not overlap in time, so that they are present in a different set of spots on the target. In addition, the matrix- peptide spot is not homogeneous and some peptides may be present in larger amounts in one area of the spot relative to another. Mass spectrometers can use stable isotopes as labels for quantitative analysis of specific compounds, but that becomes more difficult in proteomics where we examine hundreds to thousands of proteins/peptides.

In proteomics, there are two strategies for introducing mass labels to samples for quantitative analysis. In one, quantifiable chemical reagents are combined with the protein or peptides. The iTRAQ system is a set of isobaric reagents that have a mass label or reporter ion that is produced in collision induced disassociation experiments. Other methods include digestion of the proteins in ^{18}O water, which results in 2 Da-shift of all peptides digested in this labeled water, but this method can produce overlapping isotope envelopes that cause difficulties in resolving the signals.

The alternative is to metabolically label the proteins. In this procedure, the organism or cells in culture are fed nutrients containing a stable isotope (e.g. ^{15}N) which will shift the mass of the proteins/peptides by 1 Da per nitrogen. This offers some advantages over labeling the peptides with iTRAQ, in that the samples can be combined before any extraction steps and protein separation techniques such as 2D gels can be used

without introducing systematic errors to the samples. This allows us to use the best separation technology we have currently for proteins, but we don't have to worry about multiple proteins in each gel band, because we are comparing the individual peptides from each protein in the mass spectrometer.

One of the main results from these lens connexin studies was the identification of phosphorylation sites on the proteins. The phosphorylation of proteins is not random, but involves coordinated cell signaling and activation of kinases. Cell signaling is a system that has both inter- and extracellular components. Cells transfer information from the exterior of the cell though a network of proteins and small molecules. The most important or at least the best studied system is the network of kinases and phosphatases that control the phosphorylation of proteins in the cell (Kholodenko, 2006). The addition of phosphate to the protein can either directly or indirectly altering a binding site (Ptashne, 2009). A prime example of creating a binding site is tyrosine phosphorylation, which can create a binding site for proteins containing a src homology 2 domain (Moran et al., 1990; Pawson & Gish, 1992). A change in a binding site can result in a greater affinity for a substrate molecule or another protein resulting in increased or decreased activation or localization to a specific region of the cell. Originally, signaling cascades were described in a linear fashion; a receptor at the cell surface binds a ligand and becomes activated and initiates a cascade of phosphorylation reactions that result in some change in cellular activity. However, it is clear that each of the canonical pathways interacts with various others, which we typically define as crosstalk. Crosstalk is not just a peripheral phenomenon, but rather, it is central to how the cell processes information. For example, activation of the transcription factor NF-κB normally results in the

inhibition of apoptosis, but in certain contexts it can induce apoptosis (Perkins & Gilmore, 2006). In the standard pathway, NF-κB is activated by the phosphorylation of the inhibitory protein (IκB) by a kinase (IKK) and subsequent degradation of the inhibitory protein. However, several other kinases can phosphorylate IκB, including casein kinase 2 and PI3K/pkB. As well, NF-κB itself can be phosphorylated by these and other kinases. This diverse variety of different phosphorylations can then affect the activity of NF-κB and alter which genes get activated by this important transcription factor.

The phenomenon of crosstalk seems to be ubiquitous, with many kinases phosphorylating many other proteins, thereby impacting many pathways. There are also indirect ways that protein activation can affect another pathway. For instance, there are far fewer phosphatases than kinases within cells, and hence they have a broader specificity than most kinases (reviewed in (Virshup & Shenolikar, 2009). However, phosphatases are often part of multiprotein complexes, which assist in their regulation and thereby provide degrees of specificity for these enzymes. Activation of a kinase can therefore cause another pathway to remain active longer, because of the competitive inhibition of the phosphatase. In addition, other enzymes such as proteases can be conditionally activated and the effect of them on a target is irreversible and can have more lasting effects on the cell.

Various kinases have different specificities; some phosphorylate a small number of targets, whereas others have many different targets (reviewed in: (Zhu et al., 2005; Ubersax & Ferrell, Jr., 2007)). The specificity of many kinases is still rather unclear. Kinases tend to be very promiscuous in *in vitro* studies, but many are thought to be less

so in the actual cell (Pinna & Ruzzene, 1996). However, some large scale studies of phosphorylation during mitosis (Olsen et al., 2010; Dephoure et al., 2008) indicate that over a 1000 proteins show unregulated phosphorylation at more than one residue within the protein. Many of the sites are putative CDK sites, and during mitosis, >80% of the putative CDK sites were observed to be phosphorylated.

For large scale analysis of the effects on protein in the cell, mass spectrometry would seem ideal, and currently it does offer the best way to obtain a closer inspection of the state of the cell's proteins, but it is still far from ideal. Firstly, there are issues of obtaining adequate spectral coverage of the proteins. Rarely do we get complete coverage of the sequence of any particular protein and often phosphorylated proteins are present in such low abundance that they are masked by the more abundant peptides. There are techniques for enriching the sample with phosphorylated proteins such as using IMAC (Cao & Stults, 1999) or titanium dioxide (Pinkse et al., 2004).

Another consideration concerning our current MS analyses is that our measurements are made over a population of cells and not each individual cell. What if the result of activating a particular receptor is a change in the amplitude or frequency of phosphorylations through the network? Some models of the MAPK pathway indicate that it is possible to have waves of kinase activity if autophosphorylation is a feature of the pathway (Munoz-Garcia & Kholodenko, 2010). As the measurements are made over a number of cells, and if they are not synchronized, there is not going to be any measurable change in phosphorylation, as the individual fluctuations will be masked by the majority. This is a problem typical of most protein and RNA expression studies, although it is now possible to conduct single cell measurements of fluorescently labeled

protein using high resolution microscopy techniques (Newman et al., 2006; Sigal et al., 2006) or extract RNA from single cells using laser-micro-dissection techniques.

A final problem is the variability due to the measurement techniques. My experience with iTRAQ indicates that there is a large degree of variation inherent in the process, which is compounded when we are looking at complex mixtures of peptides. At the protein level, there was some variation from replicate to replicate, but individual peptides of a protein showed a very large degree of variation between measurements. This could be due to a number of factors. One possibility is just random variation in reporter ion counts especially for low abundance peptides. When the ion counts are low there is the possibility of large errors in comparing the ratios of the reporter peaks. Small differences in quantity of protein can result in differences in missed cleavages. Another possibility is that samples have differences in post- translational modifications, which results in differences in specific peptides.

The shotgun approach and systems biology

Our investigation of connexin phosphorylation followed a shotgun type approach, in that the cells were processed and crude extracts were digested with proteases to get a large number of different peptides from several proteins. In an ancillary study, I have likewise begun examining the proteomes of wheat and a related model monocot plant, *Brachypodium distachyon,* using mass spectrometry techniques, but in this plant study, I am using iTRAQ techniques in an attempt to quantify differences in protein expression in plants subjected to different treatments.

