

Analysis of modifications and behaviours of a hybrid beta-hexosaminidase and other biomolecules by mass spectrometry and liquid chromatography

by

Taylor Battellino

A Thesis submitted to the Faculty of Graduate Studies of the

University of Manitoba

in partial fulfillment of the requirement of the degree of

Master of Science

Department of Chemistry

Copyright © 2022 by Taylor Battellino

Abstract

The study of post-translational modifications is a large focus in proteomics experiments as these modifications are often what drive signalling and trafficking of proteins within a cell or an organism. In the immune system, the glycosylation of immunoglobulins is a key point of the study of xenotransplantation. In lysosomes, the glycosylation or phosphorylation of both an enzyme and its substrate is highly important in degradation pathways, and will result in disease if interrupted. Immunoglobulin G and Tay-Sachs disease are respective examples of these instances, and both are discussed in detail in this thesis. Subtypes of porcine immunoglobulin G were studied by liquid chromatography and mass spectrometry in order to determine classification by glycoproteome analysis. Variations of these techniques, particularly electrospray-ionization and matrix assisted laser desorption/ionization mass spectrometry, were used to study the glycosylation and phosphorylation profiles of a hybrid beta-hexosaminidase enzyme which was designed to be an improved contender for enzyme replacement therapy for Tay-Sachs disease. The substrate of this enzyme, a ganglioside, was also analyzed using these techniques in order to determine detectability and relative quantification in mice brains.

Liquid chromatography and mass spectrometry are the techniques most commonly utilized in proteomic and glycoproteomic experiments. Complex samples can be analyzed by nuanced techniques such as two-dimensional liquid chromatography, and this information can be used to create algorithms to aid in identification of post-translationally modified peptides. The Sequence-Specific Retention Calculator is an example of an algorithm which was created in part for this purpose, and it is explored in this thesis in two different experiments. Phosphorylated peptides are more difficult to detect in mass spectrometry, and only make up a small portion of peptides in a cell. Due to this, it is useful to utilize liquid chromatography to both simplify a proteomic sample and create another dimension in which identification can be performed. Development of these techniques includes the optimization of a mobile phase, which was explored using both formic and acetic acid.

Acknowledgements

First and foremost, I'd like to thank my research supervisors Dr. H  l  ne Perreault and Dr. Oleg Krokhin not only for supervising me in this work, but for giving me every opportunity to grow as a researcher and for always being understanding and appreciative.

I'd also like to thank my collaborators, former lab instructors, and particularly my lab members Baylie Gigolyk and Milan Teraiya for their support and guidance, as well as Dr. Teresa De Kievit for her help and encouragement on my committee.

Lastly, I want to thank my friends and family for supporting me in my research endeavors and for encouraging me to continue pursuing my passions.

Table of Contents

Abstract	ii
Acknowledgements	iii
List of Tables	vi
List of Figures	vii
List of Abbreviations	ix
1 Chapter 1: General Introduction	2
1.1 Proteins and Peptides	3
1.2 Liquid Chromatography	9
1.3 Mass Spectrometry.....	11
1.4 Biological molecules investigated in this thesis	17
1.5 Objectives	18
1.6 References.....	19
2 Chapter 2: Glycoproteomic study of porcine IgG subtypes by LC-MS/MS	23
2.0 Contributions of Authors	25
2.1 Abstract.....	26
2.2 Introduction.....	27
2.3 Experimental	31
2.4 Results and Discussion	34
2.5 Conclusion	48
2.6 Acknowledgments.....	49
2.7 Supplementary Information	50
2.8 References.....	56
3 Chapter 3: Influence of ion pairing agent in LC-MS proteomic experiments.....	61
3.0 Contributions of Authors	63
3.1 Abstract.....	64
3.2 Introduction.....	65
3.3 Experimental	67
3.4 Results.....	70
3.5 Discussion.....	76
3.6 Supporting Information.....	78
3.7 References.....	84

4	Chapter 4: Retention time prediction for phosphopeptides in LC-MS.....	88
4.0	Contributions of Authors	90
4.1	Abstract	91
4.2	Introduction.....	92
4.3	Materials and Methods.....	95
4.4	Results and Discussion	98
4.5	Conclusions.....	112
4.6	Acknowledgements.....	113
4.7	Supplementary Information	114
4.8	References.....	116
5	Chapter 5: Analysis of glycosylation and phosphorylation profiles of HexM and GM2....	119
5.1	Introduction.....	120
5.2	Experimental.....	122
5.3	Results and Discussion	124
5.4	Conclusions.....	132
5.5	References.....	133
6	Chapter 6: General Conclusion.....	134

List of Tables

Table 1.1 Amino acids and their abbreviations.	4
Table 2.1 Pig IgG heavy chain constant protein sections identified by X!Tandem from RPLC-MS/MS data for three different tryptic digestion samples run separately (WT, DKO, and CKO).	33
Table 2.2 List of proteins other than IgG heavy chains identified by X!Tandem.	38
Table 2.3 List of the most abundant glycopeptides observed in this study.	40
Table 4.1 Identification summary for the analyses of various phosphorylated isolates.	99
Table 5.1 MALDI-MS analysis of phosphorylation of high mannose N-linked glycosylation sites of recombinant HexA and recombinant HexM.....	126
Table 5.2 Relative quantitation of GM2 ganglioside in WT, KO, and DKO mice brains.....	131

List of Figures

Figure 1.1 Structures of common carbohydrates and O- and N-linked glycans.	7
Figure 1.2 Structures of common gangliosides and their respective degradation pathways.	8
Figure 1.3 Diagram of the laser desorption/ionization process which occurs in a MALDI instrument.	12
Figure 1.4 Schematic of a TOF/TOF mass analyzer.....	13
Figure 1.5 Diagram of ionization process in LC-ESI-MS.	14
Figure 1.6 Schematic of a triple-quadrupole TOF hybrid mass analyzer.	15
Figure 1.7 Diagram of the QExactive Orbitrap mass analyzer.	16
Figure 2.1 Workflow used to identify pig IgG subtypes in each sample analyzed in this study..	29
Figure 2.2 Sequence alignments for all proteins listed in Table 1.	36
Figure 2.3 Total ion current obtained from the high-performance liquid chromatography/mass spectrometry (HPLC/MS) analysis of the tryptic digest of the CKO pig IgG sample.	42
Figure 2.4 Tandem mass spectra for two glycopeptides ionized as $(M + 3H)^{3+}$ ions.....	43
Figure 2.5 Examples of N-glycan structures observed in this study of pig IgG.....	45
Figure 3.1 Variation of separation selectivity between formic and acetic acid-based eluent systems in peptide RPLC.	71
Figure 3.2 The effect of ion pairing modifier on LC-MS/MS ID output and detection sensitivity	73
Figure 4.1 Experimental procedures used for retention data collection in 1D and 2D LC-MS/MS proteomic experiments targeting retention time prediction of phosphorylated peptides.....	100
Figure 4.2 Effect of ion pairing modifier on separation selectivity of modified – non-modified peptide pairs.	102
Figure 4.3 Composition and sequence specific features driving variations in dHI – retention shift upon phosphorylation.....	105
Figure 4.4 Retention time prediction for phosphorylated peptides.....	109
Figure 5.1 Compared amino acid sequences of the HexA α -subunit and HexM μ -subunit.....	124
Figure 5.2 MALDI-MS spectra of glycosylated peptide GTFFINKTE from the Glu-C digestion of a) HexA, b) HexM and c) hyperphosphorylated HexM, showing the presence of high mannose species and enhanced phosphorylation.	125

Figure 5.3 ESI MS (A, B) and MS/MS (C) spectra of phosHexM fractions containing high mannose conjugates of FINKTEIEDF..... 127

Figure 5.4 MALDI-TOF mass spectra of permethylated ganglioside extracts from brains of wild-type mice..... 128

Figure 5.5 MALDI-TOF mass spectra of permethylated ganglioside extracts from brains of knockout (B) and double knockout (A, C) mice..... 129

Figure 5.6 MALDI-TOF mass spectra of permethylated ganglioside extracts from brains of WT (A), DKO (B), and KO (C) mice spiked with an isotopically-labelled standard..... 130

List of Abbreviations

AA	Acetic acid
ACN	Acetonitrile
BBB	Blood-brain barrier
CMAH	Cytidine monophosphate- <i>N</i> -acetylneuraminic acid hydroxylase
CNS	Central nervous system
Da	Dalton
DHB	2,5-dihydroxybenzoic acid
DKO	Double knockout
DNA	Deoxyribonucleic acid
DTT	Dithiothrietol
ER	Endoplasmic reticulum
ERT	Enzyme replacement therapy
ESI	Electrospray ionization
FA	Formic acid
Fab	Antigen-binding fragment
Gal	Galactose
GD1	GD1 ganglioside (disialic)
GD2	GD2 ganglioside (disialic)
Glc	Glucose
GlcNAc	<i>N</i> -acetyl glucosamine
GM1	GM1 ganglioside (monosialic)
GM2	GM2 ganglioside (monosialic)
GM2AP	GM2 activator protein
GM3	GM3 ganglioside (monosialic)
HC	Heavy chain
HexA	Beta-hexosaminidase A
HexM	Beta-hexosaminidase M
HFBA	Heptafluorobutyric acid
HILIC	Hydrophilic interaction liquid chromatography
IAA	Iodoacetamide

IgA	Immunoglobulin A
IgG	Immunoglobulin G
IgM	Immunoglobulin M
KO	Knockout
KU	Kyoto University
LC	Liquid chromatography
LSD	Lysosomal storage disorder
M6P	Mannose-6-phosphate
MALDI	Matrix-assisted laser desorption/ionization
Man	Mannose
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
NeuAc	N-acetyl neuraminic acid (sialic acid)
phosHexM	Hyper-phosphorylated beta-hexosaminidase M
PTM	Post-translational modification
QTOF	Quadrupole time-of-flight
SAX	Strong anion exchange
SCX	Strong cation exchange
SSRCalc	Sequence-specific retention calculator
TFA	Trifluoroacetic acid
TIC	Total ion chromatogram
TOF	Time-of-flight
UM	University of Manitoba
WT	Wild-type

1 Chapter 1: General Introduction

1.1 Proteins and Peptides

Proteomics is the study of proteins and their interactions in a system. The proteomic profile of a cell or other entity is referred to as the proteome, which is typically complex and dynamic. Proteins can bear many modifications, some called post-translational modifications (PTMs), such as phosphorylation and glycosylation. These modifications alter the behaviour of both the specific amino acids they affect, as well as the protein as a whole. It is important, therefore, to understand these behavioural changes on both a small and a large scale.

Complex proteins such as enzymes can include multiple subunits, with various modifications. Each feature is important in terms of properties such as folding capability, lability, and trafficking. The study of individual amino acids in a sequence is essential to understand interactions between enzyme and substrate, or protein and receptor. Proteomics employs techniques such as liquid chromatography (LC) and mass spectrometry (MS) to study these systems either using a bottom-up or top-down approach. Bottom-up proteomics involves the digestion of purified proteins and subsequent analysis of peptides, whereas top-down proteomics involves the analysis of intact proteins and their instrumental fragmentation.¹ Both approaches provide a comprehensive analysis of the protein or proteins in question and can be applied quantitatively.

Techniques such as LC and MS are also useful in the analysis of other biomolecules such as gangliosides.² This type of analysis works similarly to top-down proteomics, as these molecules can also be fragmented into smaller pieces for further analysis. This is useful for the study of glycosylation, for example, as it is with post-translationally modified peptides.

1.1.1 Amino acids

There are twenty canonical amino acids which exist in nature. Modifications such as glycosylation, phosphorylation, alkylation, and oxidation may only occur with particular amino acids due to their unique properties. These modifications as well as the abbreviations of these amino acids can be found in **Table 1.1**.

Table 1.1 Amino acids and their abbreviations.³

Amino acid	3-letter abbreviation	1-letter abbreviation	Relevant possible modification(s)
Alanine	Ala	A	
Arginine	Arg	R	
Asparagine	Asn	N	N-glycosylation
Aspartic acid	Asp	D	
Cysteine	Cys	C	Alkylation
Glutamic acid	Glu	E	
Glutamine	Gln	Q	
Glycine	Gly	C	
Histidine	His	H	
Isoleucine	Ile	I	
Leucine	Leu	L	
Lysine	Lys	K	
Methionine	Met	M	
Phenylalanine	Phe	F	
Proline	Pro	P	
Serine	Ser	S	Phosphorylation, O-glycosylation
Threonine	Thr	T	Phosphorylation, O-glycosylation
Tryptophan	Trp	W	
Tyrosine	Tyr	Y	Phosphorylation
Valine	Val	V	

Serine, threonine, and tyrosine are the three amino acids which can be subjected to phosphorylation. This is due to them including a particular attachment site for these negatively charged groups. Serine, threonine, and asparagine are the amino acids which can be subjected to glycosylation. Alkylation of cysteine is another important modification, as it is required in the enzymatic digestion of proteins.⁴

Hydrophilic amino acids include asparagine, cysteine, glutamine, serine, and threonine. Hydrophobic amino acids include alanine, isoleucine, leucine, methionine, phenylalanine, tyrosine, and valine. Negatively charged amino acids include glutamic acid and aspartic acid, and positively charged amino acids include lysine, arginine, and histidine. However, the overall charge of an amino acid or peptide largely depends on the pH of the environment in which it exists.⁵ Lastly, lysine and arginine are particularly important as they are terminal amino acids in tryptic peptides.

When studying peptides by MS, a technique called tandem-MS (MS/MS) can be used to identify amino acid sequences.⁵ This is because fragment ions are created which correspond to the mass of individual amino acid residues. The exact sequence of a peptide is important for understanding peptide behaviour in both LC and MC systems, as the assembly of its amino acids can affect features such as ion-pairing and helicity. These are critical features in chromatographic peptide separations. LC is a very useful tool for this type of analysis, as peptide behaviour can be analyzed by observing retention in various systems. The pH of the mobile phase, addition of an ion-pairing modifier, and functional groups on the column resin are all factors which will affect the retention of a peptide due to the impact on individual amino acids. As post-translational modifications heavily influence amino acid and peptide behaviour, analysis by LC is useful for modified peptides as well.⁶

1.1.2 Phosphorylation

Phosphorylation is arguably the most important post-translational peptide modification, as it is a major feature in signalling pathways. The phosphorylation or dephosphorylation of a protein by an enzyme is often the first step in pathways such as those leading to cell differentiation.⁷ For this reason, it is also the most commonly found PTM. However, phosphorylated peptides only make up a small number of peptides in a typical proteomic sample. Therefore, enrichment techniques such as titanium dioxide (TiO₂)⁸ and strong cation exchange (SCX)⁹ are often used prior to studying phosphopeptides by LC-MS.

Phosphorylation of a peptide or protein involves the addition of a hydrophilic, negatively charged phosphate group, and consequent addition of 79.9 mass units. Both of these features make phosphorylation a detectable modification by both components of LC-MS. In LC, the phosphorylation of peptides can be observed by a shift in retention, regardless of the system.¹⁰ The specifics of the chromatographic system, however, will determine both the direction and magnitude of this shift. In MS, phosphorylation is detectable by the mass increase. The issue is that the addition of a negatively charged group results in an increase in lability, particularly in matrix-assisted laser desorption/ionization (MALDI) MS. In all kinds of MS, another issue is that the signal created from a phosphorylated peptide is approximately 30% smaller than that from a peptide without modifications.¹¹ Phosphopeptide enrichment is therefore also important for this reason, as the removal of dephosphorylated or unphosphorylated peptides will result in a more easily analyzable mass spectrum. Another way to combat this difficulty in phosphopeptide analysis by MS is to create a way in which these peptides can be identified in LC. Using the sequence-specific retention shifts that occur as a result of phosphorylation, an algorithm can be created which will identify phosphopeptides both by mass and retention in a chromatographic system.

Phosphorylation is also a modification that can occur on glycoproteins. In this case, the phosphate group attaches itself to an oxygen atom on a sugar moiety and/or on an amino acid. Lability of these molecules, especially in MALDI-MS is even more of an issue here, as these groups tend not to remain intact during the ionization process.¹² As with the phosphorylation of peptides, the phosphorylation of glycopeptides is also very important in signalling pathways and

trafficking. An example of this is the mannose-6-phosphate (M6P) pathway which is crucial in lysosomal enzyme trafficking.¹³

1.1.3 Glycosylation

Glycosylation is similar to phosphorylation as a PTM, as it also plays a role in signalling and trafficking pathways.¹³ The major difference is that glycosylation, the addition of sugar, involves the addition of many different variations of a hydrophilic group. The majority of these sugars carry no charge, however there are some which carry a negative charge. The most common example of this is sialic acid, denoted as a purple diamond in **Figure 1.1**.¹⁴

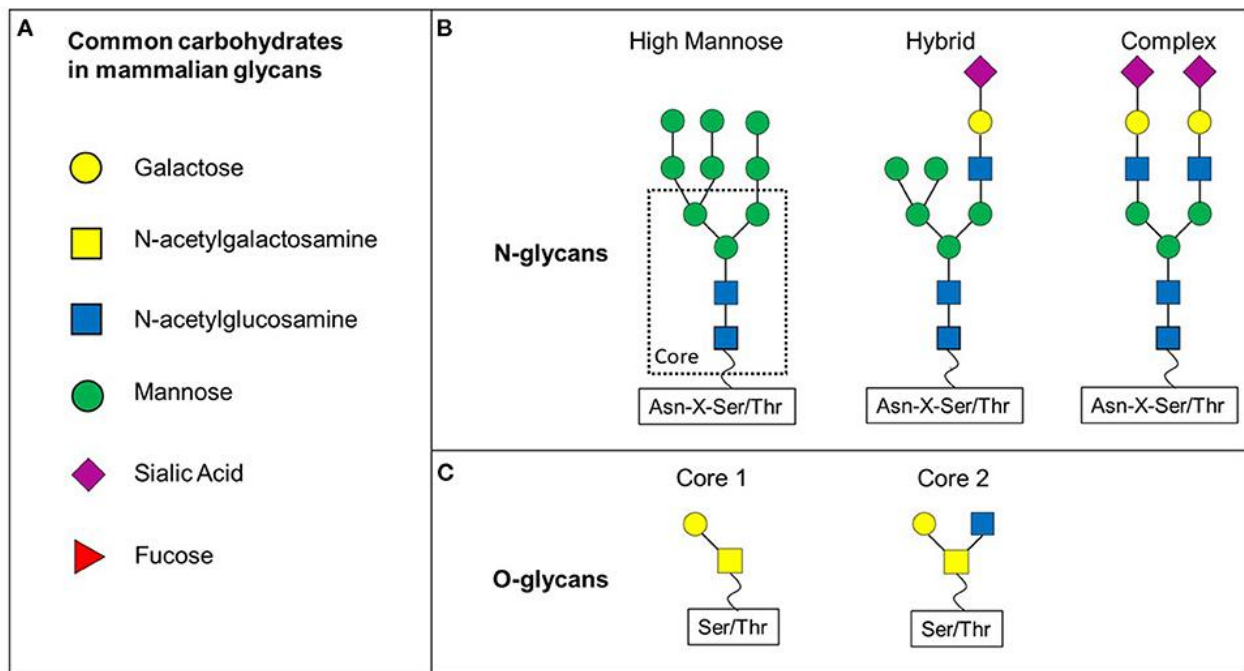


Figure 1.1 Structures of common carbohydrates and O- and N-linked glycans. Amended from reference 14.¹⁴

Hexose sugars such as glucose, galactose, and mannose are often denoted by an H, whereas N-acetylglucosamine and N-acetylgalactosamine are denoted by an N. The high mannose glycan in **Figure 1.1** would be described as H9N2.

Though these sugars are typically added one-by-one via enzymes, they can form structures called glycans. Types of glycans include complex and high-mannose, and they can either be O-linked or N-linked. O-linked glycans attach themselves to serine or threonine, whereas N-linked

glycans attach to asparagine.¹⁵ The specific amino acid patterns can be seen in **Figure 1.1**. N-linked glycans are more commonly found in proteins and are especially important in proteins such as immunoglobulins.¹⁶ Glycoproteomic profiles of immunoglobulins are important for the classification of types and corresponding subtypes in different species.

The same lability and decrease in signal in MS that occurs with phosphopeptides also occurs with glycopeptides.⁶ Similarly to phosphorylation, glycopeptide enrichment techniques exist which can aid in detection.¹⁷ However, this process is a little more difficult due to the variation in glycan composition.

Glycosylation is a modification that can also occur on other types of biomolecules, such as gangliosides.¹⁸ Gangliosides are molecules which are essentially glycosylated sphingolipids, and their glycosylation patterns are displayed in **Figure 1.2**.¹⁹

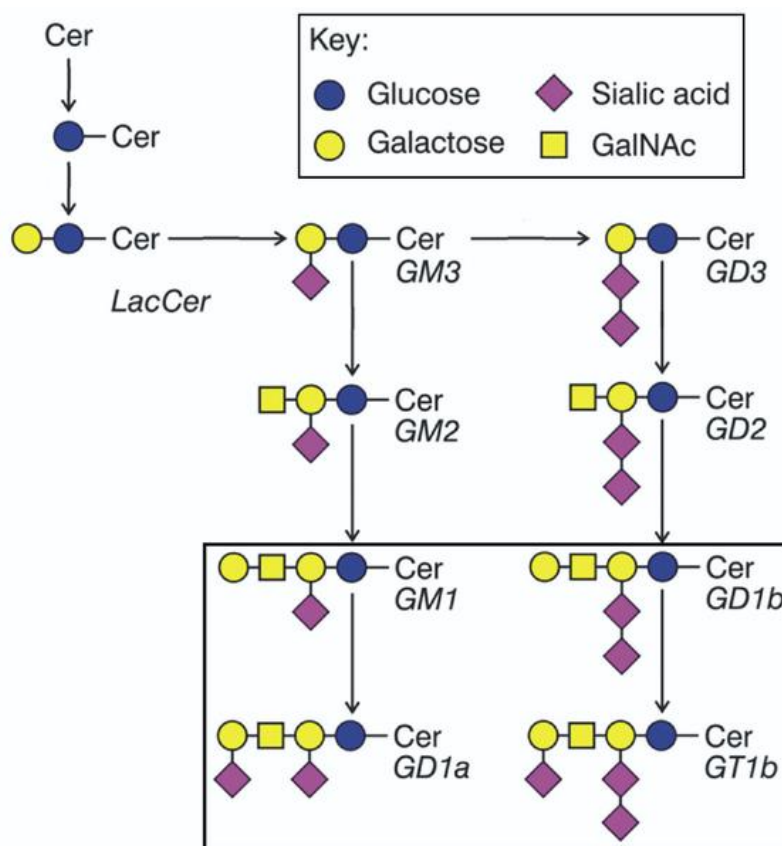


Figure 1.2 Structures of common gangliosides and their respective degradation pathways. Amended from reference 19.¹⁹

Enzymes such as beta-hexosaminidase A are responsible for the degradation of these gangliosides in the lysosomes of cells. The glycosylation of these enzymes is also important in this process due to the existence of the mannose-6 pathway.¹³

1.2 Liquid Chromatography

Reversed-phase liquid chromatography is a useful and dynamic tool which can be used for many different types of proteomic analyses. It is often hyphenated to MS for detection. A common ionization technique that is used in LC-MS is called electrospray ionization (ESI) which translates the liquid-phase sample into a gas phase suitable for MS.²⁰ RPLC in this case is used primarily to separate complex proteomic samples for more reliable identification via MS or MS/MS. However, it can also be used prior to RPLC-MS analysis, which is called 2D LC.²¹

The first dimension of 2D LC is typically a separation via a different type of LC, such as hydrophilic interaction liquid chromatography (HILIC) or reversed-phase pH 10 separation. This type of LC is used to further simplify samples before LC-MS analysis by separating and fractionating them in another system. The second dimension is almost exclusively reversed-phase LC-ESI-MS. This includes the use of a C18 column or nano-column and an increasing acetonitrile gradient.²¹

1.2.1 Enrichment

Prior to one-dimension or two-dimensional analysis, LC can be used as an enrichment technique for complex proteomic samples. An example of this is strong cation exchange (SCX), which is used in phosphopeptide experiments. SCX uses a column bearing negatively charged groups, usually sulfonic acid. As this type of chromatography involves the exchange of cations, charged peptides will elute from least positive to most positive as the salt gradient in the mobile phase is increased.⁹

Trypsin is the enzyme most commonly used in protein digestion. Due to its cleavage occurring after positively charged amino acids lysine and arginine, tryptic peptides carry a +2 charge at pH 7.²² This also accounts for the positively charged N-terminus. Phosphopeptides carry an approximate +1 charge due to the addition of the negatively charged phosphate group. In SCX, phosphopeptides will elute in the form of a sharp peak on a chromatogram, followed by

a very broad peak containing other tryptic peptides. This separation allows the collection of phosphopeptides for further analysis without the involvement of uncharged peptides.⁹

1.2.2 First dimension

The first dimension of 2D LC is typically done on a micro-flow instrument and employs a different separation system from the second dimension. Two examples of common separation methods for first dimension analysis are hydrophilic interaction ion chromatography (HILIC) and reversed-phase LC at pH 10. These conditions are used to separate complex proteomic samples and fractionate them into smaller components for further analysis in a different system.²³

Like normal-phase LC separations, HILIC involves the use of an increasing water gradient. HILIC columns include features such as zwitterions or amide groups which are linked to a silica resin, and it is this which causes the hydrophilic interactions with peptides. Peptides elute from this type of column in order of increasing polarity.²⁴

Separation at pH 10, in comparison, uses a reversed-phase system. This involves the use of a non-polar column, typically C18, and an increasing acetonitrile gradient.²⁵ The high pH comes from the addition of a salt to the mobile phase, such as ammonium formate. This change in pH results in peptides carrying smaller positive charges compared to a pH 7 environment.

1.2.3 Second dimension

The second dimension of 2D LC is often done on a nano-flow instrument, employing reversed-phase separation with a C18 column and an increasing acetonitrile gradient. Fractions collected from the first-dimension separation are run individually for simpler analysis of complex samples.²³

Two-dimensional (2D) experiments end with MS or MS/MS and the chromatograph used for the final dimension is typically hyphenated to a mass analyzer. Both the chromatographic and spectrometric data can be analyzed and used in both the detection and identification of peptides or other biomolecules in a sample. By spiking samples analyzed by LC-MS/MS with a mixture of standard peptides with known masses and retention indices in a chromatographic system, results can be adjusted to account for instrument variation and this information can be used to create algorithms for more reliable peptide identification.²⁶

1.3 Mass Spectrometry

Mass spectrometry is a universal technique used for the detection and identification of compounds in a sample. Under a high vacuum, MS measures the mass-to-charge (m/z) ratio of ions and outputs data in the form of a spectrum. There are various types of mass analyzers which are commonly used in proteomics, such as quadrupole, triple quadrupole, ion trap, and time-of-flight analyzers. Mass analyzers can be hyphenated to a LC or coupled to a different technique entirely.²⁷ Due to the high vacuum and therefore gas phase conditions of MS, an ionization technique must be implemented between LC and MS. An example of this is electrospray ionization (ESI).²⁰ A MS technique which is not able to be coupled with LC is MALDI, which involves the ionization of samples crystallized on a metal plate.²⁸

Tandem-MS or MS/MS is the process by which pre-selected ions in a sample can be fragmented into smaller pieces for further analysis. This technique is often used to sequence peptides or identify attached glycans.²⁹

1.3.1 Matrix-Assisted Laser Desorption/Ionization (MALDI)

MALDI is a technique which is used off-line from LC separation. Prior to MALDI analysis, samples must be mixed with a matrix such as 2,5-dihydroxybenzoic acid. After being spotted on a metal target, this matrix co-crystallizes with the analyte, and it is this which becomes ionization when the sample is irradiated with a laser, as shown in **Figure 1.3**.³⁰

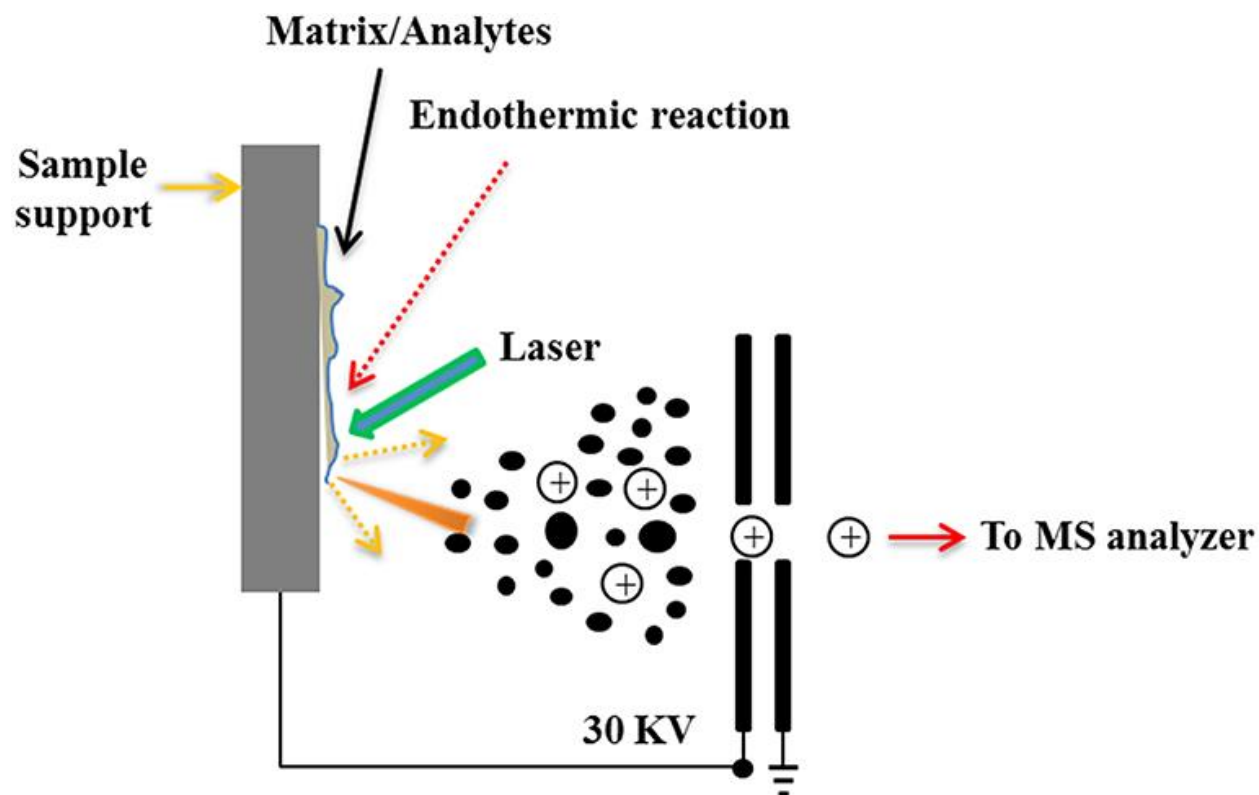


Figure 1.3 Diagram of the laser desorption/ionization process which occurs in a MALDI instrument. Amended from reference 30.³⁰

In a MALDI-TOF instrument, ions resulting from this process become accelerated and travel through a flight tube to the detector. The time in which it takes for an ion to reach a detector is transformed into a m/z ratio, and this is reported in the resulting spectrum. The Bruker ultrafleXtreme MALDI-TOF/TOF is an example of an instrument which uses a TOF/TOF analyzer, as depicted in **Figure 1.4**.³¹ It also uses a 2 kHz smartbeam-II laser, which is important to note in m/z calculations.³²

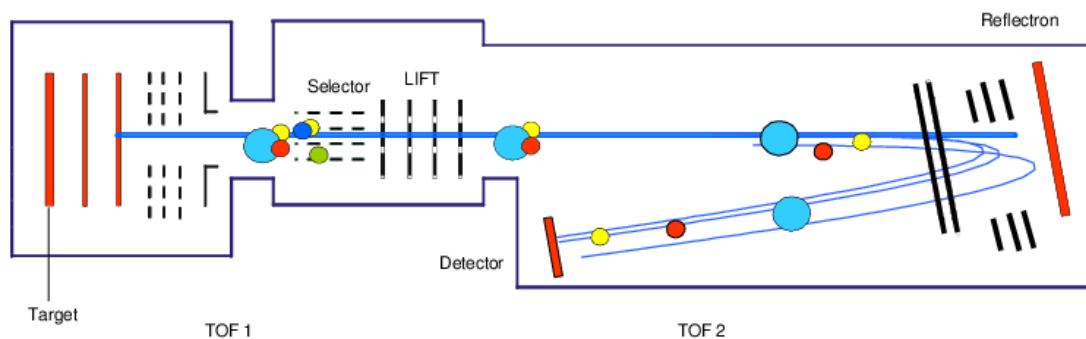


Figure 1.4 Schematic of a TOF/TOF mass analyzer. Amended from reference 31.³¹

MALDI is typically used in the analysis of small peptides and other compounds. In positive mode, ions are most commonly detected in a protonated form, denoted as $[M+H]^+$, though salt adducts with sodium and potassium may also exist. In negative mode, ions are detected as $[M-H]^-$ or as chlorine adducts. Occasionally, ions may be detected in doubly or singly protonated or deprotonated forms. The presence of a reflectron (**Figure 1.4**) allows for the use of reflector mode, which ensures higher mass accuracy and is therefore useful for small molecule analysis in particular. This instrument was used for some of the work reported in this thesis.

Using parameters such as the frequency of the applied laser and the length of flight tube, the instrument can translate information obtained from the ions into mass spectra displaying m/z values. MS/MS experiments can be carried out using LIFT mode in MALDI-TOF analysis. This mode involves a timed ion selector that separates ions based on their mass-dependent kinetic energy, resulting in the fragmentation of ions with a specific mass only.³³ This is the mode which is typically used in peptide sequencing due to its targeted fragmentation procedure.

1.3.2 ESI

ESI functions by use of an ion source, which vaporizes liquid-phase samples from LC and transforms them into gas-phase ions for MS analysis. This is done via a capillary needle, and this process is demonstrated in **Figure 1.5**.³⁰

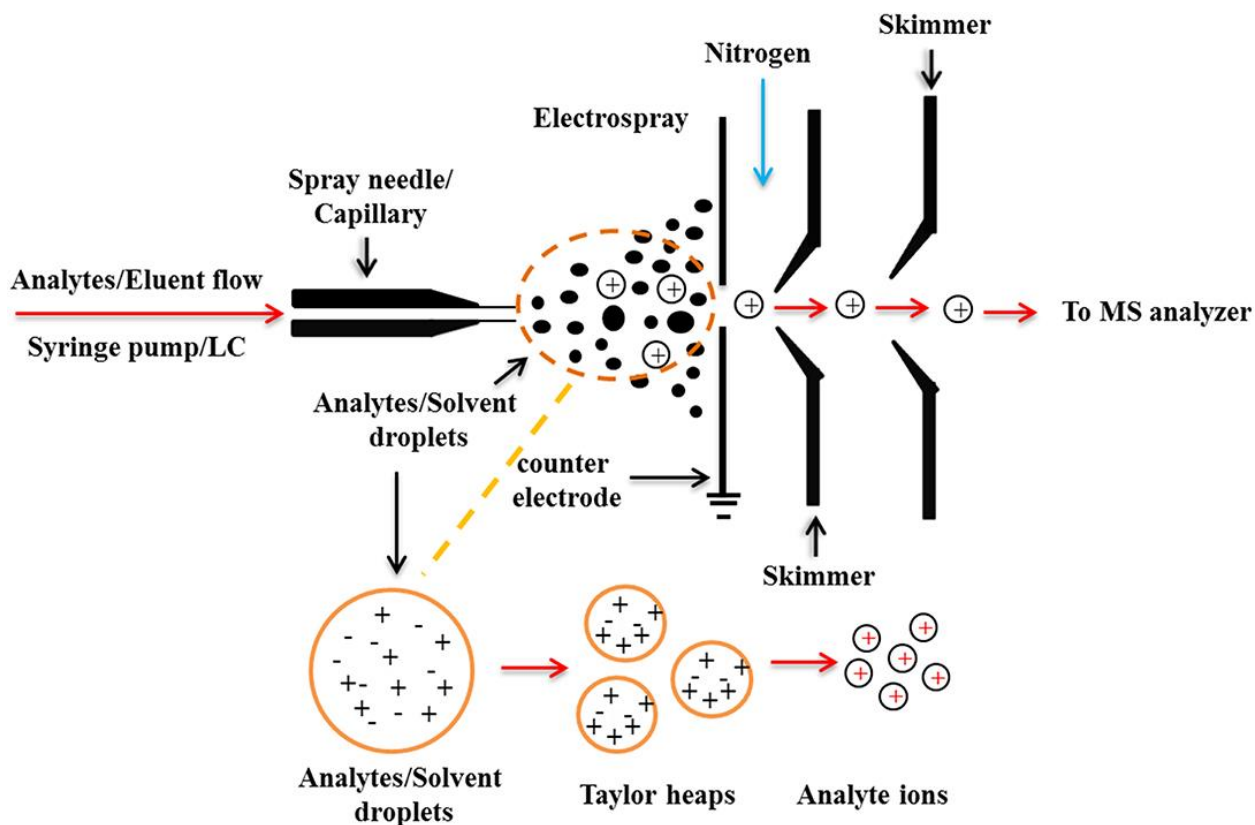


Figure 1.5 Diagram of ionization process in LC-ESI-MS. Amended from reference 30.³⁰

ESI is most commonly used with nano-flow LC instruments. This ionization method results in both singly and multiply charged droplets, which translates to singly and multiply charged ions. These are denoted as $[M+2H]^{2+}$, $[M+3H]^{3+}$, and so on. Based on these values, the molecular weight (MW) of the analyte can be calculated. The commonality of these multiply charged analytes is the primary reason for the large mass range of ESI, since multiply charged ions have much smaller m/z ratios.

