Therapeutic Protein and Glycoprotein Production, Optimization, and Analysis Methods

By Melinda L. Toumi

Ph.D., University of Kansas 2010

Submitted to the Department of Chemistry and the
Faculty of the Graduate School of the University of Kansas
In partial fulfillment of the requirements for the degree of
Doctor of Philosophy

_____

Chairperson

_____

_____

_____

_____

Date defended:_____

The Dissertation Committee for Melinda L. Toumi
certifies that this is the approved version of the following dissertation:

Therapeutic Protein and Glycoprotein Production, Optimization, and Analysis Methods

_____

Chairperson

_____

_____

_____

_____

Date approved:_____

ABSTRACT

Proteins and glycoproteins have the potential to improve health when used as therapeutic products for prevention or treatment of disease, but methods are needed for solving some remaining challenges to production, optimization, and analysis. In this dissertation, novel glycosylation engineering strategies are used to solve challenges in conventional protein production and are used to improve therapeutic protein stability by blocking asparagine deamidation, a ubiquitous cause of protein degradation. Glycan structures on therapeutic glycoproteins must be optimized to avoid negative impacts on pharmacological properties. Methods to optimize glycans are described within, using novel, extracellular glycan trimming reactions performed with glycosidases that can be implemented without harm to protein activity or stability. Finally, analysis of proteins with mass spectrometric peptide mass fingerprinting techniques can be complex, so mass defect filters used for data analysis were improved by determining the correct filter size, using experimental human protein data.

ACKNOWLEDGEMENTS

Thank you to my grandmother, Roberta, for providing genetic proof that personality is not completely forged by nurture. I am glad to be like you.

Please note: all women with children face the same barrier to achieving academic and career goals - the lack of adequate child care in the United States. Adequate means safe and affordable. It is *impossible* to focus all day on solving the world's problems, if you are not sure of the safety and care of your own children. I have been extremely fortunate in that my children were able to attend Hilltop Child Development Center at KU. Pat Pisani, Director, and every one of the early childhood education instructors and staff at Hilltop are the reason that parents like me are able to reach their academic goals. Thank you to my fellow students for paying the student fees that support Hilltop every semester. Thank you to KU's Student Senate for scholarships that assisted with cost. I implore our University to continue supporting Hilltop, so that parent-students are not faced with inadequate child care, or worse – that women are not relegated to the days of being forced to choose between being educated and having a family. Thank you to the teachers and staff at Hillcrest Elementary, who have supported my children as people, while educating them.

Thank you to the scientists who came before me, especially those who were persecuted for their ideas, those whose credit was due but never arrived, and those whose dedication to science left for little else in their lives.

Science is a love in its own right.

TABLE OF CONTENTS   PAGE

0

CHAPTER 1 INTRODUCTION

Proteins and glycoproteins are used as the active therapeutic agent for treating or preventing many different diseases, and despite advances in therapeutic product development, challenges remain to be addressed. Areas in the field that will benefit from demonstration of new methods include protein production, pharmacological property optimization, and protein analysis. The contiguous pages encompass an introduction to protein and glycoprotein therapeutic composition, describe some examples of Food and Drug Administration approved therapeutic proteins, and introduce recombinant technology and lectin affinity chromatography techniques used to produce and purify the products. Then, analytical methods that are used to determine the composition and structure of therapeutic proteins are discussed. Finally, subsequent chapters present work that is intended for making further improvements to protein and glycoprotein therapeutic products.

1.1 PROTEIN AND GLYCOPROTEIN THERAPEUTICS

1.1.1 Overview of therapeutic proteins, glycoproteins, and glycosylation

Proteins may be developed into a therapeutic product, after the specific function or effect of the protein is identified. Proteins consist of an amino acid sequence that is encoded by DNA, which is used to express the protein using recombinant technology,[1,2] discussed below. The protein is produced and purified in soluble, folded, and active form, and must be optimized with regard to a range of pharmacological properties.[1] To

be a high quality therapeutic product, the recombinant protein must, for example, contain the correct post-translational modifications (PTMs), such as the glycosylation structures that are known to provide optimal therapeutic effects.[3-5]

N-linked glycans are appended to a protein when the amino acid consensus sequence (N-X-S/T) is present in the protein primary structure, where X is any amino acid other than proline.[3,4] The type of N-linked glycan that will be present at a particular glycosylation site is determined by the cofactors, glycosyltransferases, and glycosidases located in the rough endoplasmic reticulum (RER) and Golgi bodies that comprise the cellular expression system.[3-5] The glycan composition is dependent upon cellular environment is very different from the direct genetic encoding that occurs for the amino acid sequence.[3-5] While protein synthesis is template driven, glycosylation is not. Instead, glycosylation on a protein depends on the activity of the glycosyltransferases and glycosidases (collectively the glycoenzymes) and availability of cofactors; these are a direct consequence of the local environment within the cell, which can vary significantly.[8]

Therapeutic glycoproteins are produced with the goal of including glycan moieties that are similar to naturally occurring human glycans.[4-7] The two most common forms of glycosylation in human proteins, N- and O- linked, both consist of the same individual carbohydrate building blocks, but are attached to the protein sequence either through the side chain of an asparagine residue (N-linked glycans), or to the hydroxyl group of serine or threonine (O-linked glycans), respectively.[3,4] N-linked glycans have been studied extensively, and are known to exist as one of three main types as described in **Figure 1.**

2
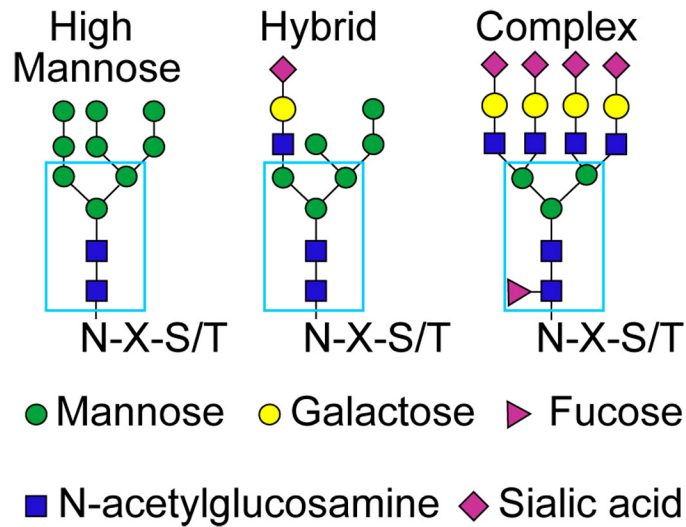
Figure 1. Three main classes of *N*-linked glycans: high mannose, hybrid, and complex. All three share the same pentasaccharide core structure, then terminate with different carbohydrate residues.

The *N*-linked glycan types, high mannose, hybrid, and complex, all share the same pentasaccharide core, as shown in the figure. The classification depicted in Figure 1 is determined by the identity of the terminal carbohydrate residues, which are acquired during further enzymatic processing in the Golgi body of a cell.[4,7] Terminal residues generally consist of only mannose residues in high mannose type, galactose or sialic acids in complex type. In a hybrid type structure, at least one branch terminates in high mannose and one branch resembles complex type glycans.[4] Figure 1 depicts triantennary and tetrantennary branching; biantennary glycans have only two branches (not shown).

Individual carbohydrate residues are linked together in various linkage types, α or β, and the linkages are referred to by the carbon number assigned around the carbohydrate ring structure.[8] In **Figure 2,** the top example shows two carbohydrate residues linked with a β(1,4) bond, and the bottom shows an α(1,2) linkage. One glycan residue can also be attached to another at more than one position, and with more than

3

one linkage type.[4,5,8] For example, sialic acid residues are often linked to galactose in the following configurations: α(2,3), α(2,6), or α(2,8). Thus, a high degree of heterogeneity may exist within the *N*-linked glycans located on a particular protein, including major glycoform structures, variable linkages, and individual structural isoforms.[4,6,9]
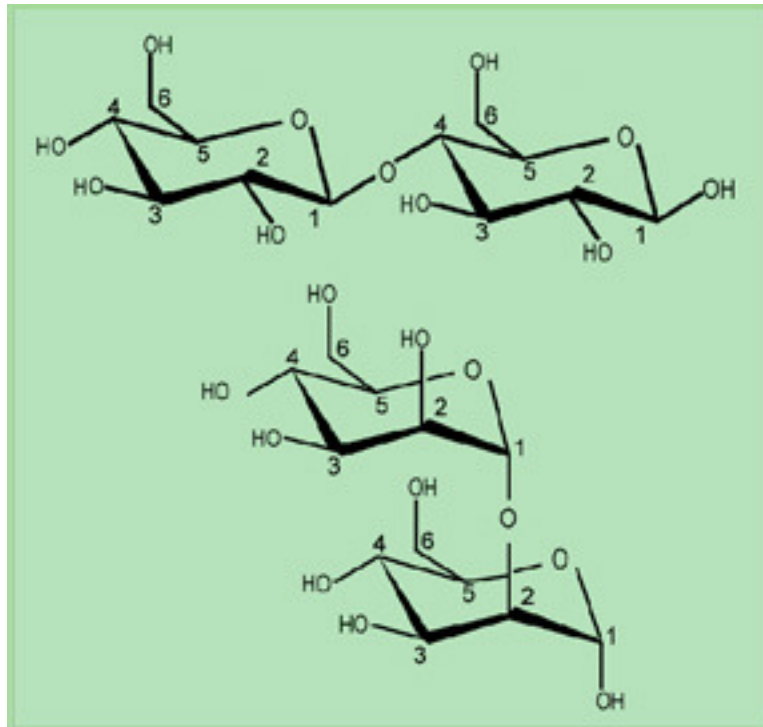
Figure 2. Carbohydrate linkage configurations. The top example shows two carbohydrate residues linked together with a β(1,4) linkage and the bottom sugars are linked with an α(1,2) linkage.

N-linked glycan variation can be problematic, because glycan composition affects a wide range of pharmacological properties, such as circulation half life, immunogenicity, receptor binding, thermal stability, folding, charge, and product potency.[5,7-8] ***These observed effects have shown that glycosylation of therapeutic glycoproteins must be in the correct location, and of the correct type, for the production of an effective and safe therapeutic product.***[5,8-12] The glycan post-

translational modifications are acquired during protein production within a chosen host cell, and are dependant upon host cell choice, which is discussed in section 1.2.1. Methods to optimize glycan composition are presented in Chapters 3 and 4.

1.1.2 Therapeutic protein and glycoprotein examples

The earliest use of proteins as therapeutics involves the purification of proteins from harvested tissues and organs. For example, the first use of the protein hormone, insulin, was demonstrated using purified insulin from dog pancreas, and the purified pancreatic insulin became available all over the Western world by the 1920's for the treatment of diabetes.[13] Just a few decades later, isolation of human growth hormone from human cadaver pituitary glands was performed for the treatment of growth hormone deficiency.[14,15] Today, therapeutic protein production is performed using recombinant DNA technology, which is discussed below in section 1.2.1. Human growth hormone, and erythropoietin, another protein therapeutic produced by recombinant DNA technology, are the key proteins studied in this dissertation, so more detail is included about these proteins.

Human growth hormone (hGH) is an endogenous 22 kDa protein that is critical to proper growth and metabolism.[16] The hGH protein is nonglycosylated in the most abundant form found in the pancreas, even though other tissues, such as the placenta, have been found to have glycosylated hGH.[17] For the children and adults who suffer from hGH deficiencies, the recombinantly expressed hGH (rhGH) has become an important therapeutic.[18-20] Growth hormone is discussed further in Chapters 2 and 5. Both insulin and growth hormone proteins do not need glycosylation to have the

intended therapeutic effect in the body, and this lack of required glycosylation allows for production in protein expression hosts such as bacteria.[21] Unfortunately, therapeutic glycoproteins do not function without the proper glycosylation moieties, as is the case for erythropoietin, and other expression host cells are required.

Recombinant, human erythropoietin (rhEPO) is a glycoprotein therapeutic that was developed after the endogenous glycoprotein hormone erythropoietin was determined to be required for red blood cell production. Endogenous EPO is produced in the kidney and circulates in serum to the bone marrow, where it differentiates erythrocyte progenitor cells for the production of red blood cells.[22] EPO naturally possesses three sialylated, complex type *N*-linked glycans, which are essential to erythropoietin's *in vivo* activity.[23-26] The complex type *N*-linked glycans of EPO have interesting effects on receptor binding. The receptor was determined to have an increased binding affinity constant for EPO after removal of the terminal sialic acid carbohydrate residues from the *N*-linked glycans.[26-28] Without the sialic acids, however, the exposed galactose residues are bound by the liver's galactose binding proteins, and EPO is cleared from circulation quickly thus non-sialylated EPO has no appreciable *in vivo* activity.[23-29] Therefore, the requirement for sialylated, complex type *N*-linked glycans must be met to produce an effective glycoprotein therapeutic, and can be accomplished by carefully choosing the protein expression host. Host selection is discussed below in section 1.2.2. New methods to further optimize the N-linked glycans of erythropoietin are described in Chapter 4.

1.2 RECOMBINANT PROTEIN AND GLYCOPROTEIN PRODUCTION FOR THERAPEUTIC USE

Therapeutic protein production was first accomplished by extracting and purifying protein hormones from tissue, such as the early use of human growth hormone and insulin.[13-15] Production of protein and glycoprotein therapeutics is now achieved through the use of recombinant DNA technology.

1.2.1 Protein expression using recombinant technology

To produce, or express, a protein using recombinant DNA, a series of steps are required. See **Figure 3** below.
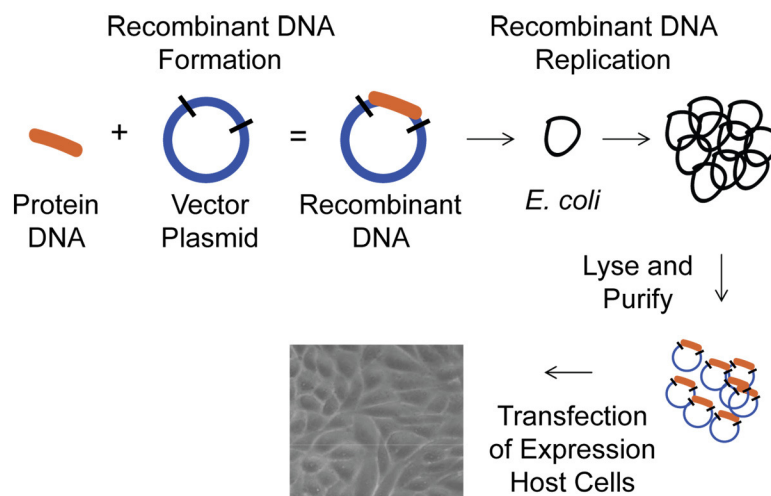
Figure 3. Protein expression work-flow diagram. Recombinant DNA technology is used to express proteins for therapeutic use.

The first step in protein production is to obtain the DNA for the protein of interest. A DNA sequence is composed of the four nucleotide bases, G, C, A, and T, covalently bound by phosphodiester bonds, arranged into groups of three bases, called codons.

These codons encode for individual amino acids; and the codons are arranged in the order necessary to generate the amino acid sequence of the desired protein.[2,30-31] The DNA sequence that encodes the protein, called a gene, is combined with a larger piece of DNA, that will act as a vector to introduce the protein gene into the chosen expression host,[2] is shown in Figure 3. The combined DNA is the "recombinant" DNA.[2] Recombinant DNA used in protein expression is usually double stranded and circular in shape and called a "plasmid". It is not a part of the host cell genetic material.[30-31]

The next step, as shown in Figure 3 is to use the vector plasmid, containing the gene for the target protein, to transform *Escherichia coli (E.coli).* The *E. coli* replicate the plasmid to produce large amounts of the recombinant DNA, provided the plasmid contains the necessary DNA sequences (such as promoter regions) that act as signals to the *E. coli*, to produce the product.[2,30-31] Finally, the *E. coli* can be lysed to purify the DNA, and the DNA can be introduced into the chosen protein expression host cell.[2,30-31] The host transcribes the protein DNA into mRNA, and then the mRNA is translated into the protein of interest.[2,30-31] The choice of expression host affects glycosylation, and this topic is discussed *vida infra*.[5,8-12]

E. coli and baculovirus-directed insect cell lines have been used extensively as protein expression hosts,[31-32] but most of the proteins are expressed in aggregate form, and they must be dissociated using a series of trial and error experiments that are designed to extract, solubilize, and refold the protein.[32-35] In addition, these expression hosts do not provide the type of glycosylation that is required for therapeutic glycoproteins such as erythropoietin that require sialylated, complex type *N*-linked glycans for activity.[23-29,36] Thus, *E. coli* and insect cell lines are generally useful for non-

glycosylated protein production only. See **Table 1** for a partial list of host cell lines that are used in protein production.

Table 1. Protein expression host cell options.[21,36-41]

Host	Examples
Bacteria	*Escherichia coli*
Insect	*Drosophila melanogaster*
Yeast	*Saccharomyces cerevisiae*
	Schizosaccharomyces pombe
	Pichia pastoris
Plant	*Nicotiana benthamiana*
Mammal	T293-K (human)
	Baby Hamster Kidney (BHK)
	NS0 (mouse)
	Chinese Hamster Ovary (CHO)

All of the protein expression host cell lines in Table 1 have been tested for therapeutic protein or glycoprotein production, with varying degrees of success. Attempts at producing human-like glycosylation patterns for antibodies and other therapeutic glycoproteins using plant cells as an expression host have produced glycan moieties that still contain some level of sugars that are specific to plant cells, such as the β(1,2)-xlyose and α(1,3)-fucose.[37] These sugar residues are immunogenic to humans.[6,37,40] Therapeutic proteins produced using *E. coli* must also be carefully evaluated for dangerous endotoxins and high levels of acetate.[42-43] Yeast strains are naturally capable of producing proteins with only high mannose type *N*-linked glycans.[44-46] Even after extensive genetic alterations to yeast strains such as *Pichia pastoris*, the

glycans contain only biantennary complex type that still lack higher degrees of branching observed in many human glycan structures.[46] For these reasons, most therapeutic glycoproteins are produced in mammalian cell lines, such as Chinese hamster ovary (CHO) or baby hamster kidney (BHK) cells.[39,40,47,48]

Even among available mammalian host cell lines such as CHO or BHK, considerable variation can be found in the PTMs for glycoprotein therapeutics. It is important to reiterate that the *N*-linked glycans of a protein cannot be directly controlled in a manner similar to the primary amino acid sequence of the protein, so optimization of cell culture conditions, such as choosing a desirable cell line and media formulation, to promote the desired glycosylation pattern, is an important focus of research in the therapeutic product industry.[5,8-12,36-41,47-49]

One example of a case where glycan optimization is important is in manufacturing erythropoietin. Erythropoietin requires sialylated, complex type *N*-linked glycans to produce any therapeutic effect, but care must be taken in optimizing the sialylation on the protein because some forms of sialic acid can create potentially dangerous immune responses. *N*-glycolylneuraminic acid (Neu5Gc), a derivative of sialic acid, can be added to the N-linked glycan, and can produce an immunogenic response in humans.[49] **Figure 4** provides the chemical structures of two sialic acids, which are structurally very similar and differ only by a single hydroxyl group, indicated in a red circle. While the structural difference is small, these two compounds are clinically distinct, where Neu5Gc can induce an adverse response, but Neu5Ac does not.
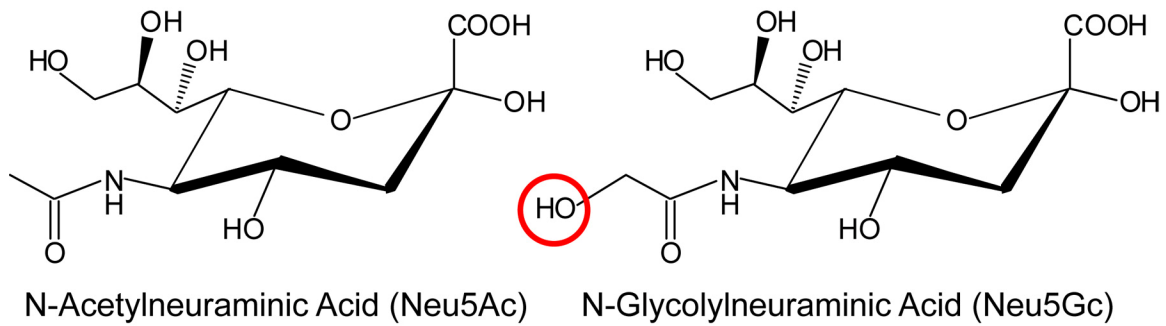
Figure 4. Chemical structure of the main sialic acid, N-Acetylneuraminic Acid (Neu5Ac), and the structurally similar, but immunogenic, derivative, N-Glycolylneuraminic Acid (Neu5Gc).

The Neu5Gc derivative is prevalent in mouse cell lines like NS0, accounting for, up to 50% of sialic acid species on N-linked glycans of recombinant proteins.[41,47] CHO cell expressed glycoproteins contain significantly less Neu5Gc, estimated to be 7-15% of sialic acid content in some studies, depending on the cell culture conditions.[47] At this time, 70% of approved glycoprotein therapeutic products are expressed in a CHO cell line due to reasons such as glycan composition,[48] making expression host cell line an important part of therapeutic production. While CHO cells diminish the incorporation of Neu5Gc into erythropoietin, they do not completely suppress its presence. Therefore, better strategies are needed to avoid this residue, in order to maximize the safety and potential for long-term use of this drug. A novel strategy for removing sugar residues that prove to be problematic to therapeutic glycoproteins is discussed in Chapters 3 and 4.

After a nonglycosylated therapeutic protein is expressed, the protein must then be purified. A wide range of currently employed affinity tags used for protein purification are available, but each type of tag can be disadvantageous to protein structure or function, depending on the application.[50] An alternative protein production and

purification platform is described in Chapter 2, for use with nonglycosylated proteins, which account for approximately 50% of human proteins.[51] The other 50% of human proteins are known to be glycosylated, and the glycans can serve as a means for purification, using lectin affinity. Glycoprotein purification with lectin affinity is discussed in section 1.2.2.
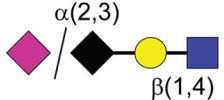
1.2.2 Purification using lectin affinity

A lectin is a protein that binds to carbohydrates.[52-54] In nature, lectins bind to sugar residues on the surface of cells, and lectins are directly involved in the processes that some pathogens, such as influenza and tuberculosis, use to invade and infect hosts.[52-54] The exploitation of the natural affinity of lectins has allowed for lectins to be used in many applications, from fabrication of protein thin films to glycoprotein purification.[55,56] Erythropoietin and other glycoprotein therapeutics are often purified using lectin affinity.[56]

Lectins can be bound to a support resin and loaded into a cartridge or column, then used as part of a purification strategy using methodology similar to other affinity chromatography techniques.[56-58] Briefly, the glycoprotein of interest is concentrated, mixed with a loading buffer, and applied to the column. The presence of carbohydrates that are recognized and bound by the lectin allow for the glycoprotein of interest to be retained, while other species are rinsed away. The glycoprotein is then eluted, often through displacement. For example, a solution containing a high concentration of a carbohydrate, for which the lectin will have greater affinity for than the glycan moiety of

the glycoprotein, is introduced so that bound glycoproteins are eluted due to displacement by the elution buffer's carbohydrate.[57-58]

The physical basis for a lectin to have affinity for a glycan is primarily through hydrogen bonding, where the negatively charged amino acids (D and E) of the lectin form hydrogen bonds with the hydroxyl groups of sugar residues. Alternatively, hydrogen bonds can form between the amide groups of amino acids N or Q, of the lectin, and the sugar hydroxyl groups.[59] Lectins with affinity for charged carbohydrate residues, such as sialic acids, also rely primarily on hydrogen bonding, where the uncharged, polar amino acids can have affinity for the negatively charged sialic acids.[59] Lectins that have been identified as having specificity for sialic acids and other complex type *N*-linked glycan terminal residues are displayed in **Figure 5** below**.**

Lectin	Sugar Residue and Linkage
Wheat germ (WGA)	β(1,4) ■—■ and all sialic acids ◆
Sambucus nigra (SNA)	α(2,6) ◆—○
Maackia amurensis (MAL)	α(2,3) ◆ / ◆—○—■ β(1,4)

Legend:
○ Galactose   ■ N-acetylglucosamine
◆ Sialic acid (Neu5Ac)   ◆ Sialic acid (Neu5Gc)

Figure 5. Lectins with affinity for complex type N-linked glycans.[52,59-61]

The plant lectins described in the first column in Figure 5 can be useful in therapeutic glycoprotein purification strategies,[56-59] since complex type *N*-linked glycans are a necessary component of many therapeutic glycoproteins. These particular plant

lectins are commercially available. Use of lectin affinity chromatography is demonstrated in Chapters 2, 4, and 5. After therapeutic proteins are purified, they need to be characterized and studied using many different analytical techniques. Mass spectrometry and circular dichroism spectroscopy are two such techniques, and are discussed below in section 1.2.3.

1.2.3 Analytical chemistry in protein and glycoprotein analysis

Potential protein and glycoprotein therapeutic products must be well characterized with respect to structure and function long before they can be developed into a viable treatment product. Mass spectrometry (MS) is an analytical method that is capable of identifying the composition of a protein and its glycans with a high degree of confidence. Techniques for protein analysis using MS are discussed in section 1.2.3.1. In addition to obtaining molecular composition information, which is afforded by mass spectrometry, the structure of a protein must be monitored, as structure and function are closely linked. A low-resolution, time sensitive spectroscopic technique such as circular dichroism (CD) is very useful in analyzing a protein for changes in structural characteristics, for monitoring the protein for changes to structure throughout an optimization process, and even for assessing major difference in thermal stability between two protein samples. CD is discussed below in section 1.2.3.2.

1.2.3.1 Mass spectrometry

Mass spectrometric techniques for protein and glycoprotein analysis rely on instrumentation that possesses three distinct parts. See **Figure 6.**
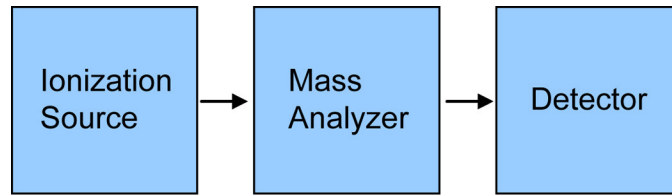
Figure 6. Mass spectrometers consist of an ionization source, mass analyzer, and a detector

For mass spectrometric analysis to occur, the analyte of interest must become a gas phase ion in the ionization source, and can possess a positive or negative charge.[62-65] The two most common ionization methods used for protein analysis are electrospray ionization (ESI) and matrix assisted laser desorption ionization (MALDI).[64] Both of these ionization methods are considered non-destructive, or "soft", because they do not produce fragments of the analyte during the ionization process.[64]

After the ions are formed, they are sent to the mass analyzer, where they are separated. The transport and separation is performed under vacuum, and this section of the instrument is where *intentional* fragmentation of the sample ions may be accomplished.[62-65] Fragmentation can be achieved using collision induced dissociation (CID). In this case, an ion is isolated and then fragmented via collisions with atoms of an inert gas, such as helium or argon.[62-65] These techniques are very valuable because the fragmentation patterns that are generated can provide significant information regarding the analyte's composition. This method is particularly useful for confirming the composition of the glycan structures on glycopeptides.[66-70] CID is used in Chapter 4 to elucidate glycan composition of a therapeutic glycoprotein. The observed fragmentation can be used to confirm or support the composition assignments made based on the

data in a full mass spectrum, where the masses of the intact analytes can be measured with a high degree of accuracy.

In the detector section of the instrument, the ions, which have been previously separated by their mass to charge ratios, are counted. The resulting ion signal is recorded, and the information is used to generate a spectrum of peaks that describe the ions present, in terms of their ratio of mass to charge (m/z).[65]

The type of information that is obtainable from mass spectrometric analyses, such as glycan mass, peptide composition, or glycosylation site location and glycan composition, are dependent upon the procedures used to prepare the sample. Strategies used to prepare proteins and glycoproteins for mass spectrometric analysis are illustrated in **Figure 7**.

To choose the best preparative strategy prior to MS analysis, one must first identify the type of information that is desired. To determine the identity of *N*-glycans of a glycoprotein, the glycans could be enzymatically released from the protein using an endoglycosidase such as Peptide-N-glycosidase F (PNGase F), and then analyzed using mass spectrometry.[71-73] When only a few glycans are present on a small glycoprotein, the whole glycoprotein can be analyzed intact. This approach is used to monitor glycan modification reactions in Chapter 3.

For larger proteins, or proteins with multiple glycosylation sites, intact analysis is often not a viable approach for obtaining information about the glycosylation. In these cases, another approach is often used for compositional analysis. This approach (often called "glycopeptide analysis") relies on proteolytic digestion, which cleaves the protein but leaves the glycans still attached to the peptides. This sample preparation step

results in a mixture of peptides and glycopeptides.[66-70] While the resulting mixture can be quite complex to analyze, this strategy provides information that other approaches cannot: specifically information about which glycans are attached to which particular glycosylation sites is obtained. This technique, which produces "glycosylation site specific" information, can even be used to identify multiple glycoforms attached to the protein at a single site.[66-70] Glycopeptide analysis may prove especially beneficial in the therapeutic development work-flow, since this technique provides value-added information, compared to other methods of glycan or peptide analysis.
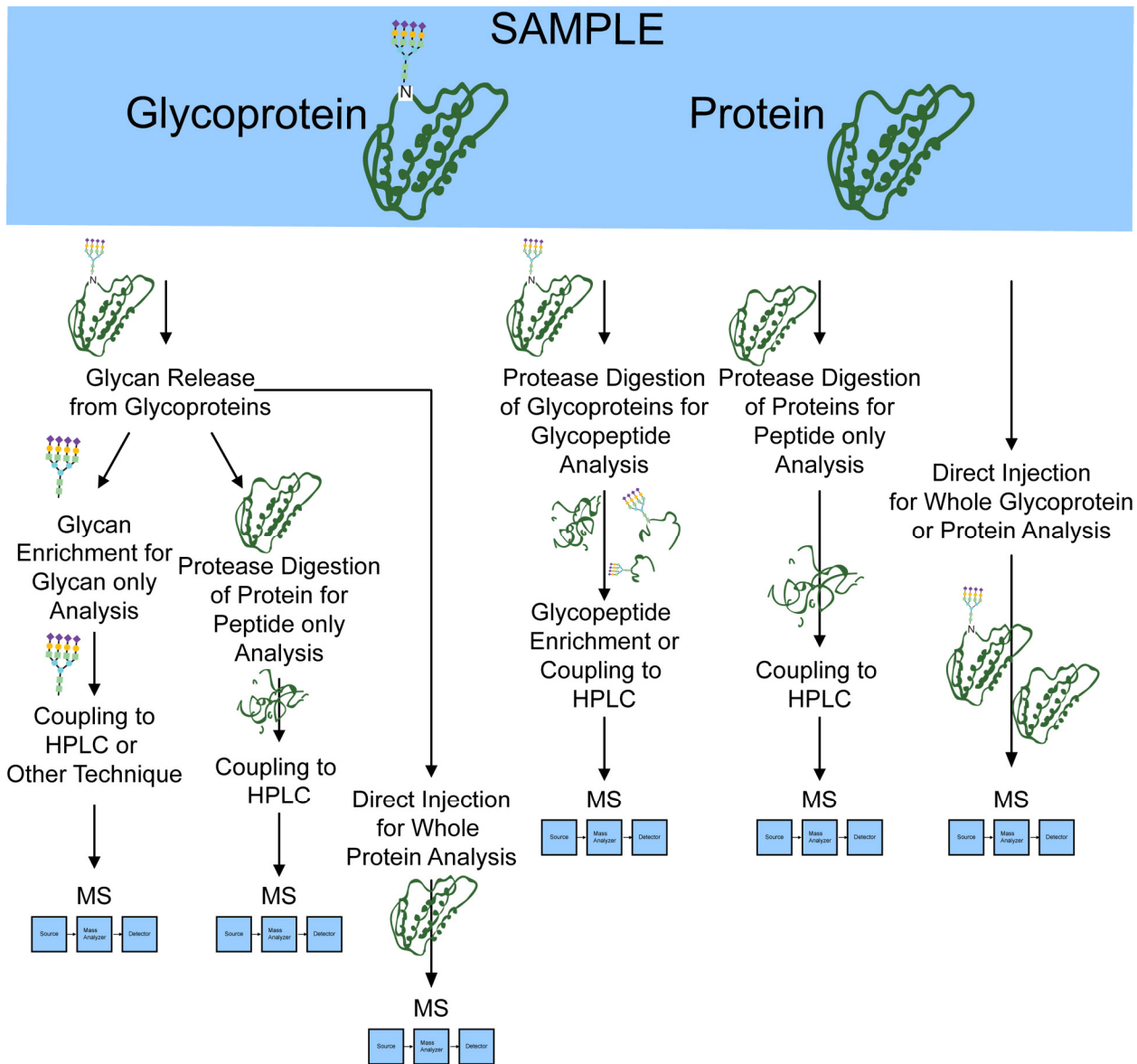
Figure 7. Strategies for protein and glycoprotein analysis using mass spectrometry.