To date, several hundred proteins have been identified within both plant species, but initial trials were fraught with a lack of consistency in the data; often the technical

replicates produced just as much variation as that seen between the two types of plants. To reduce magnitude of these systematic errors, *Brachypodium* plants were subsequently grown with ^{15}N-labelled media. This method permitted the samples to be combined at the beginning of the analysis, with proteins extracted and digested together to avoid differences in sample preparations. The heavy isotope-labeled plants showed a 90% incorporation rate, which gives a sufficient separation between the peaks, and enabled a much more accurate quantitative analysis to be conducted.

One of the problems with shotgun analysis of proteins such as connexins is that it only gives us a narrow view of the PTMs over several different connexins. There are multiple levels of organization to consider beyond the level of individual residues. For example, are the phosphorylated residues from Cx50's C- terminal scattered randomly, or are only a small set of proteins phosphorylated in multiple locations? At higher levels of organization, the connexins form hexamers, so it would be ideal to determine the stoichiometry of modifications in this case, i.e. are multiple connexins phosphorylated in one connexon or is only one connexin modified? In addition, modified connexins might be distributed across wide regions of the cells or may be localized only to certain plaques.

There are some ways to address issues such as stoichiometry. One is to use native PAGE methods such as blue native PAGE (Schagger & von, 1991; Zheng et al., 2011) to purify individual connexons, and the connexons are then subjected to a second dimension SDS PAGE to separate the individual connexins. It may be possible then to identify both the proportion of Cx46 and 50 as well as whether only a few connexins are phosphorylated. An alternative mass spectrometry approach would be to use ESI of the

whole complex. Some workers have successfully ionized large membrane complexes using this technique (Thangaraj et al., 2010). It is then possible to use a variety of fragmentation methods such as collision activated, electron capture, electron transfer, or infrared multi-photon dissociation techniques to fragment the protein and determine the extent of post- translational modifications (McLafferty et al., 2007). Presently, it is not clear whether stoichiometry of connexin phosphorylations regulates the internalization of gap junctions (i.e. how many connexins/connexons need to be phosphorylated to induced endocytosis of the gap junctions?) but even that question may be addressed in the future by using chemical imaging techniques such as Raman microscopy (Evans et al., 2005).

Despite the lack of information on the stoichiometry of connexin PTMs, shotgun methods still provide a great deal of information on the phosphorylation states of the proteins. Knowledge of specific sites of Cx43 allowed the production of antibodies to Cx43 with serine 268 phosphorylated. This highly specific antibody has proven exceptionally useful for localizing the phosphorylated form of the connexin and for measuring its abundance in cells under different conditions (Solan et al., 2003; Richards et al., 2004). It is possible that phospho-specific antibodies to the lens connexin could be useful as well. It is also possible that many of the phosphorylations are independent of each other, and particular sites could represent the activities of particular kinases. As our knowledge of kinase activity improves, we may ultimately be able to assess the kinase activity of a cell by shotgun- type methods.

General Discussion and Conclusions

The use of mass spectrometry in the study of proteins has grown immensely since the discovery of soft ionization techniques. The study of membrane proteins is difficult, because many of the methods biochemists have developed are incompatible with mass spectrometry, especially the use of detergents. This study used lens fibers to study the primary protein component of vertebrate gap junctions, the connexins. The bovine lens was shown to be a good system for the study of connexins, as the plasma membrane was easy to purify and there was ample protein for the mass spectrometric analysis. I was able to identify several post- translation modifications to both types of lens fiber connexins.

The mouse lens proved to be more difficult to study. I was able to obtain some information on the post- translational modifications of each connexin, but it was difficult to get as good quality spectra as was obtained for the cow lens analyses. The mouse has a central role in biological research of vertebrates and it would be much more useful to have information on the protein functions in mouse. More effort will be required to develop techniques to reduce or eliminate some of the most abundant contaminating proteins from the lens protein extracts, to improve the ability to acquire spectra over a wider dynamic range. Some of the spectra from Cx50 were very strong, and it should be possible with better separations to get better coverage of the protein.

The usefulness of retention time for analyzing samples of moderate complexity is also clear. When I began this work before sMART was developed, I spent

considerable time acquiring spectra of peptides from proteins that were not my primary interest. However, some of the post-translational modifications such as the truncations and probably the acetylations would not have been found if I had not obtained a large number of tandem mass spectra instead of focusing on what sMART was able to identify. As with all technological applications, improvements in mass spectrometry continue to be developed, and much of that data may also have been better acquired using ESI-MS with automatic data acquisition instead of manually using MALDI MS.

Ultimately, the main limitation of this work is the lack of information on biological function. Knowledge of the presence of phosphorylation is only the first step, and further research is required to see if there is any functional reason for the modifications that were observed. That was, and still is the reason it is important to use an appropriate model organism, such as the mouse.

Reference List

Abraham, V., Chou, M. L., George, P., Pooler, P., Zaman, A., Savani, R. C. et al. (2001). Heterocellular gap junctional communication between alveolar epithelial cells. *Am.J.Physiol Lung Cell Mol.Physiol, 280,* L1085-L1093.

Aebersold, R. & Mann, M. (2003). Mass spectrometry-based proteomics. *Nature, 422,* 198-207.

Ahmad, S. & Evans, W. H. (2002). Post-translational integration and oligomerization of connexin 26 in plasma membranes and evidence of formation of membrane pores: implications for the assembly of gap junctions. *Biochem.J., 365,* 693-699.

Arora, A., Minogue, P. J., Liu, X., Addison, P. K., Russel-Eggitt, I., Webster, A. R. et al. (2008). A novel connexin50 mutation associated with congenital nuclear pulverulent cataracts. *Journal of Medical Genetics, 45,* 155-160.

Banks, E. A., Toloue, M. M., Shi, Q., Zhou, Z. J., Liu, J., Nicholson, B. J. et al. (2009). Connexin mutation that causes dominant congenital cataracts inhibits gap junctions, but not hemichannels, in a dominant negative manner. *J.Cell Sci., 122,* 378-388.

Baranova, A., Ivanov, D., Petrash, N., Pestova, A., Skoblov, M., Kelmanson, I. et al. (2004). The mammalian pannexin family is homologous to the invertebrate innexin gap junction proteins. *Genomics, 83,* 706-716.

Bauer, R., Loer, B., Ostrowski, K., Martini, J., Weimbs, A., Lechner, H. et al. (2005). Intercellular communication: the Drosophila innexin multiprotein family of gap junction proteins. *Chem.Biol., 12,* 515-526.

Beardslee, M. A., Laing, J. G., Beyer, E. C., & Saffitz, J. E. (1998). Rapid turnover of connexin43 in the adult rat heart. *Circulation Research, 83,* 629-635.

Behnia, R., Panic, B., Whyte, J. R., & Munro, S. (2004). Targeting of the Arf-like GTPase Arl3p to the Golgi requires N-terminal acetylation and the membrane protein Sys1p. *Nat.Cell Biol., 6,* 405-413.