As the eluent flows into the charged capillary needle, solvent evaporates, and analytes become ionized when the surface tension of the droplets is maximized. This is known as a

“Coulombic explosion” and is a soft-ionization technique. Before reaching the mass analyzer, skimmers are used to focus ions.³⁴

The ABSciex TripleTOF5600 mass analyzer is an example of an instrument which is commonly attached to a nano-flow LC and equipped with an ion source. This instrument combines a triple quadrupole with a TOF analyzer for higher mass accuracy. The arrangement of this hybrid analyzer is shown in **Figure 1.6**.³⁵ This instrument has been used for the work presented in this thesis.

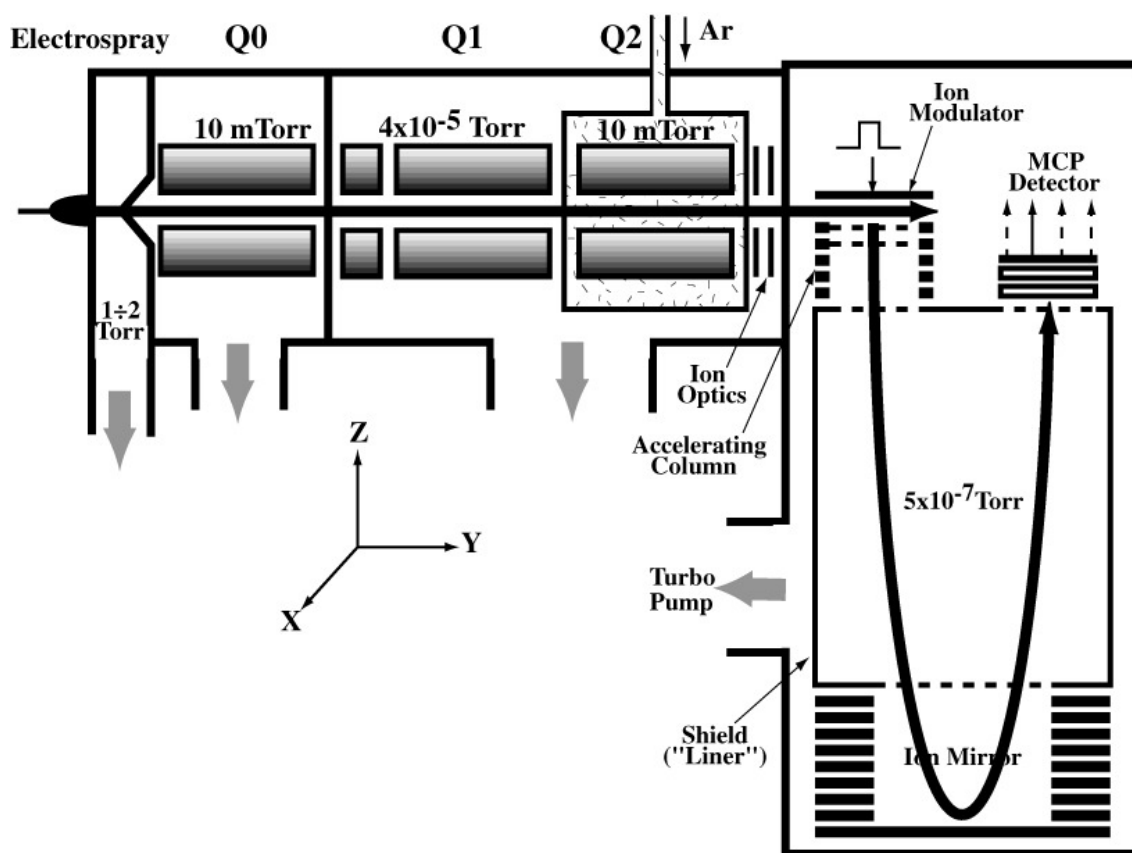


Figure 1.6 Schematic of a triple-quadrupole TOF hybrid mass analyzer. Amended from reference 35.³⁵

Triple-quad instruments offer an extra process of ion filtration compared to single-quad analyzers, as ions of interest can be isolated and fragmented separately. In MS/MS experiments, Q1 is used to select parent ions for further fragmentation in Q3. In this case, the equivalent of the third quadrupole is replaced with a TOF analyzer. Resulting MS/MS data is very useful in determining the identity of analytes in a sample as TOF analyzers offer high mass resolution.³⁵

Another example of an ESI-MS mass analyzer used in proteomics (and in this thesis) is the Thermo Fisher QExactive HF-X hybrid quadrupole-Orbitrap mass spectrometer. Unlike TOF analyzers, Orbitrap analyzers use a Fourier transformation to convert frequency information into m/z ratios. Ions are trapped in an electrostatic field in an orbital rotation in this mass analyzer, as shown in **Figure 1.7**.³⁶

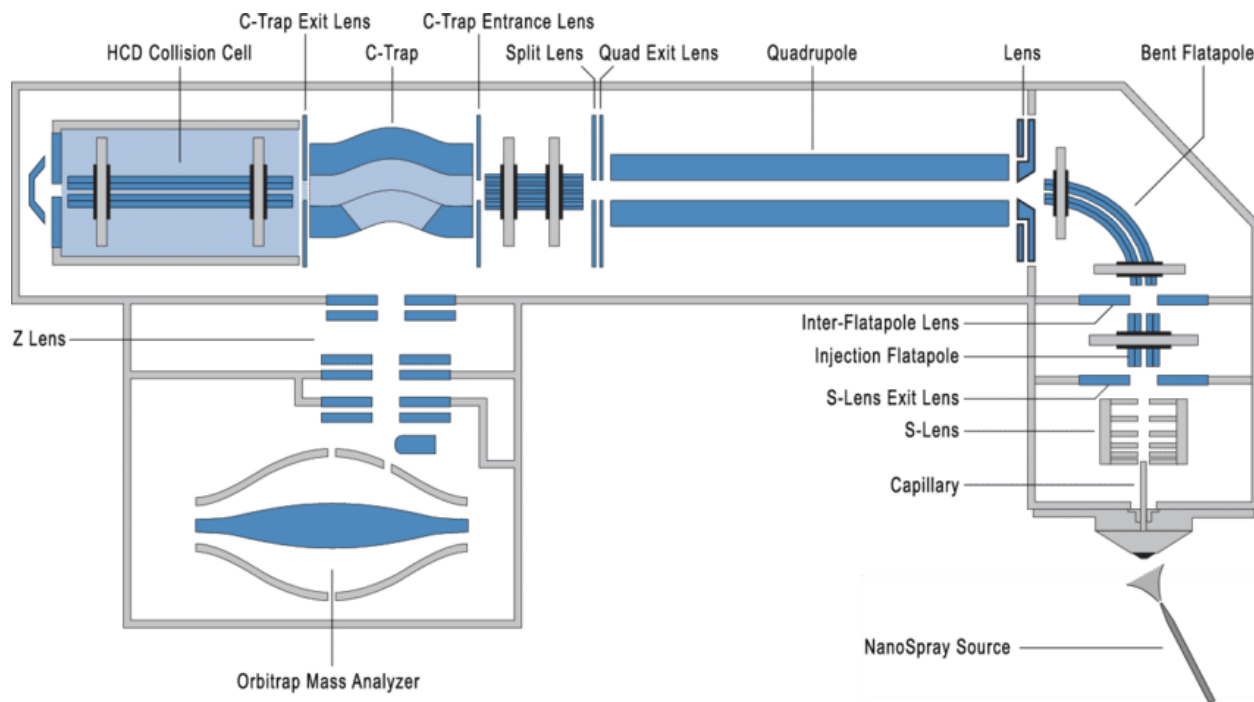


Figure 1.7 Diagram of the QExactive Orbitrap mass analyzer. Amended from reference 36.³⁶

The frequency at which an ion rotates is dependent on its m/z ratio, and therefore ions can be separated based on their mass while in orbit.³⁷ By using this frequency as a means to measure mass, this instrument offers high mass accuracy and high resolution but without the use of a magnet as in Fourier-transform ion cyclotron resonance MS.

Both TOF and Orbitrap mass analyzers may cause ions to fragment due to the strong acceleration necessary for injection, and so they are more useful for detection of ions which have already been fragmented in another analyzer.³⁷ This is the reason they are often part of hybrid instruments like those used in this thesis.

1.4 Biological molecules investigated in this thesis

Two of the chapters in this thesis focus on the analysis of specific biomolecules rather than whole-cell proteomic samples. The first chapter revolves around porcine IgG, as this antibody and its subclasses are important in understanding xenotransplantation and the concept of antibody-mediated rejection. The 11 different subtypes of porcine IgG differ in both peptide and glycopeptide structure, therefore both proteomic and glycoproteomic studies were conducted on these samples. As the primary difference between the glycoproteomic profiles of human and porcine IgG is the inclusion of unique glycans in porcine antibodies,³⁸ samples purified from both wild-type and knockout pigs were analyzed, where the knockout pigs are missing these unique sugars.

In the other chapter, both a hybrid β -hexosaminidase enzyme and its ganglioside substrate are analyzed. This enzyme, called HexM, was designed as a more compact version of HexA, which breaks down GM2 ganglioside in the lysosome.³⁹ The goal of HexM is to alleviate the neurological effects of patients with Tay-Sachs disease, which results from absent or defective HexA. Current treatment for this disease includes enzyme replacement therapy, however the administered native enzyme is too large to cross the blood-brain barrier, and therefore it cannot break down GM2 in neurons. Both HexA and HexM rely on the mannose-6-phosphate pathway for trafficking to the lysosome, therefore the presence of phosphorylated glycans on this enzyme is important. The study in this thesis which explores HexM focuses on both its phosphorylation and glycosylation using mass spectrometry.

Lastly, GM2 ganglioside extracted from the brains of mice is studied for relative quantification purposes. Gangliosides are glycosphingolipids which exist within lysosomes and vary based on the structure of their glycans. Sialic acid in particular is important in terms of gangliosides, as the M in GM2 refers to 'monosialo'. Ganglioside structures can be found in **Figure 1.2**.

1.5 Objectives

This thesis is separated into chapters corresponding to different projects and includes published works. In Chapter 2, the glycosylation profiles and peptide sequences of porcine IgG are explored and classified into subtypes using MALDI and LC-ESI-MS. This is done to further research in immunology, as the structure of these immunoglobulins is important in terms of xenotransplantation studies. It is also used to simplify the existing data on IgG subtypes in common databases.

In Chapters 3 and 4, 1D and 2D LC proteomic experiments are used to create adapted versions of the SSRCalc algorithm to account for different chromatographic conditions and peptide modifications. Chapter 3 examines the use of formic and acetic acid as ion pairing modifier in the mobile phase of LC experiments and compares their efficiency in both LC and MS. Chapter 4 deals with phosphorylated peptides and the effect of phosphorylation on retention time. Both studies discuss various sequence-specific peptide effects and the shifts in retention that occur as a result.

Chapter 5 focuses on Tay-Sachs disease and a hybrid enzyme which was developed as a potential treatment through enzyme replacement therapy. The modifications of both this enzyme and its substrate are analyzed by MALDI and ESI-MS and the importance of these modifications is emphasized. A procedure for the relative quantification of the substrate, GM2 ganglioside, is developed using MALDI and brain extracts from mice.

1.6 References

1. Catherman AD, Skinner OS, Kelleher NL. Top Down proteomics: facts and perspectives. *Biochem Biophys Res Commun.* **2014.** 445(4):683-693.
2. Lee J, Hwang H, Kim S, et al. Comprehensive Profiling of surface gangliosides extracted from various cell lines by LC-MS/MS. *Cells.* **2019.** 8(11):1323.
3. IUPAC-IUB Joint Commission on Biochemical Nomenclature. Nomenclature and symbolism for amino acids and peptides. *Eur. J. Biochem.* **1984.** 138:9-37.
4. Suttapitugsakul S, Xiao H, Smekens J, Wu R. Evaluation and optimization of reduction and alkylation methods to maximize peptide identification with MS-based proteomics. *Mol Biosyst.* **2017.** 13(12):2574-2582.
5. D. Shamshurin, V. Spicer, O.V. Krokhin. Defining intrinsic hydrophobicity of amino acids' side chains in random coil conformation. Reversed-phase liquid chromatography of designed synthetic peptides vs. random peptide data sets. *J. Chromatogr. A.* **2011.** 1218:6348-6355.
6. Ang E, Neustaeter H, Spicer V, Perreault H, Krokhin O. Retention Time Prediction for Glycopeptides in Reversed-Phase Chromatography for Glycoproteomic Applications. *Anal Chem.* **2019.** 91(21):13360-13366.
7. Wiedłocha A, Nilsen T, Wesche J, Sørensen V, Małecki J, Marcinkowska E, Olsnes S. Phosphorylation-regulated nucleocytoplasmic trafficking of internalized fibroblast growth factor-1. *Mol Biol Cell.* **2005.**16(2):794-810.
8. Ishihama Y, Wei FY, Aoshima K, Sato T, Kuromitsu J, Oda Y. Enhancement of the efficiency of phosphoproteomic identification by removing phosphates after phosphopeptide enrichment. *J Proteome Res.* **2007.** 6(3):1139-44.
9. S. Mohammed and A. J. Heck. Strong cation exchange (SCX) based analytical methods for the targeted analysis of protein post-translational modifications. *Current Opinion in Biotechnology.* **2010.**
10. Kim J, Petritis K, Shen Y, Camp DG 2nd, Moore RJ, Smith RD. Phosphopeptide elution times in reversed-phase liquid chromatography. *J Chromatogr A.* **2007.** 1172(1):9-18.

11. Hsu CC, Xue L, Arrington JV, et al. Estimating the Efficiency of Phosphopeptide Identification by Tandem Mass Spectrometry. *J Am Soc Mass Spectrom.* **2017.** 28(6):1127-1135.
12. Yang Y, Liu F, Franc V, Halim LA, Schellekens H, Heck AJ. Hybrid mass spectrometry approaches in glycoprotein analysis and their usage in scoring biosimilarity. *Nat Commun.* **2016.** 7:13397.
13. Čaval T, Zhu J, Tian W, Remmelzwaal S, Yang Z, Clausen H, Heck AJR. Targeted analysis of lysosomal directed proteins and their sites of mannose-6-phosphate modification. *Mol Cell Proteomics.* **2019.** 18(1):16-27.
14. Krautter F, Iqbal AJ. Glycans and glycan-binding proteins as regulators and potential targets in leukocyte recruitment. *Front. Cell Dev. Biol.* **2021.** 9:624082: e624082.
15. Breitling J, Aebi M. N-linked protein glycosylation in the endoplasmic reticulum. *Cold Spring Harb Perspect Biol.* **2013.** 5(8):a013359.
16. Boune S, Hu P, Epstein AL, Khawli LA. Principles of N-Linked Glycosylation Variations of IgG-Based Therapeutics: Pharmacokinetic and Functional Considerations. *Antibodies (Basel).* **2020.** 9(2):22.
17. Yang W, Shah P, Hu Y, et al. Comparison of Enrichment Methods for Intact N- and O-Linked Glycopeptides Using Strong Anion Exchange and Hydrophilic Interaction Liquid Chromatography. *Anal Chem.* **2017.** 89(21):11193-11197.
18. Ghidoni R, Sonnino S, Chigorno V, Venerando B, Tettamanti G. Occurrence of glycosylation and deglycosylation of exogenously administered ganglioside GM1 in mouse liver. *Biochem J.* **1983.** 213(2):321-329.
19. Vajn K, Viljetić B, Degmečić IV, Schnaar RL, Heffer M. differential distribution of major brain gangliosides in the adult mouse central nervous system. **2013.** *PLOS ONE* 8(9): e75720.
20. Ho CS, Lam CW, Chan MH, et al. Electrospray ionisation mass spectrometry: principles and clinical applications. *Clin Biochem Rev.* **2003.** 24(1):3-12.

21. Yeung D, Mizero B, Gussakovsky D, Klaassen N, Lao Y, Spicer V, Krokhin OV. Separation Orthogonality in Liquid Chromatography-Mass Spectrometry for Proteomic Applications: Comparison of 16 Different Two-Dimensional Combinations. *Anal Chem.* **2020.** 92(5):3904-3912.
22. Sun W, Wu S, Wang X, Zheng D, Gao Y. A systematical analysis of tryptic peptide identification with reverse phase liquid chromatography and electrospray ion trap mass spectrometry. *Genomics Proteomics Bioinformatics.* **2004.** 2(3):174-183.
23. Spicer V, Ezzati P, Neustaeter H, Beavis RC, Wilkins JA, Krokhin OV. 3D HPLC-MS with Reversed-Phase Separation Functionality in All Three Dimensions for Large-Scale Bottom-Up Proteomics and Peptide Retention Data Collection. *Anal Chem.* **2016.** 88(5):2847-55.
24. Buszewski B, Noga S. Hydrophilic interaction liquid chromatography (HILIC)--a powerful separation technique. *Anal Bioanal Chem.* **2012.** 402(1):231-247.
25. Wang Z, Ma H, Smith K, Wu S. Two-Dimensional Separation Using High-pH and Low-pH Reversed Phase Liquid Chromatography for Top-down Proteomics. *Int J Mass Spectrom.* **2018.** 427:43-51.
26. Krokhin OV, Spicer V. Peptide retention standards and hydrophobicity indexes in reversed-phase high-performance liquid chromatography of peptides. *Anal Chem.* **2009.** 81(22):9522-30.
27. Haag AM. Mass Analyzers and Mass Spectrometers. *Adv Exp Med Biol.* **2016.** 919:157-169.
28. Ou YM, Kuo SY, Lee H, Chang HT, Wang YS. An efficient sample preparation method to enhance carbohydrate ion signals in matrix-assisted laser desorption/ionization mass spectrometry. *J Vis Exp.* **2018.** (137):57660.
29. Swearingen KE, Eng JK, Shteynberg D, et al. A tandem mass spectrometry sequence database search method for identification of o-fucosylated proteins by mass spectrometry. *J Proteome Res.* **2019.** 18(2):652-663.

30. Chen G, Fan M, Liu Y, Sun B, Liu M, Wu J, Li N, Guo M. Advances in MS based strategies for probing ligand-target interactions: focus on soft ionization mass spectrometric techniques. *Front. Chem.* **2019**. 7:703: e00703
31. Pusch W, Flocco MT, Leung SM, Thiele H, Kostrzewa M. Mass spectrometry-based clinical proteomics. *Pharmacogenomics.* **2004**. 4:463: e22753.
32. Wang CC, Lai YH, Ou YM, Chang HT, Wang YS. Critical factors determining the quantification capability of matrix-assisted laser desorption/ionization- time-of-flight mass spectrometry. *Philos Trans A Math Phys Eng Sci.* **2016**. 374(2079):20150371.
33. Suckau, D., Resemann, A., Schuerenberg, M. *et al.* A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. *Anal Bioanal Chem.* **2003**. 376, 952–965.
34. 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**. 2012:282574.
35. Chernushevich IV, Loboda AV, Thomson, BA. An introduction to quadrupole–time-of-flight mass spectrometry. *J. Mass Spectrom.* **2001**. 36: 849-865.
36. Michalski A, Damoc E, Hauschild JP, Lange O, Wiegand A, Makarov A, Nagaraj N, Cox J, Mann M, Horning S. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol Cell Proteomics.* **2011**. 10:9: M111.011015.
37. Zubarev RA, Makarov A. Orbitrap is the newest addition to the family of high-resolution mass spectrometry analyzers. *Anal Chem.* **2013**. 85:11;5288-5296.
38. Salama A, Conchon S, Perota A, et al. Abstracts of the IPITA-IXA-CTS 2015 joint congress November 15-19, 2015, Melbourne, Australia. *Xenotransplantation.* **2015**. 22(S1): S2-S47.
39. Benzie G, Bouma K, Battellino T, Cooper S, Hemming R, et al. Increased phosphorylation of HexM improves lysosomal uptake and potential for managing GM2 gangliosidosis. *BBA Advances.* **2021**. 100032.

2 Chapter 2: Glycoproteomic study of porcine IgG subtypes by LC-MS/MS

Liquid chromatography–tandem mass spectrometry glycoproteomic study of porcine IgG and detection of subtypes

Taylor Battellino¹, Raymond Bacala¹, Baylie Gigolyk¹, Gideon Ong¹, Milan V. Teraiya¹, H  l  ne Perreault¹

¹Department of Chemistry, University of Manitoba, Winnipeg, Manitoba, Canada

Published:

Rapid Communications in Mass Spectrometry

February 3, 2021

Volume 35, Issue 9

<https://doi-org.uml.idm.oclc.org/10.1002/rcm.9063>

2.0 Contributions of Authors

BG, EA, TB and GO prepared the samples for MS. BG, HN and GO analyzed the samples by MS. RB wrote the Excel Macro used for sequence alignments. TB, GO, MVT, OK and HP worked on data treatment. HP and OK supervised all students involved in this work.

2.1 Abstract

Rationale: While high-throughput proteomic methods have been widely applied to monoclonal antibodies and human immunoglobulin gamma (IgG) samples, less information is available on porcine IgG. As pigs are considered one of the most suitable species for xenotransplantation, it is important to characterize IgG amino acid sequences and glycosylation profiles, which is the focus of this study.

Methods: Three different purified porcine IgG samples, including wild-type and knockout species, were digested with trypsin and enriched for glycopeptides. Digestion mixtures were spiked with a mixture of six standard peptides. Analysis was performed using electrospray ionization liquid chromatography–tandem mass spectrometry (MS/MS) in standard MS/MS data-dependent acquisition mode on a hybrid triple quadrupole time-of-flight mass spectrometer.

Results: To facilitate the classification of subtypes detected experimentally, UniprotKB database entries were organized using comparative alignment scores. Sequences were grouped based on 11 different subtypes as translated from GenBank entries. Proteomic searches were accomplished automatically using specialized software, whereas glycoprotein searches were performed manually by monitoring the extracted chromatograms of diagnostic MS/MS glycan fragments and studying their corresponding mass spectra; 40–50 non-glycosylated peptides and 4–5 glycosylated peptides were detected in each sample, with several glycoforms per sequence.

Conclusions: Proteomic analysis of porcine IgG is complicated by factors such as the presence of several subtypes, redundant heavy chain (HC) sequences in protein databases, and the lack of consistent cross-referencing between databases. Aligning and comparing HC sequences were necessary to eliminate redundancy. This study highlights the complexity of pig IgG and shows the importance of MS in proteomics and glycoproteomics.

2.2 Introduction

The glycosylation of proteins in swine has attracted considerable attention in recent studies that focus on xenotransplantation.¹⁻³ Antibody-mediated rejection constitutes an obstacle in this area of clinical application and research, and xenoantigens play an important role in this process. Well-characterized xenoantigens are galactose linked in α 1–3 (α Gal) and *N*-glycolyl neuraminic acid (Neu5Gc), a hydroxylated version of *N*-acetyl neuraminic acid (Neu5Ac).⁴ For instance, immune reactions induced by these specific glycosylation motifs were highlighted in an article by Yoon and coworkers,¹ who used anti- α Gal IgG (immunoglobulin gamma) and IgM as markers for pig to nonhuman transplantations with positive responses. Also, Li et al studied carbohydrate antigen expression and anti-pig antibodies in capuchin monkeys after grafts from triple knockout (KO) pigs (α Gal KO, CMAH-KO [CKO] and *Sda* antigen KO), where CMAH is cytidine monophosphate-*N*-acetylneuraminic acid hydroxylase and *Sda* antigen is glycosylation pattern containing *N*-acetylgalactosamine (GalNAc).² Indeed, knockout of pig β 1,4-*N*-acetylgalactosaminyltransferase 2 (β 1,4GalNAcT-II) led to a decrease in binding between human serum antibodies and pig cells.⁵

Polyclonal pig antibodies—which are glycoproteins—have been used in an attempt to prevent severe infectious diseases, although they have shown high immunogenicity toward humans.⁶ IgG from pigs bears α Gal and Neu5Gc. Reynard et al have investigated the effect of IgG molecules deprived of these deleterious epitopes in passive immunotherapy.⁶ Pig antibodies also hold promise in the development of therapeutic agents against viral diseases. With this aim, Dong et al isolated two pig IgG variants specific to the classical swine fever virus.⁷ The quantitative analysis of pig antibodies in this regard has become of importance, as emphasized in a report by Zhao et al who developed a new method for the detection of pig IgG based on fluorescence via an immuno-sensing strategy using selective pull-down by magnetic beads.⁸

For several years, there has been evidence that pig IgG is not homogeneous. Rather, it comprises subclasses, designated by some authors as subtypes. The term “subtypes” will be used for the rest of this article. Back in 1986, Mota published a monograph on IgG subtypes in animal species.⁹ During the same year, a study by Bokhout et al described the isolation of two pig IgG subtypes and discussed the possibility of the existence of more subtypes.¹⁰ A year later, Franek published a study on subtype heterogeneity of pig IgG, finding that functionally different IgG molecules may be isolated from sera collected at various stages of the immune response.¹¹ In

1992, Huang and coworkers explored the preferential mammary storage and secretion of IgG subtypes in swine¹² and in 2003, Crawley studied the IgG1:IgG2 “isotype” ratio in pigs in the presence of an extracellular bacterium, *Actinobacillus pleuropneumoniae*. Crawley also noted the different effects of all isotypes considered (IgG2a, 2b, 3, and 4) on the ability to activate complement C'.¹³

Butler's work on the genetics of pig IgG over the years has had a large impact on better defining the subtypes of pig IgG. In 2009 in a paper entitled “Porcine IgG: Structure, Genetics and Evolution,” he and his team reported 11 genomic porcine C γ gene sequences, representing 6 putative subtypes.¹⁴ In a later article, Butler and Eguchi-Ogawa sequenced the DNA of the whole immunoglobulin heavy constant (IGHC) region from a single porcine haplotype and found the genes for 7 subtypes, although their results were lined up against 12 subtypes previously entered in databases.¹⁵ Since then it has been demonstrated by Butler and Sinkora that in swine T-cells, Ig heavy chains (HCs) may be completely rearranged,¹⁶ which could explain a flexibility in the IGHG gene region and the lack of observation of four genes in a particular haplotype.

Given this complex genomic information, proteomic studies are not simple either. Our group published a study in 2016¹⁷ attempting to sequence the HC of wild-type (WT) pig IgG using mass spectrometry (MS) based on the proteins translated from the 11 gene subtypes presented by Butler et al in 2009¹⁴ and then updated by Eguchi-Ogawa et al in 2012.¹⁵ By performing *in silico* tryptic digestion, lists of tryptic peptides were generated for all potential IgG subtypes, and peptides unique to specific subtypes were used to confirm the presence of these subtypes in IgG samples. Glycosylation patterns were also of interest, showing both α Gal and Neu5Gc xenoantigens. In a second study by our group, the glycosylation of WT pig IgG HC was compared with that of the IgG from two KO pig species, α Gal KO and CKO.¹⁸

In an exhaustive study of glycosylation in similar KO pigs, Burlak et al used MS to identify all glycans detached enzymatically from IgG,¹⁹ but no specific glycosylation sites were identified. A third study by our group considered not only the HC but also glycosylation and proteomics of the whole WT pig IgG. The strategy used was to fragment IgG into Fab and Fc using papain hydrolysis before performing tryptic digestion and MS analysis. This approach, although it did not allow the identification of any glycans in the Fab portion, proved useful for specific site glycosylation analysis of Fc.²⁰ Proteomic searches were performed on pig serum

proteins by two different teams in other laboratories; however, this was after removal of IgG and albumin, and therefore, no information about IgG was available.^{21, 22} Marco-Ramell et al performed a proteomic study on pig IgG and IgA, and for these proteins their report focused more on glycosylation than protein sequencing.²³

The studies on proteomics and glycosylation of pig IgG performed by our group^{17, 18, 20} relied mostly on matrix-assisted laser desorption/ionization MS (MALDI-MS),²⁴ with sample preparation methods optimized for this technique.²⁵ Following MALDI-MS studies with manual data interpretation,^{17, 18} the goals of the present study are to (a) assess the ability of high-performance liquid chromatography (HPLC)/MS/MS coupled with an automatic proteomic database search tool to identify IgG subtypes from their representative peptides; (b) assess the potential of HPLC/MS/MS to detect Fc glycosylation relative to MALDI-MS; and (c) determine if HPLC/MS/MS allows the detection of other *N*- and/or *O*-glycosylation sites in the pig IgG proteins.

For this article, the workflow (**Figure 2.1**) was refined to be consistent for all samples analyzed: tryptic digestion followed by glycopeptide enrichment and then reversed-phase high-performance liquid chromatography (RPLC) coupled with MS and tandem MS (MS/MS). As the mass spectrometer conditions were optimized to produce useful fragmentation patterns for peptides and glycopeptides,²⁶ both proteomic and glycoproteomic approaches could be applied. It is clear that the glycopeptide enrichment protocol used in this work also pulled down a considerable number of peptides, some of which were identified as belonging to pig IgG. The search engine used could not directly identify glycopeptides due to mass increase brought by glycosylation. However, based on reproducible MS/MS glycopeptide fragmentation patterns, both peptide and glycan portions could be characterized in most cases.

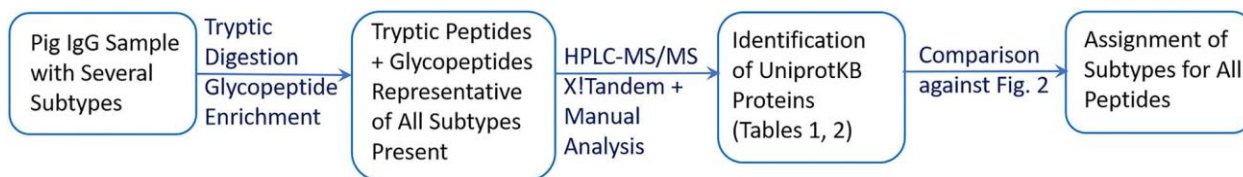


Figure 2.1 Workflow used to identify pig IgG subtypes in each sample analyzed in this study. Tables 2.1 and 2.2 and Figure 2.2 are described further in the text.

While manually searching for glycopeptide amino acid backbones identified in this study, it became obvious that some UniprotKB pig IgG HC entries are redundant or unspecific as to which subtype of pig IgG they correspond to. Using an in-house written Excel macro routine, we were able to screen through entries and eliminate some redundancy. This helped our search for specific subtypes through the observation of peptides that are unique to these subtypes. This information will enable better characterization of polyclonal pig IgG samples in future studies. As it was importantly pointed out by Vidarsson in an article on human IgG,²⁷ each subclass exhibits its own properties in terms of immune complex formation, antigen binding affinity, effector cell activation, transport into the placenta, complement activation, and half-life.

2.3 Experimental

2.3.1 Chemicals

Iodoacetamide (IAA), dithiothreitol (DTT), formic acid (FA), ammonium bicarbonate (NH_4HCO_3), and trifluoroacetic acid (TFA) were purchased from Sigma (St. Louis, MO, USA). Sequencing-grade trypsin was purchased from Promega (Madison, WI, USA). All solvents used were HPLC grade. Deionized water was obtained using a Milli-Q filtration system (EMD-Millipore, Burlington, MA, USA) supplied by a reverse-osmosis feedstock.

2.3.2 Pig IgG

Porcine IgG was obtained from Dr J. P. Soulillou's laboratory, Université de Nantes (France). Three IgG samples were investigated: WT, DKO, and CKO. Wild-type (WT) pigs used were of the Dutch Landrace Large White strain (25 kg) (EARL du Pont Romain, Surzur, France) and housed at the Large Animal Facility of the INSERM UMR1064 (Nantes University Hospital, France, agreement number: D44011). The animals were immunized against the human Jurkat T-cell line (clone E6.1, ATCC TIB-152). One subcutaneous injection was made, followed by two intravenous injections of 30×10^6 cells, and each injection was made under anesthesia at a 2-week interval. Whole blood was sampled at day 35 of the immunization protocol. Serum was stored at -80°C until further use for purification.¹⁷ Double knockout (DKO) pigs were obtained by targeting the swine CMAH gene using TALEN²⁸ on the GGTA12/2 background (obtained from Dr Soulillou through Dr D. H. Sachs, Massachusetts General Hospital, Boston, MA, USA)²⁹ and by somatic cell nuclear transfer.^{28, 30} The CKO pigs were generated in Dr C. Galli's laboratory³¹ by breeding DKO male founders to WT sows. As mutations are transmitted in a Mendelian fashion, these researchers selected CKO-only animals in the second generation. As such, Neu5Gc null females were bred to DKO boar counterparts from which CKO species were obtained.

2.3.3 Protein digestion and sample preparation

IgG (100 μg) was suspended in 100 mM ammonium bicarbonate. The samples were reduced (10 mM DTT, 56°C for 40 min) and alkylated with IAA (50 mM, room temperature, 45 min). Excess IAA was quenched by adding 16 mM DTT. Trypsin was used at a ratio of 1:50 enzyme:substrate for overnight digestion at 37°C . The ProteoExtract kit (EMD-Millipore) was

used for glycopeptide enrichment. The digestion mixtures were spiked with a mixture of six standard peptides P1–P6³² and subjected to LC/MS/MS analysis (totaling three samples). These peptides are used as retention time markers based on their hydrophobicity index (or percentage of acetonitrile in eluent) at the time of elution.²⁶

2.3.4 Chromatographic conditions

An Ultra LC system (Eksigent, Dublin, CA, USA) was used for nano-LC separations. Both eluents, A (water) and B (9:1 acetonitrile:water), contained 0.1% FA as ion pairing modifier. The system delivered buffers A and B through a 100 μm \times 200 mm analytical column packed with 3 μm Luna C18(2) (Phenomenex, Torrance, CA, USA) at a flow rate of 500 nL/min. Spiked samples were loaded on a 300 μm \times 5 mm PepMap 100 trap column (Thermo Fisher, San Jose, CA, USA). A linear increase from 0.4% to 31% buffer B in 77 min was observed, followed by 5 min at 80% B and 8 min at 0.4% B for equilibration of the column (90 min total analysis time).

2.3.5 MS and data analysis

A TripleTOF5600 mass spectrometer (ABSciex, Concord, Ontario, Canada) was used in standard MS/MS data-dependent acquisition mode. MS spectra (250 ms) were collected (m/z 370–1250), along with up to 19 MS/MS measurements, on the most abundant precursor ions (400 counts/s threshold, +2 to +5 charge state, m/z 100–1500 mass range for MS/MS, 100 ms each). Previously targeted parent ions were excluded from repetitive MS/MS acquisition for 12 s (50 milli m/z units mass tolerance).

2.3.6 Identification of proteins and peptides

Raw spectrum files were converted to Mascot generic format. The following search parameters for the identification of X!Tandem (<https://www.thegpm.org/tandem/>) were used for the identification of non-glycosylated peptides: 10 ppm and ± 50 milli m/z units mass tolerance for precursor and product ions, respectively; constant modification of Cys with IAA; potential modification of asparagine by deamidation, peptide identification confidence $\log(e) < -3$ (see note of **Table 2.1** for definition). For protein alignment comparisons, the ExPASy SIM tool was used (<https://web.expasy.org/sim/>) in conjunction with an Excel macro developed in-house using Visual Basic.

Table 2.1 Pig IgG heavy chain constant protein sections identified by X!Tandem from RPLC-MS/MS data for three different tryptic digestion samples run separately (WT, DKO, and CKO).

Rank	Log(<i>e</i>)	%Cov	Accession	Type	Match	Rank	%Cov	Log(<i>e</i>)	Accession	Type	Match
1	-425.6		L8B0U3	1a	100	9		-218.1	L8AXL9	2b	99.1
1	-490.6	47	L8B0U3	1a	100	9		-258.2	L8B0Y0	2b	100
1	-500.2		L8B0U3	1a	100	8	28	-213.0	L8B0Y0	2b	100
2	-417.0		L8B0U8	1a	100	8		-229.3	L8B0Y0	2b	100
2	-473.9		L8B0U8	1a	100	10		-238.6	L8B180	2b	96.3
4	-382.3		L8B139	1a	99.4	11	22	-155.2	L8B130	3	97.6
3	-450.0		L8B139	1a	99.4	10		-194.5	L8B130	3	97.6
5	-379.1		L8B173	1a	97.9	7		-317.1	L8B0W0	4a	100
2	-420.0		L8B173	1a	97.9	6		-332.9	L8B0W9	4a	96.4
5	-347.0		L8B173	1a	97.9	4	32	-275.8	L8B0W9	4a	96.4
3	-409.0		L8B165	1b	99.7	4		-402.6	L8B0W9	4a	96.4
3	-399.8	36	L8B165	1b	99.7	12		-117.3	L8B0S2	5a	92.9
8	-263.1		L8B0W5	2a	100	10	28	-164.8	L8B0S2	5a	92.9
6	-239.3	33	L8B0W5	2a	100	11		-188.6	L8B0S2	5a	92.9
7	-265.8		L8B0W5	2a	100	7	32	-230.5	L8B0R9	6a	97.8
5	-253.0		L8B0X5	2a	99.1	6	25	-336.6	L8B0R9	6a	97.8

Notes. Log(*e*) is log of the expectation that any particular protein assignment was made at random; type is subtype from Eguchi-Ogawa et al.¹⁵ that matched the accession number the best; Match is alignment percentage as determined by the ExPasy SIM tool and compiled using in-house Excel macro. %Cov is protein coverage through peptides detected.

Abbreviations: DKO, double knockout; MS/MS, tandem mass spectrometry; RPLC, reversed-phase high-performance liquid chromatography; WT, wild type.

2.3.7 Identification of glycopeptides

The elution of glycopeptides was monitored by the m/z 204.08 ± 0.05 trace corresponding to GlcNAc⁺ ions obtained by collision-induced dissociation in MS/MS. The spectra of all peaks on that trace were characterized manually, and where possible the peptide and glycan portions were identified by mass, sequence, and/or composition.