After the MS experiment has been completed, data analysis begins, which is sometimes extensive, complicated, and time consuming. Peptide data analysis is often simplified by filtering the data to remove peaks that are unlikely to be identified as a peptide, based on residual mass, or mass defect.[71,74] Then the data is submitted to an analysis program such as Mascot.[75,76] Currently employed data filters used in peptide mass fingerprinting experiments are based on values that were calculated using a set of theoretical peptides.[75,76] In Chapter 6, calculations from human protein data sets are presented to improve selectivity of the mass defect data filters. Mass spectrometric analysis of glycoproteins can also be aided during data analysis, using an analysis tool such as GlycoPepDB and GlycoPepID.[67-68]

In addition to mass spectrometric analysis, a structural analysis method is needed for use when developing therapeutic proteins and glycoproteins. Circular dichroism is one commonly used method, and it is discussed below.

1.2.3.2 Circular dichroism

Circular dichroism (CD) is a spectroscopic technique that can be used for low resolution assessment of protein or glycoprotein structure. CD cannot discern protein structure to the near Ångström level, as can techniques like nuclear magnetic resonance (NMR) and X-ray crystallography, but some agreement has been shown between NMR and CD data results.[77,78] Instead of high resolution structural data, CD is used for its time sensitivity, where data may take 10 minutes to collect and analyze, while protein NMR and crystallography data collection may take days or weeks. Data analysis of CD data is often completed in a matter of minutes, whereas NMR and x-ray

data can take months to analyze. CD spectroscopy, when performed using far UV wavelengths, is used to estimate protein secondary structure, to monitor changes in structure that may occur due to processing, or to monitor the protein for loss of structure in response to thermal stress.[79,80] CD can be performed using other experimental parameters to gain different types of protein information. If one uses wavelengths in the near UV region information on tertiary structure can be obtained, based on aromatic amino acid absorption.[81,82] The experiments described in Chapters 2-5 use CD with far UV wavelengths; therefore, far UV wavelength CD experiments are discussed below.

CD is performed by subjecting a sample to far UV wavelengths (190-250 nm), and collecting a signal that is generated from the difference in absorption between the left and right circularly polarized light.[80] In the far UV region, the absorption is primarily due to the electronic transitions on the protein backbone, where π to π* and n to π* transitions are thought to occur along the amide groups of the protein.[77,81,83] Differences in secondary structural characteristics of peptides and proteins can be studied using far UV CD, because changes in structure cause changes in the wavelength of light absorbed. The electronic transitions that can occur in an alpha helix produce a signal at approximately 222 and 209 nm; beta sheets absorb at 218 nm, and less ordered structure, around 195 nm.[77] These signals do overlap[77], but data analysis algorithms used to deconvolute the signals are available. Different algorithms produce different results, because they are based on different sets of protein structure calculations. Also, they use different basis sets, often consisting of protein structure determined by NMR or X-ray crystal data.[84-88] For these reasons, it is recommended that algorithms and basis

sets are carefully chosen based on similarities to the protein or peptide being analyzed.[86-88]

Although CD cannot determine protein secondary structure with high precision, the technique can be especially useful for monitoring a protein's loss in secondary structure in response to thermal stress; such a change is useful to monitor because the temperature at which the protein loses its secondary structure, its T_m, can be used to describe a protein's stability.[80] Experiments are performed by increasing the temperature of the sample, and scanning the sample at a single wavelength, and the change in signal is monitored. Using a two state protein folding assumption, the protein in solution will transition from being in the folded state, to the unfolded state, and the midpoint between the two is calculated to be the melt temperature (T_m) in °C.[89,90] CD is used in this way in Chapters 2, 4, and 5, to compare thermal stabilities of native and modified therapeutic protein samples.

1.3 SUMMARY OF SUBSEQUENT CHAPTERS

Chapter 2 introduces a new concept for protein production and purification. In this new method, a protein that is not natively glycosylated can be produced in soluble and folded form, by modifying the protein DNA to include a non-native, N-linked glycosylation consensus sequence: N – X – S/T. The modified DNA is paired with an expression host that recognizes the signal for glycosylation, and the host glycosylates the protein. The non-native glycan structure is used as a temporary protein purification tag using lectin affinity purification. Enzymatic cleavage of the entire glycan may then be performed to completely remove the purification tag. The tagging and removal process has the advantage over currently used protein purification tags, because N-linked glycan removal does not require multiple non-native amino acid residues to be part of the final protein product. Two unique mutant human growth hormone proteins were designed, expressed, and studied to illustrate this concept. This method will mitigate commonly encountered difficulties during protein production, such as producing misfolded protein and the need to extract proteins from aggresomes. Analysis by circular dichroism indicates the method can be used to express and purify large amounts of properly folded protein.

Chapter 3 describes the use of glycosidases to modify the glycan moieties of glycoproteins, a method referred to as "glycan trimming". Removal of either an entire glycan or removal of only specific sugar residues may prove useful when optimizing glycan properties of therapeutic glycoproteins. A model protein, Ribonuclease B, was used to demonstrate that trimming of the glycan structure does not destroy the

enzymatic activity of the protein, nor does it destroy the secondary structure of the protein. A novel enzymatic activity assay using Atmospheric Pressure-Matrix Assisted Laser Desorption Ionization Mass Spectrometry (AP-MALDI-MS) to detect the formation of products produced by the assay is described and is used to show the enzyme retained its activity after glycan trimming. These studies indicate glycan trimming may be broadly implemented without destruction of protein function and secondary structure.

Chapter 4 presents the application of glycan trimming methods introduced in Chapter 3, to glycoproteins containing sialic acid residues. A sialylated, therapeutic glycoprotein, erythropoietin, was expressed, purified, and then modified using glycan trimming to remove sialic acid residues. Removal of sialic residues that may be immunogenic will prove beneficial in the production of therapeutic glycoproteins like erythropoietin, and the studies presented in Chapter 4 show that sialic acid removal can be achieved using glycan trimming. Furthermore, the data indicate that the glycan trimming method can be used without decreasing erythropoietin's thermal stability. Three additional, commercially available, sialylated glycoproteins were also subjected to the same glycan trimming and stability studies to illustrate the procedure can be broadly applied to sialylated glycoproteins without negatively impacting the thermal stability of the protein. The data clearly indicate that the increase in melt temperature observed for erythropoietin post glycan-trimming is not just a feature unique to this protein.

Chapter 5 highlights the use of glycan trimming as one key component of a new method designed to prevent a common pathway of protein degradation: asparagine

deamidation. Deamidation is a major cause of aggregation in some vulnerable

therapeutic proteins, such as human growth hormone. Three different human growth

hormone variants were designed, expressed, and purified using methods similar to

those described in Chapter 2, where non-native glycosylation was introduced using site

directed mutagenesis. The glycosylation sites were placed at asparagine residues of

growth hormone that are known to deamidate; the glycans provide steric bulk, which

hinders the deamidation reaction. Secondary structure analysis of the mutant proteins

indicated that the bulky glycan was negatively impacting the protein structure, prompting

the use of enzymatic glycan trimming to trim the majority of the glycan from the protein.

The resulting protein was found to have improved secondary structure after glycan

trimming. The mutant proteins with trimmed glycans displayed improved thermal

stability compared to a growth hormone standard.

Chapter 6 addresses a protein analysis problem encountered during data

analysis of peptide mass fingerprinting experiments. Mass defect, the difference

between nominal and monoisotopic mass, has been used to identify peptide peaks in

mass spectra, when the mass defect value is within the range of values identified for

peptides. The range can be used to filter the data, where any value outside of the

defined range is rejected, since it does not have the correct mass defect to be a

peptide. The range of possible mass defect values used in peptide data filters was

calculated previously using a set of theoretical peptides, where every amino acid was

used 5% of the time at random. To make mass defect filters more selective for real

human peptides, two different sets of human proteins were digested *in silico*, and mass

defect values were calculated. The range determined using the human biofluid data was found to be narrower than the range of values previously established in the proteomic field. Using human protein data will increase the selectivity of mass defect peptide data filters used in proteomic data analysis.

1.4 REFERENCES

1.   Walsh, G. Biopharmaceuticals: recent approvals and likely directions. *Trends Biotechnol.* **2005,** *23,* 553-558.

2.   Wong, D. W. S. *The ABCs of Gene Cloning;* Chapman & Hall: New York, 1997.

3.   Jensen, P. H.; Kolarich, D.; Packer, N. H. Mucin-type O-glycosylation – putting the pieces together. *FEBS J.* **2010,** *277,* 81-94.

4.   Czlapinski, J. L.; Bertozzi, C. R. Synthetic glycobiology: Exploits in the Golgi compartment. *Curr. Opin. Chem. Biol.* **2006,** *10,* 645-651.

5.   Li, H.; d'Anjou, M. Pharmacological significance of glycosylation in therapeutic proteins. *Curr. Opin. Biotechnol.* ***2009,*** *20,* 678-684.

6.   Ohtsubo, K.; Marth, J. D. Glycosylation in cellular mechanisms of health and disease. *Cell.* **2006,** *126,* 855-867.

7.   Tabbarra, I. A. Erythropoietin. Biology and clinical applications. *Arch. Int. Med.* **1993,** *153,* 298-304.

8.   Walsh, G.; Jefferis, R. Post-translational modification in the context of therapeutic proteins. *Nat. Biotechnol.* **2006,** *24,* 1241-1252.

9.   Jenkins, N.; Parehk, R. B.; James, D. C. Getting the glycosylation right: Implications for the biotechnology industry. *Nat. Biotechnol.* **1996,** *14,* 975-981.

10.   Hossler, P.; Khattak, S. F.; Li, S. J. Optimal and consistent protein glycosylation in mammalian cell culture. *Glycobiology.* **2009,** *19,* 936-949.

11.   Beck, A.; Wagner-Rousset, E.; Bussat, M.-C.; Lokteff, M.; Klinguer-Hamour, C.; Haeuw, J.-F.; Goetsch, L.; Wurch, T.; Dorsselaer, A. V.; Corvaia, N. Trends in glycosylation, glycoanalysis and glycoengineering of therapeutic antibodies and Fc-fusion proteins. *Curr. Pharm. Biotechnol.* **2008,** *9,* 482-501.

12.   Anthony, R. M.; Nimmerjahn, F.; Ashline, D. J.; Reinhold, V. N.; Paulson, J. C.; Ravetch, J. V. Recapitulation of IVIG anti-inflammatory activity with a recombinant IgG Fc. *Science.* **2008,** *320,* 373-376.

13.   Banting, F. G.; Campbell, W. R.; Fletcher, A. A. Further clinical experience with insulin (pancreatic extracts) in the treatment of diabetes mellitus. *British Medical Journal.* **1923,** *Part 1,* 8-12.

14. du Bois, M. B.; Rappaport, R.; Royer, P. Difficultés liées à l'utilisation de la somathormone humaine du type Raben, et résultats préliminaires d'une nouvelle technique d'extraction. *Archives françaises de pédiatrie.* **1971,** *28,* 673.

15. Soyka, L. F.; Ziskind, A.; Crawford, J. D. Treatment of short stature in children and adolescents with human pituitary growth hormone (Raben). Experiences with thirty-five cases. *N. Engl. J. Med.* **1964,** *271,* 754-764.

16. Patrick, G. L. *An Introduction to Medicinal Chemistry,* 2nd ed.; Oxford University Press: New York, 2001.

17. Bustamante, J. J.; Gonzalez, L.; Carroll, C. A.; Weintraub, S. T.; Aguilar, R. M.; Munoz, J.; Martinez, A. O.; Haro, L. S. *O*-Glycosylated 24 kDa human growth hormone has a mucin-like biantennary disialylated tetrasaccharide attached at Thr-60. *Proteomics.* **2009,** *9,* 3474-3488.

18. Laron, Z. Short stature due to genetic defects affecting growth hormone activity. *N. Engl. J. Med.* **1996,** *334,* 463-466.

19. Takahashi, Y.; Kaji, H.; Okimura, Y.; Goji, K.; Abe, H.; Chihara, K. Brief report: Short stature caused by a mutant growth hormone. *N. Engl. J. Med.* **1996,** *334,* 432-436.

20. Salomon, F.; Cuneo, R. C.; Hesp, R.; Sonksen, P. H. The effects of treatment with recombinant human growth hormone on body composition and metabolism in adults with growth hormone deficiency. *N. Engl. J. Med.* **1989,** *321,* 1797-1803.

21. Gerngross, T. U. Advances in the production of human therapeutic proteins in yeasts and filamentous fungi. *Nat. Biotechnol.* **2004,** *22,* 1409-1414.

22. Graber, S. E.; Krantz, S. B. Erythropoietin and the control of red cell production. *Ann. Rev. Med.* **1978,** *29*, 51-66.

23. Dordal, M. S.; Wang, F. F.; Goldwasser, E. The role of carbohydrate in erythropoietin action. *Endocrinology.* **1985,** *116*, 2293-2299.

24. Takeuchi, M.; Takasaki, S.; Shimada, M.; Kobata, A. Role of sugar chains in the *in vitro* biological activity of human erythropoietin produced in recombinant Chinese hamster ovary cells. *J. Biol. Chem.* **1990,** *265*, 12127-12130.

25. Goldwasser, E.; Kung, C. K. H.; Eliason, J. On the mechanism of erythropoietin-induced differentiation. XIII. The role of sialic acid in erythropoietin action. *J. Biol. Chem.* **1974,** *249*, 4202-4206.

26. Tsuda, E.; Kawanishi, G.; Ueda, M.; Masuda, S.; Sasaki, R. The role of carbohydrate in recombinant human erythropoietin. *Eur. J. Biochem.* **1990,** *188*, 405-411.

27. Imai, N.; Higuchi, M.; Kawamura, A.; Tomonoh, K.; Oh-eda, M.; Fujiwara, M.; Shimonaka, Y.; Ochi, N. Physicochemical and biological characterization of asialoerythropoietin. Suppressive effects of sialic acid in the expression of biological activity of human erythropoietin *in vitro. Eur. J. Biochem.***1990,** *194*, 457-462.

28. Narhi, L. O.; Arakawa, T.; Aoki, K. H.; Elmore, R.; Rohde, M. F.; Boone, T.; Strickland, T. W. The effect of carbohydrate on the structure and stability of erythropoieitin. *J. Biol. Chem.* **1991,** *266*, 23022-23026.

29. Morell, A. G.; Gregoriadis, G.; Scheinberg, I. H.; Hickman, J.; Ashwell, G. The role of sialic acid in determining the survival of glycoproteins in the circulation. *J. Biol. Chem.* **1971,** *246*, 1461-1467.

30. Hunter, L. E. *The Processes of Life. An Introduction to Molecular Biology*; MIT Press: Cambridge, 2009.

31. Makrides, S. C. Strategies for achieving high-level expression of genes in *Escherichia coli. Microbiol. Rev.* **1996,** *60,* 512-538.

32. Hunt, I. From gene to protein: a review of new and enabling technologies for multi-parallel protein expression. *Protein Expr. Purif.* **2005,** *40,* 1-22.

33. Bajorunaite, E.; Sereikaite, J.; Bumelis, V.-A. L-Arginine suppresses aggregation of recombinant growth hormones in refolding process from *E. coli* inclusion bodies. *Protein J.* **2007,** *26,* 547-555.

34. Fujii, T.; Ohkuri, T.; Onodera, R.; Ueda, T. Stable supply of large amounts of human Fab from the inclusion bodies in *E. coli. J. Biochem.* **2007,** *141,* 699-707.

35. Chayen, N. E. Turning protein crystallisation from an art into a science. *Curr. Opin. Struct. Biol.* **2004,** *14,* 577-583.

36. Durocher, Y.; Butler, M. Expression systems for therapeutic glycoprotein production. *Curr. Opin. Biotechnol.* **2009,** *20,* 700-707.

37. Ko, K.; Brodzik, R.; Steplewski, Z. Production of antibodies In plants: Approaches and perspectives. *Current topics in microbiology and immunology.* **2009,** *332,* 55-78.

38. Greene, J. J. Host cell compatibility in protein expression. In Recombinant Gene Expression: Reviews and Protocols, *Methods in Molecular Biology, 2nd ed.;* Vol 267. Balbás, P.; Lorence, A. Eds. Humana Press Inc., Totowa, 2004.

39. Grabenhorst, E.; Schlenke, P.; Pohl, S.; Nimtz, M.; Conradt, H. S. Genetic engineering of recombinant glycoproteins and the glycosylation pathway in mammalian host cells. *Glycoconjugate J.* **1999,** *16,* 81-97.

40. Butler, M. Optimisation of the cellular metabolism of glycosylation for recombinant proteins produced by mammalian cell systems. *Cytotechnology.* **2006,** *50,* 57-76.

41. Raju, T. S.; Briggs, J. B.; Borge, S. M.; Jones, A. J. S. Species-specific variation in glycosylation of IgG: Evidence for the species-specific sialylation and branch specific galactosylation and importance for engineering recombinant glycoprotein therapeutics. *Glycobiology.* **2000,** *10,* 477-486.

42. Dayan, A. D. Safety evaluation of biological and biotechnology-derived medicines. *Toxicology.* **1995,** *105,* 59-68.

43. Tripathi, N. K.; Sathyaseelan, K.; Jana, A. M.; Rao, P. V. L. High yield production of heterologous proteins with *Escherichia coli. Defense Science Journal.* **2009,** *59,* 137-146.

44. Ziegler, F. D.; Gemmill, T. R.; Trimble, R. B. Glycoprotein synthesis in yeast. *J. Biol. Chem.* **1994,** *269,* 12527-12535.

45. Herscovics, A.; Orlean, P. Glycoprotein biosynthesis in yeast. *FASEB J.* **1993,** *7,* 540-550.

46. Hamilton, S. R.; Davidson, R. C.; Sethuraman, N.; Nett, J. H.; Jiang, Y.; Rios, S.; Bobrowicz, P.; Stadheim, T. A.; Li, H.; Choi, B.-K.; Hopkins, D.; Wischnewski, H.; Roser, J.; Mitchell, T.; Strawbridge, R. R.; Hoopes, J.; Wildt, S.; Gerngross, T. U. Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science.* **2006,** *313,* 1441-1443.

47. Baker, K. N.; Rendell, M. H.; Hills, A. E.; Hoare, M.; Freedman, R. B.; James, D. C. Metabolic control of recombinant protein *N*-glycan processing in NS0 and CHO cells. *Biotechnol. Bioeng.* **2001,** *73,* 188-202.

48. Higgins, E. Carbohydrate analysis throughout the development of a protein therapeutic. *Glycoconjugate J.* **2010,** *27,* 211-225.

49. Noguchi, A.; Mukuria, C. J.; Suzuki, E.; Naiki, M. Immunogenicity of N-glycolylneuraminic acid-containing carbohydrate chains of recombinant human erythropoietin expressed in Chinese hamster ovary cells. *J. Biochem.* **1995,** *117,* 59-62.

50. Waugh, D. S. Making the most of affinity tags. *Trends Biotechnol.* **2005,** *23,* 316-320.

51. Wong, C.-H. Protein glycosylation: new challenges and opportunities. *J. Org. Chem.* **2005,** *70,* 4219-4225.

52. Shi, W. X.; Shen, Z. M.; Sun, C.; Yang, J. T. Conformation and activity of *Phaseolus coccineus* var. rubronanus lectin. *Journal of Protein Chemistry.* **1993,** *12,* 153-157.

53. Moller, M.; de Wit, E.; Hoal, E. G. Past, present and future direction in human genetic susceptibility to tuberculosis. *FEMS Immunol. Med. Microbiol.* **2010,** *58,* 3-26.

54. Varghese, J. N.; Laver, W. G.; Colman, P. M. Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution. *Nature.* **1983,** *303,* 35-40.

55. Hoshi, T.; Akase, S.; Anzai, J.-I. Preparation of multilayer thin films containing avidin through sugar-lectin interaction and their binding properties. *Langmuir.* **2002,** *18,* 7024-7028.

56. Mellado, M. C.; Curbelo, D.; Nobrega, R.; Castilho, L. R. Comparison of affinity membrane adsorbers for the purification of recombinant human erythropoietin (rhEPO). *J. Chem. Technol. Biotechnol.* **2007,** *82,* 636-645.

57. Prasadarao, N V. Identification of *Escherichia coli* outer membrane protein A receptor on human brain microvascular endothelial cells. *Infect. Immun.* **2002,** *70,* 4556-4563.

58. Zhao, J.; Simeone, D. M.; Heidt, D.; Anderson, M. A.; Lubman, D. M. Comparative serum glycoproteomics using lectin selected sialic acid glycoproteins with mass spectrometric analysis: Application to pancreatic cancer serum. *J. Proteome Res.* **2006,** *5,* 1792-1802.

59. Weis, W. I.; Drickamer, K. Structural basis of lectin-carbohydrate recognition. *Annu. Rev. Biochem.* **1996,** *65,* 441-473.

60. Knibbs, R. N.; Goldstein, I. J.; Ratcliffe, R. M.; Shibuya, N. Characterization of the carbohydrate binding specificity of the leukoagglutinating lectin from *Maackia amurensis. J. Biol. Chem.* **1991,** *266,* 83-88.

61. Babal, P.; Pindak, F. F.; Wells, D. J.; Gardner Jr., W. A. Purification and characterization of a sialic acid-specific lectin from *Tritrichomonas mobilensis. Biochem. J.* **1994,** *299,* 341-346.

62. Siuzdak, G. *The Expanding Role of Mass Spectrometry in Biotechnology*, 2nd ed.; MCC Press: San Diego, 2006.

63. de Hoffmann, E.; Stroobant, V. *Mass Spectrometry. Principles and Applications,* 2nd ed.; John Wiley & Sons, LTD: West Sussex, 2002.

64. Liebler, D. C. *Introduction to Proteomics. Tools for the New Biology;* Humana Press: Totowa, 2002

65. Dass, C. *Fundamentals of contemporary mass spectrometry*; John Wiley & Sons, Inc.: Hoboken, 2007.

66. Dalpathado, D. S.; Irungu, J.; Go, E. P.; Butnev, V. Y.; Norton, K.; Bousfield, G. R.; Desaire, H. Comparative glycomics of the glycoprotein follicle stimulating hormone: Glycopeptide analysis of isolates from two mammalian species. *Biochemistry.* **2006,** *45,* 8665-8673.

67. Go, E. P.; Rebecchi, K. R.; Dalpathado, D. S.; Bandu, M. L.; Zhang, Y.; Desaire, H. GlycoPep DB: A tool for glycopeptide analysis using a smart search. *Anal. Chem.* **2007,** *79,* 1708-1713.

68. Irungu, J.; Go, E. P.; Dalpathado, D. S.; Desaire, H. Simplification of mass spectral analysis of acidic glycopeptides using GlycoPep ID. *Anal. Chem.* **2007,** *79,* 3065-3074.

69. Go, E. P.; Irungu, J.; Zhang, Y.; Dalpathado, D. S.; Liao, H.-X.; Sutherland, L. L.; Alam, S. M.; Haynes, B. F.; Desaire, H. Glycosylation site-specific analysis of HIV envelope proteins (JR-FL and CON-S) reveals major differences in glycosylation site occupancy, glycoform profiles, and antigenic epitopes' accessibility. *J. Proteome Res.* **2008,** *7,* 1660-1674.

70. Zhang, Y.; Go, E. P.; Desaire, H. Maximizing coverage of glycosylation heterogeneity in MALDI-MS analysis of glycoproteins with up to 27 glycosylation sites. *Anal. Chem.* **2008,** *80,* 3144-3158.

71. Levander, F.; Rognvaldsson, T.; Samuelsson, J.; James, P. Automated methods for improved protein identification by peptide mass fingerprinting. *Proteomics.* **2004,** *4,* 2594-2601.

72. Novotny, M. V.; Mechref, Y. New hyphenated methodologies in high-sensitivity glycoprotein analysis. *J. Sep. Sci.* **2005,** *28,* 1956-1968.

73. Plasencia, M. D.; Isailovic, D.; Merenbloom, S. I.; Mechref, Y.; Novotny, M. V.; Clemmer, D. E. Resolving and assigning *N*-linked glycan structural isomers from ovalbumin by IMS-MS. *J. Am. Soc. Mass Spectrom.* **2008,** *19,* 1706-1715.

74. Ding, Q.; Xiao, L.; Xiong, S.; Jia, Y.; Que, H.; Guo, Y.; Liu, S. Unmatched masses in peptide mass fingerprints caused by cross-contamination: An updated statistical result. *Proteomics.* **2003,** *3,* 1313-1317.

75. Mann, M. Useful Tables of Possible and Probable Peptide Masses. *Abstract from the 43rd Annual Conference on Mass Spectrometry and Allied Topics*; Atlanta, GA, 1995. (American Society for Mass Spectrometry).

76. Dodds, E. D.; An, H. J.; Hagerman, P. J.; Lebrilla, C. B. *J. Proteome Res.* **2006,** *5,* 1195-1203.

77. Greenfield, N.; Fasman, G. D. Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry.* **1969,** *8,* 4108-4116.

78. Tsai, A. M.; van Zanten, J. H.; Betenbaugh, M. J. I. Study of protein aggregation due to heat denaturation: A structural approach using circular dichroism spectroscopy, nuclear magnetic resonance, and static light scattering. *Biotechnol. Bioeng.* **1998,** *59,* 273-280.

79. Allen, D. L.; Pielak, G. J. Baseline length and automated fitting of denaturing data. *Protein Sci.* **1998,** *7,* 1262-1263.

80. Benjwal, S.; Verma, S.; Rohm, K.-H.; Gursky, O. Monitoring protein aggregation during thermal unfolding in circular dichroism experiments. *Protein Sci.* **2006,** *15,* 635-639.

81. Sreerama, N.; Woody, R. W. Computation and analysis of protein circular dichroism spectra. *Methods in Enzymology. Numerical Computer Methods, Part D.* **2004,** *383,* 318-351.

82. Bulheller, B. M.; Miles, A. J.; Wallace, B. A.; Hirst, J. D. Charge-transfer transitions in the vacuum-ultraviolet of protein circular dichroism spectra. *J. Phys. Chem. B.* **2008,** *112,* 1866-1874.

83. Bam, N. B.; Cleland, J. L.; Randolph, T. W. Molten globule intermediate of recombinant human growth hormone: Stabilization with surfactants. *Biotechnol. Prog.* **1996,** *12,* 801-809.

84. Johnson, Jr., W. C. Protein secondary structure and circular dichroism: a practical guide. *Proteins: Structure, Function, and Genetics.* **1990,** *7,* 205-214.

85. de Vos, A. M.; Ultsch, M.; Kossiakoff, A. A. Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science.* **1992,** *255,* 306-312.

86. Sreerama, N.; Woody, R. W. A self-consistent method for the analysis of protein structure from circular dichroism. *Anal. Biochem.* **1993,** *209*, 32–44.

87. Lobley, A.; Whitmore, L.; Wallace, B. A. DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics.* **2002,** *18,* 211–212.

88. Lees, J. G.; Miles, A. J.; Wien, F.; Wallace, B. A. A reference database for circular dichroism spectroscopy covering fold and secondary structure space. *Bioinformatics.* **2006,** *22,* 1955–1962.

89. Ahmad, F. On the estimation of stability parameters from heat-induced conformational transition curves of proteins. *Journal of the Iranian Chemical Society.* **2004,** *1,* 99-105.

90. Karpenko, V.; Kaupova, M.; Kodicek, M. The conformation and stability of human Zn-α_2-glycoprotein in aqueous and methanolic solutions. *Biophys. Chem.* **1997,** *69,* 209-217.

CHAPTER 2

THE NEW PROTEIN PURIFICATION HANDLE

Protein production via recombinant technology is an important tool across different industries and in life science studies. Producing large amounts of purified, folded protein, however, is a considerable challenge, and significant improvements to the purification schemes currently in use are needed to take full advantage of this powerful technology. We have designed a new protein purification platform that circumvents some of the most challenging complications that relate to purification, such as affinity tag removal, and generating natively folded, human proteins. In this novel protein purification scheme, a target protein sequence is modified to include a non-native glycosylation site, and the protein is glycosylated during expression in eukaryotic cells. The glycan structure serves as a unique affinity tag, affording purified protein by lectin affinity chromatography. After protein isolation, the glycan may be removed, and the native protein sequence is regenerated. We have demonstrated this new method by expressing and purifying two novel human growth hormone mutants. Each protein contained a single, nonnative glycan used as an affinity tag, which was later removed. The resulting protein was analyzed for secondary structure using circular dichroism. This new method improves protein production by providing a new affinity purification tag used in generating purified, properly folded protein.

2.1 INTRODUCTION

Recombinant protein production is pivotal in multiple industries, *e.g.* pharmaceuticals, biomedical diagnostics, and life sciences research.[1] In any setting, producing adequate quantities of highly purified protein is a formidable challenge. Recombinantly expressed proteins are commonly purified using one or more affinity tags, paired with a chromatographic method such as immobilized metal affinity chromatography (IMAC).[2] Affinity tags usually consist of amino acid residues selected for their chemical behavior or binding epitopes, and are engineered into the protein sequence, commonly at the C or N terminus.[3] The most widely used affinity tag is the polyhistidine tag (His-tag), consisting of a repeating sequence of H residues.[2,3]

His-tags may require additional amino acids near the tag location to direct tag removal using an exo- or endopeptidase.[2-4] The enzymes used in tag removal from the C-terminus often leave 5 or more amino acids remaining at the cleavage site, which are not native to the protein sequence.[2-4] His-tags may be partially or completely buried within the protein structure, making removal more difficult.[4] These small tags have been reported to negatively impact the structure, activity, or binding of some proteins.[5-10] Unfortunately, some classes of protein are excluded from His-tag-IMAC purification, such as proteins with metal centers, whereas the affinity resin may bind the metal.[5] These observations direct researchers to search for alternative purification strategies.

Other affinity tags currently in use may be significantly larger than His-tags, such as the maltose-binding protein (MBP).[11] These larger tags may be used in conjunction with a His-tag, or alone, but each tag is associated with unique structure or function effects, in addition to problematic tag removal procedures and protein solubility

issues.[3,12] The development of a new strategy that overcomes commonly encountered problems in protein purification has the potential to make significant contributions to science and medicine.