Benedetti, E. L. & Emmelot, P. (1968). Hexagonal array of subunits in tight junctions separated from isolated rat liver plasma membranes. *J.Cell Biol., 38,* 15-24.

Bennett, M. V., Barrio, L. C., Bargiello, T. A., Spray, D. C., Hertzberg, E., & Saez, J. C. (1991). Gap junctions: new tools, new answers, new questions. *Neuron, 6,* 305-320.

Bergoffen, J., Scherer, S. S., Wang, S., Oronzi Scott, M., Bone, L. J., Paul, D. L. et al. (1993). Connexin mutations in X-linked Charcot-Marie-Tooth disease. *Science, 262,* 2039-2042.

Berry, V., Mackay, D., Khaliq, S., Francis, P. J., Hameed, A., Anwar, K. et al. (1999). Connexin 50 mutation in a family with congenital 'zonular nuclear' pulverulent cataract of Pakistani origin. *Human Genetics, 105,* 168-170.

Berthoud, V. M., Bassnett, S., & Beyer, E. C. (1999). Cultured chicken embryo lens cells resemble differentiating fiber cells in vivo and contain two kinetic pools of connexin56. *Exp.Eye Res., 68,* 475-484.

Berthoud, V. M., Minogue, P. J., Laing, J. G., & Beyer, E. C. (2004). Pathways for degradation of connexins and gap junctions. *Cardiovascular Research, 62,* 256-267.

Beyer, E. C., Paul, D. L., & Goodenough, D. A. (1987). Connexin43: a protein from rat heart homologous to a gap junction protein from liver. *J.Cell Biol., 105,* 2621-2629.

Biswas, S. K., Lee, J. E., Brako, L., Jiang, J. X., & Lo, W. K. (2010). Gap junctions are selectively associated with interlocking ball-and-sockets but not protrusions in the lens. *Mol.Vis., 16,* 2328-2341.

Blom, N., Gammeltoft, S., & Brunak, S. (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J.Mol.Biol., 294,* 1351-1362.

Blonder, J., Terunuma, A., Conrads, T. P., Chan, K. C., Yee, C., Lucas, D. A. et al. (2004). A proteomic characterization of the plasma membrane of human epidermis by high-throughput mass spectrometry. *J.Invest Dermatol., 123,* 691-699.

Bodenstine, T. M., Vaidya, K. S., Ismail, A., Beck, B. H., Cook, L. M., Diers, A. R. et al. (2010). Homotypic gap junctional communication associated with metastasis suppression increases with PKA activity and is unaffected by PI3K inhibition. *Cancer Res., 70,* 10002-10011.

Bukauskas, F. F., Jordan, K., Bukauskiene, A., Bennett, M. V. L., Lampe, P. D., Laird, D. W. et al. (2000). Clustering of connexin 43-enhanced green fluorescent protein gap junction channels and functional coupling in living cells. *Proceedings of the National Academy of Sciences of the United States of America, 97,* 2556-2561.

Caldwell, K. K., Sosa, M., & Buckley, C. T. (2006). Identification of mitogen-activated protein kinase docking sites in enzymes that metabolize phosphatidylinositols and inositol phosphates. *Cell Commun.Signal., 4,* 2.

Cameron, S. J., Malik, S., Akaike, M., Lerner-Marmarosh, N., Yan, C., Lee, J. D. et al. (2003). Regulation of epidermal growth factor-induced connexin 43 gap junction communication by big mitogen-activated protein kinase1/ERK5 but not ERK1/2 kinase activation. *J.Biol.Chem., 278,* 18682-18688.

Cao, P. & Stults, J. T. (1999). Phosphopeptide analysis by on-line immobilized metal-ion affinity chromatography-capillary electrophoresis-electrospray ionization mass spectrometry. *J.Chromatogr.A, 853,* 225-235.

Colussi, C., Rosati, J., Straino, S., Spallotta, F., Berni, R., Stilli, D. et al. (2011). Nepsilon-lysine acetylation determines dissociation from GAP junctions and

lateralization of connexin 43 in normal and dystrophic heart. *Proc.Natl.Acad.Sci.U.S.A, 108,* 2795-2800.

Cooper, C. D., Solan, J. L., Dolejsi, M. K., & Lampe, P. D. (2000). Analysis of connexin phosphorylation sites. *Methods, 20,* 196-204.

Craig, R. & Beavis, R. C. (2004). TANDEM: matching proteins with tandem mass spectra. *Bioinformatics., 20,* 1466-1467.

Dahl, G., Azarnia, R., & Werner, R. (1981). Induction of cell--cell channel formation by mRNA. *Nature, 289,* 683-685.

de Boer, T. P. & van der Heyden, M. A. (2005). Xenopus connexins: how frogs bridge the gap. *Differentiation, 73,* 330-340.

del Castillo, F. J., Cohen-Salmon, M., Charollais, A., Caille, D., Lampe, P. D., Chavrier, P. et al. (2010). Consortin, a trans-Golgi network cargo receptor for the plasma membrane targeting and recycling of connexins. *Hum.Mol.Genet., 19,* 262-275.

Delamere, N. A. & Tamiya, S. (2004). Expression, regulation and function of Na,K-ATPase in the lens. *Progress in Retinal and Eye Research, 23,* 593-615.

Dephoure, N., Zhou, C., Villen, J., Beausoleil, S. A., Bakalarski, C. E., Elledge, S. J. et al. (2008). A quantitative atlas of mitotic phosphorylation. *Proc.Natl.Acad.Sci.U.S.A, 105,* 10762-10767.

Dermietzel, R., Leibstein, A., Frixen, U., Janssen-Timmen, U., Traub, O., & Willecke, K. (1984). Gap junctions in several tissues share antigenic determinants with liver gap junctions. *EMBO J., 3,* 2261-2270.

Dewey, M. M. & Barr, L. (1962). Intercellular Connection between Smooth Muscle Cells: the Nexus. *Science, 137,* 670-672.

Dole, M., Mack, L. L., & Hines, R. L. (1968). Molecular Beams of Macroions. *Journal of Chemical Physics, 49,* 2240-&.

Driessen, H. P., Ramaekers, F. C., Vree Egberts, W. T., Dodemont, H. J., de Jong, W. W., Tesser, G. I. et al. (1983). The function of N alpha-acetylation of the eye-lens crystallins. *Eur.J.Biochem., 136,* 403-406.

Duffy, H. S., Delmar, M., & Spray, D. C. (2002). Formation of the gap junction nexus: binding partners for connexins. *J.Physiol Paris, 96,* 243-249.

Evans, C. L., Potma, E. O., Puoris'haag, M., Cote, D., Lin, C. P., & Xie, X. S. (2005). Chemical imaging of tissue in vivo with video-rate coherent anti-Stokes Raman scattering microscopy. *Proc.Natl Acad.Sci.U.S.A, 102,* 16807-16812.

Fallon, R. F. & Goodenough, D. A. (1981). Five-hour half-life of mouse liver gap-junction protein. *J.Cell Biol., 90,* 521-526.