2.4 Results and Discussion

2.4.1 General comments on previous studies by this group

Two previous studies on WT pig IgG used MALDI-MS¹⁷ and HPLC/MS/MS²⁰ and did not offer comparison between WT and KO IgG samples. These studies used manual data interpretation. A third study¹⁸ compared IgG from WT, GTKO, and DKO animals using only MALDI-MS and focused on glycosylation profiles with manual data interpretation, not on glycoproteomics. The present study is the first to conduct a full proteomic study on different types of porcine IgG (WT, DKO, and CKO) with a combination of automatic proteomic searches and manual interpretation enabled by diagnostic MS/MS product ions. The results found previously were mostly reproduced in this present study, which provides much additional and complementary proteomic/glycoproteomic information.

2.4.2 Proteomics of IgG HCs

It is important to note that samples investigated in this study were prepared by isolating glycopeptides from full tryptic digests, that is, they were originally meant only for a glycoproteomic experiment. A first round of visual inspection of the LC/MS/MS data revealed the presence of many non-glycosylated tryptic peptides in each sample, so it was decided to conduct proteomic searches. Glycopeptide enrichment methods always involve a carryover of peptides,²⁵ which in this case were sufficient in number and abundance to provide protein identification. Proteomic data obtained from the RPLC/MS/MS analyses of WT, DKO, and CKO pig IgG tryptic digest samples were compiled and compared by listing all proteins identified by X!Tandem. This software uses tryptic peptide MS/MS fragmentation data to sequence the peptides and match to proteins listed in the UniprotKB database (<https://www.uniprot.org>). Most identification hits corresponded to HC IgG (molecular mass ~50 kDa), and several accession numbers were obtained for these. X!Tandem used the most recent accession numbers, as there are many older redundant entries in the database. **Table 2.1** presents the results for IgG HC (gamma) components, with different colors used for WT (black), DKO (red), and CKO (blue) results. Abundances are higher for low-ranking numbers.

For WT IgG, eight subtypes were observed in this experiment: IgG1a, 1b, 2a, 2b, 3, 4a, 5a, and 6a. For the CKO sample, the same subtypes except for IgG 1b were detected. For the DKO IgG sample, 1a, 1b, 2a, 2b, 4a, 5a, and 6b were observed. The only subtypes not

characterized in the three samples were IgG4b and 5b, in accordance with the observations of Eguchi-Ogawa et al.¹⁵ These authors had observed an additional IgG6c allelic variant, which was not detected here. In the present study, assignment of subtypes was achieved according to the workflow shown in **Figure 2.1**, as explained later in the text.

A similar variation in the populations of subtypes in different samples had been noted in a previous study of WT and DKO IgG using MALDI-MS, where qualitatively IgG6a had low abundance in DKO relative to WT.¹⁸ This was determined by measuring the relative abundance of tryptic glycopeptide EAQFNSTYR originating from Ig6a only, as other subtypes contain the sequence EEQFNSTYR.^{17, 18} As the X!Tandem search engine targets the fragmentation of bare peptides and not of glycopeptides, the latter cannot contribute to protein identification in the **Table 2.1** search and are treated separately and manually in this study. This will be described in a later section.

The ranking numbers in **Table 2.1** place the proteins in order of $\log(e)$, which is the log of the expectation that any particular protein assignment was made at random. As protein identification depends on the quality of MS/MS data obtained, where signal-to-noise ratios generally increase with concentration, it is reasonable to assume a decreasing concentration order with increasing ranking position; however, **Table 2.1** is not strictly quantitative. Some percentage coverage values were calculated only as indicators.

It is interesting to observe that four accession numbers were reached for IgG1a, with 100%, 100%, 99.4%, and 97.9% alignment with the IgG1a sequence translated from GenBank (<http://www-ncbi-nlm-nih-gov.uml.idm.oclc.org/genbank/>) and Eguchi-Ogawa et al.¹⁵ and used by us in Lopez et al.¹⁷ Similar comments apply to IgG2a, 2b, and 4a, which all count more than one accession number. The GenBank protein accession numbers did not provide direct access to their UniprotKB equivalents. For example, for IgG6a's GenBank protein sequence (#ABY85807.2), there was no corresponding protein in UniprotKB. The protein listed in **Table 2.1** as IgG6a is L8B0R9 and has 97.8% similarity with GenBank's ABY8507.2. The latter was translated from a DNA sequence and the former from an mRNA sequence, and no direct protein characterization of these two entries was found in the literature.

As for L8B0U3, which is identical at 100% to GenBank's IgG1a (AAA52219.1), there is no link between the two accession numbers from the different databases. AAA52219.1 has a

cross-reference in the European Molecular Biology Laboratory (EMBL) database but not in UniprotKB. L8B0U3 is cross-referenced to GenBank protein BAM75543, which lines up perfectly with L8B0U3 (IgG1a) but has 142 extra amino acid residues on the N-terminus, as it is the IgG1a HC precursor. From this observation it followed that most L8B UniprotKB entries, when cross-referenced to GenBank, pointed to IgG HC precursors identified at the nucleotide level only and not to HC-only sequences such as those published in Eguchi-Ogawa et al.¹⁵ As the X!Tandem used a built-in fasta file compiled with UniprotKB data (latest in the *Sus scrofa* proteome), for this study our search was limited to comparisons based on this fasta file. In the future, similar searches could be conducted with fasta files built with *Sus scrofa* proteins featured in GenBank, EMBL, or another database.

Figure 2.2 compiles all alignment values between HC proteins found in these experiments and between these proteins and GenBank translated sequences. This table was built using a home-written Excel macro that called upon the ExPasy SIM alignment tool (see “Experimental” section). It shows that low alignment values are obtained between different subtypes, for example, 84.1% between IgG5a and 6a, whereas closer subtypes, for example, IgG6a and 6b, have 92.7% similarity. “Match” values featured in **Table 2.1** correspond to those highlighted in red in **Figure 2.2**, that is, the highest-matching numbers between Uniprot and GenBank sequences of concern in this work.

	L8AXL9	L8B0R9	L8B0S2	L8B0U3	L8B0U8	L8B0W0	L8B0W5	L8B0W9	L8B0X5	L8B0Y0	L8B130	L8B139	L8B165	L8B173	L8B180	IgG1a	IgG1b	IgG2a	IgG2b	IgG3	IgG4a	IgG4b	IgG5a	IgG5b	IgG6a	IgG6b	
L8AXL9	100.0																										
L8B0R9	85.0	100.0																									
L8B0S2	79.1	82.3	100.0																								
L8B0U3	85.2	85.9	84.0	100.0																							
L8B0U8	84.5	84.3	81.6	93.2	100.0																						
L8B0W0	87.9	87.8	80.0	86.4	86.2	100.0																					
L8B0W5	91.7	87.7	82.7	88.9	86.6	90.6	100.0																				
L8B0W9	86.9	87.8	80.7	85.4	83.5	89.0	89.0	100.0																			
L8B0X5	92.4	86.7	80.7	86.9	85.9	88.0	92.8	88.8	100.0																		
L8B0Y0	90.9	85.8	80.0	86.7	86.4	88.1	92.6	88.2	92.4	100.0																	
L8B130	79.3	81.2	78.5	81.4	78.7	79.5	82.6	80.4	81.5	79.9	100.0																
L8B139	83.5	83.5	80.4	91.9	91.8	85.2	85.6	83.5	84.6	84.1	78.7	100.0															
L8B165	82.6	84.4	81.9	90.2	89.5	85.1	84.9	82.9	83.4	83.1	78.9	90.3	100.0														
L8B173	85.2	84.9	83.2	93.6	91.7	87.0	88.1	84.4	85.7	86.5	80.4	90.2	91.3	100.0													
L8B180	85.1	90.9	81.2	85.3	83.1	84.2	86.8	85.5	86.3	85.1	79.9	82.9	82.2	84.2	100												
IgG1a	89.9	86.6	85.1	100.0	100.0	90.5	89.9	88.4	90.2	90.2	80.2	99.4	95.7	97.9	87.2	100											
IgG1b	87.5	86.0	84.5	96.0	96.0	88.4	87.5	87.5	87.8	87.8	80.2	95.4	99.7	97.6	86.2	96.0	100.0										
IgG2a	98.8	89.6	82.3	89.9	89.9	94.8	100.0	93.3	99.1	99.1	81.7	89.3	87.5	89.3	89.6	89.9	87.5	100.0									
IgG2b	99.1	89.9	82.0	90.2	90.2	95.7	99.1	93.6	99.4	100.0	81.4	89.6	87.8	90.2	89.9	90.2	87.8	99.1	100								
IgG3	81.7	80.2	75.4	80.8	80.8	82.0	81.7	81.6	82.0	81.4	97.6	80.5	81.1	80.5	79.5	80.8	81.4	81.7	81.4	100							
IgG4a	94.2	89.2	80.1	89.5	89.5	100.0	94.9	96.4	94.6	94.9	81.2	88.8	87.0	89.2	87.3	89.5	87.0	94.9	94.9	82.3	100						
IgG4b	92.8	90.6	82.1	88.7	88.7	97.2	92.8	95.9	93.1	93.1	82.4	88.1	90.9	88.4	88.7	88.7	90.9	92.8	93.1	82.7	97.4	100.0					
IgG5a	80.8	83.3	92.9	87.2	87.2	82.0	81.4	82.0	81.1	81.1	75.7	86.6	86.9	87.2	83.3	87.2	86.9	81.4	81.1	75.4	79.8	81.4	100.0				
IgG5b	80.5	79.6	89.6	83.3	83.3	79.9	80.5	77.7	80.8	80.8	74.3	82.7	83.3	84.0	81.1	83.3	83.3	80.5	80.8	75.2	79.4	78.9	90.3	100.0			
IgG6a	91.5	97.8	85.0	88.4	88.4	92.1	91.5	93.9	91.8	91.8	82.4	87.7	87.1	87.7	94.9	88.4	87.4	91.5	91.8	81.1	91.4	91.5	84.1	79.6	100.0		
IgG6b	88.0	94.1	86.2	86.5	86.5	87.4	88.0	88.8	88.3	88.3	79.1	86.2	85.2	85.8	96.3	86.5	85.5	88.0	88.3	78.5	85.4	86.8	84.4	83.0	92.7	100.0	

Figure 2.2 Sequence alignments for all proteins listed in Table 1.

As proteins were identified based on the MS sequencing of their tryptic peptides, subtle differences in data leading to the identification of, for example, L8BOU3 versus L8B0U8 for IgG1a are provided in **Supplementary Table 2.1** (supporting information), where peptides are classified by the corresponding IgG subtype. This table provides detailed information on all detected peptides pertaining to proteins in **Table 2.1**, although it does not offer a fair representation of all HC gamma proteins identified in these experiments. Indeed, these HCs have several peptides in common, but only one protein is listed per peptide, that is, the protein with the highest score in the ensemble. This is the output format used by X!Tandem for conciseness. For more details, the X!Tandem output must be consulted one protein at a time.

When comparing L8B0U3 and L8B0U8 IgG1a HCs, they have most identified peptides in common except for a few, for example, the substitution at position 23 (leucine vs valine, difference in mass of 14 Da) LVESGGGLVQPGGSLR (L8B0U3)/VVESGGGLVQPGGSLR (L8B0U8). Another example of variation is the detection of peptide (K)DLPAPITR for IgG4a versus VNNVDLPAPITR for IgG1a, where the latter does not have K to allow tryptic hydrolysis.

Overall, these results point to the complexity of pig IgG samples, as not only does each IgG sample contain several subtypes, but each subtype itself has variants due to the substitution of amino acids. Such detailed analysis is possible with the high performance of MS measurements.

2.4.3 Proteins other than from IgG HCs

A number of other proteins identified in the searches described earlier were examined. **Table 2.2** provides a list of these proteins with a brief description from Uniprot in order of relevance. Most proteins in the first half of the hits were “uncharacterized” in the database, although some were described as having a role in the immune response or as Ig-like domains. The pig IgG samples were purified in Dr Soulillou's laboratory by protein A affinity chromatography, and these uncharacterized proteins could have been pulled down along with IgG.¹⁷ Also, some proteins identified in the searches may originate from constant Fab sections of IgG or from IgA/IgM. The fact that A0A286Z190 identified by X!Tandem (second line in **Table 2.2**) is described as an uncharacterized protein leads to confusion, especially because this same sequence (100% alignment) had been entered at an earlier date as K7ZJP7 and clearly

described as pig IgM. It was then attempted to align the uncharacterized IgG-like proteins with IgA and the constant lambda section of pig IgG (accession number P01846); however, this exercise was not successful.

Table 2.2 List of proteins other than IgG heavy chains identified by X!Tandem.

Rank	Accession number	Description	M_r (kDa)	Peptides with possible ^N -glycosylation site (M + H) ⁺
9	A0A287A8U7	IgG-like domain	23.1	
12	A0A286ZI90	Uncharacterized (membrane protein)	67.1	NSSK (434.21); ESLNISWTR (1105.56); VSSQNIQDFPSVLR (1589.82); TSIVFSEIYANGTFGA (1676.82)
13	A0A287AE24	Ig-like domain	23.5	
13	A0A287B0L7	Uncharacterized protein (triggering of complement)	14.1	
14	A0A287A1M4	Ig-like domain	26.9	
14	A0A287BF37	Ig-like domain	12.8	
15	A0A287A5Q8	Ig-like domain	12.9	
15	A0A287ALS7	Ig-like domain	12.5	SSANATVM (779.35)
16	A0A287AV26	Uncharacterized protein (triggering of complement)	16.4	
17	A0A075B7I6	Uncharacterized (immune response)	14.6	
18	A0A075B7J0	Uncharacterized (immune response)	10.2	
18	A0A287B3W7	Uncharacterized protein (triggering of complement)	10.7	
19	A0A287ALJ6	Uncharacterized protein (triggering of complement)	17.1	
20	A0A286ZWZ0	Uncharacterized protein (triggering of complement)	13.3	
20	A0A287AU93	Uncharacterized protein (triggering of complement)	16	
21	I3L728	Uncharacterized protein (triggering of complement)	15	
22	A0A075B7I9	Uncharacterized protein (triggering of complement)	12.9	SSANATVM (779.35)
23	A0A287AY91	Spermatogenesis-defective protein 39 homolog isoform	251.6	TKGDEEEYWNSK (1571.67); IDAVLNSSQIR (1214.66)
24	A0A287ANU5	Uncharacterized protein (triggering of complement)	15.7	
25	A0A075B7I5	Uncharacterized protein (triggering of complement)	114	
26	P83049	D-Xylulose reductase (xylose metabolism)	11.9	
26	A0A286ZLN6	Ig-like domain	102	
26	A0A287BEL8	Uncharacterized (zinc finger)	14.8	DNDTK (591.25); ALEQQLNNTTR (1415.70); VAVIDTEGLLGATVNLSQR (1955.07); SHTLNHTFFGR (1315.64)
27	I3LPU3	Uncharacterized (GTPase activator activity)	16.5	
28	A0A287AUT1	Uncharacterized (semaphorin receptor activity)	80.3	
28	F1RYP5	Uncharacterized protein (RNA binding)	162	
29	F1RM54	Uncharacterized (GTPase activator activity)	56.2	

29	A0A287ADH7	Olfactory receptor	203	NLSQAGR (744.39)
29	A0A287ALF1	Uncharacterized (multifunction)	105	
30	A0A287BL72	Beta-defensin (innate immune response against bacteria)	168	
31	A0A287BSV2	Uncharacterized protein (protein folding)	35	VCLWLVALLAANLSIR (1812.06)
31	A0A287AKC9	Ig-like domain	386	
31	F1RL06	Uncharacterized (antigen binding)	8.6	
32	A0A287BAI9	Myb/SANT DNA binding domain containing 1	20.3	ARNWTDAEM (1092.47); ITNMTFQYR (1172.56)
33	F1RPE5	Pescadillo homolog (ribonucleoprotein complex)	11	

Note. The third column shows potential tryptic *N*-glycopeptide backbones in these proteins, with their $[M + H]^+$ m/z values.

Many IgG-unrelated proteins are also listed toward the second half of **Table 2.2**. Whether or not proteins were relevant to IgG, most were examined in detail for possible *N*-glycosylation sites (consensus sequence NXT or NXS, where X = any amino acid except P). The corresponding tryptic peptide backbones are provided in **Table 2.2** along with their theoretical $[M + H]^+$ m/z values. Some of these will be referred to in the next section that focuses on the glycoproteomic aspect of this study.

Overall, X!Tandem gave high ranks to IgG (**Table 2.1**) relative to other proteins (**Table 2.2**), suggesting lower amounts of the latter. As no quantitation was performed in this study, at this stage the level of “contamination” of IgG samples by the proteins in **Table 2.2** cannot be determined with certainty.

2.4.4 *N*-Glycoproteomics

For IgG from WT and DKO animals, the predicted glycoforms were found. WT IgG contained complex glycans with Neu5Gc and α -Gal, and DKO IgG had Neu5Ac and no α -Gal. However, for the CKO IgG sample, no sialylation was observed, and no α -Gal patterns were found. We attributed this to lower sample concentration and low signals of sialylated and α -Gal glycopeptides from IgG in general.

The glycopeptides were identified manually and are listed in **Table 2.3**. Identification was conducted by tracing the MS/MS ion chromatogram of m/z 204.08 ± 0.025 produced upon the fragmentation of all *N*-glycopeptides. **Figure 2.3** shows the total ion current (TIC) obtained for the analysis of the CKO sample (top) aligned with the selected ion chromatogram (SIC) for MS/MS products at m/z 204 (bottom). It is clear by comparing the two chromatograms that

glycopeptides were not predominant in this sample, as was also the case for WT and DKO samples (not shown). Most peaks visible in the TIC correspond to non-glycosylated tryptic peptides. Although the ionization efficiency of a glycopeptide is close or equivalent to that of its bare peptide,³³ splitting one peptide into several glycoforms— thus into several peaks— reduces the signal observed for each particular glycoform.

Table 2.3 List of the most abundant glycopeptides observed in this study.

Retention time (min)	m/z	z	M + H tot	M + H pep	Sequence	Mass glycan	Glycan id	Notes
15.035	1014.9149	2	2028.8225	787.3770	FNSTYR	1241.4455	G0F-N	N = GlcNAc
17.494	749.9748	3	2247.9098	787.0000	FNSTYR	1460.9098	G1	
17.518	852.6794	3	2556.0236	787.3713	FNSTYR	1768.6523	G2FS	S = NeuGc, WT
17.526	798.6614	3	2393.9696	787.3765	FNSTYR	1606.5931	G1F	
17.547	695.9582	3	2085.8600	787.3842	FNSTYR	1298.4758	G0	
17.563	1095.9501	2	2190.8929	787.3750	FNSTYR	1403.5179	G1F-N	N = GlcNAc
17.634	744.6425	3	2231.9129	787.3763	FNSTYR	1444.5366	G0F	
17.694	1116.4629	2	2231.9185	787.3789	FNSTYR	1444.5396	G0F	
17.699	1197.4905	2	2393.9737	787.3896	FNSTYR	1606.5841	G1F	
17.855	744.6425	3	2231.9129	787.3763	FNSTYR	1444.5366	G0F	
18.780	841.3478	3	2522.0288	915.4314	QFNSTYR	1606.5974	G1F	
18.904	787.3285	3	2359.9709	915.4335	QFNSTYR	1444.5374	G0F	
19.340	1025.4366	3	3074.2952	1467.6499		1606.6453	G1F	
19.425	971.4181	3	2912.2397	1467.6990		1444.5407	G0F	
19.522	922.7305	3	2766.1769	1467.6940		1298.4829	G0	
19.886	955.0435	3	2863.1159	787.3896	FNSTYR	2075.7263	G2FS	S = NeuGc, WT
20.021	901.0238	3	2701.0568	787.3896	FNSTYR	1913.6672	G1FS	S = NeuGc, WT
20.179	949.7082	3	2847.1100	787.3632	FNSTYR	2059.7468	G2FS	S = NeuAc, DKO
20.407	895.6912	3	2685.0590	787.3670	FNSTYR	1897.6920	G1FS	S = NeuAc, DKO
20.542	967.4163	3	2900.2343	1455.6839		1444.5504	G0F	
20.574	1021.4341	3	3062.2877	1455.6847		1606.6030	G1F	
20.685	962.0568	3	2884.1560	1115.5160	EAQFNSTYR	1768.6400	G2F	
21.890	908.0402	3	2722.1060	1115.5157	EAQFNSTYR	1606.5903	G1F	
21.920	1034.7262	3	3104.1620	1173.5140	EEQFNSTYR	1930.6532	G3F	Alpha-Gal, WT
21.950	981.3848	3	2942.1400	1173.5140	EEQFNSTYR	1768.6426	G2F	
21.995	927.3776	3	2780.1182	1173.5162	EEQFNSTYR	1606.6020	G1F	
22.015	854.0237	3	2560.0565	1115.5170	EAQFNSTYR	1444.5395	G0F	

22.104	1207.8898	2	2414.7723	1173.5235	EEQFNSTYR	1241.2488	G0F-N	N = GlcNAc
22.237	873.3592	3	2618.0630	1173.5165	EEQFNSTYR	1444.5465	G0F	
22.274	878.6896	3	2634.0542	1173.5109	EEQFNSTYR	1460.5433	G1	
22.284	890.9967	3	2670.9755	1173.5200	EEQFNSTYR	1497.4555	G1-N + 3S	S = sulfate
22.289	927.3776	3	2780.1182	1173.5162	EEQFNSTYR	1606.6020	G1F	
22.486	824.6742	3	2472.0080	1173.5268	EEQFNSTYR	1298.4812	G0	
22.496	890.9967	3	2670.9755	1173.5253	EEQFNSTYR	1497.4502	G1-N + 3S	S = sulfate
22.558	1207.9965	2	2414.9857	1173.4800	EEQFNSTYR	1241.5057	G0F-N	
22.602	946.3842	3	2837.1380	1230.5443	EEQFNSTYR	1606.5937	G1F	Overalkylation of peptide
22.617	873.6927	3	2619.0635	1174.5290	EEQFNSTYR	1444.5345	G0F	Deamidation of peptide
22.780	892.3654	3	2675.0816	1230.5429	EEQFNSTYR	1444.5387	G0F	Overalkylation of peptide
24.181	1038.4081	2	2075.8089	631.2727	FNSTY	1444.5362	G0F	
24.261	692.6000	3	2075.7854	631.2709	FNSTY	1444.5145	G0F	
24.299	1010.4083	3	3029.2105	1115.5160	EAQFNSTYR	1913.6945	G1FS	S = NeuGc, WT
24.370	1083.7603	3	3249.2663	1173.5139	EEQFNSTYR	2075.7524	G2FS	S = NeuGc, WT
24.416	813.0714	4	3249.2638	1173.5170	EEQFNSTYR	2075.7468	G2FS	S = NeuGc, WT
24.638	1029.7438	3	3087.2168	1173.5204	EEQFNSTYR	1913.6964	G1FS	S = NeuGc, WT
27.169	994.4012	3	2981.1890	1212.5490	EEQFNSTYR	1768.6400	G2F	Overalkylation, dehydration of peptide
27.245	940.3814	3	2819.1296	1212.5356	EEQFNSTYR	1606.5940	G1F	Overalkylation, dehydration of peptide
27.511	886.3651	3	2657.0807	1212.5325	EEQFNSTYR	1444.5482	G0F	Overalkylation, dehydration of peptide
27.556	904.0007	3	2709.9875	1212.5326	EEQFNSTYR	1497.4549	G1-N + 3S	Overalkylation, dehydration (pep), S = sulfate
27.880	867.0287	3	2599.0715	1154.5264	EAQFNSTYR	1444.5451	G0F	Overalkylation, dehydration of peptide
27.973	837.6783	3	2511.0203	1212.5379	EEQFNSTYR	1298.4824	G0	Overalkylation, dehydration of peptide
29.006	867.3569	3	2600.0561	1155.5106	EEQFNSTYR	1444.5455	G0F	Dehydration of peptide
30.887	908.0274	3	2722.0676	1115.5179	EAQFNSTYR	1606.5497	G1F	
40.784	1055.7705	3	3165.2969	1105.5679	ESLNISWTR	2059.7290	G2FS	S = NeuAc (peptide from pig IgM), DKO
49.167	1233.2000	3	3697.5854	1515.7927		2181.7927		NeuAc, DKO

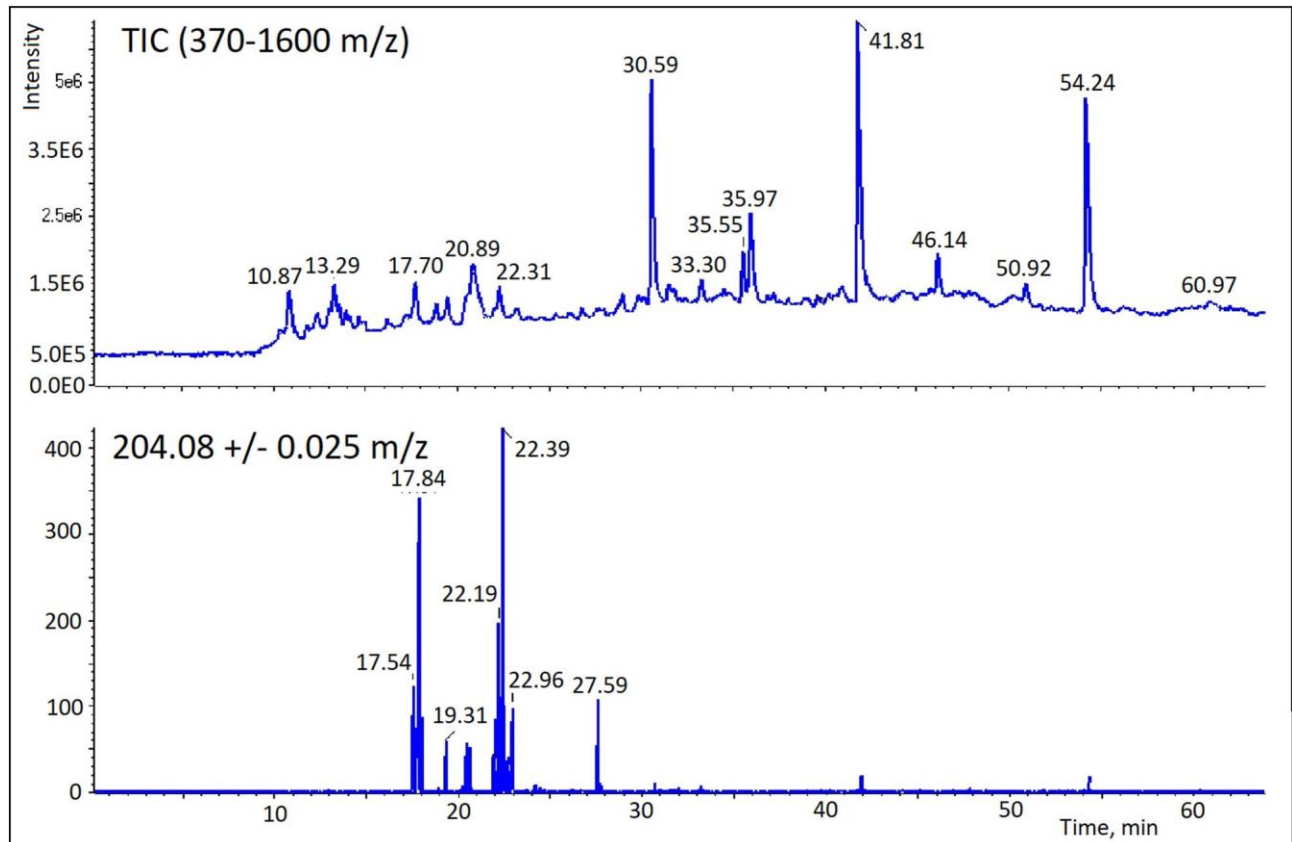


Figure 2.3 Total ion current obtained from the high-performance liquid chromatography/mass spectrometry (HPLC/MS) analysis of the tryptic digest of the CKO pig IgG sample (top): one of the seven selected ion chromatograms of the MS² product ions at m/z 204 for the same sample (bottom).

As specified in the experimental portion, the MS/MS data-dependent mode allowed the acquisition of up to 19 product ion spectra per HPLC peak, but in practice for this particular experiment there were a maximum of 7 precursor ions present at any given time with sufficient abundance to trigger MS/MS acquisition. As a result, for each sample seven m/z 204 chromatograms were inspected manually while noting the precursor ion m/z value, charge (z), and fragmentation patterns to find the $[M + H]^+$ precursor ions of product ions corresponding to the bare peptide backbones. **Figure 2.4** shows examples of two MS/MS spectra obtained for glycopeptides listed in **Table 2.3**. In **Figure 2.4A**, the peptide backbone (Q)FNSTYR is not specific to tryptic digestion as one hydrolysis event occurred at the C-terminus of glutamine. In

Figure 2.4B, the full tryptic peptide is observed with the same glycosylation site, EEQFNSTYR, which is characteristic of all subtypes except IgG6a. Tandem MS fragmentation of glycopeptides has been described in detail,³⁴ and the caption of **Figure 2.4** provides some information on spectral interpretation.

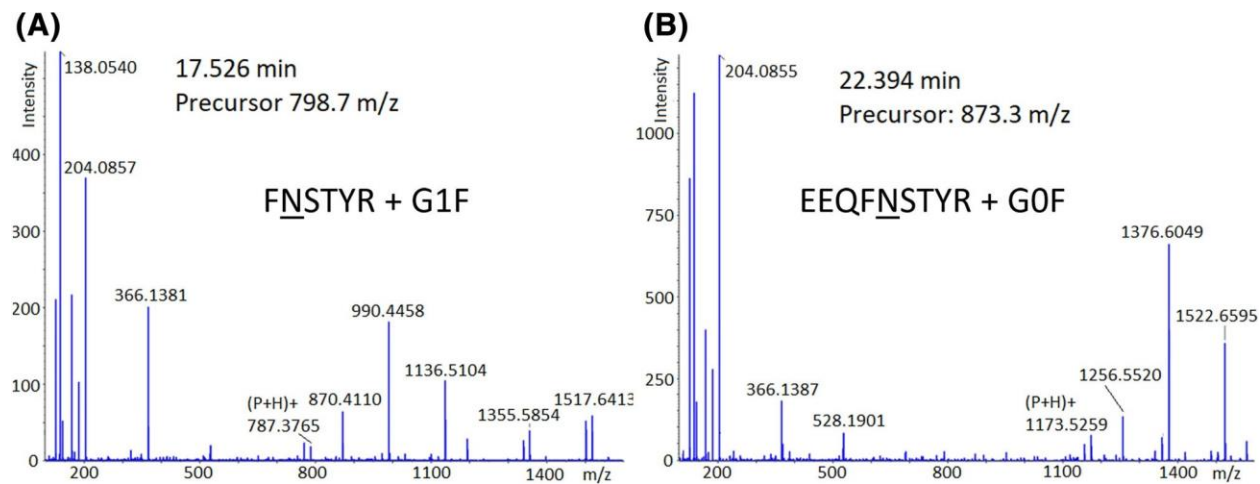


Figure 2.4 Tandem mass spectra for two glycopeptides ionized as $(M + 3H)^{3+}$ ions. These spectra highlight typical fragmentation patterns of glycopeptides. Glycan: m/z 204 (GlcNAc^+), 366 ($204 + \text{Hex}$), 528 ($366 + \text{Hex}$). Peptide: $(P + H)^+$, $(P + H + \text{GlcNAc})^+$, $(P + H + \text{GlcNAcFuc})^+$. The m/z values corresponding to these three types of ions are (A) 787, 990, and 1136 and (B) 1173, 1376, and 1522. For the G0F and G1F glycan structures, see **Figure 2.5**.

Table 2.3 provides another nonspecific peptide, still containing the same glycosylation site, (Q)NFSTY(R). Some of the other peptide backbones were tentatively identified, such as m/z 1212, 1230, 1154, and 1155, resulting from common chemical modifications of EEQFNSTYR. As IAA was used to alkylate cysteines in the protein before digestion, overalkylation of hydroxyl groups can occur, causing a positive mass shift of 57 m/z units.

This is what happened to form m/z 1230 ions, as $1173 + 57 = 1230$, which subsequently lost water. It has been observed that upon alkylation, lactonization by loss of water can take place between the two glutamic acid moieties,³⁵ which would explain m/z 1212 peptide backbones. The m/z 1155 backbones could be explained by dehydration of m/z 1173, and m/z 1154 by overalkylation and dehydration of m/z 1115. The glycopeptide eluting at retention time 40.78 min with the backbone ESLNISWTR (m/z 1105) is from the “uncharacterized” A0A286ZI90 protein in **Table 2.1** (second entry). A peptide search of ESLNISWTR spanning

the whole UniprotKB *Sus scrofa* proteome produced a better-defined pig IgM (accession number K7ZJP7). The calculated alignment level between these two proteins is 100%. The late elution of the IgM glycopeptides can be explained by the presence of NeuAc and hydrophobic tryptophan.²⁶ The last glycopeptide in **Table 2.3** at retention time 49.17 min also contains NeuAc; however, it remains unidentified. These two last glycopeptides were observed in the DKO sample but not in the others.

At this time, it was not possible to find assignments for the backbones with m/z 1455, 1467, and 1515, which were recurrent from sample to sample. These $[M + H]^+$ values were compared against those in **Table 2.3** from proteins other than IgG HC, but no matches were found. Having observed non-trypsin-specific hydrolysis in the HC, it is likely that the same phenomenon occurred for any other proteins present in the original samples. Moreover, all subtypes have noncharacterized variable sections, and unknown glycopeptides may originate from these, making their identification very difficult. Peptides with identified sequences point to IgG6a (EAQFNSTYR) and all other subtypes (EEQFNSTYR) except IgG3, which has the latter sequence but preceded by methionine instead of lysine.

The main glycosylation profile differences that were expected were the presence/absence of α -Gal epitopes and the replacement of Neu5Gc (WT) by Neu5Ac (DKO). The glycans observed and featured in **Table 2.3** (see **Figure 2.5** for reference) are of the *N*-complex biantennary type according to the MS/MS fragmentation patterns. Sialylation was ubiquitous in the WT sample (NeuGc only) and in the DKO sample (NeuAc only), but no sialylated IgG species were detected and identified in the CKO sample. The glycans GF1S, G2FS, and G2FS2 with NeuAc had also been observed by MALDI-MS in a previous study of another DKO pig IgG sample.¹⁸ Tracking down sialylated glycopeptides was possible by examining the SIC of the m/z 290 (NeuGc) and 274 (NeuAc) ions. In general, IgG molecules bearing sialylated *N*-glycans are known to represent only a small portion of bulk IgG samples, yet the importance of Fc sialylation in anti-inflammatory activity has been described.³⁶

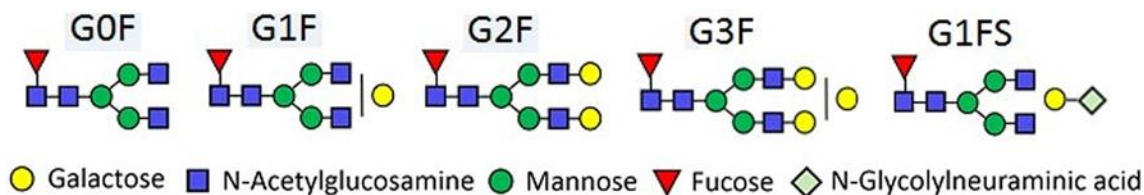


Figure 2.5 Examples of N-glycan structures observed in this study of pig IgG.

Only the WT sample exhibited a species with an α -Gal-bearing glycan, designated by the symbol G3F in **Table 2.3**. This echoed previous results obtained using MALDI-MS.^{17, 18} Moreover, sulfation was observed in a few instances and had been reported in pig proteins other than IgG by von Witzendorff et al for a *Zona pellicula* glycoprotein³⁷ and by de Waard et al for thyroglobulin.³⁸ An exhaustive study by Wang and coworkers pioneered the characterization of sulfated *N*-glycans in human IgG and tracked these acidic glycans as biomarkers for rheumatoid arthritis.³⁹ It has been suggested that sulfation of *N*-glycans can have a profound effect on biological recognition and allow the fast clearance of glycoproteins from the body.^{40, 41}

In many instances the glycan composition was assigned using ExPASy's Glycomod software (<https://web.expasy.org/glycomod/>), and some of these structures contained xylose. These are not listed here, as the glycan compositions did not obey *N*-glycosylation biosynthetic rules.⁴² Xylose in pig IgG *N*-glycans had been observed by Burlak et al,¹⁹ and the xylosylated structures hypothetically found here will be investigated in future research.

Overall, no significant variation was observed in the nature of *N*-glycans between the WT, DKO, and CKO samples. As the present study is not quantitative, it was difficult to estimate the relative abundances of different glycans in each sample. Direct glycosylation profiling was more easily achieved using MALDI-MS in a previous semiquantitative study comparing glycopeptides from the same pig IgG WT and DKO samples studied here.¹⁸ There were noticeable differences in relative Fc glycan abundances between these samples, although the glycan species were the same except for the aforementioned presence of Neu5Ac versus Neu5Gc and of α -Gal. We had also initiated a study on the glycosylation profiles of IgG from WT pigs at different times after immunization (see “Experimental” section), with no significant variations observed (unpublished results).

In this respect, it has been documented that the IgG functionality can be regulated by changes in antibody subtype and by modifications in the *N*-linked glycans in the CH2 Fc domain. Although subclass selection is regulated on infection, the “tuning” of antibody glycosylation in a signal-specific or pathogen-specific manner is still under investigation. For example, Mahan and coworkers have shown human IgG glycosylation to be determined in an antigen- and pathogen-specific manner during HIV infection.⁴³ These authors have also shown that, despite significant variations existing in the glycosylation of bulk IgG among individuals from different parts of the world, immunization can overcome these variations. Thus, in their study the antigen-specific IgG elicited have similar Fc glycosylation profiles. Their results suggest that the immune system drives IgG glycosylation in an antigen-specific manner.⁴³

In the present study, as all possible subtypes except IgG6a and IgG3 have the EEQFNSTFR glycopeptide, it was not possible to distinguish glycosylation profiles for each individual subtype represented. In our previous studies using MALDI-MS^{17, 18}, the glycosylation of IgG6a followed that of other mixed subtypes very closely. The same observation is valid in the present study according to the results shown in **Table 2.3**.