To accomplish the aim of producing significant amounts of pure, folded protein, and to mitigate the complications described above, we have developed a new protein purification affinity tag. This method consists of incorporating a nonnative, asparagine-linked (N-linked) glycan site into a nonglycosylated protein sequence. This change triggers enzymes to glycosylate the protein during eukaryotic expression, and the glycan provides a unique affinity tag, which can be exploited for protein purification. The protein is purified using lectin affinity chromatography (LAC). The glycan affinity tag may then be removed in its entirety using chemical or enzymatic means to produce the nonglycosylated product. A work-flow depiction is provided in **Figure 1**.

To illustrate the utility of this new protein purification approach, we designed two different, novel recombinant human growth hormone (rhGH) mutant proteins. Found endogenously, hGH is a non-glycosylated 22 kDa protein and is critical to proper growth and metabolism.[13] Some children and adults have hGH deficiencies, and recombinantly expressed hGH has been an important biopharmaceutical treatment for these individuals.[14-16] In the study herein, the two mutants were generated by adding a single N-linked glycosylation site into the primary sequence of each mutant, using mutagenesis. The two mutants, D136N and D153N,N155T were designed to introduce consensus glycosylation sequences (N-X-T/S) at a disordered loop in the protein. When this consensus sequence is present in mammalian cell proteins, a glycan is attached at

the N.[17] The D136N mutant's glycosylation sequence is N(136)GS. The second

mutant's glycosylation sequence is N(153)TT.
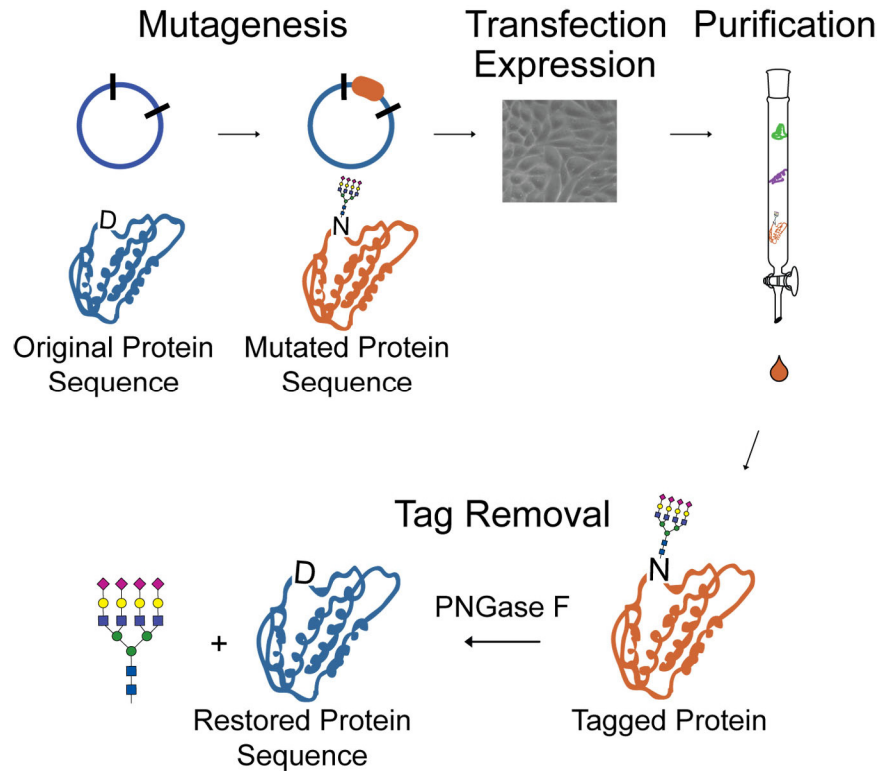
Figure 1. Protein production and purification scheme using an *N*-linked glycan as a temporary protein purification tag. The tag removal step uses PNGase F to completely remove the purification tag without leaving nonnative amino acids at the site, and also to restore the amino acid sequence to the naturally occurring D residues present in the native growth hormone sequence.

These mutant proteins were expressed in Chinese hamster ovary cells (CHO),

the most commonly used mammalian expression hosts, which are known for producing

complex N-linked glycans at the consensus sequence.[17] Purification was performed

using lectin affinity chromatography (LAC), and the glycan was removed using Protein

N-glycanase F (PNGase F). This endoglycosidase is known to cleave N-linked glycans

from proteins, and in the process the N is converted to a D[18], thus restoring the original

human growth hormone sequence, as shown in Figure 1. The pure, nonglycosylated product was characterized using gel electrophoresis (SDS-PAGE) and circular dichroism (CD) and was found to possess the desired molecular weight and secondary structure.

2.2 EXPERIMENTAL PROCEDURES

2.2.1 Materials

The CHO-K1 cells were a gift from Dr. Jeff Krise (University of Kansas, Department of Pharmaceutical Chemistry, Lawrence, KS). Water was purified using the Millipore Direct Q-3 system (Millipore, Billerica, MA). All other materials were obtained from Sigma unless otherwise specified. All solutions were passed through a 0.2 µm filter prior to use.

2.2.2 Construction of Recombinant Template Plasmid

To create the rhGH plasmid, the hGH gene in mammalian vector (pCMV6-XL5) was purchased from Origene Technologies Inc. (number SC-3300088). The 5' native signal sequence was included in the construct to enable secretion of the protein. The DNA encoded for a polyhistidine tag consisting of four histidine residues at the N terminus. Competent XL-1 Blue *E. coli* cells were transformed using the plasmid, then plated on Luria broth (LB) agar plates containing 100 mg/mL ampicillin. A colony was selected and used to inoculate 10 mL of selective LB medium. The resultant DNA was purified using a QIAprep Spin Miniprep Kit (QIAGEN, Valencia, CA) and quantified using a standard absorption assay on a Nanodrop® ND-1000 Spectrophotometer (Thermo Scientific, Wilmington, DE).

2.2.3 Site-Directed Mutagenesis of Glycosylation Site

The plasmid was used as a template for modification to include an N-linked glycan. Primers (Integrated DNA Technologies, Inc., Coralville, IA) were designed to generate the rhGH-D136N mutant: (5' to 3') atg ggg agg ctg gaa AAT ggc agc ccc cgg act g (leading) and c agt ccg ggg gct gcc ATT ttc cag cct ccc cat (reverse), where the mutated bases are indicated using capital letters. Mutagenesis was performed using the QuikChange mutagenesis kit (Stratagene, La Jolla, CA). Primers were combined with the DNA template and the manufacturer's protocol was implemented.

The mutagenesis product was digested with 1 µL of DpnI enzyme (New England Biosciences, Ipswich, MA) to selectively destroy the template DNA. The product was then used to transform competent XL-1 Blue *E. coli* cells. Transformed cells were plated onto LB agar plates containing ampicillin, grown overnight at 37 °C, and a single colony was selected to inoculate liquid LB medium. After overnight growth, the resultant DNA was isolated and purified using the QIAprep Spin Miniprep Kit. All miniprep DNA was quantified as described above and sequenced by Northwoods DNA Inc., Bemidji, MN. The desired sequence was verified prior to transfection.

To generate the rhGH-D153N,N155T mutant, the same procedure was used, except two mutations were performed using one set of primers that encoded both mutations. The primer used to create the second mutant: (5' to 3') cc tac agc aag ttc AAC aca ACC tca cac aac gat gac gc (leading) and gc gtc atc gtt gtg tga GGT tgt GTT gaa ctt gct gta gg (reverse). DNA sequence was verified prior to transfection.

2.2.4 Cell Culture and Transfection

CHO-K1 cells were seeded and maintained using complete high glucose Dulbecco's Modified Eagle Medium (DMEM) (Thermo Scientific Hyclone, Logan, UT). The medium also contained supplemental MEM Non-Essential Amino acids 100x (ATCC, Manassas, VA), 10% fetal bovine serum (Mediatech, Inc. Manassas, VA), 100X penicillin/streptomycin (ATCC, Manassas, VA), and 150 mM L-Proline.[19] Cells were maintained in T-75 flasks (BD, Franklin Lakes, NJ) with 24 mL of medium and kept at 37 °C and 5% CO_2. Passage was accomplished by two wash steps using pre-warmed phosphate buffered saline (PBS) alone and PBS containing trypsin to remove the adherent cells from the surface of the flask. After three minutes, the suspension was centrifuged at 3800 x g for 2 minutes in a 15 mL Falcon tube, the supernatant was removed via vacuum, and the cells were resuspended via gentle pipetting into 6 mL of the culture medium described above.

Cells were allowed to grow to 50% confluency and then transfection was accomplished with 4 µg of mutated DNA and 20 µL of Turbofectin 8.0 transfection reagent (OriGene Technologies, Inc., Rockville, MD) per T-75 flask, per mutant protein. The supernatant was decanted after 24 hours and the medium replaced. After an additional 24 hours the supernatant was collected and concentrated to 0.5 mL using centrifugal filtration devices (Millipore, Billerica, MA) with a 10 kDa molecular weight cut off (MWCO) at 9000 x g.

2.2.5 Lectin Affinity Chromatography (LAC) Purification

A cartridge (5-mL) containing *Maackia amurensis* leukoagglutinin (MAL) lectin resin (Qiagen, Valencia, CA) was equilibrated according to the manufacturer's directions using the indicated binding/wash buffer. The concentrated supernatant (0.5 mL) from a single supernatant collection was mixed with an equal volume of binding/wash buffer and loaded onto the column at 1 mL/minute. The column was washed using 50 mL of binding/wash buffer at 4 mL/min, and the rhGH was eluted into 50 mL of elution buffer containing 200 mM lactose. The elution fraction was concentrated to approximately 1 mL as described above. The protein content was quantified using a standard Bradford assay. The rhGH protein was dialyzed using Slide-A-Lyzer dialysis cassette with a 10 kDa MWCO (Pierce, Rockford, IL) into 10 mM sodium citrate and 150 mM sodium chloride, with 3 (1 L) exchanges.

2.2.6 Immobilized Metal Affinity Chromatography (IMAC) Purification

A prepared 5-mL column containing Chelating resin (GE Healthcare) charged with nickel was equilibrated. A sample of rhGH-D136N, which was purified using LAC and quantified with a standard Bradford assay, was mixed with an equal volume of loading buffer and was loaded onto the IMAC column. Loading buffer containing 50 mM TRIS-HCl, and 150 mM sodium chloride was adjusted with sodium hydroxide to pH 8.0 and was filtered (0.2 μm). Rinsing buffer was the same, except it also contained 20 mM imidazole. The elution buffer was the same as the loading buffer, except it also contained 200 mM imidazole. The column was washed using 100 mL of loading buffer; rinsed using 50 mL of rinse buffer, and elution was performed via syringe using 16 mL

of elution buffer. The collected fractions were analyzed using SDS-PAGE and their protein content quantified using a standard Bradford assay. The procedure was repeated for the rhGH-D153N,N155T protein.

2.2.7 SDS-PAGE Analysis

12% SDS-PAGE TRIS/glycine gels were generated for electrophoretic analysis. Protein samples were mixed with 2X reducing Laemmli loading buffer and water (1:3:3) and boiled for 15 minutes. The samples were loaded, and electrophoresis was performed at 123 V. Precision Plus unstained protein ladder (Bio-Rad, Hercules, CA) was utilized in separate lane(s) for molecular weight approximation. Gels were first stained using Coomassie (R-250) and then fully destained (using acetic acid:methanol 2:3) prior to silver staining.

2.2.8 Purification Tag Removal

Purified, glycosylated rhGH was treated with Protein N-Glycosidase F (PNGase F) from *Elizabethkingia meningosepticum* (CalBioChem, San Diego, CA) by adding 1 μL of 500 units/mL to approximately 2 mg of protein, and incubating for 24 hours at 37 ℃. An aliquot of rhGH protein was loaded onto the MAL lectin cartridge using the procedure described above. The collected fractions were analyzed using both SDS-PAGE and a Bradford assay to determine the final amount of purified, nonglycosylated rhGH.

2.2.9 CD Secondary Structure Estimation

The purified protein was diluted to 0.5 mg/mL in 10 mM sodium citrate and 150 mM sodium chloride and placed in a jacketed, quartz, 1.0-mm pathlength sample cell. Far UV CD measurements were made using a Jasco J-715 spectropolarimeter (Tokyo, Japan) with an attached circulating water bath that was built in-house. Experiments were performed at 10 °C, with a sensitivity of 100 mdeg and scan speed of 10 nm/min. The experiments were conducted under constant nitrogen flow. Scans were performed from 260 to 180 nm. Multiple spectra were averaged, smoothed and baseline corrected by subtracting signal from background solutions prior to analysis. The publicly available DICHROWEB server was used, with the CDSSTR algorithm, to calculate the percent protein secondary structure.

2.3 RESULTS AND DISCUSSION

Protein production methods aim to generate large amounts of pure and natively folded protein. This is a difficult goal to achieve. Challenges common to protein purification exist, such as optimizing the purification method and ensuring proper protein folding . To elude these issues, we propose a novel protein purification strategy to support the production of pure, folded protein. The strategy consists of introducing a nonnatural glycosylation site into the protein sequence, expressing the protein in eukaryotic cells to connect glycan tag and protein, then purifying the protein using lectin affinity. Figure 1 illustrates these steps.
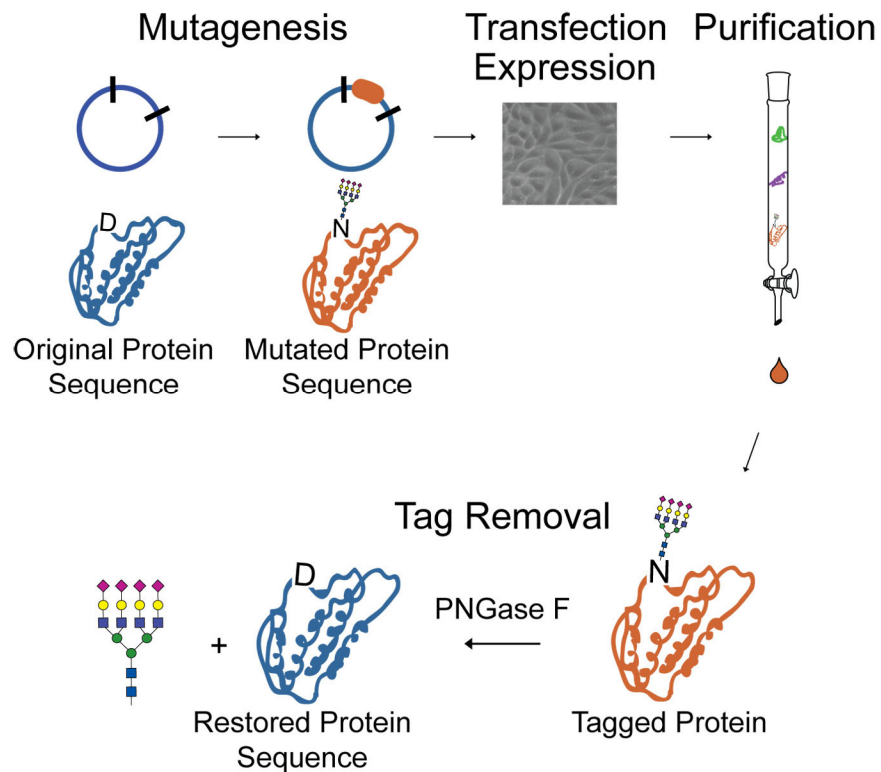
Figure 1. Protein production and purification scheme using an *N*-linked glycan as a temporary protein purification tag. The tag removal step uses PNGase F to completely remove the purification tag without leaving nonnative amino acids at the site, and also to restore the amino acid sequence to the naturally occurring D residues present in the native growth hormone sequence.

2.3.1 Novel Protein Purification Platform

To demonstrate the utility of purifying protein using this approach, we expressed two mutant, glycosylated forms of recombinant human growth hormone in CHO-K1 cells, and purified the protein using lectin affinity chromatography. In **Figure 2**, the sequence of the native hGH and the two glycosylated mutants generated in this study is shown. The two mutants each contain the "consensus sequence" (N-X-S/T), which is recognized by glycosylating enzymes within eukaryotic cells as a signal to attach, and then modify, an N-linked glycan.[20-21] The resulting glycan was used as an affinity tag to purify the growth hormone with lectin affinity chromatography. After purification, the glycan was enzymatically removed using PNGase F, as illustrated in the Tag Removal step in Figure 1 above. The tag removal step relies on the inherent enzymatic activity of PNGase F, which converts an N residue to a D during the *N*-linked glycan removal reaction.[18] The conversion to an aspartic acid actually restores the purified rhGH-D136N protein to its native sequence, which naturally contains a D at that amino acid position. The original protein sequence and genetic modifications are given in Figure 2.

```
MAHHHHFPTIPLSRLFDNAMLRAHRLHQLAFDTYQ
EFEEAYIPKEQKYSFLQNPQTSLCFSESIPTPSNR
EETQQKSNLELLRISLLLIQSWLEPVQFLRSVFAN
 Mutant I                         N
SLVYGASDSNVYDLLKDLEEGIQTLMGRLEDGSPR
 Mutant II            N  T
TGQIFKQTYSKFDTNSHNDDALLKNYGLLYCFRKD
MDKVETFLRIVQCRSVEGSCGF
```

Figure 2. Verified DNA sequence of modified rhGH mutants. The first mutant, rhGH-D136N, was modified to include the consensus sequence for N-linked glycan incorporation, using a single point mutation. The second mutant, rhGH-D153N,N155T, was modified to include the consensus sequence with two point mutations.

After glycan removal, Mutant I contained the native hGH sequence. Mutant II, the rhGH-D153N,N155T protein, contained only a single amino acid variant from the native sequence, with T-155 in place of an N residue. This modification is necessary to meet the requirements of the consensus sequence for *N*-linked glycosylation. These changes made to the protein sequence to engineer the glycan affinity tag require significantly fewer modifications than other purification tags currently in use, which often require many nonnative amino acids to be incorporated with the protein sequence.[2-4]

2.3.2 Protein Analysis

The glycosylated, rhGH proteins were analyzed before and after purification by lectin affinity, using SDS-PAGE. In **Figure 3a,** the lane to the right of the molecular weight marker lane represents the media prior to purification. A very large band at approximately 70 kDa is visible along with several other bands, including a faint band between the 20 kDa and 25 kDa markers, which matches the molecular weight of glycosylated growth hormone. The largest band is the correct molecular weight for bovine serum albumin, which is added to the cell growth medium. This lane clearly indicates, as expected, that prior to purification the protein of interest is present in a complex mixture of other proteins, whose concentrations are high, compared to the expressed protein. The rhGH bands positioned slightly above the 20 kDa marker are more evident after purification using LAC, as observed in the two lanes on the right in Figure 3a. Lane 3 is the band resulting from analysis of rhGH-D136N, and lane 4 contains rhGH-D153N,N155T.

A second analysis of rhGH-D136N using SDS-PAGE was silver stained for increased sensitivity (**Figure 3b**). There is a single, distinct band in the lane containing rhGH-D136N, at the expected molecular weight for glycosylated rhGH. No other significant bands were visible using the more sensitive silver staining procedure, indicating a purified protein product is present.

Figure 3a.
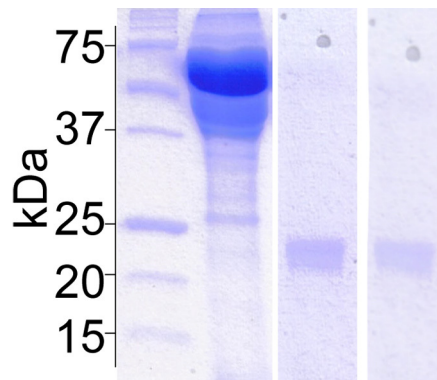
Figure 3b.
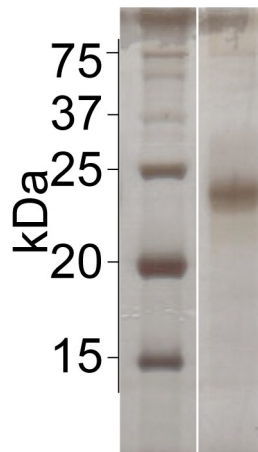
Figure 3. SDS-PAGE analysis of novel rhGH proteins. (a) Sample in lane 2 is an aliquot of collected cellular supernatant, lanes 3 and 4 contain rhGH-D136N and rhGH-D153N,N155T that have been purified using lectin affinity chromatography, respectively. Coomassie stain was used. (b) Right lane is a sample of purified rhGH-D136N, silver stain procedure was used.

2.3.3 Protein Quantitation and Comparison of Purification Methods

Purified rhGH proteins were quantified using a standard Bradford assay. The resulting values are displayed in **Table 1.** From a single 24-hour collection of the cell growth media (supernatant), 1.67 mg of glycosylated rhGH protein was recovered from 24 mL of medium, as determined after LAC purification.

Table 1. Bradford Assay Results for Protein Concentration

rhGH Sample (in mg/mL)	LAC (elution)	IMAC (wash I)	IMAC (elution)	LAC (wash)*
D136N	1.67	1.27	n/d	1.01
D153N,N155T	1.82	1.54	n/d	1.24

The notation of n/d indicates not detected, * indicates the protein sample had been deglycosylated using PNGase F.

To compare this affinity purification method to purification with a polyhistidine tag, the purified rhGH-D136N, which contained a polyhistadine tag at the N-terminus, was also purified using an IMAC column. The collected fractions were concentrated as described above. No protein was detected in the second wash fraction, the rinse fraction or the elution fraction, as indicated by n/d for not detected within Table 1. Analysis of the first wash IMAC wash fraction (wash I) showed the rhGH-D136N protein concentration to be 1.27 mg/mL, after concentration to approximately one mL. This study indicates that the metal affinity method was ineffective at binding the polyhistidine tagged protein, since the tag should have bound to the column and only eluted during the "elution" step.

The rhGH-D136N protein sample was deglycosylated using PNGase F, and then again loaded onto the LAC cartridge. Any protein that was still glycosylated after

reaction with PNGase F would bind to the column, as demonstrated above. Therefore, the LAC method could be used to separate glycosylated from deglycosylated protein and quantify the amount of protein that had been deglycosylated. The only detectable protein was found in the LAC wash fraction, and after sample concentration to one mL the amount of protein detected was 1.01 mg/mL. Similar results were obtained for the rhGH-D153,N155T protein and can be found in Table 1.

2.3.4 Secondary Structural Measurements

Secondary structure of the deglycosylated protein, containing the native hGH sequence, was characterized using far-UV circular dichroism. The spectra are shown in **Figure 4**. Minima in the spectra are observed at wavelengths 209 nm and also at 222 nm, indicating alpha helical structure.
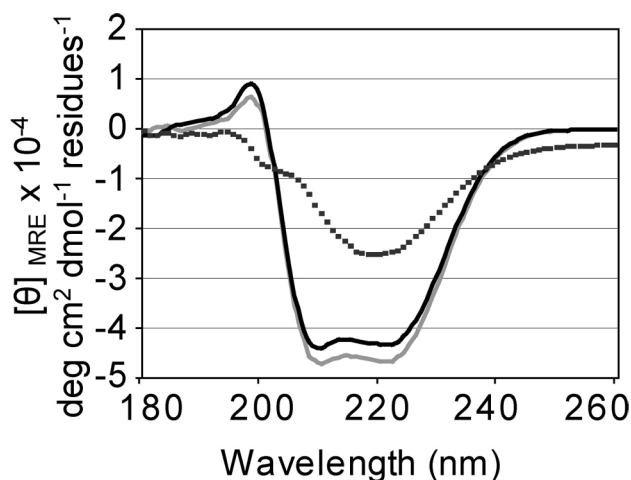
Figure 4: Far-UV CD data for glycosylated human growth hormone (▪▪) and deglycosylated proteins rhGH-D136N (in gray) and rhGH-D153N,N155T (in black). The curves corresponding to the deglycosylated proteins indicate alpha helical character is dominant after removal of the glycan purification tag.

The far-UV scan of the rhGH protein was compared to data in the literature.[21-24] The spectra for the deglycosylated proteins indicate that a large portion of the secondary structure present in the protein sample is alpha helical in character. Endogenous hGH is known to have alpha helices accounting for 50-60% of secondary structural characteristics.[21-24] To quantify the helical content of the rhGH sample, the CD data was analyzed using the publicly available, web-based DICHROWEB program.[25-28] The helical secondary structure calculation results for deglycosylated rhGH-D136 are 54% and rhGH-D153,N155T, 59%, as displayed in **Table 2.** These values are within the range of acceptable values found in literature for other hGH samples.[21-24]

Table 2: Secondary structure of deglycosylated rhGH as predicted by DICHROWEB.

rhGH	Helix	Strand	Turns	Unordered	Total
D136N	54%	9%	13%	24%	100%
D153N,N155T	59%	16%	8%	17%	100%

2.3.5 Advantages of this new method

The approach to protein production and purification described above has several benefits over other purification methods that are currently in use, such as using IMAC and His-tags. Here, we have generated a protein purification affinity tag, but the tag is not restricted to the protein's termini, as glycans can be incorporated into the protein at any point, including the termini, if desired. The tag could be specifically inserted into a disordered loop of the protein, or far away from active sites, or at the termini. Other affinity tags are practically restricted to either the N or C terminus. The ability to select the location of the purification tag is advantageous because proteins such as hGH may have binding motifs close to or including the terminal parts of the protein.[23-25]

51

In addition, other purification tags may require additional amino acid sequences to direct removal of the tag by proteases, but glycans can be partially or completely removed without this additional requirement. In this method, PNGase F was used to remove the glycan purification tag, leaving no additional amino acids on the protein. This tag removal process may prove beneficial over currently employed tag removal methods, where several nonnative amino acids often remain on the protein sequence after the tag removal steps are completed, and in some cases the additions negatively impact protein folding or activity.[5-10]

2.4 CONCLUDING REMARKS

The utility of this novel protein purification method has been demonstrated by producing two properly folded, purified, and novel human growth hormone mutants. The proteins were produced in soluble and folded form using an eukaryotic host, as opposed to the misfolded and aggregated forms often encountered using *E. coli.*[32] The lectin affinity purification method, used in conjunction with the glycan affinity tag, produced highly purified protein in good yield; by comparison, the 4-His-tag on the N terminus of the protein was an ineffective tag for IMAC purification. Finally, once the glycan affinity tag was removed, the protein adopted the proper degree of alpha helical content, supporting the assertion that the native structure was retained.

The use of a glycan purification tag has several benefits over current affinity labels. It can be engineered into the protein sequence where desired, instead of being forced onto one of the termini. Additionally, the tag addition site can be chosen such that after tag removal, no nonnative amino acids will remain on the final protein product. This approach to protein purification has the potential to make a significant impact on medicine and life science research by providing an alternative strategy for protein production and purification.

2.5 FUTURE WORK

Both human growth hormone protein mutants described in this chapter will be subjected to proteolytic digestion and high resolution mass spectrometric analysis. These studies will provide ancillary support that we have purified the protein of interest, and they can be used to characterize the glycan affinity tag. The sample preparative steps will employ glycopeptide and peptide analysis, as described in Chapter 1, section 1.2.3.1.

2.6 REFERENCES

1. Carlson, B. *Genetic Engineering & Biotechnology News.* **2009,** *29,* 8.

2. Arnau, J.; Lauritzen, C.; Petersen, G. E.; Pedersen, J. *Protein Expr. Purif.* **2006,** *48,* 1-13.

3. Waugh, D. S. *Trends Biotechnol.* **2005,** *23,* 316-320.

4. Narmandakh, A.; Bearne, S. L. *Protein Expr. Purif.* **2010,** *69,* 39-46.

5. Terpe, K. *Appl. Microbiol. Biotechnol.* **2003,** *60,* 523-533.

6. Chant, A.; Kraemer-Pecore, C. M.; Watkin, R.; Kneale, G. G. *Protein Expr. Purif.* **2005,** *39,* 152-159.

7. Freydank, A.-C.; Brandt, W.; Drager, B. *Proteins.* **2008,** *72,* 173-183.

8. Goel, A.; Colcher, D.; Koo, J.-S.; Booth, B. J. M.; Pavlinkova, G.; Batra, S. K. *Biochim. Biophys. Acta.* **2000,** *1523,* 13-20.

9. Ohana, R. F.; Encell, L. P.; Zhao, K.; Simpson, D.; Slater, M. R.; Urh, M.; Wood, K. V. *Protein Expr. Purif.* **2009,** *68,* 110-120.

10. Wu, J.; Filutowicz, M. *Acta Biochim. Polon.* **1999,** *46,* 591-599.

11. di Guan, C.; Li, P.; Riggs, P.D.; Inouye, H. *Gene.* **1988,** *67,* 21-30.

12. Esposito, D.; Chatterjee, D. K. *Curr. Opin. Biotech.* **2006,** *17,* 353-358.

13. Patrick, G. L. *An Introduction to Medicinal Chemistry,* 2nd ed.; Oxford University Press: New York, 2001.

14. Laron, Z. *N. Engl. J. Med.* **1996,** *334,* 463-466.

15. Takahashi, Y.; Kaji, H.; Okimura, Y.; Goji, K.; Abe, H.; and Chihara, K. *N. Engl. J. Med.* **1996,** *334,* 432-436.

16. Salomon, F.; Cuneo, R. C.; Hesp, R.; Sonksen, P. H. *N. Engl. J. Med.* **1989,** *321,* 1797-1803.

17. Hossler, P.; Khattak, S. F.; Li, Z. J. *Glycobiology.* **2009,** *19,* 936-949.

18. Plummer, T. H. Jr.; Elder, J. H.; Alexander, S.; Phelan, A. W.; Tarentino, A. L. *J. Biol. Chem.* **1984,** *259,* 10700-10704.

19. Kao, F.-T., Puck, T. T. *Genetics.* **1967,** *55,* 513-524.

20. Devasahayam, M. *Indian J. Med. Res.* **2007,** *126,* 22-27.

21. Helenius, A.; Aebi, M. *Science.* **2001,** *291,* 2364-2369.

22. Bam, N. B.; Cleland, J. L.; Randolph, T. W. *Biotechnol. Prog.* **1996,** *12,* 801-809.

23. Wells, J. A.; de Vos, A. M. *Annu. Rev. Biophys. Biomol. Struct.* **1993,** *22,* 329-351.

24. Abdel-Meguid, S. S.; Shieh, H.-S.; Smith, W. W.; Dayringer, H. E.; Violand, B. N.; Bentle, L. A. *Proc. Natl. Acad. Sci. USA* **1987,** *84,* 6434-6437.

25. de Vos, A. M.; Ultsch, M.; Kossiakoff, A. A. *Science.* **1992,** *255,* 306-312.

