

THE SEPARATION AND RETENTION TIME PREDICTION OF MODIFIED AND
UNMODIFIED PEPTIDES USING HILIC-MS

by

MAJORS JOHN BADGETT

(Under the Direction of Ron Orlando)

ABSTRACT

Liquid chromatography paired with mass spectrometry (LC-MS) is one of the most commonly used analytical techniques for the study of proteins and their modifications. Many of these modifications play an important role in the structure, function, or interaction of proteins, and need to be separated and identified to understand their impact or abundance. While reversed-phase (RP) chromatography had long been the chromatographic method of choice for proteomics, we show that hydrophilic interaction liquid chromatography (HILIC) is not only capable of separating polar peptides that do not retain on RP columns, but that it can allow for accurate retention time prediction.

This work describes the development and validation of a HILIC peptide prediction model that is capable of predicting the retention of native peptides or peptides with common hydrophilic modifications such as glycosylation (N- and O-linked), deamidation of asparagine, isomerization of aspartic acid, and oxidation of methionine. Coefficients describing the extent of hydrophilicity were derived for all of the amino acids and modifications, and it was found that the prediction of peptide retention is incredibly accurate ($R^2 = 0.946$). Predicted retention times can be calculated by summing the coefficients for a peptide sequence, and this has the potential

for quicker analyses, as peptides can be identified by their mass-to-charge ratio (m/z) and retention time, and can also increase in the confidence of identifications including isomeric structures as well as allow for less time spent looking for the peptide of interest in targeted approaches.

The separation of peptides with these modifications is also detailed, as HILIC is capable of fully separating native peptides from their modified forms to allow for quantitation. While some of the modifications add a substantial mass to the peptide, others, such as deamidation of asparagine or isomerization of aspartic acid, have minimal to zero mass difference and barely change the overall structure of the peptide, which can be difficult to identify simply using MS. Thus, chromatography needs to be utilized to separately identify the similar analytes. In addition to showing complete separation between native and modified peptides, we show that the separation using HILIC is consistent and predictable.

INDEX WORDS: Liquid Chromatography, Mass Spectrometry, Hydrophilic Interaction
Liquid Chromatography, Proteomics, Post-Translational Modifications,
Peptide Retention Prediction, Deamidation, Oxidation, Glycosylation

THE SEPARATION AND RETENTION TIME PREDICTION OF MODIFIED AND
UNMODIFIED PEPTIDES USING HILIC-MS

by

MAJORS JOHN BADGETT

BS, Sewanee: The University of the South, 2012

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial
Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2017

© 2017

MAJORS JOHN BADGETT

All Rights Reserved

THE SEPARATION AND RETENTION TIME PREDICTION OF MODIFIED AND
UNMODIFIED PEPTIDES USING HILIC-MS

by

MAJORS JOHN BADGETT

Major Professor:	Ron Orlando
Committee:	Jon Amster
	Josh Sharp

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2017

DEDICATION

To my two biggest fans: Mary Elizabeth and mom. Mary Elizabeth, it is a testament to your dedication and constant support that you are willing to be with me after spending four long years apart. You mean everything to me and have shown me how to be a better person in every way. The countdown is over, the bells will be rung, and our life anew has just begun. Mom, you raised two out-of-control boys all by yourself and instilled valuable life lessons along the way. From playing my music in the kitchen to be willing to patiently listen to my research, I could not have accomplished anything in life without you. Both of you are the smartest, most loving, and hardest working people that I have ever known. I am forever thankful that you are a part of my life and that you have molded me into the man I am today.

ACKNOWLEDGEMENTS

First of all, I would like to thank my adviser, Ron Orlando, for allowing me to be a part of his research and encouraging me to think outside of the box. Some of my fondest memories during my time as a graduate student have been the impromptu conversations about anything and everything, and I am constantly amazed at your seemingly endless knowledge. You have helped me grow into a scientist that is excited at the prospect of solving an impossible challenge, and I cannot thank you enough for the way you have guided me. To Barry Boyes, thank you for sharing your vast knowledge of chromatography with me and helping me with almost all of my projects. I would also like to thank Josh Sharp, who asked some of the hardest questions I have ever been faced with that helped me look at problems in different ways. To Jon Amster, thank you for challenging me to go above and beyond.

To DJ, thank you for setting me up in a position to succeed and being patient as I made mistake after mistake. You did the best possible job of teaching me the fundamentals of my research. To Judy Willis, for her support and willingness to work with me for over four years. To all my family and friends who were willing to listen to the ramblings of a mad scientist.

Thank you to the University of Georgia Chemistry Department for taking a chance on me when my future was uncertain. To the Complex Carbohydrate Research Center for providing me with the resources necessary for me to obtain this degree in an unbelievable facility. Thank you to all the former and current group members that have stayed late with me, helped me in their spare time, and have become my friends.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	v
CHAPTER	
1 INTRODUCTION.....	1
2 LITERATURE REVIEW	5
3 PEPTIDE RETENTION PREDICTION USING HYDROPHILIC INTERACTION LIQUID CHROMATOGRAPHY (HILIC) COUPLED TO MASS SPECTROMETRY	29
4 THE SEPARATION AND QUANTITATION OF PEPTIDES WITH AND WITHOUT OXIDATION OF METHIONINE AND DEAMIDATION OF ASPARAGINE USING HYDROPHILIC INTERACTION LIQUID CHROMATOGRAPHY WITH MASS SPECTROMETRY (HILIC-MS).....	55
5 THE USE OF A HILIC PEPTIDE RETENTION PREDICTION MODEL TO ANALYZE THE ISOMERIZATION OF ASPARTIC ACID AND HOW IT IS AFFECTED BY THE ADJACENT C-TERMINAL RESIDUE.....	80
6 PREDICTING THE RETENTION BEHAVIOR OF SPECIFIC O-LINKED GLYCOPEPTIDES	96
7 RETENTION TIME PREDICTION OF N-LINKED GLYCOPEPTIDES FROM HUMAN IMMUNOGLOBULIN GS USING HILIC-MS/MS	112
8 CONCLUSIONS	129

TABLE OF CONTENTS (CONT.)

	Page
REFERENCES	132

CHAPTER 1

INTRODUCTION

Liquid chromatography paired with mass spectrometry (LC-MS) is an essential technique for the analysis of proteins and carbohydrates.¹⁻⁴ However, there will always be a need for more efficient separations, faster sample analysis, or methods that will increase the confidence in biomolecular identifications. Reversed-phase (RP) chromatography is the go-to separation method of choice for proteomic LC-MS analysis due to its ability to separate hydrophobic peptides in a reproducible manner, and many models have been made that accurately predict the retention of peptides based on amino acid composition.⁵⁻⁷ Hydrophilic interaction liquid chromatography (HILIC) has recently shown that it is capable of separating peptides that do not retain on RP columns as well as peptides that have common hydrophilic modifications, such as the oxidation of methionine, the deamidation of asparagine, or N-linked and O-linked glycosylation.^{8,9} These modifications can affect a protein's function, stability, or activity, so the detection, separation and quantitation of these modified peptides are vital for protein biotherapeutics. The purpose of the work presented here is twofold: first, to use HILIC to analyze the separation of unmodified peptides as well as peptides with biologically relevant hydrophilic modifications, and second, to facilitate faster and more confident biomolecular identifications by predicting chromatographic retention using HILIC.

Chapter 3 details the creation of a peptide retention model using a penta-HILIC column that has a high correlation coefficient (0.946). Linear regression analysis was used to derive amino acid coefficients from a dataset of 297 peptides and an equation was created that allows

for the summation of coefficients to predict the retention of peptides, which in turn can enable faster analyses, more confident identifications, and even isomeric identifications. These coefficients are related to the hydrophilicity of the amino acid, as lysine, arginine, and histidine influenced retention the most and phenylalanine, tryptophan, and leucine influenced retention the least. It was discovered that hydrophobic peptides at the N-terminus and one residue from the N-terminus eluted earlier than predicted, and optimized coefficients were created to account for this. In addition, the size of the peptide has an influence on peptide retention, as peptides larger than 15 amino acids in length deviated 3-4 times more from their predicted retention time values than shorter peptides.^{6,10-12} Finally, dextran was used as a retention time calibrant so that the model could be universal, and although this was tested on a separate LC-MS system with a different column size, column temperature, flow rate, gradient slope, and length of analysis, the retention times between the two systems were within only 3.73% of each other.

In Chapter 4, two hydrophilic post-translational modifications were analyzed: the deamidation of asparagine and the oxidation of methionine. Both of these modifications can accumulate over time, and their identification is vital because they can affect protein function, structure, and stability.^{13,14} It was found that HILIC is able to baseline separate peptides with both modifications from their unmodified counterparts, and coefficients describing the extent of their hydrophilicity were derived and incorporated into the peptide prediction model created in Chapter 3. While RP chromatography experienced nearest-neighbor effects that resulted in inconsistent shifts in retention time from modified to unmodified peptides, HILIC separation of the same peptides was shown to be consistent. Deamidation of asparagine can result in the isomeric products aspartic acid (*n*-Asp) and isoaspartic acid (isoAsp), which can be difficult to individually identify in a mass spectrometer without sufficient separation. HILIC was found to

be able to separate synthetic peptides with the same sequence that were either unmodified, modified with *n*-Asp or modified with isoAsp, and it was found that the peptide with isoAsp was the most hydrophilic, eluting the latest.

As discussed in Chapter 4, deamidation of asparagine residues can form *n*-Asp or isoAsp, but aspartic acid can also isomerize to form isoAsp, albeit at a slower rate.^{15,16} This can lead to negative effects, such as age-related or neurodegenerative diseases.¹⁷⁻¹⁹ The rate of isomerization is also affected by the neighboring amino acid on the C-terminal side of the aspartic acid residue, as amino acids that are small with low steric hindrance such as glycine, or amino acids that are hydrophilic will increase the rate.^{20,21} Chapter 5 focuses on the analysis of the extent of isomerization of aspartic acid, as peptides from immunoglobulin-Gs (IgGs) and standard proteins such as cytochrome C, myoglobin, and transferrin were run on a HILIC column. Actual retention times of peptides with aspartic acid residues were compared to predicted ones derived in Chapters 3 and 4, and it was found that the majority of peptides with “DG” motifs correlated with the presence of isoAsp, as well as peptides containing the hydrophilic “DE” and “DS” motifs, and the hydrophobic “DL” motif. Selected reaction monitoring (SRM) experiments were performed to verify the presence of either *n*-Asp and/or isoAsp, and examples of DG peptides were shown that only contained one peak that matched up with isoAsp predicted times.

Chapter 6 focuses on O-linked glycosylation, as peptides with N-acetylglucosamine (GlcNAc), N-acetylgalactosamine (GalNAc), or fucose were separated using HILIC, and coefficients describing the extent of their hydrophilicity were derived and incorporated into the peptide model discussed in Chapter 3. Glycans in general are incredibly hydrophilic, so they interact with the hydrophilic stationary phase in HILIC to a high degree. This interaction allows for the separation of peptides that have these modifications from their native counterparts, which

can heighten glycopeptide analysis. It was found that O-GlcNAc and O-GalNAc were the most hydrophilic, followed by O-fucose. Because O-GlcNAc and O-GalNAc are similar in structure and hydrophilicity, their coefficients were very close together. However, all of these modifications were on par with the most hydrophilic amino acids in the peptide retention prediction model.

N-glycosylation on IgGs was analyzed in Chapter 7 using HILIC columns. Here, N-glycosylated peptide retention times were predicted using the peptide RT model discussed in Chapter 3 in conjunction with an in-house model that predicts native glycan retention. Predicting the retention for these glycopeptides could allow for isomeric and low-abundance identification, as well as being able to reducing the overall complexity of glycopeptide identification. Our previous peptide retention prediction model described in Chapter 3 was shown to work for unglycosylated IgG peptides on two separate LC-MS systems with vastly different chromatographic conditions. Predicted times for N-glycosylated peptides were shown to be consistently 1.881 glucose units (GU) less than actual retention times for all but one glycan structure, A3G1, which has a bisecting structure. Nevertheless, there was a low average deviation (0.368 GU) of predicted to actual retention times when accounting for the two GU difference, and GalNAc linkage isomers were able to be separated and identified using the combined HILIC prediction models.

CHAPTER 2

LITERATURE REVIEW

Mass Spectrometry

Mass spectrometry (MS) is a versatile analytical technique that measures a molecule's mass-to-charge ratio (m/z) for qualitative and/or quantitative detection. There are several important requirements for detection, most notably that the analyte has to carry a charge and must be in the gaseous phase. A mass spectrometer is comprised of three main parts: an ion source, which ionizes the molecules in a sample, a mass analyzer, which separates ions based on their m/z ratio, and a detector, which electronically produces a signal representative of the separated ions.^{22,23} There are numerous types of each part that allow for various kinds of experiments or measurements to be performed.

Tandem MS allows for multiple steps of analysis either in space or in time. For tandem MS in space, separate mass analyzers can be used in the same experiment, with separation, selection, or fragmentation occurring in each one. Tandem MS in time techniques trap ions in only one mass analyzer, typically a quadrupole ion trap (QIT) or a Fourier transform ion cyclotron resonance (FT-ICR) instrument, which allows for both the separation and fragmentation in the same space.²⁴ Both of these techniques can provide the selection, fragmentation, and detection of both precursor and fragment ions of analytes using either one or more mass analyzers.

The fragmentation of ions can provide useful structural information that can even help with isomeric identification. While many hard ionization techniques generate in-source

fragmentation, the most popular fragmentation techniques for proteomics are collisional-induced dissociation (CID), and ion-electron or ion-ion interaction techniques such as electron capture dissociation (ECD) and electron transfer dissociation (ETD).^{25,26} In CID, accelerated ions collide with a neutral media, typically argon, nitrogen, or helium gas, and this converts their kinetic energy to internal energy.²⁷ After enough collisions, the peptide bond is broken, producing b and y ions.²⁸ The other techniques mentioned utilize ion interactions with electrons or other ions to induce fragmentation, either by capturing low energy electrons that result in radical cations (ECD), or transferring electrons from donor anions (typically fluoranthene) in a fragmentation cell to positively charged analytes (ETD).²⁹⁻³¹ In contrast to CID, EXD fragmentation techniques produce c and z ions, which allows for the two different techniques to be complimentary for protein identification.

Ionization Techniques

Ionization can either be hard, which produces a large amount of fragmentation, or soft, which produces a low amount of fragmentation. Many of the earlier ionization techniques such as electron ionization (EI) or chemical ionization (CI) are hard. These were primarily used for the detection of small organic molecules and were not ideal for proteomics, where identification of the molecular weight is vital for analysis.^{2,32} The development of soft ionization techniques, such as electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI) opened the door for mass spectrometric analysis of large molecules due to the ability to generate ionized species without a high degree of in-source fragmentation.^{1,33} It also promoted database searching of proteins, where MS/MS data could be searched against theoretical masses for proteins and peptides in a database.^{25,33}

The most common ionization technique used for online LC-MS is ESI because of its ability to quickly evaporate analytes in a solvent as well as that it is performed at atmospheric pressure. In this technique, analytes in a solvent pass through a capillary with a high voltage applied to it, which generates a plume of ions that is introduced into the mass spectrometer.^{23,34} The formation of ions requires ample desolvation, so organic solvents are typically mixed with water to facilitate this. Two theories explain ion formation in ESI: the charge residue model (CRM) and the ion evaporation model (IEM). Dole proposed the CRM in 1968 when he first discovered ESI.³⁵ In this theory, each charged droplet that is formed experiences enough solvent evaporation and Coulomb fission to allow for the resulting smaller droplets to contain one analyte ion each, which are then released into the mass spectrometer after all of the solvent is evaporated.³⁵⁻³⁷ The second theory (IEM) was proposed by Iribane and Thompson in 1976 and details the formation of charged droplets that emit analyte ions when the charge-to-surface area ratio of the droplet becomes high enough.³⁷⁻³⁹ While both of these theories adequately explain ion formation in ESI, the IEM is generally linked to smaller molecules and the CRM is generally linked to macromolecules.^{23,37} ESI regularly produces doubly and triply charged ions, which allows for the analysis of larger biomolecules at smaller m/z ratios.

The other commonly used soft ionization technique for the study of biological samples is MALDI, which utilizes an organic matrix that absorbs radiation from a pulsed laser and transfers a charge to analytes embedded in the matrix.⁴⁰ The pulsed laser causes molecules in the solid phase to desorb into the gaseous phase, and then the generated ions are focused into the mass spectrometer for detection. Proteomic identification using MALDI can be more straightforward than ESI due to the large amount of singly charged species generated.^{40,41} The two most commonly used matrices for proteomics are α -cyano-4-hydroxycinnamic acid (CHCA) and 2,5-

dihydroxybenzoic acid (DHB), which can both produce varying amounts of fragmentation.⁴²⁻⁴⁴

The most common mass spectrometer MALDI is paired with is the time-of-flight (TOF) instrument due to its large m/z range and the ability to accurately time the pulse of the laser with the measurement of an ion's flight time.⁴⁵

Mass Analyzers

The mass analyzer portion of a mass spectrometer is where the separation of ions based on their m/z ratio occurs, and there are many different types of analyzers that can accomplish this. Each one has different characteristics that either makes them desirable or disadvantageous for a specific application. However, there is not one mass analyzer that is above all others for every application.

Linear quadrupole mass analyzers utilize four circular metal rods with fixed direct current (DC) and alternating radio frequency (RF) voltages to allow for the analysis of ions with specified m/z ratios. These rods act as a mass filter and generate an electric field in which stable ions can oscillate. The ion(s) of interest can then either be scanned into the detector or into another mass analyzer by changing the voltage. Ions outside of the specified m/z range will have an unstable trajectory and not reach the detector by colliding with the rods.⁴⁶ The most commonly used quadrupole instrument is the triple quad (QQQ), which has three sets of quadrupoles; the first for selecting precursor ions, the second for fragmentation using CID, and the third for filtering the fragment ions.^{47,48} QQQ instruments provide exceptional selectivity of analytes and have a high reproducibility, but can have limited resolution.⁴⁷

A similar mass analyzer is the quadrupole ion trap (QIT), which can either be linear or three-dimensional. The 3D version of the quadrupole has a ring electrode between two end cap electrodes. Stable ions are trapped between these electrodes that have applied DC and RF

potentials and then can be axially ejected into the detector by adjusting the RF potential.^{49,50} An inert buffer gas, usually helium, is pumped into the trap to dampen the energy of the ions so that they do not hit any of the walls. One of the beneficial traits of the QIT is that it allows for ion accumulation, which can increase the sensitivity. The two dimensional QIT has the same setup as the linear quadrupole but includes front and rear trapping plates that change potentials to allow for ion accumulation. Ions are accumulated by ramping the potential on the rear plate up so that they are repelled if they try to exit the trap.⁴⁶ After enough ions enter the trap, the front plate potential is also ramped up, trapping the ions inside the quadrupole.²² The rear plate potential is then reduced to attract ions so that they exit for the next stage of analysis. This technique also utilizes collisional cooling with a buffer gas and can inject ions either axially or radially depending on the design.^{22,46}

Linear time-of flight (TOF) instruments measure the time it takes ions to travel in a straight line from the ion source to the detector in a flight tube of known length. This measurement is based on the ion's m/z ratio, as smaller ions will reach the detector faster than larger ones. After ionization, ions are accelerated by a voltage towards a detector, which is usually a microchannel plate (MCP) due to the spread of the ions traversing the flight tube.²² Although quadrupole and magnetic sector instruments overshadowed TOF instruments after they were first commercialized, the emergence of MALDI as a pulsed ionization technique in the 1990s helped vault TOF back into the mainstream.⁵¹ MALDI made it possible for TOF mass spectrometers to generate fast spectra from ions pulsed at the same time, which is vital for accurate measurements. There are many benefits to TOF machines, including the ability to rapidly generate spectra, a massive m/z range, and excellent ion transmission.^{22,52-55}

The development of two techniques has helped improve resolution in TOF instruments by correcting for the distribution of energies in ions: delayed extraction, and the use of a reflectron. Using the delayed extraction technique, ions generated from the ion source are delayed momentarily before they are accelerated in the flight tube. This makes up for the initial position and velocity of ionized species, as lower kinetic energy ions start closer to the electrode applying the acceleration voltage and have faster initial velocities that make them reach the detector at roughly the same time as higher kinetic energy ions that have a slower initial velocities.^{51,56} Reflectron instruments utilize a retarding electrostatic field, in which pulsed ions penetrate until they lose their kinetic energy and travel in the opposite direction to reach a detector. This corrects for the distribution of kinetic energies, as ions with more energy travel farther into the reflectron and reach the detector at the same time as ions with lower energies that do not penetrate the reflectron to the same degree.^{51,55,56}

Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometers use large magnetic fields to measure the orbit (or cyclotron frequency) of ions that are trapped. Smaller ions travel in faster orbits than larger ones. After RF pulses excite the ions, they will travel close to detector plates that will produce specific image currents for each ion. These image currents can be amplified and turned into unique m/z ratios by performing a Fourier transform. Although these machines provide the highest mass resolution out of any type of mass analyzer, they have a limited dynamic range and require extremely low-pressure systems.⁵⁷⁻⁵⁹

Orbitrap instruments are similar to FT-ICR mass spectrometers, as an outer electrode detects the image current of trapped ions that are orbiting a spindle-like electrode and the resulting signal is converted by performing a Fourier transform.^{60,61} Although only recently commercialized in 2005, Orbitrap instruments differ from FT-ICR instruments in the most

expensive requirement: they do not require a magnetic field. Injected ions are electrostatically attracted to the central electrode, but experience radial and axial fields that make the ions travel in an elliptical-like path around the central electrode.²² This mass analyzer is commonly paired with linear ion traps so that a high mass accuracy of precursor ions can be generated in the Orbitrap portion and fast fragmentation data can be generated in the other portion.⁶¹

Liquid Chromatography

Chromatography is the process of separating a mixture into individual components by utilizing analyte interactions with a stationary phase and a mobile phase. These interactions can be chemical, such as an analyte adsorbing to an adsorbent surface, or physical, such as a small analyte easily penetrating a porous particle. There are many considerations when choosing the type of chromatography to use in a separation, such as the system used, the stationary phase, mobile phase composition, size of the column, type and size of the particles, isocratic or gradient elution, flow rate, analyte structure and composition, among others.⁶² All of these considerations need to be applied to generate a sufficient separation for a specific mixture.

Although there are numerous types of stationary phases such as normal phase, HILIC, and hydrophobic interaction chromatography (HIC), reversed-phase (RP) chromatography is the most commonly used stationary phase for proteomic separations based on its ability to retain non-polar analytes.⁵⁻⁷ However, HILIC has recently become more popular as a method of separating polar analytes that do not retain on RP columns.^{63,64} The high complexity of biological samples requires the need for efficient separation strategies to maximize identification, which can be accomplished by fine-tuning chromatographic conditions or using additional modes of separation. For example, the selectivities of HILIC and RP are not in direct opposition to one another so they can be used in tandem to generate more complex separations.^{8,9,65}

The measurement of a column's efficiency can be put in terms of band broadening, which describes how a group of the same type of analyte moves apart as it advances through a column.⁶⁶ Van Deemter created an equation to calculate this in 1956, and related column efficiency with the height equivalent to a theoretical plate (H), or a measurement of the equilibration of a sample between stationary and mobile phases:

$$H = A + \frac{B}{u} + Cu \quad \text{Equation 1}$$

In this equation, there are three terms that influence the height of a plate: Eddy diffusion (A), longitudinal diffusion (B), and resistance to mass transfer (C), with mobile phase velocity (u) also playing a role. For a column to have a high efficiency, the plate height needs to be low, which would require the factors that influence band broadening needing to be minimized.⁶⁷ Eddy diffusion describes the multiple paths that an analyte can take through a column. These paths can differ based on the column packing or particle size, which results in multiple paths and a wider distribution of analyte elution.⁶⁸⁻⁷⁰ This term can be reduced with well-packed columns as well as particles that are consistently small. The second term, longitudinal diffusion, describes the diffusion of an analyte from the center of a band to the edges. Higher flow rates will reduce this term and limit the amount of band broadening that results.⁷¹ Resistance to mass transfer is the final term that influences band broadening, and it represents the movement of an analyte between the stationary and mobile phases, or a transfer of masses between the phases. For example, some analytes will move from phase to phase while others will be predominantly in the mobile phase. This will increase the width of the band, but can be reduced by using lower flow rates, using smaller particles, or increasing the temperature.⁷¹⁻⁷³

The mobile phase velocity (flow rate) is very important to consider as it can simultaneously increase and decrease the amount of band broadening through longitudinal diffusion and resistance to mass transfer. Flow rates can range from milliliters to hundreds of nanoliters with the sensitivity of detection increasing at the lower rates. However, the minimal amount of flow can lead to complications such as increased time of separation, poor peak resolution, and not being able to detect the presence of a leak. There is a clear give-and-take relationship in terms of flow rate between the time of analysis, the amount of band broadening, the sensitivity of detection, and peak resolution, so it is important to decide what aspects of the separation matter most.⁷³

Plate height can also be calculated using the equation below, where L is the column length and N is the number of plates:^{67,74}

$$H = \frac{L}{N} \quad \text{Equation 2}$$

A higher number of plates and lower column length will result in a small plate height, which means that the column has a high efficiency. However, to determine the plate height, the number of plates needs to be calculated first:^{74,75}

$$N = 5.54 \left(\frac{t_R}{w_b} \right)^2 \quad \text{Equation 3}$$

This equation takes into account the peak width at half height (w_b) for an analyte with a specific retention time (t_R). Columns that have narrow peaks have a larger number of plates and higher column efficiencies.

While column efficiency is necessary in determining the degree of separation within a column, being able to separate two adjacent peaks is just as important. The extent of this separation is called resolution, which is defined as the ability to differentiate two peaks that are next to each other from one another:⁶⁷

$$R_s = \frac{2(t_{R,2} - t_{R,1})}{(w_{b,1} + w_{b,2})} \quad \text{Equation 4}$$

The difference in retention for two peaks is divided by the peak widths to determine the degree of resolution. Peaks that are narrower and have higher differences in retention times will result in a larger resolution because they most likely will not be overlapping. It is commonly recognized that a resolution of 1.5 or over will result in baseline separation between the two peaks, while coelution will occur with resolutions under 1.5.⁷⁶ For the analysis of biological samples with LC-MS, it is important to get significant separation so that different peptides will not coelute and cloud the mass spectra.

Reversed-Phase Chromatography

Reversed-phase (RP) chromatography has long been the most commonly used chromatographic technique paired with mass spectrometry and the go-to separation method of choice for proteomics.⁵⁻⁷ This is due to its high reproducibility and resolution, as well as the simplicity with which the column's selectivity can be adjusted through changes in the mobile phase.⁷⁷ Another beneficial trait is that the stationary phase consists of long chain hydrocarbons (typically octadecyl moieties) covalently bound to silica particles, which adequately separate nonpolar species, including predominantly hydrophobic peptides.⁷⁸ RP chromatography can

rapidly and efficiently separate complex digested peptides prior to MS/MS analysis, allowing for high-throughput proteomic experiments.

The separation mechanism of RP chromatography is based on the hydrophobic adsorption of the analytes onto the surface of the stationary phase as well as the partitioning of fully embedded analytes within the hydrocarbon chains.^{79,80} More hydrophobic analytes will retain longer than less hydrophobic analytes, and can be eluted by the addition of organic solvent (typically acetonitrile or methanol) in the mobile phase using either an isocratic or gradient separation. Acetonitrile is primarily used as the organic solvent in RP due to its compatibility with electrospray ionization (ESI)-MS, its low viscosity that reduces the amount of backpressure, and low absorbance for UV applications.⁶⁴

Buffering salts are commonly used in RP chromatography to increase the ionic strength of analytes, which in turn increase the overall separation. Ammonium formate and ammonium acetate are most commonly used due to their accord with MS, but the retention behavior of analytes can vary as a product of which salt is used.⁶⁴ Another common addition to buffers are mobile phase modifiers such as formic acid, which be used to adjust the pH of the buffer and change the charge state of analytes. This is useful for proteomics, where pH conditions change the interactions of charged amino acids with the stationary phase and can induce different retentions and more complex separations. These pH modifiers can also provide an increased sensitivity in MS from enhanced protonation of analytes.⁸¹

Hydrophilic Interaction Liquid Chromatography

Hydrophilic interaction liquid chromatography (HILIC) combines the mobile phases used in RP chromatography with the stationary phases used in normal phase chromatography (NP) to allow for the separation of polar or charged species. Although HILIC is not as characterized as

RP chromatography, it has shown to be able to separate polar analytes that are not retained on RP columns, and is a useful and complimentary approach that can be used in conjunction with RP chromatography.^{63,64,82} The most commonly used stationary phases for HILIC include bare silica or silica derivatized with polar stationary groups, as well as zwitterionic stationary phases that can induce even more complex separations.⁸³⁻⁸⁶ Due to HILIC's ability to separate polar analytes, new stationary phases with different structures and charges can help improve the coverage of more complex analytes that can be separated by this technique.⁸⁶

Adjacent to the bulk organic mobile phase is a water-rich layer that exists at the surface of the stationary phase. Analytes are retained three different ways: partitioning between this water-rich layer and the bulk mobile phase, adsorption onto the hydrophilic stationary phase, and a combination of partitioning and adsorption.^{84,87,88} However, the partitioning mechanism is still relatively unknown.⁸³ More polar analytes will be retained longer due to the strong interactions with the stationary phase in the water-rich layer, which include electrostatic interactions, hydrogen bonding and van der Waals forces.^{64,88}

Acetonitrile is the main organic buffer of choice for HILIC for the same reasons it is the primary organic buffer for RP chromatography. However, because the bulk mobile phase will primarily be organic in HILIC rather than aqueous in RP, faster mobile phase velocities can occur, which reduces longitudinal diffusion (B-term in Van Deemter equation) without sacrificing performance. This heightened viscosity can also lead to longer columns being used with lower back pressures as well as enhanced sensitivity with ESI-MS.⁸⁴ One drawback to using acetonitrile as the main solvent for HILIC is that many analytes are insoluble at high percentages of acetonitrile, potentially making sample injection rather difficult.