Fan, J., Fariss, R. N., Purkiss, A. G., Slingsby, C., Sandilands, A., Quinlan, R. et al. (2005). Specific interaction between lens MIP/Aquaporin-0 and two members of the gamma-crystallin family. *Mol.Vis., 11,* 76-87.

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., & Whitehouse, C. M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science, 246,* 64-71.

Fenyo, D. & Beavis, R. C. (2003). A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. *Anal.Chem., 75,* 768-774.

Fort, A. G., Murray, J. W., Dandachi, N., Davidson, M. W., Dermietzel, R., Wolkoff, A. W. et al. (2011). In Vitro Motility of Liver Connexin Vesicles along Microtubules Utilizes Kinesin Motors. *J.Biol.Chem., 286,* 22875-22885.

Forte, G. M., Pool, M. R., & Stirling, C. J. (2011). N-terminal acetylation inhibits protein targeting to the endoplasmic reticulum. *PLoS.Biol., 9,* e1001073.

Fraser, S. E., Green, C. R., Bode, H. R., & Gilula, N. B. (1987). Selective disruption of gap junctional communication interferes with a patterning process in hydra. *Science, 237,* 49-55.

Fujimoto, K., Nagafuchi, A., Tsukita, S., Kuraoka, A., Ohokuma, A., & Shibata, Y. (1997). Dynamics of connexins, E-cadherin and á-catenin on cell membranes during gap junction formation. *Journal of Cell Science, 110,* 311-322.

Furakawa, T. & Furshpan, E. J. (1963). Two inhibitory mechanisms in the Mauthner neurons of goldfish. *J.Neurophysiol., 26,* 140-176.

Furshpan, E. J. & Potter, D. D. (1959). Transmission at the giant motor synapses of the crayfish. *J.Physiol, 145,* 289-325.

Gao, J., Sun, X., Martinez-Wittinghan, F. J., Gong, X., White, T. W., & Mathias, R. T. (2004). Connections between connexins, calcium, and cataracts in the lens. *Journal of General Physiology, 124,* 289-300.

Gautschi, M., Just, S., Mun, A., Ross, S., Rucknagel, P., Dubaquie, Y. et al. (2003). The yeast N(alpha)-acetyltransferase NatA is quantitatively anchored to the ribosome and interacts with nascent polypeptides. *Mol.Cell Biol., 23,* 7403-7414.

Ghoshroy, S., Goodenough, D. A., & Sosinsky, G. E. (1995). Preparation, characterization, and structure of half gap junctional layers split with urea and EGTA. *J.Membr.Biol., 146,* 15-28.

Giepmans, B. N., Verlaan, I., & Moolenaar, W. H. (2001). Connexin-43 interactions with ZO-1 and alpha- and beta-tubulin. *Cell Commun.Adhes., 8,* 219-223.

Gong, X., Baldo, G. J., Kumar, N. M., Gilula, N. B., & Mathias, R. T. (1998). Gap junctional coupling in lenses lacking +¦3 connexin. *Proceedings of the National Academy of Sciences of the United States of America, 95,* 15303-15308.

Gonzalez, F. A., Raden, D. L., & Davis, R. J. (1991). Identification of substrate recognition determinants for human ERK1 and ERK2 protein kinases. *J.Biol.Chem., 266,* 22159-22163.

Goodenough, D. A. (1974). Bulk isolation of mouse hepatocyte gap junctions. Characterization of the principal protein, connexin. *J.Cell Biol., 61,* 557-563.

Goodenough, D. A. & Stoeckenius, W. (1972). The isolation of mouse hepatocyte gap junctions. Preliminary chemical characterization and x-ray diffraction. *J.Cell Biol., 54,* 646-656.

Govindarajan, R., Chakraborty, S., Johnson, K. E., Falk, M. M., Wheelock, M. J., Johnson, K. R. et al. (2010). Assembly of connexin43 into gap junctions is regulated differentially by E-cadherin and N-cadherin in rat liver epithelial cells. *Mol.Biol.Cell, 21,* 4089-4107.

Han, J., Little, M., David, L. L., Giblin, F. J., & Schey, K. L. (2004). Sequence and peptide map of guinea pig aquaporin 0. *Mol.Vis., 10,* 215-222.

Han, J. & Schey, K. L. (2004a). Proteolysis and mass spectrometric analysis of an integral membrane: aquaporin 0. *J.Proteome Res., 3,* 807-812.

Han, J. & Schey, K. L. (2004b). Proteolysis and mass spectrometric analysis of an integral membrane: aquaporin 0. *J.Proteome Res., 3,* 807-812.

Harris, A. L. (2001). Emerging issues of connexin channels: biophysics fills the gap. *Q.Rev.Biophys., 34,* 325-472.

Hershko, A., Heller, H., Eytan, E., Kaklij, G., & Rose, I. A. (1984). Role of the alpha-amino group of protein in ubiquitin-mediated protein breakdown. *Proc.Natl.Acad.Sci.U.S.A, 81,* 7021-7025.

Herve, J. C., Bourmeyster, N., Sarrouilhe, D., & Duffy, H. S. (2007). Gap junctional complexes: from partners to functions. *Prog.Biophys.Mol.Biol., 94,* 29-65.

Heys, K. R., Friedrich, M. G., & Truscott, R. J. (2008). Free and bound water in normal and cataractous human lenses. *Invest Ophthalmol.Vis.Sci., 49,* 1991-1997.

Hodgkin, A. L. & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology, 117,* 500-544.

Hudder, A. & Werner, R. (2000). Analysis of a Charcot-Marie-Tooth disease mutation reveals an essential internal ribosome entry site element in the connexin-32 gene. *J.Biol.Chem., 275,* 34586-34591.

Hwang, C. S., Shemorry, A., & Varshavsky, A. (2010). N-terminal acetylation of cellular proteins creates specific degradation signals. *Science, 327,* 973-977.

Iribarne, J. V. & Thomson, B. A. (1976). Evaporation of Small Ions from Charged Droplets. *Journal of Chemical Physics, 64,* 2287-2294.

Jacobs, M. D., Soeller, C., Sisley, A. M., Cannell, M. B., & Donaldson, P. J. (2004a). Gap junction processing and redistribution revealed by quantitative optical measurements of connexin46 epitopes in the lens. *Invest Ophthalmol.Vis.Sci., 45,* 191-199.

Jacobs, M. D., Soeller, C., Sisley, A. M. G., Cannell, M. B., & Donaldson, P. J. (2004b). Gap Junction Processing and Redistribution Revealed by Quantitative Optical

Measurements of Connexin46 Epitopes in the Lens. *Investigative Ophthalmology and Visual Science, 45,* 191-199.

Karas, M. & Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular mass exceeding 10,000 daltons. *Anal.Chem., 60,* 2299-2301.

Kelsell, D. P., Dunlop, J., Stevens, H. P., Lench, N. J., Liang, J. N., Parry, G. et al. (1997). Connexin 26 mutations in hereditary non-syndromic sensorineural deafness. *Nature, 387,* 80-83.