26. Sreerama, N.; Woody, R. W. *Anal. Biochem.* **1993,** *209*, 32–44.

27. Lobley, A.; Whitmore, L.; Wallace, B. A. *Bioinformatics.* **2003,** *18,* 211–212.

28. Lees, J. G.; Miles, A. J.; Wien, F.; Wallace, B. A. *Bioinformatics.* **2006,** *22,* 1955–1962.

29. Whitmore, L.; Wallace, B. A. *Nucleic Acids Res.* **2004,** *32,* W668–W673.

30. Sola, R. J.; Griebenow, K. *J. Pharm. Sci.* **2009,** *98,* 1223-1245.

31. Gut, A.; Kappeler, F.; Hyka, N.; Balda, M. S.; Hauri, H.-P.; Matter, K. *EMBO J.* **1998,** *17,* 1919-1929.

32. Thomas, J. G.; Ayling, A.; Baneyx, F. *Appl. Biochem. Biotechnol.* **1997**, *66,* 197-238.

CHAPTER 3

POST-PROTEIN EXPRESSION ENZYMATIC GLYCAN TRIMMING

Recombinant glycoproteins present unique challenges to biopharmaceutical development, especially when efficacy is affected by glycosylation. In these cases, optimizing the protein's glycosylation is necessary, but difficult, since the glycan structures cannot be genetically encoded, and glycosylation in non-human cell lines can be very different from human glycosylation profiles. We are exploring a potential solution to this problem by designing enzymatic glycan optimization methods to produce proteins with useful glycan compositions. To demonstrate viability of this new approach to generating glycoprotein-based pharmaceuticals, the *N*-linked glycans of a model glycoprotein, ribonuclease B (RNase B), were modified using an α-mannosidase to produce a new glycoprotein with different glycan structures. The secondary structure of the native and modified glycoproteins was retained, as monitored using circular dichroism. An assay was also developed using an RNA substrate to verify that RNase B had indeed retained its function after being subjected to the necessary glycan modification conditions. This is the first study that verifies both activity and secondary structure of a glycoprotein after enzymatic glycan trimming for use in glycoprotein therapeutic development methods. The evidence of preserved structure and function for a modified glycoprotein indicates that extracellular enzymatic modification methods could be implemented in producing designer glycoproteins for therapeutic use.

3.1 INTRODUCTION

Glycoprotein therapeutics are currently administered for treatment and/or prevention of diseases. Trends indicate that glycoproteins will continue to account for the majority of new biopharmaceuticals.[1] Protein and glycoprotein-based therapeutics are being pursued in part, due to their known, specific mechanisms of action. Glycoprotein therapeutic use has been made possible by the rapid development in recombinant glycoprotein expression techniques over the last few decades. Mammalian cell lines, used for glycoprotein expression, are capable of producing proteins with complex-type asparagine linked (*N*-linked) glycans, but the glycan compositions are heterogeneous.[2-3] Other expression hosts such as yeast and insect cells are not equipped with the cellular machinery to create the complex *N*-linked glycans found in human glycosylation profiles without genetic modifications.[4-5]

N-linked glycans are of particular importance for glycoprotein pharmaceutical production, because they have shown to directly influence the glycoprotein's properties such as stability, folding, charge, and even function.[6-9] For example, glycodelin in human tissues exists with two main glycoforms: glycodelinA (GdA), which is fully sialylated, and glycodelinS (GdS), which lacks sialic acid, but possesses numerous fucose residues.[10] It is proposed that the negatively charged sialic acid residues of GdA activate the T-cell suppressing activity of this glycoprotein, whereas the neutral glycans of GdS merely mask the active site of the amino acid backbone and therefore prevent immune suppression.[11-12] The addition of sialic acid residues to *N*-linked glycans is particularly important in glycoprotein therapeutic construction as these residues have been shown to not only promote activity, but increase *in vivo* circulation half-life as well,

such is the case with erythropoietin.[13-14] These examples demonstrate that a glycoprotein developed for pharmaceutical applications must possess the *N*-linked glycan motifs that will support the desired pharmacological properties.

Researchers interested in radically changing glycan structures of glycoproteins, to control pharmacological properties, currently use methods that modify the intracellular glycosylation machinery of the expression host to include non-native glycosylating enzymes (glycoenzymes) into the cell's Golgi apparatus.[15-16] Recently Hamilton, *et al.*, engineered the yeast *Pichia pastoris* to produce glycoproteins with biantennary, complex *N*-linked glycans.[17] This is a seminal advance that shows yeast expression systems could be used to generate glycoproteins with "human-like" glycosylation; however, the glycans consist of biantennary, complex *N*-linked structures, and lack the higher branching observed in some naturally occurring mammalian glycoproteins.[18] Another obstacle to pursuing this type of intracellular glycosylation remodeling includes genetically engineering a non-native glycoenzyme to localize correctly within the cellular machinery. This task is not trivial, and the cellular location for some glycosylating enzymes hasn't been determined precisely.[19] The majority of glycoenzymes are type II membrane proteins that need to become anchored into the Golgi membrane.[16] This localization event is based on the *N*-terminus of the enzyme and is not always predictable.[20] Assuming the new glycoenzyme is properly localized, there may be additional genetic work that must be done, such as silencing the genes for competing glycoenzymes, and some gene disruptions cause the host to die.[21] While intracellular glycosylation remodeling is a powerful and important approach to engineering glycoproteins with optimal glycosylation, it still has many barriers to cross

before it can be implemented routinely and cost effectively. For these reasons, we propose a complementary approach to designing glycoprotein therapeutics with novel glycosylation; this strategy is described in **Figure 1**.

1. Glycan Remodeling

1. Reaction I
2. Reaction II

2. Compositional Analysis

3. Structural Analysis

4. Functional Assay

+ Required reactants → Function of native glycoprotein

+ Required reactants → Function of improved glycoprotein
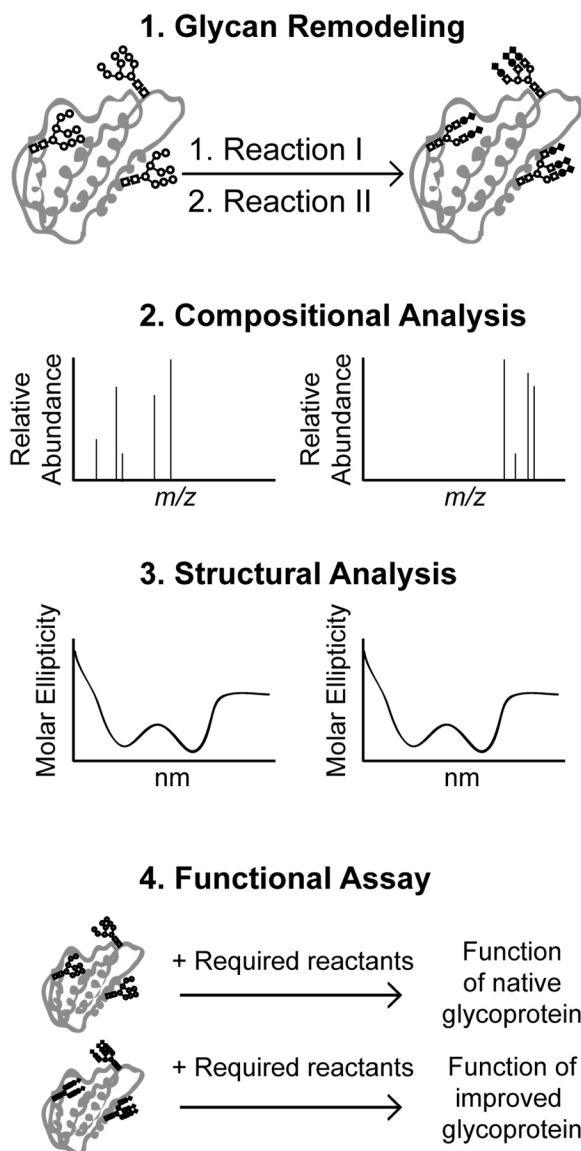
Figure 1: Extracellular enzymatic glycan optimization work-flow for glycoprotein therapeutic agents.

In this new approach, glycoenzymes, and their required cofactors, are used to optimize glycans of glycoproteins in a step-wise fashion after protein expression and purification has been completed. The basic premise for extracellular enzymatic glycan optimization has been demonstrated previously. Most notably, it was completed by Witte, *et al.*, who engineered a sialyl Lewis X moiety on the single glycosylation site of ribonuclease B in place of its natural high mannose glycan.[22] We implemented a similar enzymatic approach as the first step in our work-flow, and follow it with structural and functional assays on the newly engineered protein (Figure 1). Direct mass spectrometric (MS) analysis ensures the desired glycan modifications occurred, while circular dichroism (CD) spectroscopic analysis is useful in monitoring for disruptions to the secondary structure of the protein portion. The final step in the work-flow consists of monitoring the glycoprotein's function to ascertain that the desired activity is indeed retained, since any glycoprotein modified for therapeutic use must verifiably retain its activity.

The study presented here uses ribonuclease B (RNase B) as a model glycoprotein to demonstrate the use of enzymatic glycan trimming methods for the production of structurally and functionally conserved protein. This protein is a useful model system that allows straightforward analysis of structure and function. RNase B is a 15 kDa enzyme that contains a single *N*-linked glycosylation site at N^{34}, and consists of both α-helices and β-strands in its structure.[23-24] Removal of mannose residues from the high mannose *N*-linked glycans on RNase B was performed with an α-mannosidase isolated from Jack Bean.[25-26] The removal of mannose residues from high-mannose *N*-linked glycans may prove to be a desirable modification, since glycoproteins with high

mannose glycans have a short *in vivo* circulation half-life.[27-28] The native and modified

glycoprotein samples were then analyzed with respect to composition, structure, and

function. A work-flow diagram of the functional assay is shown in **Figure 2.** The

retention of activity is the cornerstone for developing useful glycoprotein therapeutic

agents with designer glycosylation. The study described herein supports previous work

that extracellular glycosylation remodeling is a useful approach to generating

glycoproteins with custom-made glycosylation; and this is the first demonstration that

enzymatic glycan trimming methods are viable for generating glycoproteins for

pharmaceutical applications.
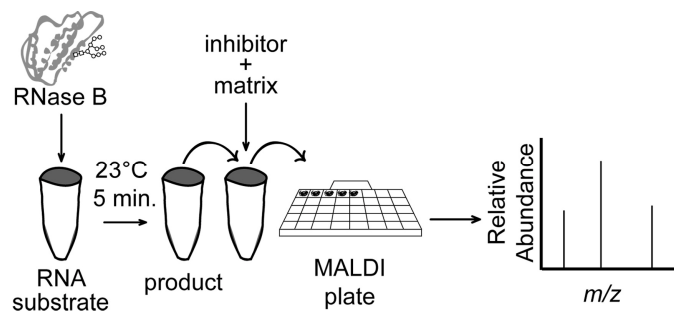
Figure 2. Novel RNase B activity scheme.

3.2 EXPERIMENTAL PROCEDURES

3.2.1 Materials

HPLC grade methanol was obtained from Thermo Fisher Scientific (Fairlawn, NJ), RNA substrate was obtained from Integrated DNA Technologies (Coralville, IA), and sodium bicarbonate was purchased from Fluka (Milwaukee, WI). Water was purified using the Millipore Direct Q-3 system (Millipore, Billerica, MA), and all other reagents were purchased from Sigma Aldrich (St. Louis, MO).

3.2.2 Glycan Modification

Approximately 500 µg of bovine pancreatic RNase B was added to 300 µL of 10 mM NH_4HCO_3 buffer, and the solution was adjusted to pH 5.0 using HCl. A commercially available α-mannosidase was added in an enzyme:glycoprotein ratio of 1:1000 (mol/mol), and incubated for 72 hours at 37°C. A second, identical sample of RNase B was treated to the same conditions, except the mannosidase was not added.

3.2.3 Mass Spectrometry

A portion (150 µL) of the glycoprotein samples were subjected to DTT in excess for 1 hour at 60°C, then diluted to 5 µM using 80:20 MeOH:H_2O, and 0.5% acetic acid. MS data were collected using a Thermo Fisher Scientific (San Jose, CA) hybrid linear ion-trap Fourier transform ion cyclotron resonance mass spectrometer (LTQ-FT-ICR-MS). Direct infusion was performed at 2 µL/min for electrospray ionization in the positive ion mode. The spray voltage was set at 2.8 kV, and the N_2 sheath gas was set to 10.0

psi. Ion desolvation was aided by heating the ion transfer tube to a temperature of 250 °C. Resolution was set to 100,000 for the ion m/z 400. All parameters were adjusted using the instrument software Excalibur, version 2.0.

3.2.4 Circular Dichroism

Far UV circular dichroism measurements were made using a Jasco J-715 spectropolarimeter (Tokyo, Japan) with an attached circulating water bath. Portions of the native and the modified glycoproteins were diluted with water to a final concentration of 71µM, and 250 µL of sample was placed in a jacketed, quartz, 1.0 mm pathlength sample cell. Experiments were performed at 10 °C, with a sensitivity of 100 mdeg and scan speed of 20 nm/min. The experiments were conducted under constant nitrogen flow. Scans were performed from 260 to 190 nm in triplicate. All spectra were smoothed and baseline corrected by subtracting signal from background solutions prior to analysis. The publicly available DICHROWEB server was used, with the SELCON3 algorithm, to estimate the percent glycoprotein secondary structure for α-helices, β-strands, β-turns, and disordered structural components.[29-32]

3.2.5 Activity Assay

RNA substrate, 5'-GGUAG-3', was diluted to 0.1 mg/mL with RNase free water. A small amount of the RNA substrate was mixed with native RNase B, in an enzyme:substrate ratio of 1:10,000 (mol/mol) and allowed to react for 5 minutes. The reaction was quenched using 0.1% (v/v) diethyl pyrocarbonate.[33] One µL of product was combined with one µL of MALDI matrix (15 mg/mL dihydroxybenzoic acid in 80:20

acetonitrile:water), spotted, and allowed to dry at 60 °C for thirty minutes. The reaction

was repeated using the modified RNase B sample. Mass spectral data were collected in

negative ion mode using a quadrupole ion trap mass spectrometer (LCQ Advantage,

Thermo Fisher Scientific, San Jose, CA) and an atmospheric pressure matrix assisted

laser desorption ionization (AP-MALDI) source (MassTech, Columbia, MD). The ion

transfer tube was heated to 250 °C to assist with desolvation of ions. The nitrogen laser

was operated at 337 nm, with a 4 nsec pulse width, and controlled using AP Maldi

Target software, version 3.4. The voltage applied to the MALDI plate was set to 3.0 kV.

A total of 333 laser shots were averaged into one spectrum prior to data analysis. Data

analysis was performed using the instrument software Excalibur, version 1.3.

3.3 RESULTS

3.3.1 Enzymatic Glycan Modification and Verification by MS

Proof of principle experiments, outlined in the work-flow depicted in Figure 1, were conducted using RNase B as a model glycoprotein. RNase B has five naturally occurring glycoforms, ranging from five to nine mannose residues (man_5-man_9), at its single *N*-linked glycosylation site.[34-35] RNase B was reacted with α-mannosidase to trim the glycans. This extracellular glycoenzyme reaction completed the first step in the glycan optimization method.

The second step in Figure 1 is to verify the effectiveness of the enzymatic reaction by obtaining glycan compositional data for the glycoprotein. The model glycoprotein's single *N*-linked glycosylation site allowed for direct analysis using mass spectrometric techniques. This was accomplished by analyzing the whole glycoprotein with electrospray ionization in the positive ion mode, using an FTICR-MS. RNase B contains four disulfide bonds, and if reduced, charge states are observed well within the range the required *m/z* range.[36] Reduction of disulfide bonds was performed with DTT in excess. Upon analysis of the reduced native and modified glycoproteins, a distribution of charge states, +8 through +11, were observed. (See insets of **Figure 3**). The +9 charge state peaks for both the native and modified RNase B samples are labeled with measured *m/z* values and respective glycoform structures in Figures 3a and 3b. The predicted and observed *m/z* values for glycoforms are listed in **Table 1**. The five naturally occurring glycoforms of the native glycoprotein were observed in all charge states detected, with man_5 accounting for the largest peak for the native glycoprotein.

Man$_5$, observed at *m/z* 1656.69 in Figure 3a, has been described as the most abundant glycoform species for RNase B, reportedly composing nearly half of the population.[37] A man$_4$ glycoform was present in the commercially available sample, as confirmed by additional mass spectrometric analysis, performed using the same procedure described in section 3.2.3.
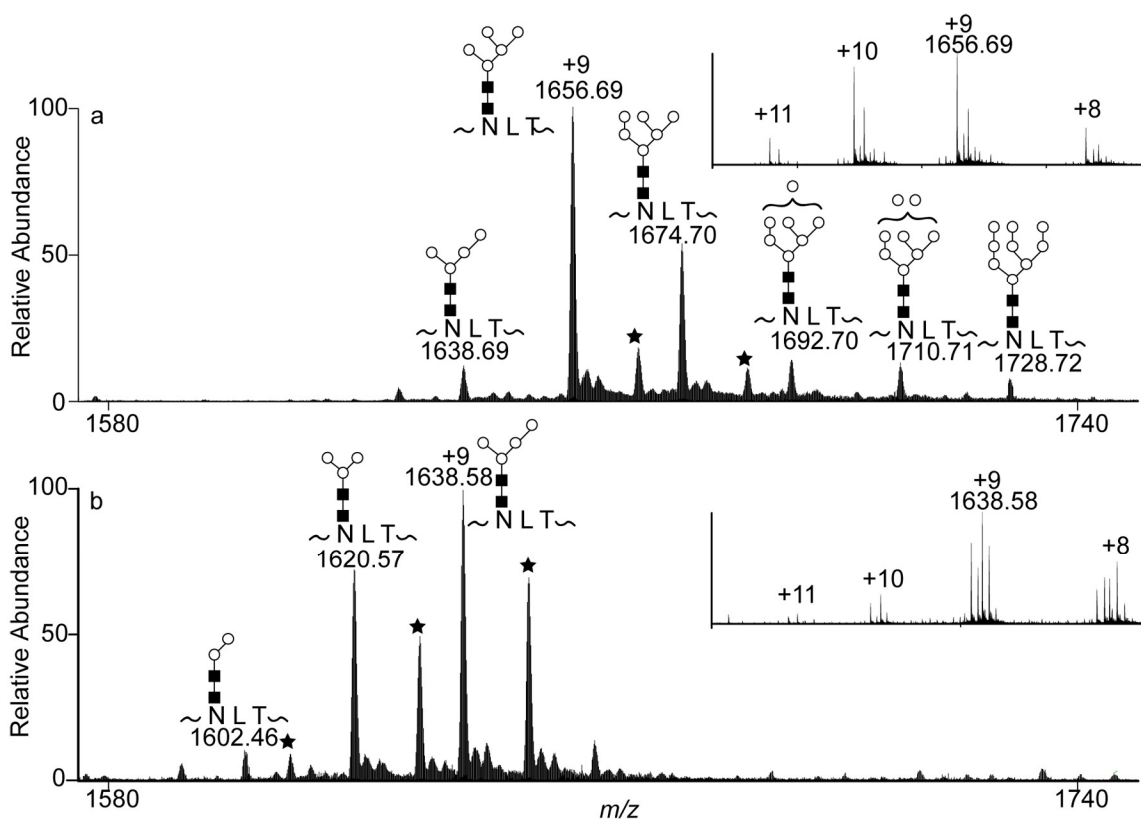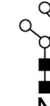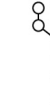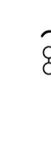
Figure 3. FT-ICR-MS data for the +9 charge state on intact RNase B. Insets: full MS data showing charge state distribution. Peaks are labeled with *m/z* values for the most abundant isotopes and glycan compositions for each of the protein's glycoforms. Symbols: ★ = sodiated adducts, ■ = N-acetylglucosamine, and O = mannose. (a) Native RNase B. (b) Modified RNase B, treated with α-mannosidase from Jack Bean for 72 hours, at 37 °C, and pH 5.0. The glycoforms Man$_5$-Man$_9$ are not observed in the modified glycoprotein.

The glycoforms observed in the modified RNase B sample were expected to differ from the native protein by possessing fewer mannose residues, due to the reaction with the exoglycosidase, α-mannosidase. This glycoenzyme was expected to

sequentially trim terminal mannose residues from the glycan structures of RNase B, based on characterization of its specific action.[25,26,38] Mass spectrometry experiments were repeated using the modified RNase B sample, and results are depicted in Figure 3b. In the +9 charge state, three glycoforms of the modified glycoprotein are detected: man_2 at m/z 1602.45, man_3 at m/z 1620.57, and man_4 at m/z 1638.57. These peaks are labeled with their assigned glycan structures, based on the calculation of exact mass for each species. The predicted and observed m/z values for the modified glycoprotein peaks in the +9 charge state are displayed (Table 1) and values for both the fully protonated and the significant sodiated adducts are included. Both spectra contained sodiated adducts of species, and these are indicated with a symbol (★).

Table 1: Predicted and observed m/z values for native and modified RNase B in the +9 charge state[a]

RNase B Glycoform Structure	Man_2	Man_3	Man_4	Man_5	Man_6	Man_7	Man_8	Man_9
Predicted m/z	1602.07	1620.08	1638.08	1656.09	1674.09	1692.10	1710.11	1728.11
Observed[N,P] m/z	n/a	n/a	1637.90	1655.92	1673.92	1691.93	1709.95	1728.06
Observed[M,P] m/z	1601.90	1619.91	1637.91	n/a	n/a	n/a	n/a	n/a
Predicted m/z	1609.55	1630.52	1648.52	1666.52	1684.59	1702.54	1720.54	1738.55
Observed[N,S] m/z	n/a	n/a	n/a	1666.47	1684.59	n/a	n/a	n/a
Observed[M,S] m/z	1609.23	1630.68	1648.58	n/a	n/a	n/a	n/a	n/a

[a]Symbols: ■ = N-acetylglucosamine, O = mannose, N = native glycoprotein, P = protonated only species, M = modified glycoprotein, S = sodiated adducts, and n/a = not applicable -- these ions were not detected.

3.3.2 Secondary Structural Measurements

The third step in the work-flow, outlined in Figure 1, consists of monitoring glycoproteins for changes in protein structure. Monitoring for these changes is a necessary step, because a functional assay alone may not be a reliable indicator of an intact glycoprotein; since some species with alterations in protein folding can still be enzymatically active.[39-40] CD spectroscopic analysis was performed on the native and modified glycoprotein samples in the far UV wavelength region (260-190 nm) to monitor for differences in secondary structure. The data points were plotted in overlay mode, as shown in **Figure 4**, and although both samples produced broad signals, an indication of an αβ protein, a slight increase in signal over the α helical wavelength region (~209 nm) is present in the modified glycoprotein sample.
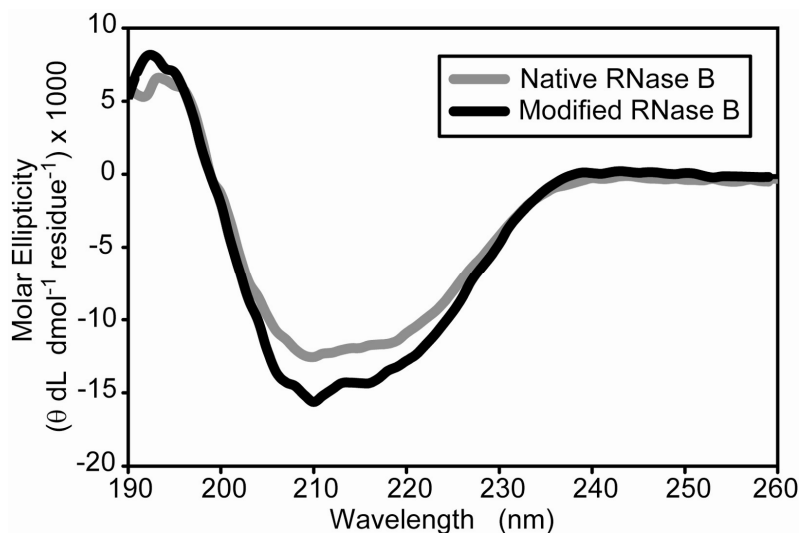
Figure 4: Far-UV CD data of overlaid spectra for native and modified RNase B glycoprotein. The modified protein retains very similar secondary structure to the native protein.

The results were analyzed using the web-based DICHROWEB server,[29-32] and the percent secondary structure estimation results are presented in **Table 2** as

calculated by the server. Native RNase B was estimated to have α-helices accounting for 12% and 10% of its overall structure, and 17% and 10% of its structure was attributed to β-strands. The modified RNase B was characterized as containing 18% and 13% α-helical and 14% and 9% β-stranded structures. The rest of the protein was estimated to contain turns and disordered structure. For each of the features identified for this protein, the percent contributions are similar for the native and modified structure, and the absolute differences in the values are within the margin of error of the measurement. These data demonstrate that the modified protein had similar secondary structure to the native protein.

Table 2. Percent secondary structure of RNase B as predicted by DICHROWEB.

RNase B	Helix	Strand	Turns	Unordered	Total
Native	22% ± 1	27% ± 1	21% ± 1	31% ± 1	101%
Modified	31% ± 8	23% ± 4	18% ± 2	27% ± 3	99%

3.3.3 RNase B Functional Assay

The final step in the glycan optimization work-flow includes a functional assay that compares the function of the modified glycoprotein to that of the native glycoprotein. This step is imperative if a glycoprotein with modified glycosylation is to become an active pharmaceutical agent. The model glycoprotein chosen for this study, RNase B, is known to catalyze the cleavage of RNA substrates via hydrolysis of the 3',5'-phosphodiester bonds. The modified RNase B sample's ability to catalyze RNA depolymerization was not expected to be negatively affected by enzymatic glycan trimming, as the removal of mannose residues is thought to promote RNase B activity.[26]

In addition, the nonglycosylated analog to RNase B, RNase A, is known to also perform similar functions.[41] Never-the-less, the functional assay is still a critical component of the work flow, because the *sample handling conditions* necessary to achieve the glycosylation remodeling could, in principle, deactivate the protein. To demonstrate retention of activity after extracellular glycan modification, an RNA substrate with the sequence 5'-GGUAG-3' was reacted with the native glycoprotein, and then the experiment was repeated with the modified glycoprotein, illustrated in Figure 2. The products produced from cleavage of the RNA substrate were detected using AP-MALDI, coupled to a quadrupole ion trap mass spectrometer. In **Figure 5**, AP-MALDI-MS data acquired in the negative ion mode are shown for the intact RNA substrate (Figure 5a), along with the RNA substrate after reacting with the native RNase B (Figure 5b). Figure 5c shows the RNA after reacting with the RNase B that has undergone enzymatic glycan trimming.

The base peak in Figure 5a, at *m/z* 1607, represents the intact substrate with sequence 5'-GGUAG-3', and its sodiated adducts, for example at *m/z* 1629, are labeled with a symbol (★). In Figure 5b, the products of the native RNase B reaction are illustrated. The peaks at *m/z* 611 and *m/z* 995 indicate that the RNA substrate was cleaved between the uridine and adenosine substituents, leaving the two nucleotides with base compositions AG and GGU respectively. No intact substrate was detected in either experiment as indicated by the arrows in Figure 5b and 5c. The data in Figure 5c matched the data in Figure 5b, thus demonstrating that the native and the modified glycoproteins were both effective at cleaving this particular RNA substrate, under the

conditions tested. These results indicate that the glycoprotein RNase B can undergo the extracellular enzymatic glycan trimming process, while retaining its desired function.
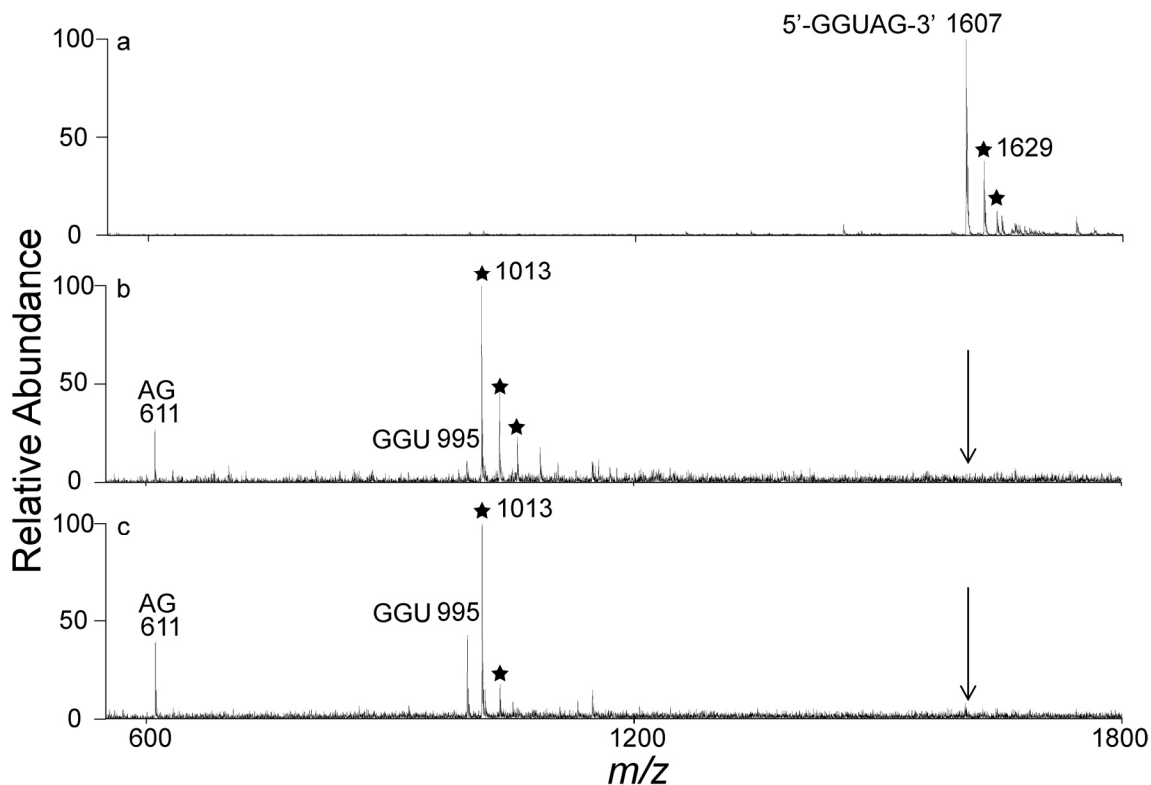
Figure 5: AP-MALDI-MS data of (a) intact RNA substrate with the sequence 5'-GGUAG-3', (b) nucleotide products from RNA substrate digestion, after reaction with native RNase B and (c) nucleotide products from RNA substrate digestion, after reaction with modified RNase B. The MS data shows that the RNase cleavage reaction was complete for both the native RNase B and modified RNase B. Symbols: ★ = sodiated adducts.

3.4 DISCUSSION

A model glycoprotein, RNase B, was chosen to demonstrate the utility of extracellular enzymatic glycan trimming. RNase B is an ideal protein for these studies because of its small size, consisting of 124 amino acids, and its single *N*-linked glycosylation site, at N^{34}. Additionally, the glycosylation on this protein is well characterized. The asparagine is linked to one of five glycoforms that consist of five to nine mannose residues with man_5 accounting for nearly half the naturally occurring population.[34-37] The glycan modification reaction, removal of mannose residues, is very relevant to pharmaceutical development, since high mannose glycans are known to contribute to a short *in vivo* circulation half life; thus enzymatic removal may be a desirable first step in glycan remodeling.[27,28]

Using the RNase B protein as a model, we demonstrated a new approach for development of designer glycoprotein therapeutics with custom-made glycosylation. The approach capitalizes on previous research that has demonstrated extracellular glycosylation remodeling of whole proteins is possible[22,25], and we implement that principle in the first step of a four-step work-flow for therapeutic glycoprotein development and characterization. In our approach, MS analysis of the whole protein (Step 2 in the work-flow) was an efficient method of characterizing the success of the enzymatic reaction. Direct analysis by ESI-FT-ICR-MS served as a useful way to monitor the enzymatic reaction and to optimize the reaction conditions, since no purification, and minimal sample preparation, are required for this analytical approach. Mass spectral analysis has been shown to be a very reliable method to characterize protein glycosylation, and even very large proteins, with multiple glycosylation sites, can

be studied using this approach, if a protease digestion step is incorporated so that glycopeptides are analyzed.[42-46]

In addition to modifying and characterizing the glycosylation composition, the process of extracellular glycosylation remodeling for therapeutic glycoprotein development must also address the need for validating that the protein structure and function are retained. Herein, we have implemented both a structural assay, using CD, and a novel functional assay that monitors the enzymatic cleavage of RNA. While the structural and functional assays could take on various forms, depending on the glycoprotein target of interest, these components are important steps in validating that the modified glycoprotein is still useful for its intended purpose.