Peptide Retention Prediction

When the first peptide retention prediction models were introduced, their purpose was to analyze the individual contributions of amino acids to the retention of a peptide in hopes of accurately predicting when peptides would elute. Many such models were created for RP and normal phase chromatography, where each amino acid would have a coefficient that described the extent of its hydrophobicity/hydrophilicity.^{8,89-101} The coefficients could then be summed together for a specific peptide (usually with an intercept) to calculate the predicted retention time for any peptide sequence. Linear regression analysis is the most commonly used method of derivation for the coefficients, but other methods such as substituting amino acids on a synthetic peptide or even using MATLAB[®] have been used.^{89-94,96-98,100} While most of these models predict retention in terms of actual time, there are some that use the percentage of organic solvent as well.⁹⁹ This is an attempt of trying to allow a model to be universal in a sense; as long as the stationary and mobile phases are the same, the model could work on different LC-MS systems, using different flow rates or gradient slopes and still accurately predict where the peptide would elute. However, amino acid coefficients can change with different stationary phases and different operating conditions such as pH, temperature, among others, so the need for new models to be created for new types of stationary phases and operating conditions still exists.

Apart from simply describing the hydrophobicity/hydrophilicity of an amino acid, peptide retention models can help further the identification process by increasing the speed of analysis, increasing the confidence of peptide identifications by eliminating false positives, and separately identifying isomeric peptides when fragmentation data is insufficient.^{91,98} These models can identify peptides based on their m/z ratio and predicted retention times in a similar way to accurate mass and time tag (AMT) proteomics, which eliminates the need for standard

database searching.¹⁰²⁻¹⁰⁴ Targeted approaches such as SRM or MRM techniques can also benefit from these models, as the time spent looking for peptides of interest can be shortened to a time range that is close to where the predicted retention time of the peptide is.

RP Models

The first peptide retention prediction models were made for RP chromatography due to its widespread use in proteomics. In 1977, Molnar and Horvath were the first people to note that it should be possible to obtain estimates of the hydrophobicities of the amino acids within a peptide, and this discovery opened the door for O'Hare and Nice to relate peptide retention to the sum of the coefficients for the most hydrophobic amino acids in a peptide.^{78,94} Shortly thereafter, Meek created the first complete peptide retention prediction model in 1980, where coefficients were derived for each individual amino acid.⁹³

Meek's model prompted other researchers to examine the many factors that influence peptide retention on two different sides: the peptide side and the analytical side.^{11,12} The peptide side contains characteristics relating to peptide composition and structure, such as the position of amino acids within the peptide, the amino acids at the N- or C- termini, the total length of the peptide, and the total size. The analytical side contains characteristics relating to the method of separation: the stationary phase, the size of the column, the mobile phase composition, the temperature, the pH, and many other characteristics. Mant found that peptides over 15 amino acids in length tend to deviate from predicted retention times to a greater extent than shorter peptides, and other studies have suggested that peptide length and observed retention time have an exponential relationship.^{11,105} This could be because longer peptides have a greater probability of forming secondary or tertiary structures, which can change their interaction with the stationary phase because certain amino acids would be shielded from contact.¹⁰⁶ One possible solution to

this problem is manipulating temperature, as higher temperatures could be used to unfold these peptides so that they interact in more of a predictable manner.¹⁰⁷

The charge of an amino acid can greatly affect retention, so pH is important to consider when performing a separation. Charged residues such as histidine, lysine or arginine can be very hydrophilic when charged, but less-so if the pH is raised. Sequence corrections, size corrections, or the incorporation of coefficients for post-translation modifications have also been added to models, expanding their application and allowing for more accurate predictions. It was even found that different amino acid retention coefficients would have to be created for each position on a peptide because peptides with differing orders of amino acid sequence but the same composition had varying retention times.¹⁰⁸ The amino acid residues that increase retention by the greatest amount in RP chromatography are the non-polar, aromatic or aliphatic ones such as tryptophan, tyrosine, phenylalanine, leucine, and isoleucine. These residues interact with the non-polar stationary phases more than small or polar residues, which either decrease the retention or have a negligible effect. Because C18-bonded silica is the predominant stationary phase used for RP chromatography, subsequent models have focused more on specific peptide or analytical characteristics and less on the type of stationary phase.

HILIC Models

Even though most of the prediction models that have been made are for RP chromatography, the number of HILIC models has increased as the number of different types of HILIC columns has also increased.^{8,91,92,101} HILIC has shown to be highly compatible with ESI-MS in addition to being able to separate peptides, which has furthered its popularity.¹⁰⁹ Yoshida introduced the first HILIC peptide retention prediction model in 1998 on an TSK Amide-80 column under normal-phase terminology.⁶⁵ This was many years after the first RP model was

made because HILIC was introduced later and was not used nearly as much in proteomics. Afterwards, other researches created HILIC models for different types of stationary phases, notably Gilar, who derived amino acid coefficients for three different types of HILIC columns: bare silica, bridge-ethyl hybrid silica and an amide modified bridge-ethyl hybrid silica.⁸ What was discovered was that the different stationary phases have different selectivities and produced coefficients that were not the same. This supports the need for the derivation of amino acid coefficients for specific mobile phase and stationary phase operation in HILIC because the types of stationary phase are much more diverse than RP chromatography. Another key discovery is that even though the derived amino acid coefficients for the HILIC model generally matched up to the inverse of the amino acid coefficients from RP models, it is not necessarily a linear correlation, indicating that HILIC and RP can be combined for more complex separations in multidimensional HPLC.⁸

Polar, charged residues such as histidine, lysine, and arginine have the greatest effect in HILIC and generally have the largest coefficients overall because they have side chains that interact more with the polar stationary phase and water-rich layer through adsorption and partitioning. The N- and C- termini also have polar characteristics, which can be result of HILIC models containing fairly high intercepts. Small residues or residues that contain both hydrophilic and hydrophobic characteristics such as glycine, valine, proline, or methionine tend to have a negligible effect on retention. However, similarly to RP models, these coefficients can change with different operating conditions.

Hydrophilic Protein Modifications

Modifications can alter the function, activity, or interaction of proteins through enzymatic cleavage or the extension of an amino acid's functional group. For protein biotherapeutics, the

quantitation and separation of modified species is vital in order to determine the extent of a protein's stability or activity because modifications can accumulate during storage or transportation.^{13,14,110} Many of these modifications increase the hydrophilicity of the altered residue, such as the deamidation of asparagine, the oxidation of methionine, or the glycosylation of serine and threonine residues. HILIC's diverse group of polar stationary phases have the potential to separate these changes in hydrophilicity while allowing for quantitation of unmodified species and their modified counterparts.

In the realm of proteomics using LC-MS analysis, many enzymatically cleaved peptides are modified either on purpose or environmentally over time. For database searching, including variable modifications can vastly increase the search space. This can lead to more false positives and require higher degrees of protein validation, but is also essential in trying to identify modified peptides.^{111,112} Some of these modifications can be virtually indistinguishable from their unmodified counterparts due to minimal differences in mass or fragmentation, which can further complicate protein identification through database searching. However, utilizing chromatography to separate modified from unmodified peaks can greatly improve identification capabilities even on lower resolution mass spectrometers.

Deamidation of Asparagine

The process of deamidation has been viewed as a molecular clock, as this modification has large implications in ageing, protein development, and protein turnover, and has been linked to cataracts, Alzheimer's disease, and β -amyloid aggregation.^{19,113-116} However, the effects of deamidation are not as well known as other common post-translation modifications. For this process, residues form a five-membered succinimide intermediate that can hydrolyze under physiological conditions to form either isoAsp or aspartic acid (*n*-Asp) in both D and L

configurations, with the products having an abundance ratio of roughly 3:1, respectively.^{15,16,115,117,118} It has been shown that identification and isolation of the cyclic intermediate is much more difficult at lower pHs (under 5) because hydrolysis of the intermediate is promoted through acid catalysis at a much faster rate than at neutral or higher pHs.^{119–122} Furthermore, asparagine residues are deamidated at a much faster rate than glutamine residues due to it being more entropically favorable to form the five-membered cyclic intermediate rather than the six, and the rate is further increased when a small residue such as glycine is on the C-terminal side of the modified residue.^{15,117,123} This is a result of low steric hindrance, which allows the residue to fold onto itself to form the cyclic intermediate. In addition, the rate can also be affected from environmental changes, such as an increase at elevated temperature and pH.

Deamidation can be a difficult modification to analyze due to a mass difference of only 1 Dalton (Da) between modified and unmodified forms. It is even more complicated when there is overlap between a modified residue and an unmodified ¹³C isotope, resulting in a 0.0152 Da difference.¹²⁴ In high-throughput experiments using lower resolution mass spectrometers, there can be a high degree of false positives due to this minimal mass difference. There is a significant need for sufficient separation before MS analysis because of this issue, and although RP chromatography can do this, the separation is not consistent due to sequence effects from neighboring residues.¹²⁴ HILIC has shown that it not only can separate modified peptides from their unmodified forms, but also can do so consistently, allowing for straightforward and predictable identification.

Isomerization of Aspartic Acid

Isoaspartic acid, or isoAsp, is most commonly thought of in reference to the process of deamidation, but can also be formed through the isomerization of aspartic acid, or *n*-Asp. The biological impact of the formation of isoAsp through isomerization is the exact same as through deamidation, as it is largely associated with ageing and age-related diseases, as well as neurodegenerative disorders such as Alzheimer's disease.^{17,18,21} Through isomerization, aspartic acid undergoes formation of the same succinimide intermediate as the deamidation of asparagine, without the shift in mass, and results in the formation of isoAsp in both D- and L-configurations. However, the rate of isomerization of aspartic acid can be up to 38 times slower than deamidation of asparagine.^{15,16} The deficiency in mass difference makes it even harder to analyze in a mass spectrometer, where the only fragmentation methods suitable for identification are electron transfer dissociation (ETD) or electron capture dissociation (ECD) that show unique fragment ions for isoAsp at *c*+57 and *z*-57. However, the intensities of these peaks can be almost 95% lower than other peaks in the MS/MS spectra.^{125,126} This promotes the need for chromatographic separation prior to MS analysis so that there is minimal ambiguity in the identification of the isomers.

It has been shown that HILIC can separate *n*-Asp and the more hydrophilic isoAsp, eliminating the need for EXD fragmentation techniques and providing separate isomeric identification in the mass spectrometer.¹²⁴ This greatly simplifies the analysis of peptides with this modification and provides a straightforward and consistent technique that allows for isolation and identification. Other techniques that have been used to analyze aspartic acid isomerization include isotopic labeling and the use of isoaspartic acid O-methyltransferase or

isoAsp-specific antibodies, but all of these techniques have problems that can lead to complications in identification.^{125,127}

Oxidation of Methionine

Oxidation is a much more studied and well-known modification, and has a larger mass difference of 16 Da between modified and unmodified species to allow for easier MS identification. Reactive oxygen species (ROS), which are oxygen-containing molecules that are created through the intake of oxygen in the human body, can oxidize protein residues when a biological system experiences induced or environmental stress.^{128–130} The rate of oxidation is dependent on the amount of ROS present, as well as the level of stress. Although this modification can occur on many different residues including lysine, tryptophan, histidine, arginine, and proline, it is only reversible for the sulfur-containing amino acids cysteine and methionine, with methionine being the most susceptible.^{129,131}

Similarly to deamidation, it is known to affect the stability, structure and activity of numerous proteins including calmodulin, lysozyme, and ribonuclease B, and has also been linked to bronchitis, Parkinson's disease, respiratory distress syndrome, diabetes, emphysema, Alzheimer's disease, and other age-related diseases.^{130,132–137} Through the oxidation mechanism, methionine (Met-S) is oxidized to form the R- and S- isomers of methionine sulfoxide (Met-SO), which can further oxidize to form methionine sulfone (Met-SO₂). Although the first step of this reaction is reversible due to the presence of methionine sulfoxide reductase (either MsrA or MsrB, which reduce the S- and R- isomers, respectively), the second step is not, and the formation of Met-SO₂ in biological systems is scarce.¹³⁰ The reversibility of Met-SO back to Met is very important, because it has been discovered that increased levels of MsrA can increase the lifespan of various mammals and insects.^{129,138,139} In Alzheimer's patients, the levels of MsrA are

significantly lower than normal individuals, and this is indicated by significantly greater amounts of Met-SO.^{129,133,137,140}

Although the difference in mass makes identification of oxidized and unoxidized peptides much easier using MS techniques, the separation of oxidized species using RP chromatography is not consistent and can be troublesome. HILIC can be useful for problem, as oxidized residues have an increased hydrophilicity and will be retained longer on polar stationary phases.

Glycosylation

Glycosylation provides a much more substantial addition of a carbohydrate to the modified residue in comparison to deamidation or oxidation, and this addition changes the m/z ratio, structure, and polarity of a glycosylated moiety to a high degree.^{141,142} Protein glycosylation can either be N-linked, meaning the carbohydrate is linked through a nitrogen atom on an asparagine residue, or O-linked, meaning the carbohydrate is linked through an oxygen atom on a serine or threonine residue. Defects in these modifications can lead to a number of diseases, referred to as congenital disorders of glycosylation (CDG), and these include cancer, heart defects, Peters-Plus syndrome, Walker-Warburg syndrome, among over 50 other diseases that have been identified.^{143,144}

N-linked glycosylation follows the consensus sequence Asn-X-Ser/Thr, where X can be any amino acid excluding proline. The majority of N-glycans linked to asparagine residues are through N-acetylglucosamine (N-GlcNAc) and have a common core pentasaccharide consisting of 3 mannose residues and 2 GlcNAc residues.¹⁴⁵ There are three different types of N-glycans: 1) high mannose, which contain mannose residues attached to the core structure; 2) complex, which can have many different types of sugar moieties attached to the core structure; 3) hybrid, which share traits from the other two classes by having branches that are high mannose or complex. For

the biosynthesis of N-linked glycoproteins, lipid-linked oligosaccharides (LLOs) are first assembled at the membrane of the endoplasmic reticulum (ER) and their structure consists of two GlcNAc, nine mannose, and three glucose residues.¹⁴⁶ Each LLO is then transferred to a nascent polypeptide in the lumen of the ER by an oligosaccharyltransferase (OGT).¹⁴⁶ This occurs as the protein is being translated, so the addition of the glycan has a massive impact to the overall structure of the protein once it is fully formed.¹⁴⁶⁻¹⁴⁸ Afterwards, the sugar moieties on the glycopeptide can be further cleaved or elongated by glycosidases and glycosyltransferases across the ER and the Golgi apparatus to form the three different types of N-glycopeptides, and the number of structural possibilities that result from this process is in the millions.¹⁴⁹ One study by Apweiler discovered that over half of the proteins in the Swiss-Prot database (almost 75,000) were N-glycosylated, and it has been found that N-glycosylation has large implications in a protein's solubility, structure, stability, folding, and cell-to-cell interactions.¹⁵⁰

In comparison to N-linked glycosylation, O-linked glycosylation does not require a consensus sequence or an oligosaccharide precursor for the transfer of the sugar to a protein. Instead, O-linked glycopeptides are formed in the Golgi apparatus post-translation on serine or threonine residues. These modifications are important for protein structure and stability, and are involved in a number of different processes, including lubrication, the aggregation of proteins, inflammation response, among others.^{151,152} The most common type of O-linked glycosylation is mucin-type, which is named for the high density of mucins or mucin-type proteins that these sugars are attached to.¹⁵³ It has been found that O-GalNAc is present on over 85% of proteins passing through the Golgi apparatus and over 10% of human proteins overall.¹⁵³ The addition of an O-GalNAc to a S/T residue is mediated by 20 different polypeptide N-acetylgalactosaminyl transferases (GalNAc-Ts) that use uridine-diphosphate GalNAc (UDP-GalNAc) as the

donor.^{151,152,154} Each glycan can be further elongated or modified, but the extent is far less than that of N-linked glycosylation. O-fucose addition is similar, as it can be elongated by thirteen different fucosyltransferases.¹⁵⁵ Along with O-glucose, O-fucose is commonly found on epidermal growth factor (EGF) like repeats and thompospondin type-1 repeats (TSR) and is heavily involved in Notch signaling.¹⁵⁶⁻¹⁶¹ Another type of O-linked glycosylation is O-GlcNAcylation, which is not further extended after addition to the protein and is added or removed from proteins by only O-GlcNAc transferase (OGT) and O-GlcNAcase (OGA).^{162,163} This modification has been found on over 1,000 proteins, and is highly dynamic as it readily cycles on and off proteins due to environmental changes such as stress.¹⁶⁴⁻¹⁶⁶ The final type of O-glycosylation is O-mannosylation, which is primarily found in yeast and is initiated in the ER instead of the Golgi apparatus, however much less is known about this O-linked modification compared to the others.^{152,167,168} O-mannose is added by protein O-mannosyltransferases (PMTs) using dolichol-phosphate-B-D-mannose as a donor, and can be elongated or branched with different sugar moieties.^{167,168}

Due to their high complexity and lack of automated identification, glycopeptide analysis is incredibly difficult and relies primarily on LC-MS techniques.¹⁶⁹ It is further complicated by microheterogeneity, which is when numerous types of glycans can occupy a glycosylation site.^{141,142} Being able to resolve native peptides from their glycosylated counterparts would lead to a heightened degree of characterization. RP chromatography has traditionally been used to separate peptide/glycopeptide samples but the very hydrophilic glycans interact minimally with the non-polar stationary phase and provide minimal separation from their unmodified versions.^{170,171} Polar stationary phases, like the ones used in HILIC, would greatly increase the

interaction with the glycan and provide a much better separation, potentially resolving peptide and glycopeptide pairs.

CHAPTER 3

PEPTIDE RETENTION PREDICTION USING HYDROPHILIC INTERACTION LIQUID CHROMATOGRAPHY (HILIC) COUPLED TO MASS SPECTROMETRY¹

¹ Badgett, M.J., Boyes, B., Orlando, R. To be submitted to *Journal of Chromatography A*

Abstract

A model that predicts retention for peptides using a HALO® penta-HILIC column and gradient elution was created. Coefficients for each amino acid were derived using linear regression analysis and these coefficients can be summed to predict the retention of peptides. This model's accuracy ($R^2 = 0.946$) is on par with previous RP and HILIC models. Apart from amino acid composition, length and location of amino acid residues on a peptide were examined and a site-specific correction for hydrophobic residues at the N-terminus as well as hydrophobic residues one spot over from the N-terminus was created.

Introduction

The use of hydrophilic interaction liquid chromatography (HILIC) columns has grown tremendously due to the various types of columns available as well as their ability to separate polar analytes. Although reversed-phase (RP) chromatography is the preferred method of choice for proteomic experiments, HILIC is able to separate peptides that are not retained on RP columns, making it a very useful and complimentary technique for proteomic experiments that has even been paired with RP for more complex separations.^{8,9,65}

Standard proteomic experiments have long used chromatography coupled to mass spectrometry for analysis. In these experiments, peptides are identified by their mass-to-charge (m/z) ratio and fragmentation data, which usually involves database searching. While this technique is very common, researchers may have trouble separately identifying peptides with the same m/z ratio in which fragmentation data is insufficient in identification. To this end, chromatography can be used to further the identification process, as retention times of peptides are related to their amino acid sequences. By predicting what the retention would be, peptides can quickly be identified by their m/z ratio as well as their retention time, and peptides with the same mass but different sequences can be identified separately due to differing retention times. This can decrease the time spent in identification as well as increase the confidence of identifications.^{10,98} Targeted approaches can also benefit from this, as the time of analysis spent looking for specific peptides can be shortened.

O'Hare and Nice were the first researchers to notice that peptide retention was directly related to amino acid composition, and this discovery in 1979 opened the door for models that were able to predict the retention of peptides.^{8,10,65,89,90,92-100,172,173} Almost all of these models have been made using RP as the means of separation, but there have been several HILIC models

that have been made recently.^{8,10,65,92} These models derive coefficients for each amino acid, describing their hydrophilic or hydrophobic behavior. When summed together, the coefficients can accurately predict where a peptide will elute on a particular column. The method of derivation of these methods can range from using linear regression analysis to substituting amino acids on a synthetic peptide, and can even include sequence corrections, size corrections, or various modifications.^{10,89,90,92,93,96,100,173} Even though most of the prediction models have been created for RP columns, the number of HILIC models has increased as the types of HILIC columns available have increased throughout the years as well. The first HILIC peptide prediction model was created by Yoshida in 1998 on an TSK Amide-80 column, and then Gilar et. al. created coefficients for three HILIC columns with different stationary phases: bare silica, bridge-ethyl hybrid silica, and an amide modified bridge-ethyl hybrid silica.^{8,65} All of these models have very high correlation coefficients (in the range of 0.92-0.97), indicating that the prediction of peptide retention times using these columns can be extremely accurate. However, they have also shown that the amino acid coefficients can change with different HILIC stationary phases and are also dependent on operating conditions such as pH. Due to this concern, new peptide retention models need to be made for new HILIC stationary phases and specific mobile phases.

In this paper, we have created a HILIC peptide retention prediction model using 297 peptides from various proteomic samples for a HALO® penta-HILIC column. Coefficients for each amino acid have been derived using linear regression analysis and the correlation coefficient is very high (0.94553), indicating the accuracy of this model. We also introduce a site-specific correction for peptides with hydrophobic amino acids at the N-terminus, criteria for peptides selection, and retention expression in glucose units (GU) so that the model can be ran

on any LC-MS system. This useful model will be able to increase protein confidence and reduce the time spent in identification by predicting the retention of peptides.

Materials and Methods

Protein Digestion

Human IgGs were separated from human serum (Sigma-Aldrich, St. Louis, MO, USA) using a HiTrap™ Protein G column (General Electric Company, Fairfield, CT, USA). Myoglobin, transferrin, concanavalin A, fetuin, cytochrome C, lysozyme, ribonuclease B, carbonic anhydride, and dextran were purchased from Sigma-Aldrich (St. Louis, MO, USA). Bovine serum albumin was purchased from Waters (Milford, MA, USA). These proteins as well as yeast proteins, mosquito cuticular proteins, and *H. pylori* proteins were reduced using 10-mM DTT and then alkylated using 55-mM IDA, which were both purchased from Sigma Aldrich (St. Louis, MO, USA). Sequencing-grade trypsin or chymotrypsin purchased from Promega (San Luis Obispo, CA, USA) was added (50:1, w/w, protein/trypsin) and protein samples were incubated overnight.

LC-MS/MS Settings and Instrumentation

Data were acquired using a Finnegan LTQ (Thermo-Fisher, San Jose, CA, USA) and an 1100 Series Capillary LC system (Agilent Technologies, Palo Alto, CA, USA) with an ESI source that used spray tips made in-house. Samples were suspended in 25% H₂O, 75% ACN and 0.1% formic acid (Sigma-Aldrich, St. Louis, MO, USA) and injected into the LC. Peptides were separated using a 200 μm x 150 mm HALO® penta-HILIC column packed with 2.7-μm diameter superficially porous particles (Advanced Materials Technology, Wilmington, DE, USA) as shown in Figure 1. The gradient used for each sample was 95-30% ACN over 90 minutes at a 2 μL/min flow rate. The mobile phase contained 0.1% formic acid (Sigma Aldrich, St. Louis,

MO, USA) and the organic solvent contained 50 mM ammonium formate (Thermo-Fisher, San Jose, CA, USA).

To make sure that this model would be universal, some of the same digested proteins were run on a 4000 Q Trap (AB Science, Chatham, NJ, USA). Peptides were separated by a 2.1 mm x 15 cm Halo penta-HILIC column packed with 2.7- μ diameter superficially porous particles using a Nexera UFLC (Shimadzu, Columbia, MD, USA). The gradient used for each sample was 78-48% ACN over 80 minutes at a 0.4-mL/min flow rate. Spectra were obtained using an ESI source.

Database Search Parameters

The resulting RAW files were converted using Trans-Proteomic Pipeline (Seattle Proteome Center, Seattle, WA, USA), then the MS/MS spectra of each sample were searched using Mascot (Matrix Scientific, Boston, MA, USA) against corresponding protein databases of theoretical MS/MS spectra. The following parameters were utilized in Mascot: a peptide tolerance of 1000 ppm, a fragment tolerance of 0.6 Da, two max missed cleavages of trypsin, and a fixed modification of carbamidomethyl (C).

Selection of Peptides for Prediction Model and Post-Run Data Analysis

All peptides that had a higher Mascot score than 10 were considered. Peptide retention times were found by hand from .RAW files from the apex of the peaks using Xcalibur software (Thermo-Fisher, San Jose, CA, USA), and resulting MS/MS data were visually inspected to verify the peptide assignments. Chromatographic peaks for each peptide had to have a peak asymmetry value of between 0.25 - 4, and peptides exhibiting peak widths greater than 5.5 minutes were excluded from analysis. Peptides had to be fewer than 15 amino acids in length. Peptide retention times in minutes were converted to glucose units based on dextran samples that

were run immediately before. Linear regression analysis using StatPlus (AnalystSoft, Walnut, CA, USA) was used to find the coefficients for each amino acid and 297 peptides overall were used in this study.

Results

Amino Acid Coefficients

Linear regression analysis was used to find coefficients for each amino acid, and these results are shown in Table 1. Using Equation 1, predicted retention times of peptides, R_T , can be calculated, where L_i is the amount of residue i in the peptide, AA_i is the amino acid coefficient of residue i , and b_0 is the intercept of the model. The predicted retention times of the 297 peptides in this model were plotted against their actual times and this is shown in Figure 1. The derived correlation coefficient is 0.94553, which expresses the high accuracy of the amino acid coefficients. This value is on the higher end of previous RP and HILIC peptide retention prediction models.^{8,10,65,89,90,92–100,172,173}

From the coefficients that were derived, lysine, arginine, and histidine dominate the hydrophilic retention, while negatively charged side chains such as aspartic acid and glutamic acid also contribute to peptide retention. Aromatic or aliphatic amino acids such as tryptophan, phenylalanine, isoleucine and leucine contribute to faster peptide elution due to their hydrophobic nature. There were many amino acids that had p-values indicating that they were statistically insignificant and did not contribute to the retention to a great extent. These amino acids are small (i.e. glycine and alanine) or had both hydrophobic and hydrophilic characteristics (i.e. proline and methionine). There is also a large intercept value, which could describe the hydrophilic nature of both the N and C termini on a peptide as well as the time it takes for the un-retained peptides to travel through the column and reach the MS.

All of the coefficients are expressed in glucose units (GU) rather than minutes so that the model can be used on any LC-MS system. Two procainamide-labeled dextran samples were run before each sample, averaged, and then the retention time of peptides in minutes was converted to GU based on the logarithmic fit for the dextran samples. These dextran samples elute in order of increasing monosaccharide linkage and provide a reference for the retention times of peptides that are in minutes, which is shown in Figure 2. This novel approach allows the model to be used regardless of LC-MS system as long as dextran is run before the samples. It also allows for modifications to a LC-MS system to happen, so that if capillary lines were to be changed in length or the LC setup was modified (excluding the actual column), it would not affect the conversion of a peptide's retention time in minutes to glucose units. To ensure that dextran would be a suitable retention time calibrant for the model, a set of peptide standards were run on different LC-MS systems over the course of a month and the retention times of the standards had minimal changes.

Hydrophobic Residues at the N-Terminus of Peptides

Site-specific trends were investigated in the dataset of 297 peptides, specifically at the N-terminus due to the use of trypsin on most of the samples. It was found that 44 out of 70 (63%) peptides with hydrophobic residues at their N-terminus eluted earlier than predicted and optimized coefficients were created for this, as shown in Table 2. Using an iterative process that maximized the correlation coefficient, it was found that a 10% decrease in the value of the original hydrophobic amino acids (phenylalanine, isoleucine, leucine, tryptophan, and tyrosine) resulted in optimized coefficients that had a R-squared value of 0.95552. With these optimized coefficients, the average deviation between actual and predicted retention times dropped from 0.255 GU to 0.246 GU, increasing the accuracy of prediction. Also, 37 out of the 70 (53%)

peptides eluted earlier than predicted, evening the distribution of predicted retention times greater and smaller than actual retention times. These coefficients are only to be used for the first hydrophobic amino acids at the N-terminus of a peptide and no others. Hydrophilic amino acids at the N-terminus were also investigated, but although there was a slight trend (40 of 73, 55%) of peptides with actual retention times larger than their predicted ones, the optimization of the coefficients would be negligible and would not help increase the correlation coefficient.

Peptides with hydrophobic residues one position over from the N-terminus were also investigated from trends, and it was found that 11 of 15 (73%) of peptides that fit this description had actual retention times that were shorter than their predicted ones. Using the same iterative process, it was found that a 5% decrease in the value of the original hydrophobic coefficients resulted in optimized coefficients that had an elevated R-squared value of 0.95563. These optimized coefficients are found in Table 3 and are only for the hydrophobic residue of a peptide that is one position over from the N-terminus, while the first residue at the N-terminus is also a hydrophobic residue. In addition to an increased R-squared value, the average deviation dropped from 0.199 GU to 0.193 GU using the optimized coefficients, with a more even distribution of predicted retention times that were greater and smaller than actual retention times (8 out of 15 (53%) peptides had actual retention times shorter than predicted ones). Hydrophilic residues in this position were also investigated, but it was found again that even though there was a significant trend (10 of 15, 67%), the optimization of the coefficients would again be negligible and would not help the correlation of the model.

Test Peptides

To test the accuracy of prediction, helicobacter pylori samples were run on the same LC-MS setup as the peptides used to create the model. From the test samples, 64 peptides fit the

selection criteria and were investigated. Figure 3 shows the actual times of the test peptides plotted against the predicted times, which yielded a high correlation coefficient of 0.96444, slightly higher than the correlation coefficient of the model itself. This shows that the model is more than capable of predicting retention times for biologically relevant samples that are not just standard proteins. Of the 64 test peptides, 38 of them had lower actual retention times than their predicted ones (59%), which were calculated using Equation 1. The average deviation from actual to predicted times was 0.35 GU, or 1.72 min, indicating the accuracy of prediction.

A 4000 Q-Trap with a Nexera UFLC system was used to test the accuracy of prediction of the model on a completely different LC-MS system. BSA and carbonic anhydrase were run on this system that had a different column size, flow rate, gradient, column temperature, and length of analysis, however peptides that were identified on both LC-MS systems only differed by an average of 2.29 minutes (0.52 GU) and were within 3.73% of each other.

Discussion

Different HILIC columns exhibit different selectivities from one another, making the creation of a new model for the penta-HILIC stationary phase a requirement in order to predict peptide retention.^{8,9} New amino acid coefficients were derived for this model, criteria for peptide selection were created, optimized coefficients for site-specific trends were derived, and other characteristics such as peptide length were analyzed. It was widely known that amino acid composition is the main characteristic that determines peptide retention, but it was shown that amino acid location has a strong effect as well.