2.4.5 Importance of studying sialylation of pig IgG

It is crucial to report information specific to sialylation in KO pig IgG. By knocking out the genes responsible for the expression of CMAH (and α -galactosyltransferase), polyclonal pig IgG will probably have a much lower immunogenic potential in humans.²⁸ The inactivation of CMAH in humans occurred 2–3 million years ago. However, many animals used for food and biological materials in medicine carry Neu5Gc. Humans regularly ingest trace amounts of Neu5Gc through diet, resulting in the incorporation of Neu5Gc into biological tissues, especially epithelial cells. As a result of medical treatments or transplants of animal-derived products (bioprosthetic heart valves, cells, and tissue replacement), immunological reactions often occur.³¹

The CKO pigs used in this study to obtain IgG samples have been developed and described by Perota and Galli,³¹ based on the main Neu5Gc-null mouse model used before to study this phenotype. For obvious reasons of size compatibility of blood vessels and organs with humans, larger animals (sheep, cattle, and swine) are now better candidates to subject to the CKO mutation. It is also important to note here that clinical xenotransplantation processes are

not limited to using organs or tissues but can also make use of specific molecules, for example, IgG²⁹ in immunotherapy.

DKO pigs with both α -Gal transferase and CMAH knocked out have been studied in more laboratories,^{29, 44, 46} and DKO IgG has been shown to modify the course of Ebola infection in guinea pigs in passive immunotherapy,⁶ with prolonged survival of EBOV-infected guinea pigs and decreased extent of EBOV replication in these animals.

2.4.6 O-Glycosylation in pig IgG

Glycosylated species other than *N*-glycopeptides were present in all three samples and were detected by tracing the m/z 163 (hexose⁺) MS/MS ion chromatograms. It is still questionable whether these peaks correspond to *O*-glycosylated peptides as identification is not as simple as that of *N*-glycosylated peptides. There is no consensus amino acid sequence for *O*-glycosylation to occur, whether it be on serine or threonine. There is also no common core such as the trimannosyl moiety that characterizes all *N*-glycans. *O*-glycans are attached mainly through a glycosidic bond with the OH group of serine or threonine. They have diverse structures and encompass various structural families of relatively heterogeneous structures. Most *O*-glycans are attached to S or T through *N*-acetyl galactosamine (GalNAc) and contain GlcNAc, Gal, sialic acid, and fucose residues. Sialylated *O*-glycans, if present in sufficient amounts, would have been detected in the m/z 274 and 290 traces for NeuAc and NeuGc, respectively, but peaks with reliable signal-to-noise ratios in these traces corresponded to sialylated *N*-glycopeptides. It is likely that *O*-glycosylation, if present, would occur in the variable regions of pig IgG. It was suggested that only up to 10%–15% of these regions can be glycosylated,²⁷ that is, *O*-glycopeptide concentrations would be at least 10 times lower than Fc *N*-glycopeptide concentrations, as the Fc regions are nearly 100% glycosylated. A future study will be dedicated to the characterization of *O*-glycans and variable-region *N*-glycans using an efficient and reliable method developed by Yang et al in 2017.⁴⁷

2.5 Conclusion

The first aim of this study was to assess the ability of RPLC/MS/MS coupled with a powerful proteomic search engine to identify IgG subtypes from their representative peptides. It was possible to do so, and this study highlighted the importance of data treatment/handling of the large amount of information obtained using this instrumental setup. The program X!Tandem allowed for the identification of the HC constant region (gamma) of several subtypes in each sample. This procedure was complex, as the HC GenBank sequences used as references in this work (IgG1a, 1b, 2a, 2b, 3, 4a, 4b, 5a, 5b, 6a, 6b, 6c) are not described as such in UniprotKB, where entries are rather classified as “IgG heavy chain” with no further subtype specification. Aligning and comparing HC sequences of interest were thus necessary. **Figure 2.2** offers only a small sample of all HC alignments generated by our in-house written Excel macro. Only the HC IgG entries found by X!Tandem are featured in this table. Otherwise, the *Sus scrofa* proteome on UniprotKB counts over 50 HC IgG entries, with much redundancy in the pool. If two protein sequences are redundant in the database and designated differently, the search engine will refer to the latest entry. The ability to characterize proteins by this method therefore strongly depends on the accuracy of protein description in the database. These challenges go along with the very scarce information in the literature regarding the study of pig IgG at a protein level.

The second aim was to characterize IgG Fc glycosylation using RPLC/MS/MS and assess the information obtained compared with the MALDI-MS results.^{17, 18} In summary, all glycoform profiles previously found using MALDI-MS were found in this study, although direct visual RPLC/MS/MS profiling was more difficult than using MALDI-MS, as the glycopeptides were separated in time and detected as 2+ and 3+ charged ions. The present study allowed the detection of several forms of the same glycopeptide with modified amino acids and different retention times.

The third aim was to characterize other *N/O* glycosylation sites in pig IgG results, but these attempts did not provide meaningful results at this time. Future work consists of screening all the RPLC/MS/MS data for *O*-glycopeptides and other oligosaccharide-containing compounds and applying more specific methods for *O*-glycan discovery to these samples.

2.6 Acknowledgments

The authors thank Dr Jean Paul Soulillou, Dr Cesare Galli, and their research teams for purifying and providing IgG samples from pig serum. The authors thank Dr Oleg Krokhin for the use of the electrospray ionization-MS/MS triple quadrupole mass spectrometer. Vic Spicer is acknowledged for installing the *Sus scrofa* proteome fasta file in X!Tandem. The authors also thank Emy Komatsu for helping in sample preparation. This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2017-00502; H.P.).

2.7 Supplementary Information

Supplementary Table 2.1 List of all non-glycosylated peptides found in this study of pig IgG

tryptic digestion mixtures. Black: wild-type; red: double knockout; blue: CMAH knockout.

Accession no.: UniprotKB.

m/z m+h	ppm	rt, min	start	sequence	modification(s)	protein	type
767.4774	-0.6	16.25	348	LPAPITR		L8B0U3	1a
767.4774	-2.3	16.25	348	LPAPITR		L8B0U3	1a
767.4774	-5.9	16.25	348	LPAPITR		L8B0U3	1a
770.3978	-2	16.64	151	PLAPCGR	C155+57;	L8B0U3	1a
770.3978	-7.1	16.64	151	PLAPCGR	C155+57;	L8B0U3	1a
826.5185	-4.8	13.88	262	IFPPKPK		L8B0U3	1a
859.3945	-0.5	16.00	289	EHAEVQF		L8B0U3	1a
859.3945	-6.8	16.00	289	EHAEVQF		L8B0U3	1a
905.4761	-2.9	15.47	281	CVVVDVSK	C281+57;	L8B0U3	1a
933.4611	-2	16.80	150	YPLAPCGR	C155+57;	L8B0U3	1a
946.4265	-2.4	16.61	289	EHAEVQFS		L8B0U3	1a
946.4265	-5.5	16.61	289	EHAEVQFS		L8B0U3	1a
962.4982	-0.8	28.97	424	GTFFLYSK		L8B0U3	1a
981.5728	-0.5	23.55	346	VDLPAPITR		L8B0U3	1a
981.5728	-1.7	23.55	346	VDLPAPITR		L8B0U3	1a
981.5728	-4.7	23.55	346	VDLPAPITR		L8B0U3	1a
983.5633	0.7	21.03	29	GLVQPGGSLR		L8B173	1a
983.5633	-1.9	21.03	29	GLVQPGGSLR		L8B173	1a
983.5633	-4.6	21.03	29	GLVQPGGSLR		L8B173	1a
995.5156	0.2	19.79	372	PPPAEELSR		L8B0U3	1a
1000.458	0.3	4.57	415	TTPPQQDVD		L8B0U3	1a
1032.53	-0.8	19.91	149	VYPLAPCGR	C155+57;	L8B0U3	1a
1032.53	-3.6	19.91	149	VYPLAPCGR	C155+57;	L8B0U3	1a
1032.53	-6.2	19.91	149	VYPLAPCGR	C155+57;	L8B0U3	1a
1095.616	1.4	23.53	345	NVDLPAPITR		L8B0U3	1a
1095.616	-0.1	23.53	345	NVDLPAPITR		L8B0U3	1a
1097.606	0.6	21.89	27	GGGLVQPGGSLR		L8B0U3	1a
1097.606	-1.4	21.89	27	GGGLVQPGGSLR		L8B0U3	1a
1097.606	-5.1	21.89	27	GGGLVQPGGSLR		L8B0U3	1a
1105.592	-2.4	19.97	279	VTCVVVDVSK	C281+57;	L8B0U3	1a
1108.6	0.4	17.03	371	LPPPAEELSR		L8B0U3	1a
1108.6	-1.2	17.03	371	LPPPAEELSR		L8B0U3	1a
1108.6	-4.6	17.03	371	LPPPAEELSR		L8B0U3	1a
1119.562	-1.1	21.24	148	SVYPLAPCGR	C155+57;	L8B0U3	1a
1126.538	-1.7	13.20	243	TKPPCPICPG	C247+57; C250+57;	L8B0U3	1a
1126.538	-1.1	13.20	243	TKPPCPICPG	C247+57; C250+57;	L8B0U3	1a

1126.538	-5.2	13.20	243	TKPPCPICPG	C247+57; C250+57;	L8B0U3	1a
1162.495	1	10.90	221	SYTCNVNHPA	C224+57;	L8B0U3	1a
1162.495	-0.7	10.90	221	SYTCNVNHPA	C224+57;	L8B0U3	1a
1162.495	-3.1	10.90	221	SYTCNVNHPA	C224+57;	L8B0U3	1a
1169.631	-1.9	27.57	58	QAPGKGLEWLA		L8B0U3	1a
1175.577	-2.6	24.65	145	TAPSVYPLAPC	C155+57;	L8B0U3	1a
1175.577	-2.3	24.65	145	TAPSVYPLAPC	C155+57;	L8B0U3	1a
1184.638	1.5	22.05	26	SGGGLVQPGGSLR		L8B0U3	1a
1184.638	0	22.05	26	SGGGLVQPGGSLR		L8B0U3	1a
1184.638	-4	22.05	26	SGGGLVQPGGSLR		L8B0U3	1a
1209.647	1.3	20.28	370	TLPPPAEELSR		L8B0U3	1a
1209.647	-0.7	20.28	370	TLPPPAEELSR		L8B0U3	1a
1209.647	-3.6	20.28	370	TLPPPAEELSR		L8B0U3	1a
1216.614	1.4	27.34	147	PSVYPLAPCGR	C155+57;	L8B0U3	1a
1216.614	-0.5	27.34	147	PSVYPLAPCGR	C155+57;	L8B0U3	1a
1216.614	-4.7	27.34	147	PSVYPLAPCGR	C155+57;	L8B0U3	1a
1225.563	-3.8	5.70	224	CNVNHPATTTK	C224+57; C224-17;	L8B0U3	1a
1225.679	2.1	22.17	132	GPGVEVVVSSAPK		L8B0U3	1a
1225.679	-1.3	22.17	132	GPGVEVVVSSAPK		L8B0U3	1a
1232.598	-1.7	24.32	145	TAPSVYPLAPCG	C155+57;	L8B0U3	1a
1247.638	-0.3	15.01	359	AIGQSREPQVY		L8B0U3	1a
1305.596	2.6	17.30	415	TTPPQQDVDGTF		L8B0U3	1a
1305.596	-1.2	17.30	415	TTPPQQDVDGTF		L8B0U3	1a
1305.596	-3.3	17.30	415	TTPPQQDVDGTF		L8B0U3	1a
1308.727	2	26.10	343	VNNVDLPAPITR		L8B0U3	1a
1308.727	1.8	26.10	343	VNNVDLPAPITR		L8B0U3	1a
1308.727	-3.2	26.10	343	VNNVDLPAPITR		L8B0U3	1a
1309.711	2.5	26.10	343	VNNVDLPAPITR	N344+1;	L8B0U3	1a
1309.711	-0.4	26.10	343	VNNVDLPAPITR	N344+1;	L8B0U3	1a
1309.711	-2.7	26.10	343	VNNVDLPAPITR	N344+1;	L8B0U3	1a
1313.681	1.8	21.50	25	ESGGGLVQPGGSLR		L8B0U3	1a
1347.592	1.9	7.10	403	SNGQPEPEGNYR		L8B0U3	1a
1347.592	0.2	7.10	403	SNGQPEPEGNYR		L8B0U3	1a
1347.592	-2	7.10	403	SNGQPEPEGNYR		L8B0U3	1a
1348.576	1.9	7.10	403	SNGQPEPEGNYR	N404+1;	L8B0U3	1a
1348.576	0.7	7.10	403	SNGQPEPEGNYR	Q406+1;	L8B0U3	1a
1348.576	-1.4	7.10	403	SNGQPEPEGNYR	N404+1;	L8B0U3	1a
1353.639	0.6	13.95	87	FTISRDNSQNTA		L8B0U3	1a
1388.699	2.8	24.05	145	TAPSVYPLAPCGR	C155+57;	L8B0U3	1a
1388.699	0.7	24.05	145	TAPSVYPLAPCGR	C155+57;	L8B0U3	1a
1388.699	-2.9	24.05	145	TAPSVYPLAPCGR	C155+57;	L8B0U3	1a
1391.619	2.7	8.17	403	SNGQPEPENTYR		L8B173	1a
1412.749	-0.6	23.22	24	VESGGGLVQPGGSLR		L8B0U3	1a
1414.751	3.5	30.14	145	TAPLVYPLAPCGR	C155+57;	L8B173	1a

1414.751	-2.1	30.14	145	TAPLVYPLAPCGR	C155+57;	L8B173	1a
1415.612	1.6	13.36	243	TKPPCPICPGCE	C247+57; C250+57; C253+57;	L8B0U3	1a
1452.664	-0.2	26.32	415	TTPPQQDVDGTFE		L8B0U3	1a
1471.779	-0.1	27.24	368	VYTLPPPAEELSR		L8B0U3	1a
1471.779	-0.2	27.24	368	VYTLPPPAEELSR		L8B0U3	1a
1511.818	-1.3	26.62	23	VVESGGGLVQPGGSLR		L8B0U8	1a
1511.818	-1.4	26.62	23	VVESGGGLVQPGGSLR		L8B0U8	1a
1525.833	3.5	27.68	23	LVESGGGLVQPGGSLR		L8B0U3	1a
1525.833	0.8	27.68	23	LVESGGGLVQPGGSLR		L8B0U3	1a
1525.833	-1.5	27.68	23	LVESGGGLVQPGGSLR		L8B0U3	1a
1560.794	4.5	23.97	275	QTPEVTCVVVDVSK	C281+57;	L8B0U3	1a
1560.794	-0.2	23.97	275	QTPEVTCVVVDVSK	C281+57;	L8B0U3	1a
1560.794	-1.8	23.97	275	QTPEVTCVVVDVSK	C281+57;	L8B0U3	1a
1565.748	4.8	33.18	415	TTPPQQDVDGTFEFL		L8B0U3	1a
1565.748	0.2	33.18	415	TTPPQQDVDGTFEFL		L8B0U3	1a
1565.748	-0.7	33.18	415	TTPPQQDVDGTFEFL		L8B0U3	1a
1593.733	0.9	11.00	221	SYTCNVNHPATTTK	C224+57;	L8B0U3	1a
1593.733	-1.3	11.00	221	SYTCNVNHPATTTK	C224+57;	L8B0U3	1a
1593.733	-4	11.00	221	SYTCNVNHPATTTK	C224+57;	L8B0U3	1a
1631.791	1217	19.67	294	YVDGVEVHTAETSPK		L8B139	1a
1631.791	1221	19.67	294	YVDGVEVHTAETSPK		L8B139	1a
1639.869	3.7	35.27	129	DLWGPVVEVVSSAPK		L8B0U3	1a
1661.937	2.5	37.67	322	VVSVLPIQHQDWLK		L8B0U3	1a
1661.937	-0.6	37.67	322	VVSVLPIQHQDWLK		L8B0U3	1a
1661.937	-1.9	37.67	322	VVSVLPIQHQDWLK		L8B0U3	1a
1693.927	2.2	37.67	322	VVSVLPIQHQDWLK	W333+32;	L8B0U3	1a
1693.927	-0.3	37.67	322	VVSVLPIQHQDWLK	W333+32;	L8B0U3	1a
1693.927	-2.7	37.67	322	VVSVLPIQHQDWLK	W333+32;	L8B0U3	1a
1718.959	2	36.90	322	VVSVLPIQHQDWLKG		L8B0U3	1a
1718.959	-1.5	36.90	322	VVSVLPIQHQDWLKG		L8B0U3	1a
1728.812	5.2	35.55	415	TTPPQQDVDGTFEFLY		L8B0U3	1a
1728.812	-0.8	35.55	415	TTPPQQDVDGTFEFLY		L8B0U3	1a
1754.813	2.8	25.71	92	DNSQNTAYLQMNSLR		L8B0U3	1a
1770.808	2.5	25.71	92	DNSQNTAYLQMNSLR	M102+16;	L8B0U3	1a
1770.808	-1.3	25.71	92	DNSQNTAYLQMNSLR	M102+16;	L8B0U3	1a
1825.933	6	29.34	365	EPQVYTLPPPAEELSR		L8B0U3	1a
1825.933	0.5	29.34	365	EPQVYTLPPPAEELSR		L8B0U3	1a
1825.933	-1.2	29.34	365	EPQVYTLPPPAEELSR		L8B0U3	1a
1826.823	2	21.36	243	TKPPCPICPGCEVAGPS	C247+57; C250+57; C253+57;	L8B0U3	1a
1826.823	-1.2	21.36	243	TKPPCPICPGCEVAGPS	C247+57; C250+57; C253+57;	L8B0U3	1a

1826.823	-2.5	21.36	243	TKPPCPICPGCEVAGPS	C247+57; C250+57; C253+57;	L8B0U3	1a
1943.939	4.4	33.54	415	TTPPQQDVDGTFFLYSK		L8B0U3	1a
1943.939	-0.2	33.54	415	TTPPQQDVDGTFFLYSK		L8B0U3	1a
1943.939	-1.4	33.54	415	TTPPQQDVDGTFFLYSK		L8B0U3	1a
1969.035	2	31.67	19	GEEKLVESGGGLVQPGGSLR		L8B0U3	1a
1969.035	500.2	31.67	19	GEEKLVESGGGLVQPGGSLR		L8B0U3	1a
2072.96	3.2	31.94	243	TKPPCPICPGCEVAGPSVF	C247+57; C250+57; C253+57;	L8B0U3	1a
2072.96	-0.3	31.94	243	TKPPCPICPGCEVAGPSVF	C247+57; C250+57; C253+57;	L8B0U3	1a
2072.96	-0.9	31.94	243	TKPPCPICPGCEVAGPSVF	C247+57; C250+57; C253+57;	L8B0U3	1a
2470.25	3.9	39.62	415	TTPPQQDVDGTFFLYSKLAVDK		L8B0U3	1a
2589.2	5	38.58	289	EHAEVQFSWYVDGVEVHTAETR		L8B0U3	1a
2697.388	4.4	37.77	415	TTPPQQDVDGTFFLYSKLAVDKAR		L8B0U3	1a
2697.388	-0.9	37.77	415	TTPPQQDVDGTFFLYSKLAVDKAR		L8B0U3	1a
2814.348	5	37.42	289	EHAEVQFSWYVDGVEVHTAETRPK		L8B0U3	1a
2814.348	0.8	37.42	289	EHAEVQFSWYVDGVEVHTAETRPK		L8B0U3	1a
2814.348	-0.6	37.42	289	EHAEVQFSWYVDGVEVHTAETRPK		L8B0U3	1a
2846.338	0.3	37.42	289	EHAEVQFSWYVDGVEVHTAETRPK	W297+32;	L8B0U3	1a
2880.461	0.3	39.22	243	TKPPCPICPGCEVAGPSVFIFPPKPK	C247+57; C250+57; C253+57;	L8B0U3	1a
2880.461	-1.4	39.22	243	TKPPCPICPGCEVAGPSVFIFPPKPK	C247+57; C250+57; C253+57;	L8B0U3	1a
2880.461	4.9	39.22	243	TKPPCPICPGCEVAGPSVFIFPPKPK	C247+57; C250+57; C253+57;	L8B0U3	1a
778.4206	-1.7	9.46	232	VGIHQPPQ		L8B165	1b
778.4206	-7.2	9.46	232	VGIHQPPQ		L8B165	1b
1391.619	-1.4	8.17	396	SNGQPEPENTYR		L8B165	1b
1392.603	-2.1	8.17	396	SNGQPEPENTYR	N397+1;	L8B165	1b
1563.741	-4.3	21.98	232	VGIHQPQTCPICPG	C240+57; C243+57;	L8B165	1b
2264.026	-1.6	27.56	232	VGIHQPQTCPICPGCEVAGPS	C240+57; C243+57; C246+57;	L8B165	1b
2510.162	-0.7	35.61	232	VGIHQPQTCPICPGCEVAGPSVF	C240+57; C243+57; C246+57;	L8B165	1b
1418.71	2.1	23.66	146	TAPSVYPLAPCSR	C156+57;	L8B0W5	2a
1418.71	-0.7	23.66	146	TAPSVYPLAPCSR	C156+57;	L8B0W5	2a
1418.71	-3	23.66	146	TAPSVYPLAPCSR	C156+57;	L8B0W5	2a
1140.554	0.6	14.90	249	TKPPCPICPA	C253+57; C256+57;	L8B0X5	2b
1497.827	-1	27.25	23	LVESGGGLVQPGGSLK		L8AXL9	2b
3164.743	-8.5	49.02	8	VVLFALLQGVQGEERLVESGGGLVQPGGSLR		L8B0Y0	2b
3164.743	-12	49.02	8	VVLFALLQGVQGEERLVESGGGLVQPGGSLR		L8B0Y0	2b
3164.743	-14.9	49.02	8	VVLFALLQGVQGEERLVESGGGLVQPGGSLR		L8B0Y0	2b

873.4312	-1.2	15.79	162	DTSGPNVAL		L8B130	3
873.4312	-1.2	15.79	162	DTSGPNVAL		L8B130	3
3622.632	13	33.87	92	DNSQNTAYLQMNSLRTEDTAQCYCTRGGIAIA	C113+57; C115+57;	L8B130	3
863.4621	-2.5	17.82	463	SIFKTPGN		L8B0W9	4a
882.5043	-1.9	19.79	348	DLPAPITR		L8B0W9	4a
882.5043	-0.6	19.79	348	DLPAPITR		L8B0W9	4a
882.5043	-5.3	19.79	348	DLPAPITR		L8B0W9	4a
999.5469	-6	17.86	58	QAPGKGLESL		L8B0W9	4a
1077.532	-0.3	9.89	362	GQTREPQVY		L8B0W9	4a
1077.532	-0.9	9.89	362	GQTREPQVY		L8B0W9	4a
1077.532	-4.1	9.89	362	GQTREPQVY		L8B0W9	4a
1126.55	0.2	17.30	244	VGRPCPCPA	C248+57; C251+57;	L8B0W9	4a
1126.55	-0.9	17.30	244	VGRPCPCPA	C248+57; C251+57;	L8B0W9	4a
1126.55	-4.7	17.30	244	VGRPCPCPA	C248+57; C251+57;	L8B0W9	4a
1138.61	-2.5	16.19	372	LPPPTEELSR		L8B0W9	4a
1138.61	-4.5	16.19	372	LPPPTEELSR		L8B0W9	4a
1140.554	-6.4	14.90	234	TKPPPCPCPA	C238+57; C241+57;	L8B0W0	4a
1239.658	1.7	19.62	371	TLPPPTEELSR		L8B0W9	4a
1239.658	-0.1	19.62	371	TLPPPTEELSR		L8B0W9	4a
1239.658	-2.8	19.62	371	TLPPPTEELSR		L8B0W9	4a
1260.56	1.4	7.16	405	NGQPEPEGNYR		L8B0W9	4a
1260.56	0.1	7.16	405	NGQPEPEGNYR		L8B0W9	4a
1260.56	-3.5	7.16	405	NGQPEPEGNYR		L8B0W9	4a
1261.544	0.7	7.16	405	NGQPEPEGNYR	N405+1;	L8B0W9	4a
1261.544	-0.2	7.16	405	NGQPEPEGNYR	N405+1;	L8B0W9	4a
1261.544	-3.4	7.16	405	NGQPEPEGNYR	N405+1;	L8B0W9	4a
1321.591	-2.6	10.29	416	TTPPQQDVDGTY		L8B0W9	4a
1337.754	1.1	21.12	344	VNNKDLPAPITR		L8B0W9	4a
1337.754	-0.3	21.12	344	VNNKDLPAPITR		L8B0W9	4a
1337.754	-5.8	21.12	344	VNNKDLPAPITR		L8B0W9	4a
1468.659	-0.6	21.57	416	TTPPQQDVDGTYF		L8B0W9	4a
1501.79	0.1	26.58	369	VYTLPPPTEELSR		L8B0W9	4a
1501.79	-2.1	26.58	369	VYTLPPPTEELSR		L8B0W9	4a
1547.833	-0.5	30.82	326	VLPIQHQDWLNGK		L8B0W9	4a
1581.743	1.1	28.55	416	TTPPQQDVDGTYFL		L8B0W9	4a
1581.743	-1.1	28.55	416	TTPPQQDVDGTYFL		L8B0W9	4a
1744.806	1.6	32.16	416	TTPPQQDVDGTYFLY		L8B0W9	4a
1744.806	-1.7	32.16	416	TTPPQQDVDGTYFLY		L8B0W9	4a
1833.002	3.3	36.57	323	VVSVLPIQHQDWLNGK		L8B0W9	4a
1833.002	0.9	36.57	323	VVSVLPIQHQDWLNGK		L8B0W9	4a
1833.002	-2.1	36.57	323	VVSVLPIQHQDWLNGK		L8B0W9	4a
1855.944	0.3	28.74	366	EPQVYTLPPPTEELSR		L8B0W9	4a
1855.944	-0.4	28.74	366	EPQVYTLPPPTEELSR		L8B0W9	4a
1855.944	5.9	28.74	366	EPQVYTLPPPTEELSR		L8B0W9	4a

1864.992	-1.4	36.57	323	VVSVLPIQHQDWLNGK	W334+32;	L8B0W9	4a
1959.933	3.7	30.13	416	TTPPQQDVDGTYFLYSK		L8B0W9	4a
1959.933	-0.8	30.13	416	TTPPQQDVDGTYFLYSK		L8B0W9	4a
1959.933	-1.6	30.13	416	TTPPQQDVDGTYFLYSK		L8B0W9	4a
2028.909	0	28.79	244	VGRPCPICPACEGPGPSAF	C248+57; C251+57; C254+57;	L8B0W9	4a
2298.172	1.1	28.93	362	GQTREPQVYTLPPPTEELSR		L8B0W9	4a
2298.172	0.2	28.93	362	GQTREPQVYTLPPPTEELSR		L8B0W9	4a
2836.41	0	37.71	244	VGRPCPICPACEGPGPSAFIFPPKPK	C248+57; C251+57; C254+57;	L8B0W9	4a
1324.686	2.5	21.59	344	VNNEDLPGPITR		L8B0S2	5a
1324.686	-1	21.59	344	VNNEDLPGPITR		L8B0S2	5a
1324.686	-4.5	21.59	344	VNNEDLPGPITR		L8B0S2	5a
1756.9	5.4	30.24	366	SPEVYTLPPPAEELSK		L8B0S2	5a
2277.183	423.8	41.03	270	DILMISRTPEVTCVVVDVSK	C282+57; M273+16;	L8B0S2	5a
1088.558	0.6	17.56	357	LSPSAEELSR		L8B0R9	6a
1189.606	-0.4	20.11	356	TLSPSAEELSR		L8B0R9	6a
1613.76	0.3	28.12	225	VDLCVGRPCPICPA	C228+57; C233+57; C236+57;	L8B0R9	6a
1650.736	12.1	15.60	211	CFTCNVNHPATTTK	C211+57; C214+57;	L8B0R9	6a
1650.736	9.7	15.60	211	CFTCNVNHPATTTK	C211+57; C214+57;	L8B0R9	6a
1662.921	2.8	38.86	308	VVSVLPIQHEDWLK		L8B0R9	6a
1662.921	-1.3	38.86	308	VVSVLPIQHEDWLK		L8B0R9	6a
1805.892	-0.2	29.20	351	EPQVYTLSPSAEELSR		L8B0R9	6a
2402.231	-0.7	27.66	345	AKGPSREPQVYTLSPSAEELSR		L8B0R9	6a
2516.119	0.2	36.35	225	VDLCVGRPCPICPACEGPGPSAF	C228+57; C233+57; C236+57; C239+57;	L8B0R9	6a
2516.119	5.6	36.35	225	VDLCVGRPCPICPACEGPGPSAF	C228+57; C233+57; C236+57; C239+57;	L8B0R9	6a
3323.62	5.4	40.96	225	VDLCVGRPCPICPACEGPGPSAFIFPPKPK	C228+57; C233+57; C236+57; C239+57;	L8B0R9	6a
1219.498	12.3	15.50	220	CFTCNVNHPA	C220+57; C223+57;	L8B180	6b
1650.736	7.3	15.60	220	CFTCNVNHPATTTK	C220+57; C223+57;	L8B180	6b

57: carbamidomethyl from iodoacetamide reaction

1: deamidation

16: oxidation

32: dioxidation

17: ammonia loss

2.8 References

1. Yoon CH, Choi SH, Lee HJ, Kang HJ, Kim MK. Predictive biomarkers for graft rejection in pig-to-non-human primate corneal xenotransplantation. *Xenotransplantation*. 2019;26(4):e12515.
2. Li Q, Shaikh S, Iwase H, et al. Carbohydrate antigen expression and anti-pig antibodies in New World capuchin monkeys: Relevance to studies of xenotransplantation. *Xenotransplantation*. 2019;26(3): e12498.
3. Ramis G, Martinez-Alarcon L, Medina-Moreno E, et al. Presence of pig IgG and IgM in sera samples from baboons after an Orthotopic liver xenotransplantation. *Transplant Proc*. 2018;50(9):2842-2846.
4. Salama A, Conchon S, Perota A, et al. Abstracts of the IPITA-IXA-CTS 2015 joint congress November 15-19, 2015, Melbourne, Australia. *Xenotransplantation*. 2015;22(S1):S2-S47.
5. Zhao C, Cooper DKC, Dai Y, Hara H, Cai Z, Mou L. The Sda and cad glycan antigens and their glycosyltransferase, beta1,4GalNAcT-II, in xenotransplantation. *Xenotransplantation*. 2018;25(2):e12386.
6. Reynard O, Jacquot F, Evanno G, et al. Anti-EBOV GP IgGs lacking alpha1-3-galactose and Neu5Gc prolong survival and decrease blood viral load in EBOV-infected Guinea pigs. *PLoS One*. 2016;11(6): e0156775.
7. Dong H, Su A, Lv D, et al. Development of whole-porcine monoclonal antibodies with potent neutralization activity against classical swine fever virus from single B cells. *ACS Synth Biol*. 2019; 8(5):989-1000.
8. Zhao Q, Wu Q, Ma P, et al. Selective and sensitive fluorescence detection method for pig IgG based on competitive immunosensing strategy and magnetic bioseparation. *Talanta*. 2019;195:103-108.
9. Mota I. IgG subclasses in animal species. *Monogr Allergy*. 1986;19: 302-312.

10. Bokhout BA, van Asten-Noordijk JJ, Stok W. Porcine IgG. Isolation of two IgG-subclasses and anti-IgG class- and subclass-specific antibodies. *Mol Immunol*. 1986;23(6):675-683.
11. Franek F. The subclass heterogeneity of pig IgG. *Vet Immunol Immunopathol*. 1987;17(1-4):79-89.
12. Huang SC, Hu ZL, Hasler-Rapacz J, Rapacz J. Preferential mammary storage and secretion of immunoglobulin gamma (IgG) subclasses in swine. *J Reprod Immunol*. 1992;21(1):15-28.
13. Crawley A, Wilkie BN. Porcine Ig isotypes: Function and molecular characteristics. *Vaccine*. 2003;21(21-22):2911-2922.
14. Butler JE, Wertz N, Deschacht N, Kacs Kovics I. Porcine IgG: Structure, genetics, and evolution. *Immunogenetics*. 2009;61(3): 209-230.
15. Eguchi-Ogawa T, Toki D, Wertz N, Butler JE, Uenishi H. Structure of the genomic sequence comprising the immunoglobulin heavy constant (IGHC) genes from *Sus scrofa*. *Mol Immunol*. 2012;52(3-4): 97-107.
16. Sinkorova J, Stepanova K, Butler JE, Sinkora M. T cells in swine completely rearrange immunoglobulin heavy chain genes. *Dev Comp Immunol*. 2019;99:103396.
17. Lopez PG, Girard L, Buist M, et al. Characterization of N-glycosylation and amino acid sequence features of immunoglobulins from swine. *Glycoconj J*. 2016;33(1):79-91.
18. Buist M, Komatsu E, Lopez PG, et al. Features of N-glycosylation of immunoglobulins from knockout pig models. *J Anal Bioanal Tech*. 2016;7(5):2.
19. Burlak C, Bern M, Brito AE, et al. N-linked glycan profiling of GGTA1/CMAH knockout pigs identifies new potential carbohydrate xenoantigens. *Xenotransplantation*. 2013;20(5):277-291.
20. Nelson C, Bacala R, Gigolyk B, et al. Characterization of Whole and Fragmented Wild-Type Porcine IgG *Intech Open*. In: Ince M, Ince OK, eds. *Recent Advances in Analytical Chemistry*; 2019:65-93.

21. Sun JF, Shi ZX, Guo HC, Li S, Tu CC. Proteomic analysis of swine serum following highly virulent classical swine fever virus infection. *Virology*. 2011;8(1):107.
22. Marco-Ramell A, Bassols A. Enrichment of low-abundance proteins from bovine and porcine serum samples for proteomic studies. *Res Vet Sci*. 2010;89(3):340-343.
23. Marco-Ramell A, Miller I, Nobauer K, et al. Proteomics on porcine haptoglobin and IgG/IgA show protein species distribution and glycosylation pattern to remain similar in PCV2-SD infection. *J Proteomics*. 2014;101:205-216.
24. Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem*. 1988;60(20): 2299-2301.
25. Komatsu E, Buist M, Roy R, et al. Characterization of immunoglobulins through analysis of N-glycopeptides by MALDI-TOF MS. *Methods*. 2016;104:170-181.
26. Ang E, Neustaeter H, Spicer V, Perreault H, Krokhin O. Retention time prediction for Glycopeptides in reversed-phase chromatography for Glycoproteomic applications. *Anal Chem*. 2019;91(21): 13360-13366.
27. Vidarsson G, Dekkers G, Rispen T. IgG subclasses and allotypes: From structure to effector functions. *Front Immunol*. 2014;5:520.
28. Sachs DH, Galli C. Genetic manipulation in pigs. *Curr Opin Organ Transplant*. 2009;14(2):148-153.
29. Salama A, Mosser M, L  v  que X, et al. Neu5Gc and α 1-3 GAL Xenoantigen knockout does not affect Glycemia homeostasis and insulin secretion in pigs. *Diabetes*. 2017;66(4):987-993.
30. Naso F, Stefanelli U, Buratto E, et al. Alpha-gal inactivated heart valve bioprostheses exhibit an anti-calcification propensity similar to knockout tissues. *Tissue Eng Part a*. 2017;23(19-20): 1181-1195.
31. Perota A, Galli C. N-Glycolylneuraminic acid (Neu5Gc) null large animals by targeting the CMP-Neu5Gc hydroxylase (CMAH). *Front Immunol*. 2019;10:2396.

32. Krokhin OV, Spicer V. Peptide retention standards and hydrophobicity indexes in reversed-phase high-performance liquid chromatography of peptides. *Anal Chem.* 2009;81(22): 9522-9530.
33. Roy R, Ang E, Komatsu E, et al. Absolute quantitation of Glycoforms of two human IgG subclasses using synthetic fc peptides and Glycopeptides. *J Am Soc Mass Spectrom.* 2018;29(6):1086-1098.
34. Krokhin O, Ens W, Standing KG, Wilkins J, Perreault H. Site-specific N-glycosylation analysis: Matrix-assisted laser desorption/ionization quadrupole-quadrupole time-of-flight tandem mass spectral signatures for recognition and identification of glycopeptides. *Rapid Commun Mass Spectrom.* 2004;18(18):2020-2030.
35. Gomes de Oliveira AG, Roy R, Raymond C, et al. A systematic study of glycopeptide esterification for the semi-quantitative determination of sialylation in antibodies. *Rapid Commun Mass Spectrom.* 2015;29 (19):1817-1826.
36. Kaneko Y, Nimmerjahn F, Ravetch JV. Anti-inflammatory activity of immunoglobulin G resulting from fc sialylation. *Science.* 2006;313 (5787):670-673.
37. von Witzendorff D, Ekhlesi-Hundrieser M, Dostalova Z, et al. Analysis of N-linked glycans of porcine zona pellucida glycoprotein ZPA by MALDI-TOF MS: A contribution to understanding zona pellucida structure. *Glycobiology.* 2005;15(5):475-488.
38. de Waard P, Koorevaar A, Kamerling JP, Vliegthart JF. Structure determination by ¹H NMR spectroscopy of (sulfated) sialylated N- linked carbohydrate chains released from porcine thyroglobulin by peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase-F. *J Biol Chem.* 1991;266(7):4237-4243.
39. Wang JR, Gao WN, Grimm R, et al. A method to identify trace sulfated IgG N-glycans as biomarkers for rheumatoid arthritis. *Nat Commun.* 2017;8(1):631.
40. Kawashima H, Fukuda M. Sulfated glycans control lymphocyte homing. *Ann N Y Acad Sci.* 2012;1253(1):112-121.
41. Taguchi T, Iwasaki M, Muto Y, et al. Occurrence and structural analysis of highly sulfated multiantennary N-linked glycan chains derived from a fertilization-associated

- carbohydrate-rich glycoprotein in unfertilized eggs of *Tribolodon hakonensis*. *Eur J Biochem.* 1996; 238(2):357-367.
42. Bieberich E. Synthesis, processing, and function of N-glycans in N-glycoproteins. *Adv Neurobiol.* 2014;9:47-70.
43. Mahan AE, Jennewein MF, Suscovich T, et al. Antigen-specific antibody glycosylation is regulated via vaccination. *PLoS Pathog.* 2016;12(3). e1005456-e1005456
44. Lutz AJ, Li P, Estrada JL, et al. Double knockout pigs deficient in N-glycolylneuraminic acid and galactose alpha-1,3-galactose reduce the humoral barrier to xenotransplantation. *Xenotransplantation.* 2013;20(1):27-35.
45. Kwon D-N, Lee K, Kang M-J, et al. Production of biallelic CMP-Neu5Ac hydroxylase knock-out pigs. *Sci Rep.* 2013;3(1): 1981-1981.
46. Li P, Estrada JL, Burlak C, et al. Efficient generation of genetically distinct pigs in a single pregnancy using multiplexed single-guide RNA and carbohydrate selection. *Xenotransplantation.* 2015;22(1):20-31.
47. Yang S, Hu Y, Sokoll L, Zhang H. Simultaneous quantification of N and O-glycans using a solid-phase method. *Nat Protoc.* 2017;12(6): 1229-1244.