In the structural experiments, CD was useful for identifying that the modified protein structure was generally very similar to that of the native protein, although a slight signal increase occurred in the α-helical region, at wavelength 209 nm, for the modified glycoprotein. Perturbations to RNase B's secondary structure with changes in glycan composition have been reported in the literature; however, any observed changes were described as being statistically insignificant.[38,47-49] Although the CD data illustrates an increase in α-helical character for the modified glycoprotein, the calculations for secondary structural characterization are within the margin of error of the data for the native protein, and thus are not considered significant.

The final and most critical aspect of extracellular glycosylation remodeling for therapeutic glycoprotein development is assessing the function of the modified glycoprotein. For RNase B, as well as for other enzymes, an enzymatic assay is an obvious choice for functional validation. We chose to detect the nucleotide products of

RNase B's enzymatic reaction with an RNA substrate using AP-MALDI-MS. MALDI

techniques are an advantageous detection strategy because they use very little sample

(1 µL in this case); they afford rapid analysis, and provide high sensitivity and selectivity.

Additionally, AP-MALDI ionization, which is used in these experiments, has shown to be

an even "softer" ionization source than conventional MALDI; thus the RNA was not

expected to fragment during the ionization process.[50] Other *in vitro* (or *in vivo*) methods

of assessing function are also possible, as well as other detection methods for

analyzing the results. In our assay, the AP-MALDI-MS data clearly and rapidly revealed

that RNase B, after enzymatic glycan trimming, retained its function, compared to the

unmodified protein.

3.5 CONCLUDING REMARKS

The studies described above lay the groundwork for using extracellular glycosylation remodeling in therapeutic glycoprotein development. This approach could show significant advantages to intracellular glycosylation remodeling (which is currently the gold standard in optimizing protein glycosylation), because the enzymatic conditions can be more carefully controlled and optimized when they are conducted on the lab bench, as opposed to inside the cell. We are currently pursuing this strategy in synthesizing several glycoprotein-based pharmaceuticals with custom-designed glycosylation profiles.

3.6 REFERENCES

1. Walsh, G. Biopharmaceuticals: recent approvals and likely directions. *Trends Biotechnol.* **2005,** *23,* 553-558.

2. Helenius, A.; Aebi, M. Roles of *N*-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* **2004,** *73,* 1019-1049.

3. Grogan, M. J.; Pratt, M. R.; Marcaurelle, L. A.; Bertozzi, C. R. Homogenous glycopeptides and glycoproteins for biological investigation. *Annu. Rev. Biochem.* **2002,** *71,* 593-634.

4. Tomiya, N.; Narang, S.; Lee, Y. C.; Betenbaugh, M. J. Comparing *N*-glycan processing in mammalian cell lines to native and engineered lepidopteran insect cell lines. *Glycoconjugate J.* **2004,** *21,* 343-360.

5. Hamilton, S. R.; Gerngross, T. U. Glycosylation engineering in yeast: The advent of fully humanized yeast. *Curr. Opin. Biotechnol.* **2007,** *18,* 387-392.

6. Narhi, L. O.; Arakawa, T.; Aoki, K. H.; Elmore, R.; Rohde, M. F.; Boone, T.; Strickland, T. W. The effect of carbohydrate on the structure and stability of erythropoietin. *J. Biol. Chem.* **1991,** *266*, 23022-23026.

7. Arnold, U.; Ulbrich-Hofman, R. Kinetic and thermodynamic thermal stabilities of ribonuclease A and ribonuclease B. *Biochemistry.* **1997,** *36,* 2166-2172.

8. Xu, G.; Narayan, M.; Scheraga, H. A. The oxidative folding rate of bovine pancreatic ribonuclease is enhanced by a covalently attached oligosaccharide. *Biochemistry.* **2005,** *44,* 9817-9823.

9. Takeuchi, M.; Takasaki, S.; Shimada, M.; Kobata, A. Role of sugar chains in the *in vitro* biological activity of human erythropoietin produced in recombinant Chinese hamster ovary cells. *J. Biol. Chem.* **1990,** *265*, 12127-12130.

10. Seppala, M.; Koistinen, H.; Koistinen, R.; Chiu, P. C. N.; Yeung, W. S. B. Glycosylation related actions of glycodelin: Gamete, cumulus cell, immune cell and clinical associations. *Hum. Reprod. Update.* **2007,** *13,* 275-287.

11. Mukhopadhyay, D.; SundarRaj, S.; Alok, A.; Karande, A. A. Glycodelin A, not glycodelin S, is apoptotically active – Relevance of sialic acid modification. *J. Biol. Chem.* **2004,** *279,* 8577-8584.

12. Jayachandran, R.; Radcliffe, C. M.; Royle, L.; Harvey, D. J.; Dwek, R. A.; Rudd, P. M. ; Karande, A. A. Oligosaccharides modulate the apoptotic activity of glycodelin. *Glycobiology.* **2006,** *16,* 1052-1063.

13. Goldwasser, E.; Kung, C. K. H.; Eliason, J. Mechanism of erythropoietin-induced differentiation. XIII. The role of sialic acid in erythropoietin action. *J. Biol. Chem.* **1974,** *249*, 4202-4206.

14. Tsuda, E.; Kawanishi, G.; Ueda, M.; Masuda, S.; Sasaki, R. The role of carbohydrate in recombinant human erythropoietin. *Eur. J. Biochem.* **1990,** *188*, 405-411.

15. Sethuraman, N.; Stadheim, T. A. Challenges in therapeutic glycoprotein production. *Curr. Opin. Biotechnol.* **2006,** *17,* 341-346.

16. Czlapinski, J. L.; Bertozzi, C. R. Synthetic glycobiology: Exploits in the Golgi compartment. *Curr. Opin. Chem. Biol.* **2006,** *10,* 645-651.

17. Hamilton, S. R.; Davidson, R. C.; Sethuraman, N.; Nett, J. H.; Jiang, Y.; Rios, S.; Bobrowicz, P.; Stadheim, T. A.; Li, H.; Choi, B.-K.; Hopkins, D.; Wischnewski, H.; Roser, J.; Mitchell, T.; Strawbridge, R. R.; Hoopes, J.; Wildt, S.; Gerngross, T. U. Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science.* **2006,** *313,* 1441-1443.

18. Bobrowicz, P.; Davidson, R. C.; Li, H.; Potgieter, T. I.; Nett, J. H.; Hamilton, S. R.; Stadheim, T. A.; Miele, R. G.; Bobrowicz, B.; Mitchell, T.; Rausch, S.; Renfer, E.; Wildt, S. Engineering of an artificial glycosylation pathway blocked in core oligosaccharide assembly in the yeast *Pichia pastoris*: Production of complex humanized glycoproteins with terminal galactose. *Glycobiology.* **2004,** *14,* 757-766.

19. Ferrara, C.; Brunker, P.; Suter, T.; Moser, S.; Puntener, U.; Umana, P. Modulation of therapeutic antibody effector functions by glycosylation engineering: Influence of Golgi enzyme localization domain and co-expression of heterologous β1,4-*N*-acetylglucosaminyltransferase III and Golgi α-mannosidase II. *Biotechnol. Bioeng.* **2006,** *93,* 851-861.

20. Wildt, S.; Gerngross, T. U. The humanization of *N*-glycosylation pathways in yeast. *Nat. Rev. Microbiol.* **2005,** *3,* 119-128.

21. Nett, J. H.; Gerngross, T. U. Cloning and disruption of the *PpURA5* gene and construction of a set of integration vectors for the stable genetic modulation of *Pichia pastoris. Yeast.* **2003,** *20,* 1279-1290.

22. Witte, K.; Sears, P.; Wong, C.-H. Enzymatic glycoprotein synthesis: preparation of ribonuclease glycoforms via enzymatic glycopeptides condensation and glycosylation. *J. Am. Chem. Soc.* **1997,** *119,* 2114-2118.

23. Xu, G.; Zhai, H.; Narayan, M.; McLafferty, F. W.; Scherage, H. A. Simultaneous characterization of the reductive unfolding pathways of RNase B isoforms by top-down mass spectrometry. *Chem. Biol.* **2004,** *11,* 517-524.

24. Wang, S.; Wang, P. G.; Qi, Q. Influence of substrate conformation on the deglycosylation of ribonuclease B by recombinant yeast peptide: *N*-glycanase. *Acta. Biochim. Biophys. Sin.* **2007,** *39,* 8-14.

25. Tarelli, E.; Byers, H. L.; Wilson, M.; Roberts, G.; Homer, K. A.; Beighton, D. Detecting mannosidase activities using ribonuclease B and Matrix-Assisted LASER Desorption/Ionization-Time of Flight mass spectrometry. *Anal. Biochem.* **2000,** *282,* 165-172.

26. Rudd, P. M.; Joao, H. C.; Coghill, E.; Fiten, P.; Saunders, M. R.; Opdenakker, G.; Dwek, R. A. Glycoforms modify the dynamic stability and functional activity of an enzyme. *Biochemistry.* **1994,** *33,* 17-22.

27. Mistry, P. K.; Wraight, E. P.; Cox, T. M. Therapeutic delivery of proteins to macrophages: Implications for treatment of Gaucher's disease. *Lancet.* **1996,** *348,* 1555-1559.

28. Wright, A.; Morrison, S. L. Effect of altered C_H2-associated carbohydrate structure on the functional properties and *in vivo* fate of chimeric mouse-human immunoglobulin G1. *J. Exp. Med.* **1994,** *180,* 1087-1096.

29. Sreerama, N.; Woody, R.W. A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal. Biochem.* **1993,** *209,* 32-44.

30. Lobley, A.; Whitmore, L.; Wallace, B. A. DICHROWEB: An interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics.* **2002,** *18,* 211-212.

31. Lees, J. G.; Miles, A. J.; Wien, F.; Wallace, B. A. A reference database for circular dichroism spectroscopy covering fold and secondary structure space. *Bioinformatics.* **2006,** *22,* 1955-1962.

32. Whitmore, L.; Wallace, B. A. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic. Acids. Res.* **2004,** *32,* W668-W673.

33. Leonard, N.; J.; McDonald, J. J.; Reichmann, M. E. Reaction of diethyl pyrocarbonate with nucleic acid components, I. Adenine. *Proc. Natl. Acad. Sci. USA.* **1970,** *67,* 93-98.

34. Temporini, C.; Perani, E.; Calleri, E.; Dolcini, L.; Lubda, D.; Caccialanza, G.; Massolini, G. Pronase-immobilized enzyme reactor: An approach for automation in glycoprotein analysis by LC/LC-ESI/MS^n. *Anal. Chem.* **2007,** *79,* 355-363.

35. Pitchayawasin, S.; Isobe, M. Mass spectrometric assignment of Smith degradation glycopeptides derived from ribonuclease B. *Biosci. Biotechnol. Biochem.* **2004,** *68* 1424-1433.

36. Reid, G. E.; Stephenson, J. L.; McLuckey, S. A. Tandem mass spectrometry of ribonuclease A and B: *N*-linked glycosylation site analysis of whole protein ions. *Anal. Chem.* **2002,** *74,* 577-583.

37. Rudd, P. M.; Scragg, I. G.; Coghill, E.; Dwek, R. A. Separation and analysis of the glycoform populations of ribonuclease B using capillary electrophoresis. *Glycoconjugate J.* **1992,** *9,* 86-91.

38. Joao, H. C.; Dwek, R. A. Effects of glycosylation on protein structure and dynamics in ribonuclease B and some of its individual glycoforms. *Eur. J. Biochem.* **1993,** *218,* 239-244.

39. White, F. H. Studies on the relationship of disulfide bonds to the formation and maintenance of secondary structure in chicken egg white lysozyme. *Biochemistry.* **1982,** *21,* 967-977.

40. Laity, J. H.; Shimotakahara, S.; Scheraga, H. A. Expression of wild-type and mutant bovine pancreatic ribonuclease A in *Escherichia coli. Proc. Natl. Acad. Sci. USA.* **1993,** *90,* 615-619.

41. Opitz, J. G.; Ciglic, M. I.; Haugg, M.; Trautwein-Fritz, K.; Raillard, S. A.; Jermann, T. M.; Benner, S. A. Origin of the catalytic activity of bovine seminal ribonuclease against double-stranded RNA. *Biochemistry.* **1998,** *37,* 4023-4033.

42. Go, E. P.; Irungu, J.; Zhang, Y.; Dalpathado, D. S.; Liao, H.-X.; Sutherland, L. L.; Alam, S. M.; Haynes, B. F.; Desaire, H. Glycosylation site-specific analysis of HIV envelope proteins (JR-FL and CON-S) reveals major differences in glycosylation site occupancy, glycoform profiles, and antigenic epitopes' accessibility. *J. Proteome Res.* **2008,** *7,* 1660-1674.

43. Zhang, Y.; Go, E. P.; Desaire, H. Maximizing coverage of glycosylation heterogeneity in MALDI-MS analysis of glycoproteins with up to 27 glycosylation sites. *Anal. Chem.* **2008,** *80,* 3144-3158.

44. Dalpathado, D. S.; Irungu, J.; Go, E. P.; Butnev, V. Y.; Norton, K.; Bousfield, G. R.; Desaire, H. Comparative glycomics of the glycoprotein follicle stimulating hormone: Glycopeptide analysis of isolates from two mammalian species. *Biochemistry.* **2006,** *45,* 8665-8673.

45. Go, E. P.; Rebecchi, K. R.; Dalpathado, D. S.; Bandu, M. L.; Zhang, Y.; Desaire, H. GlycoPep DB: A tool for glycopeptide analysis using a smart search. *Anal. Chem.* **2007,** *79,* 1708-1713.

46. Irungu, J.; Go, E. P.; Dalpathado, D. S.; Desaire, H. Simplification of mass spectral analysis of acidic glycopeptides using GlycoPep ID. *Anal. Chem.* **2007,** *79,* 3065-3074.

47. O'Conner, S. E.; Imperiali, B. Modulation of protein structure and function by asparagine-linked glycosylation. *Chem. Biol.* **1996,** *3,* 803-812.

48. Wormald, M. R.; Dwek, R. A. Glycoproteins: Glycan presentation and protein-fold stability. *Structure.* **1999,** *7,* R155-R160.

49. Williams, R. L.; Greene, S. M.; McPherson, A. The crystal structure of ribonuclease B at 2.5 Å resolution. *J. Biol. Chem.* **1987,** *262,* 16020-16031.

50. Laiko, V. V.; Baldwin, M. A.; Burlingame, A. L. Atmospheric pressure matrix-assisted laser desorption/ionization mass spectrometry. *Anal. Chem.* **2000,** *72,* 652-657.

CHAPTER 4

OPTIMIZING SIALIC ACID CONTENT VIA GLYCAN TRIMMING:

RETAINING THERAPEUTIC GLYCOPROTEIN THERMAL STABILITY

Optimized protein modifications, such as N-linked glycosylation structures, are necessary to attain desired pharmacological properties of some biopharmaceuticals. Sialic acid residues and derivatives, including N-glycolylneuraminic acid (Neu5Gc), are glycan building blocks often incorporated into the N-linked glycans during production of therapeutic glycoproteins. Neu5Gc is problematic because it is immunogenic to humans, and can induce the formation of antibodies to complex type N-linked glycan epitopes. Attempts at preventing Neu5Gc incorporation have focused on gene silencing within the expression host, or avoiding expression host systems known to produce large amounts of Neu5Gc, such as mouse-derived NS0 cells. We propose an alternate method to remove Neu5Gc that employs enzymatic glycan trimming. Four sialylated glycoproteins, including the FDA approved glycoprotein therapeutic, erythropoietin, were treated with a sialidase to trim off all sialylated glycans, including Neu5Gc. Glycopeptide analysis using mass spectrometry verified sialic acid removal. Both the native and modified proteins were also subjected to thermal stress studies, to determine melt temperatures, T_m, using far UV circular dichroism. The desialylated glycoproteins had higher T_m values, compared to the native (sialylated) proteins, for all four proteins studied. The T_m data indicates that the method of enzymatic glycan trimming can be implemented into glycoprotein production work-flows, without harming protein thermal stability. Additionally, glycan trimming could be used on therapeutic glycoproteins to prevent adverse immunogenic responses induced by Neu5Gc exposure.

4.1 INTRODUCTION

Pharmacological properties of therapeutic glycoproteins are heavily influenced by post-translational modifications (PTMs) such as N-linked glycosylation.[1-3] Complex type N-linked glycans that terminate in sialic acid residues have shown to be crucial to the efficacy of some therapeutic products, such as recombinant human erythropoietin (rhEPO).[4] Erythropoietin, used for treating anemia, is also an endogenous protein in human serum that circulates to the bone marrow, where it is necessary for erythrocyte progenitor cells to differentiate for the production of erythrocytes.[5] rhEPO exhibits complex behavior in relation to sialic acid residues, in that removal of these terminal residues from the N-linked glycans will produce an increased binding affinity for the EPO receptor found on the erythrocyte progenitor cells.[6-8] Yet, without sialic acid , the exposed galactose residues remaining on the N-linked glycans are bound by the liver's galactose binding protein, rhEPO is quickly cleared from circulation, and has no *in vivo* activity.[9-12] Sialic acid residues are thus an important part of therapeutic glycoprotein composition.

More than one sialic acid structure is known to exist in non-human mammalian cells. The most common form of sialic acid, N-Acetylneuraminic acid (Neu5Ac), is added to the terminal positions on complex type N-linked glycans by the glycan processing enzymes within Golgi of the expression hosts.[13,14] A sialic acid derivative, N-Glycolylneuraminic acid (Neu5Gc), can also be added to the glycans. Neu5Gc differs in structure from Neu5Ac by just a single hydroxyl group, and is immunogenic to humans.[15] Neu5Gc is not found in healthy human adults,[16] and its introduction into the human body can instigate an immune response where antibodies against complex N-

linked glycan epitopes are produced.[17] More than one neutralizing antibody has been reported against Neu5Gc, including the anti-Paul-Bunnell antibody and the anti-Hanganutziu-Deicher antibody.[18-19] Some patients receiving rhEPO therapy for anemia have developed such a severe immune response to conformational epitopes of rhEPO complex glycans, their anemia worsened severely and they became dependent upon higher risk treatments, including blood transfusions.[20-23]

Optimizing the N-linked glycans of therapeutic glycoproteins is, therefore, an important part of preventing adverse responses to this entire class of therapeutics. The amount of Neu5Gc varies by expression host cell line, accounting for up to 50% of sialic acid species in mouse derived NS0 cells, and 7-15% of sialic acid in Chinese hamster ovary (CHO) cell lines.[24-25] Due to requirements such as sialylation, the majority of currently approved glycoprotein therapeutics are produced in CHO cells and not in the mouse derived NS0.[26] Efforts in preventing Neu5Gc being added to therapeutic glycoproteins rely on a method using anti-sense RNA technology to reduce production of the sialic acid derivative, but only a reduction in Neu5Gc content has been reported.[27] To ensure the safety of glycoprotein therapeutic products, alternate methods are needed.

A method for removing specific glycans from existing N-linked glycan structures in glycoproteins has been demonstrated recently; the method (glycan trimming) cleaved mannose monosaccharides from a glycoprotein, without altering the protein's structure or function.[28] If this method were also applicable to the removal of Neu5Gc from glycoproteins, this approach could open up a new pathway for producing therapeutic glycoproteins with reduced immunogenicity.

To investigate this application of glycan trimming, recombinant human erythropoietin (rhEPO) was expressed in CHO cells, purified using lectin affinity, and then subjected to trimming using Sialidase A, to remove all the sialic acid residues. The success of the trimming reaction was verified by mass spectrometry. Both native and enzymatically modified proteins were subjected to thermal stress while using circular dichroism (CD) to monitor the loss of secondary structure. A diagram of the work-flow is provided in **Scheme 1.** After promising results with rhEPO were obtained, the glycan trimming method was also applied to three commercially available sialylated glycoproteins, to determine the generality of the method. This is the first study to demonstrate that glycoprotein therapeutic products do not lose their thermal stability when glycan trimming methods are used to remove sialic acid residues, and this approach offers a strategy for completely removing immunogenic glycan structures on therapeutic glycoproteins.

Scheme 1. Experimental work-flow for removing sialic acid residues from the N-linked glycans of therapeutic glycoproteins and the characterization methods that follow glycan optimization. Symbols: □ N-Acetylglucosamine, ● galactose, ○ mannose, ◆ N-Acetylneuraminic acid, ◇ N-Glycolylneuraminic acid.
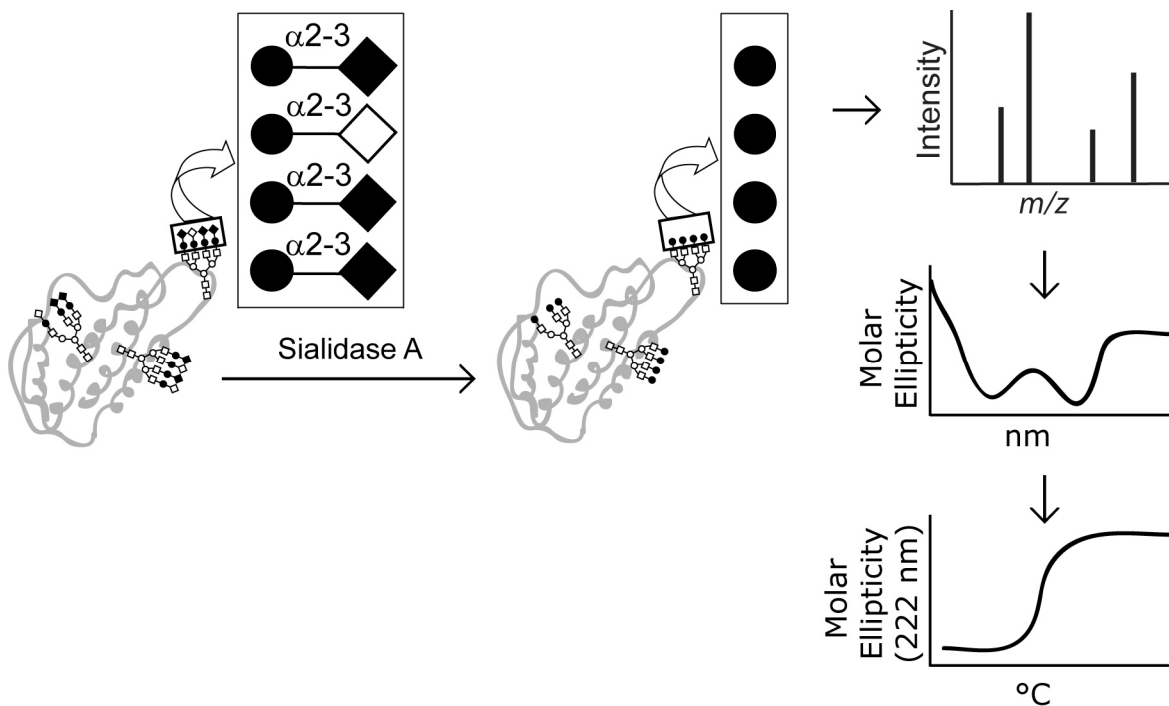
4.2 EXPERIMENTAL PROCEDURES

4.2.1 Materials

The CHO-K1 cells were a gift from Dr. Jeff Krise (University of Kansas,

Department of Pharmaceutical Chemistry, Lawrence, KS). Water was purified using the

Millipore Direct Q-3 system (Millipore, Billerica, MA). All other materials were obtained

from Sigma Aldrich (St. Louis, MO) unless otherwise specified. All solutions were

passed through a 0.2 µm filter prior to use.

4.2.2 Cloning and Construction of the Expression Vector for rhEPO

The cDNA of human erythropoietin (hEPO) was obtained from ATCC

(NM_000799) (Manassas, VA). The hEPO gene was amplified from this template in a

thermocycler using the following primers: 5'- GGACTTTCCAAAATGTCG -3' (forward

primer) and 5'-

CGCGCG<u>TCTAGA</u>TCA**GTGGTGGTGGTG**TCTGTCCCCTGTCATGCAGGC-3' (reverse

primer with <u>XbaI</u> restriction site and **His4-tag**), (Integrated DNA Technologies, Inc.,

Coralville, IA). Due to the lack of a compatible restriction site in hEPO, the forward

primer corresponds to an upstream sequence embedded in the template plasmid. Use

of the upstream sequence permits inclusion of a close, in-frame EcoR1 site for use in

subcloning of the amplified gene product. Four histidine residues were added to the C-

terminal end of the sequence, preceding the stop codon, as an affinity tag for use with

IMAC purification. The amplified insert and the pCMV6-XL5 vector (Origene

Technologies Inc., Rockville, MD) were digested with the corresponding EcoRI and XbaI

endonucleases. Each DNA reaction was separated, visualized with ethidium bromide and purified from agarose gels, and the insert and vector were ligated together by overnight incubation with T4 Ligase (Promega, Madison, WI) at 16 °C. The ligation reaction was then transformed into competent BL21(DE3) *E. coli* (Novagen, Gibbstown, NJ) using a standard heat shock protocol. Transformed colonies were obtained by overnight growth at 37 °C on selective LB agar plates containing 100 µg/mL ampicillin. Individual colonies were selected and used to inoculate 10 mL Luria Broth (LB) containing 100 µg/mL ampicillin. Liquid cultures were grown overnight in an incubating shaker at 37 °C. Plasmid DNA was isolated from each of the overnight cultures using the QIAprep Spin Miniprep Kit (QIAGEN, Valencia, CA). The isolated DNA samples were quantified using UV absorbance at 260 and examined for purity using 260/280 nm. Constructs were checked for the incorporation of an insert of the correct size; each was incubated with the aforementioned nucleases and separated using agarose gel electrophoresis. DNA constructs were verified to be correct by bidirectional sequencing (Northwoods DNA Inc., Bemidji, MN) using the forward (5'-GGACTTTCCAAAATGTCG-3') and reverse (5'-ATTAGGACAAGGCTGGTGGG-3') primers included with the pCMV6 vector from Origene Technologies Inc.

4.2.3 Cell Culture and Transfection

CHO-K1 cells were seeded and maintained using complete high glucose Dulbecco's Modified Eagle Medium (DMEM) (Thermo Scientific Hyclone, Logan, UT). To the medium, supplemental MEM Non-Essential Amino acids 100x (ATCC, Manassas, VA), 10% fetal bovine serum (Mediatech, Inc. Manassas, VA), 100X

penicillin/streptomycin (ATCC, Manassas, VA), and 150 mM L-Proline were added.

Cells were maintained in T-75 flasks (BD, Franklin Lakes, NJ) with 24 mL of medium

and kept at 37 °C and 5% CO_2. Passage was performed using two wash steps with pre-

warmed phosphate buffered saline (10 % PBS) alone and then PBS containing trypsin

(0.25% trypsin with 0.1% EDTA) to remove the adherent cells from the surface of the

flask. After three to five minutes, the suspension was centrifuged at 3800 x g for 2

minutes in a 15 mL tube, the supernatant was removed via vacuum, and the cells were

resuspended via gentle pipetting into 6 mL of the culture medium described above, then

reseeded.

Cells were allowed to grow to 50% confluency and then transfection was

performed with 4 µg of hEPO DNA and 20 µL of Turbofectin 8.0 transfection reagent

(OriGene Technologies, Inc., Rockville, MD) per T-75 flask. The supernatant was

decanted after 24 hours and the medium replaced. After an additional 24 hours, the

supernatant was collected and concentrated to 0.5 mL using centrifugal filtration

devices (Millipore, Billerica, MA) with a 10 kDa molecular weight cut off (MWCO) at

9000 x g.

4.2.4 Lectin Affinity Chromatography (LAC) Purification

A cartridge (5-mL) containing *Maackia amurensis* leukoagglutinin (MAL) lectin

resin (Qiagen, Valencia, CA) was equilibrated according to the manufacturer's

directions using the indicated binding/wash buffer. The concentrated supernatant (0.5

mL) containing the rhEPO glycoprotein, from a single supernatant collection, was mixed

with an equal volume of binding/wash buffer and loaded onto the column at 1

mL/minute. The column was washed using 50 mL of binding/wash buffer at 4 mL/min, and the rhEPO was eluted into 50 mL of elution buffer containing 200 mM lactose. The elution fraction was concentrated to approximately 1 mL as described above. The rhEPO protein content was quantified using a standard Bradford assay.

4.2.5 SDS-PAGE Analysis

SDS-PAGE 4-12 % Bis-TRIS Midi gels (Invitrogen, Carlsbad, CA) were used for electrophoretic analysis. The rhEPO protein sample was concentrated to approximately 10 mg/mL using a molecular weight cut off filter, mixed with 2X reducing Laemmli loading buffer, and boiled for 10 minutes. The samples were loaded, and electrophoresis was performed in MES buffer (Invitrogen) per manufacturer's directions. Precision Plus Protein Dual Color stained protein ladder (Bio-Rad, Hercules, CA) was utilized in a separate lane for molecular weight approximation. Gels were stained using Coomassie (R-250) in excess, overnight. Destaining was accomplished with 20% acetic acid in 50/50 methanol/H_2O until background was clear.

4.2.6 Enzymatic Glycan Trimming

Approximately 300 µg of each commercially available sialylated glycoprotein, human alpha-1-acid glycoprotein (AGP1), human serotransferrin (TF), and ovine follicle stimulating hormone (oFSH), were dissolved in 10 mM TRIS-HCl, pH 6.0, to a final protein concentration of 2 mg/mL. Four µL of Sialidase A (Prozyme, Hayward, CA) were added to the protein solution and then incubated at 37 °C for 48 hours. The enzyme:glycoprotein ratio used was 1:1500 (mol/mol) for AGP1 and oFSH, and 1:660

for TF. The purified erythropoietin (rhEPO) was treated in the same manner, with a ratio of enzyme:glycoprotein at 1:1500 (mol/mol). Sialylated protein samples were made to the same protein concentration, and 4 µL of water were added in place of enzyme.

4.2.7 Mass Spectrometric Verification of Glycan Removal

Aliquots of glycoprotein samples were exposed to either 4 M or 6 M urea, reduced using dithiothreitol (DTT) at 15 mM, and then incubated at 37 °C for one hour. Iodoacetamide (IAA) was added for alkylation at 25 mM, and the samples were incubated in the dark at room temperature for one hour. DTT was again added to neutralize any excess IAA, for a final DTT concentration of 40 mM. Solutions were then diluted using 10 mM TRIS-HCl until urea concentration was approximately 1.5 M and the pH was adjusted to 7.5. Sequencing grade trypsin (Promega, Madison, WI) was added in a 1:30 (wt/wt) enzyme:glycoprotein ratio. Protease digestion was allowed to proceed for 18 hours at 37 °C and stopped by adding 1 µL of glacial acetic acid for every 100 µL of digested glycoprotein. Samples were stored at -20 °C until analysis.