The amino acid residues that have positively charged side chains (arginine, lysine, and histidine) have a positive effect on retention and have the largest effect overall, which is consistent with other studies.^{8,10,65,98,101} These side-chains interact with the stationary phase to a

greater extent and increase the retention of the peptides. Aspartic acid and glutamic acid have negatively charged side chains that also increase retention, but they do not have as great of an effect as the positively charged side chains. This is because the pH of the mobile phase (around 3) is lower than that of the pK_a of both residues (3.86 for aspartic acid and 4.07 for glutamic acid), making them neutral and thus interact less strongly with the stationary phase than a charged species. However, these amino acids do have the fourth and fifth largest coefficients besides arginine, lysine, and histidine, indicating that they still are significant. The large, aromatic residues such as phenylalanine, tryptophan, and tyrosine all decreased the retention of peptides due to the hydrophobic nature of the side chains minimally interacting with the highly polar stationary phase. The coefficients for these residues, among others, match up to the inverse of coefficients from reverse phase models. While this was expected, Gilar, et. al. showed that it is not necessarily a linear correlation, and that HILIC and RP can be combined in multidimensional HPLC for more complex separations.⁸

Peptide Retention Prediction Purpose and Correlation with Database Searching

The purpose for peptide retention prediction is threefold. First, it can provide a quicker data analysis as peptides can be identified from their m/z ratio as well as their retention time, eliminating the need for database searching. Second, retention prediction is able to filter out false positives and lead to more confident identifications by comparing actual retention times to theoretical retention times. Finally, it can help in isomeric identification. For example, if two peptides differ only by leucine and isoleucine, they would be indistinguishable in the mass spectrometer. However, the retention coefficients for these two amino acids are different so the peptides would elute at different times, allowing for identification of both species. When MS2 data is insufficient for separately identifying different species, retention time prediction can help.

In database searching, peptides are scored based on the “match” between experimental data and their database sequence. The higher the score of the peptide match, the less likely it is a random match. To test if our model was similar in this aspect, namely that peptides with lower deviations from actual to predicted retention times would have a lower probability of being a random match, 100 peptides from *H. Pylori* proteins were ran on the same LTQ setup as the peptides used to create the model, and their deviations were compared to their Mascot scores. Figure 5 shows this comparison, as peptides were grouped based on their Mascot score and plotted against their average deviations. The resulting data shows an agreement with Mascot score and deviation from predicted to actual retention times, as the peptides with lower deviations have higher Mascot scores and vice versa. It also shows that peptides with lower Mascot scores and higher deviations between actual and predicted times have much larger standard deviations. This indicates that peptides with actual retention times that are very close to theoretical ones are less likely to be false positives.

The Effect of Peptide Length

Although amino acid composition contributes the most to peptide retention, other models have shown that length has an effect as well.^{6,10-12} Mant, et. al. showed that the retention of peptides that have 15 or more residues in their sequence deviated more than expected and cannot be overlooked.^{6,11} Table 4 shows peptides from standard proteins that were not used in this model due to their length, and the average deviations (1.06 GU, or 4.80 min.) are 3-4 times higher than peptides with shorter sequences that were used in the model. A potential reason for this could be due to longer peptides more easily forming second order structures and interacting with the stationary phase in a way that cannot be predicted accurately. This consideration was applied to the creation of this model, as the cutoff for peptide size was 15 amino acids in length.

Applying an elevated column temperature could disrupt a peptide's secondary structure so that it interacts in a more predictable manner with the stationary phase. Long peptides from human IgGs, BSA, transferrin, concanavalin A, lysozyme, and cytochrome C were run at column temperatures of 25°C and 60°C, and the data are shown in Table 5. It is clear from this data that the higher column temperature decreases the deviation from predicted times. However, some peptides run at 25°C were the closer to the predicted times, suggesting that not all of the longer peptides may have had second order structure. Regardless, applying the column temperature decreases the deviation from 1.06 GU to 0.60 GU, increasing the accuracy of prediction for peptides over 15 amino acids in length. Human IgGs were also run at 40°C and 80°C, but both of those temperatures had higher deviations than long peptides ran at 60°C.

It is also evident from the dataset that long peptides with actual retention times closer to predicted retention times at 25°C had a smaller average deviation (0.904 GU) than long peptides with closer retention times at 60°C (1.291 GU). This indicates that applying the elevated column temperature produces a more significant change in structural interaction with the stationary phase, and further supports our reasoning that many of these long peptides have second order structure that unravels at higher temperatures. There were only 3 cases out of 18 where a peptide had a longer retention time at 65°C in comparison to 25°C, and in all cases they were closer to the predicted times. Applying a higher temperature to a column will decrease the retention times to peptides without higher order structure, but in these cases there is significant evidence that their structure and/or interaction with the stationary phase changed due to the increase in retention times.

The Effect of Amino Acid Location

Another contribution to peptide retention is the location of residues in the peptide sequence. It was found that hydrophobic residues directly at the N-terminus and one residue away from the N-terminus elute earlier than expected, and optimized coefficients were derived to account for this difference and make the model even more accurate. Hydrophilic residues at the N-terminus were also investigated, and although there was a slight trend (40 out of 73 were retained longer than predicted), adjusting the coefficients would have a minimal effect on the accuracy of prediction. A potential reason that the hydrophobic residues are having a greater impact than the hydrophilic residues at the N-terminus is due to the fact that the N-terminus is already charged and hydrophilic, allowing hydrophobic residues to change the interactions with the stationary phase to a greater extent than the hydrophilic residues. There have been some models that have incorporated optimized coefficients that are based on the distance from the termini, but excluding the coefficients derived from hydrophobic residues one spot over from the N-terminus, there were no other identified trends that suggested that doing the same would improve the accuracy of the model.^{10,92,98}

Summary

A peptide retention model based on amino acid composition was created using a HALO® penta-HILIC column with gradient elution. This model was shown to be very accurate ($R^2 = 0.946$), on par with previously reported RP and HILIC models. It also includes optimized coefficients for hydrophobic residues at the N-terminus and hydrophobic residues one residue over from the N-terminus. The use of dextran as a retention time calibrant was essential for making this model capable of being used on any LC-MS system.

We are currently deriving coefficients for peptides with post-translational modifications that can be separated from unmodified peptides using the HILIC column. Many of these modifications cannot be separated by RP chromatography and they include oxidation, deamidation, and O-GlcNAcylation, among others.¹²⁴ We hope to develop a glycopeptide retention prediction model by combining this model with a glycan retention prediction model that is currently being developed in our laboratory.

Acknowledgements

We would like to thank Rudradatt Persaud for his help with the early visualization of the project. We would also like to thank Dr. Mary Elizabeth Thompson and Dr. T. Colin Campbell for their help with the data sorting and regression analysis. Finally, we would like to thank Clayton Seigel for his help with the data processing.

Table 3.1

Derived coefficients for each amino acid. Red amino acids are hydrophilic, blue amino acids are hydrophobic, and the contribution to retention for green amino acids did not achieve statistical significance.

Amino Acid	Coefficient
Alanine (A)	0.164
Cysteine (C)*	0.293
Aspartic Acid (D)	0.800
Glutamic Acid (E)	0.719
Phenylalanine (F)	-0.967
Glycine (G)	0.233
Histidine (H)	1.564
Isoleucine (I)	-0.615
Lysine (K)	2.121
Leucine (L)	-0.799
Methionine (M)	-0.337
Asparagine (N)	0.610
Proline (P)	0.129
Glutamine (Q)	0.703
Arginine (R)	1.828
Serine (S)	0.334
Threonine (T)	0.357
Valine (V)	-0.306
Tryptophan (W)	-1.138
Tyrosine (Y)	-0.430
Intercept	1.535
R-Squared Value	0.94553

Table 3.2

Optimized coefficients for the first hydrophobic amino acid at the N-terminus

Amino Acid	Coefficient
Phenylalanine (F)	-1.063
Isoleucine (I)	-0.676
Leucine (L)	-0.879
Tryptophan (W)	-1.252
Tyrosine (Y)	-0.473
R-Squared Value	0.94620

Table 3.3

Optimized coefficients for the second hydrophobic amino acid at the N-terminus

Amino Acid	Coefficient
Phenylalanine (F)	-1.015
Isoleucine (I)	-0.646
Leucine (L)	-0.839
Tryptophan (W)	-1.195
Tyrosine (Y)	-0.451
R-Squared Value	0.94600

Table 3.4**Retention of peptides with 15 or more amino acids**

Peptide	Length	Deviation (min)	Deviation (GU)
RPCFSALTPDETYVPK	16	6.01	1.88
LFTFHADICTLPDTEK	16	8.90	2.35
NTDGSTDYGILQINSR	16	0.58	0.16
EDLIWELLNQAQEHFGK	17	0.37	0.09
GITWGEETLMEYLENPK	17	5.96	0.98
VYACEVTHQGLSSPVTK	17	24.24	4.33
TTPPVLDSDGSFFLYSK	17	6.06	0.95
GITWGEETLMEYLENPKK	18	7.54	1.79
TVAAPSVFIFPPSDEQLK	18	3.33	0.56
RTVAAPSVFIFPPSDEQLK	19	1.20	0.29
AAPSVTLFPPSSEELQANK	19	0.16	0.04
ANPTVTLFPPSSEELQANK	19	0.91	0.24
EVQLVQSGGGLVQPGGSLR	19	5.45	1.28
DLILQGDATTGTDGNLELTR	20	3.84	1.10
VDNALQSGNSQESVTEQDSK	20	0.36	0.16
GLVLIAFSQYLQQCPFDEHVK	21	7.96	1.64
GFYPSDIAVEWESNGQPENNYK	22	0.89	0.27
SPDShPADGIAFFISNIDSSIPSGSTGR	28	2.61	0.89

Table 3.5**Retention of long peptides with and without a column oven**

Peptide	Length	Predicted RT	RT (Without	RT (With 60°
		(GU)	Oven)	Oven)
RPCFSALTPDETYVPK	16	6.39	8.28	6.63
LFTFHADICTLPDTEK	16	5.05	7.40	4.56
NTDGSTDYGILQINSR	16	6.27	6.43	6.35
EDLIWELLNQAQEHFGK	17	5.47	5.39	5.05
GITWGEETLMEYLENPK	17	4.33	3.35	4.04
VYACEVTHQGLSSPVTK	17	6.69	2.37	4.07
TTPPVLDSDGSFFLYSK	17	3.19	4.14	3.18
GITWGEETLMEYLENPKK	18	6.45	4.66	6.52
TVAAPSVFIFPPSDEQLK	18	3.65	4.22	3.92
RTVAAPSVFIFPPSDEQLK	19	5.49	5.78	4.92
AAPSVTLFPPSSEELQANK	19	5.77	5.81	5.18
ANPTVTLFPPSSEELQANK	19	6.24	6.00	5.87
EVQLVQSGGGLVQPGGSLR	19	4.84	6.11	4.60
DLILQGDATTGTDGNLELTR	20	6.27	7.38	5.73
VDNALQSGNSQESVTEQDSK	20	10.70	10.54	10.04
GLVLIAFSQYLQQCPFDEHVK	21	4.01	5.65	4.69
GFYPSDIAVEWESNGQPENNYK	22	6.82	7.08	6.13
SPDSHPADGIAFFISNIDSSIPSGSTGR	28	7.65	8.54	5.62

Equation 3.1

Calculation of Predicted Retention Times

$$R_T = \sum (L_i AA_i) + b_0$$

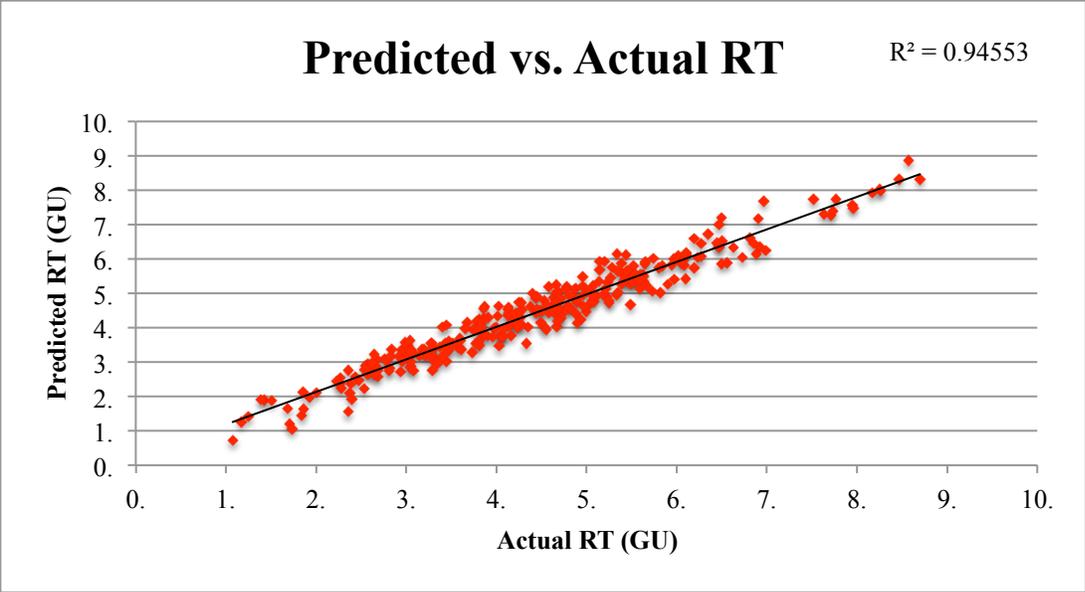


Figure 3.1: Predicted vs. actual times of the 297 peptides used in the study

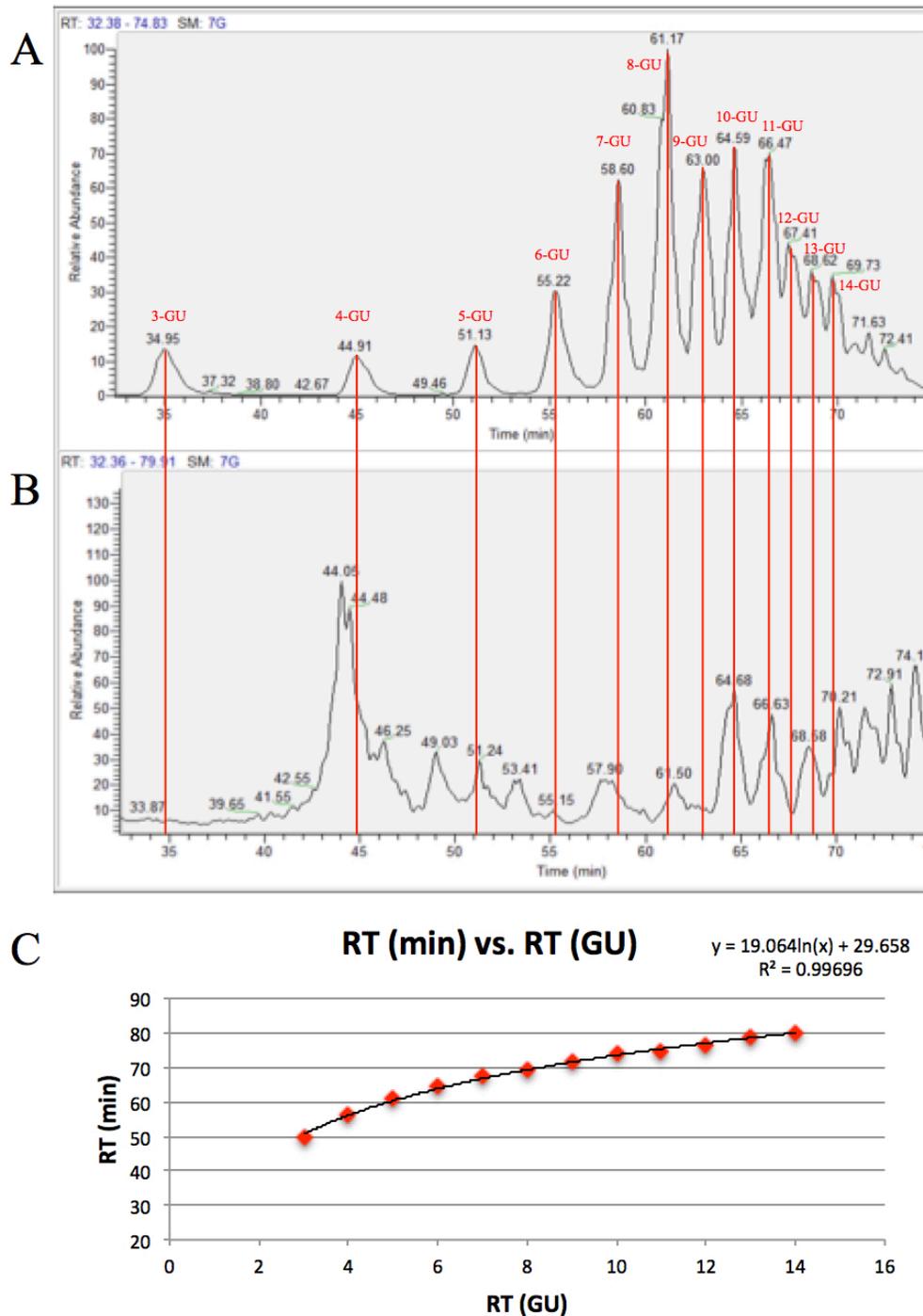


Figure 3.2: Procainamide-labeled dextran samples served as a retention time calibrant to the peptides used in the model. Monosaccharides elute in terms of increasing linkage (A) and then peptide retention times (B) were converted from minutes to glucose units (GU) using the logarithmic fit of the dextran units (C).

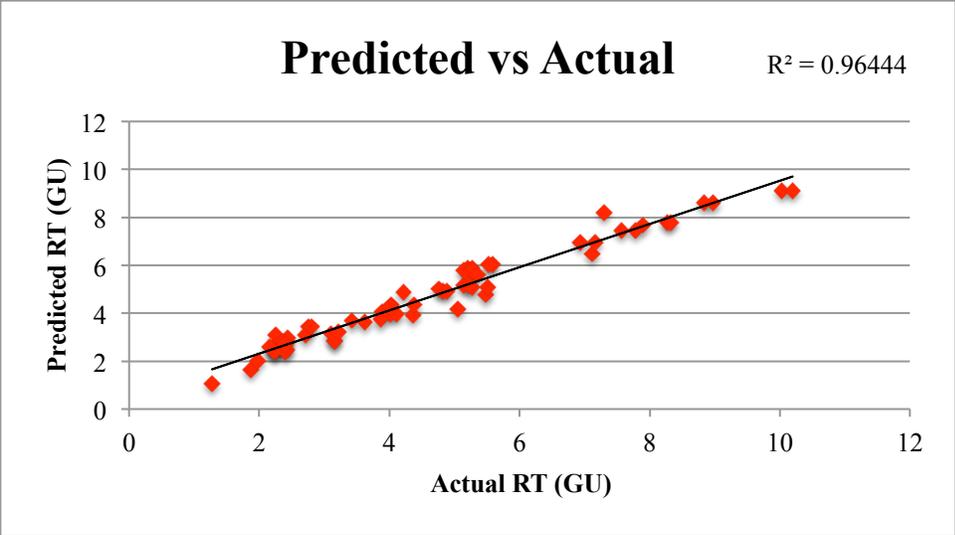


Figure 3.3: Predicted vs. actual times of helicobacter pylori test peptides

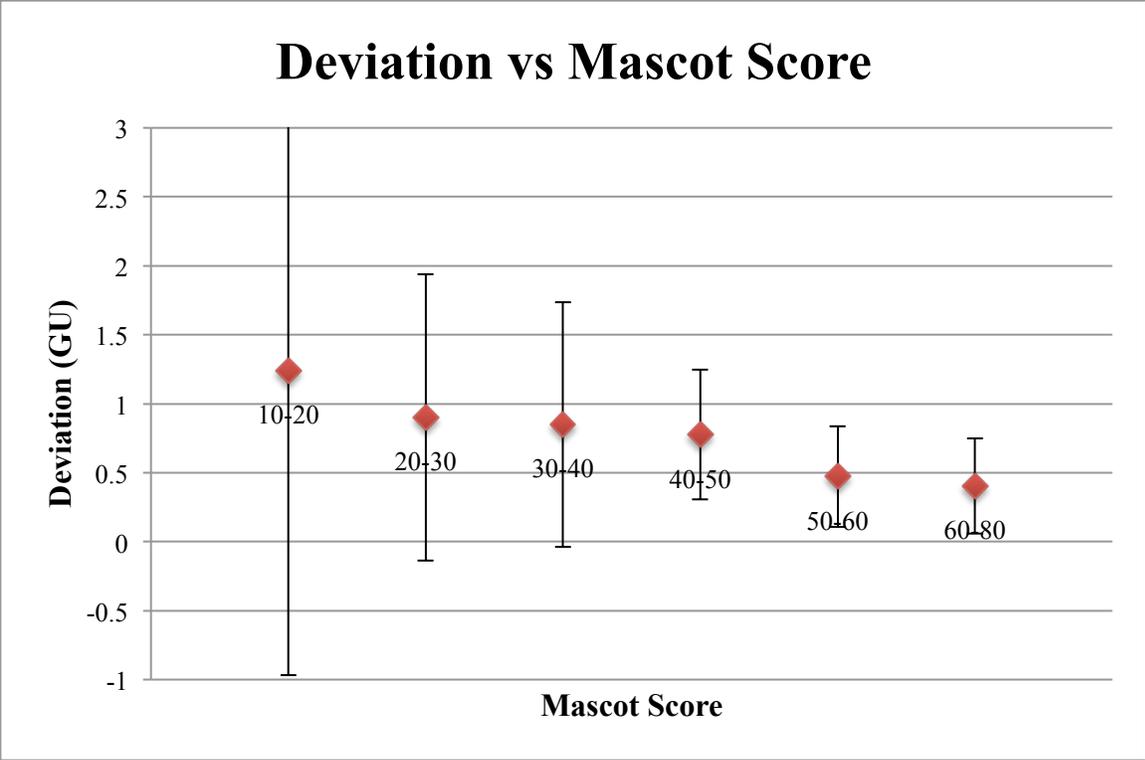


Figure 3.4: Deviation of actual retention times and theoretical retention times plotted against Mascot score

CHAPTER 4

THE SEPARATION AND QUANTITATION OF PEPTIDES WITH AND WITHOUT OXIDATION OF METHIONINE AND DEAMIDATION OF ASPARAGINE USING HYDROPHILIC INTERACTION LIQUID CHROMATOGRAPHY WITH MASS SPECTROMETRY (HILIC-MS)²

² Badgett, M.J., Boyes, B., Orlando, R. The Separation and Quantitation of Peptides With and Without Oxidation of Methionine and Deamidation of Asparagine Using Hydrophilic Interaction Liquid Chromatography With Mass Spectrometry (HILIC-MS). *Journal of the American Society for Mass Spectrometry*. 2017. Reprinted here with permission of publisher.

Abstract

Peptides with deamidated asparagine residues and oxidized methionine residues are often not resolved sufficiently to allow quantitation of their native and modified forms using reversed phase (RP) chromatography. The accurate quantitation of these modifications is vital in protein biotherapeutic analysis because they can affect a protein's function, activity, and stability. We demonstrate here that hydrophilic interaction liquid chromatography (HILIC) adequately and predictably separates peptides with these modifications from their native counterparts. Furthermore, coefficients describing the extent of the hydrophilicity of these modifications have been derived and were incorporated into a previously made peptide retention prediction model that is capable of predicting the retention times of peptides with and without these modifications.

Introduction

Many of the chemical modifications that accumulate in biotherapeutic agents during bioprocessing, purification, storage, or other stages increase the hydrophilicity of the amino acid side chain on the altered residue(s). These modifications can include the oxidation of methionine and the deamidation of asparagine, among others. The separation and quantitation of peptides that have these modifications is of paramount importance in protein biotherapeutics because the modifications can contribute to a loss of stability or activity.^{13,14,110}

Although little is known about the effects that deamidation of asparagine have on protein function, it is known that deamidation is involved in protein degradation and development.^{15,117,174,175} This reaction is spontaneous and non-enzymatic, where asparagine residues undergo formation of a five-membered succinimide ring intermediate from an intramolecular attack, and then subsequently hydrolyze under physiological conditions to form either aspartyl or isoaspartyl peptides which can be found in both the D and L configurations (Figure 1). Deamidation occurs at a much faster rate (up to 70 times) when an unhindered amino acid residue such as glycine is on the C-terminal side of an asparagine in the primary sequence (XXX-Asn-Gly-XXX), but its rate is also affected by other conditions and characteristics such as temperature, pH, and protein structure.^{13,15,16,19,117,118,176,177} As deamidation changes the peptide/protein structure and conformation, it can significantly affect the function and stability of proteins. For example, deamidation of an Asn-Gly site in hemoglobin alters its affinity for oxygen, while the same modification alters the proteolytic cleavage of human growth hormone (hGH).^{113,116} Deamidation has been studied using different analytical techniques, such as isoelectric focusing, capillary electrophoresis, and a variety of LC-MS/MS techniques, but they all have limitations that make analyzing deamidation a challenge.^{13,175} The biggest challenge for

mass spectrometric analysis of deamidated proteins is that there is only a one Dalton mass shift between the modified and native forms, which causes the deamidated species to overlap with the mass-to-charge (m/z) ratios of the ^{13}C isotopes of the unmodified species.¹⁷⁸ Without employing a separation technique that can fully distinguish the modified and unmodified versions, mass spectrometric analysis of deamidation can be highly challenging.

More is known about the oxidation of methionine compared to the deamidation of asparagine, presumably because the larger mass difference between modified and unmodified peptides makes it easier to study this post-translational modification (PTM) by mass spectrometry. The oxidation of methionine has been shown to affect the structure, stability and biological functions of a variety of proteins and is a major instability factor of protein pharmaceuticals including monoclonal antibodies.^{130,134,179} It is also associated with the development of several diseases, including Alzheimer's disease, emphysema, and respiratory distress syndrome, among others.^{130,132,133,135,180} Methionine (Met-S) can oxidize to form methionine sulfoxide (Met-SO) via a formal oxygen transfer, which can further oxidize to form methionine sulfone (Met-SO₂) as shown in Figure 2. Met-SO can be reduced back to methionine using methionine sulfoxide reductase A (MsrA), which is found in most cells. MsrA has been shown to be important in Alzheimer's disease, as the levels of MsrA in the brain of Alzheimer's disease patients is significantly lower than in the brain of normal individuals, and this is reflected by increased levels of Met-SO in these regions.^{129,133,137,140,179} Oxidation of methionine is similarly significant in "normal" aging, as a decline in MsrA activity lead to a 40% decrease in the maximum life span of mice, and overexpression of MsrA in *Drosophila* greatly extends their life span.^{129,138,139} Just like deamidation of asparagine, oxidation of methionine affects protein structure, which in turn can lead to negative effects such as reduced protein activity or stability.

Altered activity due to oxidation of methionine has been discovered in a plethora of different proteins including, but not limited to, chymotrypsin, ribonuclease B, lysozyme, and pepsin.

179,181–184

Using conventional reversed-phase (RP) approaches, oxidized peptides can sometimes be separated from their native forms, but deamidated peptides are often not resolved from their unmodified counterparts. Hao, et. al. used a multidimensional RP-ERLIC-MS/MS approach to collect a triad of deamidated products together and then subsequently separated them based on their pI to allow for identification.¹¹⁵ High resolution hydrophilic interaction liquid chromatography (HILIC) can also be a solution to this problem, as the change in hydrophilicity of amino acid side chains resulting from these modifications may change the selectivity of the peptides that have these modifications sufficiently to allow for chromatographic separation, which could enable their quantitation.

Here, we demonstrate the capacity of HILIC-MS to separate and quantitate modified peptides and their native counterparts for the analysis of human immunoglobulin Gs (IgGs), and other standard proteins. Previously, we have created a peptide retention prediction model using HILIC that is based on the summation of amino acid coefficients.¹⁸⁵ Herein the utility of this model is expanded by derivation of coefficients for the oxidation of methionine and for the deamidation of asparagine, which are now incorporated into the previous retention model. Modified and unmodified peptides can quickly and easily be identified by their predicted relative retention times in conjunction with their m/z ratio. This will provide an easier and consistent assessment of the extent of modifications in biotherapeutic agents, as well as allow for the separation, characterization, and potential isolation of peptides with these modifications.

Materials and Methods

Protein Digestion

Human IgGs were separated from human serum (Sigma-Aldrich, St. Louis, MO, USA) using a HiTrap™ Protein G column (General Electric Company, Fairfield, CT, USA). Cytochrome C, lysozyme, transferrin, and dextran were purchased from Sigma-Aldrich. Bovine serum albumin was purchased from Waters (Milford, MA, USA). These proteins as well as yeast and mosquito cuticular proteins were reduced using 10 mM dithiothreitol (DTT) and then alkylated using 55 mM iodoacetamide (IDA), both purchased from Sigma Aldrich. Sequencing-grade trypsin or chymotrypsin purchased from Promega (San Luis Obispo, CA, USA) was added at 50:1 (w/w, protein/trypsin) for incubation overnight in 50 mM ammonium bicarbonate (pH 7.0) at 37°C. Three synthetic peptides with the same sequence of GFYPSDIAVEWESNGQPENNYK were purchased from Bachem (Bubendorf, Switzerland). One peptide was unmodified, one had a *n*-Asp modification at the 14th residue, and the last one had an isoAsp modification at the 14th residue.

LC-MS/MS Settings and Instrumentation

Data were acquired using a Finnegan LTQ (Thermo-Fisher, San Jose, CA, USA) in series with an 1100 Series Capillary LC system (Agilent Technologies, Palo Alto, CA, USA) with an ESI source that used spray tips made in-house. Samples were suspended in 25% H₂O, 75% ACN and 0.1% formic acid (Sigma-Aldrich, St. Louis, MO, USA) for direct injection into the LC system. Peptides were separated using a 200- μ m x 150-mm HALO® Penta-HILIC column packed with 2.7 μ m diameter superficially porous particles that have a 90 Å pore diameter (Advanced Materials Technology, Wilmington, DE, USA) at room temperature. The gradient elution conditions employed a linear increase in aqueous solvent from 5-70% over 90 minutes at

a 2 μ L/min flow rate, using the column at room temperature. The (strong) aqueous solvent contained 0.1% formic acid (Sigma Aldrich, St. Louis, MO, USA) with 50 mM ammonium formate (Thermo-Fisher, San Jose, CA, USA) and the organic solvent was acetonitrile with 0.1% formic acid. The settings for the mass spectrometer included taking the 5 most intense ions from each full mass spectrum for fragmentation using collision-induced dissociation (CID), and the resulting MS/MS spectra were recorded.