3 Chapter 3: Influence of ion pairing agent in LC-MS proteomic experiments

Formic or acetic acid for peptide LC-MS: is the debate still going?

Taylor Battellino¹, Kosuke Ogata², Victor Spicer³, Yasushi Ishihama², Oleg Krokhin^{*3,4}

¹Department of Chemistry, University of Manitoba, Winnipeg, MB, Canada

²Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto 606-8501, Japan

³Manitoba Centre for Proteomics and Systems Biology, Department of Internal Medicine, University of Manitoba, Winnipeg, MB, Canada.

⁴Department of Internal Medicine, University of Manitoba, Winnipeg, MB, Canada

To be submitted to the Journal of Proteome Research, January 2022.

3.0 Contributions of Authors

YI and OK were responsible for creating and supervising the project. TB, KO, and OK prepared samples and performed subsequent analysis. TB, KO, OK, and VS worked on data treatment. TB and KO prepared figures. TB, OK, KO, and YI contributed to the writing and editing of the manuscript.

3.1 Abstract

Despite the general acceptance of formic acid as the additive of choice for peptide reversed-phase LC-MS/MS applications, some still argue that the selection of acetic acid represents a better option. To settle this debate, we investigated both the difference in MS sensitivity and chromatographic behaviour of peptides between these two systems. This interlaboratory study was performed using different MS setups and C18 separation media employing both 0.1% formic and 0.5% acetic acid as ion pairing modifier. Relative to formic acid, we find an overall ~2.2-2.5x increase in MS signal and slight decrease in RP LC retention (-0.7% acetonitrile on average) for acetic acid conditions. While these two features have opposing effects on peptide detectability, we find that acetic acid produces up to 60% higher peptide ID output depending on the type of sample. The drop in RPLC retention increases with peptide net charge at acidic pH. MS signal is dependent on the difference between charge of the precursor ion and charge of the peptide in solution, favoring species with a low pI. Lower peptide retention under acetic acid conditions demonstrates its higher hydrophilicity and as expected, leads to composition and sequence-dependent character of the observed retention shift. A custom version of the SSRCalc retention prediction model has been developed to accommodate these charges, with formic and acetic acid models showing identical results independent of the origin of data collection.

3.2 Introduction

The accelerated development of bottom-up proteomics in the last three decades has led to the establishment of specific sets of methods/protocols commonly applied in research laboratories around the world¹. They can be roughly divided into sample preparation, peptide separation, mass spectrometry (MS), and identification subcategories and usually aim to utilize the best practices available. Selection of MS equipment and, to a large extent, identification/quantitation software practices are mostly dependent on MS vendor offerings, except for the cases where significant investments in customized equipment or new computer applications^{2,3} can be afforded by advanced research labs. The field of sample preparation for proteomics is driven by multiple independent labs targeting new procedures for protein extraction and digestion, as well as peptide labeling, and remains the most dynamic proteomics subdivision in terms of development⁴. Peptide separation in its standard 1D format or in the second dimension of 2D techniques, on the other hand, has eluded ground-breaking developments for a long time. This is due to the long history of peptide reversed-phase liquid chromatography (RPLC) prior to the proteomic era, which introduced only relatively minor changes to this technique.

Proteomic RPLC-MS applications favor carboxylic acids (formic (FA), acetic (AA)) over perfluorinated carboxylic acids (trifluoroacetic (TFA), heptafluorobutyric (HFBA)), which suppress ionization efficiency and are typical for analytical and preparative RPLC with UV detectors^{5,6}. Significant efforts had been directed towards the optimization of eluent additives in the late 90s and early 2000s to perfect both separation efficiency and sensitivity (ESI efficiency)⁶⁻⁹. Formic and acetic acids prevail in the latter, while TFA and HFBA provide better retention and peak shape. The application of mixed modifiers¹⁰ and post-column addition of sensitivity-boosting agents¹¹ have been suggested as possible solutions. In this regard we can very clearly observe a constant search of optimal RPLC conditions, which can be found in J. Yates, III and co-workers' publications¹²⁻¹⁵. Their MudPit technology transitioned from using 0.5% AA¹² in 1999 - to 0.02 HFBA¹³ - to 0.5% AA / 0.012 HFBA¹⁴ - to 0.1% FA¹⁵ in 2006. Further improvements to the quality of RPLC columns have made separation efficiency less of an issue. Consequently, early reports on the superior sensitivity of FA relative to AA⁷⁻⁹ as well as the presence of Fe-acetic acid complexes¹⁶ potentially interfering with peptide detection tilted the preferences towards formic acid.

We surveyed Volume 19 (2020) of the *Journal of Proteome Research* and established that most publications with LC-MS peptide data employ various fully porous (~100-150Å) C18 phases with standard trimethylsilyl or hydrophilic end-capping and volatile carboxylic acids as ion pairing modifiers. 95% of the 187 reports using peptide RPLC-MS utilized formic acid; mostly 0.1% with a few variations up to 0.2%. Acetic acid was listed in only 3% of cases despite its early popularity in some of the most advanced laboratories^{12,17}. Thus, M. Mann's lab had typically employed 0.5% AA based eluents before switching to 0.1% FA in 2013-2014¹⁷⁻¹⁹. At the same time, there are researchers who insist on the superior performance of acetic acid based eluents^{20,21}.

The motivation for this work was to establish differences in peptide detection sensitivity and separation selectivity between two standard eluent systems (0.1% FA and 0.5% AA) employed in two collaborating laboratories. The lab in Kyoto University (KU, Y. Ishihama) favors 0.5% acetic acid – a choice made in early 2000 due to the better stability of ESI on a QStar Pulsar I instrument. The University of Manitoba lab (UM, O. Krokhin) traditionally uses 0.1% FA following the recommendation of their LC-MS vendor, Sciex/LC-Packings. A detailed evaluation of peptide RPLC separation selectivity in terms of the comparison between these two systems has not been reported before. This information would help in expanding the application of retention prediction models and tabulated peptide retention data acquired using alternative LC settings.

3.3 Experimental

3.3.1 Materials and digest preparation

All chemicals in the UM lab were of analytical chemistry grade and purchased from Sigma-Aldrich (St-Louis, MO) or Thermo Fisher (Waltham, MA), unless mentioned otherwise. LC-MS grade deionized water, acetonitrile, formic acid (Optima LC-MS), and acetic acid (Optima LC-MS, both from Thermo Fisher) were used for eluent preparation. All chemicals in the KU lab were purchased from Fujifilm Wako (Osaka, Japan) unless otherwise specified. A comprehensive description of sample preparation procedures is provided in the Supplementary information. The UM lab has performed a standard in-solution tryptic digestion of a protein extract from human cell-line K562 (V6941, Promega, Madison, WI) using sequencing grade modified trypsin (Promega). The KU lab used whole HeLa cell digests prepared using a phase transfer surfactant aided trypsin digestion protocol^{22,23} with sequencing grade modified trypsin (Promega) or lysargiNase (Millipore). Phosphopeptide and N-terminal acetylated peptide isolation was performed as described previously^{20,23}. Half of the phosphopeptides' isolate was dephosphorylated using alkaline phosphatase (Fujifilm Wako). In-house designed peptide retention standards P1-P6²⁴ and B1-B5 were synthesised by BioSynthesis (Lewisville, TX).

3.3.2 Micro-flow LC separations

An Agilent 1100 series with UV detection at 214 nm and manual 100 μ L loop injection was used for the separation of standard peptide mixtures and first dimension peptide fractionation in 2D LC-MS/MS. The latter was done using standard high pH RP conditions on a 1x100 mm 5 μ m Xterra column (Waters) with pairwise concatenation into 20 fractions²⁵. Standard peptides were separated using a 1x50 mm 3 μ m Luna C18(2) column with a 1% per minute water/ acetonitrile gradient (0-30%) with either 0.1% FA or 0.5% AA as the eluent additive.

3.3.3 Nano-flow LC-MS settings

Whole cell digests were analysed by 1D (both labs) and 2D LC-MS/MS (UM). Analysis of phosphorylated/dephosphorylated and acetylated peptide fractions was done by 1D LC-MS/MS in KU. Both laboratories employed both 0.1% FA and 0.5% AA eluent additives with

linear water/acetonitrile gradients with identical injection amounts and otherwise identical LC-MS conditions for each lab.

The University of Manitoba used an Ultra 2D-nano (Eksigent, Dublin, CA) with a Triple TOF 5600 (Sciex, Concord, ON) hyphenated through an IonSpray III source. Nano-LC configuration included a 0.3x5 mm trap Pepmap C18 column (Thermo Fisher) and an in-house packed 100 μm x 20 cm 3 μm Luna C18(2) (Phenomenex, Torrance, CA), terminated by the standard stainless-steel fittings and 2 μm stainless steel screens from Valco (Houston, TX). 10 μL of the samples containing ~ 1 μg of peptides (both whole cell digests and high pH fractions) were spiked with ~ 400 fmole of P1-P6 standard peptides, injected on the trap columns (5 min at 15 $\mu\text{L}/\text{min}$), and eluted using a linear water/acetonitrile 0.5-35% ACN gradient at 500 nL/min flow rate.

Kyoto University configuration included an Ultimate 3000 RSLCnano (Thermo Fisher) nano LC system and 100 μm x 15 cm in-house packed Reprosil-Pur C18 AQ 1.9 μm (Dr. Maisch, Germany) column maintained at 30 $^{\circ}\text{C}$ and an Orbitrap Fusion Lumos (HCD-iontrap) mass spectrometer. A 5 μL loop was used for direct injection (250 ng of peptides, phosphopeptides enriched from 50 μg peptides, or N-terminal peptides enriched from 10 μg peptides: spiked with ~ 200 fmol of each standard peptide). Peptides were separated using linear 1% to 32% ACN gradients over 62 min (0.5% per min) at a 500 nL/min flow rate. Details of MS settings for the Triple TOF 5600 and Orbitrap Fusion Lumos mass spectrometers are outlined in the Supplementary Information.

3.3.4 Protein/peptide identifications and evaluation of MS detection sensitivity

UM employed the X!Tandem search engine for peptide identification with a 20 and 50 ppm mass tolerance for precursor and product ions, respectively. The search parameters included fixed modification of Cys (carbamidomethylation), tryptic digestion settings with up to 2 missed cleavages allowed, and a common set of post-translational modifications (Met oxidation, deamidation, N-terminal cyclization (Gln, Cys) and acetylation). All identifications (peptides with $\log(e) \leq -1$) were additionally filtered using the SSRCalc retention time prediction to ensure high quality data for retention modeling. KU used the MaxQuant26 (ver.1.6.7.0) for peptide identification. Carbamidomethylation (C) was set as the fixed modification. Oxidation (M) and acetylation (Protein N-term) were set as variable modifications. For phosphoproteome

analysis, phosphorylation (STY) was additionally allowed as a variable modification. Strict Trypsin/P (C-terminal to K/R) or lysargiNase (N-terminal to K/R) specificity with up to 2 missed cleavages was allowed.

Retention times from RPLC runs were converted into hydrophobicity index (HI, % acetonitrile) units using linear regression against tabulated retention values of standard peptides spiked into each sample/fraction (**Supplementary Table 3.1**). Detection sensitivity was assessed based on MS2 signals for the constant time 100 ms MS/MS acquisition and MS1 peak areas of respective peptide pairs for the Triple TOF 5600 (UM) and Orbitrap Fusion Lumos (KU), respectively.

3.4 Results

Preliminary assessment of FA-AA RPLC separation selectivity has been done using micro-flow RPLC separation of standard peptide mixtures. Formic acid is stronger than acetic (pKa 3.75 vs. 4.76), therefore a higher concentration of the latter is usually used, as it provides a similar pH value of the mobile phase. **Figure 3.1a** shows micro-flow RP HPLC separations of two standard calibration mixtures routinely used in the UM laboratory. P2-P6 peptides²⁵ are of similar mass, length (7-10 residues), and carry a 2+ charge at acidic pH. B1-B5 peptides also span a wide range of hydrophobicities, are 15-17 residues long, possess nearly identical mass, and carry a 2+ or 3+ charge. Retention shifts of ~-0.6 % ACN were observed for the 2+ peptides, except for an unchanged retention for B2. 3+ peptides exhibited a larger ~-1% ACN decrease upon switching to AA (**Figure 3.1b**). To confirm possible charge dependence, four mixtures of 1+, 2+, 3+, 4+ species (4 peptides each) have been analysed, showing the same trend – that negative retention shifts increase with charge (**Supplementary Table 3.2**). The B2 peptide (**Figure 3.1b**) represents an exception, which is likely occur due to the Asp residue in its N-terminal position. This results in a lower basicity of the N-terminal amine²⁷ and consequential altering of its ion-pairing environment.

Identification output of LC-MS/MS acquisitions serves as an ultimate criterion for the selection of experimental conditions. In the context of ion-pairing modifiers for RPLC separation, detection sensitivity and chromatographic behaviour (overall retention and peak shape) represent the most important features determining identification output. **Figure 3.1 a, b** show identical peak shapes and slightly lower retention values upon switching to AA. This should lead to decreased ID output as some hydrophilic peptides will not retain and likely be missed when using 0.5% AA as ion-pairing modifier. **Figure 3.2** compares the number of unique peptide identifications obtained in the UM and KU labs for various samples investigated in this study and highlights the difference in detection sensitivity between two ion-pairing systems.

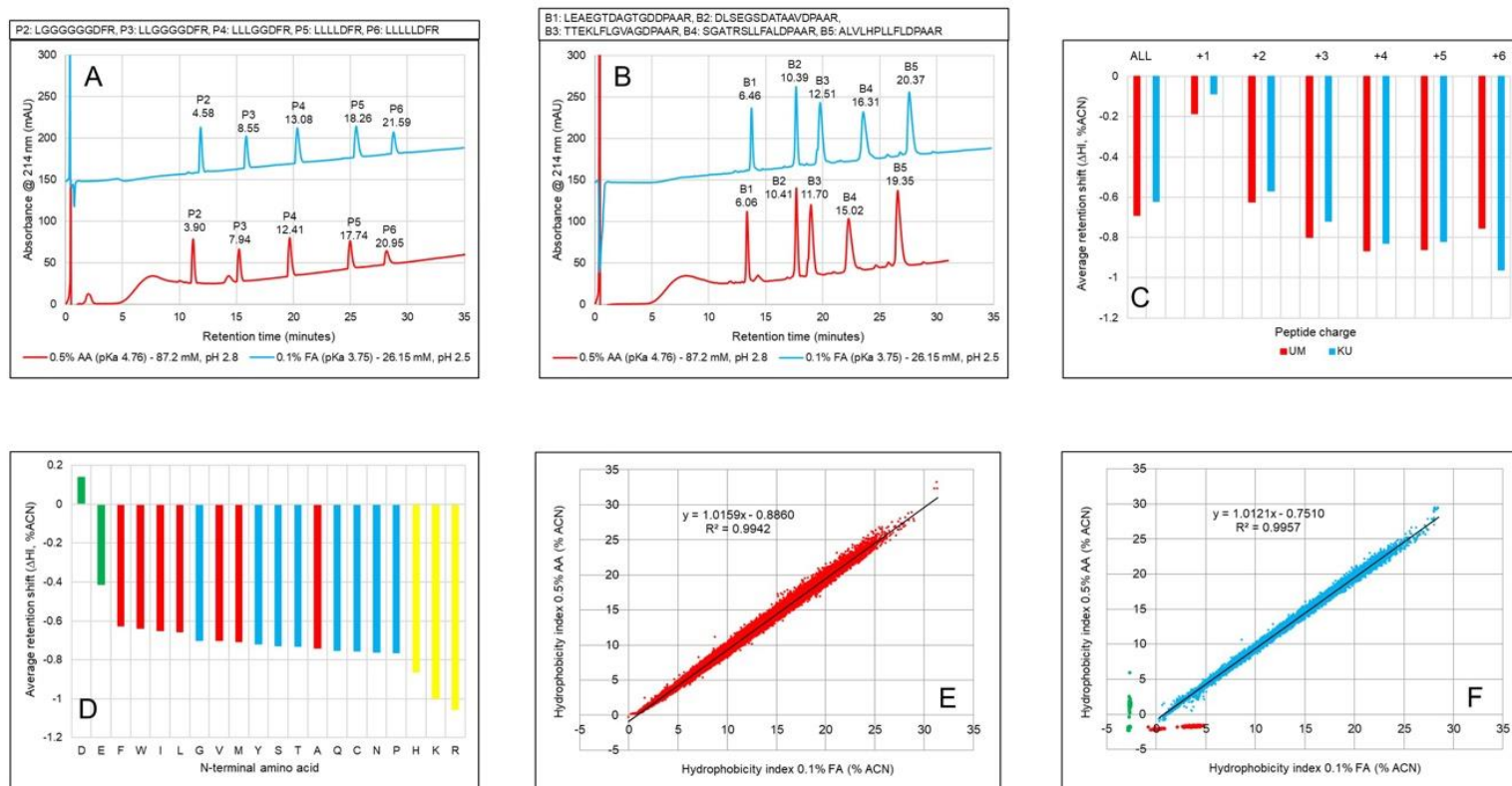


Figure 3.1 Variation of separation selectivity between formic and acetic acid-based eluent systems in peptide RPLC. A, B – separation of P2-P6 and B1-B5 standard peptide mixtures in micro-flow RPLC using 0.1% FA and 0.5% AA as eluent additives; C – average and charge-dependent retention shifts observed in both laboratories; D – average retention shifts observed for 2+ peptides depending on the N-terminal residue; E, F – correlations between retention times of peptides separated under FA and AA conditions.

Both 1D and 2D LC-MS/MS acquisitions using a Triple TOF 5600 showed a moderate increase of 10-15% in peptide identification (**Figure 3.2a**). Considering the average decrease in peptide retention (-0.7% ACN, **Figure 3.1c**), this likely suggests higher detection sensitivity upon switching to AA. The UM lab used the MS2 signal intensity for the constant duration of 100 ms MS/MS acquisitions and found that peptides under 0.5% acetic acid conditions exhibit on average a 2.5-times increase in sensitivity (**Figure 3.2b**).

The efficiency of ESI in the KU study has been estimated using MS1 peak area and showed a 2.15-fold increase (**Figure 2b**). This led to unchanged number of peptide IDs for the whole cell HeLa digest. Switching to the analyses of less complex mixtures of N-terminal peptides from lysargiNase digest and enriched phosphorylated peptides on a Orbitrap Fusion Lumos further confirmed increased sensitivity of AA conditions; the number of unique peptide IDs increased by 5 and 63%, respectively (**Figure 3.2a**). The reason for the better performance of AA-based eluents for the latter samples can be found in the dependence of the signal increase from the difference between precursor charge and the net charge in solution value (**Figure 2c**). Clearly, peptides with low pI values such as acetylated and phosphorylated species benefit the most from switching ion pairing chemistry from FA to AA.

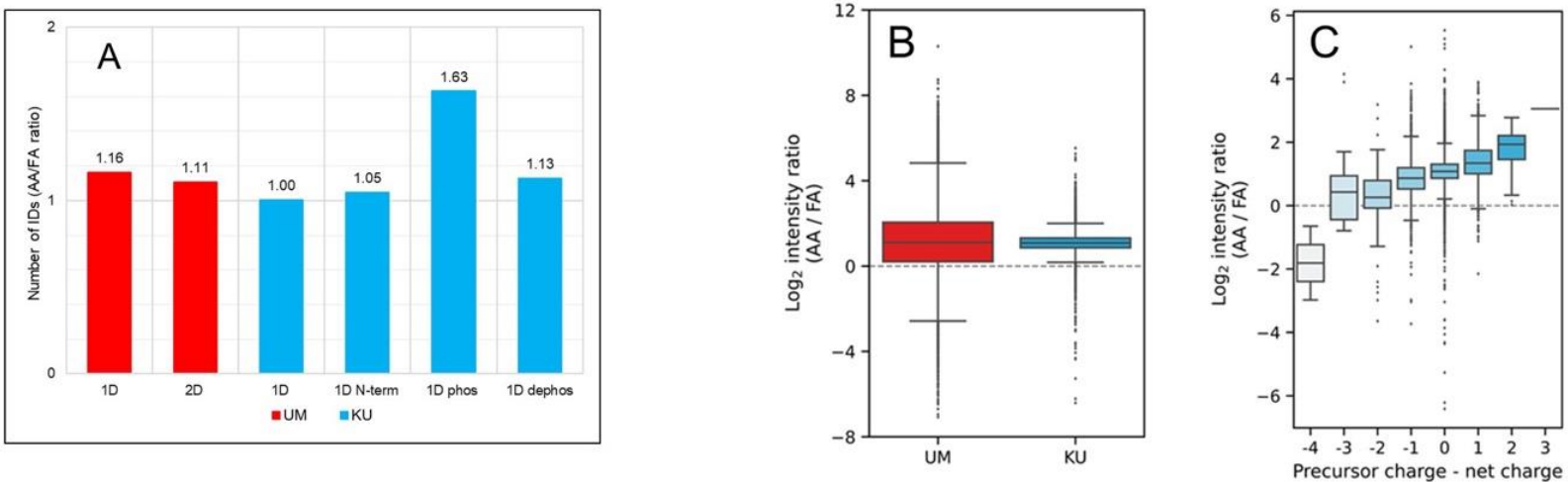


Figure 3.2 The effect of ion pairing modifier on LC-MS/MS ID output and detection sensitivity. A – the number of peptide IDs observed for the 1D and 2D LC-MS/MS analyses depending on ion-pairing chemistry; B – overall change in detection sensitivity observed in UM and KU laboratories; C – variation of peptide MS1 signal in FA-AA system depending on the difference between charge in gas phase and net charge in solution.

Retention behaviour of peptides in nano-RPLC with FA and AA as ion pairing modifiers has been assessed using 41,117 and 12,643 peptide pairs detected in both FA-AA 2D (UM) and 1D (KU) runs, respectively. We find a very similar average retention decrease of -0.69 and -0.62% ACN for the two labs (**Figure 3.1c**). By analysing retention shifts for peptides of different net charge, we confirm the trend found earlier in micro-flow LC experiments: larger negative shifts have been observed for highly charged peptides, albeit with a flattening of this trend for +5 and +6 species.

The correlation between retention values for peptides retained under 0.1% FA and 0.5% AA conditions look virtually identical for both labs (**Figure 3.1 e,f**), confirming the similarity in separation selectivity variations. These plots feature negative 0.8-0.9 intercepts (overall retention decrease), 1.014-1.016 slopes (smaller amplitude of the retention shifts for more hydrophobic peptides), and R^2 -value correlations of 0.994-0.996, indicating very similar separation selectivity between the two ion pairing modifiers. KU used a direct injection set-up resulting in early elution of some hydrophilic peptides under both FA (76 peptides highlighted in green) and AA conditions (136 peptides highlighted in red in **Figure 3.1 f**). Given the usual inconsistency with retention times of hydrophilic species with direct injection, these numbers are in agreement with lower peptide retention upon switching to AA. It was of interest to evaluate the separation selectivity between these modifiers when both were applied at 0.1% concentrations.

Supplementary Figure 3.1 shows a more drastic change in peptide retention with correlation dropping to a R^2 value of 0.983. We attribute this shift to the higher pH value of the 0.1% AA solution, which alters the dissociation status of carboxyl groups. This was confirmed by analysing the most significant outliers, which were represented by the most acidic (retention increase in 0.1% AA) and the most basic peptides.

Composition and sequence-specific factors driving retention shifts between FA and AA based eluents. Adaptation of the SSRCalc model²⁸ for interlaboratory use and retention data collection was one of the major goals of this work. Comparing HI values between two labs showed virtually identical plots for both FA and AA conditions with R^2 -values of ~0.999 (**Supplementary Figure 3.2**). This confirms negligible variation in selectivity due to switching the type of C18 separation media. This also means that the models developed for UM's Luna C18(2) system should be transferrable to KU's Reprisil-Pur C18 AQ settings.

SSRCalc adaptation to AA conditions has been achieved by developing a prediction model for the value of the retention shift $\Delta HI = HI_{AA} - HI_{FA}$, where respective HI values are both expressed in acetonitrile % units derived using spiked peptide standards. Predicted values for peptide retention under AA conditions is then calculated as $SSRCalc HI_{AA} = SSRCalc HI_{FA} + \Delta HI$, following the approach we introduce for peptides carrying post-translational modifications²⁹. Similar to other SSRCalc models, we incorporate composition and sequence-dependent features into the structure of the algorithm²⁸. The former includes the number of positively charged residues, peptide length and hydrophobicity (**Supplementary Figure 3.3**). The latter is dominated by the influence of N-terminal chemistry.

Alterations in ion-pairing environment play an important role in apparent hydrophobicity values for terminal residues and are always being assessed during the optimization of SSRCalc models. **Figure 3.1d** shows average retention shifts observed for the 2+ peptides subset of the UM lab's 2D-LC/MS data, depending on the N-terminal residue. Confirming the observation for the B2 peptide in **Figure 3.1b**, N-terminal Asp and Glu on average show minute positive and negative shifts, respectively. The largest negative shifts are characteristic for positively charged (His, Arg, Lys) and more hydrophilic residues. Hydrophobic amino acids showed intermediate retention shifts, likely indicating a lower degree of ion-pairing shielding from interactions with the C18 stationary phase when using acetic acid.

The core SSRCalc model for C18 FA conditions showed a 0.979 R²-value correlation for the entire collection of 41,117 tryptic peptides (**Supplementary Figure 4.4a**). Following experimental values' plots (**Figure 3.1 e,f**), the SSRCalc(FA) application to 0.5%AA separation showed a similar slope and negative intercept corresponding overall decrease of retentions, and a slightly lower prediction accuracy of 0.975 R²-value (**Supplementary Figure 4.4b**). Once adjusted to AA conditions, the SSRCalc model exhibits identical 0.979 R²-value correlations for acetic acid data in both labs (**Supplementary Figure 4.4 c,d**), with slope 1 and intercept 0, as both predicted and experimental values are expressed in the same ACN % units.

3.5 Discussion

The application of acetic acid as an ion-pairing modifier reduces peptide retention by 0.7% acetonitrile on average, indicating a possible loss of some hydrophilic species during separation. Albeit being important for the analysis of hydrophilic peptides, this would not make a significant impact for most labs, which prefer 3-5% ACN starting conditions in RPLC gradients. In our hands, this disadvantage was compensated by more than 2-fold increase in MS signal, resulting in a higher number of peptide IDs. Depending on the type of sample, we observed up to 60% more identification with the best results found for phosphorylated peptides. Analysis of common whole cell lysate tryptic digests showed a moderate increase, in part due to the superior sensitivity of modern mass spectrometers; a sample load of a few hundred nanogram to microgram of peptides provides enough signal to saturate the available peptide MS/MS capacity, independent of variations in MS signal. Our results also show that the AA retention decrease is less profound for acetylated and phosphorylated peptides. Thus, the average retention of phosphorylated peptides detected in both systems decreased only by 0.2% ACN, demonstrating a further advantage of 0.5% AA for phosphopeptide analysis.

Acetic acid behaves as a more hydrophilic ion-pairing modifier compared to formic acid despite the presence of additional methylene group. This contradiction is observed due to the difference in hydration radius of formate and acetate: 0.24 vs. 0.26 nm, respectively³⁰. This suggests possible steric hindrance between larger acetate counterions, which leads to a relative increase in hydrophobic interactions of terminal residues as shown in **Figure 3.1d**.

Our results on comparing detection sensitivity between FA and AA -based eluents in nano-flow LC have confirmed the recent conclusion of Lenco et al.²¹ for micro-flow format on the superior performance of acetic acid as ion-pairing modifier for the analysis of complex peptide mixtures. Analysis of the current literature, on the other hand, indicates the dominant role of formic acid. Reviewing prior literature on this subject, one may find an early indication of superior sensitivity of FA when applied to ESI MS of proteins and peptides⁷⁻⁹. Another possible argument for the selection of FA in the past can be attributed to the fact that AA-based eluents exhibit common background contaminants (500-600 m/z range) identified as [acetic acid – Fe – O] complexes of various composition¹⁶. The presence of prominent contaminant peaks in the mass range typical for most peptides' m/z could have been a critical deterring factor against selecting AA. On the other hand, gradual improvement in the quality of reagents and

introduction of metal-free LC-MS systems in recent years may have changed this situation. We did not observe any additional background peaks in the AA-based system for the Ultimate 3000 RSLC-nano instrument operating with a metal-free Reprosil-Pur C18 AQ column. LC settings used at the UM lab featured Eksigent's nano 2D Ultra and a home-made Luna C18(2) column with a single stainless steel outlet frit. The latter has likely contributed to the appearance of background peaks (520-550 m/z, **Supplementary Figure 3.5**). Their intensity, however, was lower than 300 counts per second selected as MS/MS acquisition threshold, thus having no impact on ID output.

It remains unclear why early studies⁷⁻⁹ showed advantages of FA contrary to our findings or those by Lenco et al.²¹. Setting aside the possible improvement of chemical purity of the reagents and its effect on detection sensitivity (which would be challenging to verify now), the question remains: will these findings affect the selection of LC conditions in proteomics labs going forward? Given the history of a very slow uptake for seemingly very advantageous procedures such as first dimension RPLC fractionation at high pH³¹ or boosting the sensitivity by DMSO addition to eluents³² – the answer is “very unlikely”. The latter contribution describes a 2-3 fold increase in detection sensitivity using no significant changes in hardware. Yet, it was used in less than 2% of instances reported 7 years later in Volume 19 of the *Journal of Proteome Research*.

3.6 Supporting Information

Supplementary Experimental procedures used in University of Manitoba and Kyoto University laboratories

Supplementary Table 3.1 Tabulated hydrophobicity index (HI, % ACN) values for standard peptides used for retention alignment.

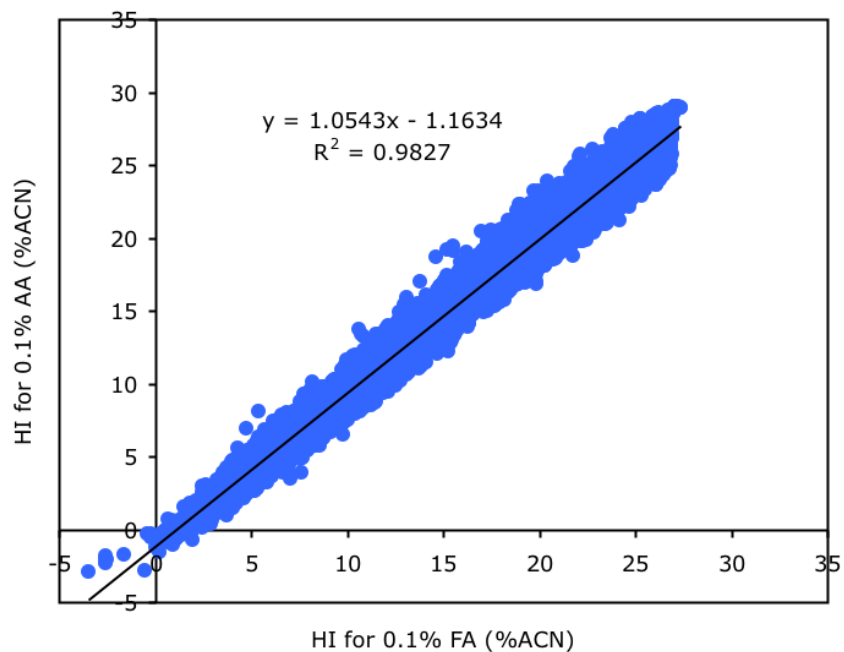
	Sequence	HI 0.1% FA	HI 0.5% AA	Δ HI (%ACN)
P1-P6	P2 – LGGGGGGDFR	4.58*	3.90	-0.68
	P3 – LLGGGGDFR	8.55*	7.94	-0.61
	P4 – LLLGGDFR	13.08*	12.41	-0.67
	P5 – LLLLDFR	18.26*	17.74	-0.52
	P6 - LLLLLDFR	21.59*	20.95	-0.64
B1-B5	B1 - LEAEGTDAGTGDDPAAR	6.46	6.06	-0.4
	B2 - DLSEGSDATAAVDPAAR	10.39	10.41	0.02
	B3 - TTEKLFLGVAGDPAAR	12.51	11.70	-0.81
	B4 - SGATRSLLFALDPAAR	16.31	15.02	-1.29
	B5 - ALVLHPLLFLDPAAR	20.37	19.35	-1.02

* - tabulated HI values from Krokhin, O. V.; Spicer, V. *Anal Chem* **2009**, 81, (22), 9522-30.

Supplementary Table 3.2 Retention times for the standard peptide mixtures of different net charge at acidic pH separated using 0.1% formic and 0.5% acetic acid ion pairing modifiers.

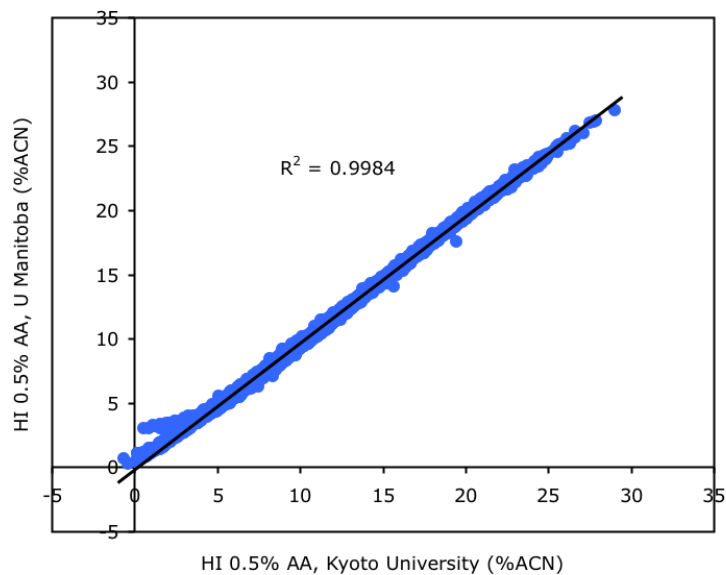
	Sequence*	RT 0.1% FA (min)	RT 0.5% AA (min)	Δ RT (Δ HI, % ACN)	Δ HI (average)
1+	LASAADFG	17.29	17.02	-0.27	-0.1
	LLSAADFG	22.49	22.36	-0.13	
	LLSLADFG	28.13	28.11	-0.02	
	LLSLDFG	31.51	31.52	0.01	
2+	LASAADFR	14.22	13.75	-0.47	-0.42
	LLSAADFR	18.72	18.24	-0.48	
	LLSLADFR	23.66	23.23	-0.43	
	LLSLDFR	26.31	26.02	-0.29	
3+	LASAAHFR	10.38	9.85	-0.53	-0.92
	LLSAAHFR	13.36	12.49	-0.87	
	LLSLAHFR	17.62	16.47	-1.15	
	LLSLHFR	21.16	20.02	-1.14	
4+	LAHAAHFR	8.16	6.61	-1.55	-1.19
	LLHAAHFR	10.4	9.67	-0.73	
	LLHLAHFR	13.93	12.76	-1.17	
	LLHLLHFR	17.16	15.84	-1.32	

* - standard peptide mixtures from Vu, H.; Spicer, V.; Gotfrid, A.; Krokhin, O. V *J Chromatogr A* **2010**, 1217, (4), 489-97.

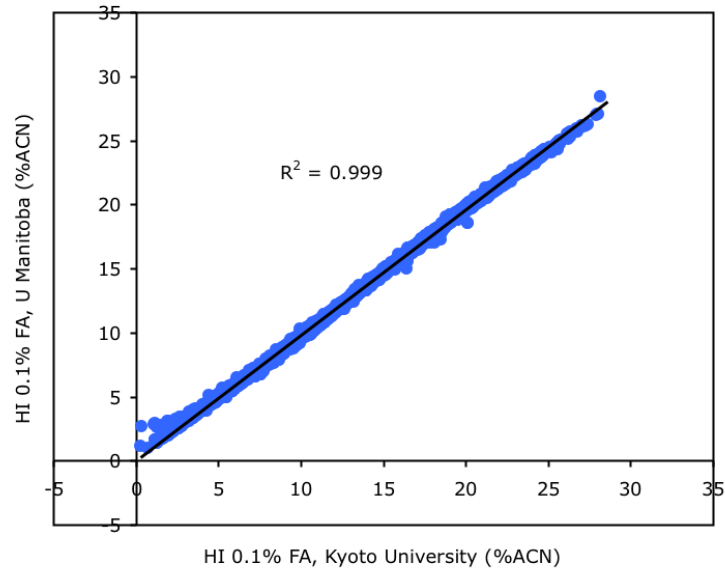


Supplementary Figure 3.1 Correlation between hydrophobicity index values (%ACN) for peptides separated using 0.1% FA and 0.1% AA as ion pairing additives (Kyoto University data).