Kholodenko, B. N. (2006). Cell-signalling dynamics in time and space. *Nat.Rev.Mol.Cell Biol., 7,* 165-176.

Kilby, G. W., Truscott, R. J., Stuchbury, G. M., & Sheil, M. M. (1996). Mass spectrometry of lens crystallins: bovine beta-crystallins. *Rapid Commun.Mass Spectrom., 10,* 123-129.

Kistler, J., Goldie, K., Donaldson, P., & Engel, A. (1994). Reconstitution of native-type noncrystalline lens fiber gap junctions from isolated hemichannels. *J.Cell Biol., 126,* 1047-1058.

Kistler, J., Kirkland, B., & Bullivant, S. (1985). Identification of a 70,000-D protein in lens membrane junctional domains. *J.Cell Biol., 101,* 28-35.

Kjaer, K. W., Hansen, L., Eiberg, H., Leicht, P., Opitz, J. M., & Tommerup, N. (2004). Novel Connexin 43 (GJA1) mutation causes oculo-dento-digital dysplasia with curly hair. *American Journal of Medical Genetics, 127 A,* 152-157.

Knochenmuss, R. (2006). Ion formation mechanisms in UV-MALDI. *Analyst, 131,* 966-986.

Krokhin, O. V. (2006). Sequence-specific retention calculator. Algorithm for peptide retention prediction in ion-pair RP-HPLC: application to 300- and 100-A pore size C18 sorbents. *Anal.Chem., 78,* 7785-7795.

Krokhin, O. V., Craig, R., Spicer, V., Ens, W., Standing, K. G., Beavis, R. C. et al. (2004). An improved model for prediction of retention times of tryptic peptides in ion pair reversed-phase HPLC: its application to protein peptide mapping by off-line HPLC-MALDI MS. *Mol.Cell Proteomics., 3,* 908-919.

Krokhin, O. V., Ying, S., Cortens, J. P., Ghosh, D., Spicer, V., Ens, W. et al. (2006). Use of peptide retention time prediction for protein identification by off-line reversed-phase HPLC-MALDI MS/MS. *Anal.Chem., 78,* 6265-6269.

Kuffler, S. W. & Potter, D. D. (1964). GLIA IN THE LEECH CENTRAL NERVOUS SYSTEM: PHYSIOLOGICAL PROPERTIES AND NEURON-GLIA RELATIONSHIP. *J.Neurophysiol., 27,* 290-320.

Kyte, J. & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J.Mol.Biol., 157,* 105-132.

Laing, J. G. & Beyer, E. C. (1995). The gap junction protein connexin43 is degraded via the ubiquitin proteasome pathway. *Journal of Biological Chemistry, 270,* 26399-26403.

Laird, D. W. (2005). Connexin phosphorylation as a regulatory event linked to gap junction internalization and degradation. *Biochimica et Biophysica Acta - Biomembranes, 1711,* 172-182.

Laird, D. W., Castillo, M., & Kasprzak, L. (1995). Gap junction turnover, intracellular trafficking, and phosphorylation of connexin43 in brefeldin A-treated rat mammary tumor cells. *Journal of Cell Biology, 131,* 119-203.

Laird, D. W., Puranam, K. L., & Revel, J. P. (1991). Turnover and phosphorylation dynamics of connexin43 gap junction protein in cultured cardiac myocytes. *Biochemical Journal, 273,* 67-72.

Lampe, P. D. & Johnson, R. G. (1989). Phosphorylation of MP26, a lens junction protein, is enhanced by activators of protein kinase C. *J.Membr.Biol., 107,* 145-155.

Lampe, P. D. & Johnson, R. G. (1990). Amino acid sequence of in vivo phosphorylation sites in the main intrinsic protein (MIP) of lens membranes. *Eur.J.Biochem., 194,* 541-547.

Lampe, P. D. & Lau, A. F. (2000). Regulation of gap junctions by phosphorylation of connexins. *Archives of Biochemistry and Biophysics, 384,* 205-215.

Lampe, P. D. & Lau, A. F. (2004c). The effects of connexin phosphorylation on gap junctional communication. *Int.J.Biochem.Cell Biol., 36,* 1171-1186.

Lampe, P. D. & Lau, A. F. (2004b). The effects of connexin phosphorylation on gap junctional communication. *Int.J.Biochem.Cell Biol., 36,* 1171-1186.

Lampe, P. D. & Lau, A. F. (2004a). The effects of connexin phosphorylation on gap junctional communication. *Int.J.Biochem.Cell Biol., 36,* 1171-1186.

Lampe, P. D., TenBroek, E. M., Burt, J. M., Kurata, W. E., Johnson, R. G., & Lau, A. F. (2000). Phosphorylation of connexin43 on serine368 by protein kinase C regulates gap junctional communication. *J.Cell Biol., 149,* 1503-1512.

Lo, W. K., Mills, A., & Kuck, J. F. (1994). Actin filament bundles are associated with fiber gap junctions in the primate lens. *Exp.Eye Res., 58,* 189-196.

Lo, W. K. & Reese, T. S. (1993). Multiple structural types of gap junctions in mouse lens. *J.Cell Sci., 106 (Pt 1),* 227-235.

Lo, W. K., Shaw, A. P., Takemoto, L. J., Grossniklaus, H. E., & Tigges, M. (1996). Gap junction structures and distribution patterns of immunoreactive connexins 46 and 50 in lens regrowths of Rhesus monkeys. *Exp.Eye Res., 62,* 171-180.

Loboda, A. V., Krutchinsky, A. N., Bromirski, M., Ens, W., & Standing, K. G. (2000a). A tandem quadrupole/time-of-flight mass spectrometer with a matrix-assisted laser desorption/ionization source: design and performance. *Rapid Commun.Mass Spectrom., 14,* 1047-1057.

Loboda, A. V., Krutchinsky, A. N., Bromirski, M., Ens, W., & Standing, K. G. (2000b). A tandem quadrupole/time-of-flight mass spectrometer with a matrix-assisted laser desorption/ionization source: design and performance. *Rapid Commun.Mass Spectrom., 14,* 1047-1057.

Locke, D., Koreen, I. V., & Harris, A. L. (2006b). Isoelectric points and post-translational modifications of connexin26 and connexin32. *FASEB J., 20,* 1221-1223.

Locke, D., Koreen, I. V., & Harris, A. L. (2006a). Isoelectric points and post-translational modifications of connexin26 and connexin32. *FASEB J.*.

Loewenstein, W. R. & Kanno, Y. (1964). STUDIES ON AN EPITHELIAL (GLAND) CELL JUNCTION. I. MODIFICATIONS OF SURFACE MEMBRANE PERMEABILITY. *J.Cell Biol., 22,* 565-586.

Maciejewski, P. M., Peterson, F. C., Anderson, P. J., & Brooks, C. L. (1995). Mutation of serine 90 to glutamic acid mimics phosphorylation of bovine prolactin. *J.Biol.Chem., 270,* 27661-27665.