To make sure that this model would be universal, some of the same digested proteins as well as the synthetic peptides were run on a 4000 Q Trap (AB Science, Chatham, NJ, USA). Peptides were separated by a 2.1 mm x 15 cm HALO® Penta-HILIC column packed with 2.7- μ diameter superficially porous particles using a Nexera UFLC (Shimadzu, Columbia, MD, USA). The temperature of the column was 60°C. The gradient used for each sample was 22-52% water over 80 minutes at a 0.4-mL/min flow rate. Spectra were obtained using an ESI source.

For RP analysis using the Finnegan LTQ and 1100 Series Capillary LC system, samples were suspended in 95% H₂O, 5% ACN and 0.1% FA (Sigma-Aldrich, St. Louis, MO, USA) and injected to the LC. Peptides were separated using a 200- μ m x 150-mm HALO® Peptide ES-C18 column packed with 5- μ m diameter superficially porous particles (Advanced Materials Technology, Wilmington, DE, USA). The column was at room temperature. The gradient was 5-75% ACN for 120 minutes at a 2 μ L/min flow rate. The mobile phase contained 0.1% formic acid and 10 mM ammonium formate (Thermo-Fisher, San Jose, CA, USA). The LC-MS/MS system and MS parameters were the same as the HILIC analysis.

Database Search Parameters

The resulting RAW files were converted using Trans-Proteomic Pipeline (Seattle Proteome Center, Seattle, WA, USA), then the MS/MS spectra of each sample were searched

using Mascot (Matrix Scientific, Boston, MA, USA) against corresponding protein databases of theoretical MS/MS spectra. The following parameters were utilized in Mascot: a peptide tolerance of 1000 ppm, a fragment tolerance of 0.6 Da, two max missed cleavages of trypsin, and a fixed modification of carbamidomethyl (C).

Selection of Peptides for Prediction Model and Post-Run Data Analysis

All peptides that had a higher Mascot score than 10 were considered. Peptide retention times were determined manually from .RAW files using the apex of the peaks displayed in Xcalibur software (Thermo-Fisher, San Jose, CA, USA), and resulting MS/MS data were visually inspected for fragmentation that was consistent with peptide assignments. Chromatographic peaks for each peptide had to have a peak asymmetry value of between 0.25 - 4, and peptides exhibiting peak widths greater than 5.5 minutes were excluded from analysis. Peptide retention times in minutes were converted to glucose units based on dextran samples that were run immediately before. Linear regression analysis using StatPlus (AnalystSoft, Walnut, CA, USA) was used to find the incremental retention coefficients for each amino acid.

Results and Discussion

Resolving Modified and Unmodified Peaks

To evaluate the ability of RP chromatography to resolve oxidized and native peptides, samples were first run on a C18 RP column. Figure 3 panel B shows the separation for the BSA peptide TVMENFVAFVDK, where the bold amino acid residue is the expected site of modification. In this figure, the oxidized and unoxidized versions are separated, which allows for the easy quantitation of these two species. However, this is not always the case with oxidation using RP chromatography, as shown by the chromatography of the BSA peptide ETYGDMA DCCEK (Figure 3A), where the oxidized version and native version are not

separated. These cases show that RP chromatography, using typical acidic mobile phase conditions, does not separate oxidized and unoxidized peptides with certainty, which can decrease detection relative sensitivity for modified versus unmodified peptides, leading to uncertainty about the significance of oxygen-driven degradation processes. The separation is also influenced by the composition of the neighboring residues to the methionine, which are different in the two peptides in terms of hydrophobicity.

Deamidation is a difficult modification to analyze because the 1Da mass increase of the modification places the molecular ion for the deamidated peptide at the same nominal mass as the one ^{13}C isotope of the unmodified species. Utilizing chromatography to resolve these two species decreases the complexity of identifying and quantitating the modified species. However, for the IgG peptide GFYPSDIAVEWESNGQPENNYK, the deamidated and non-deamidated peptides coelute from a RP column, and this is shown in Figure 4. There were no cases using RP chromatography in any of the peptide samples in which the deamidated peptide and the native peptide had baseline separation. The majority of the peptides coeluted and some had peak shoulders, but there was never enough separation to quantitate the peaks. The selectivity differences for the asparagine and iso-aspartic acid side chain functional groups are very small under typical low pH separation conditions, leading to minimal resolution capabilities. This separation problem is in addition to the similar masses that the modified and unmodified peptides possess, leading to significant errors in detection of the modified versus unmodified peptides. The mass difference between the unmodified ^{13}C peptide and modified peptide is 0.0152 Daltons, which for the deamidated IgG peptide GFYPSDIAVEWESNGQPENNYK would require a mass spectrometer with a resolution of 167,377 to be able to resolve from its unmodified ^{13}C version. In essence, using chromatographic techniques that are not able to

separate deamidated peptides from their native forms requires very high-resolution mass spectrometers, able to detect the minute mass difference. Even then, it still is difficult to identify the presence of a modified peptide with high certainty, and extremely difficult to determine relative abundance of the peptide pairs.

From the HILIC peptide retention model that we previously created there is a substantial difference between asparagine and aspartic acid coefficients, which indicates that deamidated peptides should be separated from their nondeamidated forms. There is also a significant difference between alanine and serine that implies adding an oxygen to methionine should be enough to effectively separate peptides with oxidized methionine residues from their native counterparts.¹⁸⁵ To test this, the same samples were analyzed using HILIC separation conditions. Figure 5 shows baseline separation for the IgG oxidized peptide KDSGFQMNQLR (panel A) and the IgG deamidated peptide GFYPSDIAVEWESNGQPENNYK (panel B). Both modified peptides and their native forms exhibited baseline separation on the HILIC column, allowing for confident quantitation of the peak areas or heights even with low-resolution mass spectrometers. These modifications increase the hydrophilicity of their respective peptides, which in turn increases the retention time. Using HILIC to analyze deamidation negates the requirement to employ a high-resolution mass spectrometer, as necessitated by the overlap of the unmodified peptide ¹³C isotope envelope, and the modified peptide mass. The separation of the modified peptides from the native structure is predictable, and occurs regardless of peptide sequence, as a variety of different types of residues were adjacent to the site of deamidation in the peptides that were used in the study. Separation selectivity factors (α) and resolution values for unmodified and modified peptide pairs that were separated and identified together are shown in Table 1.

From the deamidation mechanism, it is clear that two potential, *n*-Asp and isoAsp, modified products can be present. To deduce the form of the deamidated products, synthetic peptides with the sequence GFYPSDIAVEWESNGQPENNYK were run on the 4000 Q Trap LC-MS system. These synthetic peptides had three versions: unmodified, *n*-Asp at the 14th residue, and isoAsp at the 14th residue, and Figure 6 shows a run with all of the versions separated. The least retained peak is the unmodified form of the peptide, while the peak in the middle is the aspartyl version, and the peak most retained is the isoaspartyl version. Comparing these results to the same deamidated peptide in Figure 7 shows that the deamidated peak at 37 minutes is the isoaspartyl version of the peptide, and peak at 35 minutes corresponds to the aspartyl version of the peptide. There is another set of three peaks around this peptide that are not labeled, most likely indicating a second deamidation site at the first asparagine in the “NN” motif in the peptide sequence. However, the separation of this peptide is consistent with the first two peptides shown in Figure 7, which only differ in the residue in the 8th position (the earlier eluting peptide has a leucine where the later eluting peptide has a valine). These peptides share the same elution order as the synthetic peptide. For all the samples run on the LTQ, it was ambiguous as to which deamidation product was present before the synthetic peptides were run because only one deamidated peak would appear. It is known that isoAsp is 2-3 times more abundant than the *n*-Asp, so the deamidation that has been seen for the peptides of the current study actually corresponds to isoAsp, and the ratio of peak abundances of *n*-Asp to isoAsp in Figure 7 is analogous to this statement.^{15,16,115,117,118} Due to the different physical properties of these two deamidation products such as pK_a (*n*-Asp: 3.9, isoAsp: 3.2), the retention times will be different in the mildly acidic conditions of separation used during the current study. A higher percentage

of isoAsp will be charged at this pH, increasing the retention in comparison to *n*-Asp. There was no indication that the D and L configurations could be resolved.

Finally, the extent of the hydrophilic retention shifts for both modifications is consistent with the HILIC column, but was shown to be inconsistent with the RP system employed. This is due to the hydrophilic modifications having a greater selectivity difference with the HILIC stationary phase than the RP stationary phase, allowing for the prediction of the retention to be heightened for HILIC.¹⁸⁶ It is also due to secondary effects in RP, such as amino acid location or neighboring residue composition, as the two peptides TVMENFVAFVDK and ETYGDMAADCCEK exhibited different retention behavior. The first peptide has a hydrophobic neighboring residue and exhibits baseline separation between modified and unmodified forms, while the second peptide has hydrophilic residues on either side and the oxidized version is not fully separated from the native peptide. RP resolution is driven by hydrophobicity differences in analytes, so the change in polarity of the first peptide due to oxidation could affect the surface interactions of the hydrophobic residue next to the methionine, leading to a better separation than obtained for the second peptide. Whatever underlies a mechanistic interpretation of the selectivity differences in separation, it is clear that reversed phase chromatography is much more sensitive to sequence effects than HILIC, for which was observed a consistent retention shift regardless of neighboring residues.

Peptide Prediction Model Coefficients

We have previously created a model that predicts peptide retention based on amino acid composition.¹⁸⁵ In this model, coefficients for each amino acid were derived using linear regression analysis of 50 unmodified peptides, and the retention time of a peptide can be predicted by using Equation 1, where R_T is the predicted retention time, L_i is the amount of

residue i in the peptide, AA_i is the amino acid coefficient of residue i , and b_0 is the intercept of the model. We have recently expanded this model using data from 297 unmodified peptides, and it has a very high correlation coefficient (0.94553), indicating accurate prediction.

The amino acid coefficients are expressed in glucose units (GU) from procainamide-labeled dextran samples that were run before each sample. This approach allows the model to be used on any LC-MS system as long as a dextran standard ladder is run before the protein sample of interest, and the retention times of peptides are then converted from minutes to GU based on the logarithmic fit for the dextran samples. Dextran elutes in order of increasing monosaccharide linkage, and provides a useful reference for the retention times of peptides. Excluding the actual stationary phase and mobile phase composition, this approach also allows for modifications to a LC-MS system to occur, such as the changing of the length of a capillary line or detector configuration, which would not affect the conversion of a peptide's retention time to GU. To ensure that dextran would be a suitable retention time calibrant, peptide standards were run on two different LC-MS systems over the course of a month, and data analysis indicated that the retention times of the standards had minimal changes. These two systems had differing column lengths, column temperatures, gradients, and flow rates, yet the retention times of peptides that were run on both systems were within 3.73% of each other and only differed by an average of 0.52 GU (2.29 minutes).

Two new coefficients were created for the isoAsp form of the deamidated asparagine residues and oxidized methionine residues to be able to predict the retention of peptides with these modifications. Twelve deamidated peptides and 27 Met-oxidized peptides were discovered and incorporated into the model. These modified peptides were from some of the samples used to create the unmodified peptide retention model (IgGs, mosquito cuticular proteins, yeast proteins,

BSA, cytochrome C, transferrin, and lysozyme), and regression analysis was used to derive these coefficients. The deamidation coefficient corresponding to the isoaspartyl form that was derived had a value of 1.409 (R-squared = 0.94186), indicating that the modification is very hydrophilic and will increase the retention time of peptides with this modification. This coefficient was on the higher end of all coefficients in the unmodified peptide retention model, only less than the three most hydrophilic residues: lysine, arginine, and histidine. The large deamidation coefficient (1.409) that was derived supports the claim that the deamidated peaks are indeed isoAsp. This coefficient is much larger than the difference between the asparagine coefficient and the aspartic acid coefficient, which would correspond to the formation of the *n*-Asp product.

The oxidized Met coefficient was also found to be hydrophilic, with a value of 0.633. This is a large difference from the unoxidized methionine coefficient (-0.337) and it was shown that this difference is sufficient for ready separation of the unmodified peptides from the modified ones. As with the value of the deamidation coefficient, the oxidized methionine coefficient is greater than expected, based on the modest difference in coefficients between alanine and serine. This comparison was made because both cases differ by the addition of an oxygen atom, so the expected difference between the alanine/serine and oxidized/unoxidized deviations should be minimal. The oxidized samples slightly increased the R-squared of the model to 0.94637, whereas the deamidated samples did the opposite, slightly decreasing the R-squared to 0.94186. However, the incorporation of both of the coefficients into the previous model barely affected the overall R-squared value, and this is because the total amount of modified peptides (39) was significantly less than the amount of unmodified peptides used to create the original model (297). The coefficients that were derived for the hydrophilic modifications do not affect the values for the unmodified amino acid coefficients because we are

more concerned with the separation of the modified and unmodified peptides rather than their actual retention times. In time there will be more instances of these modifications and we can gather a better understanding of the impact the coefficients have to the overall fit of the model. For now, we have found that both of these modifications are hydrophilic, with deamidation being one of the most hydrophilic coefficients in the model, and the coefficients explain why we are able to see sufficient separation between the modified and unmodified peptides using the HILIC mode of separation.

Summary

Deamidated asparagine residues and oxidized methionine residues were shown to be resolvable from their native forms using HILIC chromatography, which allows for individual peak quantitation. This is particularly useful for deamidation, where the mass differences between a peptide containing an unmodified ^{13}C isotope and a deamidated asparagine residue are too small to resolve from one another without using a high-resolution mass spectrometer. In the current examples, analyses were conducted using an LTQ instrument, with only limited mass resolution capabilities. By being able to fully separate peptides with and without these modifications, the identification process can be heightened and peptides with modifications can be more easily quantitated, which is vital in protein biotherapeutics where the quantitation of analytes with modifications needs to be known. Additionally, coefficients describing each modification's hydrophilicity were derived and incorporated into a peptide retention prediction model that was previously presented. Both coefficients were shown to be very hydrophilic and did not affect the already high R-squared value of the original model by a significant amount.

Acknowledgements

Support for this work comes from NIH Grant GM0 93747 to BEB. We would like to thank Clayton Seigel for his help with the data processing.

Table 4.1**Selectivity factor (α) and resolution values for modified and unmodified peptide pairs.****Modified residues are underlined in red**

Sequence	Modification	Selectivity (α)	Resolution
GFYPSDIAVEWES <u>NG</u> QPENNYK	Asn/(Iso)Asp14	1.099	1.928
<u>GN</u> PTVEVELTTEKGVFR	Asn/(Iso)Asp2	1.087	2.006
LLGVAGGQAFEGAPT <u>N</u> VEIAR	Asn/(Iso)Asp16	1.134	2.088
<u>NP</u> VILADACCSR	Asn/(Iso)Asp1	1.098	1.431
TVDYTADDV <u>NG</u> FNAVVS	Asn/(Iso)Asp10	1.062	1.424
VVEEYTADPV <u>NG</u> FNAVVHR	Asn/(Iso)Asp11	1.073	1.348
VVSVLTVLHQDWL <u>NG</u> K	Asn/(Iso)Asp14	1.119	1.579
VVSVLTVVHQDWL <u>NG</u> K	Asn/(Iso)Asp14	1.123	1.874
IET <u>M</u> R	Met/MetSO4	1.097	1.364
KDSGFQ <u>M</u> NQLR	Met/MetSO7	1.058	1.288
<u>M</u> PCTEDYLSLILNR	Met/MetSO1	1.198	1.990
TV <u>M</u> ENFVAFVVK	Met/MetSO3	1.139	1.932

Equation 4.1

Calculation of Predicted Retention Times

$$R_T = \sum (L_i AA_i) + b_0$$

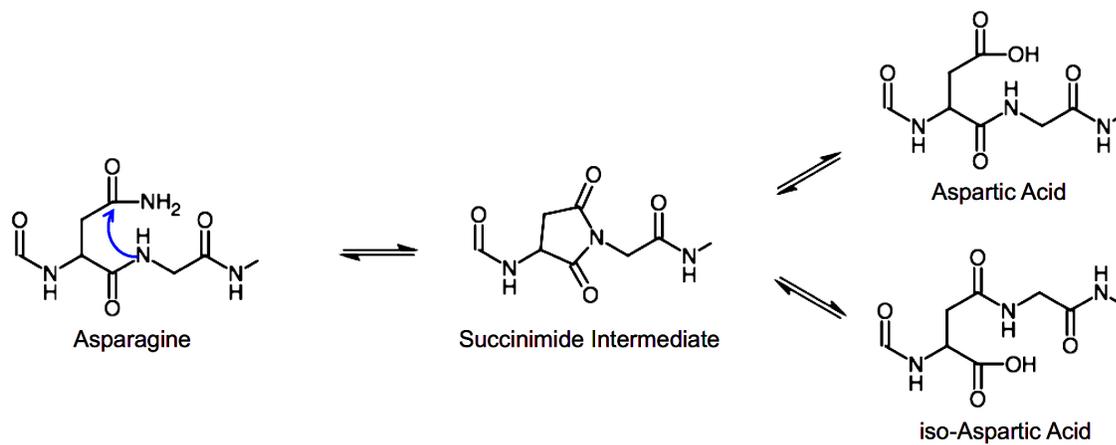


Fig. 4.1: The deamidation of asparagine mechanism. Asparagine forms a five-membered succinimide ring intermediate from an intramolecular attack, and then hydrolyzes to form either aspartyl and isoaspartyl peptides (Created using ChemDoodle® by iChemLabs).

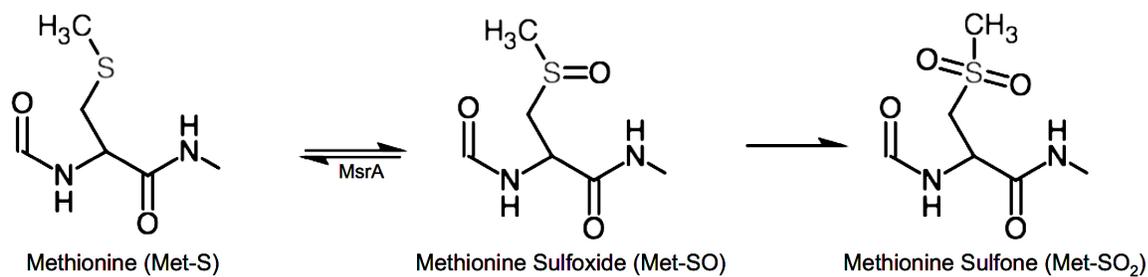


Fig. 4.2: The oxidation of methionine mechanism. Methionine (Met-S) oxidizes to form methionine sulfoxide (Met-SO), which can further oxidize to form methionine sulfone (Met-SO₂). Met-SO can be reduced back to methionine using methionine sulfoxide reductase A (MsrA) (Created using ChemDoodle® by iChemLabs).

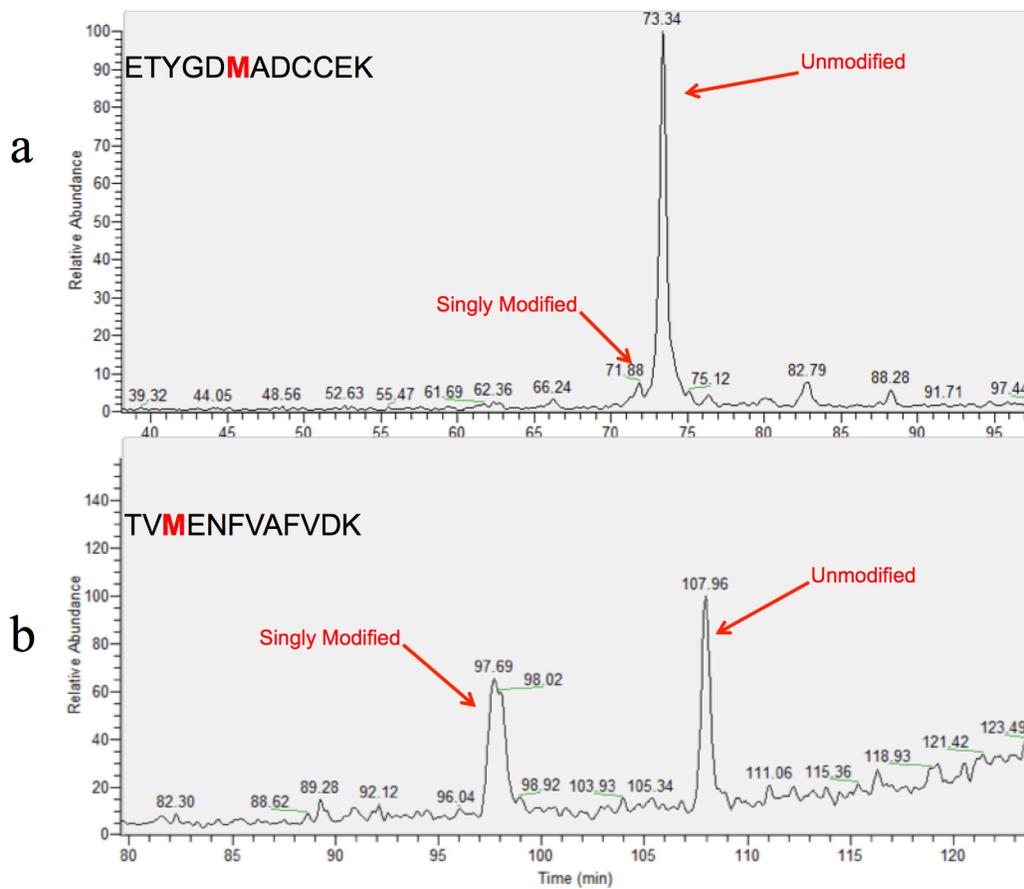


Fig. 4.3: The separation of oxidized peptides and their native forms using a C18 column. Oxidized peptides are not consistently separating from their unmodified counterparts in a predictable fashion

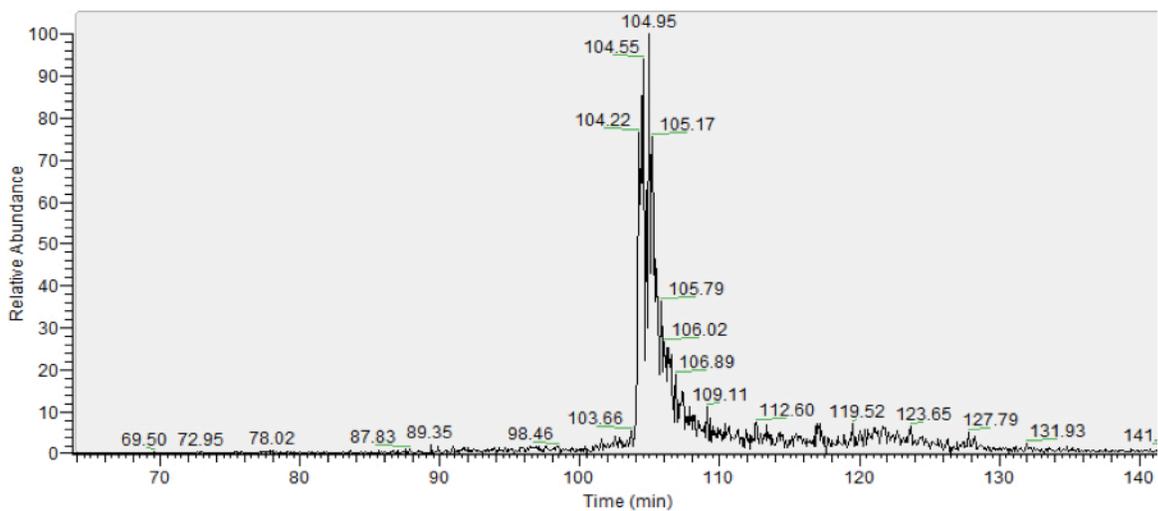


Fig. 4.4: The chromatography of the IgG deamidated peptide

GFYPSDIAVEWESNGQPENNYK and its native forms using a C18 column. Both modified and native forms coeluted around 105 minutes.

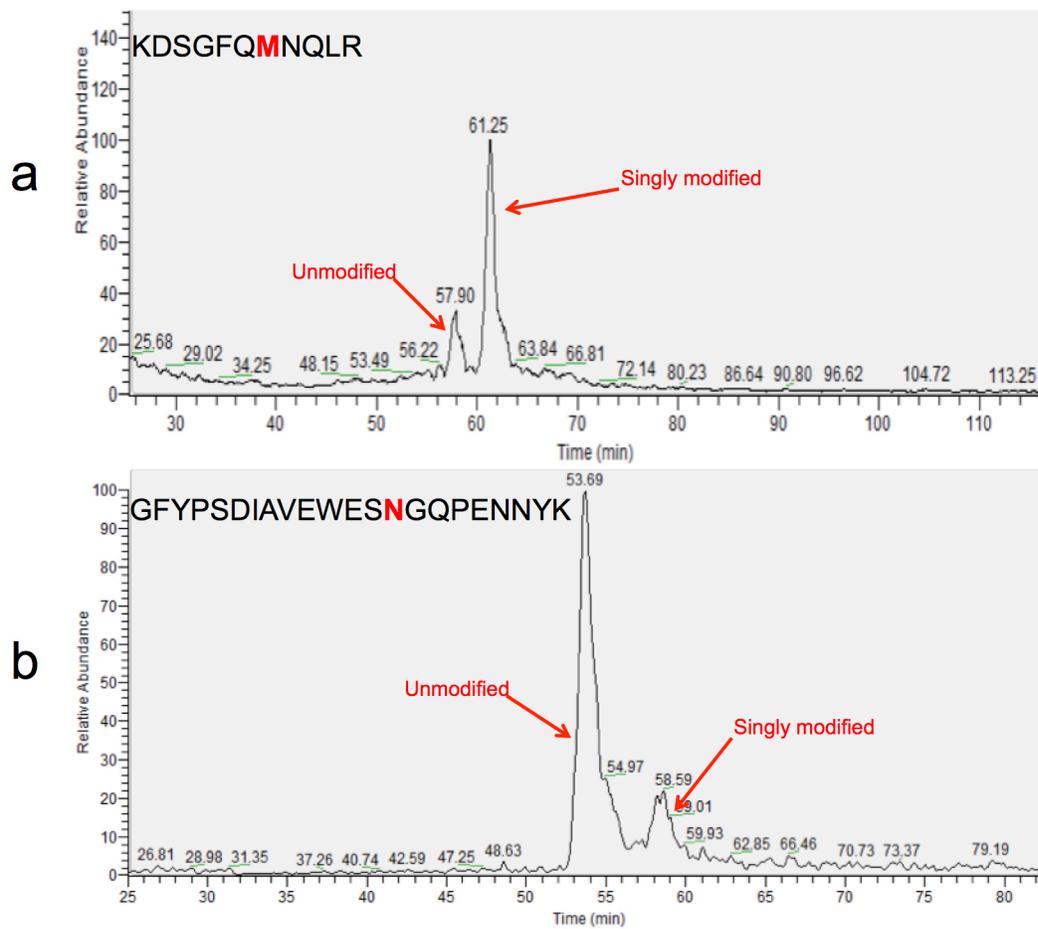


Fig. 4.5: The separation of oxidized and deamidated peptides from their native forms using a HILIC column. Deamidated and oxidized peptides are adequately separated enough from their unmodified counterparts to allow for quantitation

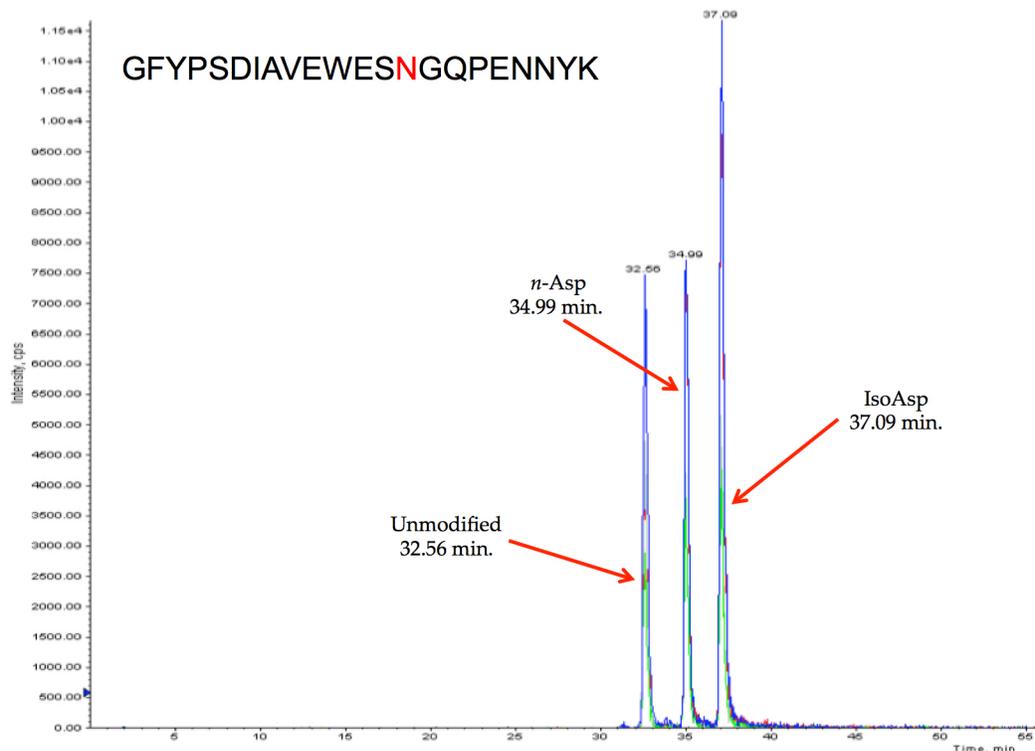


Fig. 4.6: The separation of the unmodified, *n*-Asp, and isoAsp versions of the synthetic peptide GFYPSDIAVEWESNGQPENNYK, with the site of modification at the residue in red. The unmodified form eluted first, followed by the *n*-Asp form, and finally the isoAsp form

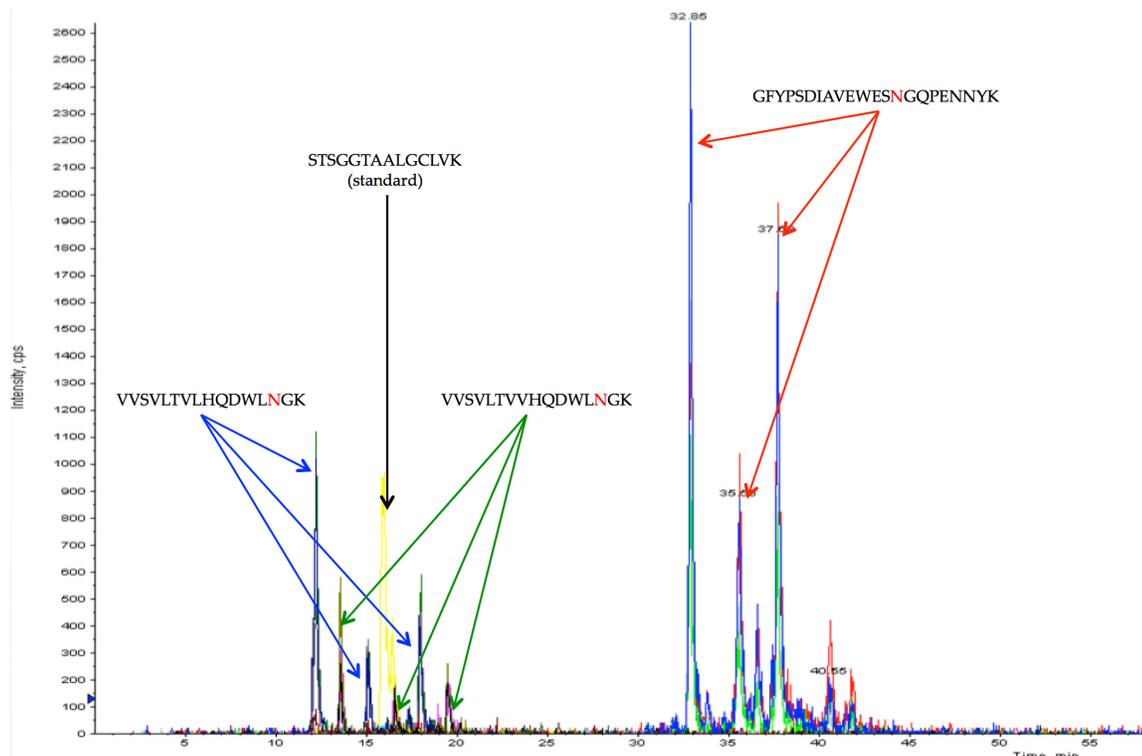


Fig. 4.7: Analysis of peptides for detecting deamidation by LCMS with the QTrap 4000. For the three peptides that have deamidation sites, there are peaks corresponding to the unmodified form, the *n*-Asp form, and the isoAsp form of the asparagine residue in red. The unmodified form of each peptide elutes first, followed by the *n*-Asp form, and finally the isoAsp form. The first two deamidated peptides differ at the 8th position, where the first peptide has a leucine and the second peptide has a valine

CHAPTER 5

THE USE OF A HILIC PEPTIDE RETENTION PREDICTION MODEL TO ANALYZE THE ISOMERIZATION OF ASPARTIC ACID AND HOW IT IS AFFECTED BY THE ADJACENT C-TERMINAL RESIDUE³

³ Badgett, M.J., Boyes, B., Orlando, R. To be submitted to *Journal of Biomolecular Techniques: JBT*.