A

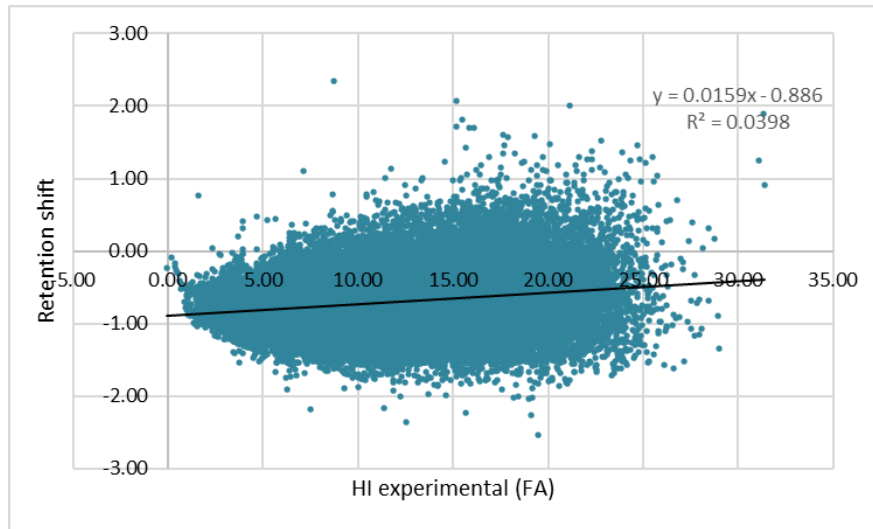


B

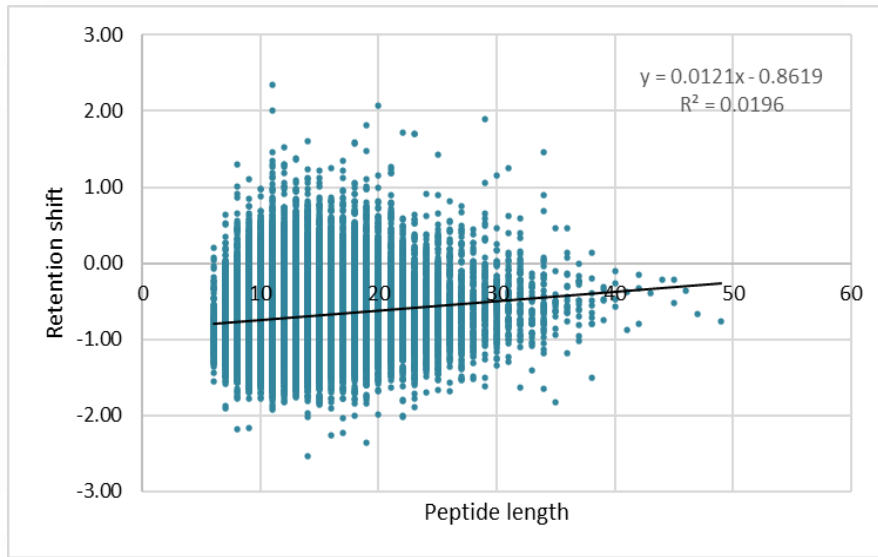


Supplementary Figure 3.2 Correlation between hydrophobicity index values (%ACN) measured in two laboratories for peptides separated using identical eluent conditions (A - 0.5% AA or B - 0.1% FA).

A

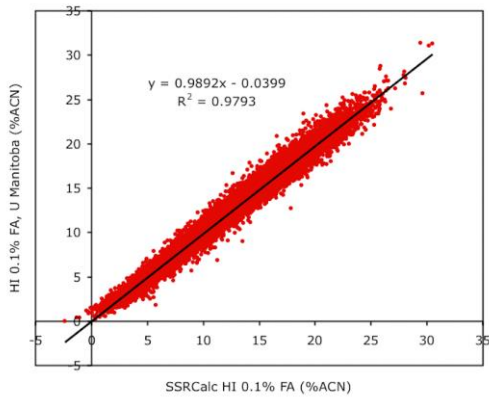


B

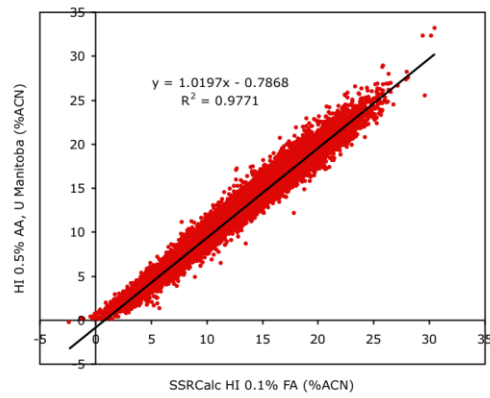


Supplementary Figure 3.3 Dependence of retention shift (DHI, % ACN) upon switching from 0.1% FA to 0.5% AA from peptide hydrophobicity (A) and length (B).

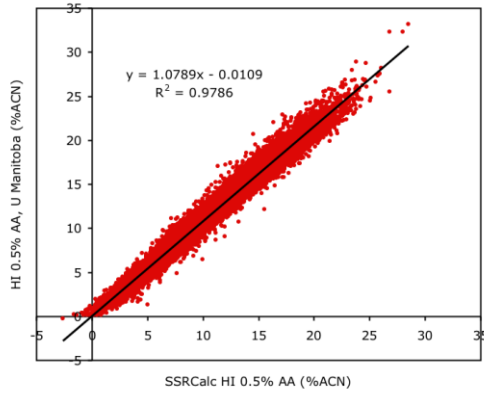
A



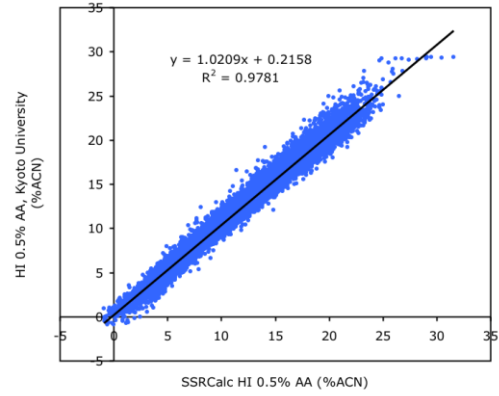
B



C

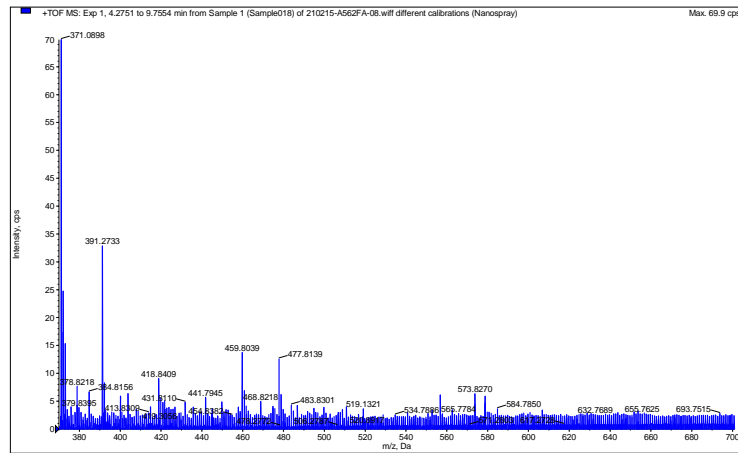


D

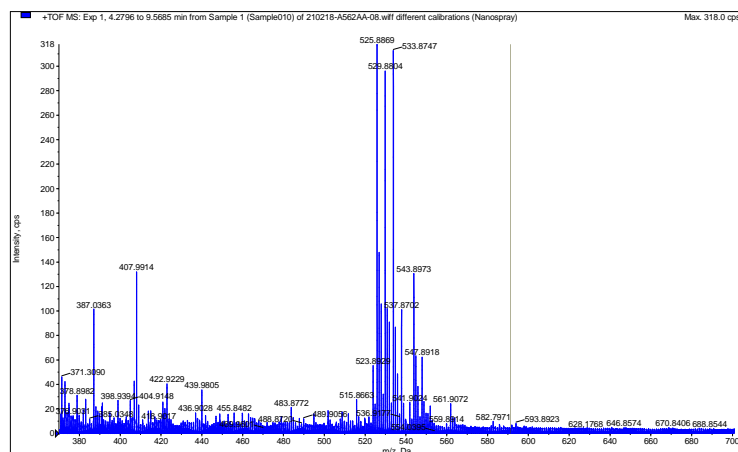


Supplementary Figure 3.4 Peptide retention time prediction using standard (0.1% FA) and modified (0.5% AA) SSRCalc models.

A



B



Supplementary Figure 3.5 Background ESI signals for FA (a) and AA (b) based eluents for Triple TOF 5600 mass spectrometer (UM).

Acknowledgements

This work was supported by the grants from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2016- 05963 O.V.K.), the JST Strategic Basic Research Program CREST No. 18070870 (Y.I.), AMED-CREST program No. JP18gm1010010 (Y.I.), JSPS Grants-in-Aid for Scientific Research No. 21H02459 (Y.I.) and No. 20K21478. (Y.I. & K.O.).

Data Availability

The MS raw data and analysis files have been deposited with the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the jPOST partner repository [33] (<https://jpostdb.org>) with the data set identifier PXD029416.

3.7 References

1. Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M. C.; Yates, J. R., 3rd, Protein analysis by shotgun/bottom-up proteomics. *Chem Rev* **2013**, 113, (4), 2343-94
2. MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J., Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010**, 26, (7), 966-8.
3. Tyanova, S.; Temu, T.; Cox, J., The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* **2016**, 11, (12), 2301-2319.
4. Rogers, J. C.; Bomgardner, R. D., Sample Preparation for Mass Spectrometry-Based Proteomics; from Proteomes to Peptides. *Adv Exp Med Biol* **2016**, 919, 43-62.
5. Ishihama, Y., Proteomic LC-MS systems using nanoscale liquid chromatography with tandem mass spectrometry. *J Chromatogr A* **2005**, 1067, (1-2), 73-83.
6. Garcia, M. C., The effect of the mobile phase additives on sensitivity in the analysis of peptides and proteins by high-performance liquid chromatography-electrospray mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* **2005**, 825, (2), 111-23.
7. Huber, C. G.; Premstaller, A., Evaluation of volatile eluents and electrolytes for high-performance liquid chromatography-electrospray ionization mass spectrometry and capillary electrophoresis-electrospray ionization mass spectrometry of proteins. I. Liquid chromatography. *J Chromatogr A* **1999**, 849, (1), 161-73.
8. Garcia, M. C.; Hogenboom, A. C.; Zappey, H.; Irth, H., Effect of the mobile phase composition on the separation and detection of intact proteins by reversed-phase liquid chromatography-electrospray mass spectrometry. *J Chromatogr A* **2002**, 957, (2), 187-99.
9. Issaq, H.J.; Fox, S.D.; Mahadevan, M.; Conrads, T.P.; Veenstra, T.D., 2003. Effect of experimental parameters on the HPLC separation of peptides and proteins. *J Liq Chromatogr Relat Technol* **2003**, 26, (14), 2255-83.
10. Chong, B. E.; Yan, F.; Lubman, D. M.; Miller, F. R., Chromatofocusing nonporous reversed-phase high-performance liquid chromatography/electrospray ionization time-of-flight mass spectrometry of proteins from human breast cancer whole cell lysates: a novel

- two-dimensional liquid chromatography/mass spectrometry method. *Rapid Commun Mass Spectrom* **2001**, 15, (4), 291-6.
11. Kuhlmann, F. E.; Apffel, A.; Fischer, S. M.; Goldberg, G.; Goodley, P. C., Signal enhancement for gradient reverse-phase high-performance liquid chromatography-electrospray ionization mass spectrometry analysis with trifluoroacetic and other strong acid modifiers by postcolumn addition of propionic acid and isopropanol. *J Am Soc Mass Spectrom* **1995**, 6, (12), 1221-5.
 12. Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., 3rd, Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* **1999**, 17, (7), 676-82.
 13. Washburn, M. P.; Wolters, D.; Yates, J. R., 3rd, Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **2001**, 19, (3), 242-7.
 14. Washburn, M. P.; Ulaszek, R.; Deciu, C.; Schieltz, D. M.; Yates, J. R., 3rd, Analysis of quantitative proteomic data generated via multidimensional protein identification technology. *Anal Chem* **2002**, 74, (7), 1650-7.
 15. Motoyama, A.; Venable, J. D.; Ruse, C. I.; Yates, J. R., 3rd, Automated ultra-high-pressure multidimensional protein identification technology (UHP-MudPIT) for improved peptide identification of proteomic samples. *Anal Chem* **2006**, 78, (14), 5109-18.
 16. Keller, B. O.; Sui, J.; Young, A. B.; Whittall, R. M., Interferences and contaminants encountered in modern mass spectrometry. *Anal Chim Acta* **2008**, 627, (1), 71-81.
 17. Ishihama, Y.; Schmidt, T.; Rappsilber, J.; Mann, M.; Hartl, F. U.; Kerner, M. J.; Frishman, D., Protein abundance profiling of the Escherichia coli cytosol. *BMC Genomics* **2008**, 9, 102.
 18. Zanivan, S.; Maione, F.; Hein, M. Y.; Hernandez-Fernaund, J. R.; Ostasiewicz, P.; Giraud, E.; Mann, M., SILAC-based proteomics of human primary endothelial cell morphogenesis unveils tumor angiogenic markers. *Mol Cell Proteomics* **2013**, 12, (12), 3599-611.

19. Kulak, N. A.; Pichler, G.; Paron, I.; Nagaraj, N.; Mann, M., Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat Methods* **2014**, 11, (3), 319-24.
20. Chang, C. H.; Chang, H. Y.; Rappsilber, J.; Ishihama, Y., Isolation of Acetylated and Unmodified Protein N-Terminal Peptides by Strong Cation Exchange Chromatographic Separation of TrypN-Digested Peptides. *Mol Cell Proteomics* **2020**, 20, 100003.
21. Lenco, J.; Vajrychova, M.; Pimkova, K.; Proksova, M.; Benkova, M.; Klimentova, J.; Tambor, V.; Soukup, O., Conventional-Flow Liquid Chromatography-Mass Spectrometry for Exploratory Bottom-Up Proteomic Analyses. *Anal Chem* **2018**, 90, (8), 5381-5389.
22. Masuda, T.; Tomita, M.; Ishihama, Y. Phase Transfer Surfactant-Aided Trypsin Digestion for Membrane Proteome Analysis. *J. Proteome Res.* **2008**, 7 (2), 731–740.
23. Sugiyama, N.; Masuda, T.; Shinoda, K.; Nakamura, A.; Tomita, M.; Ishihama, Y. Phosphopeptide Enrichment by Aliphatic Hydroxy Acid-Modified Metal Oxide Chromatography for Nano-LC-MS/MS in Proteomics Applications. *Mol. Cell. Proteomics* **2007**, 6 (6), 1103–1109.
24. Krokhin, O. V.; Spicer, V., Peptide retention standards and hydrophobicity indexes in reversed-phase high-performance liquid chromatography of peptides. *Anal Chem* **2009**, 81, (22), 9522-30.
25. Dwivedi, R. C.; Spicer, V.; Harder, M.; Antonovici, M.; Ens, W.; Standing, K. G.; Wilkins, J. A.; Krokhin, O. V., Practical implementation of 2D HPLC scheme with accurate peptide retention prediction in both dimensions for high-throughput bottom-up proteomics. *Anal Chem* **2008**, 80, (18), 7036-42.
26. Cox, J.; Mann, M. MaxQuant Enables High Peptide Identification Rates, Individualized P.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification. *Nat. Biotechnol.* **2008**, 26 (12), 1367–1372.
27. Krokhin, O. V.; Anderson, G.; Spicer, V.; Sun, L.; Dovichi, N. J., Predicting Electrophoretic Mobility of Tryptic Peptides for High-Throughput CZE-MS Analysis. *Anal Chem* **2017**, 89, (3), 2000-2008.
28. Krokhin, O. V., Sequence-specific retention calculator. Algorithm for peptide retention prediction in ion-pair RP-HPLC: application to 300- and 100-Å pore size C18 sorbents. *Anal Chem* **2006**, 78, (22), 7785-95.

29. Villacres, C.; Spicer, V.; Krokhin, O. V., Confident Identification of Citrullination and Carbamylation Assisted by Peptide Retention Time Prediction. *J Proteome Res* **2021**, 20, (3), 1571-1581.
30. Kelewou, H.; Lhassani, A.; Merzouki, M.; Drogui, P.; Sellamuthu, B. Salts retention by nanofiltration membranes: Physicochemical and hydrodynamic approaches and modeling. *Desalination* **2011**, 277, 106–112.
31. Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C., Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. *J Sep Sci* **2005**, 28, (14), 1694-703.
32. Hahne, H.; Pachl, F.; Ruprecht, B.; Maier, S. K.; Klaeger, S.; Helm, D.; Medard, G.; Wilm, M.; Lemeer, S.; Kuster, B., DMSO enhances electrospray response, boosting sensitivity of proteomic experiments. *Nat Methods* **2013**, 10, (10), 989-91.
33. Okuda, S.; Watanabe, Y.; Moriya, Y.; Kawano, S.; Yamamoto, T.; Matsumoto, M.; Takami, T.; Kobayashi, D.; Araki, N.; Yoshizawa, A. C.; Tabata, T.; Sugiyama, N.; Goto, S.; Ishihama, Y. jPOSTrepo: An International Standard Data Repository for Proteomes. *Nucleic Acids Res.* **2017**, 45 (D1), D1107–D1111.

4 Chapter 4: Retention time prediction for phosphopeptides in LC-MS

Peptide retention time prediction for peptides with post-translational modifications:
phosphorylation

Taylor Battellino¹, Darien Yeung², Vic Spicer³, Kosuke Ogata⁴, Yasushi Ishihama⁴, Oleg V. Krokhin^{1,2,3,5*}

¹Department of Chemistry, University of Manitoba, 360 Parker Building, 144 Dysart Road, Winnipeg, R3T 2N2, Canada

²Department of Biochemistry and Medical Genetics, University of Manitoba, 336 BMSB, 745 Bannatyne Avenue, Winnipeg, R3E 0J9, Canada

³Manitoba Centre for Proteomics and Systems Biology, 799 JBRC, 715 McDermot Avenue, Winnipeg, R3E 3P4, Canada

⁴Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto 606-8501, Japan

⁵Department of Internal Medicine, University of Manitoba, 799 JBRC, 715 McDermot Avenue, Winnipeg, R3E 3P4, Canada

To be submitted to the Journal of Chromatography A, January 2022.

4.0 Contributions of Authors

TB, DY, KO, and OK prepared samples and TB, OK, and KO performed their analysis. TB, KO, VS, and OK worked on data treatment. TB and OK contributed to the writing and editing of the manuscript. OK supervised the project.

4.1 Abstract

Development of a peptide retention prediction model for reversed-phase chromatography applications in proteomics is reported for peptides carrying phosphorylated Ser, Thr and Tyr residues. The major retention features have been assessed using collection of 10000+ phosphorylated/non-phosphorylated peptide pairs identified in a series 1D and 2D LC-MS/MS acquisitions using formic acid conditions. Single modification event on average results in increased peptide retention for phosphorylation of Ser (+ 1.46), Thr (+1.33), Tyr (+0.93 % ACN) on a chromatographic scale. At the same time, we established several composition and sequence specific features, which drive deviations from these average values. Thus, single phosphorylation of serine results in retention shifts ranging from -2.4 to 5.5 % ACN depending on position of the residue, nature of nearest neighbour residues, peptide length, hydrophobicity and pI value, and its propensity to form amphipathic helical structures. We established that the altered ion-pairing environment upon phosphorylation is detrimental for this variability. Hydrophobicity of the ion-pairing modifier directly informs the magnitude of expected shifts: (most hydrophilic) 0.5% acetic acid (larger positive shift upon phosphorylation) > 0.1% formic acid (positive) > 0.1% trifluoroacetic (negative) > 0.1% heptafluorobutyric acid (larger negative shift). 2D LC-MS analysis allowed evaluation of phospho-peptide behaviour in high pH reversed-phase and HILIC separation on amide phase (pH 4.5), showing decreased and increased retention of phosphorylated peptides, respectively. The effect of phosphorylation has been also evaluated for six different HILIC columns, strong cation- and strong anion exchange separations using previously acquired 2D LC-MS/MS analyses of yeast digests.

Keywords: post-translational modifications, phosphorylation, peptide retention prediction

4.2 Introduction

Phosphorylation is one of the most common reversible post-translational modifications (PTM), which has been intensively studied in the past two decades due to its biological relevance as a regulator of protein stability, activity and signal transduction.¹⁻³ Although phosphorylation can occur at many amino acid residues, it is most commonly seen on the hydroxyamino acids, serine (Ser, S), threonine (Thr, T) and tyrosine (Tyr, Y) in approximately a 89:10:1 ratio.⁴ Each addition of a phosphate group results in a +79.96633 Da mass increase easily seen by MS and MS/MS analyses. Problems with its detection occur due to lower detection sensitivity (MS signal of more acidic phosphorylated peptides) and the sub-stoichiometric nature of this modification.⁵ As a result, the MS signal produced by phosphopeptides makes up a tiny portion of the total ion count generated in usual LC-MS/MS runs. To deal with this issue, a collection of peptide enrichment protocols has been developed based on IMAC, TiO₂, and SCX enrichment methods.⁶⁻⁹

Regardless of upstream protein/peptide purification steps, reversed-phase HPLC with MS detection (RPLC-MS) remained the method of choice for the final analysis step. Because of this, studying chromatographic behavior of phosphorylated peptides became a hot topic in proteomic and chromatographic literature. The importance of taking RPLC separation properties into account can be illustrated by the results of the ABRF 2003 proteomic case study.¹⁰ Participants were asked to identify protein and detect two phosphorylated peptides from its sequence in provided tryptic digest. All labs but one had failed to find the second phosphorylated peptide – simply because of its hydrophilic nature. Successful analysis has been done in the lab which never specialized in phosphorylation but used a rather unconventional RPLC-MALDI MS combination with 0.1% trifluoroacetic acid (TFA) as ion pairing modifier. This provided increased peptide retention compared to formic or acetic acid (FA, AA), which are generally accepted as the modifiers proving highest detection sensitivity and were used by most participants of that study.

The hydrophilic character of the phosphate group led to the original assumption that this modification should result in decreased retention on RPLC columns. While some reports confirmed this, others found higher retention of phosphorylated species. Steen et al.¹¹ pointed out that this obvious misconception was caused by switching ion-pairing modifiers. TFA based eluents favor retention of non-modified peptides and FA promotes higher retention of the

phosphorylated counterpart in respective pairs. The importance of hydrophobicity of ion-pairing modifiers for phosphopeptide analysis was stressed in several publications.¹²⁻¹⁵ Gorshkov et al.¹² showed the inversion of retention order when FA or TFA were applied to the analysis of phosphorylated/de-phosphorylated peptide pairs. Ogata et al.¹⁶ showed the same trend for AA – TFA modifiers on C18 column, and at the same time – they found lower retention of phosphorylated species was typical for separation on a C4 column using AA. Several publications aimed at finding specific features of peptide sequence and amino acid composition, which will explain observed retention shifts and thus predict expected phosphopeptide behaviour.^{13,14,17} These attempts were largely unsuccessful due to a couple of major reasons. First, the Kim et al. report utilized a limited number of peptides,¹³ thus reducing the scope of generalization. Second, these studies used various ion-pairing agents (FA in different concentrations, TFA) and various stationary phases, thus masking observed features. Lastly, we believe that some of these studies^{14,17} could be affected by low reproducibility of retention time (RT) measurements in nano-flow format. Marcantonio et al.¹⁴ have composed a collection of retention values for ~750 phosphorylated/dephosphorylated peptide pairs using peptide enrichment and alkaline phosphatase dephosphorylation. Yet, they found a great variability of retention shifts: both positive and negative using 0.2% FA as eluent additive. The prevalence of positive retention shifts upon phosphorylation for modification sites closer to the N-termini was the only feature they have established. The authors explained it by formation of salt bridges between positively and negatively charged functional groups. Marx et al.¹⁷ have reported on LC-MS/MS analysis of a synthetic peptide library containing potentially more than 100,000 non-modified – phosphorylated pairs in equal Ser:Thr:Tyr proportion. Analysing ~30,000 pairs detected in one set of the experiments, they concluded that retention time shifts upon phosphorylation (ΔRT) do not depend on peptide length and increase in the order: Tyr < Thr < Ser. 70 % of peptides in this study increased retention upon phosphorylation, 24% decreased, and 6 % remained unchanged. Findings by Kim et al.¹³ confirmed a more significant drop in retention for Tyr phosphorylation, found no dependence of retention time shifts upon phosphorylation and peptide hydrophobicity, and saw a larger amplitude of variations of retention shifts for smaller peptides vs. larger ones. The authors also obtained a 0.977 R² correlation for their SSRCalc¹⁸ retention time prediction when applied to tryptic peptides in their dataset, vs. 0.922 value for phosphorylated species.

The question of RPLC measurement's standardization has been addressed by Krokhin and Spicer,¹⁹ who introduced retention calibrating peptides and proposed a hydrophobicity scale for peptide separation in RPLC based on acetonitrile percentage (hydrophobicity units, HI). Compared to unitless scales it provides direct information on expected retention shifts upon peptide modification and has been used in several PTM/chemical modification chromatographic studies from this lab.

Despite significant efforts undertaken previously, the question of which features of peptide composition and sequence determine a change in RPLC peptide retention following phosphorylation remains open. The aim of this work was to evaluate RPLC retention behaviour of phosphorylated/dephosphorylated peptide pairs using a large dataset composed using strictly controlled (calibrated) RPLC separation and introduce retention prediction for phosphorylated peptides into the SSRCalc algorithm. We opted to use Marcantonio et al.¹⁴ sample preparations approach (peptide enrichment and enzymatic de-phosphorylation followed LC-MS analysis) instead of costly methods based on synthetic peptides. We employed both the most common 0.1% formic and 0.5% acetic acid modifiers and standard peptide retention alignment to ensure the accuracy of RPLC measurements. A general overview of chromatographic behaviour of phosphorylated peptides in other popular HPLC techniques: HILIC, SCX, high pH RPLC, strong anion exchange (SAX) has been also targeted in this study.

4.3 Materials and Methods

4.3.1 Materials, protein digestion, chromatographic columns

All chemicals were analytical chemistry or LC-MS grade and sourced from Sigma Chemicals (St. Louis, MO) unless noted otherwise. Deionized water and HPLC-grade acetonitrile were used for preparation of the eluents. Sequencing grade modified trypsin (Promega, Madison, WI) was used for the digestion. Tryptic digests of human HEK, RPMI 8866, K562, and Jurkat cells were prepared as described elsewhere.²⁰ The Kyoto University lab used whole HeLa cell digests prepared using a phase transfer surfactant aided trypsin digestion protocol.²¹ In both cases Cys residues were reduced and alkylated with iodoacetamide. The resulting mixture of tryptic peptides were purified using reversed-phase HPLC and lyophilized. Alkaline phosphatase was sourced from Thermo Fisher Scientific (Waltham, MA) and Fujifilm Wako, respectively. Designed standard peptides P1-P6¹⁹ and non-modified/phosphorylated peptides (SGVVVGDGFGGR, pSGVVVGDGFGGR) were synthesized by BioSynthesis Inc. (Lewisville, TX).

A home-packed 1x100 mm Luna C18(2) 5 μ m (Phenomenex, Torrance, CA) was used for the separation of standard peptides. 1 \times 100 mm packed with 5 μ m Xterra C18 (Waters, Milford, MA), and 3 \times 50 mm 3 μ m XBridge Amide (Waters, Milford, MA) columns were used for high pH RP and HILIC first dimension separations, respectively. Second dimension LC-MS settings at the University of Manitoba featured a 100 μ m \times 200 mm column packed with 3 μ m Luna C18 (2) (Phenomenex, Torrance, CA). Comparison of formic and acetic acid separation selectivity for phosphopeptide analysis at Kyoto University lab was done using 100 μ m x 15 cm in-house packed Reprisil-Pur C18 AQ 1.9 μ m (Dr. Maisch, Germany). Previously acquired 2D LC-MS/MS data for yeast digests²²⁻²⁵ using various columns in the first dimension have been re-analysed to extract retention data of phosphorylated peptides: Luna HILIC 3 μ m 200 \AA (3x50 mm), Luna (2) Silica 3 μ m 100 \AA (3x50mm) (all Phenomenex), Atlantis Silica 3 μ m 100 \AA (3x50) (Waters), ZIC-HILIC 3.5 μ m 100 \AA (2.1x100), ZIC-cHILIC 3 μ m 100 \AA (2.1x100) (Sequant, Merck, Darmstadt, Germany), IonPac AS24 (2.1x250mm), (Thermo Fisher, Sunnyvale, CA), PolySulfoethyl A 5 μ m (2.1x100 mm), (Poly LC, Columbia, MD).

4.3.2 First dimension separation in 2D LC-MS runs

First dimension separations have been performed using an Agilent 1100 series HPLC system with a manual 100 μ L loop injector and UV detection at 214 nm. HILIC separation (XBridge Amide) employed a 1% water per min linear gradient starting with 10% water – 90% ACN, with both eluents containing 20 mM pH 4.5 ammonium formate at a 300 μ L/min flow rate.²² Thirty-one 1 min fractions were collected with a 11-41 min peptide elution range. High pH RP (XTerra C18) (Waters) used 10 mM pH 10 ammonium formate as eluent additive with a 1.8% ACN per min linear gradient at 150 μ L/min.²⁶ Eighteen 1 min fractions were collected within a 6-22 min elution window. Fractions were lyophilized and redissolved in buffer A prior to the second dimension LC-MS/MS.

4.3.3. Peptide enrichment and dephosphorylation

Major steps of experimental procedures are outlined in **Figure 1**. Whole-cell tryptic digests (~200 μ g for each experiment) of various cell lines (HEK, RPMI 8866, K562, Jurkat) were enriched for phosphopeptides using SCX or TiO₂ (Titansphere Phos-TiO₂ Kit from GL Sciences) according to manufacturer procedure. Strong cation exchange (SCX) chromatography (PolySULFOETHYL A from PolyLC Inc.) was used to process ~4 mg (20 injections) of Jurkat digest for each of 2D LC-MS/MS runs using experimental conditions²⁴ described in Supplementary Information. Phosphopeptide isolation in Kyoto University was performed as described previously.²⁷ One third of the enriched phosphopeptide mixture was dephosphorylated using alkaline phosphatase. Phosphorylated and de-phosphorylated samples were spiked with the standard peptides¹⁹ and analysed separately by LC-MS/MS. The dephosphorylated portion of the SCX isolate (from ~4 mg of starting material) was re-combined with phosphorylated peptides prior to 2D LC-MS/MS analyses. Each fraction for 2D acquisitions were spiked with the standard peptides for retention alignment purposes.

4.3.4. Second dimension LC-MS/MS

The University of Manitoba lab used an Easy Nano LC system with a QExactive HF-X hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fisher) with a gradient of 1% to 50% B (A: 100% water, B: 80:20 ACN: water, both containing 0.1% formic acid) in 107 min, followed by 1 min rapid increase to 100%B and 12 min wash (500 nL/min flow rate). Kyoto University

lab employed an Ultimate 3000 RSLCnano and an Orbitrap Fusion Lumos (HCD-iontrap) mass spectrometer (Thermo Fisher). Peptides were separated using a linear 1% to 32% ACN gradient over 62 min (0.5% per min) at a 500 nL/min flow rate. Details of data-dependent acquisitions settings for the QExactive HF-X and Orbitrap Fusion Lumos mass spectrometers are given in the Supplementary Information.

4.3.5 Data analysis and peptide identifications

The X!Tandem search engine was used for peptide identifications. Mass errors for parents and fragment ions were set to ± 10 ppm and 20 ppm, respectively. Static Cys modification (alkylation with iodoacetamide, +57.021 Da) was used together with potential phosphorylation (+79.996 Da) for Ser, Thr and Tyr residues. All identifications with $\log(e) < -2$ and unequivocal assignment of phosphorylation sites have been additionally filtered using SSRCalc retention time prediction. Combined identifications results yielded ~10,000 unique pairs of phosphorylated peptides and their non-phosphorylated analogs. Retention values were expressed in Hydrophobicity Index (HI, % acetonitrile units) using linear regression against tabulated HI values for standard peptides in each sample (fraction). Retention shifts upon phosphorylation were calculated as: $\Delta HI = HI_{PO4} - HI_{NM}$, where HI_{PO4} and HI_{NM} are retention values for phosphorylated and non-modified peptides, respectively. Retention times in the first (HILIC, high pH RP) dimension were assigned based on the fraction number in which this peptide was found and converted into %ACN or % water using experimental gradient slope settings. pI values were calculated using an online IPC 2.0 prediction tool.²⁸

4.4 Results and Discussion

4.4.1. Phosphopeptide enrichment and identification in 1D and 2D LC-MS runs.

Table 4.1 shows the summary of peptide identification for series of 1D and 2D LC-MS runs performed for various enrichment protocols. Optimization of the latter or/and attaining a record number of phosphopeptide identifications was beyond the scope of this work. However, we wanted to explore one additional option for enrichment enabled by SCX separation at acidic pH. SCX separations feature a single sharp peak between 1+ and 2+ subpopulations of peptides (**Figure 4.1**). We observed it during SCX modeling study using yeast digests²⁴ and even more so when separating digests of human cell lines. To determine the content of this peak, we collected this fraction and performed LC-MS/MS analysis. 61% of identified peptides were phosphorylated – similar to some results obtained using a TiO₂- based strategy. Surprisingly, 92% of this population consisted of singly phosphorylated tryptic species without internal basic residues (Lys, Arg, His) or doubly phosphorylated ones with one basic residue: i.e. peptides carrying a 1+ net charge at acidic pH. Such selectivity represents a stark contrast when compared to ~40-50% of 1+ phosphorylated species found in TiO₂ isolates.

The reason why these molecules elute as one sharp peak under SCX conditions remains unclear. SCX retention is driven by peptide charge and size, thus each charge group is spread over the SCX elution scale. All peptides in this fraction behave as molecules of identical size, carrying an intermediate charge of ~+1.5. Multiple SCX isolations (using ~4 mg of starting digest material) of this peak have been used to collect a sufficient amount of phosphorylated peptides with the idea to increase identification output in 2D LC-MS analyses. However, the final results were largely disappointing (**Table 4.1**): the number of identified peptide pairs did not scale up according to increased analysis time and sample load. Most likely this happens due to the loss of phosphopeptides during first dimension separations and handling of fractions. Using high pH RP in the first dimension (standard first dimension application for proteomics) may additionally result the peptide loss due to hydrophilic character of phosphate at higher pH.

Table 4.1 Identification summary for the analyses of various phosphorylated isolates.

Analysis	Unique phospho-peptides	Unique un phospho-peptides	% phospho-peptides	% of 1+ net charge in phosphorylated subset
TiO ₂ -based enrichment				
Jurkat (#1)	2733	467	85.4	48.2
Jurkat (#2)	3186	389	89.1	41.2
H8866	3352	1756	65.6	41.5
K562	3034	2284	57.1	41.3
HeLa (FA)*	1933	939	67.3	51.7
HeLa (AA)*	3266	1282	71.8	46.6
SCX-based enrichment				
Jurkat (#3)	3007	1899	61.3	92.4**
2D Jurkat (high pH RP)	2148	3362	38.9***	89.6**
2D Jurkat (HILIC)	6157	7491	45.1***	92.0**

*analyses of identical phospho-isolates performed in Kyoto University using 0.1% formic or 0.5% acetic acid ion-pairing modifiers in RPLC;

**SCX enrichment of phospho-peptides showed superior selectivity towards peptides carrying +1 charge in solution;

***2D LC-MS analyses have been performed using a mixture of phosphorylated/dephosphorylated peptides, thus resulting in lower percentage of identified phosphorylated species.