Mackay, D., Ionides, A., Kibar, Z., Rouleau, G., Berry, V., Moore, A. et al. (1999). Connexin46 mutations in autosomal dominant congenital cataract. *American Journal of Human Genetics, 64,* 1357-1364.

Maeda, S., Nakagawa, S., Suga, M., Yamashita, E., Oshima, A., Fujiyoshi, Y. et al. (2009). Structure of the connexin 26 gap junction channel at 3.5 A resolution. *Nature, 458,* 597-602.

Manthey, D., Banach, K., Desplantez, T., Lee, C. G., Kozak, C. A., Traub, O. et al. (2001). Intracellular domains of mouse connexin26 and -30 affect diffusional and electrical properties of gap junction channels. *J.Membr.Biol., 181,* 137-148.

Marquart, K. H. (1977). So called annular gap junctions in bone cells of normal mice. *Experientia, 33,* 270-272.

Mathias, R. T., Rae, J. L., & Baldo, G. J. (1997). Physiological properties of the normal lens. *Physiological Reviews, 77,* 21-50.

McAvoy, J. W., Chamberlain, C. G., de Iongh, R. U., Hales, A. M., & Lovicu, F. J. (1999). Lens development. *Eye (Lond), 13 (Pt 3b),* 425-437.

McLafferty, F. W., Breuker, K., Jin, M., Han, X., Infusini, G., Jiang, H. et al. (2007). Top-down MS, a powerful complement to the high capabilities of proteolysis proteomics. *FEBS J., 274,* 6256-6268.

Moran, M. F., Koch, C. A., Anderson, D., Ellis, C., England, L., Martin, G. S. et al. (1990). Src homology region 2 domains direct protein-protein interactions in signal transduction. *Proc.Natl.Acad.Sci.U.S.A, 87,* 8622-8626.

Munoz-Garcia, J. & Kholodenko, B. N. (2010). Signalling over a distance: gradient patterns and phosphorylation waves within single cells. *Biochem.Soc.Trans., 38,* 1235-1241.

Musil, L. S. & Goodenough, D. A. (1993). Multisubunit assembly of an integral plasma membrane channel protein, gap junction connexin43, occurs after exit from the ER. *Cell, 74,* 1065-1077.

Narita, K. (1958). Isolation of acetylpeptide from enzymic digests of TMV-protein. *Biochim.Biophys.Acta, 28,* 184-191.

Newman, J. R., Ghaemmaghami, S., Ihmels, J., Breslow, D. K., Noble, M., DeRisi, J. L. et al. (2006). Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature, 441,* 840-846.

Nicholson, B. J., Gros, D. B., Kent, S. B., Hood, L. E., & Revel, J. P. (1985). The Mr 28,000 gap junction proteins from rat heart and liver are different but related. *J.Biol.Chem., 260,* 6514-6517.

Nielsen, P. A., Baruch, A., Shestopalov, V. I., Giepmans, B. N., Dunia, I., Benedetti, E. L. et al. (2003). Lens connexins alpha3Cx46 and alpha8Cx50 interact with zonula occludens protein-1 (ZO-1). *Mol.Biol.Cell, 14,* 2470-2481.

Noble, D. (2002). Modeling the heart--from genes to cells to the whole organ. *Science, 295,* 1678-1682.

Ogden, A. T., Nunes, I., Ko, K., Wu, S., Hines, C. S., Wang, A. F. et al. (1998). GRIFIN, a novel lens-specific protein related to the galectin family. *J.Biol.Chem., 273,* 28889-28896.

Olsen, J. V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M. L., Jensen, L. J. et al. (2010). Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci.Signal., 3,* ra3.

Pal, J. D., Liu, X., Mackay, D., Shiels, A., Berthoud, V. M., Beyer, E. C. et al. (2000). Connexin46 mutations linked to congenital cataract show loss of gap junction channel function. *American Journal of Physiology - Cell Physiology, 279,* C596-C602.

Pappin, D. J., Hojrup, P., & Bleasby, A. J. (1993). Rapid identification of proteins by peptide-mass fingerprinting. *Curr.Biol., 3,* 327-332.

Patterson, S. D. & Aebersold, R. H. (2003). Proteomics: the first decade and beyond. *Nat.Genet., 33 Suppl,* 311-323.

Paul, D. L. (1986). Molecular cloning of cDNA for rat liver gap junction protein. *J.Cell Biol., 103,* 123-134.

Paul, D. L., Ebihara, L., Takemoto, L. J., Swenson, K. I., & Goodenough, D. A. (1991). Connexin46, a novel lens gap junction protein, induces voltage-gated currents in nonjunctional plasma membrane of Xenopus oocytes. *J.Cell Biol., 115,* 1077-1089.

Pawson, T. & Gish, G. D. (1992). SH2 and SH3 domains: from structure to function. *Cell, 71,* 359-362.

Peracchia, C. (1973a). Low resistance junctions in crayfish. I. Two arrays of globules in junctional membranes. *J.Cell Biol., 57,* 66-76.

Peracchia, C. (1973b). Low resistance junctions in crayfish. II. Structural details and further evidence for intercellular channels by freeze-fracture and negative staining. *J.Cell Biol., 57,* 54-65.

Perkins, N. D. & Gilmore, T. D. (2006). Good cop, bad cop: the different faces of NF-kappaB. *Cell Death.Differ., 13,* 759-772.

Pestana, A. & Pitot, H. C. (1975). Acetylation of nascent polypeptide chains on rat liver polyribosomes in vivo and in vitro. *Biochemistry, 14,* 1404-1412.

Phelan, P., Bacon, J. P., Davies, J. A., Stebbings, L. A., Todman, M. G., Avery, L. et al. (1998). Innexins: a family of invertebrate gap-junction proteins. *Trends Genet., 14,* 348-349.

Piehl, M., Lehmann, C., Gumpert, A., Denizot, J. P., Segretain, D., & Falk, M. M. (2007). Internalization of large double-membrane intercellular vesicles by a clathrin-dependent endocytic process. *Mol.Biol.Cell, 18,* 337-347.

Pinkse, M. W., Uitto, P. M., Hilhorst, M. J., Ooms, B., & Heck, A. J. (2004). Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal.Chem., 76,* 3935-3943.

Pinna, L. A. & Ruzzene, M. (1996). How do protein kinases recognize their substrates? *Biochim.Biophys.Acta, 1314,* 191-225.

Polevoda, B., Cardillo, T. S., Doyle, T. C., Bedi, G. S., & Sherman, F. (2003). Nat3p and Mdm20p are required for function of yeast NatB Nalpha-terminal acetyltransferase and of actin and tropomyosin. *J.Biol.Chem., 278,* 30686-30697.

Ptashne, M. (2009). Binding reactions: epigenetic switches, signal transduction and cancer. *Curr.Biol., 19,* R234-R241.