Full mass spectrometric scans (MS[1]) and tandem mass spectrometry scans (MS[2]) of the peptide/glycopeptide digestion samples were collected by performing online high performance liquid chromatography electrospray ionization mass spectrometric (HPLC/ESI-MS) analysis using a hybrid, linear ion trap Fourier transform ion cyclotron resonance mass spectrometer (LIT-FTICR-MS), (Thermo Fisher Scientific, San Jose, CA). Chromatography was controlled using the UltiMate capillary LC system (Dionex, Sunnyvale, CA) and FAMOS autosampler. LC mobile phase was (A) 99.9% H_2O and (B) 99.9% acetonitrile, both with 0.1% formic acid. Flow rate was set to 5

µL/min for the capillary column (C_{18} 300 µm i.d. × 100 mm, 300 Å, by Micro-Tech, Vista, CA), onto which 5 µL of digested product was loaded. MS^1 scans were collected with the spray voltage set to 2.8 kV in positive ion mode, ion transfer tube heated to 200 °C, and nebulizing gas (N_2) at 10 psi. The resolution was set to 100,000 for the ion *m/z* 400. The MS^2 data was collected using data dependent mode where the 6 most intense ions in the MS^1 scans were selected for fragmentation using collision induced dissociation (CID). Additional parameters for MS^2 experiments included using a 2 Da isolation width, activation q = 0.25, and collision energy = 30%, as defined by the Xcalibur software (version 2.0, Thermo Fisher Scientific, San Jose, CA). A wash and a blank run were included in between each sample analysis. Data were interpreted with the aid of GlycoPepDB[29] and GlycoPepID[30].

4.2.8 Circular Dichroism Studies

The commercially available proteins were diluted using 10 mM TRIS-HCl to a final concentration of 0.8 or 1.0 mg/mL, and the rhEPO samples were dialyzed using 3 one L buffer exchanges, into 10 mM TRIS-HCl, final concentration at 0.5 mg/mL as determined using a standard Bradford assay. The protein samples were placed in a jacketed, quartz, 1.0-mm pathlength sample cell and sealed tightly. Far UV CD measurements were made using a Jasco J-715 spectropolarimeter (Tokyo, Japan) with an attached circulating water bath that was built in-house. Experiments were performed with a sensitivity of 100 mdeg and scan speed of 10 nm/min, under constant nitrogen flow. Background solutions contained 4 µL of enzyme for the desialylated (trimmed) samples. Thermal unfolding experiments and replicate studies were completed using

either the Jasco J-715 spectropolarimeter or a Jasco J-810 spectropolarimeter while monitoring changes in ellipticity at 222 nm, in response to increasing temperature (25-90 °C). Temperature was increased using an attached Peltier temperature controller. Scans were collected using a data pitch of 0.2°C, at a rate of 12°C/hour. Scan speed was 10 nm/min and a 2 second response time was allowed. The transition temperature was calculated using a sigmoidal fit algorithm within Origin (version 6.0), and the sialylated and de-sialylated samples were compared for each protein. A total of 5 separate preparation and analysis experiments were performed for each commercially available protein. Three replicates were performed on rhEPO.

4.3 RESULTS

Some sialic acid derivatives on therapeutic glycoproteins are known to be immunogenic to humans.[15,17] Enzymatic glycan trimming was employed to remove the sialic acids from a therapeutic glycoprotein, recombinant human erythropoietin (rhEPO), using a sialidase. Mass spectrometric analysis was used to verify that the glycan trimming reaction was successful, then thermal melt studies were performed using far UV CD to determine if enzymatic glycan trimming inflicts damage to the thermal stability of the therapeutic glycoprotein. The thermal scans indicated the trimmed glycoproteins were more stable than the native sample, which was unexpected. Three commercially available, sialylated glycoproteins were then studied using the same methods to ensure that the improved thermal stability of desialylated glycoproteins was not an unique feature to rhEPO.

4.3.1 SDS-PAGE Analysis of Recombinant Human Erythropoietin

In **Figure 1**, the results from SDS-PAGE analysis on the purified and concentrated rhEPO glycoprotein are shown. In the lane containing the expressed protein, a large, round band appears between the 37 and 50 kDa protein standards, and there are no other significant bands present in this concentrated protein sample. The gel was over-loaded with protein intentionally, so that any small impurities would also be concentrated and detected; however, no other significant bands appeared. This result indicates that a purified protein is present. The molecular weight is expected to be between 37-40 kDa for the recombinant form of EPO, which is heavily glycosylated and sialylated when produced in CHO cells.[31] The rhEPO possesses three N-linked glycan

sites, and one commonly utilized O-linked glycan, and these heavily branched and sialylated glycans contribute to the carbohydrates comprising 40% of the glycoprotein's molecular weight.[31,32]
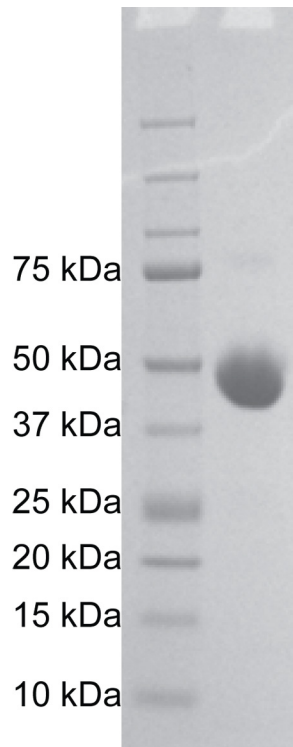
Figure 1. SDS-PAGE analysis of recombinant human erythropoietin at high concentration. No other significant bands are present.

4.3.2 Enzymatic Glycan Trimming and Verification

After purification and dialysis, a portion of the rhEPO sample was treated with a commercially available sialidase to remove the sialic acids. To confirm that the sialidase catalyzed the removal of sialic acids from the N-linked glycans of rhEPO, mass spectrometric analysis of glycopeptides was conducted. To conduct these experiments, the glycoproteins were digested using proteomics grade trypsin, which produces a mixture of peptides and glycopeptides. Separation of the mixture was achieved by

performing online liquid chromatography during LC/ESI-MS analysis. **Figure 2a** displays a full (MS) mass spectrum obtained from analyzing the sialylated form of rhEPO, with a peak at m/z 1547.18 in the 4+ charge state, corresponding to the mass of a glycopeptide of rhEPO containing one missed cleavage within the peptide portion, and three sialic acid residues. The peptide and N-linked glycan composition is shown near the peak in Figure 2a. Tandem mass spectrometric experiments were also performed on this peak. The product ions provide additional confirmation of the glycan composition. In Figure 2b, the detected product ions are shown, the predominant fragmentation pattern includes the successive loss of sialic acid residues. This fragmentation pattern is typical of multiply charged glycopeptides, as demonstrated previously.[30,33]

The desialylated rhEPO sample was analyzed in the same way. In Figure 2c, a full mass spectrum is shown using the same m/z window as the sialylated sample. In this spectrum, no sialylated glycopeptide is detected. We ruled out the possibility of this peak being absent because of slight chromatographic drifts in retention time by also searching for this peak in the MS data at ± 5 minutes from the retention time of the peak in Figure 2A, but even when searching for the sialylated glycopeptide peak with an expanded retention time window, the peak at m/z 1547.18 is not observed. Instead, a peak at m/z 1328.84 in the 4+ charge state is present, corresponding to the same peptide and N-linked glycan, except the sialic acids are no longer present. See Figure 2d. The modified N-linked glycan composition obtained by using enzymatic glycan trimming via sialidase is indicated near the peak in the figure. After confirmation of sialic

acid removal, the sialylated and desialylated rhEPO samples were subjected to thermal stress and monitoring using circular dichroism (CD).
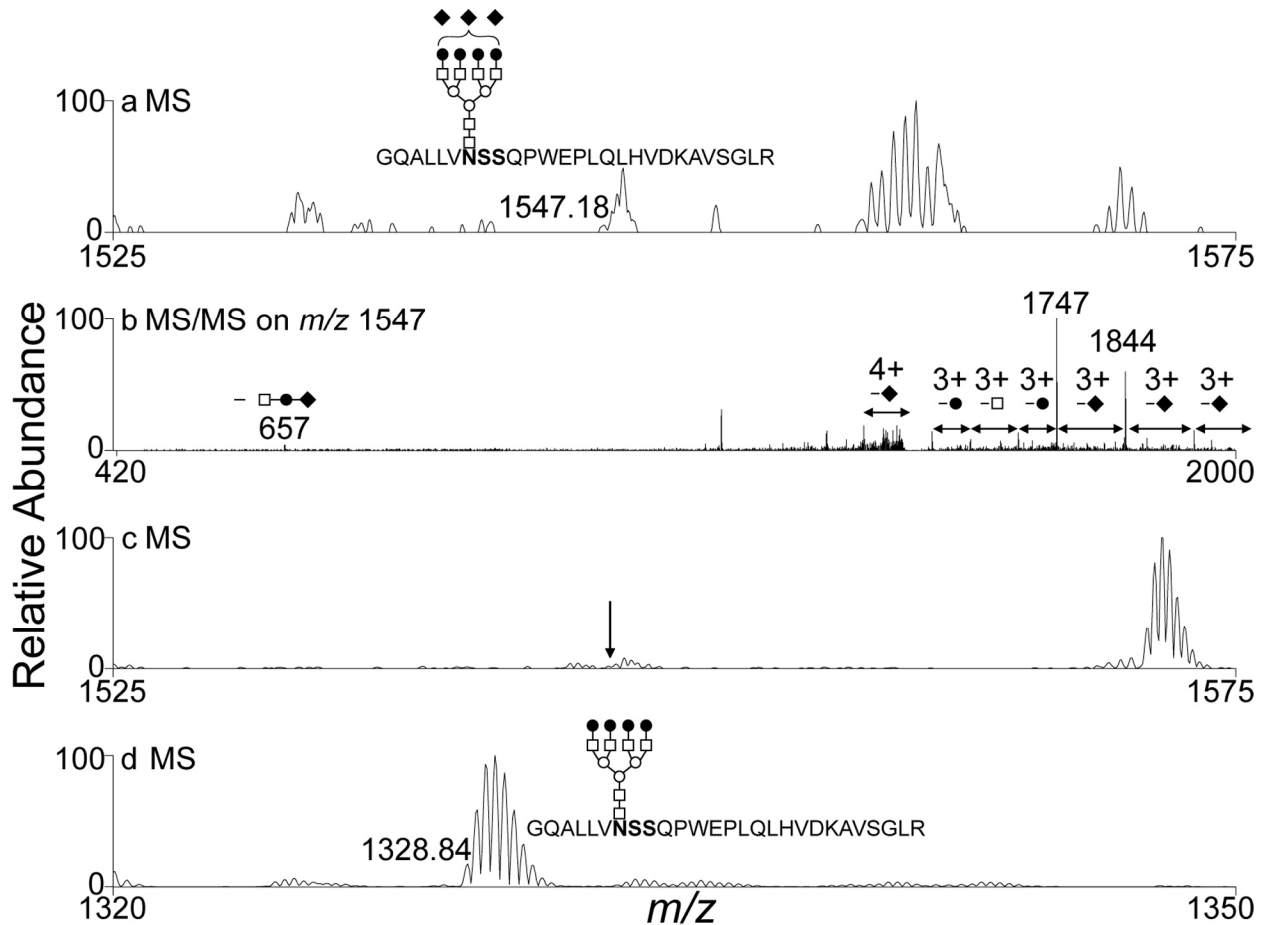
Figure 2. Glycopeptide analysis of erythropoietin. (a) Full mass spectrum with a glycopeptide in the 4+ charge state. Glycan composition is shown next to the monoisotopic peak, m/z 1547.18. (b) Tandem MS (MS/MS) was used to fragment the peak m/z 1547 and fragments produced in the 4+ and the 3+ charge states indicate the loss of three sialic acid residues from the N-linked glycan. (c) In the desialylated erythropoietin sample, the m/z 1547.18 is no longer present, at the same retention time or ± 5 minutes, indicating this sialylated glycopeptide species is no longer present. Instead, a peak at m/z 1328.84 is present in the 4+ charge state in spectrum (d), indicating that the glycopeptide is present, except the sialic acids have been enzymatically removed. Symbols: □ N-Acetylglucosamine, ● galactose, ○ mannose, ◆ N-Acetylneuraminic acid.

4.3.3 Circular Dichroism Monitoring Loss of Secondary Structure

The sialylated and desialylated rhEPO samples were subjected to increasing

temperature and monitored for the loss of secondary structure using Far UV CD at 222

nm. In **Figure 3a**, data is shown for a single preparation of rhEPO, where the data for

the sialylated form is shown in black, and the desialylated data is shown in gray. The

desialylated sample is noted to have a higher melt temperature (T_m) than the sialylated

rhEPO sample, where the transition occurs at a lower temperature. The experiments

were repeated in triplicate with separate sample preparations, and each time the

desialylated sample produced the higher T_m value. Average T_m values for the rhEPO

samples are given in **Table 1.**
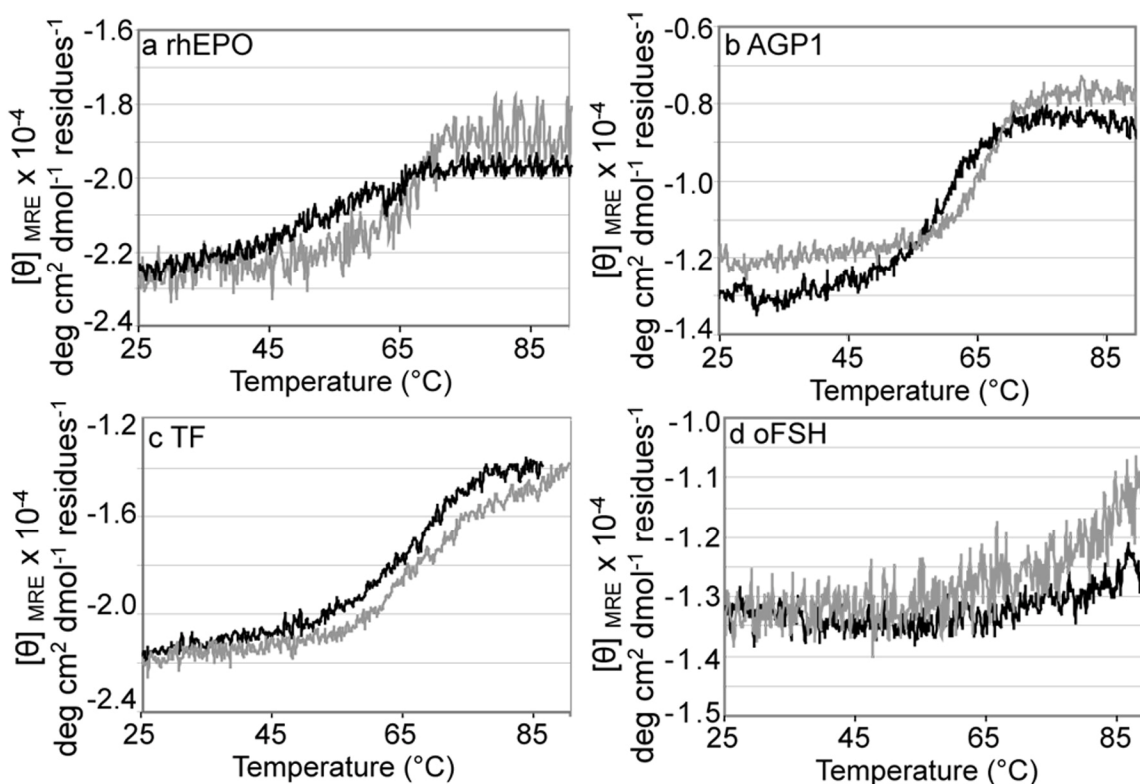
Figure 3. Temperature dependant loss of secondary structure monitored at 222 nm using Far UV CD, for (a) rhEPO, (b) AGP1, (c) TF, (d) oFSH Symbols: **----** sialylated glycoproteins and ----- desialylated glycoproteins.

To investigate the general applicability of the observation that enzymatic glycan trimming produces more stable proteins, when sialic acids are trimmed from sialylated glycoproteins, three commercially available glycoproteins were obtained and subjected to the same experiments as the rhEPO samples (enzymatic glycan trimming, followed by melt studies using CD). The three glycoproteins were human alpha-1-acid glycoprotein (AGP1), human serotransferrin (TF), and ovine follicle stimulating hormone (oFSH), and the results for these samples are given in **Figure 3b-3d.** In Figure 3b, the sialylated AGP1 scans produced a lower T_m value than the desialylated samples. Experiments were repeated until N=5, and average T_m results are given in Table 1. The TF and oFSH results, Figure 3c and 3d respectively, also produced similar results, where in each study the transition occurred at the lower temperature for the sialylated samples.

Table 1. T_m values for each of the four sialylated glycoproteins, before and after enzymatic glycan trimming. N = 5 for AGP, TF, and oFSH, where each experiment was performed separately with individual sample preparations. N = 3 for rhEPO.

Protein	rhEPO	AGP1	TF	oFSH
Sialylated T_m (℃)	58	56	61	65
De-sialylated T_m (℃)	66	61	65	76

4.4 DISCUSSION

Removal of glycan residues using enzymatic glycan trimming has been demonstrated previously for mannose residues.[(Toumi)] This new study was designed to accomplish two goals: use the glycan trimming method to solve a therapeutic glycoprotein problem, and quantify any loss of thermal stability the glycoprotein might experience because of glycan trimming. The chosen therapeutic glycoprotein was recombinant human erythropoietin (rhEPO), which is used to treat anemia. The N-linked glycans of rhEPO terminate with sialic acid residues, but some derivatives of the sialic acid, especially the Neu5Gc species, are immunogenic to humans and can cause adverse responses to treatment.[20-23] We propose using enzymatic glycan trimming with a sialidase to remove sialic acids as an additional step in the therapeutic glycoprotein production work flow, to prevent adverse immunological effects of therapeutic glycoprotein administration.

The rhEPO was expressed in CHO cells, purified using lectin affinity, and then treated with a sialidase to remove all the sialic acid residues. The choice of enzyme for the glycan trimming work is important, because it is dependant upon the enzyme's specificity for both the targeted glycan's identity and linkage with other sugars. The sialidase chosen for this work, Sialidase A, is known to have broad specificity for sialic acid in the main form, Neu5Ac, as well as common derivatives such as Neu5Gc.[34] The linkage specificity of this glycosidase was also optimal for these experiments. The expression host glycosylation machinery determines the linkage of the sialic acid to the rest of the glycan, and it has been demonstrated that CHO cells produce sialic acids with an α-(2,3) linkage to the corresponding galactose residue, and not an α-(2,6),

which is more common in humans.[4,25,31] The linkage specificity for the activity of the chosen sialidase is known to be approximately the same for α-(2,3) and α-(2,6) linkages between sialic acids and galactose,[34] making Sialidase A useful choice for this study.

This study focused on enzymatic removal of sialic acid residues for use in optimizing glycans on therapeutic glycoproteins, and investigating what effects sialic acid removal may impart on glycoprotein stability. The rhEPO, with and without sialic acid, was subjected to thermal stress and monitored for loss of secondary structure using Far UV CD. Since sialic acids have been demonstrated to be very important to rhEPO *in vivo* activity,[9-12] it was predicted that removal of sialic acid could negatively impact the glycoprotein's thermal stability. Instead, the desialylated samples of rhEPO were observed to have a higher melt temperature (T_m) value than the unmodified glycoprotein. To determine if this was a property unique to rhEPO, three commercially available sialylated glycoproteins were obtained and subjected to the same experimental methods. In each case, the T_m values indicated that the sialic acid trimming method did not harm the thermal stability of the glycoproteins, but instead improved it. These results indicate that the trimming method can be implemented into therapeutic glycoprotein production methods without harming protein stability.

The study described above was designed to solve the known immunogenicity problems with sialic acid derivatives (such as Neu5Gc) present on therapeutic glycoproteins. Removal of Neu5Gc will improve immunogenicity profiles, but therapeutic products such as rhEPO do not exert a therapeutic effect without some form of sialic acid.[9-12] Methods to add the desired sialic acid onto N-linked glycan structures using glycosyltransferases have been demonstrated by Paulson *et al.*[35] Enzymatic glycan

trimming of sialic acid could be used in conjunction with the glycosyltransferase method to produce therapeutic glycoproteins with desirable sialic acid content. Another potential application for sialic acid glycan trimming could be to remove the α-(2,3) linked sialic acids found on N-linked glycans produced in CHO cells, in conjunction with glycosyltransferase methods to add only α-(2,6) linked sialic acids to IgG used to treat autoimmune disorders, as the α-(2,6) configuration of sialic acid on IVIG has been identified as responsible for anti-inflammatory properties.[36] Improving the functional activity of a therapeutic glycoprotein such as IVIG could lead to lower doses and less frequent administration.[36]

4.5 CONCLUDING REMARKS

To demonstrate the practicality of using enzymatic glycan trimming to remove sialic acids from therapeutic glycoprotein, rhEPO was generated and treated with a sialidase. During study of the glycoprotein thermal stability, unexpected results were observed, where the desialylated samples produced a higher melt temperature (T_m) than the sialylated form. To determine if this was a property unique to only one protein, three other sialylated glycoproteins were treated with a sialidase, and analyzed. The T_m values in all cases indicated that the sialic acid trimming method does not negatively impact the stability of the glycoprotein; in fact it improves the stability. These results indicate that the sialic acid glycan trimming method can be implemented without damaging the thermal stability of glycoprotein therapeutics such as rhEPO. Use of this method to remove immunogenic sugars will reduce the likelihood of adverse immunogenic responses to glycoprotein therapeutic treatments.

4.6 REFERENCES

1.   Li, H.; d'Anjou, M. Pharmacological significance of glycosylation in therapeutic proteins. *Curr. Opin. Biotechnol.* **2009,** *20,* 678-684.

2.   Walsh, G.; Jefferis, R. Post-translational modification in the context of therapeutic proteins. *Nat. Biotechnol.* **2006,** *24,* 1241-1252.

3.   Jenkins, N.; Parehk, R. B.; James, D. C. Getting the glycosylation right: Implications for the biotechnology industry. *Nat. Biotechnol.* **1996,** *14,* 975-981.

4.   Takeuchi, M.; Takasaki, S.; Shimada, M.; Kobata, A. Role of sugar chains in the *in vitro* biological activity of human erythropoietin produced in recombinant Chinese hamster ovary cells. *J. Biol. Chem.* **1990,** *265*, 12127-12130.

5.   Graber, S. E.; Krantz, S. B. Erythropoietin and the control of red cell production. *Ann. Rev. Med.* **1978,** *29,* 51-66.

6.   Tsuda, E.; Kawanishi, G.; Ueda, M.; Masuda, S.; Sasaki, R. The role of carbohydrate in recombinant human erythropoietin. *Eur. J. Biochem.* **1990,** *188*, 405-411.

7.   Imai, N.; Higuchi, M.; Kawamura, A.; Tomonoh, K.; Oh-eda, M.; Fujiwara, M.; Shimonaka, Y.; Ochi, N. Physicochemical and biological characterization of asialoerythropoietin. Suppressive effects of sialic acid in the expression of biological activity of human erythropoietin *in vitro. Eur. J. Biochem.* **1990,** *194*, 457-462.

8.   Narhi, L. O.; Arakawa, T.; Aoki, K. H.; Elmore, R.; Rohde, M. F.; Boone, T.; Strickland, T. W. The effect of carbohydrate on the structure and stability of erythropoieitin. *J. Biol. Chem.* **1991,** *266*, 23022-23026.

9.   Morell, A. G.; Gregoriadis, G.; Scheinberg, I. H.; Hickman, J.; Ashwell, G. The role of sialic acid in determining the survival of glycoproteins in the circulation. *J. Biol. Chem.* **1971,** *246*, 1461-1467.

10.   Goldwasser, E.; Kung, C. K. H.; Eliason, J. On the mechanism of erythropoietin-induced differentiation. XIII. The role of sialic acid in erythropoietin action. *J. Biol. Chem.* **1974,** *249*, 4202-4206.

11.   Dordal, M. S.; Wang, F. F.; Goldwasser, E. The role of carbohydrate in erythropoietin action. *Endocrinology.* **1985,** *116*, 2293-2299.

12.   Fukuda, M. N.; Sasaki, H.; Lopez, L.; Fukuda, M. Survival of recombinant erythropoietin in the circulation: The role of carbohydrates. *Blood.* **1989,** *73*, 84-89.

13. Grabenhorst, E.; Schlenke, P.; Pohl, S.; Nimtz, M.; Conradt, H. S. Genetic engineering of recombinant glycoproteins and the glycosylation pathway in mammalian host cells. *Glycoconjugate J.* **1999,** *16,* 81-97.

14. Czlapinski, J. L.; Bertozzi, C. R. Synthetic glycobiology: Exploits in the Golgi compartment. *Curr. Opin. Chem. Biol.* **2006,** *10,* 645-651

15. Noguchi, A.; Mukuria, C. J.; Suzuki, E.; Naiki, M. Immunogenicity of N-glycolylneuraminic acid-containing carbohydrate chains of recombinant human erythropoietin expressed in Chinese hamster ovary cells. *J. Biochem.* **1995,** *117,* 59-62.

16. Irie, A.; Koyama, S.; Kozutsumi, Y.; Kawasaki, T.; Suzuki, A. The molecular basis for the absence of N-glycolylneuraminic acid in humans. *J. Biol. Chem.* **1998,** *273*, 15866-15871.

17. Higashi, H.; Naiki, M.; Matuo, S.; Okouchi, K. Antigen of "serum sickness" type of heterophile antibodies in human sera: Identification as gangliosides with N-glycolylneuraminic acid. *Biochem. Biophys. Res. Commun.* **1977,** *79*, 388-395.

18. Merrick, J. M.; Zadarlik, K.; Milgrom, F. Characterization of the Hanganutziu-Deicher (serum-sickness) antigen as gangliosides containing N-glycolylneuraminic acid. *Int. Archs Allergy appl. Immun.* **1978,** *57*, 477-480.

19. Golaszewska, E.; Kurowska, E.; Duk, M.; Koscielak, J. Paul-Bunnell antigen and a possible mechanism of formation of heterophile antibodies in patients with infectious mononucleosis. *Acta Biochimica Polonica* **2003,** *50*, 1205-1211.

20. Bergrem, H.; Danielson, B. G.; Eckardt, K. U.; Kurtz, A.; Strisberg, M. A case of antierythropoietin antibodies following recombinant human erythropoietin treatment. In: *Erythropoietin: molecular physiology and clinical application*; Marcel Dekker: New York, 1993.

21. Pece, R.; de la Torre, M.; Alcazar, R.; Urra, J. M. Antibodies against recombinant human erythropoietin in a patient with erythropoietin-resistant anemia. *N. Engl. J. Med.* **1996,** *335,* 523-524.

22. Prabhakar, S.; Muhlfelder, T. Antibodies to recombinant human erythropoietin causing pure red cell aplasia. *Clin. Nephrol.* **1997,** *47,* 331-335.

23. Casadevall, N.; Nataf, J.; Viron, B.; Kolta, A.; Kiladjian, J.; Martin-Dupont, P.; Michaud, P.; Papo, T.; Ugo, V.; Teyssandier, I.; Varet, B.; Mayeux, P. Pure red-cell aplasia and antierythropoietin antibodies in patients treated with recombinant erythropoietin. *N. Engl. J. Med.* **2002,** *346*, 469-475.

24. Raju, T. S.; Briggs, J. B.; Borge, S. M.; Jones, A. J. S. Species-specific variation in glycosylation of IgG: Evidence for the species-specific sialylation and branch specific galactosylation and importance for engineering recombinant glycoprotein therapeutics. *Glycobiology.* **2000,** *10,* 477-486.

25. Baker, K. N.; Rendell, M. H.; Hills, A. E.; Hoare, M.; Freedman, R. B.; James, D. C. Metabolic control of recombinant protein *N*-glycan processing in NS0 and CHO cells. *Biotechnol. Bioeng.* **2001,** *73,* 188-202.

26. Higgins, E. Carbohydrate analysis throughout the development of a protein therapeutic. *Glycoconjugate J.* **2010,** *27,* 211-225.

27. Chenu, S.; Gregoire, A.; Malykh, Y.; Visvikis, A.; Monaco, L.; Shaw, L.; Schauer, R.; Marc, A.; Goergen, J. L. Reduction of CMP-*N*-acetylneuraminic acid hydroxylase activity in engineered Chinese hamster ovary cells using an antisense-RNA strategy. *Biochim. Biophys. Acta* **2003,** *1622*, 133-144.

28. Toumi, M. L.; Go, E. P.; Desaire, H. Development of fully functional proteins with novel glycosylation via enzymatic glycan trimming. *J. Pharm. Sci.* **2009**, *98,* 2581-2591.

29. Go, E. P.; Rebecchi, K. R.; Dalpathado, D. S.; Bandu, M. L.; Zhang, Y.; Desaire, H. GlycoPepDB: A tool for glycopeptide analysis using a "smart search". *Anal. Chem.* **2007**, *79*, 1708–1713.

30. Irungu, J.; Go, E. P.; Dalpathado, D. S.; Desaire, H. Simplification of mass spectral analysis of acidic glycopeptides using GlycoPep ID. *Anal. Chem.* **2007**, *79*, 3065–3074.

31. Skibeli, V.; Nissen-Lie, G.; Torjesen, P. Sugar profiling proves that human serum erythropoietin differs from recombinant human erythropoietin. *Blood.* **2001,** *98,* 3626-3634.

32. Takeuchi, M.; Takasaki, S.; Miyazaki, H.; Kato, T.; Hoshi, S.; Kochibe, N.; and Kobata, A. Comparative study of the asparagine-linked sugar chains of human erythropoietins purified from urine and the culture medium of recombinant Chinese hamster ovary cells. *J. Biol. Chem.* **1988,** *263,* 3657-3663.

33. Irungu, J.; Go, E. P.; Zhang, Y.; Dalpathado, D. S.; Liao, H.-X.; Haynes, B. F.; Desaire, H. Comparison of HPLC/ESI-FTICR MS versus MALDI-TOF/TOF MS for glycopeptide analysis of a highly glycosylated HIV envelope glycoprotein. *J. Am. Soc. Mass. Spectrom.* **2008,** *19,* 1209-1220.

34. Uchida, Y.; Tsukada, Y.; Sugimori, T. Enzymatic properties of neuraminidases from *Arthrobacter ureafaciends. J. Biochem. (Tokyo)* **1979,** *86,* 1573-1585.

35. Paulson, J. C.; Bayer, R. J.; Sjoberg, E. Practical *in vitro* Sialylation of Recombinant Glycoproteins. U.S. Patent 7,220,555 B2, May 22, 2007.

36. Anthony, R. M.; Nimmerjahn, F.; Ashline, D. J.; Reinhold, V. N.; Paulson, J. C.; Ravetch, J. V. Recapitulation of IVIG anti-inflammatory activity with a recombinant IgG Fc. *Science.* **2008,** *320,* 373-376.

CHAPTER 5

ENHANCING HUMAN GROWTH HORMONE STABILITY USING DESIGNER GLYCOSYLATION

Degradation pathways, such as asparagine deamidation, negatively impact the stability of therapeutic proteins like recombinant human growth hormone (rhGH), leading to decreases in efficacy, safety, and functionality. We have developed a novel approach to limiting protein deamidation by introducing a nonnative N-linked glycosylation site at asparagine residues that are known to readily deamidate. The sites are glycosylated by a eukaryotic expression host, and the glycan is enzymatically trimmed *in vitro* to contain a single monosaccharide residue. This moiety is designed to sterically hinder deamidation. To demonstrate this technology, three mutant, glycosylated rhGH proteins were produced. Circular dichroism (CD) was used to study secondary structure, and the proteins were found to possess a higher degree of alpha helical character after glycan trimming. The proteins' stability was also assessed by studying loss of secondary structure, under conditions that promote asparagine deamidation (using far UV CD). The superior stability of the glycosylated proteins was demonstrated by comparing melt temperature (T_m) values for the three rhGH mutants to T_m values obtained using a nonglycosylated growth hormone standard. Using this glycosylation engineering strategy to improve protein stability will lead to better therapeutic protein products.