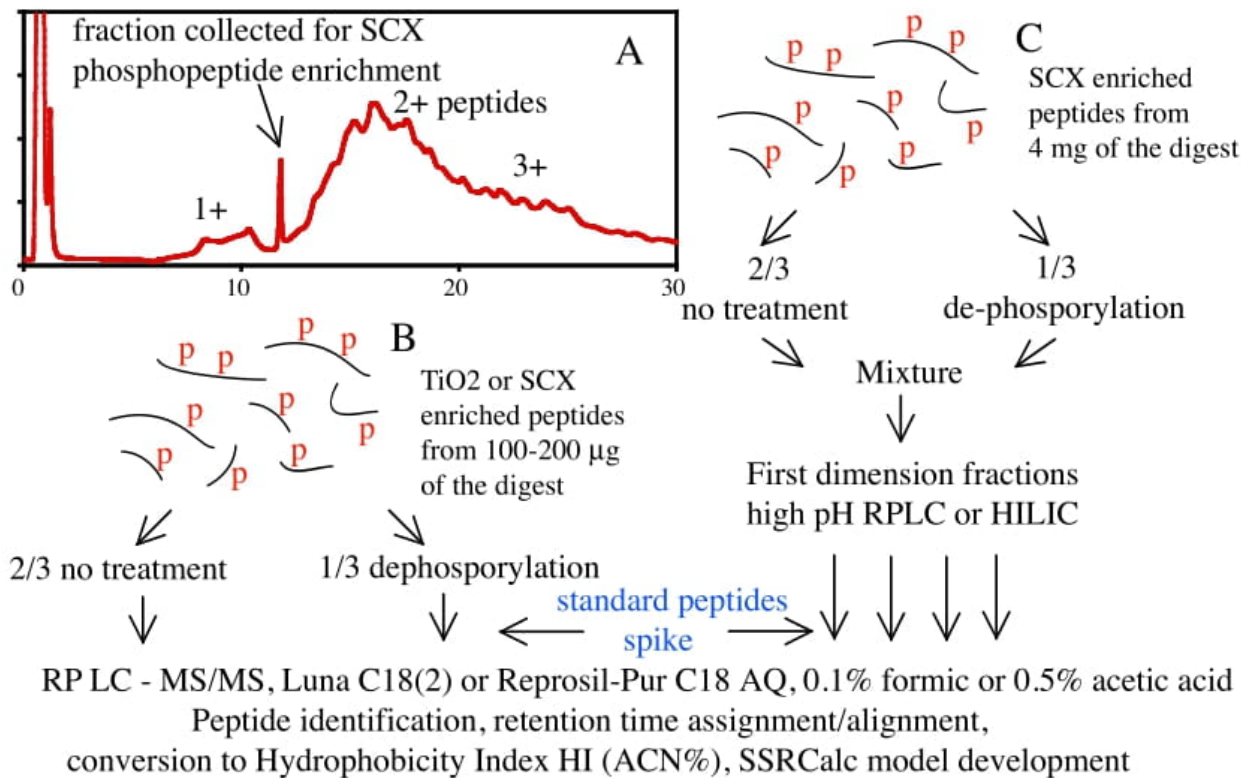


Figure 4.1 Experimental procedures used for retention data collection in 1D and 2D LC-MS/MS proteomic experiments targeting retention time prediction of phosphorylated peptides. A – SCX separation of whole Jurkat cell lysate tryptic digest; B, C – sample processing and LC-MS workflow for 1D and 2D experiments.

The combined redundant ID list included 20,000+ peptide pairs. **Supplementary Table 4.1** shows examples of redundant identifications across several runs and illustrates a high reproducibility of retention value measurements when alignment protocols have been applied. Averaging retention shifts while removing redundant identifications resulted in 10,000+ unique modified/non-modified peptide pairs, which were used for retention modeling.

4.4.2. Modified-non-modified pairs retention selectivity: the effect of ion-pairing modifiers.

It was of interest to confirm previous findings on the effect of ion pairing modifier on separation selectivity of non-modified – phosphorylated pairs. **Figure 4.2A** shows a steady increase in peptide retention in the FA < TFA < HFBA sequence. The latter is characterized by a large negative shift upon phosphorylation, while application of formic acid leads to increased retention of the pSGVVVGDGFGGR peptide compared to its non-modified analog. TFA

exhibits an intermediate (slightly positive) shift. This data agrees with previous reports on the effect of hydrophobicity of ion-pairing modifiers. Moreover, recently we studied (manuscript in preparation) separation selectivity variations upon switching between 0.1% formic and 0.5% acetic acid. The latter was more hydrophilic resulting in a -0.7% acetonitrile retention shift on average for usual tryptic digests. Analysis of phosphorylated – non-modified peptides pairs in these systems confirmed that 0.5% acetic acid further increases positive retention shifts upon phosphorylation (**Figure 4.2B**).

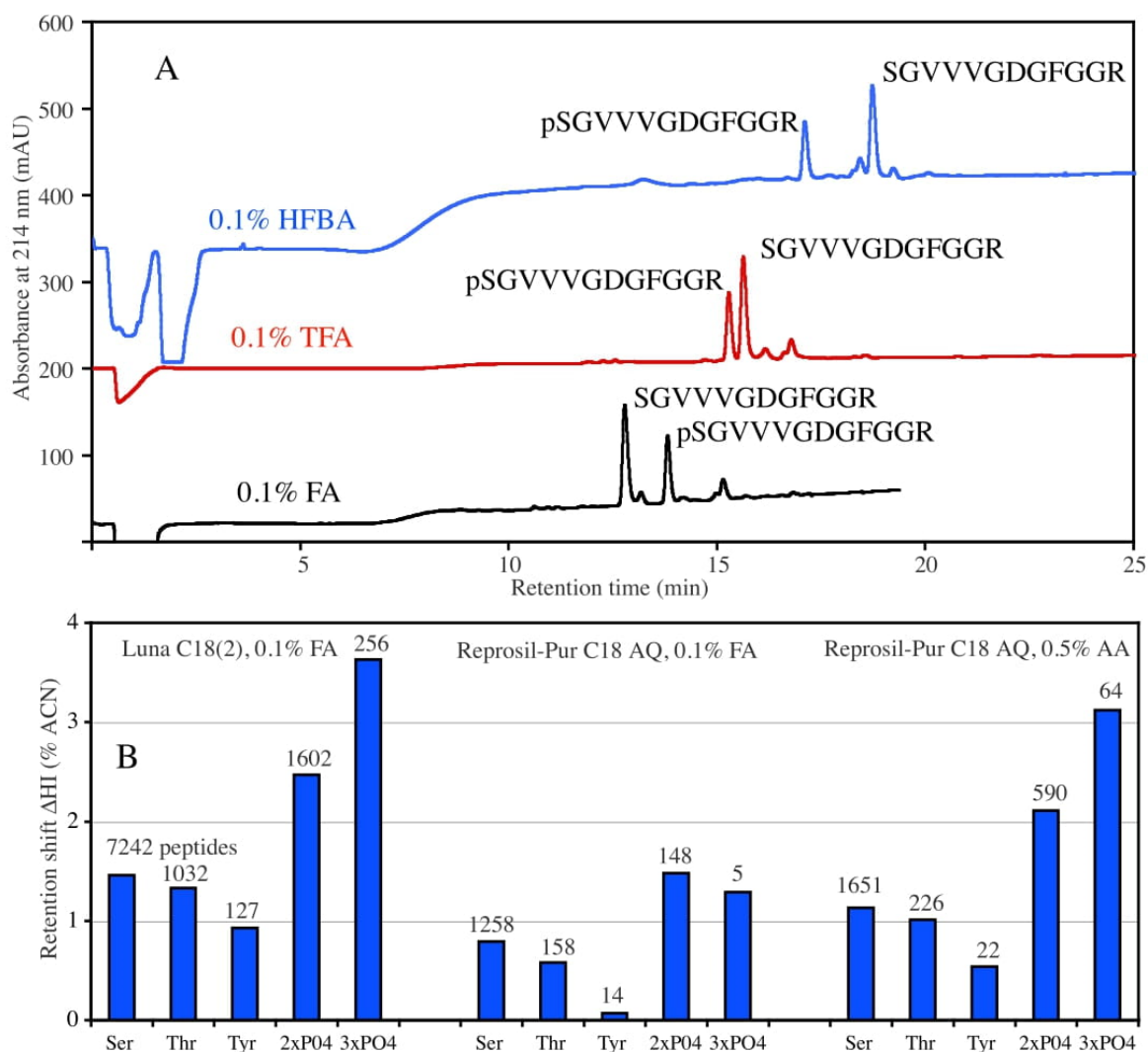


Figure 4.2 Effect of ion pairing modifier on separation selectivity of modified – non-modified peptide pairs. A – RPLC separation of SGVVVGDGFGGR-pSGVVVGDGFGGR pair using 0.1% FA, 0.1% TFA and 0.1% HFBA with identical 2% ACN/min gradient; B – average retention shifts upon phosphorylation for Luna C18(2) 0.1% FA, Reprisil-Pur C18 AQ using 0.1% FA and 0.5% AA.

This allows us to conclude that the change in RPLC peptide retention upon phosphorylation is governed by the hydrophobicity of the ion-pairing modifier. Our previous studies on RPLC retention modeling^{18,29} all indicate the importance of taking into account the ion pairing mechanism. Addition of negatively charged functional groups such as phosphate reduces the basicity of primary amines (N- terminus, side chain of Lys) resulting in a lower number of associated counter ions. The impact of this on peptide retention depends on the hydrophobicity

of ion-pairing counterions and the nature of neighboring residues. The HFBA counter-ion is the most hydrophobic – its loss results in decreased RPLC retention. Application of hydrophilic FA and AA as ion-pairing modifiers leads to increased retention of phosphorylated peptides. Notably, when studying the effect of asparagine glycosylation³⁰ we found very similar trend: despite hydrophilic character of monosaccharide moieties, the addition of sialic acid led to increased retention of glycopeptides in RPLC system with 0.1% FA as eluent additive.

4.4.3. The effect of phosphorylation on peptide retention in proteomic RPLC-MS experiments: 0.1% formic acid as ion pairing modifier

Formic acid is the most popular ion-pairing modifier applied in proteomics due to its superior ESI efficiency (sensitivity) and satisfactory chromatographic characteristics (separation efficiency, peak capacity). **Figure 4.2B** shows average retention shifts depending on modified residue and number of phosphorylation sites for Luna C18(2) and Reprosil-Pur C18 AQ stationary phases used in this work. On average, single phosphorylation results in a 1.44% ACN retention increase for Luna C18(2). Retention shifts increase in the order of decreased hydrophobicity: Tyr < Thr < Ser. The same trend was established for Reprosil Pur C18 AQ, however the amplitude of the average shift for single phosphorylation was found two-times smaller: 0.76% ACN. Peptides with multiple phosphorylations exhibit larger retention increases.

While both Luna C18(2) and Reprosil-Pur C18 AQ show similar retention trends, the difference in the amplitude is noticeable. Only 203 of our 8401 singly phosphorylated peptides (2.4%) decreased in retention upon phosphorylation for Luna C18(2). Correspondent numbers for Reprosil Pur C18 AQ were 9.4% and 24% in this work and a report by Marx et al.,¹⁷ respectively. All together this suggests that the effect of peptide phosphorylation is driven by both composition of the mobile phase and properties of the stationary phase. The latter must be considered when prediction models are intended to cover various types of chromatographic columns.

4.4.4. Composition and sequence specific effects of peptide phosphorylation

Previous findings on this subject were limited to Kim et al.¹³ and Marcantonio et al.¹⁴ reports. The former established no effect of peptide hydrophobicity, and a decreased amplitude of retention shifts with peptide length. The latter showed that phosphorylation near the N-termini increases the amplitude of positive retention shifts upon phosphorylation. These two publications used 33 and 750 correspondent peptide pairs. Our data include more than unique 10,000 Δ HI measurements. Since retention shifts depend on the type of residue, it is reasonable to assess the behaviour of respective subpopulations. Our data contains the information on 7242 Ser and 1031 Thr single phosphorylations - sufficient for such analysis, albeit conclusions made for Ser subpopulation would be more confident.

All SSRCalc peptide retention time prediction models consider overall peptide hydrophobicity, length, isoelectric point (all composition dependent), nature of N-terminal residues, nearest neighbour effects, and contribution of peptide propensity to form helical structures (sequence dependent properties). **Figure 4.3** and the following discussion summarizes our major findings:

- the magnitude of the retention increase (Δ HI) is lower for longer peptides (**Figure 4.3A**);
- the dependence of retention shifts from peptide hydrophobicity is generally flat (slight increase on average) with larger Δ HI characteristic for peptides of medium hydrophobicity (**Figure 4.3B**);
- acidic peptides exhibit smaller retention change upon phosphorylation as shown in **Figure 4.3C** (Δ HI vs. pI plot);
- the magnitude of retention increase systematically drops with the distance from N-termini for both, Ser and Thr (**Figure 4.3D**);
- larger retention shifts are typical for peptides featuring hydrophobic N-termini and hydrophobic residues flanking phosphorylation site (**Figure 4.3E-G**);
- extremely large shifts are observed for relatively short peptides with amphipathic helical sequences starting at N-termini (**Figure 4.3H**).

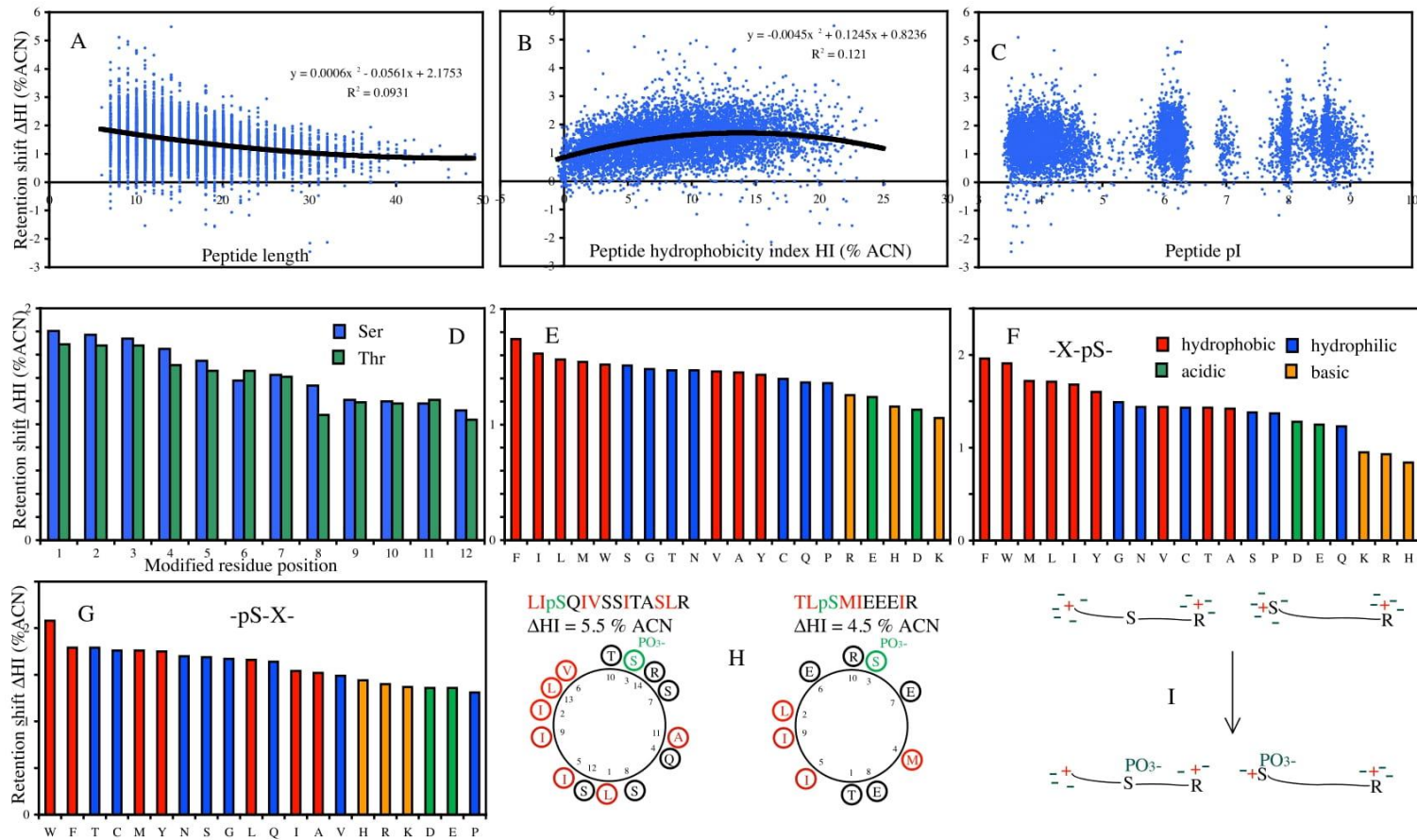


Figure 4.3 Composition and sequence specific features driving variations in dHI – retention shift upon phosphorylation. A, B, C – dependence of dHI for a single Ser phosphorylation from peptide length, hydrophobicity and pI; D – effect of the distance of modified residue from peptide N-termini for Ser and Thr (average values); E – G – average retention shifts depending on N-terminal, preceding and residue following modified Ser; H – axial helical projections for peptides with extremely high retention shifts; I – alterations in ion pairing environment leading to variation of RPLC retention following phosphorylation.

Considering these observations and explaining them shed light on underlying mechanisms responsible for the variations of ΔHI values. The vast majority of them are driven by the alteration of the ion-pairing environment following phosphorylation (**Figure 4.3I**). Larger acidity of modified peptides leads to a reduced number of associated counter-ions. This in turn reduces their shielding effect and causes a larger increase for peptides with hydrophobic N-termini. A similar mechanism explains the gradual reduction of retention shift when the modification site is moved further from the N-terminus. Peptides with a lower pI (large number of acidic residues Asp, Glu) are affected by phosphorylation to a smaller degree since their ion pairing association is smaller to begin with. **Figure 4.3C** shows a slight increase in ΔHI for peptides with a larger pI. Considering 100 peptides with the smallest and largest shifts confirms this trend: the former have 5.34 and the latter 6.24 average pI values, respectively. Similar to many PTMs or chemical modifications, the effect on RPLC retention is smaller for longer peptides due to simple steric effects. All together this creates a situation in which the smallest (sometimes negative) ΔHI s are characteristic for long relatively acidic peptides.

The largest ΔHI values have been observed for relatively short peptides featuring hydrophobic residues in positions 1, 2, and 5, 6 (4, 5), plus phosphorylation in positions 2 (predominantly) and 3. Such combinations correspond to the potential formation of an amphipathic helical structure at the N-termini. In non-modified peptides, stabilization of such structures is suppressed due to the presence of hydrophilic formate counterions near the N-terminus. Phosphorylation reduces the number of associated counterions and thus increases interactions of the amphipathic helix with the C18 phase leading to increased retention (**Figure 4.3H**).

The nearest neighbour effect was evaluated by calculating average ΔHI values depending on amino acids flanking the modification site. Contrary to our expectation that negatively charged hydrophilic functional groups would reduce hydrophobic contribution for neighboring residues, we observed the largest increase for hydrophobic (especially aromatic) residues. Acidic and basic neighbours (K, R, H, D, E) led to lower ΔHI s. The only exception from this rule was when modified Ser (Thr) was followed by Pro. It should be noted that SP, TP sequences represent the most common phosphorylation site motifs; approximately 50% of phosphorylation in our dataset carried the SP/TP sequence. At the same time SP/TP motifs are some of the most

common helix N-cap stabilization sequences.³¹ In proteins these motifs serve as a start signal for alpha-helices. Being located prior to an amphipathic helical stretch, they also increase the retention of peptides in RP HPLC. N-cap stabilization occurs due to unique ability of SP/TP sequences to form hydrogen bonds with otherwise unpaired CO and NH groups at the edge of the helix. Hydrogen bonding reduces the accessibility of hydrophilic functional groups and should result in increased RPLC retention. We find that phosphorylation leads to the lowest retention increase when the modification site is followed by Pro. Therefore, phosphorylation diminishes the N-capping effect.

4.4.5. Modified version of SSRCalc model taking phosphorylation into account

Figure 4 A and B shows the retention time prediction for both dimensions of 2D HILIC-RPLC MS runs performed in this work using 4000 peptide pairs identified. Increased retention of phosphorylated species leads to increased intercepts and changing slopes when the non-modified model is applied to phosphorylated species. Similar to other PTMs^{30,32,33} our approach to phosphopeptide retention modeling is based on predicting retention shift following modification: $SSRCalc_{PO4} = SSRCalc_{HI_{NM}} + \Delta HI_{PO4}$, where $SSRCalc_{HI_{NM}}$ represents a core RPLC model for non-modified peptides at 0.1% formic acid conditions. A predictive model for ΔHI_{PO4} , was built using a dataset of 10,258 experimental ΔHI values.

The ΔHI_{PO4} model coded every phosphorylation site and its N and C-terminal neighbouring amino acids [X@X], with N-terminal phospho-sites coding the next two amino acids [@XX]. Position-scaling variables accounted for the site's distance from the N or C terminal of the sequence and the nature of N-terminal amino acid. As in our prior correction models, these scaled coefficient values were summed to form the core correction ΔHI_{PO4} . The correction also included offset variables to account for the presence of acidic residues aspartic acid and glutamic acid (Asp, Glu) in the sequence, and sequence length factor. Optimization of the variable space was done using a hill-climbing approach, where a variable and signed offset was randomly selected. If the offset- modified variable value yielded an improved correlation between predicted $SSRCalc_{HI_{MOD}}$ and observed HI values, the optimization continued moving the variable by the offset value until the correlation stopped improving.

Figure 4.4C shows the performance of the modified SSRCalc RPLC model adapted to take phosphorylation into account. The resulting correlation for modified peptides has the same slope and intercept as for their unmodified counterparts. The accuracy drops to R^2 0.962 compared to non-modified peptides, likely due to our limited ability to model amphipathic helicity.

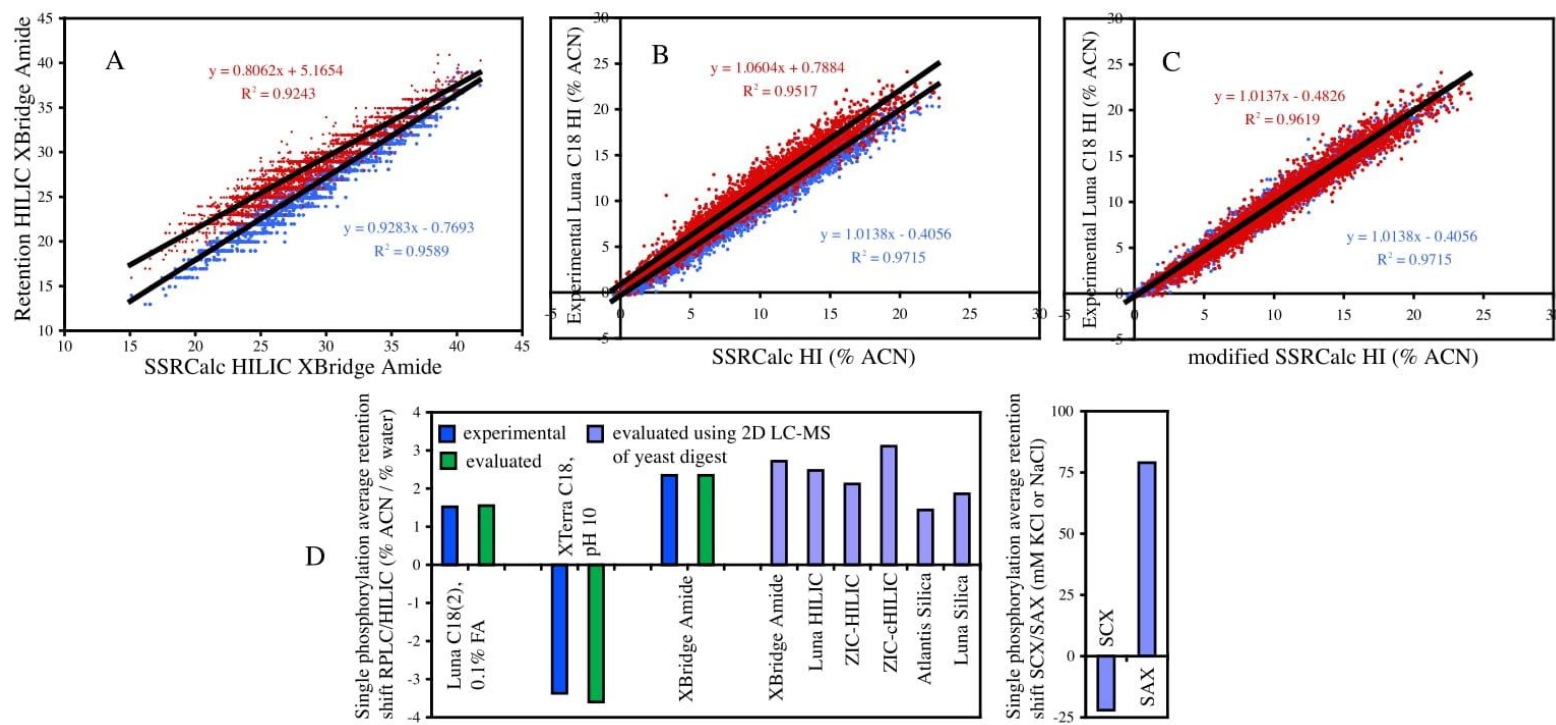


Figure 4.4 Retention time prediction for phosphorylated peptides. A – retention time prediction for HILIC Xbridge amide separation using SSRCalc HILIC model; B, C - retention time prediction for RPLC (formic acid) using core SSRCalc (formic acid) and model adjusted to take into account phosphorylation; D – Retention shifts observed for various peptide separation modes.

4.4.6. The effect peptide phosphorylation on retention in various separation modes

Assigning retention values for both modified and non-modified peptides represents the most straightforward and accurate way to determine retention shifts ($\Delta RT = RT_{PO_4} - RT_{NM}$). However, such capability is not always available. Most proteomics analyses are done for unmodified digests without targeted addition or removal of specific modifications to generate representative peptide pairs. In this case retention shifts can be estimated based on the experimental measurements for modified peptides and retention modeling of non-modified peptides. Taking first dimension HILIC separation on an XBridge Amide column (**Figure 4.4A**) as an example, ΔRT can be calculated using predicted values as $\Delta RT = RT_{PO_4} - (0.9283 * SSRCalc_{HILIC} - 0.7693)$. Based on this calculation we found the average deviation for a single phosphorylation was equal to 2.35 min or 2.35% water (for 1% per minute gradient). The average ΔRT calculated using experimental values was identical. 1.52% vs. 1.55% ACN average retention increase was found for single phosphorylation under RPLC formic acid conditions (**Figure 4.4B, D**). Addition of a single hydrophilic PO_4 group under basic pH RPLC conditions reduces peptide retention by ~3.5% ACN: -3.37 experimental vs. 3.6 calculated.

All together this attests the high accuracy of previously developed SSRcalc models and allows the evaluation of the effect of phosphorylation on peptide retention in several additional separation modes investigated previously in our lab.²²⁻²⁵ These analyses were performed for untreated yeast digests, thus resulting in lower number of phosphorylated peptides, which in many cases did not have respective non-modified counterparts. **Figure 4.4D** summarizes the influence of phosphorylation on these columns/separation conditions. Expectedly, all HILIC separations at pH 4.5 showed a retention increase due to the addition of a hydrophilic phosphate group. At the same time there was a specific effect of distribution of charged functional groups for silica and zwitterionic stationary phases. Both silica columns exhibited lower ΔRT due to electrostatic repulsion of negatively charged silanol groups. A similar situation was found for ZIC-HILIC columns featuring a distal negative charge location of the zwitterionic pair. Conversely, ZIC-cHILIC exhibit largest retention increase due to the distal positioning of positively charged groups.

The effect of phosphorylation in ion-exchange separations is governed by electrostatic interactions and led to a ~80 mM (NaCl) retention increase in strong anion exchange separations.

SCX retention decreased by ~22 mM (KCl). It should be noted that the latter is affected by the co-elution of singly phosphorylated peptides carrying 1+ net charge at acidic pH as discussed in section 4.3.1.

4.5 Conclusions

Peptide retention time prediction represents one of the most popular subfields of separation science in recent years. This interest was driven by both classical chromatography specialists and proteomic practitioners. The former aimed at better understanding separation mechanisms and the latter were searching for the advanced tools to improve confidence of peptide identifications by taking into account retention data. Applied to detection of phosphorylation events, the most common question was: is it possible to unequivocally assign phosphorylation site (e.g. distinguish between neighboring residues) in the absence of sufficient MS/MS fragmentation information based on chromatographic data? Based on large collection of retention data acquired in this work, the answer is “most likely no”. An average retention shifts upon phosphorylation for Ser and Thr differ by 0.13% acetonitrile, which is significantly lower current accuracy of retention prediction models. While positional specificity relative to peptide N-termini was also clear, the differences in the shifts for closely spaced residues were also very small. Extremely large retention shifts have been found for peptides with amphipathic helical sequences at N-termini. On the other hand, these situations were relatively rare.

At the same time, this first detail study of phosphorylation further confirmed major features of RPLC separation mechanism in which most of the sequence specificity is driven by the changes in ion-pairing environment. Phosphorylation reduces positive net charge of peptides separated under common RPLC conditions at acidic pH. The reduction of ion-pairing association changes peptide retention according to hydrophobic properties of ion-pairing modifiers used. Overall peptide retention increases in the order: AA < FA < TFA < HFBA. While retention shifts induced by phosphorylation reduce in the same order from positive to negative. The same effect of ion-pairing environment results in larger retention shifts for peptides with hydrophobic N-termini and for modification sites located closer to N-termini. The addition of phosphorylation to the SSRCalc model continues the trend for comprehensive coverage of post-translational modification by prediction algorithms and further uptake of retention prediction methods by proteomic community.

4.6 Acknowledgements

This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2016-05963; O.V.K.). The authors also want to thank Dr. L. Kozłowski for help with calculating pI values.

4.7 Supplementary Information

Supplementary Table 4.1 The effect of peptide phosphorylation on chromatographic retention in different separation modes.

Separation mode (number of peptide pairs)	Experimental DHI = HI _{Ac} - HI _{NL}	Predicted DHI = HI _{Ac} - SSRCalcHI _{NL}
RP HPLC, 0.1% formic acid		
1 Ser (7242)	1.46% ACN	
1 Thr (1032)	1.33% ACN	
1 Tyr (127)	0.93% ACN	
1 Phos (8401)	1.43% ACN	
2 Phos (1602)	2.47% ACN	
3 Phos (256)	3.63% ACN	
1 N-term Ser (464)	1.80% ACN	
All peptides (10259)		
RP HPLC, 0.1% FA (Kyoto)		
1 Ser (1258)	0.79% ACN	
1 Thr (158)	0.58% ACN	
1 Tyr (14)	0.07% ACN	
2 Phos (148)	1.48% ACN	
3 Phos (5)	1.29% ACN	
All peptides (1583)		
RP HPLC, 0.5% AA (Kyoto)		
1 Ser (1651)	1.13% ACN	
1 Thr (226)	1.01% ACN	
1 Tyr (22)	0.54% ACN	
2 Phos (590)	2.11% ACN	
3 Phos (64)	3.12% ACN	
All peptides (2553)		

RP HPLC, pH 10		
1 Ser (1260)	-3.41 %ACN	
1 Thr (245)	-3.24 %ACN	
1 Tyr (20)	-3.53 %ACN	
2 Phos (292)	-4.14 %ACN	
All peptides (1817)		
XBridge amide, pH 4.5		
1 Ser (3436)	2.36% water	
1 Thr (545)	2.17% water	
1 Tyr (48)	3.76% water	
2 Phos (398)	2.69% water	
All peptides (4427)		
Single Phosphorylation		
Luna HILIC (83)		2.48% water
XBridge amide (133)		2.72% water
ZIC-HILIC (167)		2.12% water
ZIC-cHILIC (176)		3.11% water
Atlantis Silica (124)		1.44% water
Luna Silica (79)		1.86% water
SCX (990)		-2.61 min (-22.2 mM KCl)
SAX (257)		9.92 min (79.4 mM NaCl)

* - experimental data from references 22-24, 34.

4.8 References

1. Lemeer, S.; Heck, A. J., The phosphoproteomics data explosion. *Curr Opin Chem Biol* **2009**, 13, (4), 414-20.
2. Hunter, T., Signaling--2000 and beyond. *Cell* **2000**, 100, (1), 113-27.
3. Riley, N. M.; Coon, J. J., Phosphoproteomics in the age of rapid and deep proteome profiling. *Anal Chem* **2016**, 88, (1), 74-94.
4. Olsen, J. V.; Blagoev, B.; Gnadt, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M., Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **2006**, 127, (3), 635-48.
5. Thingholm, T. E.; Jensen, O. N.; Larsen, M. R., Analytical strategies for phosphoproteomics. *Proteomics* **2009**, 9, (6), 1451-68.
6. Kweon, H. K.; Hakansson, K., Selective zirconium dioxide-based enrichment of phosphorylated peptides for mass spectrometric analysis. *Anal Chem* **2006**, 78, (6), 1743-9.
7. Zhou, H.; Ye, M.; Dong, J.; Corradini, E.; Cristobal, A.; Heck, A. J.; Zou, H.; Mohammed, S., Robust phosphoproteome enrichment using monodisperse microsphere-based immobilized titanium (IV) ion affinity chromatography. *Nat Protoc* **2013**, 8, (3), 461-80.
8. Tsai, C. F.; Wang, Y. T.; Yen, H. Y.; Tsou, C. C.; Ku, W. C.; Lin, P. Y.; Chen, H. Y.; Nesvizhskii, A. I.; Ishihama, Y.; Chen, Y. J., Large-scale determination of absolute phosphorylation stoichiometries in human cells by motif-targeting quantitative proteomics. *Nat Commun* **2015**, 6, 6622.
9. Beausoleil, S. A.; Jedrychowski, M.; Schwartz, D.; Elias, J. E.; Villen, J.; Li, J.; Cohn, M. A.; Cantley, L. C.; Gygi, S. P., Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci U S A* **2004**, 101, (33), 12130-5.
10. Krokhin, O. V.; Ens, W.; Standing, K. G., MALDI QqTOF MS combined with off-line HPLC for characterization of protein primary structure and post-translational modifications. *J Biomol Tech* **2005**, 16, (4), 429-40.
11. Steen, H.; Jebanathirajah, J. A.; Rush, J.; Morrice, N.; Kirschner, M. W., Phosphorylation analysis by mass spectrometry: myths, facts, and the consequences for qualitative and quantitative measurements. *Mol Cell Proteomics* **2006**, 5, (1), 172-81.

12. Perlova, T. Y.; Goloborodko, A. A.; Margolin, Y.; Pridatchenko, M. L.; Tarasova, I. A.; Gorshkov, A. V.; Moskovets, E.; Ivanov, A. R.; Gorshkov, M. V., Retention time prediction using the model of liquid chromatography of biomacromolecules at critical conditions in LC-MS phosphopeptide analysis. *Proteomics* **2010**, 10, (19), 3458-68.
13. Kim, J.; Petritis, K.; Shen, Y.; Camp, D. G., 2nd; Moore, R. J.; Smith, R. D., Phosphopeptide elution times in reversed-phase liquid chromatography. *J Chromatogr A* **2007**, 1172, (1), 9-18.
14. Marcantonio, M.; Trost, M.; Courcelles, M.; Desjardins, M.; Thibault, P., Combined enzymatic and data mining approaches for comprehensive phosphoproteome analyses: application to cell signaling events of interferon-gamma-stimulated macrophages. *Mol Cell Proteomics* **2008**, 7, (4), 645-60.
15. Tholey, A.; Toll, H.; Huber, C. G., Separation and detection of phosphorylated and nonphosphorylated peptides in liquid chromatography-mass spectrometry using monolithic columns and acidic or alkaline mobile phases. *Anal Chem* **2005**, 77, (14), 4618-25.
16. Ogata, K.; Krokhin, O. V.; Ishihama, Y., Retention order reversal of phosphorylated and unphosphorylated peptides in reversed-phase LC/MS. *Anal Sci* **2018**, 34, (9), 1037-1041.
17. Marx, H.; Lemeer, S.; Schliep, J. E.; Matheron, L.; Mohammed, S.; Cox, J.; Mann, M.; Heck, A. J.; Kuster, B., A large synthetic peptide and phosphopeptide reference library for mass spectrometry-based proteomics. *Nat Biotechnol* **2013**, 31, (6), 557-64.
18. Krokhin, O. V., Sequence-specific retention calculator. Algorithm for peptide retention prediction in ion-pair RP-HPLC: application to 300- and 100-Å pore size C18 sorbents. *Anal Chem* **2006**, 78, (22), 7785-95.
19. Krokhin, O. V.; Spicer, V., Peptide retention standards and hydrophobicity indexes in reversed-phase high-performance liquid chromatography of peptides. *Anal Chem* **2009**, 81, (22), 9522-30.
20. Spicer, V.; Ezzati, P.; Neustaeter, H.; Beavis, R. C.; Wilkins, J. A.; Krokhin, O. V., 3D HPLC-MS with reversed-phase separation functionality in all three dimensions for large-scale bottom-up proteomics and peptide retention data collection. *Anal Chem* **2016**, 88, (5), 2847-55.

21. Masuda, T.; Tomita, M.; Ishihama, Y., Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. *J Proteome Res* **2008**, 7, (2), 731-40.
22. Spicer, V.; Krokhin, O. V., Peptide retention time prediction in hydrophilic interaction liquid chromatography. Comparison of separation selectivity between bare silica and bonded stationary phases. *J Chromatogr A* **2018**, 1534, 75-84.
23. Yeung, D.; Klaassen, N.; Mizero, B.; Spicer, V.; Krokhin, O. V., Peptide retention time prediction in hydrophilic interaction liquid chromatography: Zwitter-ionic sulfoalkylbetaine and phosphorylcholine stationary phases. *J Chromatogr A* **2020**, 460909.
24. Gussakovsky, D.; Neustaeter, H.; Spicer, V.; Krokhin, O. V., sequence-specific model for peptide retention time prediction in strong cation exchange chromatography. *Anal Chem* **2017**, 89, (21), 11795-11802.
25. Yeung, D.; Mizero, B.; Gussakovsky, D.; Klaassen, N.; Lao, Y.; Spicer, V.; Krokhin, O. V., Separation orthogonality in liquid chromatography-mass spectrometry for proteomic applications: comparison of 16 different two-dimensional combinations. *Anal Chem* **2020**, 92, (5), 3904-3912.
26. Dwivedi, R. C.; Spicer, V.; Harder, M.; Antonovici, M.; Ens, W.; Standing, K. G.; Wilkins, J. A.; Krokhin, O. V., Practical implementation of 2D HPLC scheme with accurate peptide retention prediction in both dimensions for high-throughput bottom-up proteomics. *Anal Chem* **2008**, 80, (18), 7036-42.
27. Wakabayashi, M.; Kyono, Y.; Sugiyama, N.; Ishihama, Y., extended coverage of singly and multiply phosphorylated peptides from a single titanium dioxide microcolumn. *Anal Chem* **2015**, 87, (20), 10213-21.
28. Kozlowski, L. P., IPC 2.0: prediction of isoelectric point and pKa dissociation constants. *Nucleic Acids Res* **2021**, 49, (W1), W285-W292.
29. Krokhin, O. V.; Craig, R.; Spicer, V.; Ens, W.; Standing, K. G.; Beavis, R. C.; Wilkins, J. A., 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), 908-19.