Abstract

Amino acids on the C-terminal side of aspartic acid (*n*-Asp) residues greatly affect the rate at which *n*-Asp isomerizes to form isoaspartic acid (isoAsp) in a similar fashion to the deamidation of asparagine. This process can affect the structure, function, and lifetime of proteins, so the separation and identification of *n*-Asp and isoAsp is significant in the biotherapeutic realm. Recently, hydrophilic interaction liquid chromatography (HILIC) has been shown to be able to fully separate these two isomers due to the difference in hydrophilicity, potentially increasing the confidence in proteomic identifications and allowing for faster analyses. This paper uses a previously made HILIC peptide retention prediction model to predict the retention times of peptides with aspartic acid residues and investigate the extent of isomerization by comparing the predicted times to the actual times. In addition, the analysis of the effect that C-terminal residues have to the isomerization of aspartic acid residues is detailed, and it was found that the DG, DE, DS, and DL motifs all had trends that promoted isomerization.

Introduction

Deamidation of asparagine residues is a well-studied post-translational modification that can cause protein degradation and affect protein development.^{114,115,122,187} Through this process, asparagine residues undergo formation of a five-membered succinimide intermediate that can subsequently hydrolyze to form either aspartyl (*n*-Asp) or isoaspartyl (isoAsp) residues that can be in the D and L configurations (Figure 1). The second step of this mechanism is reversible, indicating that either product can fold back into the succinimide intermediate and undergo further structural changes, such as an aspartyl residue converting to an isoaspartyl residue, or vice versa. Because the *n*-Asp deamidation product is an aspartic acid residue, this suggests that aspartic acid can isomerize to isoaspartic acid without having being converted from asparagine initially. This can easily be overlooked due to succinimide formation from aspartic acid at neutral pH being 13-36 times slower than from asparagine, and because isoAsp is largely associated with deamidation of asparagine and not isomerization of aspartic acid.^{126,188}

The formation of isoAsp through deamidation, isomerization or racemization is largely associated with ageing and can be viewed as a molecular clock in this regard.^{19,122} In addition, it is commonly related to neurodegenerative diseases such as Alzheimer's disease, as well as linked to β -amyloid aggregation.^{17,18,21} The rate of formation of isoAsp is affected by conditions and characteristics such as temperature, pH and protein structure, where the rate of succinimide formation and subsequent hydrolysis is significantly reduced at lower pHs and temperatures.^{13-16,20,110,117-119,176,177} IsoAsp is particularly important in protein biotherapeutics, where the modification can accumulate over time and indicate the loss of protein activity or stability.¹²⁵ Because of this, it is vital to be able to fully isolate and identify the formation of isoAsp for

analysis, but can be difficult due to minimal mass differences and minimal structural changes between precursors and products.

Current techniques used to analyze isoAsp include isotopic labeling, isoAsp specific antibodies, and the use of protein isoaspartic acid O-methyltransferase, among others, but the need for a consistent, straightforward technique that can isolate and analyze isoAsp still exists.^{125,189} Common mass spectrometric methods that successfully analyze the isomerization of aspartic acid use fragmentation techniques that include ion-electron or ion-ion interactions, such as electron transfer dissociation (ETD) or electron capture dissociation (ECD), which are typically paired with RP chromatography.^{125,175,190} Using these techniques, unique fragment ions from isoAsp (*c*+57 and *z*-57) can be identified from the MS/MS spectra, but the intensities for these peaks can be severely low; almost 95% lower than other fragment ions in the spectrum.^{125,126} IsoAsp identification can also be particularly challenging in high-throughput proteomics experiments where many false positives can be reported, and separation of the isomers can be inconsistent depending on the LC conditions.^{125,175} However, HILIC has been shown to provide baseline separation between these isomers, allowing for sufficient quantitation even on a low-resolution mass spectrometer.¹²⁴

Previously, we have created a model that predicts peptide retention using HILIC and derived a coefficient for the deamidation of asparagine that is related to its overall hydrophilicity.^{124,185} What was discovered is that the derived deamidation coefficient was much more hydrophilic than the difference between the asparagine and aspartic acid coefficients in the model (corresponding to the *n*-Asp product), and this indicated that almost all of the deamidation that we were seeing resembled the formation of the isoAsp product, which is 2-3 times more abundant than *n*-Asp according to the literature.^{15,16,115,118,121,124} This discovery prompted us to

take a closer look at aspartic acid residues because the deamidation mechanism suggests that it can readily isomerize to form isoAsp. It has been described in the literature that both the N-terminal and C-terminal residues next to the modified residue affect the rate of deamidation, but that the C-terminal residue is more important.¹⁹ The motif we decided to initially look at was “DG” because asparagine undergoes deamidation at much faster rates when smaller residues such as glycine are on the C-terminal side due to low steric hindrance, but we decided to look at other prevalent motifs as well.²¹

In this paper, we show that the “DG” motif readily facilitates isomerization of *n*-Asp to isoAsp as most of the peptides (5 out of 6) that were analyzed with this motif have retention times that correlate to the isoAsp form. Our peptide prediction model was used to predict the retention of peptides as well as the isomerization. We also investigated trends from other motifs and found that peptides with “DE”, “DL”, and “DS” motifs mostly had retention times that were in correlation to the formation of isoAsp.

Materials and Methods

Protein Digestion

Human IgGs were separated from human serum (Sigma-Aldrich, St. Louis, MO, USA) using a HiTrap™ Protein G column (General Electric Company, Fairfield, CT, USA). Cytochrome C, myoglobin, transferrin, and dextran were purchased from Sigma-Aldrich. Yeast plant tissue was purchased from Thermo-Fisher (San Jose, CA, USA). Proteins were reduced using 10 mM dithiothreitol (DTT) and then alkylated using 55 mM iodoacetamide (IDA), both purchased from Sigma Aldrich. Sequencing-grade trypsin or chymotrypsin purchased from Promega (San Luis Obispo, CA, USA) was added at 50:1 (w/w, protein/trypsin) for incubation overnight in 50 mM ammonium bicarbonate (pH 7.0) at 37°C.

LC-MS/MS Settings and Instrumentation

Data were acquired using a Finnegan LTQ (Thermo-Fisher, San Jose, CA, USA) in series with an 1100 Series Capillary LC system (Agilent Technologies, Palo Alto, CA, USA) with an ESI source that used spray tips made in-house. Samples were suspended in 25% H₂O, 75% ACN and 0.1% formic acid (Sigma-Aldrich, St. Louis, MO, USA) for direct injection into the LC system. Peptides were separated using a 200- μ m x 150-mm HALO® Penta-HILIC column packed with 2.7 μ m diameter superficially porous particles that have a 90 Å pore diameter (Advanced Materials Technology, Wilmington, DE, USA) at room temperature. The gradient elution conditions employed a linear increase in aqueous solvent from 5-70% over 90 minutes at a 2 μ L/min flow rate. The (strong) aqueous solvent contained 0.1% formic acid (Sigma Aldrich, St. Louis, MO, USA) with 50 mM ammonium formate (Thermo-Fisher, San Jose, CA, USA) and the organic solvent was acetonitrile with 0.1% formic acid. The settings for the mass spectrometer included taking the 5 most intense ions from each full mass spectrum for fragmentation using collision-induced dissociation (CID), and the resulting fragmentation spectra were recorded.

Some of the same digested proteins were run on a 4000 Q Trap (Sciex, Framingham, MA, USA) with a selected reaction monitoring (SRM) method. Peptides were separated by a 2.1 mm x 15 cm HALO® Penta-HILIC column packed with 2.7- μ diameter superficially porous particles using a Nexera UFLC (Shimadzu, Columbia, MD, USA). The temperature of the column was 60°C. The gradient used for each sample was 22-52% water over 80 minutes at a 0.4-mL/min flow rate. Spectra were obtained using an ESI source.

Database Search Parameters

The resulting RAW files were converted using Trans-Proteomic Pipeline (Seattle Proteome Center, Seattle, WA, USA), then the MS/MS spectra of each sample were searched using Mascot (Matrix Scientific, Boston, MA, USA) against corresponding protein databases of theoretical MS/MS spectra. The following parameters were utilized in Mascot: a peptide tolerance of 1000 ppm, a fragment tolerance of 0.6 Da, two max missed cleavages of trypsin, and a fixed modification of carbamidomethyl (C).

Selection of Peptides for Prediction Model and Post-Run Data Analysis

All peptides that had a higher Mascot score than 10 were considered. Peptide retention times were determined manually from .RAW files using the apex of the peaks displayed in Xcalibur software (Thermo-Fisher, San Jose, CA, USA), and resulting MS/MS data were visually inspected for fragmentation that was consistent with peptide assignments. Peptide retention times in minutes were converted to glucose units based on dextran samples that were run immediately before.

Results and Discussion

“DG” Motif

It is important to note that we previously discovered that isoAsp is more hydrophilic than *n*-Asp, and that peptides with these two modifications can be fully separated from one another.¹²⁴ The difference in hydrophilicities comes from the position of the CH₂ group in either the peptide backbone (isoAsp) or the side chain (*n*-Asp) (see Figure 1), and we were able to derive coefficients so that we could predict the retention of peptides with this isomeric modification. With the CH₂ group in the peptide backbone, the carboxylic acid side chain of isoAsp will be

more polar than *n*-Asp and will interact with the polar stationary phase and water-rich layer to a greater extent.

Six peptides in our samples that contained a “DG” motif were identified with single and clear chromatographic peaks. The retention times of these peptides were found by hand and were compared to predicted retention times that correlate to either the presence of *n*-Asp or isoAsp (Table 1). These retention times were converted from minutes into glucose units (GU) from procainamide-labeled dextran samples that were run before and after the peptide samples as a retention time calibrant. Five of these peptides had actual retention times closer to the isoAsp prediction, and one of these peptides had actual retention times closer to the *n*-Asp prediction. Although it is a relatively small sample size, the peptides that were closer to the isoAsp predictions had an average difference of 0.160 GU, or 2.10 minutes. This small difference indicates the prediction accuracy of the model, but also that the aspartic acids are readily isomerizing to isoaspartic acid with a glycine residue on the C-terminal side at a pH of around 8. Similarly to deamidation, the small residue allows the asparagine or aspartic acid to fold onto itself and form the succinimide intermediate with low steric hindrance. It is worth noting that the lone peptide that was closer to the *n*-Asp prediction, VLAVGDGIAR, had around 2-3 times higher of a difference in predicted versus actual retention (0.486 GU or 4.76 min) than the average of the five peptides that were closer to the isoAsp prediction. This could be due to almost all of the amino acid residues being hydrophobic or small and thus inhibiting the peptide from interacting with the stationary phase. The peptide that had the highest GU difference, AVDDFLLSLDGTANK, has two sites of isomerization, DD, and DG, in addition to a site of deamidation, NK. If the predicted times were adjusted to account for all three sites forming isoAsp, the difference goes from 1.568 GU to 0.349 GU. This is entirely possible because

hydrophilic residues are known to promote deamidation of asparagine, and the same could be said about isomerization of aspartic acid.

To further investigate the presence of *n*-Asp and/or isoAsp, two peptides (TTPPVLDSDGSFFLYSK and FNWYVDGVEVHNAK) from human IgG samples were run on a 4000 Q Trap with a selected reaction monitoring (SRM) method. For each peptide, only one chromatographic peak was found and the actual retention times matched up to the formation of isoAsp (Figure 2). This suggests that the peptides are readily isomerizing. There is a very small peak in the extracted ion chromatogram for the FNWY peptide, which could represent a minimal amount of *n*-Asp.

Other Motifs

Apart from glycine being a residue that promotes deamidation, other residues that are either small and/or have polar functional groups such as threonine, serine and histidine have been shown to increase it as well. Very large and nonpolar residues such as tryptophan, tyrosine, or phenylalanine have low rates of deamidation because their size hinders the intermolecular attack to form the succinimide intermediate.^{19,21,191} Stephenson and Clark investigated the rate of deamidation of asparagine for five different adjacent C-terminal residues at a pH of 7.4 and found the order of decreasing rate to be glycine, serine, alanine, leucine and proline.¹⁸⁸ Because the hydrophilic serine had a faster rate than the small alanine, they hypothesized that the hydroxyl group could increase the nucleophilicity of the nitrogen of the peptide bond that attacks the functional group by deprotonating it, or it could bond to either the nitrogen or oxygen atoms on the side chain of asparagine to increase the electrophilicity of the carbon atom in the side chain.¹⁸⁸ In either case, the hydrophilic residue provides chemical interactions that promote the formation of the succinimide intermediate to a greater extent than smaller residues (excluding

glycine) and hydrophobic residues. To see if the residues that affect the rate of asparaginyl deamidation similarly affect the rate of aspartic acid isomerization, we also investigated the retention of peptides with other residues besides glycine.

Table 2 shows the amount of peptides that were closer to the *n*-Asp or isoAsp prediction separated out by motif. The motifs that have clear trends are DE, DL, and DS. For many of the motifs, there were not enough instances to determine if there was a trend. However, all peptides with the DE motif (8 out of 8) were closer to the isoAsp prediction, which supports the hypothesis that the isomerization of aspartic acid is promoted by hydrophilic residues on the C-terminal side. This was also the case with DS, as 4 out of the 5 peptides that had this motif were closer to the isoAsp prediction as well. Both motifs had some of the lowest deviations from predicted to actual retention times (1.26 min. for DE and 0.96 min for DS), which shows the accuracy of prediction for the peptides containing these motifs.

Curiously, DL also had a trend (5 out of 6) that implies that leucine residues on the C-terminal side promote the isomerization of aspartic acid. Our previously made HILIC retention model indicates that it is the third most hydrophobic amino acid residue overall (Badgett). This is especially curious given that the retention times of the only two peptides with a DI motif were closer to *n*-Asp predictions. In addition, Stephenson and Clarke's paper found that deamidation forms at a much slower rate with a leucine residue at the C-terminal side of the asparagine, at almost 50 times slower in comparison to glycine and around nine times slower in comparison to serine.¹⁸⁸ A possible explanation for the existence of this trend is that peptides with the DL motif had one of the highest deviations from actual to predicted times (3.27 min or 1.64 GU) out of any motif, and this suggests that the prediction of some or most of the peptides containing the DL motif could lack in accuracy. The difference in hydrophilicity between *n*-Asp and isoAsp is

0.609 GU, less than half the deviation of the DL peptides. Given the small sample size of this and most of the motifs, we would like to analyze many more peptides with these motifs to see if they actually do promote isomerization.

Summary

The main aspect of this study to consider is that the analysis of isomerization was based on how the peptides matched up to predicted retention times. We did not employ a fragmentation method such as ETD or ECD to individually distinguish between *n*-Asp or isoAsp versions of all of the peptides because we wanted to show that our predictive model could be fully capable of accurately predicting whether peptides had *n*-Asp or isoAsp as well as to investigate the affect that different C-terminal residues had to retention. Thus, while the isomerization trends we saw for the DG, DE and DS motifs were analogous to previous findings in the literature, the dependence on retention time prediction for modification assignment showed a flaw, as it was found that the DL motif also promoted isomerization of aspartic acid as well. Nevertheless, the model was very useful in determining which peptides matched up to *n*-Asp or isoAsp predictions, enabling quick analysis of the modification with the potential to further the confidence in modified peptide identifications.

Acknowledgements

Support for this work comes from NIH Grant GM0 93747 to BEB.

Table 5.1

Retention times of peptides containing the “DG” motif were compared to predicted times corresponding to the presence of *n*-Asp or isoAsp. Differences were calculated by subtracting the actual retention times (RT) from the predicted RT that was the closest.

Numbers in red were closest to the actual retention times.

Peptide	Actual RT (GU)	Predicted RT (<i>n</i>-Asp)	Predicted RT (isoAsp)	Difference (GU)
TTPPVLDSDGSFFLYSK	4.001	3.195	3.804	0.197
FNWYVDGVEVHNAK	5.493	4.903	5.512	0.019
VLAVGDGIAR	2.345	2.931	3.540	0.586
AVDDFLLSLDGTANK	6.423	4.246	4.855	1.568
TALVHDGLAR	5.112	4.739	5.359	0.247
TFIAVKPDGVQR	6.072	5.677	6.286	0.214

Table 5.2

Peptides from a wide range of motifs that had actual retention times either closer to *n*-Asp or isoAsp predicted times. Bold motifs were considered a trend.

Motif	# of Peptides Closer to <i>n</i> -Asp Prediction	# of Peptides Closer to isoAsp Prediction
DA	3	3
DE	0	8
DF	0	1
DI	2	0
DK	0	1
DL	1	5
DN	0	1
DP	1	3
DQ	0	2
DS	1	4
DT	1	1
DV	2	2
DY	1	0

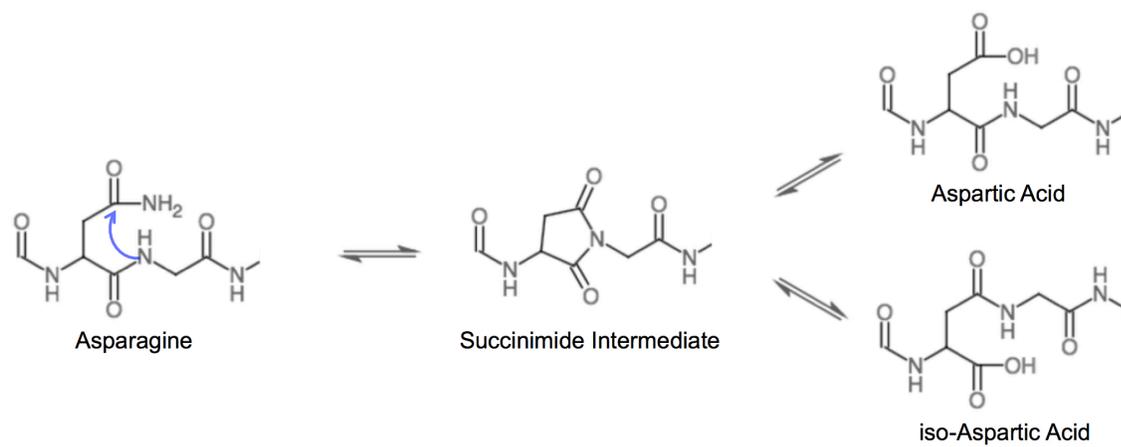


Figure 5.1: The deamidation of asparagine mechanism

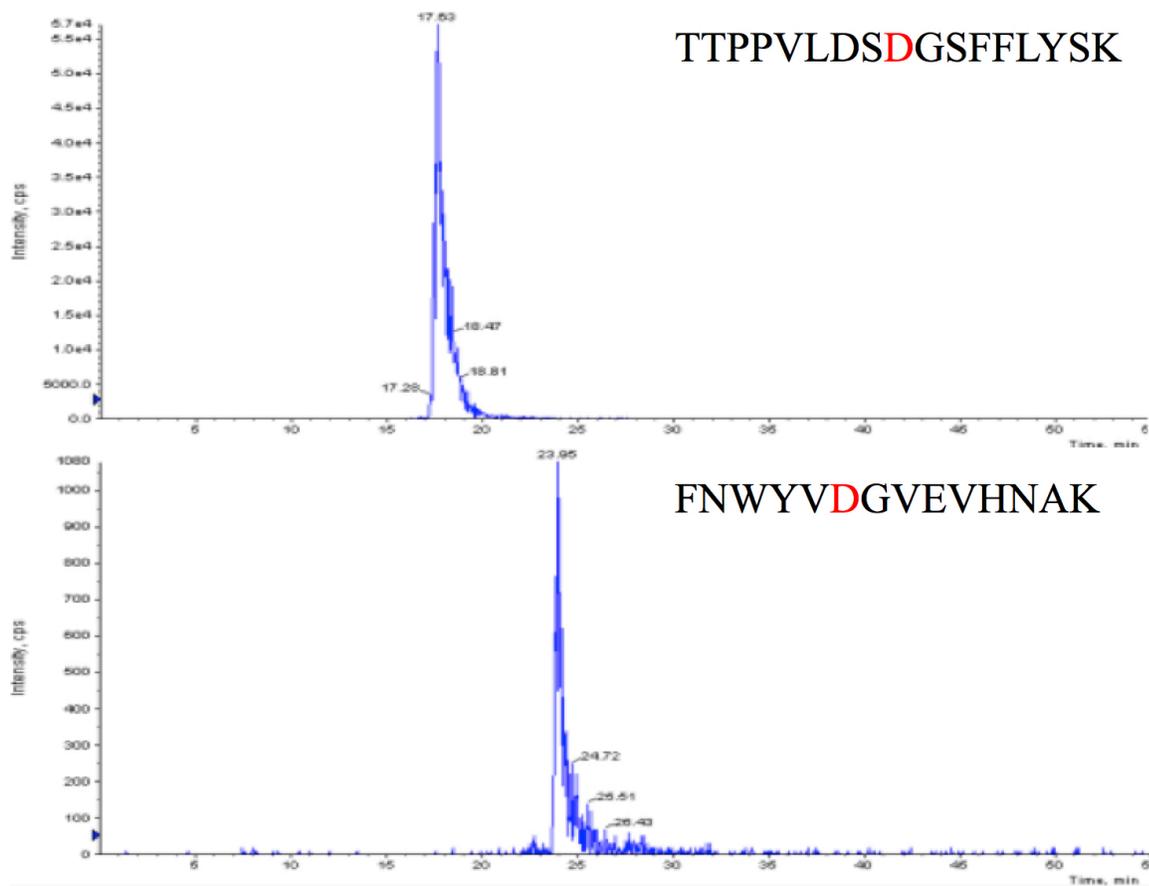


Figure 5.2: Extracted ion chromatograms for two peptides TPPVLSDGSFFLYSK and FNWYVDGVEVHNAK run on the 4000 Q Trap with a SRM method. The red letter indicates the site of isomerization.

CHAPTER 6
PREDICTING THE RETENTION BEHAVIOR OF SPECIFIC O-LINKED
GLYCOPEPTIDES⁴

⁴ Badgett, M.J., Boyes, B., Orlando, R. To be submitted to *Journal of Biomolecular Techniques: JBT*.

Abstract

O-linked glycosylation is a prevalent post-translational modification that changes the overall structure, polarity, and function of proteins. Current chromatographic techniques used to analyze O-glycosylated peptides and their native forms primarily rely on reverse-phase (RP) chromatography, which generates minimal separation. Hydrophilic interaction liquid chromatography (HILIC) can be a solution to this problem, as the polar glycan addition would greatly interact with the polar stationary phase and could potentially generate enough separation to separately identify the peptide from its modified form (something about it being easier to identify). In this paper, HILIC is employed to separate peptides with O-GalNAc, O-GlcNAc, and O-fucose additions from their native forms, and coefficients representing the extent of hydrophilicity were derived using linear regression analysis as a means to predict the retention times of peptides with these modifications.

Introduction

The ability to resolve glycosylated peptides from their native forms would greatly facilitate glycoproteomics since this would enable glycan characterization while connected to a peptide tag that provides the glycan's location. Reversed-phase liquid chromatography (RP-LC) is typically used to separate peptide/glycopeptide mixtures, where hydrophobic interaction drives the retention. Hence the hydrophilic glycans, which do not interact significantly with the RP stationary phase, play little to no role in the separation. This behavior makes it difficult to resolve glycopeptides from their unmodified counterparts by RP, which is particularly true when the glycan consists of a single monosaccharide. Alternatively, glycans interact extensively with the stationary phases used in hydrophilic-interaction liquid chromatography (HILIC). Therefore, HILIC should be capable of resolving these peptide/glycopeptide pairs.

The attachment of a single monosaccharide to the side chain of serine and threonine residues has been found to play a wide range of biological functions, including inflammation response, lubrication, or protein stability.^{151,152} Two examples of glycans analyzed in the present study are O-linked N-acetylglucosamine (O-GlcNAc) and O-linked N-acetylgalactosamine (O-GalNAc), which are isomeric and attached to serine or threonine residues in proteins. The attachment of these monosaccharides plays an important role in numerous biological diseases. O-GlcNAc is involved in enzymatic activity, protein function, and disease-relevant signaling, and has been found on over 1,000 proteins.^{162,192-194} It can rapidly be added or removed from target proteins due to the presence of O-GlcNAc transferase (OGT) and O-GlcNAcase (OGA) and the cycling of O-GlcNAc structures on proteins is due to environmental changes including many forms of stress.¹⁶⁴⁻¹⁶⁶ Abnormal O-GlcNAc modification is associated with numerous diseases, such as Alzheimer's disease, Huntington's disease, diabetes, systematic lupus erythematosus,

and Parkinson's disease, as well as many types of cancer.^{163,165,192,194–196} O-GalNAc modification is referred to as the Tn antigen and is often found on mucins or mucin-like proteins associated with tumor cells. This modification is present in more than 10% of human proteins and over 85% of proteins passing through the Golgi apparatus.¹⁵³ In contrast to O-GlcNAc, O-GalNAc modification is initiated by 20 polypeptide GalNAc transferases (GalNAc-Ts) that control where the site of glycan attachment in proteins.^{151,153,154,197,198} However, similarly to O-GlcNAc, aberrant modification of O-GalNAc is associated with a wide range of diseases such as acute coronary disease, congenital heart disease, various types of cancer, and many others.^{154,198–200}

Another monosaccharide that is found attached to proteins through an O-linkage to a serine or threonine residue is fucose, commonly known as O-fucosylation. This modification was first discovered by Hallgren, et. al. in 1975 and is commonly found on epidermal growth factor (EGF) like repeats as well as thrombospondin type-1 repeats (TSR).^{156,159,161,201} Similarly to O-GalNAcylation, other monosaccharides are frequently added to the core O-fucose and this elongation can be critical for protein activity.²⁰² The attachment of O-linked fucose to proteins is catalyzed by eleven different fucosyltransferases that have guanosine diphosphate-fucose (GDP-fucose) as the donor substrate, and these fucosyltransferases most commonly reside in the Golgi apparatus but have also been found in the endoplasmic reticulum.^{156,157,203–205} Although fucose is found to play an essential role in physiological processes such as blood type determination and Notch signaling as well as other signal transduction processes, it is linked to many different cancers, rheumatoid arthritis, cystic fibrosis, and leukocyte adhesion deficiency type II (LAD II), among others.^{155,157,159,204,206–213}

We have previously devised a model that predicts peptide retention using HILIC that provides amino acid coefficients describing their hydrophilicity, and are currently devising a

model that predicts the retention of glycopeptides in HILIC separations.¹⁸⁵ Dextran is used as a retention calibrant, enabling peptide and glycopeptide retention to be expressed in glucose units (GU) that permits a comparison of peptide retention across different LC-MS systems. Combining retention models of HILIC separations of peptides and glycans is directed to prediction of retention of previously uninvestigated glycopeptides. Here, we show that glycopeptides can be resolved from the native peptide, and retention coefficients that represent the hydrophilicity of O-GlcNAc, O-GalNAc, and O-fucose are derived. This will permit a more straightforward determination of the extent and type of glycosylation for the glycans analyzed in this study.

Methods and Materials

Glycopeptides

Synthetic peptide-glycopeptide pairs were obtained from GlycoScientific LLC (Athens, GA, USA).