Purnick, P. E., Oh, S., Abrams, C. K., Verselis, V. K., & Bargiello, T. A. (2000). Reversal of the gating polarity of gap junctions by negative charge substitutions in the N-terminus of connexin 32. *Biophys.J., 79,* 2403-2415.

Revel, J. P. & Karnovsky, M. J. (1967b). Hexagonal array of subunits in intercellular junctions of the mouse heart and liver. *J.Cell Biol., 33,* C7-C12.

Revel, J. P. & Karnovsky, M. J. (1967a). Hexagonal array of subunits in intercellular junctions of the mouse heart and liver. *J.Cell Biol., 33,* C7-C12.

Richard, G. (2005). Connexin disorders of the skin. *Clinics in Dermatology, 23,* 23-32.

Richards, T. S., Dunn, C. A., Carter, W. G., Usui, M. L., Olerud, J. E., & Lampe, P. D. (2004). Protein kinase C spatially and temporally regulates gap junctional communication during human wound repair via phosphorylation of connexin43 on serine368. *J.Cell Biol., 167,* 555-562.

Robertson, J. D. (1963a). THE OCCURRENCE OF A SUBUNIT PATTERN IN THE UNIT MEMBRANES OF CLUB ENDINGS IN MAUTHNER CELL SYNAPSES IN GOLDFISH BRAINS. *J.Cell Biol., 19,* 201-221.

Robertson, J. D. (1963b). THE OCCURRENCE OF A SUBUNIT PATTERN IN THE UNIT MEMBRANES OF CLUB ENDINGS IN MAUTHNER CELL SYNAPSES IN GOLDFISH BRAINS. *J.Cell Biol., 19,* 201-221.

Sarma, J. D., Wang, F., & Koval, M. (2002). Targeted gap junction protein constructs reveal connexin-specific differences in oligomerization. *Journal of Biological Chemistry, 277,* 20911-20918.

Sasaki, T. & Garant, P. R. (1986). Fate of annular gap junctions in the papillary cells of the enamel organ in the rat incisor. *Cell and Tissue Research, 246,* 523-530.

Sasakura, Y., Shoguchi, E., Takatori, N., Wada, S., Meinertzhagen, I. A., Satou, Y. et al. (2003). A genomewide survey of developmentally relevant genes in Ciona intestinalis. X. Genes for cell junctions and extracellular matrix. *Dev.Genes Evol., 213,* 303-313.

Schagger, H. & von, J. G. (1991). Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Anal.Biochem., 199,* 223-231.

Schey, K. L., Fowler, J. G., Schwartz, J. C., Busman, M., Dillon, J., & Crouch, R. K. (1997a). Complete map and identification of the phosphorylation site of bovine lens major intrinsic protein. *Invest Ophthalmol.Vis.Sci., 38,* 2508-2515.

Schey, K. L., Fowler, J. G., Schwartz, J. C., Busman, M., Dillon, J., & Crouch, R. K. (1997b). Complete map and identification of the phosphorylation site of bovine lens major intrinsic protein. *Invest Ophthalmol.Vis.Sci., 38,* 2508-2515.

Schey, K. L., Fowler, J. G., Schwartz, J. C., Busman, M., Dillon, J., & Crouch, R. K. (1997c). Complete map and identification of the phosphorylation site of bovine lens major intrinsic protein. *Invest Ophthalmol.Vis.Sci., 38,* 2508-2515.

Schey, K. L., Little, M., Fowler, J. G., & Crouch, R. K. (2000). Characterization of human lens major intrinsic protein structure. *Invest Ophthalmol.Vis.Sci., 41,* 175-182.

Schiavi, A., Hudder, A., & Werner, R. (1999). Connexin43 mRNA contains a functional internal ribosome entry site. *FEBS Lett., 464,* 118-122.

Schubert, A. L., Schubert, W., Spray, D. C., & Lisanti, M. P. (2002). Connexin family members target to lipid raft domains and interact with caveolin-1. *Biochemistry, 41,* 5754-5764.

Setty, S. R., Strochlic, T. I., Tong, A. H., Boone, C., & Burd, C. G. (2004). Golgi targeting of ARF-like GTPase Arl3p requires its Nalpha-acetylation and the integral membrane protein Sys1p. *Nat.Cell Biol., 6,* 414-419.

Shaw, R. M., Fay, A. J., Puthenveedu, M. A., von, Z. M., Jan, Y. N., & Jan, L. Y. (2007). Microtubule plus-end-tracking proteins target gap junctions directly from the cell interior to adherens junctions. *Cell, 128,* 547-560.

Sigal, A., Milo, R., Cohen, A., Geva-Zatorsky, N., Klein, Y., Liron, Y. et al. (2006). Variability and memory of protein levels in human cells. *Nature, 444,* 643-646.

Singh, D. & Lampe, P. D. (2003). Identification of connexin-43 interacting proteins. *Cell Commun.Adhes., 10,* 215-220.

Smith, J. B., Thθvenon-Emeric, G., Smith, D. L., & Green, B. (1991). Elucidation of the primary structures of proteins by mass spectrometry. *Analytical Biochemistry, 193,* 118-124.

Solan, J. L., Fry, M. D., TenBroek, E. M., & Lampe, P. D. (2003). Connexin43 phosphorylation at S368 is acute during S and G2/M and in response to protein kinase C activation. *J.Cell Sci., 116,* 2203-2211.

Solan, J. L. & Lampe, P. D. (2007). Key connexin 43 phosphorylation events regulate the gap junction life cycle. *J.Membr.Biol., 217,* 35-41.

Sosinsky, G. E., Boassa, D., Dermietzel, R., Duffy, H. S., Laird, D. W., Macvicar, B. et al. (2011). Pannexin channels are not gap junction hemichannels. *Channels (Austin.), 5,* 193-197.

Srivastava, K., Chaves, J. M., Srivastava, O. P., & Kirk, M. (2008). Multi-crystallin complexes exist in the water-soluble high molecular weight protein fractions of aging normal and cataractous human lenses. *Exp.Eye Res., 87,* 356-366.

Strous, G. J., Berns, A. J., & Bloemendal, H. (1974). N-terminal acetylation of the nascent chains of alpha-crystallin. *Biochem.Biophys.Res.Commun., 58,* 876-884.

Suga, M., Maeda, S., Nakagawa, S., Yamashita, E., & Tsukihara, T. (2009). A description of the structural determination procedures of a gap junction channel at 3.5 A resolution. *Acta Crystallogr.D.Biol.Crystallogr., 65,* 758-766.

Swamy-Mruthinti, S. & Schey, K. L. (1997a). Mass spectroscopic identification of in vitro glycated sites of MIP. *Curr.Eye Res., 16,* 936-941.

Swamy-Mruthinti, S. & Schey, K. L. (1997b). Mass spectroscopic identification of in vitro glycated sites of MIP. *Curr.Eye Res., 16,* 936-941.

Sweeney, M. H. & Truscott, R. J. (1998). An impediment to glutathione diffusion in older normal human lenses: a possible precondition for nuclear cataract. *Exp.Eye Res., 67,* 587-595.