5.1 INTRODUCTION

Recombinant DNA technology has become a vital tool for the pharmaceutical industry, allowing for mass production of therapeutic hormones and other proteins.[1] Despite impressive progress in biopharmaceutical formulation methods, some proteins continue to suffer from common degradation pathways, such as deamidation.[2,3] Recombinant human growth hormone (rhGH) is a prime example.[4]

Human growth hormone (hGH) is an endogenous protein required for proper growth.[5] Growth hormone deficiencies were previously treated using growth hormone extracted and purified from the pituitary glands of human cadavers.[6] The yield of hGH obtained from this source was insufficient to meet rising demand, while the extraction techniques used were overly complex.[6,7] This problem has been partly solved by treating growth hormone deficiencies with recombinant human growth hormone (rhGH).[8-10] Although administration of rhGH has proven beneficial, the protein is susceptible to chemical degradation (via deamidation in particular), and the recombinant protein is not stable.[4] As a result, the shelf life and circulation half-life of the medication is limited, and daily injections are required.[11] Optimizing the stability of proteins such as rhGH would be advantageous for more efficient drug storage, dosing schedules, and patient benefit. To accomplish the goal of improving the stability of rhGH, we designed a strategy to prevent the most common degradation pathway for this protein, deamidation.

Nonenzymatic deamidation is the primary means by which rhGH degrades, most often occurring at N149 and N152.[12-14] (Amino acid positions are described according to accession number P01241, *www.uniprot.org*.) Deamidation can be an important component in regulating biological systems, by acting as a "molecular clock".[15] In the

clinical setting, deamidation of therapeutic proteins has a negative affect on higher order structure that results in a decrease in the drug's activity, half-life *in vivo*, as well as shelf life.[4,12-14,16-19] When susceptible asparagines (N) deamidate, the amino acid is converted into aspartic acid (D).[20,21] The N to D shift is accompanied by the replacement of an NH_2 group with an OH , which increases the protein's molecular weight by one Dalton.[20] This modification causes further rearrangements in the protein, with the most destructive change, in terms of its effect on the protein structure and function, being the addition of an extra carbon to the peptide backbone of the protein.[20,21] Deamidation is associated with reduced rhGH stability and function[2,4,14] and induces protein aggregation, further hampering the pharmaceutical efficacy of rhGH.[22,23]

Previous efforts to block deamidation-based damage to proteins have focused on the chemical residues where the change occurs – asparagines. Chen, *et al.*, replaced vulnerable asparagine residues of glucoamylase, from *Aspergillus awamori*, using site directed mutagenesis.[24] Results were mixed, where enzyme secretion, thermal stability, and activity were impacted depending on the N residue position being mutated, and whether the introduced amino acid was Q or D.[24]

Instead of removing asparagine residues at these locations in rhGH, we sought to block or slow deamidation by adding an asparagine linked (N-linked) glycan to the vulnerable asparagine residues. Asparagine deamidation occurs through a cyclic imide intermediate, and demaidation can only occur if the amide nitrogen is not sterically constrained.[21,25] Therefore, it should be possible to prevent deamidation by adding a bulky group to the asparagine residue that will sterically hinder the formation of the intermediate species.

Incorporation of a glycan at the vulnerable asparagines was accomplished using site-directed mutagenesis to introduce the N-linked glycosylation consensus sequence, N-X-S/T, where X is any amino acid other than proline. The consensus sequence was attained by mutating the residues after the asparagine. The workflow is depicted in **Scheme 1**. The mutant proteins were expressed in Chinese Hamster Ovary (CHO-K1) cells that are known to recognize the N-linked consensus sequence as a signal to add complex glycans to the lengthening polypeptide chain.[27] The glycosylated mutant rhGH proteins were purified using lectin affinity chromatography and analyzed using SDS-PAGE. Glycans were trimmed enzymatically using endo-β-N-acetylglucosaminidase F (Endo F3), to ensure that a single N-acetylglucosamine was attached to the asparagine residue on the rhGH protein. The secondary structure of the resultant mutant proteins was studied using circular dichroism (CD), and thermal stability studies were performed under conditions that encourage deamidation.

Scheme 1. Discouraging deamidation using glycosylation. The amino acids near two asparagine residues known to readily deamidate are modified to include N-linked glycosylation sites. N-linked glycans are acquired during expression with a eukaryotic host cell line, purified using lectin affinity chromatography, and enzymatically trimmed to remove the majority of the glycan mass. The remaining monosaccharide residue is intended to sterically block the nonenzymatic deamidation reaction.
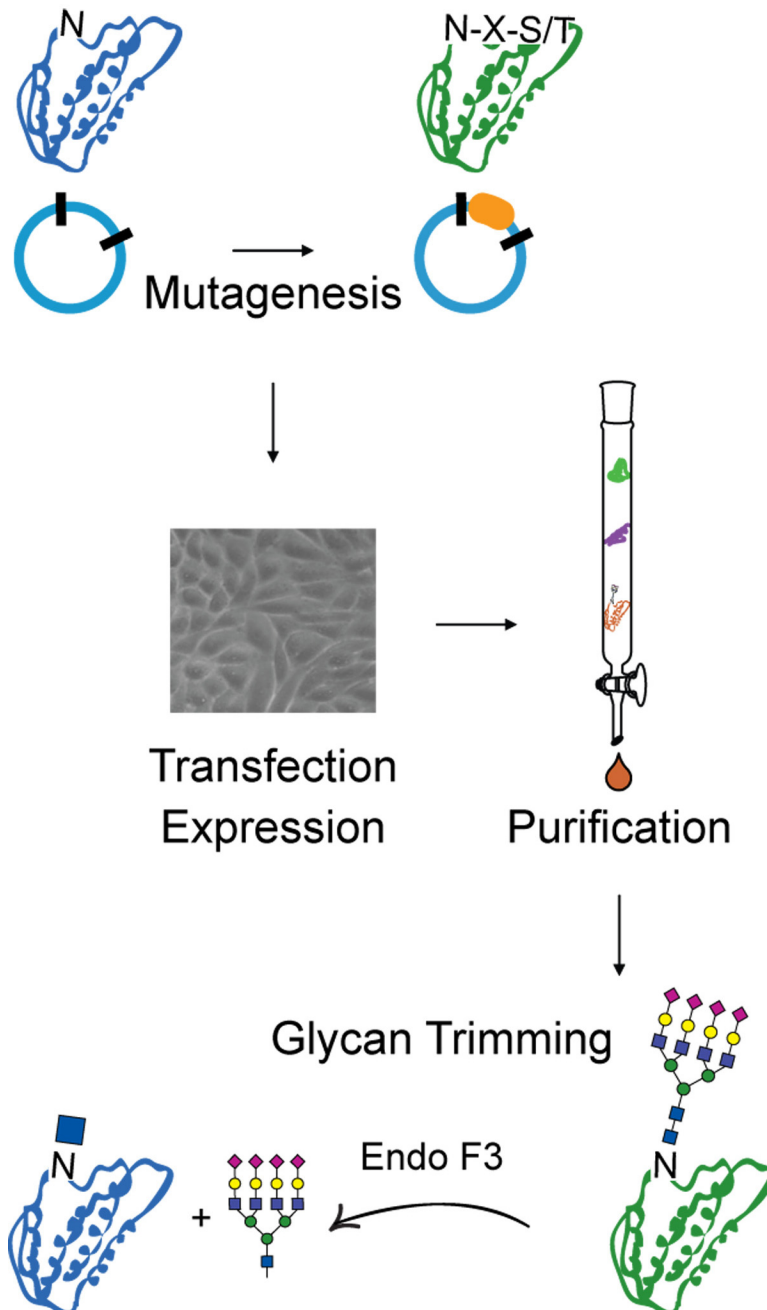
5.2 EXPERIMENTAL PROCEDURES

5.2.1 Materials

The CHO-K1 cells gift from Dr. Jeff Krise (University of Kansas, Department of Pharmaceutical Chemistry, Lawrence, KS). All water was purified using a Millipore Direct Q-3 system (Millipore, Billerica, MA). Other materials were obtained from Sigma in the highest grade available, unless otherwise specified. All solutions were sterilized by filtration prior to use.

5.2.2 Construction of Recombinant Template Plasmid

The mammalian vector pCMV6-XL5, which contained the human hGH gene, was purchased from Origene Technologies Inc.(number SC-3300088), and used as the template for PCR based site-directed mutagenesis. Features of this plasmid include an N-terminal peptide secretion signal to enable secretion of the protein into the media, and an N-terminal His4-tag to ease purification. To facilitate plasmid production in *E. coli*, the plasmid also contains the ampicillin resistance gene and ColE1 origin of replication.

Competent XL-1 Blue E. coli cells (Novagen, Gibbstown, NJ) were transformed using a standard heat shock procedure, then plated on Luria broth (LB) with 2% agar plates containing 100 mg/mL ampicillin. A colony was selected and used to inoculate 10 mL of selective LB medium. The resultant DNA was purified using the QIAprep Spin Miniprep Kit (QIAGEN, Valencia, CA) and quantified using the Nanodrop® ND-1000

Spectrophotometer (Thermo Scientific, Wilmington, DE) according to the manufacturer's instructions.

5.2.3 Site-Directed Mutagenesis of Glycosylation Site

The vector plasmid was used as a template for amino acid modification to incorporate an N-linked glycosylation site at the asparagine residue(s) of interest. Mutagenesis was performed using the QuikChange mutagenesis kit (Stratagene, La Jolla, CA) using the manufacturer's instructions. DNA site directed mutagenesis primers (Integrated DNA Technologies, Inc., Coralville, IA) were designed to generate the rhGH-H157S mutant: (5' to 3') c agc aag ttc gac aca aac tca **AG**c aac gat gac gca cta c (leading) and g tag tgc gtc atc gtt g**CT** tga gtt tgt gtc gaa ctt gct g (reverse), where the mutated bases are indicated using capital letters. The rhGH-D160S mutant was generated with primers (5' to 3') c agc aag ttc gac aca aac tca gac aac gat **AG**c gca cta c (leading) and g tag tgc gtc atc gtt g**CT** tga gtt tgt gtc gaa ctt gct g (reverse), and the rhGH-H157S,D160S mutant was generated using (5' to 3') c agc aag ttc gac aca aac tca **AG**c aac gat **AG**c gca cta c (leading) and g tag tgc g**CT** atc gtt g**CT** tga gtt tgt gtc gaa ctt gct g (reverse).

The mutagenesis product was digested with 1 µL of DpnI enzyme (New England Biosciences, Ipswich, MA) to selectively destroy the template DNA. The product was used to transform competent XL-1 Blue *E. coli* cells as described above. Transformed cells were selected after overnight growth at 37 °C on LB agar containing 100mg/mL ampicillin. Plasmids containing the vector containing the recombinant rhGH gene were harvested using a QIAprep Spin Miniprep Kit according to manufacturer's instructions. All miniprep DNA was quantified as described above and sequenced by Northwoods

DNA Inc., Bemidji, MN. The desired sequences were each verified prior to transfection, with DNA sequence results analyzed using the FinchTV application.

5.2.4 Cell Transfection and Culture

CHO-K1 cells were seeded and maintained using complete high glucose Dulbecco's Modified Eagle Medium (DMEM) (Thermo Scientific Hyclone, Logan, UT). The medium was supplemented with MEM Non-Essential Amino acids 100x (ATCC, Manassas, VA), 10% fetal bovine serum (Mediatech, Inc. Manassas, VA), 100X penicillin/streptomycin (ATCC, Manassas, VA), and 150 mM L-Proline.[28]

Cells were maintained in T-75 flasks (BD, Franklin Lakes, NJ) with 24 mL of medium and kept at 37 °C in the presence of 5% CO_2. Passage was accomplished by two wash steps using pre-warmed 10% phosphate buffered saline (PBS) whereupon the cells were trypsinized for 3 min to remove the adherent cells from the surface of the flask, using 0.25% trypsin with 0.1% EDTA. Then, the cells were harvested by centrifugation (3800 x g, 2 min); the supernatant was removed by vacuum aspiration, and the cells were then gently resuspended in 6 mL of the culture medium, described above.

Cells were grown to 50% confluency and transfected with 4 µg of mutated DNA and 20 µL of Turbofectin 8.0 transfection reagent (OriGene Technologies, Inc., Rockville, MD) per T-75 flask, per mutant protein, according to Origene's recommendations. The supernatant was removed by decanting and replacing the media every 24 hrs. After 48 hrs of growth, the supernatant was collected, filtered using a 0.2 µm pore size filter, and concentrated to 0.5 mL using cellulose centrifugal filtration devices (Millipore, Billerica,

MA) with a 10 kDa molecular weight cut off (MWCO) at 8800 x g in a fixed angle (35°)

rotor.

5.2.5 Lectin Affinity Chromatography (LAC) Purification

A cartridge (5-mL) containing *Maackia amurensis* leukoagglutinin (MAL) lectin resin

(QIAgen, Valencia, CA) was equilibrated according to the manufacturer's directions

using the indicated binding/wash buffer. The concentrated supernatant (0.5 mL) from a

single supernatant collection was mixed with an equal volume of binding/wash buffer

and loaded onto the column at 1 mL/minute. The column was washed using 50 mL of

binding/wash buffer at 4 mL/min, and the rhGH was eluted into 50 mL of elution buffer

containing 200 mM lactose. The elution fraction was concentrated to approximately 1

mL as described above. The protein content was quantified using a standard Bradford

assay. The rhGH protein was dialyzed using Slide-A-Lyzer dialysis cassette with a 10

kDa MWCO (Pierce, Rockford, IL) into 10 mM sodium citrate and 150 mM sodium

chloride, pH adjusted to 5.5 with HCl, with 3 (1 L) exchanges.

5.2.6 SDS-PAGE Analysis

12% SDS-PAGE TRIS/glycine gels were generated for electrophoretic analysis.

Protein samples were mixed with 2X reducing Laemmli loading buffer and water (1:3:3)

and boiled for 15 minutes. The samples were loaded, and electrophoresis was

performed at 123 V. Precision Plus unstained protein ladder (Bio-Rad, Hercules, CA)

was utilized in separate lane(s) for molecular weight approximation. Gels were stained

using Coomassie (R-250), overnight. Destaining was accomlished with 20% acetic acid

and 50/50 methanol/H$_2$O until background was clear.

5.2.7 Enzymatic Glycan Trimming

Purified, glycosylated rhGH was treated with Endo β-*N*-acetylglucosaminidase F

(Endo F3) (CalBioChem, San Diego, CA) by adding enzyme in a 1:16000

enzyme:protein molar ratio, and incubating for 72 hours at 37 °C.

5.2.8 CD Secondary Structure Estimation

The purified protein was diluted to 0.5 mg/mL in 10 mM sodium citrate and 150

mM sodium chloride and placed in a jacketed, quartz, 1.0-mm pathlength sample cell.

Far UV CD measurements were made using a Jasco J-715 spectropolarimeter (Tokyo,

Japan) with an attached circulating water bath that was built in-house. Experiments

were performed at 10 °C, with a sensitivity of 100 mdeg and scan speed of 10 nm/min.

The experiments were conducted under constant nitrogen flow. Scans were performed

from 260 to 180 nm. Multiple spectra were averaged, smoothed and baseline corrected

by subtracting signal from background solutions prior to analysis. The publicly available

DICHROWEB server was used, with the CDSSTR algorithm, to calculate the percent

protein secondary structure. Secondary structure was studied for the fully glycosylated

and the enzymatically trimmed glycosylated proteins.

5.2.9 CD Thermal Stability Quantitation

Each rhGH solution was adjusted to pH 8.0 using NaOH, and subjected to increasing temperature, from 22-90°C, with a data pitch of 0.2°C, and at a rate of 12°C/hour. The ellipticity was monitored at 222 nm. The transition temperature was calculated using a sigmoidal fit algorithm within Origin (version 6.0), and compared to a completely nonglycosylated, native rhGH protein standard. The experiments were performed in triplicate.

5.2.10 Human Growth Hormone Standard

A rhGH standard protein was produced as described above, but with the following modifications. The standard was generated by introducing a single point mutation, D136N, into the gene sequence using primers: (5' to 3') atg ggg agg ctg gaa AAT ggc agc ccc cgg act g (leading) and c agt ccg ggg gct gcc ATT ttc cag cct ccc cat (reverse). Protein verification, purification and handling was the same, except the glycan was removed completely using PNGase F, added in a enzyme:protein ratio of 1:10800 (mol/mol). In addition to cleaving the glycan, the enzyme converts the N to a D, which restores the native protein sequence. The CD thermal stability experiments on this standard were used as a control for comparison to the mutants stabilized with glycosylation.

5.3 RESULTS

5.3.1 Mutations to Acquire N-linked Glycans

Recombinant human growth hormone (rhGH) was modified to include N-linked glycan moieties at asparagine residues that are known to deamidate readily. The growth hormone sequence is shown in **Figure 1**. Mutations were completed to attain the N-linked glycan consensus sequence N-X-S/T, which acts as a signal in eukaryotic cells to add a glycan structure to the growing polypeptide chain. Mutant I contains a serine residue in place of histidine, and Mutant II contains a serine in place of aspartic acid. Both serine residues are necessary to meet the requirements of the consensus sequence. In Mutant III, both mutations were performed to obtain two N-linked glycosylation sites on the rhGH protein. The mutated DNA was verified by Northwoods DNA Inc., Bemidji, MN, prior to transfection in CHO-K1 cells. Media was collected and the mutant proteins were purified using lectin affinity chromatography.

```
MAHHHHFPTIPLSRLFDNAMLRAHRLHQLAFDTYQ
EFEEAYIPKEQKYSFLQNPQTSLCFSESIPTPSNR
EETQQKSNLELLRISLLLIQSWLEPVQFLRSVFAN
SLVYGASDSNVYDLLKDLEEGIQTLMGRLEDGSPR
```
Mutant I   NS**S**
Mutant II   ND**S**
Mutant III   NS**S**ND**S**
```
TGQIFKQTYSKFDTNSHNDDALLKNYGLLYCFRKD
MDKVETFLRIVQCRSVEGSCGF
```

Figure 1. Protein sequence of recombinant human growth hormone and the three mutant protein sequences. Mutant I rhGH-H157S. Mutant II rhGH-D160S. Mutant III rhGH-H157S,D160S. The mutations introduce N-linked glycosylation site(s) into the sequence. The resulting glycans' function is to sterically block deamidation of N155 and N157. Amino acid positions are used according to the gene sequence in the figure, with M being position 1.

5.3.2 SDS-PAGE Analysis

The molecular mass of the glycosylated mutant rhGH proteins was estimated by SDS-PAGE to be between 20 and 25 kDa; within the range expected for the glycosylated form of rhGH. (**Figure 2.**) Native hGH is known to have a molecular mass of 22 kDa.[29] The purity of the rhGH produced was qualitatively assessed by SDS-PAGE, in lanes 1-3 of Figure 2, no other significant bands are present, indicating mutant proteins have been produced to sufficient purity for further study. The mutant proteins were treated with the Endo F3 enzyme to trim the majority of the glycan from the protein and leave only an N-acetylglucosamine carbohydrate residue attached to the asparagine.
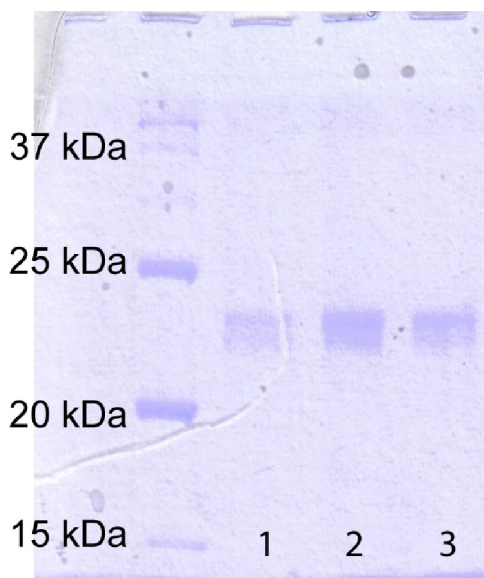

Figure 2. Coomassie stained, 12% acrylamide Tris/Glycine SDS-PAGE gel showing the three recombinant human growth hormone mutants. Lanes (1) rhGH-H157S,D160S, (2) rhGH-D160S, (3) rhGH-H157S.

5.3.3 Circular Dichroism

The secondary structure of the rhGH mutant proteins I-III was studied after purification, and the proteins did not possess secondary structure similar to native hGH, which contains a high degree of alpha helical character. In **Figure 3,** plot a, the far UV CD data for one of the purified rhGH-D160S is shown as a quintessential example. The mutant proteins all contain a large N-linked glycan that is not native to the protein, which clearly causes distortions in the secondary structure. To remedy this problem, an enzyme was used to remove the majority of the glycan. After the glycan was trimmed, the mutant proteins more closely resembled the nonglycosylated rhGH standard. The trimmed sample of rhGH-D160S is shown as plot (b) in Figure 3. Both the overall amount of secondary structure is increased, in going from plot a to plot b, and plot b is a closer match to the expected data for native growth hormone.
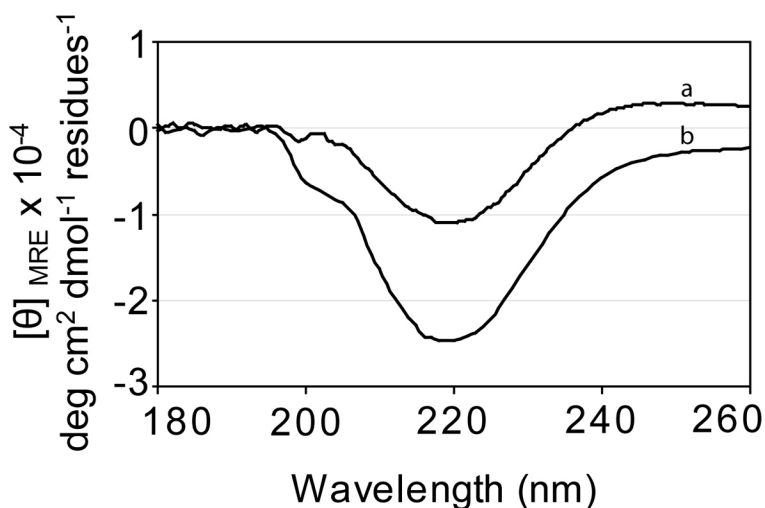

Figure 3. Far UV CD secondary structure. The rhGH-D160S protein was studied after purification (a) and was found to not resemble the structure of the rhGH standard protein. After glycan trimming to remove the majority of the glycans, the rhGH-D160S sample (b) demonstrated an increase in secondary structure and was more similar to the rhGH standard protein.

All mutant proteins were subjected to thermal stress and increased pH, which are conditions that facilitate asparagine deamidation.[30] Loss of secondary structure induced by these conditions was monitored at 222 nm using CD (**Figure 4).** The rhGH standard is noted to have a transition that starts earlier than the glycosylated protein, and loses considerably more secondary structure as temperatures continued to increase. The melt temperature (T_m) was calculated for each scan, and average results are given in **Table 1** below.
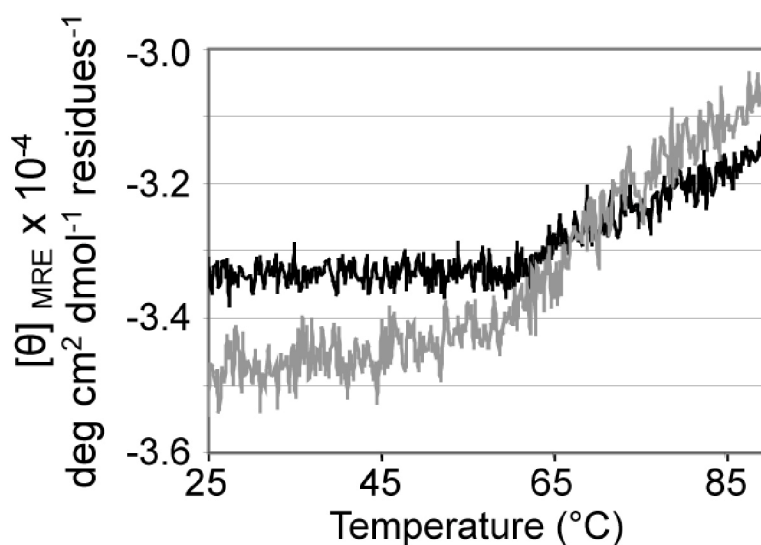
Figure 4. Secondary structure loss monitored at 222 nm, in mean residue ellipticity, using far UV CD with increasing temperature. The rhGH-H157S,D160S after glycan trimming (in black) is compared to the nonglycosylated rhGH standard nonglycosylated (in gray).

Table 1. Secondary structure loss was monitored at 222 nm using far UV CD to calculate melt temperature (T_m) values for each of the three rhGH mutant proteins. T_m values are then compared to a nonglycosylated rhGH standard protein. Each value represents an average of three separate experiments. The RSD for each measurement is less than 3 %.

Protein	rhGH-H157S	rhGH-D160S	rhGH-H157S,D160S	Standard
T_m (℃)	77.1	72.2	74.9	66.1

5.4 DISCUSSION

Recombinant human growth hormone (rhGH) is used to treat intrinsic deficiencies in growth hormone production, but these therapeutic proteins have been shown to suffer from protein instability. The major cause of instability for rhGH is thought to be deamidation of asparagine residues, especially at N155 and N157.[12] To improve the stability of rhGH proteins, a stabilization scheme was designed using non-native N-linked glycosylation at the vulnerable asparagine residues. The process by which the asparagine residues were modified is summarized in Scheme 1. The work-flow depicts mammalian cell expression hosts, CHO-K1, which are often used to incorporate N-linked glycans onto therapeutic proteins.[27,31]

The purpose of the N-linked glycans is to provide steric hindrance to slow or completely inhibit the nonenzymatic reaction that occurs to convert asparagine to aspartic acid. It has been demonstrated that asparagine deamidation is slowed when bulky or large amino acids near the asparagine are present, or when the tertiary structure reduces the amount of space surrounding the asparagine residue.[21,25,32] Our new approach uses non-native glycosylation to provide the steric hindrance required to slow or block asparagine deamidation. The large glycan structures acquired in CHO-K1 expression hosts are considerably larger than amino acids, and far UV secondary structure data indicated the glycan negatively impacted the protein structure. To reduce the amount of bulk, an enzymatic method of glycan trimming was used.[33] The secondary structure of the trimmed protein resemebled the structural data from the rhGH standard protein.

To determine if this protein stabilization scheme is effective for rhGH, the mutant rhGH proteins were subjected to conditions that promote deamidation and their melt temperatures were compared to that of a growth hormone standard. Resulting T_m values from thermal stability scans indicated that the glycosylated mutant rhGH proteins are more thermodynamically stable than the nonglycosylated standard protein. Statistical analysis using the Student's t-test indicated the differences in T_m values between each are statistically significant at the 90% confidence interval. These data indicate non-native glycosylation may improve the stability of the rhGH proteins by blocking asparagine deamidation.

5.5 CONCLUDING REMARKS

Protein instability induced by asparagine deamidation may be slowed or prevented by the inclusion of a bulky functional group at or near the asparagine residue. This concept was used to develop a novel protein stabilization scheme where N-linked glycosylation sites are mutated into the protein sequence, at the asparagine residue that is vulnerable to deamidation. To demonstrate this new approach, recombinant mutant human growth hormone proteins were generated with N-linked glycosylation at two such asparagine residues. Enzymatically trimming all the glycan except the monosaccharide attached to the asparigine reduced structural changes to the protein, yet still provides steric hindrance to block deamidation and increase thermostability. Secondary structure scans using circular dichroism indicate that the glycosylated mutants are more thermodynamically stable than a non-glycosylated growth hormone control. The method of including an N-linked glycan at vulnerable asparagine residues is a new approach to improving protein stability, and this approach could be used in developing more stable therapeutic proteins.

Kansas, undergraduate student) also assisted me in protein production, purification, and analysis of the proteins.

5.6 FUTURE WORK

The human growth hormone protein mutants described in this chapter will be subjected to proteolytic digestion and high resolution mass spectrometric analysis. These studies will provide ancillary support that we have purified the protein of interest, and that enzymatic modification of the large N-linked glycan has been achieved. The sample preparative steps will employ glycopeptide and peptide analysis, as described in Chapter 1, section 1.2.3.1.

5.7 REFERENCES

1.   Walsh, G. Biopharmaceuticals: recent approvals and likely directions. *Trends Biotechnol*. **2005,** *23,* 553-558.

2.   Hermeling, S.; Crommelin, D. J. A.; Schellekens, H.; Jiskoot, W. Structure-Immunogenicity relationships of therapeutic proteins. *Pharm. Res.* **2004**, *21*, 897-903.

3.   Sola, R. J.; Griebenow, K. *J.* Effects of glycosylation on the stability of protein pharmaceuticals. *Pharm. Sci.* **2009,** *98,* 1223-1245.

4.   Wilhelmsen, T. W.; Skibeli, V.; Arntzen, F. C. Stability study of somatropin by capillary zone electrophoresis. *Procedia Chemistry* **2010,** *2,* 34-45.

5.   Patrick, G. L. *An Introduction to Medicinal Chemistry,* 2nd ed.; Oxford University Press: New York, 2001.

6.   du Bois, M. B.; Rappaport, R.; Royer, P. Difficultés liées à l'utilisation de la somathormone humaine du type Raben, et résultats préliminaires d'une nouvelle technique d'extraction. *Archives françaises de pédiatrie.* **1971,** *28,* 673.

7.   Soyka, L. F.; Ziskind, A.; Crawford, J. D. Treatment of short stature in children and adolescents with human pituitary growth hormone (Raben). Experiences with thirty-five cases. *N. Engl. J. Med.* **1964,** *271,* 754-764.

8.   Salomon, F.; Cuneo, R. C.; Hesp, R.; Sonksen, P. H. The effects of treatment with recombinant human growth hormone on body composition and metabolism in adults with growth hormone deficiency. *N. Engl. J. Med.* **1989,** *321,* 1797-1803.

9.   Laron, Z. Short stature due to genetic defects affecting growth hormone activity. *N. Engl. J. Med.* **1996,** *334,* 463-466.

10.   Takahashi, Y.; Kaji, H.; Okimura, Y.; Goji, K.; Abe, H.; Chihara, K. Brief report: Short stature caused by a mutant growth hormone. *N. Engl. J. Med.* **1996,** *334,* 432-436.

11.   Rosenbloom, A. L.; Knuth, C.; Shulman, D. Growth hormone by daily injection in patients previously treated for growth hormone deficiency. *Southern Medical Journal.* **1990,** *83,* 653-655.

12.   Lewis, U. J.; Singh, R. N. P.; Bonewald, L. F.; Seavey, B. K. Altered proteolytic cleavage of human growth hormone as a result of deamidation. *J. Biol. Chem.* **1981,** *256,* 11645-11650.

13.   Johnson, B. A.; Shirokawa, J. M.; Hancock, W. S.; Spellman, M. W.; Basa, L. J.; Aswad, D. W. Formation of isoaspartate at two distinct sites during *in vitro* aging of human growth hormone. *J. Biol. Chem.* **1989,** *264,* 14262-14271.

14.   Tavornvipas, S.; Tajiri, S.; Hirayama, F.; Arima, H.; Uekama, K. Effects of hydrophilic cyclodextrins on aggregation of recombinant human growth hormone. *Pharm. Res.* **2004,** *21,* 2369-2376.

15.   Robinson, N. E.; Robinson, A. B. *Molecular Clocks: Deamidation of Asparaginyl and Glutaminyl Residues in Peptides and Proteins;* Althouse Press: Cave Junction, OR, 2004.

16.   Cleland, J. L.; Powell, M. F.; and Shire, S. J. The development of stable protein formulations: A close look at protein aggregation, deamidation, and oxidation. *Crit. Rev. Ther. Drug Carrier Syst.* **1993,** *10,* 307-377.