LC-MS/MS Settings and Instrumentation

Data were acquired using a Finnegan LTQ (Thermo-Fisher, San Jose, CA, USA) in series with an 1100 Series Capillary LC system (Agilent Technologies, Palo Alto, CA, USA) with an ESI source that used spray tips made in-house. Samples were suspended in 25% H₂O, 75% ACN and 0.1% formic acid (Sigma-Aldrich, St. Louis, MO, USA) for direct injection into the LC system. Samples were separated using a 200- μ m x 150-mm HALO® Penta-HILIC column packed with 2.7- μ m diameter superficially porous particles (Advanced Materials Technology, Wilmington, DE, USA) at room temperature. The gradient elution conditions employed a linear increase in aqueous solvent from 5-70% over 90 minutes at a 2 μ L/min flow rate, using the column at room temperature. The mobile phases contained 0.1% formic acid (Sigma Aldrich, St. Louis, MO, USA) in acetonitrile, and the (strong) aqueous solvent contained 50 mM ammonium

formate (Thermo-Fisher, San Jose, CA, USA). The settings for the mass spectrometer included taking the 5 most intense ions from each full mass spectrum for fragmentation using collision-induced dissociation (CID), and the resulting MS/MS spectra were recorded.

Selection of Peptides for Prediction Model and Post-Run Data Analysis

Retention times were determined manually from .RAW files using the apex of the peaks displayed in Xcalibur software (Thermo-Fisher, San Jose, CA, USA), and resulting MS/MS data were visually inspected for fragmentation that was consistent with peptide assignments.

Glycopeptide retention times in minutes were converted to glucose units based on dextran samples that were run immediately before. Linear regression analysis using StatPlus (AnalystSoft, Walnut, CA, USA) was used to find retention coefficients for the modifications.

Results and Discussion

The Separation of Glycopeptides from their native forms

The ability to resolve the native peptide from the species with a single monosaccharide is shown by the HILIC chromatogram of the peptide GTTPSPVPTTSTTSAP, with the underlined amino acid residues representing the sites of O-GalNAcylation (Figure 1). The unmodified version of the peptide elutes first, followed by peptide with only one modification at the third threonine residue, followed by the peptide with both modifications at the third and thirteenth residues. All the peaks are baseline separated, demonstrating the ability of the HILIC column to separate this highly hydrophilic modification from the native peptide. The difference between the unmodified peptide and singly modified peptide in this figure is 5.28 minutes, which is higher than the difference between the singly modified peptide and doubly modified peptide at 3.70 minutes. Taking into account the additional replicates, the average difference between the first two peaks is 1.58 minutes larger than the average difference between the last two peaks. This

suggests that the first addition primarily drives the interaction with the stationary phase and water rich layer, and that the hydrophilic addition of the second sugar does not change the overall hydrophilicity of the peptide as much as the first. There was no retention time difference between peptides that only had the third residue modified versus peptides that only had the thirteenth residue modified.

The separation of the O-GlcNAcylated peptide VPTTAASTPDAVDK is shown in Figure 2. Similarly to the first peptide, the modified form elutes later than the native form due to the hydrophilicity of the modification. These peaks are also baseline separated, allowing for quantitation. Due to the near-identical structure of O-GlcNAc to O-GalNAc, the retention time difference between peptide and glycopeptide are analogous to the native and singly modified peaks in the previous figure, and is discussed in more detail below.

O-fucosylation is a slightly smaller monosaccharide addition in comparison to O-GalNAcylation and O-GlcNAcylation, but retains in a similar way on the penta-HILIC column. Figure 3 shows this behavior, as the native and modified forms of the peptide CQNGGTCHNTH are fully separated, with the O-fucosylated form eluting later. However, the retention time difference of this peptide-glycopeptide pair is smaller than the other two types of O-linked glycosylation. This is expected as fucose lacks the amide and acetyl groups that O-GlcNAc and O-GalNAc have, which are both hydrophilic and increase the retention.

All three of these examples show that HILIC is not only capable of separating O-glycosylated peptides from their native counterparts, but that it can baseline resolve the peaks, allowing for quantitation and facilitating quicker and easier identifications.

Retention Time Prediction of O-glycans

The prediction of retention for peptides with these modifications can further the identification process, but will only work if the retention times are reproducible. All of the retention times of the native and modified peptides over the course of four runs are shown in Table 2. Their retention times in minutes were converted to glucose units (GU) from procainamide-labeled dextran samples run immediately before and after the protein samples. This enables the comparison of retention across LC-MS systems, and provides a retention time calibrant for these samples. The differences in retention times from run to run are miniscule, indicating the reproducibility and potential predictability of retention using the HILIC column. This has large implications in targeted selected reaction monitoring (SRM) experiments, where the time range for looking for an analyte of interest can be reduced. Standard deviations from the four runs for each sample are also low, with the average standard deviation being 0.094 GU, or 0.357 min. This is extremely precise for a method that uses a 90 minute long LC-MS gradient at a 2 μ L/min flow rate.

To analyze how the modifications are influencing the overall hydrophilicity of the peptides, linear regression analysis of the retention times of peptides with and without modification was used to derive coefficients (Table 2). These coefficients represent the hydrophilicity of the modification, which are almost as great as the most hydrophilic amino acids at a low pH using the same HILIC column: histidine, lysine and arginine.¹⁸⁵ The coefficients are also displayed in GU, and their high coefficient values were expected due to the size and polar characteristics of the sugar additions. O-GlcNAc and O-GalNAc only differ by the position of some of their hydroxyl groups, so the isomeric glycans should have similar coefficients, which is the case. However, O-fucose does not have the amide or acetyl group that the other two

modifications have in addition to one of the hydroxyl groups being substituted out for a terminal methyl group. This is why the coefficient is the lowest of all the monosaccharide additions.

It is important to note that the three peptides that have O-GlcNAc modifications on a serine residue, KRGRKESYSIYVYK, RGGVKRISGLIYEE, and CKRGRKESYSIYVYK have an average retention time difference of 0.706 GU from their unmodified forms, whereas the two peptides that have an O-GlcNAc modification on a threonine residue, VPTTAASTPDAVDK and CKSAPATGGVKK, have an average retention time difference of 0.990 GU from their native forms. While this could be attributed to the difference in amino acid structures, the theoretical difference between the coefficients for serine and threonine derived in our HILIC peptide retention prediction model (0.334 and 0.357, respectively) suggest that there should be a minimal change in retention. In comparison with the retention time differences in GU, the retention time difference in minutes between the peptides listed above with the O-GlcNAc modifications either on a serine or threonine residue is small (3.21 min. and 3.66 min., respectively). While using dextran as a retention time calibrant is vital for the use of these predictive coefficients across LC-MS systems, there can be contrasts depending on where the analytes elute. If the peptide-glycopeptide pairs elute earlier in the gradient, they will have smaller retention time differences than if they elute later in the gradient because of the logarithmic way that dextran elutes. However, the 0.2 GU difference between O-GlcNAc modification on a serine or threonine residue is still small enough for us to be able to predict the retention of these peptides within about half a minute, showing that despite dextran's logarithmic elution, it is fully capable of standardizing our results.

Summary

The attachment of a single monosaccharide to a peptide increased the overall hydrophilicity and led to the glycosylated species being more retained on a HILIC column, as expected. The shift in HILIC retention caused by the addition of these monosaccharides were determined to be comparable to amino acid coefficients that were the most hydrophilic in a previously made peptide retention prediction model. This work showed the ability of the HILIC column to fully separate these peptide/glycopeptide pairs reproducibly and in a predictable manner, which will allow for easier identification, characterization, and quantification. While there were no major differences between O-GalNAc and O-GlcNAc, O-fucosylated peptides were slightly less hydrophilic.

Acknowledgements

Support for this work comes from NIH Grant GM0 93747 to BEB.

Table 6.1

Retention times of unmodified and O-glycosylated peptides over four runs. Amino acids in red represent the site(s) of modification.

Peptide	Run 1 RT (GU)	Run 2 RT (GU)	Run 3 RT (GU)	Run 4 RT (GU)	Avg. RT (GU)	St. Dev.
VPTTAASTPDAVDK	6.680	6.790	6.841	6.871	6.796	0.084
VPT T AASTPDAVDK	7.960	8.123	8.120	8.105	8.077	0.078
GTPSPVPTTSTTSAP	5.111	5.026	5.174	5.174	5.121	0.070
GTPSPVPTTST T SAP	6.517	6.581	6.607	6.529	6.559	0.042
G T PSPVPTTSTTSAP	6.517	6.492	6.563	6.483	6.514	0.036
G T PSPVPTTST T SAP	7.712	7.866	7.787	7.684	7.762	0.082
KRGRKESYSIYVYK	8.412	8.366	8.171	8.306	8.314	0.105
KRGRK E SYSIYVYK	9.177	8.879	8.742	8.646	8.861	0.231
RGGVKRISGLIYEE	3.412	3.226	3.341	3.421	3.350	0.090
RGGVKR I SGLIYEE	6.167	3.968	4.068	4.241	4.111	0.119
CKLLGRV T IAQGG	4.352	4.434	4.405	4.438	4.407	0.040
CKRGRK E SYSIYVYK	7.877	7.956	7.867	7.916	7.904	0.041
CKRGRKESYSIYVYK	7.105	7.272	7.167	6.832	7.094	0.188
CELAKHAV S EGTKA	6.821	6.982	6.852	7.011	6.916	0.094
CKSAPAT G GVKK	7.191	7.312	7.275	7.360	7.284	0.072
CKSAPATGGVKK	6.480	6.614	6.584	6.661	6.585	0.076
CFNGGTCVDGIN	2.794	2.781	2.738	2.576	2.722	0.100
CFNGG T CVDGIN	3.322	3.130	3.314	3.386	3.288	0.110
CQNGGTCHNTH	6.073	6.175	6.124	6.180	6.138	0.050
CQNGG T CHNTH	6.821	7.022	6.940	6.898	6.920	0.084

Table 6.2

Retention time coefficients for the three modifications compared against the three most hydrophilic modifications in a previously made peptide retention prediction model using the same column and chromatographic conditions.

Modification/Amino Acid Residue	Coefficient (GU)
O-GlcNAcylation	1.637
O-GalNAcylation	1.758
O-fucosylation	1.438
Histidine (H)	1.564
Lysine (K)	2.121
Arginine (R)	1.828

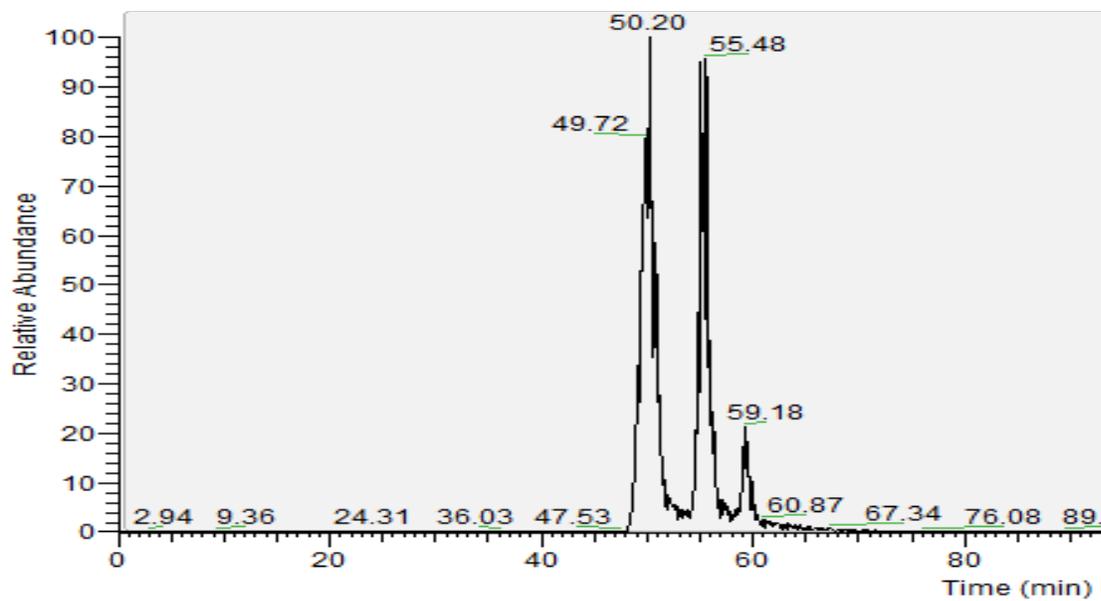


Figure 6.1: The separation of the O-GalNAcylated peptide GTPSPVPTTSTTSAP, where the earliest eluting peak is unmodified, the middle peak is modified at the third residue only and the most retained peak has both modifications.

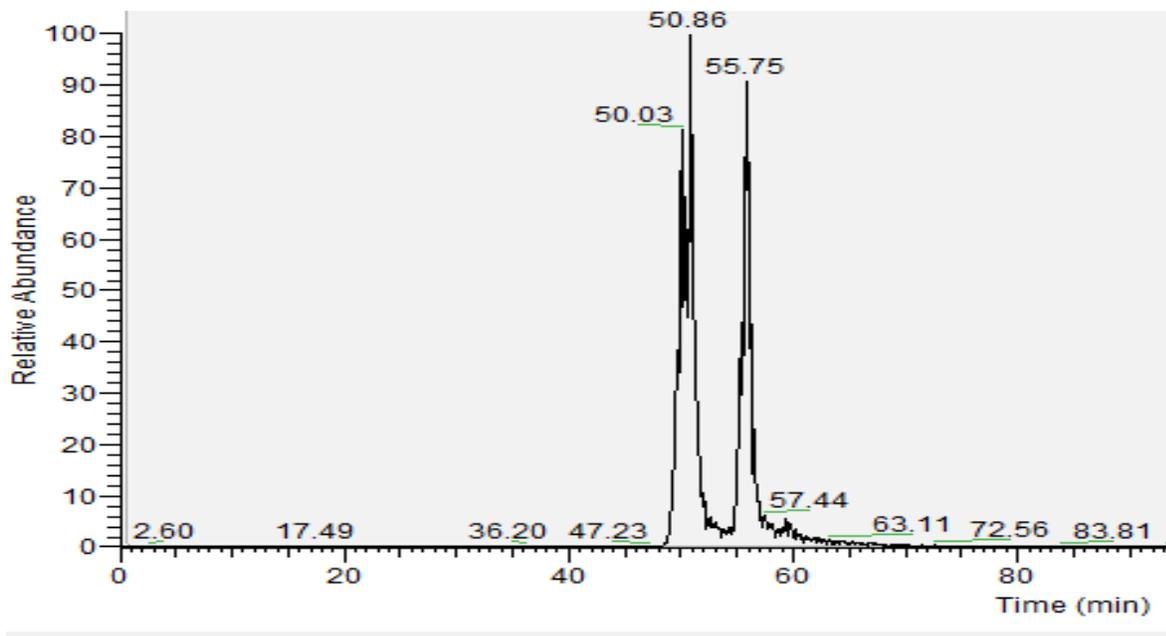


Figure 6.2: The separation of the O-GlcNAcylated peptide VPTTAASTPDAVDK, where the earlier peak is the native version of the peptide and the later peak is the modified version.

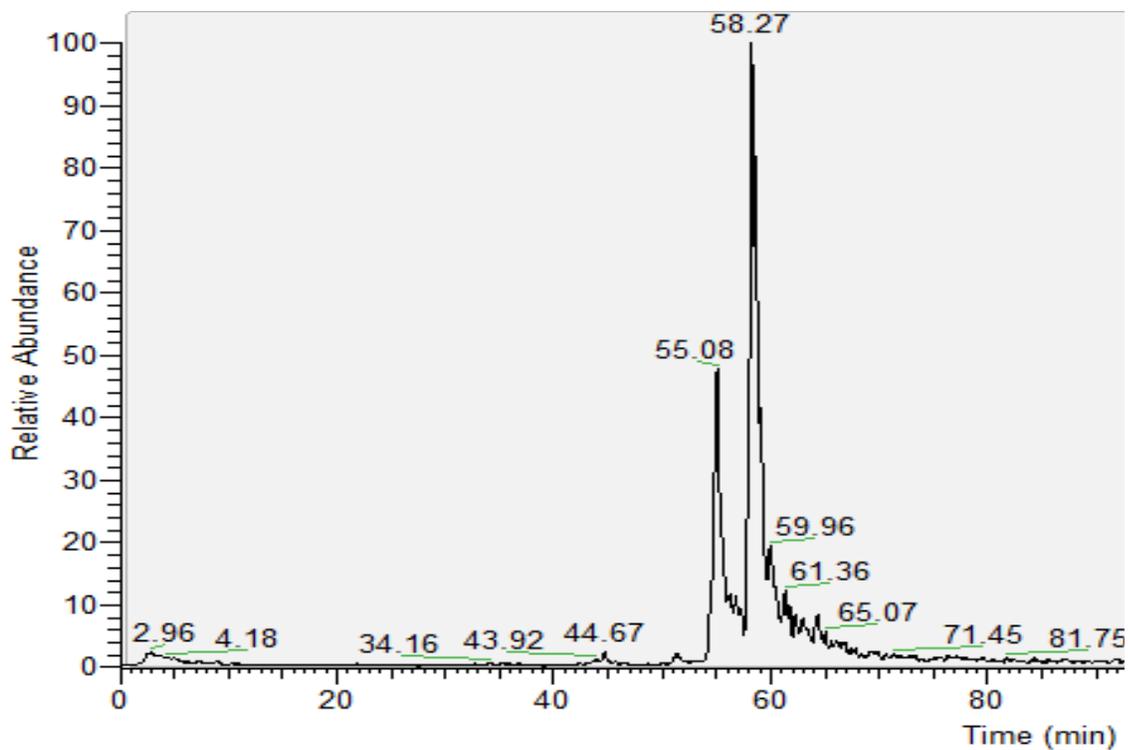


Figure 6.3: The separation of the O-fucosylated peptide CQNGGTCHNTH, where the earlier peak is the native version of the peptide and the later peak is the modified version.

CHAPTER 7

RETENTION TIME PREDICTION OF N-LINKED GLYCOPEPTIDES FROM HUMAN IMMUNOGLOBULIN GS USING HILIC-MS/MS⁵

⁵ Badgett, M.J., Betchy, E., Boyes, B., Orlando, R. To be submitted to *Analytical Chemistry*.

Abstract

Retention time prediction would greatly facilitate the identification and characterization of glycoproteins including more analytically challenging types such as isoforms and low abundance species. This is important in biotherapeutic industry, where the type and amount of glycosylation affects the folding, interaction, and function of proteins. Here, we combine two existing peptide and glycan prediction models using hydrophilic interaction liquid chromatography (HILIC) to predict the retention of glycopeptides from human IgGs. We show that our previously made peptide model is capable of accurately predicting the retention of native IgG peptides on two completely different LC-MS systems, and that in conjunction with the glycan model and an intercept, glycopeptide retention can be predicted as well.

Introduction

Glycosylation is one of the most common co- or post-translational modifications, as over 50% of eukaryotic proteins are glycosylated.^{150,214,215} This modification can affect the structure, function, interaction, and folding of proteins, and is linked to numerous diseases including rheumatoid arthritis (RA), various types of cancer, Crohn's disease, and tuberculosis, among over 50 others.^{143,144,216–223} N-linked glycosylation, meaning the carbohydrate is linked through a nitrogen atom on an asparagine residue, follows the consensus sequence Asn-X-Ser/Thr, where X can be any amino acid residue except proline, and this modification adds a substantial carbohydrate to the modified protein, increasing the polarity and mass of the protein to a high degree. The analysis and characterization of glycans moieties is essential to understanding their function, as there are numerous structural possibilities from the three types of N-glycans: high mannose, complex and hybrid.¹⁴⁹

Several important examples of glycosylated proteins in humans include immunoglobulin Gs (IgG), which comprise 75% of the antibodies circulating in human blood serum.²²⁴ IgGs have been essential in the biotherapeutic realm, as they have been the foundation of many engineered monoclonal antibodies (mAb) that treat diseases. It is important for these mAbs to agree with the human body and have long serum half-lives, therefore the analysis and characterization of the glycosylation of IgGs is imperative.^{219,220,224} There are four subclasses of IgGs (IgG1, 2, 3 and 4) that have minimal differences in their constant region (over 90% homology), but have a glycosylation site at the N297 position, allowing for binding to Fc gamma receptors (FcγR). The majority of the glycans at this position have a complex biantennary structure that is core-fucosylated, with some having bisecting structures or varying degrees of sialylation, however the degree and structure of the glycosylation varies based on a human's physiological

conditions.^{218,220,225,226} One such example is age, as the level of galactosylation changes with age in addition to a decrease in sialylation the older one gets.^{3,224} Another is pregnancy, which leads to an increase in both sialylation and galactosylation.^{219,220,224} These examples highlight the importance of knowing what glycans are present on the IgGs.

The microheterogeneity and diversity of glycans makes identification challenging, especially among structural or linkage isomers.²¹⁶ Tandem mass spectrometry (MS/MS) has emerged as a vital tool for glycan analysis, as it is able to provide structural information that can help in identification. However, isomeric identification can be extremely difficult without employing a method of separation before MS analysis. Since glycosylation is a highly hydrophilic addition, hydrophilic interaction liquid chromatography (HILIC) has been shown to help in this regard, and provides a consistent, predictable retention.^{124,185} This would aid in identification of relevant sialic acid linkage isomers in IgGs that contribute to anti-inflammatory responses. It has been shown that α 2-6 linked sialylation increases anti-inflammatory activity and that α 2-3 linked sialylation does not, and this important difference could be more easily identified with a method of separation that utilizes the change in hydrophilicity based on sialic acid linkage.^{214,219,222,224,227–229}

Previously, we have created a model that predicts peptide retention on the basis of amino acid composition using HILIC and have also created another model that predicts glycan retention with the same column.¹⁸⁵ Our hope is that we can pair them together to try and predict the retention of glycopeptides. This would help reduce the complexity of glycan identification and characterization, as well as help identify structural or linkage isomers if they can be separated. It would also help identify low abundance glycopeptide glycoforms through targeted selected reaction monitoring (SRM) experiments that would be built around the predicted retention times.

These SRM experiments could also reduce the range of time spent searching for the glycopeptide of interest. Because the hydrophilic glycan will interact more strongly with the hydrophilic stationary phase than the peptide, we assume that the prediction would be weighted more towards the glycan model, but that the models could possibly be added together for glycopeptide retention prediction.

This article details the analysis of the actual retention of glycopeptides from human IgGs on a penta-HILIC column and provides a comparison with predicted retention from the peptide and glycan models we have previously created. The majority of glycopeptide structures agreed with predicted retention times from the combined models as long as an intercept was applied. Only one glycopeptide deviated from the prediction, and we hypothesize that it is due to its bisecting structure.

Materials and Methods

Glycoprotein Separation and Digestion

Human IgGs were separated from human serum (Sigma-Aldrich, St. Louis, MO, USA) using a HiTrap™ Protein G column (General Electric Company, Fairfield, CT, USA). Proteins were reduced using 10 mM dithiothreitol (DTT) and then alkylated using 55 mM iodoacetamide (IDA), both purchased from Sigma Aldrich. Sequencing-grade trypsin or chymotrypsin purchased from Promega (San Luis Obispo, CA, USA) was added at 50:1 (w/w, protein/trypsin) for incubation overnight in 50 mM ammonium bicarbonate (pH 7.0) at 37°C.

LC-MS/MS Settings and Instrumentation

IgG samples were analyzed on a 4000 Q Trap (Sciex, Framingham, MA, USA). Samples were suspended in 25% H₂O, 75% ACN and 0.1% formic acid (Sigma-Aldrich, St. Louis, MO, USA) for direct injection into the LC system. Peptides were separated by a 2.1 mm x 15 cm

HALO® Penta-HILIC column packed with 2.7- μ diameter superficially porous particles that have a 90 Å pore diameter (Advanced Materials Technology, Wilmington, DE, USA) using a Nexera UFLC (Shimadzu, Columbia, MD, USA). The temperature of the column was 60°C. The gradient used for each sample was 22-52% water over 80 minutes at a 0.4-mL/min flow rate. The aqueous solvent contained 0.1% formic acid and 50 mM ammonium formate (Thermo-Fisher, San Jose, CA, USA), and the organic solvent was pure acetonitrile. Spectra were obtained using an ESI source. A SRM method was used to select precursor and fragment masses for both peptides and glycopeptides of interest.

Data were also acquired using a Finnegan LTQ (Thermo-Fisher, San Jose, CA, USA) in series with an 1100 Series Capillary LC system (Agilent Technologies, Palo Alto, CA, USA) with an ESI source that used spray tips made in-house. Peptides were separated using a 200- μ m x 150-mm HALO® Penta-HILIC column packed with 2.7 μ m diameter superficially porous particles at room temperature. The gradient elution conditions employed a linear increase in aqueous solvent from 5-70% over 90 minutes at a 2 μ L/min flow rate. The aqueous solvent contained 0.1% formic acid with 50 mM ammonium formate and the organic solvent was acetonitrile with 0.1% formic acid. The settings for the mass spectrometer included taking the 5 most intense ions from each full mass spectrum for fragmentation using collision-induced dissociation (CID), and the resulting MS/MS spectra were recorded.

Glycopeptide Retention Analysis

Glycopeptide retention times were determined manually using the apex of the peaks displayed in Analyst software (Sciex, Framingham, MA, USA). Peptide retention times in minutes were converted to GU from procainamide-labeled dextran samples that were run before and after the samples so that prediction can be carried out on different LC-MS systems with

different chromatographic conditions. Actual retention times were compared to predicted ones using the peptide and glycan models created in-house.

Results and Discussion

Peptide Retention Model Comparison

Human IgGs 1-4 yield glycopeptides with three different amino acid sequences at the glycosylation site of interest (N297) after trypsin digestion, as IgG2 and IgG3 have the same sequence. The unglycosylated forms of these peptides were identified in our IgG samples after digestion, and their actual retention times were compared with predicted retention times from a previously made model (Table 1). The peptide model is based on amino acid composition and is able to sum amino acid coefficients related to their hydrophilicities with an intercept to predict retention.¹⁸⁵ The retention times are expressed in glucose units (GU) from procainamide-labeled dextran samples that were used as retention time calibrants. This enables the comparison of retention on different LC-MS systems with various chromatographic conditions. The deviations from actual times and predicted times in Table 1 are extremely low, indicating the accuracy of prediction. These peptides only differ by the substitution of a phenylalanine (F) or tyrosine (Y) residue, which have coefficients in our previously made peptide retention prediction model of -0.967 GU and -0.430 GU, respectively.¹⁸⁵ Their negative values indicate that they are hydrophobic, and peptides with these residues will elute earlier on the HILIC column. Substituting one phenylalanine residue for a tyrosine residue (IgG1 to IgG4) would result in a predicted difference of 0.537 GU from the coefficients, and the actual difference was 0.550 GU (a 0.013 GU difference). Substituting two phenylalanine residues for two tyrosine residues (IgG1 to IgG2/3) would result in a predicted difference of 1.074 GU, and the actual difference was 1.202 GU (a 0.128 GU difference). Both of these comparisons demonstrate that the peptide

model is fully capable of accurately predicting the retention times of native peptides that are very similar in composition.

Other peptides from IgG samples were analyzed on UPLC-QTrap and capillary HPLC-LTQ systems to test the accuracy of the peptide retention model using different chromatographic conditions (size of the column, temperature of the column, gradient slope, buffer composition, and flow rate), and the actual retention times from both systems are shown in Table 2. The average deviation from system to system was 0.408 GU (1.64 min.), indicating that the model is capable of providing accurate retention time predictions for peptides ran on completely different systems with many different chromatographic conditions. Furthermore, there was no trend in elution, as almost half (45%) of peptides eluted earlier on the 4000 Q Trap.

Glycopeptide Retention Prediction

Glycosylated forms of the peptides in Table 1 were identified in the IgG samples, and their structures and retention times were analyzed (Table 3). Retention times in yellow are for IgG1, retention times in green are for IgG2/3, and retention times in blue are for IgG4. The number of glycopeptide identifications for each subclass is in direct correlation with their abundances in human serum, as IgG1 has around a 66% abundance, IgG2 and 3 have around a combined 30% abundance, and IgG4 has around a 4% abundance.²³⁰ The N-linked glycans studied herein are comprised of several retention effecting elements, namely N-acetylglucosamine (GlcNAc), mannose (Man), galactose (Gal), and core fucose (Fuc). These chromatographically influencing elements, in combination with the individual influences of the peptide amino acids, affect retention in a reproducible fashion, allowing for the creation of a predictive model. The retention of glycopeptides is determined by the interaction of hydrophilic functional groups of the N-linked glycan and the peptide with the HILIC stationary phase and

water-rich layer, and changes in glycopeptide structure or composition will result in greater or lesser retention on the HILIC column. The resolution of isomeric glycoforms, such as α 2-3- or α 2-6-linked sialic acid species, is possible because of these differences in the degree of interaction between the glycans and the HILIC column, allowing for separate analysis of not only individual glycan species but also of their structural isomers.²³¹

Eight glycopeptides with different glycan structures were analyzed, and their abbreviations are labeled under each structure in Figure 1. The A2 structure, which has two GlcNAc moieties, has the shortest retention, followed by the A3G1 and A2G1 glycopeptides. This shows that the retention increases as the glycan chain is extended because of the hydrophilic monosaccharides. The glycopeptides with bi-antennary structures (A2G2 and F1A2G2) had the largest retention times, with the core-fucosylated bi-antennary structure having the largest retention time overall. Each addition of a core Fuc increased retention by 0.702 GU. Glycopeptides that included a G1 structure had doublets corresponding to the linkage of GalNAc. These linkages can have GalNAc in two types of orientation, leading to different chromatographic interactions that vary by an average of 0.211 GU and can be baseline separated.

To compare the actual retention times of IgG glycoproteins with predicted ones, separate predictions for the peptide composition and the glycan structure were calculated using previously made models and then summed together. IgG1 was individually analyzed because it had the most glycopeptide identifications of all the IgG subclasses. This data is shown in Table 4, and the differences between the predicted sums and actual retention times are shown in the far right column. Almost all of actual retention times of the glycopeptides deviate by about 2 GU from predicted times, with an average deviation of 1.881 GU. In addition, all of the predicted times were lower than the actual times. It was expected that there should be some deviation from actual

to predicted times due to the glycan changing the interactions of adjacent amino acid residues with the stationary phase, but it is encouraging that most of the predictions were consistently offset by the same amount. Because of this, a simple addition of 1.881 GUs to predicted retention times of glycopeptides would enable far more accurate prediction than without the addition, and the average deviation from predicted to actual retention times would drop to 0.368 GU, and only 0.200 GU if the A3G1 structures are not included. The A3G1 glycopeptide was an outlier, as it deviated much less than the other structures.