5 Chapter 5: Analysis of glycosylation and phosphorylation profiles of HexM and GM2

This chapter includes information from a published manuscript in *BBA Advances*⁸ as well as a poster published in the *Proceedings of the 68th ASMS Conference on Mass Spectrometry and Allied Topics*.¹⁰

5.1 Introduction

Tay-Sachs disease is an autosomal recessive disorder known as a lysosomal storage disorder (LSD). LSDs result from the inability of the lysosome to degrade substrates such as gangliosides.¹ In unaffected individuals, β -hexosaminidase A (HexA) interacts with GM2-activator protein (GM2AP) to hydrolyze GM2 ganglioside into GM3 ganglioside. HexA is synthesized from *HEXA* and *HEXB* genes, where *HEXA* encodes the α -subunit and *HEXB* encodes the β -subunit.² In Tay-Sachs patients, mutations in the *HEXA* gene yield non-functional α -subunits, preventing HexA from effectively hydrolyzing GM2. There have been over 100 different mutations identified in the *HEXA* gene.³ As a result of any of these mutations, HexA becomes dysfunctional and GM2 accumulates to toxic levels in cells, notably neurons.⁴ This resulting central nervous system (CNS) deterioration is ultimately what leads to the fatality of Tay-Sachs disease and many other LSDs.

The severity and progression of Tay-Sachs disease differs based on the nature of the inherited mutation, but the average life expectancy of approximately five years remains the same. Enzyme replacement therapy (ERT) is currently the most universal treatment for LSDs, however it cannot relieve the neurological effects of Tay-Sachs disease as the large lysosomal enzymes cannot cross the blood-brain barrier (BBB). ERT typically consists of weekly, life-long intravenous injections of recombinant lysosomal enzymes. It is known that ERT proves more powerful when started at a young age, but unknown if there is a threshold at which CNS deterioration can be prevented.⁵ Other treatments for LSDs include chaperone therapy and substrate reduction therapy, though these treatments are limited as well by the rapid progression of these diseases.

Aiming for ERT which can alleviate the neurological effects of Tay-Sachs disease, a smaller, recombinant, hybrid enzyme called HexM was engineered using the most important features of the HexA α - and β -subunits. HexM is composed of a single subunit, denoted as the μ -subunit, which includes the active site of HexA's α -subunit. It also includes the components of HexA's β -subunit which are necessary for interaction with GM2AP.⁶ A key feature of this

enzyme is the presence of mannose-6-phosphate (M6P), which serves as a tag to direct HexM to the lysosome.¹ ERT relies on the M6P pathway in order to deliver HexM to the lysosome where it can metabolize GM2. Therefore, the presence of phosphorylated high-mannose N-glycans is crucial. These glycans are modified in the Golgi apparatus by an enzyme called GlcNAc-1-phosphotransferase in order to include the M6P component. A hyper-phosphorylated variation of HexM called phosHexM was then created by upregulating this enzyme in cells.

As HexM acts on GM2 ganglioside as a substrate, it is important to monitor both HexM activity and GM2 levels in experimental analysis. While HexM activity can be monitored via enzyme assays using other detectable substrates, GM2 levels in the brains of mice with *HEXA* knockouts can be assessed using relative quantitation by MALDI-MS.

5.2 Experimental

5.2.1 Chemicals

Ammonium bicarbonate, dithiothreitol (DTT), iodoacetamide (IAA), trifluoroacetic acid (TFA), acetonitrile (ACN), iodomethane, d-iodomethane, chloroform, and 2,5-dihydroxybenzoic acid (DHB) were purchased from Sigma-Aldrich (St. Louis, MO). C18 cartridges were purchased from Phenomenex (Torrance, CA). Endoproteinase GluC was purchased from New England Biolabs (Ipswich, MA).

5.2.2 Protein digestion and sample preparation

HexM, phosHexM (100 µg) were suspended in 50 mM ammonium bicarbonate (pH 8.5). Samples were reduced with 10 mM DTT (56°C, 45 min) and alkylated with 50 mM IAA (room temperature, in the dark, 30 min). Excess IAA was quenched by adding 15 mM DTT. Samples were cleaned and fractionated (9 fractions per sample) using C18 cartridges. GluC was added at a 1:50 enzyme to substrate ratio for overnight digestion (16-18 h) at 37°C.⁸

5.2.3 MALDI

Fractions were lyophilized and resuspended in 20 µL of 0.1% TFA in 30:70 ACN to water (TA30). They were mixed with DHB matrix solution (20 mg/mL in TA30) at a 1:1 ratio. Each mixture (1 µL) was spotted onto the stainless-steel MALDI target and allowed to dry. The following peptide calibration mixture (American Peptide Company, Vista, CA) was used: Bradykinin(1–7) 757.3992; angiotensin II 1046.542; angiotensin I 1296.685; substance P 1346.735; bombesin 1619.822; ACTH clip (1–17) 2093.086; ACTH clip (18–39) 2465.198; ACTH (1–39) 4539.267, where ACTH = adrenocorticotrophic hormone and numbers are calculated m/z values of $[M+H]^+$ ions. 1 µL of a TA30 solution containing 0.125 µg of each protein was mixed at a 1:1 ratio with DHB matrix solution and spotted onto the target. An ultrafleXtreme™ mass spectrometer (Bruker Daltonics, Bremen, Germany) was used in reflector positive mode for analysis.⁸

5.2.4 ESI

Fractions were lyophilized and resuspended in 0.1% FA. Approximately 1 µg of protein per fraction was injected into a Ultra 2D-nano LC system (Eksigent, Dublin, CA) with a Triple TOF 5600 (Sciex, Concord, ON) hyphenated through an IonSpray III source. Samples were

spiked with ~400 fmole of P1-P6 standard peptides and eluted using a linear water/acetonitrile 0.5-35% ACN gradient at 500 nL/min flow rate. The instrument was operated in standard MS/MS data-dependent acquisition mode.

5.2.5 Ganglioside permethylation and sample preparation

Gangliosides extracted from 100 mg of brain tissue from mice were provided in 20 μ L of a 1:1 chloroform to methanol solution. Wild-type (WT), HEXA-knockout (KO) and HEXA/Neu3-knockout (DKO) samples were permethylated with 0.1 mL CH_3I and magnetically stirred for 10 minutes. 100 nmol of purified GM2 ganglioside from Tay-Sachs patients was permethylated in the same manner, but using CD_3I instead of CH_3I . Samples were permethylated according to the procedure specified in reference 7.⁷ This permethylation procedure was repeated a second time for the tissue extracts.

5.2.6 Quantification

Samples were lyophilized and resuspended in 10 μ L ACN. They were mixed with a 10 mg/mL DHB matrix solution (4:1 water to ACN). 0.25 μ L of either standard or sample was mixed with 0.75 μ L matrix solution and spotted onto the stainless steel target. Spiked samples were comprised of 0.25 μ L standard (0.25 nmol), 0.25 μ L sample, and 0.50 μ L matrix solution. The peptide calibration standard and MALDI settings used are the same as mentioned above.

5.3 Results and Discussion

5.3.1 HexM

Mass spectrometric analysis of both pepsin and Glu-C digests of HexM, hyperphosphorylated, and HexA were carried out in order to analyze their glycoproteomic profiles. There are three known glycosylation sites in these proteins, all of which are N-glycosylated and therefore contain an Asn attachment site. In the μ -subunit of HexM, these sites are N115, N157, and N295. Their equivalents in the α -subunit of HexA are N115, N157, and N296. The full sequence of HexA compared to HexM is shown in **Figure 5.1**.⁶ HexM peptides resulting from Glu-C digestion which include these sites are SVE¹¹⁵NYTLTINDDQCLLLSE (with a MW of 2169), GTFFI¹⁵⁷NKTE (MW 1055) and FDTPGHTLSWGPGIPGLLTPCYSGSEPSGTFGPNPSL²⁹⁵NNTYE (MW 4463).

Glu-C was selected as the proteolytic enzyme over trypsin for MALDI analysis, as trypsin produces peptides which are too large for analysis by MALDI. Analysis by ESI was done using previously created peptic digests.

α -subunit	1	MTSSRLWFSLLLAFAFRATALWPPQNFQTSQRYVLYPNNFQFQYDVSSAAQPGCSV	60
μ -subunit	1	MTSSRLWFSLLLAFAFRATALWPPQNFQTSQRYVLYPNNFQFQYDVSSAAQPGCSV	60
α -subunit	61	LDEAFQRYRDLLFGSGSWPRPYLTGKRHTLEKNLVVSVVTPGCNQLPTLESVENYTLTI	120
μ -subunit	61	LDEAFQRYRDLLFGSGSWPRPYLTGKRHTLEKNLVVSVVTPGCNQLPTLESVENYTLTI	120
α -subunit	121	NDDQCLLLSETVWGALRGLETFSQLVWKSAGETFFI NKTEIEDFPRFPHRGLLLDTSRHY	180
μ -subunit	121	NDDQCLLLSETVWGALRGLETFSQLVWKSAGETFFI NKTEIEDFPRFPHRGLLLDTSRHY	180
α -subunit	181	LPIKSSILDITLDMAYNKLNVFHWHLVDDPSFPYESFTFPELMRKGSYNPVTHIYTAQDVK	240
μ -subunit	181	LPIKSSILDITLDMAYNKLNVFHWHLVDDPSFPYESFTFPELMRKGSYS-LSHIYTAQDVK	239
α -subunit	241	EVIEYARLRGIRVLAEFDTPGHTLSWGPGIPGLLTPCYSGSEPSGTFGPNPSL NNTYEF	300
μ -subunit	240	EVIEYARLRGIRVLAEFDTPGHTLSWGPGIPGLLTPCYSGSEPSGTFGPNPSL NNTYEF	299
α -subunit	301	MSTFFLEVSSVFPDFYLHLGGDEVDFTCWKSNIQDFMRKKGFGEDFKQLESFYIQTLL	360
μ -subunit	300	MSTFFLEVSSVFPDFYLHLGGDEVDFTCWKSNIQDFMRKKGFGEDFKQLESFYIQTLL	359
α -subunit	361	DIVSSYGKGYVVWQEVFDNKVKIQPDTIIQVWREDIPVNYMKELELVTKAGFRALLSAPW	420
μ -subunit	360	DIVSSYGKGYVVWQEVFDNKVKIQPDTIIQVWREDIPVNYMKELELVTKAGFRALLSAPW	419
α -subunit	421	YLNRISSGPDWRFYIYEPLAFEGTPEQKALVIGGEACMWGEYVDTNINLVPRLWPRAGAV	480
μ -subunit	420	YLNRISSGPDWRFYIYEPLAFEGTPEQKALVIGGEACMWGEYVDTNINLVPRLWPRAGAV	479
α -subunit	481	AERLWSNKLTSDLTFAYERLSHFRCEILRRGVAAQPLNVGFCFQEFQET	529
μ -subunit	480	AERLWSNKLTRMDDAYDRLSHFRCEILRRGVAAQPLVAGYCNQEFQET	528

Figure 5.1 Compared amino acid sequences of the HexA α -subunit and HexM μ -subunit. N-glycosylation sites are in yellow, blue and red. Amino acid mutations are circled or framed.

Amended from reference 6.⁶

The peptide GTFFINKTE, containing site N157, was the most prominent peptide detected by MALDI analysis in all three samples. The addition of a Man-7 or H7N2 glycan to this peptide results in a mass increase of 1541 units. The entire glycosylated peptide has a mass of 2596 and can be observed in **Figure 4.1**. Other variations of this peptide with Man-6 and Man-8 glycans were also detected by MALDI. The other two glycosylation-site containing peptides and their glycosylated counterparts were not observed.

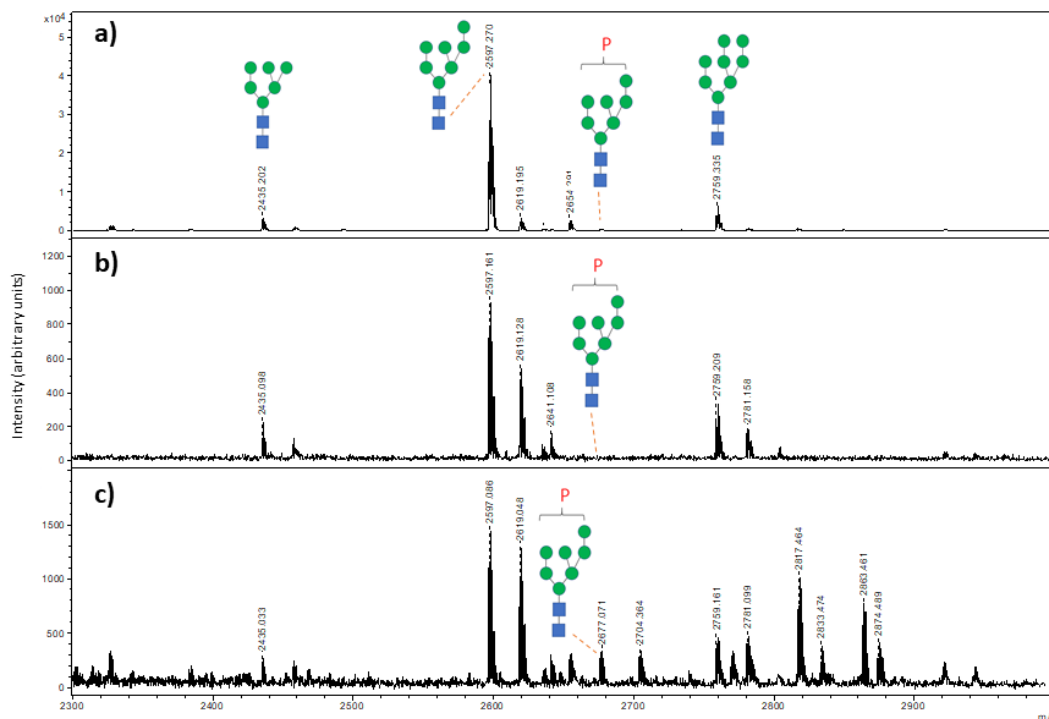


Figure 5.2 MALDI-MS spectra of glycosylated peptide GTFFINKTE from the Glu-C digestion of a) HexA, b) HexM and c) hyperphosphorylated HexM, showing the presence of high mannose species and enhanced phosphorylation. Amended from reference 8.⁸

The phosphorylated version of the GTFFINKTE glycopeptide was also observed in the three samples, but most abundantly in the hyperphosphorylated HexM sample. Qualitatively, the level of phosphorylation appears to be approximately 20%, relative to the unphosphorylated glycopeptide signal. In total, only four glycopeptide variants were detected at the N157 site, as shown in **Table 5.1**.

Table 5.1 MALDI-MS analysis of phosphorylation of high mannose N-linked glycosylation sites of recombinant HexA and recombinant HexM. Amended from reference 8.⁸

Glycosylation site	HexA	HexM
N157	H6N2	H6N2
	H7N2	H7N2
	H7N2P ^a	H7N2P ^a
	H8N2	H8N2

^aThe presence of phosphorylation, denoted by the letter P, was found at very low signal intensities for both protein samples. Hyperphosphorylated HexM showed a higher level of phosphorylation.

The low level of phosphorylation observed was attributed to the lability of the sugar-phosphate bond under the acidic conditions necessary for MALDI. For this reason, experiments were continued using ESI-MS. Samples run using this instrument were digested using pepsin, as only one peptide containing a glycosylation site was observed using Glu-C. However, this same N157 site was the only one detected in ESI-MS, this time in corresponding peptic peptide FINKTEIEDF (MW 2796).

This peptide was observed in phosHexM the form of both $[M+2H]^{2+}$ and $[M+3H]^{3+}$ ions, as shown in **Figure 5.2**.

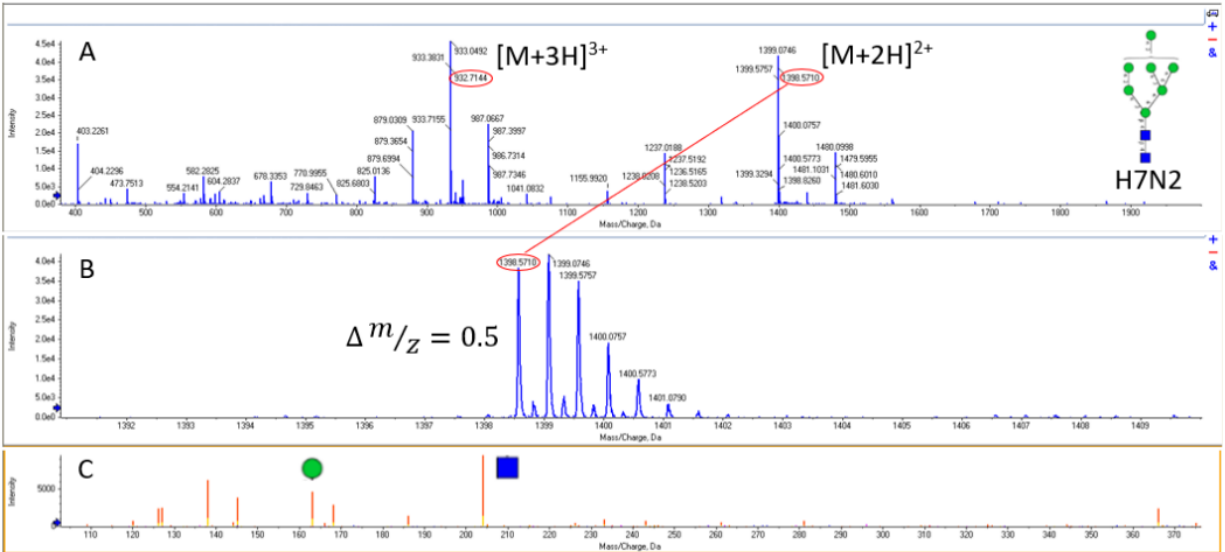


Figure 5.3 ESI MS (A, B) and MS/MS (C) spectra of phosHexM fractions containing high mannose conjugates of FINKTEIEDF.

H7N2 is the glycosylation pattern which was most commonly observed on this peptide, though variations such as H5N2 and H6N2 were identified as well. Panel A shows the +2 and +3 ions and Panel B confirms their charged states. Panel C shows a MS/MS spectrum from this sample which identifies mannose (green circle) and GlcNAc (blue square) fragments.

5.3.2 GM2

As GM2 ganglioside is the substrate which is modified by HexM, analysis of this biomolecule was conducted using MALDI. Due to the samples being extracted from mice brains, they were not purely GM2 ganglioside. Other gangliosides that exist in the samples include GM1, GD1, and GD2. The “M” in these names refers to the presence of a single sialic acid residue, and the “D” refers to the presence of two sialic acid residues. GM1 is the precursor to GM2 in its degradation pathway.

WT, KO, and DKO samples extracted from mice brains were permethylated and analyzed via MALDI-MS. The process of permethylation involves the conversion of OH and NH groups to O-CH₃ and N-CH₃ groups, respectively, and therefore results in a mass increase of 14 per substitution. This is done to optimize stability during the ionization process and facilitate identification. Since the experimental samples included various gangliosides other than GM2, this procedure was performed twice to ensure that GM2 was fully permethylated.

Sodium adducts with masses 1645 and 1849 were observed in each sample (**Figure 5.4**), corresponding to permethylated GM2 and GM1, respectively. The mass difference of 204 represents a permethylated galactose residue, which in this case is the structural difference between GM1 and GM2. Ganglioside structures can be found in **Figure 1.2**.

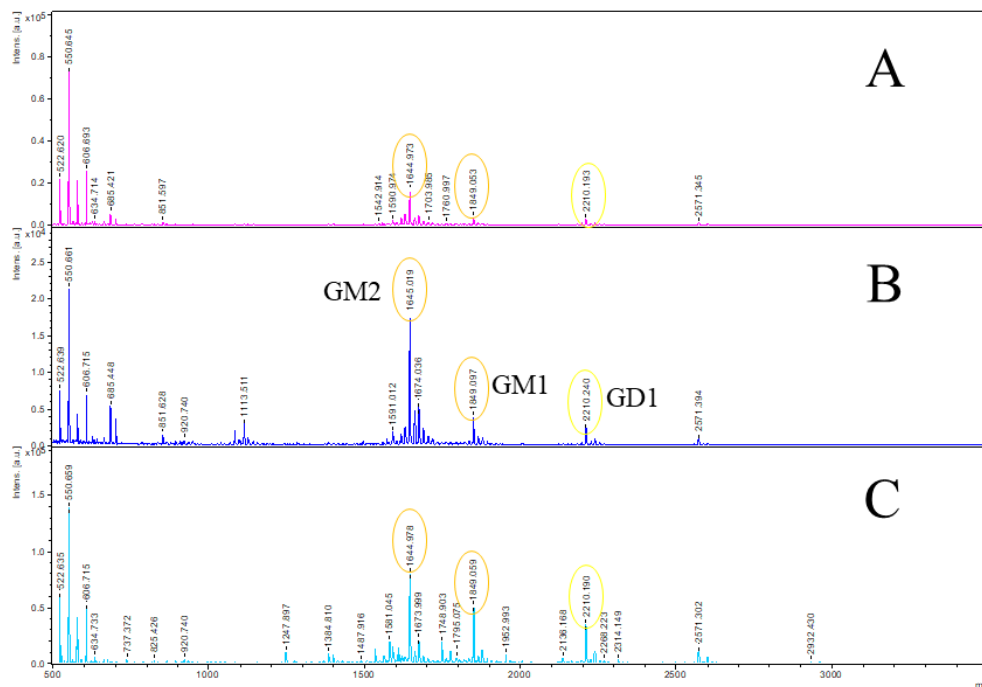


Figure 5.4 MALDI-TOF mass spectra of permethylated ganglioside extracts from brains of wild-type mice.

A third sodium adduct, corresponding to GD1 ganglioside, was also observed in the wild-type samples with a mass of 2210. The mass difference between GD1 and GM1 is 361, which corresponds to a permethylated sialic acid residue. As the performed permethylated procedure involves the use of a sodium salt and MALDI only produces one type of ion at a time, it is not surprising that all observed ions were sodium adducts. Based on relatively peak intensity alone, the level of GM2 in these samples is greater than the level of GM1.

Knockout and double knockout samples were analyzed similarly, with the expectation that they would contain a higher level of GM2 than the wild-type samples due to the removal of the *HEXA* gene. DKO samples also have a gene encoding a murine sialidase knocked out, as this enzyme has been shown to facilitate GM2 degradation in the cell.⁹

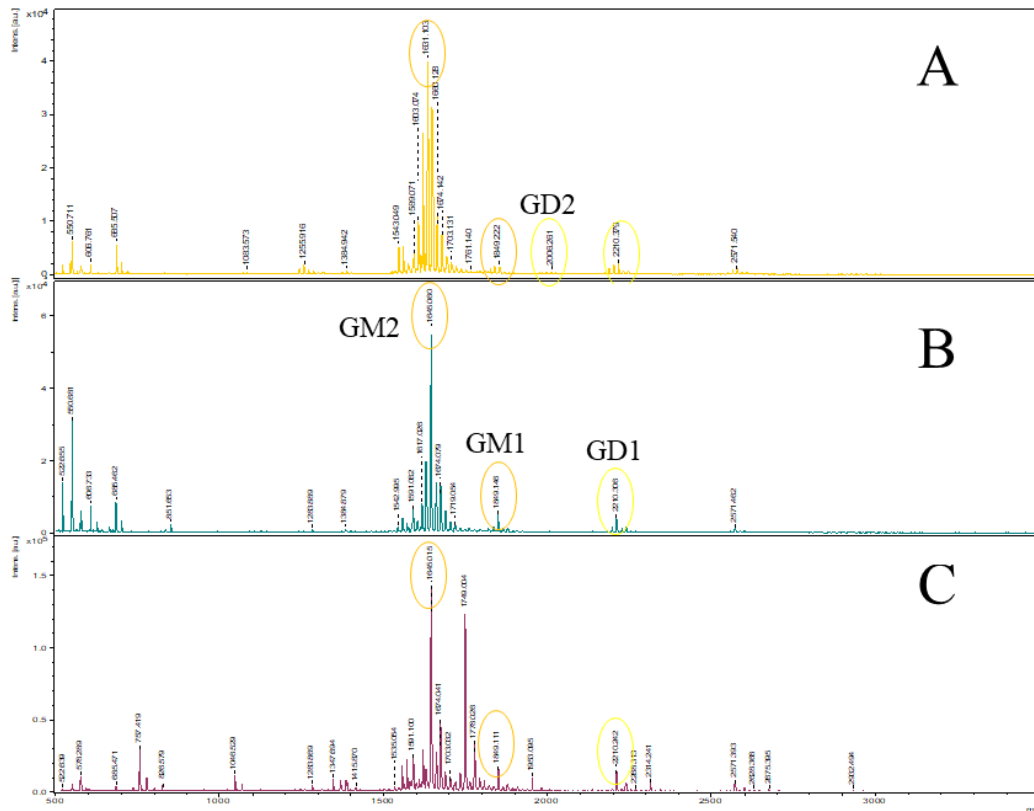


Figure 5.5 MALDI-TOF mass spectra of permethylated ganglioside extracts from brains of knockout (B) and double knockout (A, C) mice.

GM2, GM1, and GD1 were detected in all three samples once again. However, one of the DKO samples also contained GD2 ganglioside. As with GM1 and GD1, the mass difference between GM2 and GD2 represents one permethylated sialic acid residue.

As the acquisition of MALDI spectra relies varies depending on the selected area of the sample in question, the use of an internal standard is required in order to perform quantification. An isotopically labelled version of GM2 ganglioside extracted from the brain of a Tay-Sachs patient was created by performing the permethylation procedure using CD₃I as opposed to CH₃I. This standard has the same behaviour as the experimental GM2, but is distinguishable from it due to the mass increase created by deuterium. d-GM2 has a mass of 1696 whereas regular GM2 has a mass of 1645. A quantity of d-GM2 corresponding to the estimated amount anticipated to observe in the WT samples was added to WT, KO, and DKO samples to be used for quantification purposes, and the resulting spectra can be seen in **Figure 5.6**.

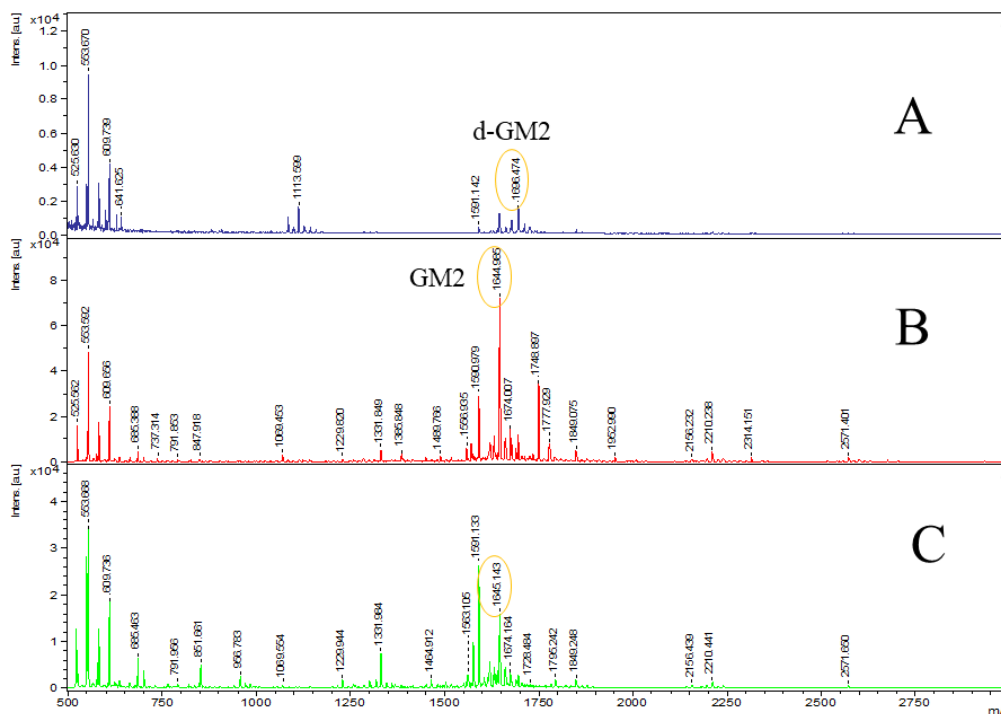


Figure 5.6 MALDI-TOF mass spectra of permethylated ganglioside extracts from brains of WT (A), DKO (B), and KO (C) mice spiked with an isotopically labelled standard.

Based on the peak intensity demonstrated in these spectra, it appears that the DKO sample contains the highest amount of GM2 ganglioside. However, due to the previously mentioned variability of data collection by MALDI, the ratio of the peak area of modified to unmodified GM2 must be calculated in order to produce consistent data. From there, the approximate amount of GM2 in the experimental samples can be calculated as they were spiked with a known quantity of the standard (**Table 5.2**).

Table 5.2 Relative quantitation of GM2 ganglioside in WT, KO, and DKO mouse brains.

GM2				d-GM2		
Sample	Peak area	nmol	µg	Peak area	nmol	µg
WT	445	0.243	0.340	458	0.250	0.350
KO	3905	1.828	2.560	354	0.250	0.350
DKO	23711	2.139	2.995	2771	0.25	0.350

As previous studies have shown the importance of the murine sialidase which is knocked out in the DKO samples,⁹ it was expected that these extracts would contain more GM2 ganglioside than the KO samples. As the WT mice were not modified to have a deficiency in HexA, those brains were expected to contain the smallest amount of GM2. This is exactly what was observed by the relative quantification performed using MALDI-MS. Data analysis was performed using the three samples (WT, KO, DKO) which produced the best spectra following permethylation. It was important to select samples which included the highest quantity of fully permethylated GM2, as the d-GM2 standard did not have any competition for the permethylation reagent and therefore was fully permethylated.

It was expected that 12-15 µg of GM2 would be present per 100 mg of WT brain tissue. Experimental results from two trials (one of which is shown in **Table 5.2**) of MALDI-MS quantification showed 13.3 ± 0.37 µg in the WT sample. The experimental results from the KO and DKO brain extracts yielded 99.3 ± 3.1 and 110.1 ± 9.7 µg of GM2 per 100 mg of brain tissue, respectively. This is consistent with the anticipated data, as mice without the *HEXA* gene should not demonstrate any HexA activity, and therefore will experience an accumulation of GM2 ganglioside in neurons.

5.4 Conclusions

As Tay-Sachs disease results from the inability of HexA to degrade GM2 ganglioside in the lysosome, both HexA and GM2 were a focus of this work. A recombinant, hybrid enzyme employing the most important features of the α - and β -subunits of HexA was created in the hopes of developing a form of enzyme replacement therapy which can alleviate the neurological effects of this disease. Current therapies are unable to do this as the lysosomal enzyme is too large to cross the blood-brain barrier, and thus HexM was designed as a small enzyme comprised of only one subunit.

Both the glycosylation and phosphorylation of HexM is important for trafficking purposes within a cell. As it makes use of the mannose-6 pathway, the presence of mannose-6-phosphate is crucial. A hyperphosphorylated version of HexM was designed in order to ensure this sugar phosphorylation was present. Both MALDI and ESI mass spectrometry techniques were used in order to analyze the modifications of HexM and phosHexM. Despite the presence of three N-glycosylation sites in the sequence of HexM and HexA, only one site was observed in all analysis. This site corresponds to an Asn residue inside of Glu-C peptide GTFFINKTE and peptic peptide FINKTEIEDF. Variations of both glycopeptides were observed, ranging from H5N2 to H9N2, with H7N2 being the most abundant. The phosphorylated version of these glycopeptides was observed as well, predominantly in MALDI-MS analysis of phosHexM. Relative to the unmodified glycopeptide, the level of phosphorylation appeared to be 20%.

Analysis of GM2 ganglioside present in the brains of wild-type, HexA-knockout, and HexA/Neu3-knockout mice was performed. Using an isotopically labelled GM2 standard modified from a Tay-Sachs patient, it was possible to perform relative quantification on these extracts. It was observed that KO brains contained approximately 7x more GM2 than the WT brains, with the DKO brains containing about 9x more than the WT brains. This was consistent with the expected results as a mutated or missing *HEXA* gene results in the inability of a cell to degrade GM2 ganglioside. The additional elimination of the *NEU3* gene enhances this effect.

Future experiments include performing relative quantification of GM2 ganglioside using the brains of mice which have been treated with HexM. Though HexM has been proven to degrade its substrate through the use of enzyme assays, this will determine its efficacy within a system and consequent potential to treat Tay-Sachs disease.

5.5 References

1. Gary-Bobo M, Nirdé P, Jeanjean A, Morère A, Garcia M. Mannose 6-phosphate receptor targeting and its applications in human diseases. *Curr Med Chem.* **2007**; 14(28): 2945-53.
2. Dersh D, Iwamoto Y, Argon Y. Tay-Sachs disease mutations in HEXA target the α chain of hexosaminidase A to endoplasmic reticulum-associated degradation. *Mol Biol Cell.* **2016**; 27(24): 3813–3827.
3. Triggs-Raine B, Mahuran DD, Gravel RR. Naturally occurring mutations in GM2 gangliosidosis: a compendium. *Adv Genet.* **2001**; 44: 199–224.
4. Macauley SL. Combination Therapies for Lysosomal Storage Diseases: A complex answer to a simple problem. *Pediatr Endocrinol Rev.* **2016**; 13: 639-48.
5. Concolino D, Deodato F, Parini R. Enzyme replacement therapy: efficacy and limitations. *Ital J Pediatr.* **2018**; 44(Suppl 2): 120.
6. Tropak MB, Yonekawa S, Karumuthil-Melethil S, Thompson P, Wakarchuk W, Gray SJ, Walia JS, Mark BL, Mahuran D. Construction of a hybrid β -hexosaminidase subunit capable of forming stable homodimers that hydrolyze GM2 ganglioside in vivo. *Mol. Ther. - Methods Clin. Dev.* **2016**; 3: 15057.
7. Gunnarsson A. N- and O- alkylation of glycoconjugates and polysaccharides by solid base in dimethyl sulphoxide/alkyl iodide. *Glycoconjugate J.* **1987**; 4:239-245.
8. Benzie G, Bouma K, Battellino T, Cooper S, Hemming R, et al. Increased phosphorylation of HexM improves lysosomal uptake and potential for managing GM2 gangliosidoses. *BBA Advances.* **2021**; 100032.
9. Seyrantepe V, Demir SA, Timur ZK, Von Gerichten J, Marsching C, Erdemli E, Oztas E, Takahashi K, Yamaguchi K, Ates N, Dönmez Demir B, Dalkara T, Erich K, Hopf C, Sandhoff R, Miyagi T. Murine Sialidase Neu3 facilitates GM2 degradation and bypass in mouse model of Tay-Sachs disease. *Exp Neurol.* **2018.** 299(Pt A): 26-41.
10. Characterization of High Mannose and Phosphorylated High Mannose Glycosylation Sites in Hybrid β -Hexosaminidase (HexM), Battellino T, Tran T, Nguyen DM, Benzie G, Krokhin O, Mark B, Perreault H, 301292, June 2, 2020. Proceedings of the 68th ASMS Conference on Mass Spectrometry and Allied Topics, Online Meeting, June 1-12, **2020**.

6 Chapter 6: General Conclusion

Through the use of various liquid chromatography and mass spectrometry techniques, several developments were made in the fields of proteomics, immunology, and molecular therapy. A focus on post-translational modifications such as phosphorylation and glycosylation facilitated the classification of porcine IgG subtypes, as well as an addition to the SSRCalc algorithm which accounts for the behaviour of phosphopeptides. It also allowed for the ability to determine the structural components of a recombinant lysosomal enzyme. These techniques were also applied to gangliosides, which could be analyzed similarly due to the properties they share with peptides.

The application of specific chromatographic techniques such as strong cation exchange and hydrophilic interaction liquid chromatography was explored, as well as electrospray ionization and matrix-assisted laser desorption/ionization mass spectrometry. These techniques were optimized for the analysis of complex proteomic samples and modified biomolecules through the development of both instrumental methods and data analysis. Different mass analyzers such as hybrid time-of-flight, quadrupole, and Orbitrap analyzers were used.

Different subtypes of porcine IgG had been previously classified due to the variability of the heavy chains of the antibody. Using experimental LC-MS/MS data and entries from the UniprotKB database, glycopeptides were organized into these subtypes and redundant entries were eliminated. Using mass spectrometry, this study emphasized the complexity of immunoglobulins, which is especially important in studies related to xenotransplantation.

Through the development of several large-scale proteomics experiments, two variations of the SSRCalc algorithm were created: one accounting for the use of acetic acid as ion pairing modifier as opposed to formic acid, and another accounting for phosphorylation. Each of these modifications alters the interactions that occur in a chromatographic system both at a peptide and an amino acid level. Sequence-specific effects were identified using large datasets produced from two different laboratories and incorporated into the development of these modified algorithms in order to maximize their accuracy. These studies also demonstrated the use of various chromatographic techniques and conditions and their roles in peptide enrichment and other procedures.

Advances were made in the field of molecular therapy by studying the modifications of hybrid beta-hexosaminidase M. The glycosylation and phosphorylation profiles of this enzyme were determined using both MALDI and ESI-MS, and this study highlighted the importance of these features. As the activity of HexM in the brain is the next step in terms of analysis leading to the development of a new variation of enzyme replacement therapy for Tay-Sachs disease, qualitative and quantitative information regarding GM2 levels was obtained. These experiments will be used as the groundwork for future studies regarding this enzyme.

All of these studies emphasized the reason for which both liquid chromatography and mass spectrometry are so widely used in proteomics. Each study employed a variety of analytical or computational techniques, with a focus on modifications such as phosphorylation and glycosylation, and many of these can be applied to a variety of future experiments.