Thangaraj, B., Ryan, C. M., Souda, P., Krause, K., Faull, K. F., Weber, A. P. et al. (2010). Data-directed top-down Fourier-transform mass spectrometry of a large integral membrane protein complex: photosystem II from Galdieria sulphuraria. *Proteomics, 10,* 3644-3656.

Thomas, K. R. & Capecchi, M. R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell, 51,* 503-512.

Thomas, T., Jordan, K., Simek, J., Shao, Q., Jedeszko, C., Walton, P. et al. (2005). Mechanism of Cx43 and Cx26 transport to the plasma membrane and gap junction regeneration. *Journal of Cell Science, 118,* 4451-4462.

Thomson, M. & Gunawardena, J. (2009). Unlimited multistability in multisite phosphorylation systems. *Nature, 460,* 274-277.

Truscott, R. J. (2005). Age-related nuclear cataract-oxidation is the key. *Exp.Eye Res., 80,* 709-725.

Ubersax, J. A. & Ferrell, J. E., Jr. (2007). Mechanisms of specificity in protein phosphorylation. *Nat.Rev.Mol.Cell Biol., 8,* 530-541.

Unger, V. M., Kumar, N. M., Gilula, N. B., & Yeager, M. (1997). Projection structure of a gap junction membrane channel at 7 A resolution. *Nat.Struct.Biol., 4,* 39-43.

Unger, V. M., Kumar, N. M., Gilula, N. B., & Yeager, M. (1999). Three-dimensional structure of a recombinant gap junction membrane channel. *Science, 283,* 1176-1180.

van der Heyden, M. A., van, E. M., Wilders, R., de Bakker, J. M., & Opthof, T. (2004). Connexin43 orthologues in vertebrates: phylogeny from fish to man. *Dev.Genes Evol., 214,* 261-266.

VanSlyke, J. K. & Musil, L. S. (2002). Dislocation and degradation from the ER are regulated by cytosolic stress. *Journal of Cell Biology, 157,* 381-394.

Vidal, M. & Legrain, P. (1999). Yeast forward and reverse n-hybrid systems. *Nucleic Acids Research, 27,* 919-929.

Virshup, D. M. & Shenolikar, S. (2009). From promiscuity to precision: protein phosphatases get a makeover. *Mol.Cell, 33,* 537-545.

Walsh, C. T. (2005). *Postranslational Modification of Proteis: Expanding Nature's Inventory*. Chicago: B. Roberts.

Warn-Cramer, B. J., Lampe, P. D., Kurata, W. E., Kanemitsu, M. Y., Loo, L. W., Eckhart, W. et al. (1996). Characterization of the mitogen-activated protein kinase

phosphorylation sites on the connexin-43 gap junction protein. *J.Biol.Chem., 271,* 3779-3786.

Warner, A. E., Guthrie, S. C., & Gilula, N. B. (1984). Antibodies to gap-junctional protein selectively disrupt junctional communication in the early amphibian embryo. *Nature, 311,* 127-131.

Wei, C. J., Francis, R., Xu, X., & Lo, C. W. (2005). Connexin43 associated with an N-cadherin-containing multiprotein complex is required for gap junction formation in NIH3T3 cells. *Journal of Biological Chemistry, 280,* 19925-19936.

Weidmann, S. (1952). The electrical constants of Purkinje fibres. *J.Physiol, 118,* 348-360.

Werner, R. (2000). IRES elements in connexin genes: a hypothesis explaining the need for connexins to be regulated at the translational level. *IUBMB Life, 50,* 173-176.

Werner, R., Miller, T., Azarnia, R., & Dahl, G. (1985). Translation and functional expression of cell-cell channel mRNA in Xenopus oocytes. *J.Membr.Biol., 87,* 253-268.

White, T. W., Bruzzone, R., Wolfram, S., Paul, D. L., & Goodenough, D. A. (1994). Selective interactions among the multiple connexin proteins expressed in the vertebrate lens: the second extracellular domain is a determinant of compatibility between connexins. *J.Cell Biol., 125,* 879-892.

White, T. W., Goodenough, D. A., & Paul, D. L. (1998). Targeted ablation of connexin50 in mice results in microphthalmia and zonular pulverulent cataracts. *J.Cell Biol., 143,* 815-825.

White, T. W. & Paul, D. L. (1999). Genetic diseases and gene knockouts reveal diverse connexin functions. *Annual Review of Physiology, 61,* 283-310.

White, T. W., Wang, H., Mui, R., Litteral, J., & Brink, P. R. (2004). Cloning and functional expression of invertebrate connexins from Halocynthia pyriformis. *FEBS Lett., 577,* 42-48.

Willecke, K., Eiberger, J., Degen, J., Eckardt, D., Romualdi, A., Guldenagel, M. et al. (2002). Structural and functional diversity of connexin genes in the mouse and human genome. *Biol.Chem., 383,* 725-737.

Wu, C. C. & Yates, J. R., III (2003). The application of mass spectrometry to membrane proteomics. *Nat.Biotechnol., 21,* 262-267.

Xia, C. H., Cheng, C., Huang, Q., Cheung, D., Li, L., Dunia, I. et al. (2006). Absence of alpha3 (Cx46) and alpha8 (Cx50) connexins leads to cataracts by affecting lens inner fiber cells. *Exp.Eye Res., 83,* 688-696.

Yu, X. S., Yin, X., Lafer, E. M., & Jiang, J. X. (2005). Developmental regulation of the direct interaction between the intracellular loop of connexin 45.6 and the C terminus of major intrinsic protein (aquaporin-0). *J.Biol.Chem., 280,* 22081-22090.

Zhang, J. T., Chen, M., Foote, C. I., & Nicholson, B. J. (1996). Membrane integration of in vitro-translated gap junctional proteins: co- and post-translational mechanisms. *Mol.Biol.Cell, 7,* 471-482.

Zheng, J., Wei, C., Zhao, L., Liu, L., Leng, W., Li, W. et al. (2011). Combining blue native polyacrylamide gel electrophoresis with liquid chromatography tandem mass spectrometry as an effective strategy for analyzing potential membrane protein complexes of Mycobacterium bovis bacillus Calmette-Guerin. *BMC Genomics, 12,* 40.

Zhou, X. W., Pfahnl, A., Werner, R., Hudder, A., Llanes, A., Luebke, A. et al. (1997). Identification of a pore lining segment in gap junction hemichannels. *Biophys.J., 72,* 1946-1953.

Zhu, G., Liu, Y., & Shaw, S. (2005). Protein kinase specificity. A strategic collaboration between kinase peptide specificity and substrate recruitment. *Cell Cycle, 4,* 52-56.

Zimmer, D. B., Green, C. R., Evans, W. H., & Gilula, N. B. (1987). Topological analysis of the major protein in isolated intact rat liver gap junctions and gap junction-derived single membrane structures. *J.Biol.Chem., 262,* 7751-7763.