We hypothesize that the A3G1 glycopeptide deviated less than the others because of its bisecting structure, which is shown in Figure 1. The different retention times for the A3G1 structure depend on the linkage of GalNAc, which is in agreement with the other G1-containing structures. The GalNAc here could also be on the bisecting GlcNAc instead of just the two antenna GlcNAc moieties, however only two peaks were identified for this structure and the retention time difference between isoforms is equal to the other G1 structures. If the GalNAc were on the bisecting GlcNAc, it would be less retained than if it were on an antenna GlcNAc because the interaction with the stationary phase would be hindered by the surrounding monosaccharides. In terms of the retention time deviation of this structure from the other glycopeptides, the bisecting GlcNAc moiety could be shielded from stationary phase interaction by the terminal GalNAc or the other monosaccharide subunits, making the glycopeptide elute earlier than predicted. The reason we think this is because the other bisecting glycopeptide, F1A3, does not have the terminal GalNAc and it agrees with the 1.881 GU deviation, albeit on the lower end. However, due to the limited availability of bisecting glycans, the glycan prediction model did not have many instances in its dataset. Using an additional coefficient of 0.70 GU for bisecting glycopeptides would allow the A3G1 glycopeptide structure to be much

closer to the 1.881 GU intercept (1.489 GU and 1.656 GU), and would enable even more accurate predictions. The F1A3 structure would be on the higher end of deviating glycopeptide structures at 2.390 GU when using the additional coefficient, but would not differ much more than what it previously was (an increase in difference of 0.080 GU from the 1.881 GU intercept).

Summary

The ability to sum the predictions from the peptide model and the glycan model demonstrate the ease of predicting glycopeptide retention as long as the 1.881 GU intercept is applied. Even though this study was only done on glycopeptides from IgG samples, this provides a significant impact on the analysis of glycopeptides in general, as it can further the identification process, help in isomeric identification, and aid in SRM experiments. The 1.881 GU intercept is intriguing, as it shows that the interactions with the hydrophilic stationary phase are higher in glycopeptides than in native peptides and glycans combined together. The structure of the glycan on the peptide was also shown to be important, as it was found that the A3G1 bisecting structure needed an additional coefficient to accurately predict retention due to possible shielding effects.

Acknowledgements

Support for this work comes from NIH Grant GM0 93747 to BEB.

Table 7.1**Actual retention times of native peptides compared to predicted retention times**

Source	Peptide	Mass	Actual RT (min)	Actual RT (GU)	Predicted RT (GU)	Deviation (GU)
IgG1	EEQYNSTYR	1189.5	54.06	5.888	5.946	0.058
IgG2/3	EEQFNSTFR	1173.5	48.82	4.687	4.872	0.185
IgG4	EEQFNSTYR	1157.5	51.81	5.339	5.409	0.070

Table 7.2

A comparison of IgG peptide retention times on the LTQ and 4000 Q Trap systems.

Peptides in red eluted earlier on the 4000 Q Trap than the LTQ (10 out of 22).

Peptide Sequence	4000 Q Trap RT (GU)	LTQ RT (GU)	Deviation
ADYEK	6.348	6.381	-0.033
DSTYLSSTLTLSK	4.422	4.120	0.306
DELTK	5.656	5.251	0.404
ALPAPIEK	4.098	3.282	0.815
EPQVYTLPPSR	4.668	4.679	-0.011
FNWYVDGVEVHNAK	5.247	5.225	0.021
GLPAPIEK	4.304	3.474	0.830
LTVLGQPK	3.120	2.137	0.982
TVAPECS	4.100	3.793	0.306
AGVETTPSK	5.467	5.968	-0.502
YAASSYLSLTPEQWK	3.131	3.227	-0.096
GPSVFPLAPSSK	3.747	3.177	0.570
STSGGTAALGCLVK	4.134	4.440	-0.306
VYACEVTHQGLSPVTK	5.438	6.226	-0.787
TVAAPSVFIFPPSDEQLK	4.565	4.108	0.458
RTVAAPSVFIFPPSDEQLK	5.428	5.996	-0.568
GPSVFPLAPCSR	3.597	2.776	0.821
STSESTAALGCLVK	4.914	5.042	-0.128
TPPVLDSGDSFFLYSK	3.704	3.873	-0.169
AAPSVTLFPPSSEELQANK	6.113	5.730	0.383
ANPTVTLFPPSSEELQANK	5.293	5.443	-0.150
EVQLVESGGGLVQPGGSLR	5.475	5.151	0.323

Table 7.3

IgG glycopeptide retention times. Boxes that contain “X” indicate that no glycopeptide was identified.

Glycan Structure	IgG1 Actual RT (min)	IgG1 Actual RT (GU)	IgG2/3 Actual RT (min)	IgG2/3 Actual RT (GU)	IgG4 Actual RT (min)	IgG4 Actual RT (GU)
A2	48.53	13.686	X	X	X	X
F1A2	49.90	14.378	47.49	13.183	48.68	13.760
A2G1	50.22	14.545	X	X	X	X
A2G1	50.70	14.799	X	X	X	X
F1A2G1	51.56	15.265	49.20	14.020	X	X
F1A2G1	51.94	15.475	49.61	14.229	X	X
A2G2	52.19	15.615	X	X	X	X
F1A3	50.90	14.906	48.45	13.647	X	X
F1A2G2	53.44	16.334	51.20	15.068	52.31	15.683
A3G1	49.07	13.955	46.56	12.748	47.84	13.350
A3G1	49.40	14.122	46.96	12.933	48.30	13.573

Table 7.4**Predicted and actual retention times of glycopeptides identified in IgG1**

Glycan	Predicted Glycan Retention (GU)	Predicted Peptide Retention (GU)	Sum of Predictions (GU)	Actual RT (GU)	Difference Between Actual and Predicted Sum (GU)
A2	5.46	5.946	11.406	13.686	2.280
F1A2	6.46	5.946	12.406	14.378	1.973
A2G1	6.41	5.946	12.356	14.545	2.189
A2G1	6.41	5.946	12.356	14.799	2.443
F1A2G1	7.41	5.946	13.356	15.265	1.909
F1A2G1	7.41	5.946	13.356	15.475	2.119
BI	7.36	5.946	13.306	15.615	2.309
F1A3	7.27	5.946	13.216	14.906	1.690
F1BI	8.36	5.946	14.306	16.334	2.028
A3G1	7.22	5.946	13.166	13.955	0.789
A3G1	7.22	5.946	13.166	14.123	0.956

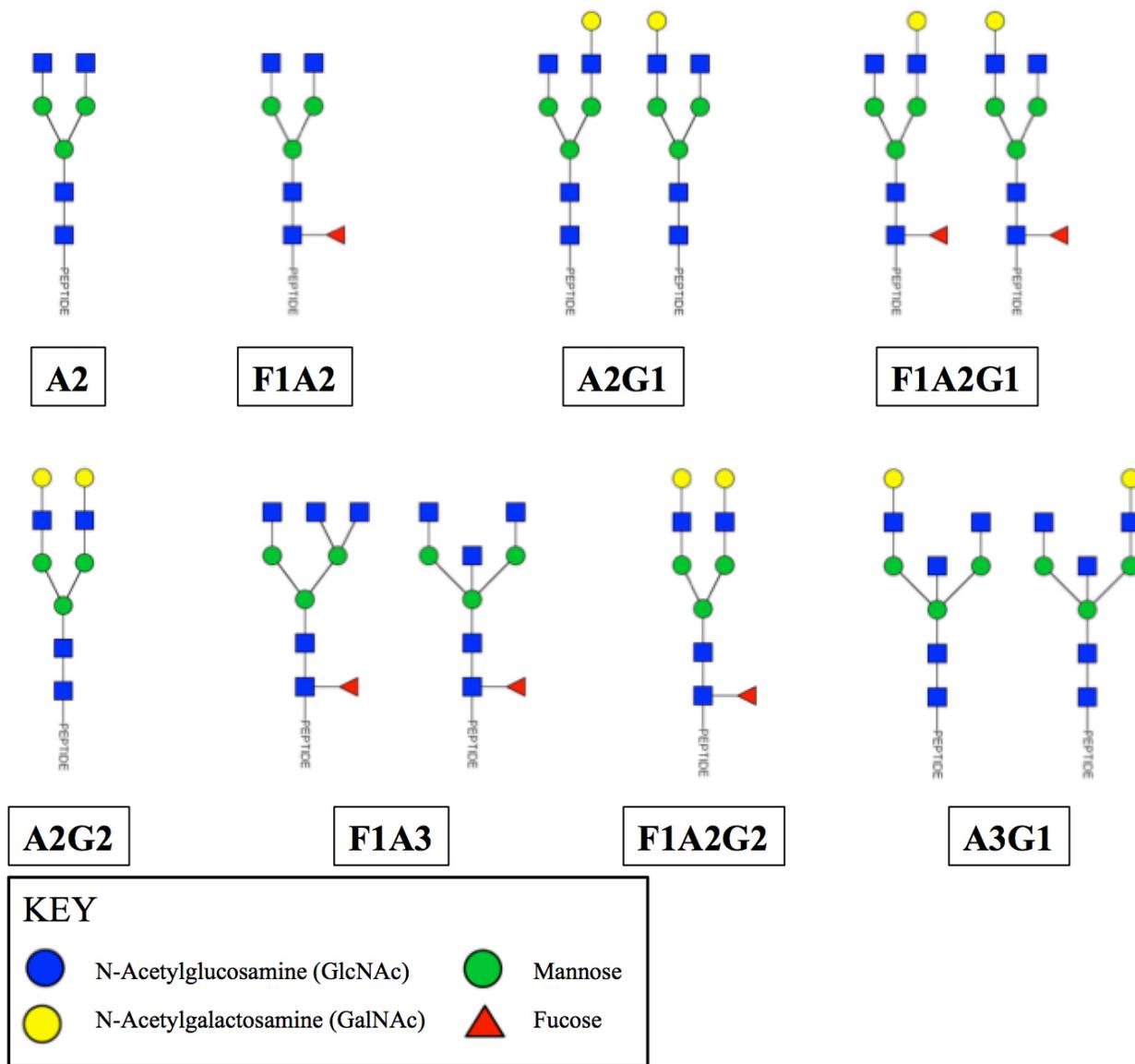


Figure 7.1: Glycan structures analyzed in the glycopeptide retention prediction model.

Each structure with a “G1” can have two possible linkages of Gal, and both isoforms are shown.

CHAPTER 8

CONCLUSIONS

The two main purposes of this work are first, to analyze the separation of peptides and hydrophilic post-translational modifications using HILIC, and second, to facilitate faster and more confident identifications by deriving coefficients that can be summed together to predict the retention of native and modified peptides. HILIC was shown of being fully capable of baseline separating unmodified peptides from their modified forms in a predictable manner.

Chapter 3 describes the creation and validation of a highly accurate HILIC model that predicts peptide retention based on amino acid composition. Coefficients describing the overall hydrophilicity of the amino acids were derived, and can be summed together with an intercept to predict the retention of peptides with any sequence. It was found that the size of the peptide had an effect on retention, as peptides longer than 15 amino acids in length deviated from predicted retention times more than shorter peptides. It was also found that location can affect peptide retention for hydrophobic residues directly at or one residue from the C-terminus, and optimized coefficients were created to account for both of these positions. Finally, dextran was shown to be a suitable retention time calibrant, as peptides run on two completely different LC-MS systems with different chromatographic conditions were within 0.52 GU (2.29 minutes) of each other. This model can help in numerous regards including decreasing the time spent in identification, increasing the confidence in identifications, allowing for isomeric identifications, or reducing the time spent searching for an analyte of interest in SRM techniques.

In Chapter 4, two hydrophilic modifications were analyzed using a HILIC column: the deamidation of asparagine and the oxidation of methionine. Both of these modifications have serious implications in the biotherapeutics realm, and thus are important to be able to fully separate out from their unmodified forms to allow for quantitation. In comparison to RP chromatography, HILIC was found to be able to baseline separate peptides with both modifications from their native forms in a predictable fashion. Coefficients describing their hydrophilicities were derived and incorporated into the model created in Chapter 3. Deamidation of asparagine resulted in one of the most hydrophilic coefficients overall, whereas oxidation of methionine resulted in a moderately hydrophilic coefficient. Because there can be two different deamidation products, *n*-Asp and isoAsp, synthetic peptides with no modification, *n*-Asp modification, and isoAsp modification were analyzed, and it was discovered that the derived deamidation coefficient corresponded to the formation of isoAsp, which was 3-4 times more abundant than *n*-Asp in proteomic samples.

In addition to the deamidation of asparagine, the isomerization of aspartic acid can also produce isoAsp, and this was examined in Chapter 5. This modification is difficult to identify simply using a mass spectrometer without ExD fragmentation techniques, so the ability to separate out the two forms before MS analysis is important. Using the predictive model discussed in Chapter 3, the retention times of peptides with aspartic acid residues were predicted and compared to actual retention times to see if they correlated with *n*-Asp or the formation for isoAsp. What was found was that the adjacent amino acid residue on the C-terminal side of the aspartic acid had a tremendous affect on the isomerization of aspartic acid, and peptides with “DG”, “DE”, “DS”, and “DL” sequences had trends that suggested that the peptides were readily

isomerizing to isoAsp. Furthermore, chromatograms containing these peptides only showed one peak instead of a mixture of *n*-Asp and isoAsp.

Chapter 6 explored the separation of three different types of O-linked glycosylation: O-GalNAcylation, O-GlcNAcylation, and O-fucosylation. All three forms of glycosylation were shown to be extremely hydrophilic, and coefficients relating to their overall hydrophilicity were derived and incorporated into the model discussed in Chapter 3. In addition, peptides with all three types of modification were shown to be baseline separated from their native counterparts, allowing for quantitation. O-GalNAc and O-GlcNAc had very similar coefficient values that were among the most hydrophilic overall. O-fucose had a slightly lower value, but was still on the higher end of the amino acid coefficients from the peptide model.

Finally, Chapter 7 examined the other type of glycosylation, N-linked, on human IgGs. The peptide model discussed in Chapter 3 was utilized to predict the retention of a native tryptic peptide at N297 with slightly different sequences between IgG variants, and was found to be extremely accurate on a LC-MS system that was completely different than the one used to make the model. This peptide model was combined with another model made in our laboratory that predicts the retention of glycans to be able to predict the retention of glycopeptides. It was found that the predicted retention times for glycopeptides from human IgGs deviated for the most part by 1.881 GU. The A3G1 structure was the only glycopeptide that did not fit this trend, and this was due to its bisected structure. However, using the 1.881 GU intercept, the retention of the majority of glycopeptides can accurately predicted. The ability to simply sum the peptide model, glycan model, and 1.881 GU intercept together for glycopeptide retention prediction has the potential to greatly facilitate glycopeptide identification.

REFERENCES

1. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature*. 2003;422(6928):198-207.
2. Domon B, Aebersold R. Mass spectrometry and protein analysis. *Science*. 2006;312(5771):212-217.
3. Wuhrer M. Glycomics using mass spectrometry. *Glycoconj J*. 2013;30(1):11-22.
4. Zaia J. Mass Spectrometry and the emerging field of glycomics. *Chem Biol*. 2008;15(9):881-892.
5. Ali I, Aboul-Enein HY, Singh P, Singh R, Sharma B. Separation of biological proteins by liquid chromatography. *Saudi Pharm J*. 2010;18(2):59-73.
6. Mant CT, Zhou NE, Hodges RS. Correlation of protein retention times in reversed-phase chromatography with polypeptide chain length and hydrophobicity. *J Chromatogr A*. 1989;476:363-375.
7. Mitulovic G, Mechtler K. HPLC techniques for proteomics analysis--a short overview of latest developments. *Brief Funct Genomic Proteomic*. 2006;5(4):249-260.
8. Gilar M, Jaworski A. Retention behavior of peptides in hydrophilic-interaction chromatography. *J Chromatogr A*. 2011;1218(49):8890-8896.
9. Wang Y, Lehmann R, Lu X, Zhao X, Xu G. Novel, fully automatic hydrophilic interaction/reversed-phase column-switching high-performance liquid chromatographic system for the complementary analysis of polar and apolar compounds in complex samples. *J Chromatogr A*. 2008;1204(1):28-34.
10. Krokhin OV, Craig R, Spicer V, Ens W, Standing KG, Beavis RC, Wilkins JA. An improved model for prediction of retention times of tryptic peptides in ion pair reversed-phase HPLC: its application to protein peptide mapping by off-line HPLC-MALDI MS. *Mol Cell Proteomics*. 2004;3(9):908-919.
11. Mant CT, Burke TWL, Black JA, Hodges RS. Effect of peptide chain length on peptide retention behaviour in reversed-phase chromatography. *J Chromatogr A*. 1988;458:193-205.
12. Meek JL, Rossetti ZL. Factors affecting retention and resolution of peptides in high-performance liquid chromatography. *J Chromatogr A*. 1981;211(1):15-28.

13. Gervais D. Protein deamidation in biopharmaceutical manufacture: Understanding, control and impact. *J Chem Technol Biotechnol*. 2016;91(3):569-575.
14. Jenkins N, Murphy L, Tyther R. Post-translational modifications of recombinant proteins: Significance for biopharmaceuticals. *Mol Biotechnol*. 2008;39(2):113-118.
15. Bischoff R, Kolbe HVJ. Deamidation of asparagine and glutamine residues in proteins and peptides: structural determinants and analytical methodology. *J Chromatogr B Biomed Sci Appl*. 1994;662(2):261-278.
16. Geiger T, Clarke S. Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. Succinimide-linked reactions that contribute to protein degradation. *J Biol Chem*. 1987;262(2):785-794.
17. Roher AE, Esh CL, Kokjohn TA, Castano EM, Van Vickle GD, Kalback WM, Patton RL, Luehrs DC, Dausg ID, Kuo YM, Emmerling MR, Soares H, Quinn JF, Kaye J, Connor DJ, Silverberg NB, Adler CH, Seward JD, Beach TG, Sabbagh MN. AB peptides in human plasma and tissues and their significance for Alzheimer's disease. *Alzheimer's Dement*. 2009;5(1):18-29.
18. Roher AE, Lowenson JD, Clarke S, Wolkow C, Wang R, Cotter RJ, Reardon IM, Zurcher-Neely HA, Heinrichson RL, Ball MJ, Greenberg BD. Structural alteration in the peptide backbone of b-amyloid core protein may account for its deposition and stability in Alzheimer's disease. *J Biol Chem*. 1993;268(5):3072-3083.
19. Robinson NE, Robinson AB. Molecular clocks. *Proc Natl Acad Sci USA*. 2001;98(3):944-949.
20. Robinson NE, Robinson ZW, Robinson BR, Robinson AL, Robinson JA, Robinson ML, Robinson AB. Structure-dependent nonenzymatic deamidation of glutaminyl and asparaginyl pentapeptides. *J Pept Res*. 2004;63(5):426-436.
21. Shimizu T, Matsuoka Y, Shirasawa T. D -amino acid biosystem biological significance of isoaspartate and its repair system. *Biol Pharm Bull*. 2005;28(9):1590-1596.
22. Gross JH. *Mass Spectrometry A Textbook*. 2nd ed. Heidelberg: Springer-Verlag; 2011.
23. Banerjee S, Mazumdar S. Electrospray ionization mass spectrometry: A technique to access the information beyond the molecular weight of the analyte. *Int J Anal Chem*. 2012;2012:1-40.
24. De Hoffmann E. Tandem mass spectrometry - a primer. *J Mass Spectrometry*. 1996;31:129-137.
25. Kim MS, Pandey A. Electron transfer dissociation mass spectrometry in proteomics. *Proteomics*. 2012;12:530-542.

26. Kim S, Mischerikow N, Bandeira N, Navarro JD, Wich L, Mohammed S, Heck AJR, Pevzner PA. The generating function of CID, ETD, and CID/ETD pairs of tandem mass spectra: applications to database search. *Mol Cell Proteomics*. 2010;9(12):2840-2852.
27. Johnson AR, Carlson EE. Collision-induced dissociation mass spectrometry: A powerful tool for natural product structure elucidation. *Anal Chem*. 2015;87(21):10668-10678.
28. Khatun J, Ramkissoon K, Giddings MC. Fragmentation characteristics of collision-induced dissociation in MALDI TOF/TOF mass spectrometry. *Anal Chem*. 2007;79(8):3032-3040.
29. Cardoza JD, Parikh JR, Ficarro SB, Marto JA. Mass spectrometry-based proteomics: Qualitative identification to activity-based protein profiling. *Wiley Interdiscip Rev Syst Biol Med*. 2012;4(2):141-162.
30. Hart-Smith G. A review of electron-capture and electron-transfer dissociation tandem mass spectrometry in polymer chemistry. *Anal Chim Acta*. 2014;808:44-55.
31. Zubarev R, Kelleher NL, McLafferty FW. Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J Am Chem Soc*. 1998;120(16):3265-3266.
32. Chowdhury SK, Katta V, Chait BT. Electrospray ionization mass spectrometric peptide mapping: a rapid, sensitive technique for protein structure analysis. *Biochem Biophys Res Commun*. 1990;167(2):686-692.
33. Aebersold R, Goodlett DR. Mass spectrometry in proteomics. *Chem Rev*. 2001;101:269-295.
34. Bakhtiar R, Tse FL. Biological mass spectrometry: a primer. *Mutagenesis*. 2000;15(5):415-430.
35. Dole M, Mack LL, Hines RL, et al. Molecular Beams of Macroions. *J Chem Phys*. 1968;49(5):2240-2249.
36. Kebarle P, Verkerk UH. Electrospray: From ions in solution to ions in the gas phase, what we know now. *Mass Spectrom Rev*. 2009;28:898-917.
37. Wilm M. Principles of electrospray ionization. *Mol Cell Proteomics*. 2011;10(7):1-8.
38. Iribarne J V, Thomson BA. On the evaporation of small ions from charged droplets. *J Chem Phys*. 1976;64(6):2287-2294.
39. Nguyen S, Fenn JB. Gas-phase ions of solute species from charged droplets of solutions. *Proc Natl Acad Sci USA*. 2007;104(4):1111-1117.

40. Singhal N, Kumar M, Kanaujia PK, Viridi JS. MALDI-TOF mass spectrometry: An emerging technology for microbial identification and diagnosis. *Front Microbiol.* 2015;6:1-16.
41. Everley RA, Mott TM, Wyatt SA, Toney DM, Croley TR. Liquid chromatography/mass spectrometry characterization of escherichia coli and shigella species. *J Am Soc Mass Spectrom.* 2008;19(11):1621-1628.
42. Gabelica V, Schulz E, Karas M. Internal energy build-up in matrix-assisted laser desorption/ionization. *J Mass Spectrom.* 2004;39(6):579-593.
43. Hazama H, Nagao H, Suzuki R, Toyoda M, Masuda K, Naito Y, Awazu K. Comparison of mass spectra of peptides in different matrices using matrix assisted laser desorption/ionization and a multi-turn time-of-flight mass spectrometer, MULTUM-IMG. *Rapid Commun Mass Spectrom.* 2008;22:1461-1466.
44. Stimson E, Truong O, Richter WJ, Waterfield MD, Burlingame AL. Enhancement of charge remote fragmentation in protonated peptides by high-energy CID MALDI-TOF-MS using “cold” matrices. *Int J Mass Spectrom Ion Process.* 1997;169-170:231-240.
45. Signor L, Boeri Erba E. Matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) mass spectrometric analysis of intact proteins larger than 100 kDa. *J Vis Exp.* 2013;(79):1-7.
46. Dawson P. Quadrupole mass analysers: Performance, design and some recent applications. *Mass Spectrom Rev.* 1986;5:1-37.
47. Dawson PH. *Quadrupole Mass Spectrometry and Its Applications.* 1st ed. Amsterdam, Netherlands: Elsevier B.V.; 1976.
48. Johnson JV, Yost RA, Kelley PE, Bradford DC. Tandem-in-space and tandem-in-time mass spectrometry: triple quadrupoles and quadrupole ion traps. *AnalChem.* 1990;62(20):2162-2172.
49. March RE. Quadrupole ion trap mass spectrometry: a view at the turn of the century. *Int J Mass Spectrom.* 2000;200:285-312.
50. March RE. Quadrupole ion traps. *Mass Spectrom Rev.* 2009;28(6):961-989.
51. Guilhaus M, Mlynski V, Selby D. Perfect timing: Time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom.* 1997;11(9):951-962.
52. Chernushevich IV, Loboda AV, Thomson BA. An introduction to quadrupole-time-of-flight mass spectrometry. *J Mass Spectrom.* 2001;36(8):849-865.

53. Fuerstenau SD, Benner WH. Molecular weight determination of megadalton DNA electrospray ions using charge detection time-of flight mass spectrometry. *Rapid Commun Mass Spectrom.* 1995;9(15):1528-1538.
54. Schueler BW. Microscope imaging by time-of-flight secondary ion mass spectrometry. *Microsc Microanal Microstruct.* 1992;3:119-139.
55. Vestal ML. Modern MALDI time-of-flight mass spectrometry. *J Mass Spectrom.* 2009;44(3):303-317.
56. Cotter RJ. Time-of-flight mass spectrometry for the structural analysis of biological molecules. *Anal Chem.* 1992;64(21):1027-1039.
57. Marshall AG, Hendrickson CL. Fourier transform ion cyclotron resonance detection: Principles and experimental configurations. *Int J Mass Spectrom.* 2002;215:59-75.
58. Nikolaev EN, Kostyukevich YI, Vladimirov GN. Fourier transform ion cyclotron resonance (FT ICR) mass spectrometry: Theory and simulations. *Mass Spectrom Rev.* 2016;35:219-258.
59. Schmid DG, Grosche P, Bandel H, Jung G. FTICR-mass spectrometry for high-resolution analysis in combinatorial chemistry. *Biotechnol Bioeng.* 2000;71(2):149-161.
60. Eliuk S, Makarov A. Evolution of orbitrap mass spectrometry instrumentation. *Annu Rev Anal Chem.* 2015;8(1):61-80.
61. Perry RH, Cooks G, Noll RJ. Orbitrap mass spectrometry: Instrumentation, ion motion and applications. *Mass Spectrom Rev.* 2008;27:661-699.
62. Bhardwaj SK, Dwivedi K, Agarwal DD. A review: HPLC method development and validation. *Int J Anal Bioanal Chem.* 2015;5(4):76-81.
63. Bensaddek D, Nicolas A, Lamond AI. Evaluating the use of HILIC in large-scale, multi dimensional proteomics: Horses for courses? *Int J Mass Spectrom.* 2015;391:105-114.
64. Boersema PJ, Mohammed S, Heck AJR. Hydrophilic interaction liquid chromatography (HILIC) in proteomics. 2008:151-159.
65. Yoshida T. Prediction of peptide retention time in normal-phase liquid chromatography. *J Chromatogr A.* 1998;811:61-67.
66. Van Deemter JJ, Zuiderweg FJ. Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography. *Chem Engng Sci.* 1956;5:271-289.
67. Snyder LR, Kirkland JJ, Dolan JW. *Introduction to Modern Liquid Chromatography.* 3rd ed. Hoboken, NJ: Wiley-Blackwell; 2009.

68. Salisbury JJ. Fused-core particles : A practical alternative to sub-2 micron particles. 2008;46:883-886.
69. Kirkland JJ, Schuster SA, Johnson WL, Boyes BE. Fused-core particle technology in high-performance liquid chromatography : An overview. *J Pharm Anal.* 2013;3(5):303-312.
70. Schuster SA, Boyes BE, Wagner BM, Kirkland JJ. Fast high performance liquid chromatography separations for proteomic applications using fused-core silica particles. 2012;1228:232-241.
71. Knox JH, Scott HP. B and C terms in the van deemter equation for liquid chromatography. *J Chromatogr.* 1983;282:297-313.
72. Neue UD. Theory of peak capacity in gradient elution. *J Chromatogr A.* 2005;1079:153-161.
73. Paliwal SK, de Frutos M, Regnier FE. Rapid separations of proteins by liquid chromatography. *Methods Enzymol.* 1996;270:133-151.
74. Naushad M, Khan MR. *Ultra Performance Mass Spectrometry Liquid Chromatography: Evaluation and Application in Food Analysis.* Boca Raton, FL: CRC Press/Taylor & Francis; 2014.
75. Carlin MG, Dean JR. *Forensic Applications of Gas Chromatography.* Boca Raton, FL: CRC Press/Taylor & Francis; 2013.
76. Teutenberg T. *High-Temperature Liquid Chromatography: A User's Guide for Method Development.* Cambridge UK: RSC Publishing; 2010.
77. Aguilar M. Reversed-phase high-performance liquid chromatography. In *HPLC of Peptides and Proteins.* 2004;251:9-22.
78. Horvath C, Molnar I. Separation of amino acids and peptides on non-polar stationary phases by high-performance liquid chromatography. *J Chromatogr A.* 1977;142:623-640.
79. Dorsey JG, Dill KA. The molecular mechanism of retention in reversed-phase liquid chromatography. *Chem Rev.* 1989;89:331-346.
80. Rafferty JL, Zhang L, Siepmann JI, Schure MR, Carlo M. Retention mechanism in reversed-phase liquid chromatography : A molecular perspective. *Anal Chem.* 2007;79:6551-6558.
81. Prathap B, Dey A, Srinivasa GH, Johnson P, Arthanariswaran P. A review - Importance of RP-HPLC in analytical method development. *Int J Nov Trends Pharm Sci.* 2013;3(1):15-23.

