Glycopeptide Preparation and Application for Alzheimer's Diagnostics and Influenza Immunization

Christopher Gibbons

Follow this and additional works at: https://scholarworks.gsu.edu/chemistry_diss

Recommended Citation
doi: https://doi.org/10.57709/26666963

This Dissertation is brought to you for free and open access by the Department of Chemistry at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Chemistry Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.
Glycopeptide Preparation and Application for Alzheimer’s Diagnostics and Influenza Immunization

by

Christopher Gibbons

Under the Direction of Jun Yin, PhD

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in the College of Arts and Sciences Georgia State University 2021
ABSTRACT

Glycoproteins and glycopeptides are ubiquitous in nature but present challenges in clinical application. From a synthetic standpoint, they combine the difficulties of glycochemistry, namely regio- and stereo-specificity limitations, with the challenges inherent in peptide production. Yet their biological significance makes them impossible to ignore.

The pathology of Alzheimer’s Disease (AD) is characterized by the extracellular deposition of amyloid β peptides (Aβ) and intracellular accumulation of hyperphosphorylated tau protein (p-tau). The hyperphosphorylation of tau is known to occur in concert with the reduction of O-GlcNAc loading. Indeed, it is known that artificially forcing an increase in O-GlcNAc results in reduced phosphorylation, and vice versa, at several sites known to be highly phosphorylated in p-tau of AD patients. While the study of p-tau is advanced with many commercial antibodies available to researchers, glycosylated tau is less well studied. Here, an array of peptides representing a relevant subset of tau is prepared. The peptides are either unmodified or bear O-GlcNAc or phosphate modifications in all possible combinations. P-tau related antibodies currently used for the diagnosis of AD are then analyzed against the microarray to reveal their binding specificities with tau peptides in different forms of modifications. The microarray-based screen enables us to establish the true epitope of clinically relevant anti-tau antibodies to guide their use in the diagnosis of AD.

In another project, a CRM197-glycopeptide conjugate was prepared for the development of an anti-influenza vaccine. The glycopeptide borne by this conjugate contained three copies of a highly conserved region of the Influenza A hemagglutinin viral glycoprotein. This glycopeptide was linked to carrier protein CRM197 via a linker at a loading level of approximately 5-6
glycopeptides per protein. This conjugate was then evaluated as a potential vaccine candidate against a strain of Influenza A.

INDEX WORDS: Glycopeptide, Tau Array, P-tau, Alzheimer’s Disease Biomarkers, Influenza, Bioconjugation
Glycopeptide Preparation and Application for Alzheimer’s Diagnostics and Influenza Immunization

by

Christopher Gibbons

Committee Chair: Jun Yin

Committee: Ming Luo
Lei Li

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
December 2021
DEDICATION

This work is dedicated to my wife, the light of my life. And also to my mother, father, and sister, all three of whom are responsible for making me the person I am today.
ACKNOWLEDGEMENTS

I would like extend my heartfelt gratitude to Dr. Jun Yin for his thoughtful guidance over the last two years. And to my committee as well, Dr. Ming Luo and Dr. Lei Li, your patience, support, and feedback over the years are deeply appreciated.

I would like to acknowledge the PI who led me through much of my time at Georgia State, Dr. Peng George Wang. Without him this dissertation would not have been possible.

Special thanks is due to Garrett Edmunds, Harmon Greenway, Jack Xiao, Ross Terrell, Madhusudhan Gadi, Shuaishuai Wang, Hailiang Zhu, Liuqing Wen, and Jiabin Zhang as well as all the other colleagues I’ve had over the years. Your input and friendship made my time in the program more productive and certainly much more sustainable.
TABLE OF CONTENTS

ACKNOWLEDGEMENTS .................................................................................................................. V

LIST OF FIGURES .......................................................................................................................... X

LIST OF ABBREVIATIONS ............................................................................................................... XII

1 INTRODUCTION ............................................................................................................................. 1

1.1 Glycoconjugates of proteins and peptides .............................................................................. 1

1.1.1 Nature of glycoconjugates ...................................................................................................... 1

1.1.2 Glycoproteins in diagnosis and drug discovery ...................................................................... 4

1.1.3 An immune role for glycopeptides in diagnosis and drug discovery ....................................... 6

1.2 Preparation of glycopeptides ..................................................................................................... 7