17.   Shultz, J.; Xu, Q.-Y.; Hurst, R.; Mezei, L.; White, D.; Betlach, M.; Stevans, J.; Klekamp, M. The ISOQUANT (TM) protein deamidation detection kit. *Promega Notes Magazine.* **1995,** *No. 53*, 22-28.

18.   Robinson, N. E.; Robinson, Z. W.; Robinson, B. R.; Robinson, A. L.; Robinson, J. A.; Robinson, M. L.; Robinson, A. B. Structure-dependent nonenzymatic deamidation of glutaminyl and asparaginyl pentapeptides. *J. Peptide Res.* **2004,** *63,* 426-436.

19.   Robinson, N. E.; Robinson, A. B. Prediction of primary structure deamidation rates of asparaginyl and glutaminyl peptides through steric and catalytic effects. *J. Peptide Res.* **2004,** *63,* 437-448.

20.   Cournoyer, J. J.; Pittman, J. L.; Ivleva, V. B.; Fallows, E.; Waskell, L.; Costello, C. E.; O'Conner, P. B. Deamidation: differentation of aspartyl from isoaspartyl products in peptides by electron capture dissociation. *Protein Sci.* **2005,** *14,* 452-463.

21.   Rivers, J.; McDonald, L.; Edwards, I. J.; Beynon, R. J. Asparagine deamidation and the role of higher order protein structure. *J. Proteome Res.* **2008,** *7,* 921-927.

22.   Maa, Y.-F.; and Hsu, C. C. Aggregation of recombinant human growth hormone by phenolic compounds. *Int. J. Pharm.* **1996,** *140,* 155-168.

23.   Wang, W. Instability, stabilization, and formulation of liquid protein pharmaceuticals. *Int. J. Pharm.* **1999,** *185,* 129-188.

24.   Chen, H.-M.; Ford, C.; Reilly, P. J. Substitution of asparagine residues in *Aspergillus awamori* glucoamylase by site-directed mutagenesis to eliminate N-glycosylation and inactivation by deamidation. *Biochem. J.* **1994,** *301,* 275-281.

25.   Clarke, S. Propensity for spontaneous succinimide formation from aspartyl and asparaginyl residues in cellular proteins. *Int. J. Pept. Protein Res.* **1987,** *30,* 808–821.

26.   Bustamante, J. J.; Gonzalez, L.; Carroll, C. A.; Weintraub, S. T.; Aguilar, R. M.; Munoz, J.; Martinez, A. O.; Haro, L. S. *O*-Glycosylated 24 kDa human growth hormone has a mucin-like biantennary disialylated tetrasaccharide attached at Thr-60. *Proteomics.* **2009,** *9,* 3474-3488.

27.   Baker, K. N.; Rendell, M. H.; Hills, A. E.; Hoare, M.; Freedman, R. B.; James, D. C. Metabolic control of recombinant protein *N*-glycan processing in NS0 and CHO cells. *Biotechnol. Bioeng.* **2001,** *73,* 188-202.

28.   Kao, F.-T.; Puck, T. T. Genetics of somatic mammalian cells. IV. Properties of Chinese hamster cell mutants with respect to the requirement for proline. *Genetics.* **1967,** *55,* 513-524.

29.   www.uniprot.org Date accessed: 4/5/10.

30.   Nilsson, M. R.; Driscoll, M.; Raleigh, D. P. Low levels of asparagine deamidation can have a dramatic effect on aggregation of amyloidogenic peptides: Implications for the study of amyloid formation. *Protein Sci.* **2002,** *11,* 342-349.

31.   Higgins, E. Carbohydrate analysis throughout the development of a protein therapeutic. *Glycoconjugate J.* **2010,** *27,* 211-225.

32.   Wright, H. T. Sequence and structure determinants of the nonenzymatic deamidation of asparagine and glutamine residues in proteins. *Protein Engineering.* **1991,** *4,* 283-294.

33.   Toumi, M. L.; Go, E. P.; Desaire, H., Development of fully functional proteins with novel glycosylation via enzymatic glycan trimming. *J. Pharm. Sci.* **2009,** *98,* 2581-2591.

CHAPTER 6

IMPROVING MASS DEFECT FILTERS FOR HUMAN PROTEINS

The mass defect of a substance can be used in mass spectral analysis to identify peaks as likely belonging to a compound class, such as peptides, if the mass defect is within the known range for that compound class. For peptides, a range of possible mass defects was calculated previously, using a set of theoretical peptides, where all possible amino acid combinations were considered (Mann, M. *Abstract from the 43rd Annual Conference on Mass Spectrometry and Allied Topics*; **1995,** ASMS). We compare that range of theoretical peptide mass defects to new values obtained from *in silico* tryptic digests of proteins that are abundant in human serum and human seminal fluid. The range of mass defect values encompassing 95% of peptides for the human protein data sets was found to be up to 50% smaller than the previously reported mass defect range for the theoretical peptides. The smaller range established for human tryptic peptides can be used to improve peptide mass defect filters by excluding more species that are not likely to be peptides, thus improving filter selectivity for peptides during proteomic data analysis.

6.1 INTRODUCTION

Mass spectral data analysis can be a daunting task, but peak identification can be enhanced by incorporating mass defect (MD) analysis into the work flow. MD analysis is used to predict the elemental composition and the identification of organic compounds,[1] metabolites,[2-4] and petrochemicals,[5,6] by using values that are explicit to each class of substances. The elemental composition of some peptides can be determined using MD calculations, but the utility of the method is generally limited to smaller peptides (<800 Da).[7] MD analysis is also used to identify and classify multiple types of peptide modifications, such as phosphorylation[8-10] or cross-linking.[11] Peptide MD can also be used to deconvolute overlapping peaks,[12] support charge state determination algorithms,[13] and assist in high throughput protein identification, such as in peptide mass fingerprinting (PMF) techniques.

PMF data analysis can be enhanced by excluding extrinsic peaks from analysis, whereas those peaks arise from substances other than the protein(s) of interest.[14-15] Masses are excluded, or *filtered*, when the MD value is not within the window that is characteristic for that analyte. See **Figure 1** below for an example of the work-flow.

The expected window, or range, of mass defect values for peptides is established from theoretical peptide masses. This range is known to encompass only selected regions, or "clusters" on the mass scale.[16-18] The values in between the clusters, referred to as the "forbidden zones", are where unmodified peptides are not found. Any peaks located within these forbidden zones are indicative of peptide modifications or non-peptide interferents. Peptide modification can occur with functional groups that force the total mass into the forbidden zone, so that the MD is noticeably

outside of the anticipated range for peptides, for example with iodine tags,[19] or in cases of phosphorylation.[8-10] PMF experiments benefit from mass defect data filtering, as evidenced by higher protein identification probability scores, fewer false positives, and increased number of identified peaks.[15,20-22] Improvements to peptide MD filters can be made, using data sets with actual amino acid usage in place of theoretical peptides, because using these data will result in more accurate peptide MD values.[18,23]
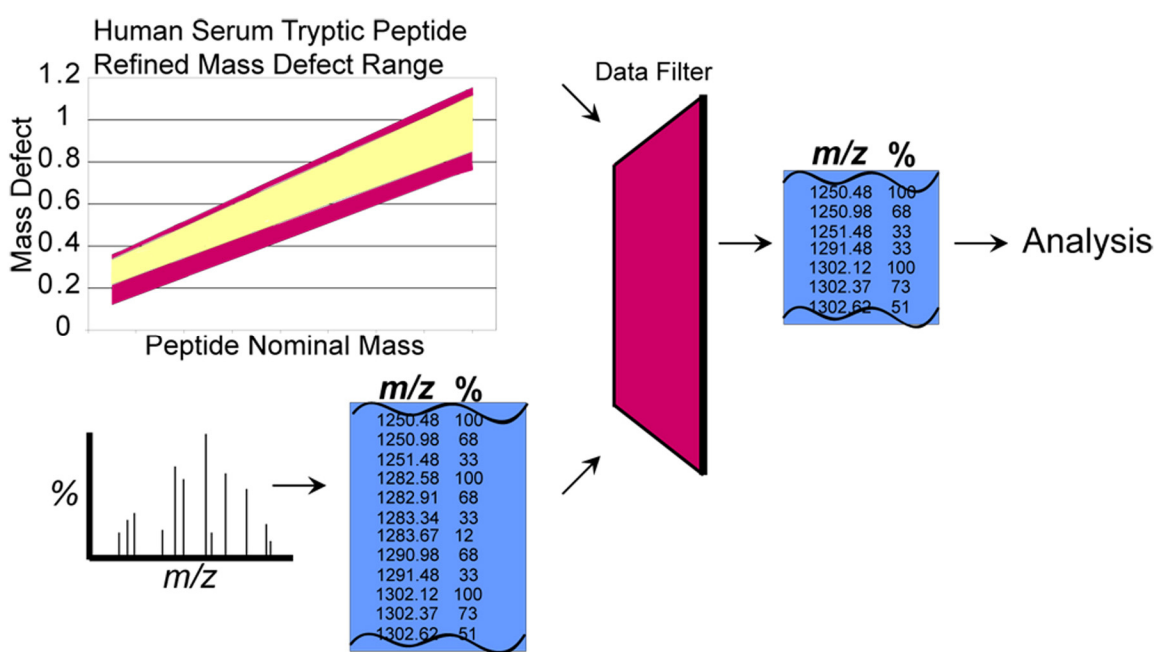

Figure 1. Work flow using a data filter that incorporates mass defect data.

Herein, we calculate mass defect values for two sets of human tryptic peptides and compare the data to theoretical peptide MD calculations. The human data sets were composed using results from *in silico* tryptic digestions, which were performed on human serum and seminal fluid proteins. Analysis of the human datasets defined the range of MD values that encompasses the middle 95% of unmodified peptides. The

breadth of this range was compared to that of the range based on theoretical peptide

calculations. Equations describing the refined MD range are presented for use in PMF

studies.

6.2 EXPERIMENTAL METHODS

6.2.1 Peptide Generation

Proteins that are abundant in human serum[24] and human seminal fluid [25] were chosen for tryptic peptide mass defect analysis. The sequences of the serum and seminal fluid protein sets were collected from the UniProt Knowledgebase, using search options or published accession numbers, respectively. Sequences containing signal peptides and propeptides were truncated so that proteins were analyzed in their relevant forms. The sequences were imported into ProteinProspector, version 5.1.8 Basic, and an *in silico* digestion was performed using the MS-Digest function. Trypsin was chosen for the protease, multiple charges and variable modifications to amino acids were omitted, and zero missed cleavages were allowed. Peptides with a minimum of five amino acids and within the mass range of 500-8000 Da were chosen for analysis.

6.2.2 Mass Defect Analysis

The mass defect was calculated for each tryptic peptide. The mass defect per unit mass (MaDPUM) was then calculated by dividing the mass defect by the monoisotopic peptide mass, and the values were sorted into 100 Da bins, based on the nominal masses of the peptides. The mean and the range of MaDPUM values that encompassed the middle 95% of peptides were established for each bin and plotted against the peptide mass. The MaDPUM calculations were performed separately for the human serum and seminal fluid data sets. A third set of MaDPUM ranges were

established for theoretical peptides, based on previously published mass defect data for theoretical peptides.[16,20]

6.3 RESULTS AND DISCUSSION

Mass defect filters benefit proteomic data analysis by removing non-peptide peaks. The removal is performed prior to submitting MS data to an analysis program such as MASCOT, and is achieved by eliminating peaks whose numerical mass defect values are not within an expected window. Peaks located in the mass ranges outside of the expected window, often referred to as the "forbidden zone", are then excluded from further analysis.

6.3.1 MaDPUM Calculations

In these experiments, we defined the range of mass defect per unit mass (MaDPUM) values that incorporates the middle 95% of tryptic peptides from two human protein data sets. The sequences of the 50 most abundant seminal fluid proteins[25] and nearly 300 human serum proteins[24] were collected from the UniProt Knowledgebase. Each protein was subjected to an *in silico* tryptic digestion using the MS-Digest function within ProteinProspector. The outputs were used to calculate the mass defect (MD) for each peptide (equation I), which is defined as the difference between the nominal (NM) and monoisotopic masses (MI).[1] Nominal mass is an integer sum consisting of the integer masses of the most abundant isotopes for each element, *e.g.* C = 12, H = 1, and O = 16.[26] The monoisotopic mass is a sum of the exact masses of the most abundant isotope for each element in the sample of a substance, *e.g.* C = 12.0000, H = 1.0078, O = 15.9949.[26] The MaDPUM was calculated by dividing the mass defect by the monoisotopic mass of the peptide, as in equation II.

(I) $$MD = MI - NM$$

(II) $$MaDPUM = \frac{MD}{MI}$$

6.3.2 Determination of Mean

The peptides within the human serum data set were grouped into 100 Da bins for analysis, according to the original nominal mass of the peptide. Within each 100 Da bin, the mean MaDPUM value was calculated. The global mean was then calculated and determined to be 0.00050 for the serum data set. In **Figure 2a**, the MaDPUM mean values are plotted for each nominal mass 100 Da bin for serum peptides. The mean MaDPUM value of 0.00050 is higher than the literature value of 0.00048,[16,20] which incorporates theoretical peptide compositions.

Since analysis of the human serum tryptic peptides produced results that differ from the literature values, the set of human seminal fluid peptides was analyzed as a second data set, to validate that the observations were not characteristic of human serum peptides alone. This data is shown in **Figure 2b**. The seminal fluid data set also showed that the global mean value was 0.00050, the same as the human serum data set. The agreement in these two data sets suggests that this mean mass defect value (0.00050) is appropriate to use for any large set of human proteins.

In Figure 2, the smallest peptides are noted to have a mean MaDPUM value greater than the average due to a mathematical bias that is based on protease specificity.[27] All tryptic peptides, except possibly the C-terminal peptide, possess an R or K residue at the peptide's C-terminus.[28] The MaDPUM for both R and K is greater than the average MaDPUM for all common amino acids, thus creating a slight bias for an increased MaDPUM for the small tryptic peptides.
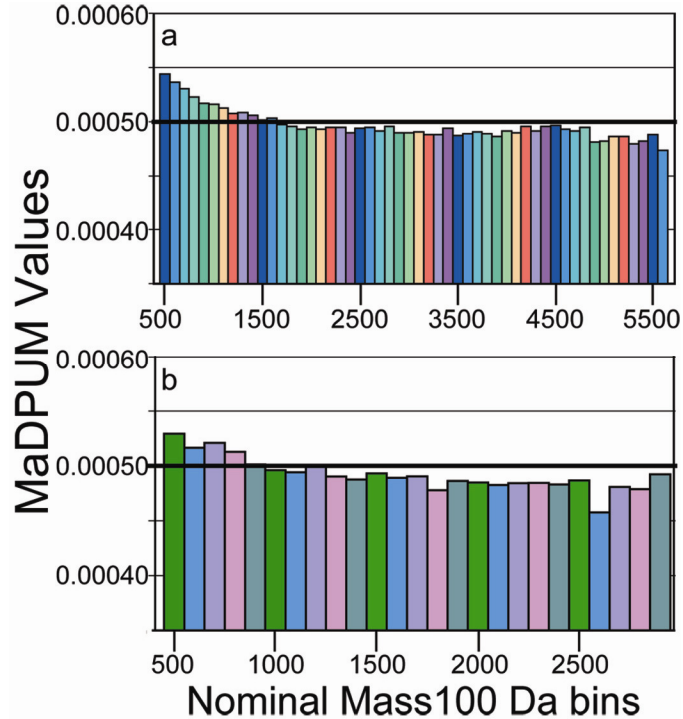
Figure 2. MaDPUM mean values as calculated for each nominal peptide mass 100 Da bin. Global mean for both (a) human serum peptides and (b) human seminal fluid peptides was determined at 0.00050, indicated by the dark, horizontal line.

To understand why global mean values for both the serum and seminal fluid data was larger than anticipated (0.00050 instead of 0.00048), the peptide compositions, including amino acid usage, were analyzed in the seminal fluid data. In **Figure 3a**, the amino acid usage is shown for the theoretical peptide data set, where the 20 common amino acids are utilized equally and would account for 5% of each peptide.
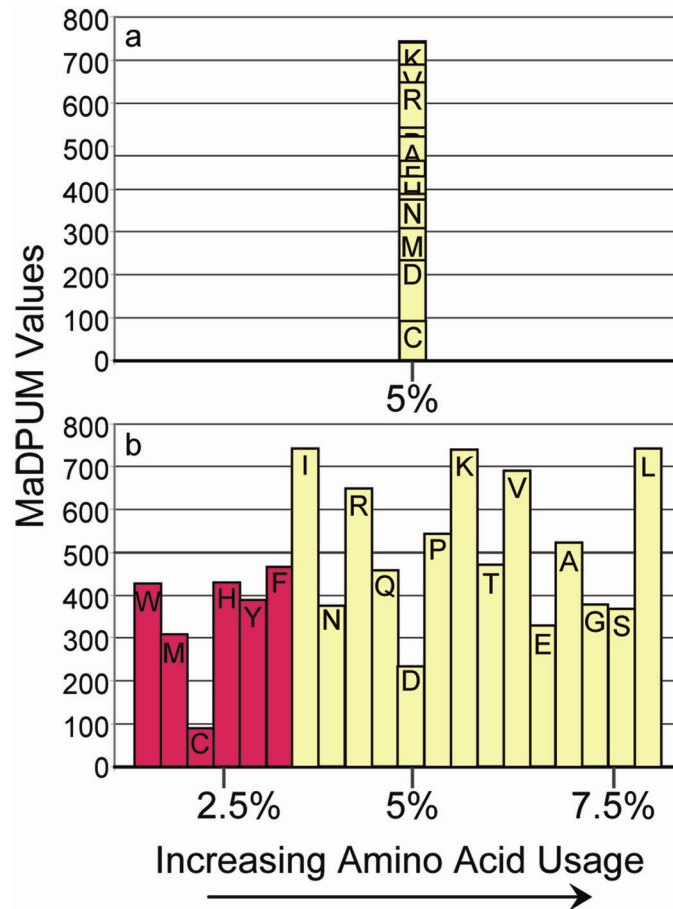
Figure 3. Amino acid usage for the 20 common amino acids. The MaDPUM value for each amino acid is indicated by bar height in ppm. In (a), the usage is shown for the theoretical protein data set. Each amino acid accounts for 5% of utilization. In (b), amino acid usage is shown for the human seminal fluid data set. The 20 common amino acids are ranked by increasing usage from left to right. The 5 least utilized amino acids, shown in red, possess a MaDPUM value that is less than the average value for all amino acids. Amino acid usage in human seminal fluid proteins determines the mean MaDPUM value of 500 ppm (or 0.00050)

The MaDPUM value for each amino acid is indicated by bar height. In **Figure 3b,** amino acid usage is shown, as calculated for the seminal fluid data set, ranked by increasing usage. The ten amino acids to the right, in Figure 3b, would be expected to comprise 50.0% of polypeptide compositions if each amino acid were used equally, but instead they account for 65.8% of amino acid usage in the seminal fluid data. The five least utilized amino acids, to the left in the figure, have MaDPUM values less than the

average for amino acids. These observations demonstrate that amino acid usage will determine the MaDPUM value for each peptide, and the differential amino acid utilization also forces the global average MaDPUM value to 0.00050, higher than previously reported. Amino acid usage may prove an important consideration when refining mass defect filters for peptides, as usage varies across species.[29]

6.3.3 MaDPUM Width for Theoretical Peptides

The theoretical peptide masses and corresponding MaDPUM values were calculated using relationships based on nominal peptide mass, as described previously.[16,20]

(III) $$MI = NM + 0.00048NM$$

(IV) $$W = 0.19 + 0.0001NM$$

The MI (monoisotopic) masses for theoretical peptides were generated using equation III, and the width (W) of the range that encompassed 95% of peptides was determined using equation IV.[16,20] We used these equations to calculate MaDPUM for the theoretical peptides, for each 100 Da bin. Results were plotted against nominal peptide mass (**Figure 4a**). These data are used to assess how similar the width is for the experimental MaDPUM data.
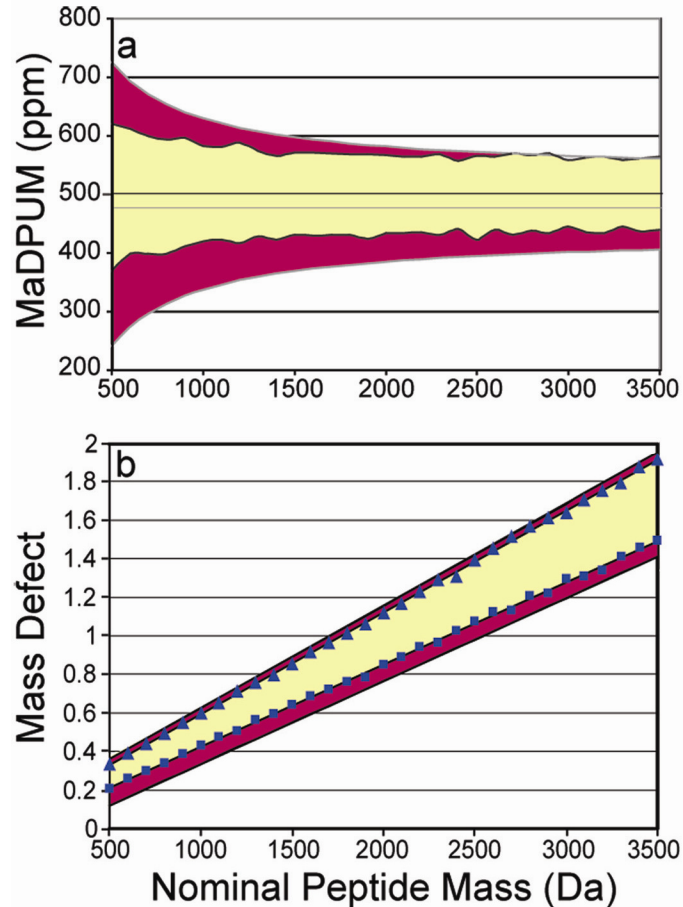
Figure 4a. MaDPUM values in ppm *vs.* nominal peptide mass in Da. The red area represents the values that are excluded from the human serum tryptic peptide dataset, but included in theoretical peptide calculations. The area in yellow represents the human serum tryptic peptide MaDPUM values. The differences in MaDPUM widths for the nominal masses shown were found to be between 23-50% for the analyzed mass range.

Figure 4b. Mass defect *vs.* nominal peptide mass in Da. The area representing mass defect values that encompass 95% of human serum tryptic peptides is yellow, which is a significantly more narrow range of mass defect values than those corresponding to the theoretical peptide data set, shaded in red. The upper and lower boundaries incorporating 95% of the serum peptides was calculated for each 100 Da bin and data points are indicated by blue symbols, ▲ upper boundary and ■ lower boundary. Best fit lines through the points were added to the data, and the equations can be found in equations V and VI for the upper and lower boundaries of the serum peptides.

6.3.4 MaDPUM Width (range) for Human Biofluid Peptides

Extensive studies of the width, or the range, of mass defect values for unmodified peptides were conducted on both the serum and seminal fluid proteins. These results indicated that no differences existed between the two human data sets for the range of

141

MaDPUM values encompassing the middle 95% of peptides. The width (range) determination studies below are presented exclusively using the serum peptides, as that set has more data points

After calculating the mass defect for each serum peptide in each 100 Da bin, we measured the mass defect width for the biofluid data, which can be defined as a range of MaDPUM values that incorporated 95% of all the human serum peptides, for each 100 Da bin. This width was calculated at each 100 Da bin by sorting the MaDPUM values in ascending order within the bin, and by manually observing the middle 95% of values. The yellow area in the center of Figure 4a represents the experimental width. The red areas in Figure 4a represent MaDPUM values that are excluded from the serum data's width, but are included in the theoretical width, according to equation IV.

In **Figure 4b**, the data is plotted differently. Here, MD values are plotted *vs.* nominal peptide mass divided into 100 Da bins. The MD widths are shown for the theoretical peptides (in red) and the human serum peptides (in the center shown in yellow). The theoretical peptide width was established using equation IV. The width for the serum peptides was defined using the experimental data to determine data points for the upper and lower boundaries containing 95% of peptides within each 100 Da bin. In figure 4b, symbols indicate the individual data points for each bin, and linear trend lines to fit the data points are shown. Equations that define the upper and lower boundary best fit lines are found in equations V and VI below, respectively.

(V) $\qquad y = 0.00052738x + 0.066015$

(VI) $\qquad y = 0.00042565x + 0.00038210$

The R^2 values for the best fit lines are 0.99921 or greater.

6.3.5 Width Analysis

The differences in width between the human serum data set and the theoretical data were analyzed to determine how significantly this data will affect the size of MD mass analysis filters. The MaDPUM width established for theoretical peptide compositions are significantly wider than the width in this study on human serum and seminal fluid tryptic peptides. For peptides with a mass of 1000 Da, the width of MaDPUM values that would include the middle 95% of serum peptides is 42.3% smaller than the width established using theoretical peptides. See **Table 1**.

Table 1. Observed MaDPUM ranges for peptide data sets. Where ws = width of the observed MaDPUM range for human serum tryptic peptides, wt = width of the observed MaDPUM range for theoretical peptides. The MaDPUM values are presented in ppm.

Peptide Mass (Da)	MaDPUM^{ws}	MaDPUM^{wt}	Size Difference
1000	162	281	42.3 %
2000	132	193	31.6 %
3000	114	162	25.6 %

At a mass of 3000 Da, the difference in width is smaller, 25.6%. These differences in the calculated size of the "forbidden zone" for mass defect analysis are largest where it makes the most impact, on the lower end of the mass scale; a majority of tryptic peptides are less than 3000 Da in mass.[30] The differences in width indicate that a significant number of additional peaks may be excluded from analysis, without losing data from unmodified peptides.

143

6.4 CONCLUDING REMARKS

To refine peptide mass defect filters, the mass defect and mass defect per unit mass (MaDPUM) values were analyzed for two sets of human proteins and compared to data from the literature based on theoretical peptide compositions. The global human MaDPUM mean value was found to be 0.00050, larger than previously reported, and analysis of amino acid usage indicates that organism usage should be considered when refining mass defect filters for peptide analysis. The width of MaDPUM values encompassing 95% of peptides within the data sets analyzed were found to be up to 50% smaller than the width previously established using the theoretical peptide data. The selectivity of peptide mass defect filters can be improved by using the refined equations presented here, describing the upper and lower boundaries of the mass defect values for unmodified peptides. Use of these equations would increase the number of excluded extraneous (non-peptide or modified peptide) peaks.

ACKNOWLEDGMENT

6.5 REFERENCES

1.   Kendrick, E. *Anal. Chem.* **1963,** *35,* 2146-2154.

2.   Mortishire-Smith, R. J.; Castro-Perez, J. M.; Yu, K.; Shockcor, J. P.; Goshawk, J.; Hartshorn, M. J.; Hill, A. *Rapid Commun. Mass Spectrom.* **2009,** *23,* 939-948.

3.   Zhu, M.; Ma, L.; Zhang, D.; Ray, K.; Zhao, W.; Humphreys, W. G.; Skiles, G.; Sanders, M.; Zhang, H. *Drug Metab. Dispos.* **2006,** *34,* 1722-1733.

4.   Rousu, T.; Pelkonen, O.; Tolonen, A. *Rapid Commun. Mass Spectrom.* **2009,** *23,* 843-855.

5.   Fu, J.; Kim, S.; Rodgers, R. P.; Hendrickson, C. L.; Marshall, A. G. *Energy & Fuels.* **2006,** *20,* 661-667.

6.   Hughey, C. A.; Hendrickson, C. L.; Rodgers, R. P.; Marshall, A. G. *Anal. Chem.* **2001,** *73,* 4676-4681.

7.   Zubarev, R. A.; Hakansson, P.; Sundqvist, B. *Anal. Chem.* **1996,** *68,* 4060-4063.

8.   Shi, Y.; Bajrami, B.; Morton, M.; Yao, X. *Anal. Chem.* **2008,** *80,* 7614-7623.

9.   Bruce, C.; Shifman, M. A.; Miller, P.; Gulcicek, E. E. *Anal. Chem.* **2006,** *78,* 4374-4382.

10.   Artemenko, K. A.; Zubarev, A. R.; Samgina, T. Y.; Lebedev, A. T.; Savitski, M. M.; Zubarev, R. A. *Anal. Chem.* **2009,** *81,* 3738-3745.

11.   Pourshahian, S.; Limbach, P. A. *J. Mass Spectrom.* **2008,** *43,* 1081-1088.

12.   Renard, B. Y.; Kirchner, M.; Steen, H.; Steen, J. A. J.; Hamprecht, F. A. *BMC Bioinformatics.* **2008,** *9,* 355.

13.   Zhang, Z.; Marshall, A. G. *J. Am. Soc. Mass Spectrom.* **1998,** *9,* 225-233.

14.   Levander, F.; Rognvaldsson, T.; Samuelsson, J.; James, P. *Proteomics.* **2004,** *4,* 2594-2601.

15.   Ding, Q.; Xiao, L.; Xiong, S.; Jia, Y.; Que, H.; Guo, Y.; Liu, S. *Proteomics.* **2003,** *3,* 1313-1317.

16.   Mann, M. Useful Tables of Possible and Probable Peptide Masses. *Abstract from the 43rd Annual Conference on Mass Spectrometry and Allied Topics*; Atlanta, GA, 1995. (American Society for Mass Spectrometry).

17.   Gay, S.; Binz, P.-A.; Hochstrasser, D. F.; Appel, R. D. *Electrophoresis.* **1999,** *20,* 3527-3534.

18.   Zubarev, R. A.; Bonddarenko, P. V. *Rapid Commun. Mass Spectrom.* **1991,** *5,* 276-277.

19. Yao, X.; Diego, P.; Ramos, A. A.; Shi, Y. *Anal. Chem.* **2008,** *80,* 7383-7391.

20. Dodds, E. D.; An, H. J.; Hagerman, P. J.; Lebrilla, C. B. *J. Proteome Res*. **2006,** *5,* 1195-1203.

21. Mann, M.; Kelleher, N. L. *PNAS.* **2008,** *105,* 18132-18138.

22. Ding, J.; Shi, J.; Poirier, G. G.; Wu, F.-X. *Proteome Science.* **2009,** *7,* 9.

23. Senko, M. W.; Beu, S. C.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom*. **1995,** *6,* 229-233.

24. Anderson, N. L.; Polanski, M.; Pieper, R.; Gatlin, T.; Tirumalai, R. S.; Conrads, T. P.; Veenstra, T. D.; Adkins, J. N.; Pounds, J. G.; Fagan, R.; Lobley, A. *Mol. Cell. Proteomics.* **2004,** *3,* 311-326.

25. Fung, K. Y. C.; Glode, L. M.; Green, S.; Duncan, M. W. *Prostate.* **2004,** *61,* 171-181.

26. Siuzdak, G. *The Expanding Role of Mass Spectrometry in Biotechnology*, 2nd Ed.; MCC Press: SanDiego, CA, 2006; p 240.

27. Barsnes, H.; Eidhammer, I.; Cruciani, V.; Mikalsen, S.-O. *Eur. J. Mass Spectrom.* **2008,** *14,* 311-317.

28. Olsen, J. V.; Ong, S.-E.; Mann, M. *Mol. Cell. Proteomics.* **2004,** *3,* 608-614.

29. Karlin, S.; Blaisdell B. E.; Bucher, P. *Protein Engineering.* **1992,** *5,* 729-738.

30. Huang, P.;Wall, D. B.; Parus, S.; Lubman, D. M. *J Am. Soc. Mass Spectrom.* **2000,** *11,* 127-135.

FIN.