82. Kohler BI, Periat A, Guillarme D, Veuthey J, Issues S. Advances in hydrophilic interaction liquid chromatography for pharmaceutical analysis. *LCGC*. 2013;1-7.
83. Buszewski B, Noga S. Hydrophilic interaction liquid chromatography (HILIC) - a powerful separation technique. *Anal Bioanal Chem*. 2012;402:231-247.
84. Heaton J, Smith NW. Advantages and disadvantages of HILIC; A brief overview. *Chrom Today*. 2012;44-47.
85. Jandera P. Stationary and mobile phases in hydrophilic interaction chromatography: A review. *Analytica Chimica Acta*. 2011;692:1-25.
86. Qiao L, Dou A, Shi X, Li H, Shan Y, Lu X, Xu G. Development and evaluation of new imidazolium-based zwitterionic stationary phases for hydrophilic interaction chromatography. *J Chromatogr A*. 2013;1286:137-145.
87. Hemström P, Irgum K. Hydrophilic interaction chromatography. *J Sep Sci*. 2006;29(12):1784-1821.
88. McCalley DV, Neue UD. Estimation of the extent of the water-rich layer associated with the silica surface in hydrophilic interaction chromatography. *J Chromatogr A*. 2008;1192(2):225-229.
89. Browne CA, Bennett HPJ, Solomon S. The isolation of peptides by high-performance liquid chromatography using predicted elution positions. *Anal Biochem*. 1982;124:201-208.
90. Guo D, Mant CT, Taneja AK, Parker JMR, Rodges RS. Prediction of peptide retention times in reversed-phase high-performance liquid chromatography I. Determination of retention coefficients of amino acid residues of model synthetic peptides. *J Chromatogr A*. 1986;359:499-518.
91. Krokhin OV, Spicer V. Predicting peptide retention times for proteomics. *Curr Protoc Bioinformatics*. 2010;1-15.
92. Le Maux S, Nongonierma AB, Fitzgerald RJ. Improved short peptide identification using HILIC-MS/MS: Retention time prediction model based on the impact of amino acid position in the peptide sequence. *Food Chem*. 2015;173:847-854.
93. Meek JL. Prediction of peptide retention times in high-pressure liquid chromatography on the basis of amino acid composition. *Proc Natl Acad Sci USA*. 1980;77(3):1632-1636.
94. O'Hare MJ, Nice EC. Hydrophobic high-performance liquid chromatography of hormonal polypeptides and proteins on alkylsilane bonded silica. *J Chromatogr A*. 1979;171:209-226.

95. Palmblad M, Ramström M, Markides KE, Håkansson P, Bergquist J. Prediction of chromatographic retention and protein identification in liquid chromatography/mass spectrometry. *Anal Chem.* 2002;74(22):5826-5830.
96. Parker JMR, Guo D, Hodges RS. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry.* 1986;25(19):5425-5432.
97. Sakamoto Y, Kawakami N, Sasagawa T. Prediction of peptide retention times. *J Chromatogr A.* 1988;442:69-79.
98. Tripet B, Cepeniene D, Kovacs JM, Mant CT, Krokhin O V., Hodges RS. Requirements for prediction of peptide retention time in reversed-phase high-performance liquid chromatography: Hydrophilicity/hydrophobicity of side-chains at the N- and C-termini of peptides are dramatically affected by the end-groups and location. *J Chromatogr A.* 2007;1141(2):212-225.
99. Tyteca E, Périat A, Rudaz S, Desmet G, Guillaume D. Retention modeling and method development in hydrophilic interaction chromatography. *J Chromatogr A.* 2014;1337:116-127.
100. Wilson KJ, Honegger A, Stötzl RP, Hughes GJ. The behaviour of peptides on reverse-phase supports during high-pressure liquid chromatography. *Biochem J.* 1981;199(1):31-41.
101. Yoshida T. Calculation of peptide retention coefficients in normal-phase liquid chromatography. *J Chromatogr A.* 1998;808:105-112.
102. Pasa-Tolic L, Masselon C, Barry RC, Shen Y, Smith RD. Proteomic analyses using an accurate mass and time tag strategy. *Biotechniques.* 2004;37(4):621-636.
103. Strittmatter EF, Ferguson PL, Tang K, Smith RD. Proteome analyses using accurate mass and elution time peptide tags with capillary LC time-of-flight mass spectrometry. *J Am Soc Mass Spectrom.* 2003;14(9):980-991.
104. Tarasova IA, Guryča V, Pridatchenko ML, Gorshkov AV, Kieffer-Jaquinod S, Evreinov VV, Masselon CD, Gorshkov MV. Standardization of retention time data for AMT tag proteomics database generation. *J Chromatogr B Anal Technol Biomed Life Sci.* 2009;877(4):433-440.
105. Chabanet C, Yvon M. Prediction of peptide retention time in reversed phased high-performance liquid chromatography. *J Chromatogr.* 1992;157:738-742.
106. Babushok VI, Zenkevich IG. Retention characteristics of peptides in RP-LC: Peptide retention prediction. *Chromatographia.* 2010;(9):781-797.

107. Chen Y, Mant CT, Hodges RS. Temperature selectivity effects in reversed-phase liquid chromatography due to conformation differences between helical and non-helical peptides. *J Chromatogr A*. 2003;1010:45-61.
108. Houghten RA, Degraw ST. Effect of positional environmental domains on the variation of high-performance liquid chromatographic peptide retention coefficients. *J Chromatogr*. 1987;386:223-228.
109. Harscoat-Schiavo C, Nioi C. Hydrophilic properties as a new contribution for computer-aided identification of short peptides in complex mixtures. *Anal Bioanal Chem*. 2012;403:1939-1949.
110. Khawli LA, Goswami S, Hutchinson R, Kwong ZW, Yang J, Wang X, Yao Z, Sreedhara A, Cano T, Tesar D, Nijem I, Allison DE, Wong PY, Kao YH, Quan C, Joshi A, Harris RJ, Motchnik P. Charge variants in IgG1: Isolation, characterization, in vitro binding properties and pharmacokinetics in rats. *MAbs*. 2010;2(6):613-624.
111. Shortreed MR, Wenger CD, Frey BL, Sheynkman GM, Scalf M, Keller MP, Attie AD, Smith LM. Global identification of protein post-translational modifications in a single-pass database search. *J Proteome Res*. 2015;14(11):4714-4720.
112. Tsur D, Tanner S, Zandi E, Bafna V. Identification of post-translational modifications by blind search of mass spectra. *Nat Biotechnol*. 2005;23(12):1562-1567.
113. Charache S, Fox J, McCurdy P, Kazazian Jr. H, Winslow R. Postsynthetic deamidation of hemoglobin providence (B82 Lys - Asn, Asp) and its effect on oxygen transport. *J Clin Invest*. 1977;59(4):652-658.
114. Dunkelberger EB, Buchanan LE, Marek P, Cao P, Raleigh DP, Zanni MT. Deamidation accelerates amyloid formation and alters amylin fiber structure. *J Am Chem Soc*. 2012;134:12658-12667.
115. Hao P, Qian J, Dutta B, Cheow ESH, Sim KH, Meng W, Adav SS, Alpert A, Sze SK. Enhanced separation and characterization of deamidated peptides with RP-ERLIC-based multidimensional chromatography coupled with tandem mass spectrometry. *J Proteome Res*. 2012;11(3):1804-1811.
116. Lewis UJ, Singh RNP, Bonewald LF, Seavey BK. Altered proteolytic cleavage of human growth hormone as a result of deamidation. *J Biol Chem*. 1981;256(22):11645-11650.
117. Catak S, Monard G, Aviyente V, Ruiz-López MF. Deamidation of asparagine residues: Direct hydrolysis versus succinimide-mediated deamidation mechanisms. *J Phys Chem A*. 2009;113(6):1111-1120.

118. Tyler-Cross R, Schirch V. Effects of amino acid sequence, buffers, and ionic strength on the rate and mechanism of deamidation of asparagine residues in small peptides. *J Biol Chem*. 1991;266(33):22549-22556.
119. Capasso S, Mazarella S. Kinetics and mechanism of succinimide ring formation in asparagine deamidation. *J Chem Soc Perkin Trans*. 1993;679-682.
120. Capasso S, Salvadori S. Effect of the three-dimensional structure on the deamidation reaction of ribonuclease A. *J Pept Res*. 1999;54(5):377-382.
121. Catak S, Aviyente V, Ruiz-lo MF, Poincare H. Reaction mechanism of deamidation of asparaginyl residues in peptides: Effect of solvent. *J Phys Chem A*. 2006;12:8354-8365.
122. Robinson NE. Protein deamidation. *Proc Natl Acad Sci USA*. 2002;99(8):5283-5288.
123. Li B, Borchardt RT, Topp EM, Vandervelde D, Schowen RL. Racemization of an asparagine residue during peptide deamidation. *J Am Chem Soc*. 2003;125:11486-11487.
124. Badgett MJ, Boyes B, Orlando R. The separation and quantitation of peptides with and without oxidation of methionine and deamidation of asparagine using hydrophilic interaction liquid chromatography with mass spectrometry (HILIC-MS). *J Am Soc Mass Spectrom*. 2017:1-9.
125. Ni W, Dai S, Karger BL, Zhou ZS. Analysis of isoaspartic acid by selective proteolysis with asp-N and electron transfer dissociation mass spectrometry. *Anal Chem*. 2010;82(17):7485-7491.
126. Sargaeva NP, Lin C, O'Connor PB. Differentiating n-terminal aspartic and isoaspartic acid residues in peptides. *Anal Chem*. 2011;83(17):6675-6682.
127. Eakin CM, Miller A, Kerr J, Kung J, Wallace A. Assessing analytical methods to monitor isoAsp formation in monoclonal antibodies. *Front Pharmacol*. 2014;5:1-9.
128. Costa V, Quintanilha A, Moradas-Ferreira P. Protein oxidation, repair mechanisms and proteolysis in *saccharomyces cerevisiae*. *Life*. 2007;59:293-298.
129. Stadtman ER, Van Remmen H, Richardson A, Wehr NB, Levine RL. Methionine oxidation and aging. *Biochim Biophys Acta - Proteins Proteomics*. 2005;1703(2):135-140.
130. Vogt W. Oxidation of methionyl residues in proteins: Tools, targets, and reversal. *Free Radic Biol Med*. 1995;18(1):93-105.
131. Verrastro I, Pasha S, Jensen K, Pitt A, Spickett C. Mass spectrometry-based methods for identifying oxidized proteins in disease: advances and challenges. *Biomolecules*. 2015;5(2):378-411.

132. Cochrane CG, Spragg R, Revak SD. Pathogenesis of the adult respiratory distress syndrome. Evidence of oxidant activity in bronchoalveolar lavage fluid. *J Clin Invest*. 1983;71(3):754-761.
133. Gabbita SP, Aksenov MY, Lovell MA, Markesbery WR. Decrease in peptide methionine sulfoxide reductase in Alzheimer's disease brain. *J Neurochem*. 1999;73(4):1660-1666.
134. Gaza-Bulseco G, Faldu S, Hurkmans K, Chumsae C, Liu H. Effect of methionine oxidation of a recombinant monoclonal antibody on the binding affinity to protein A and protein G. *J Chromatogr B Anal Technol Biomed Life Sci*. 2008;870(1):55-62.
135. Gladstone Jr. IM, Levine RL. Oxidation of proteins in neonatal lungs. *Pediatrics*. 1994;93(5):764-768.
136. Pryor WA, Dooley MM, Church DF. Inactivation of human A-1-proteinase inhibitor by gas-phase cigarette smoke. *Biochem Biophys Res Commun*. 1984;122(2):676-681.
137. Stadtman ER, Berlett BS. Neurotoxicity and disease. *Society*. 1997;485-494.
138. Moskovitz J, Bar-Noy S, Williams WM, Requena J, Berlett BS, Stadtman ER. Methionine sulfoxide reductase (MsrA) is a regulator of antioxidant defense and lifespan in mammals. *Proc Natl Acad Sci USA*. 2001;98(23):12920-12925.
139. Ruan H, Tang XD, Chen ML, Joiner MLA, Sun G, Brot N, Weissbach H, Heinemann SH, Iverson L, Wu CF, Hoshi T. High-quality life extension by the enzyme peptide methionine sulfoxide reductase. *Proc Natl Acad Sci USA*. 2002;99(5):2748-2753.
140. Liang X, Kaya A, Zhang Y, Le DT, Hua D, Gladyshev VN. Characterization of methionine oxidation and methionine sulfoxide reduction using methionine-rich cysteine-free proteins. *BMC Biochem*. 2012;13(1):21.
141. Chandler KB, Pompach P, Goldman R, Edwards N. Exploring site-specific N-glycosylation microheterogeneity of haptoglobin using glycopeptide CID tandem mass spectra and glycan database search. *J Proteome Res*. 2013;12(8):3652-3666.
142. Roth Z, Yehezkel G, Khalaila I. Identification and quantification of protein glycosylation. *Int J Carbohydr Chem*. 2012;2012:1-10.
143. Hennet T. Diseases of glycosylation beyond classical congenital disorders of glycosylation. *Biochim Biophys Acta - Gen Subj*. 2012;1820(9):1306-1317.
144. Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME. *Essentials of Glycobiology*. 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1999.

145. Taylor, M.E. & Drickamer K. *Introduction to Glycobiology*. 2nd ed. Great Clarendon, Oxford: Oxford Univ Press; 2006.
146. Breitling J, Aebi M. N-linked protein glycosylation in the endoplasmic reticulum. *Cold Spring Harb Perspect Biol*. 2013;5(8):1-15.
147. Aebi M. N-linked protein glycosylation in the ER. *Biochim Biophys Acta - Mol Cell Res*. 2013;1833(11):2430-2437.
148. Lodish H, Berk A, Kaiser C, Krieger M, Scott MP, Bretscher A, Ploegh H, Matsudaira P. *Molecular Cell Biology*. 6th ed. New York, NY: W. H. Freeman and Company; 2008.
149. Drickamer K, Taylor ME. Evolving views of protein glycosylation. *Trends Biochem Sci*. 1998;23(9):321-324.
150. Apweiler R, Hermjakob H, Sharon N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta - Gen Subj*. 1999;1473(1):4-8.
151. Hang HC, Bertozzi CR. The chemistry and biology of mucin-type O-linked glycosylation. *Bioorganic Med Chem*. 2005;13(17):5021-5034.
152. Van den Steen P, Rudd PM, Dwek RA, Opdenakker G. Concepts and principles of O-linked glycosylation. *Crit Rev Biochem Mol Biol*. 2008;33(3):151-208.
153. Gill DJ, Clausen H, Bard F. Location, location, location: New insights into O-GalNAc protein glycosylation. *Trends Cell Biol*. 2011;21(3):149-158.
154. Bennett EP, Mandel U, Clausen H, Gerken TA, Fritz TA, Tabak LA. Control of mucin-type O-glycosylation: A classification of the polypeptide GalNAc-transferase gene family. *Glycobiology*. 2012;22(6):736-756.
155. Becker DJ, Lowe JB. Fucose: Biosynthesis and biological function in mammals. *Glycobiology*. 2003;13(7).
156. Luo Y, Koles K, Vorndam W, Haltiwanger RS, Panin VM. Protein O-fucosyltransferase 2 adds O-fucose to thrombospondin type 1 repeats. *J Biol Chem*. 2006;281(14):9393-9399.
157. Luo Y, Haltiwanger RS. O-fucosylation of notch occurs in the endoplasmic reticulum. *J Biol Chem*. 2005;280(12):11289-11294.
158. Matsumoto K, Ayukawa T, Ishio A, Sasamura T, Yamakawa T, Matsuno K. Dual roles of O-glucose glycans redundant with monosaccharide O-fucose on Notch in Notch trafficking. *J Biol Chem*. 2016;291(26):13743-13752.

159. Mormann M, Maček B, De Peredo AG, Hofsteenge J, Peter-Katalinić J. Structural studies on protein O-fucosylation by electron capture dissociation. *Int J Mass Spectrom.* 2004;234:11-21.
160. Rana NA, Nita-Lazar A, Takeuchi H, Kakuda S, Luther KB, Haltiwanger RS. O-Glucose trisaccharide is present at high but variable stoichiometry at multiple sites on mouse notch1. *J Bio Chem.* 2011;286:31623-31637.
161. Wang Y, Lee GF, Kelley RF, Spellman MW. Identification of a GDP-L-fucose:polypeptide fucosyltransferase and enzymatic addition of O-linked fucose to EGF domains. *Glycobiology.* 1996;6(8):837-842.
162. Bond MR, Hanover JA. A little sugar goes a long way: The cell biology of O-GlcNAc. *J Cell Biol.* 2015;208(7):869-880.
163. Bond MR, Hanover JA. O- GlcNAc cycling: A link between metabolism and chronic disease. *Annu Rev Nutr.* 2013;33(1):205-229.
164. Kearse KP, Hart GW. Lymphocyte activation induces rapid changes in nuclear and cytoplasmic glycoproteins (O-linked N-acetylglucosamine/nucleoplasmic and cytoplasmic glycosylation/Pro-Glu-Ser/Thr regions). *Biochemistry.* 1991;88:1701-1705.
165. Slawson C, Copeland RJ, Hart GW. O-GlcNAc signaling: A metabolic link between diabetes and cancer? *Trends Biochem Sci.* 2010;35(10):547-555.
166. Zeidan Q, Hart GW. The intersections between O-GlcNAcylation and phosphorylation: Implications for multiple signaling pathways. *J Cell Sci.* 2010;123:13-22.
167. Neubert P, Halim A, Zauser M, Essig A, Joshi HJ, Zatorska E, Larsen ISB, Loibl M, Castells-Ballester J, Aebi M, Clausen H, Strahl S. Mapping the O-mannose glycoproteome in *saccharomyces cerevisiae*. *Mol Cell Proteomics.* 2016;15(4):1323-1337.
168. Praissman JL, Willer T, Sheikh MO, Toi A, Chitayat D, Lin YY, Lee H, Stalnakker SH, Wang S, Prabhakar PK, Nelson SF, Stemple DL, Moore SA, Moremen KW, Campbell KP, Wells L. The functional O-mannose glycan on A-dystroglycan contains a phospho-ribitol primed for matriglycan addition. *Elife.* 2016;5:1-28.
169. Kailemia MJ, Ruhaak LR, Lebrilla CB, Amster IJ. Oligosaccharide analysis by mass spectrometry: A review of recent developments. *Anal Chem.* 2015;86(1):196-212.
170. Higel F, Demelbauer U, Seidl A, Friess W, Sörgel F. Reversed-phase liquid-chromatographic mass spectrometric N-glycan analysis of biopharmaceuticals. *Anal Bioanal Chem.* 2013;405(8):2481-2493.
171. Vreeker GCM, Wuhrer M. Reversed-phase separation methods for glycan analysis. *Anal Bioanal Chem.* 2016:1-20.

172. Su SJ, Grego B, Niven B, Hearn MTW. Analysis of group retention contributions for peptides separated by reversed phase high performance liquid chromatography. *J Liq Chromatogr.* 1981;4(10):1745-1764.
173. Sasagawa T, Okuyama T, Teller DC. Prediction of peptide retention times in reversed-phases high-performance liquid chromatography during linear gradient elution. *J Chromatogr.* 1982;240:329-340.
174. Liu DT. Deamidation: a source of microheterogeneity in pharmaceutical proteins. *Trends Biotechnol.* 1992;10:364-369.
175. Yang H, Zubarev RA. Mass spectrometrics analysis of asparagine deamidation and aspartate isomerizaion in polypeptides. *Electrophoresis.* 2011;31(11):1764-1772.
176. Patel J, Kothari R, Tunga R, Ritter NM, Tunga BS. Stability considerations for biopharmaceuticals, Part 1: Overview of protein and peptide degradation pathways. *Bioprocess Int.* 2011;9(1):20-31.
177. Scotchler JW, Robinson AB. Deamidation of glutaminyl residues: Dependence on pH, temperature, and ionic strength. *Anal Biochem.* 1974;59(1):319-322.
178. Robinson NE, Zabrouskov V, Zhang J, Lampi KJ, Robinson AB. Measurement of deamidation of intact proteins by isotopic envelope and mass defect with ion cyclotron resonance Fourier transform mass spectrometry. *Rapid Commun Mass Spectrom.* 2006;20:3535-3541.
179. Levine R, Moskovitz J, Stadtman E. Oxidation of methionine in proteins: Roles in antioxidant defense and cellular regulation. *IUBMB Life.* 2001;50(4):301-307.
180. Carney JM, Smith CD, Carney AM, Butterfield DA. Aging and oxygen-induced modifications in brain biochemistry and behavior. *Ann NY Acad Sci.* 1994;738(1):44-53.
181. Jori G, Galiazzo G, Marzotto A, Scoffone E. Selective and reversibe photo-oxidation of the methionyl residues in lysozyme. *J Biol Chem.* 1968;243(16):4272-4278.
182. Ray WJ, Latham HG, Katsoulis M, Koshland DE. Evidence for involvement of a methionine residue in the enzymatic action of phosphoglucomutase and chymotrypsin. *J Am Chem Soc.* 1960;82(17):4743-4744.
183. Kido K, Kassell B. Oxidation of methionine residues of porcine and bovine pepsins. *Biochemistry.* 1975;14(3):631-635.
184. Schachter H, Dixon GH. Identification of the methionine involved in the active center of chymotrypsin. *Biochem Biophys Res Commun.* 1962;9:132-137.

185. Badgett MJ, Boyes B, Orlando R. Prediction of peptide retention times in hydrophilic interaction liquid chromatography (HILIC) based on amino acid composition. *Chromatogr Today*. 2015;39-42.
186. Alpert AJ. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *J Chromatogr A*. 1990;499:177-196.
187. Hao P, Ren Y, Datta A, Tam JP, Sze SK. Evaluation of the effect of trypsin digestion buffers on artificial deamidation. *J Proteome Res*. 2015;14:1308-1314.
188. Stephenson RC, Clarke S. Succinimide formation from aspartyl and asparaginyl peptides as a model for the spontaneous degradation of proteins. *J Biol Chem*. 1989;264(11):6164-6170.
189. Reissner KJ, Aswad DW. Deamidation and isoaspartate formation in proteins: Unwanted alterations or surreptitious signals? *Cell Mol Life Sci*. 2003;60(7):1281-1295.
190. Cournoyer JJ, Pittman JL, Ivleva VB, et al. Deamidation: Differentiation of aspartyl from isoaspartyl products in peptides by electron capture dissociation. *Protein Sci*. 2005;14:452-463.
191. Konuklar FAS, Aviyente V, Sen TZ, Bahar I. Modeling the deamidation of asparagine residues via succinimide intermediates. *J Mol Model*. 2001;7(5):147-160.
192. Zhu Y, Shan X, Yuzwa SA, Vocadlo DJ. The emerging link between O-GlcNAc and Alzheimer disease. *J Biol Chem*. 2014;289(50):34472-34481.
193. Hart GW, Akimoto Y. Chapter 18: The O-GlcNAc modification. In *Essentials of Glycobiology*. 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1999.
194. Ma J, Hart GW. O-GlcNAc profiling: from proteins to proteomes. *Clin Proteomics*. 2014;11(1):8.
195. Ngoh GA, Facundo HT, Zafir A, Jones SP. O-GlcNAc signaling in the cardiovascular system. *Circ Res*. 2010;107(2):171-185.
196. Yuzwa SA, Shan X, Macauley MS, Clark T, Skorobogatko Y, Vosseller K, Vocadlo DJ. Increasing O-GlcNAc slows neurodegeneration and stabilizes tau against aggregation. *Nat Chem Biol*. 2012;8(4):393-399.
197. Joshi HJ, Steentoft C, Schjoldager KT, BG, Vakhrushev SY, Wandall HH, Clausen H. Protein O-GalNAc glycosylation: Most complex and differentially regulated PTM. *Glycosci Biol Med*. 2015:1-1569.
198. Hanisch FG. O-glycosylation of the mucin type. *Biol Chem*. 2001;382(2):143-149.

199. O'Halloran AM, Patterson CC, Horan P, Maree A, Curtin R, Stanton A, McKeown PP, Shields DC. Genetic polymorphisms in platelet-related proteins and coronary artery disease: Investigation of candidate genes, including N-acetylgalactosaminyltransferase 4 (GALNT4) and sulphotransferase 1A1/2 (SULT1A1/2). *J Thromb Thrombolysis*. 2009;27(2):175-184.
200. Fakhro KA, Choi M, Ware SM, Belmont JW, Towbin JA, Lifton RP, Khokha MK, Brueckner M. Rare copy number variations in congenital heart disease patients identify unique genes in left-right patterning. *Proc Natl Acad Sci USA*. 2011;108(7):2915-2920.
201. Hallgren P, Lundblad A, Svensson S. A new type of carbohydrate-protein linkage in a glycopeptide from normal human urine. *J Biol Chem*. 1975;250(14):5312-5314.
202. Serth K, Schuster-Gossler K, Kremmer E, Hansen B, Marohn-Köhn B, Gossler A. O-fucosylation of DLL3 is required for its function during somitogenesis. *PLoS One*. 2015;10(4):1-21.
203. Moloney DJ, Lin AI, Haltiwanger RS. The O-linked fucose glycosylation pathway. *J Biol Chem*. 1997;272(30):19046-19050.
204. Luo Y, Nita-Lazar A, Haltiwanger RS. Two distinct pathways for O-fucosylation of epidermal growth factor-like or thrombospondin type 1 repeats. *J Biol Chem*. 2006;281(14):9385-9392.
205. Okajima T, Xu A, Lei L, Irvine KD. Chaperone Activity of Protein. *Science*. 2005;307:1599-1603.
206. Ricketts LM, Dlugosz M, Luther KB, Haltiwanger RS, Majerus EM. O-fucosylation is required for ADAMTS13 secretion. *J Biol Chem*. 2007;282(23):17014-17023.
207. Müller J, Rana NA, Serth K, Kakuda S, Haltiwanger RS, Gossler A. O-fucosylation of the NOTCH ligand mDLL1 by POFUT1 is dispensable for ligand function. *PLoS One*. 2014;9(2):1-9.
208. Becker DJ, Lowe JB. Leukocyte adhesion deficiency type II. *Biochim Biophys Acta - Mol Basis Dis*. 1999;1455:193-204.
209. Flögel M, Lauc G, Gornik I, Maček B. Fucosylation and galactosylation of IgG heavy chains differ between acute and remission phases of juvenile chronic arthritis. *Clin Chem Lab Med*. 1998;36(2):99-102.
210. Gornik I, Maravić G, Dumić J, Flögel M, Lauc G. Fucosylation of IgG heavy chains is increased in rheumatoid arthritis. *Clin Biochem*. 1999;32(8):605-608.

211. Kim ML, Chandrasekharan K, Glass M, Shi S, Stahl MC, Kaspar B, Stanley P, Martin PT. O-fucosylation of muscle agrin determines its ability to cluster acetylcholine receptors. *Mol Cell Neurosci.* 2008;39(3):452-464.
212. Orntoft TF, Vestergaard EM. Clinical aspects of altered glycosylation of glycoproteins in cancer. *Electrophoresis.* 1999;20(2):362-371.
213. Scanlin TF, Glick MC. Terminal glycosylation in cystic fibrosis. *Biochim Biophys Acta - Mol Basis Dis.* 1999;1455(2-3):241-253.
214. Zhang P, Woen S, Wang T, Liao B, Zhao S, Chen C, Yang Y, Song Z, Wormald MR, Yu C, Rudd PM. Challenges of glycosylation analysis and control: An integrated approach to producing optimal and consistent therapeutic drugs. *Drug Discov Today.* 2016;21(5):740-765.
215. Khoury GA, Baliban RC, Floudas CA. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci Rep.* 2011;1:90.
216. Devakumar A, Mechref Y, Kang P, Novotny MV, Reilly JP. Identification of isomeric N-glycan structures by mass spectrometry with 157 nm laser-induced photofragmentation. *J Am Soc Mass Spectrom.* 2008;19(7):1027-1040.
217. Bianco GA, Toscano MA, Ilarregui JM, Rabinovich GA. Impact of protein-glycan interactions in the regulation of autoimmunity and chronic inflammation. *Autoimmun Rev.* 2006;5(5):349-356.
218. Arnold JN, Wormald MR, Sim RB, Rudd PM, Dwek RA. The impact of glycosylation on the biological function and structure of human immunoglobulins. *Annu Rev Immunol.* 2007;25:21-50.
219. Lin CW, Tsai MH, Li ST, Tsai TI, Chu KC, Liu YC, Lai MY, Wu CY, Tseng YC, Shivatare SS, Wang CH, Chao P, Wang SY, Shih HW, Zeng YF, You TH, Liao JY, Tu YC, Lin YS, Chuang HY, Chen CL, Tsai CS, Huang CC, Lin NH, Ma C, Wu CY, Wong CH. A common glycan structure on immunoglobulin G for enhancement of effector functions. *Proc Natl Acad Sci USA.* 2015;112(22):201513456.
220. Huhn C, Selman MHJ, Ruhaak LR, Deelder AM, Wuhrer M. IgG glycosylation analysis. *Proteomics.* 2009;9(4):882-913.
221. Parekh RB, Dwek RA, Sutton BJ, Fernandes DL, Leung A, Stanworth D, Rademacher TW, Mizuochi T, Taniguchi T, Matsuta K, Takeuchi F, Nagano Y, Miyamoto T, Kobata A. Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG. *Nature.* 1985;316(6027):452-457.
222. Abès R, Teillaud JL. Impact of glycosylation on effector functions of therapeutic IgG. *Pharmaceuticals.* 2010;3(1):146-157.

223. Varki A. Biological roles of oligosaccharides: All of the theories are correct. *Glycobiology*. 1993;3(2):97-130.
224. Shade KT, Anthony R. Antibody glycosylation and inflammation. *Antibodies*. 2013;2(3):392-414.
225. Wuhrer M, Stam JC, Van De Geijn FE, Koeleman CAM, Verrips CT, Dolhain RJEM, Hokke CH, Deelder AM. Glycosylation profiling of immunoglobulin G (IgG) subclasses from human serum. *Proteomics*. 2007;7(22):4070-4081.
226. Endo T, Kochibe N, Kobata A. Structural study of the carbohydrate moieties of two human immunoglobulin subclasses (IgG2 and IgG4). *Glycoconj J*. 1989;6(1):57-66.
227. Mancera-Arteu M, Gimnez E, Barbosa J, Sanz-Nebot V. Identification and characterization of isomeric N-glycans of human alfa-acid-glycoprotein by stable isotope labelling and ZIC-HILIC-MS in combination with exoglycosidase digestion. *Anal Chim Acta*. 2016;940:92-103.
228. Toegel S, Pabst M, Wu SQ, Grass J, Goldring MB, Chiari C, Kolb A, Altmann F, Viernstein H, Unger FM. Phenotype-related differential α -2,6- or α -2,3-sialylation of glycoprotein N-glycans in human chondrocytes. *Osteoarthr Cartil*. 2010;18(2):240-248.
229. Anthony RM, Nimmerjahn F, Ashline DJ, Reinhold VN, Paulson JC, Ravetch J V. A recombinant IgG Fc that recapitulates the antiinflammatory activity of IVIG. *Science*. 2008;320(1):373-376.
230. Hashira S, Okitsu-Negishi S, Yoshino K. Placental transfer of IgG subclasses in a Japanese population. *Pediatr Int*. 2000;42(4):337-342.
231. Tao S, Huang Y, Boyes BE, Orlando R. Liquid chromatography-selected reaction monitoring (LC-SRM) approach for the separation and quantitation of sialylated N-glycans linkage isomers. *Anal Chem*. 2014;86(21):10584-10590.