1.2.1 Extraction from natural resources .......................................................................................... 7

1.2.2 Preparation by chemical synthesis of peptides ....................................................................... 7

1.2.3 Chemoenzymatic synthesis .................................................................................................... 11

1.2.4 Challenges in the synthesis of glycopeptides ......................................................................... 12

1.3 Approach of this work ................................................................................................................ 13

1.3.1 Automatic parallel synthesis of glycopeptides of large quantities ......................................... 13

1.3.2 Glycopeptide microarrays ..................................................................................................... 14

1.3.3 Bioconjugate vaccines .......................................................................................................... 16

2 PREPARATION OF A TAU PEPTIDE LIBRARY ...................................................................... 19

2.1 Introduction ............................................................................................................................... 19
2.1.1 A library to support diagnostics and antibody screening .................. 19

2.1.2 Design considerations ........................................................................ 20

2.2 Plan of synthesis .................................................................................... 21

2.2.1 Sequence choice considerations .......................................................... 23

2.3 Experimental ......................................................................................... 24

2.3.1 Building block synthesis .................................................................... 24

2.3.2 Peptide synthesis ................................................................................ 28

2.3.3 First attempt ....................................................................................... 29

2.3.4 Final sequences .................................................................................. 31

2.3.5 Purification and finalization ................................................................. 33

2.4 Discussion .............................................................................................. 33

2.5 Supporting Spectra ................................................................................ 35

3 A TAU ARRAY FOR ALZHEIMER’S DIAGNOSTICS .................................. 48

3.1 Introduction ............................................................................................ 48

3.1.1 P-tau .................................................................................................... 49

3.1.2 pT181, pS202, and pT205 Antibodies .................................................... 50

3.2 Methods .................................................................................................. 50

3.3 Results .................................................................................................... 52

3.3.1 Preliminary observations ..................................................................... 55

3.3.2 AT270 binding site ............................................................................. 56
### 3.4 Discussion

3.4.1 Binding constants ................................. 58
3.4.2 Future work .......................................... 59
3.4.3 Conclusions .......................................... 59

### 4 A PATH TO A UNIVERSAL INFLUENZA VACCINE

4.1 Introduction ............................................ 60
4.1.1 The influenza viruses .................................. 60
4.1.2 Public health impact ................................... 61
4.1.3 Current vaccines ..................................... 64
4.1.4 Stalk-targeting antibodies ............................ 65
4.1.5 A stalk-targeting vaccine ............................. 65
4.1.6 A conserved epitope ................................... 66

4.2 Experimental ............................................. 67
4.2.1 The target .............................................. 67
4.2.2 Glycopeptide preparation ............................ 69
4.2.3 Bioconjugation ....................................... 70
4.2.4 Sample Preparation .................................. 71
4.2.5 Initial mouse study structure ........................ 72

4.3 Discussion ................................................ 74

REFERENCES ................................................................ 77
LIST OF FIGURES

Figure 1. The monosaccharides referenced herein .............................................................. 2
Figure 2. Forms of N- and O-linked glycans .................................................................. 3
Figure 3. A) Overall Scheme of SPPS, B) The resin used for all SPPS herein ......................... 9
Figure 4. 1) A schmidt donor, 2) a thioglycoside, and 3) a glycosyl halide ......................... 13
Figure 5. Overall scheme of arrays for antibody binding analysis ....................................... 15
Figure 6. Conjugation to a carrier boosts response ............................................................. 18
Figure 7. Overall plan of array preparation ......................................................................... 22
Figure 8. Synthesis scheme of O-GlcNAc Ser/Thr building blocks .................................... 24
Figure 9. The initial peptides chosen ................................................................................. 29
Figure 10. Example from the first set of peptides ................................................................. 30
Figure 11. The final sequences used .................................................................................... 31
Figure 12. Example of shorter peptides .............................................................................. 32
Figure 13. GlcNTroc (3x Ac) ............................................................................................. 35
Figure 14. GlcNTroc Trichloroacetimidate Donor ............................................................... 36
Figure 15. Fmoc-Thr(3xAc GlcNTroc)-OBn ........................................................................ 37
Figure 16. Fmoc-Ser(3xAc GlcNAc)-OBn .......................................................................... 38
Figure 17. Fmoc-Thr(3xAc GlcNAc)-OBn .......................................................................... 39
Figure 18. NMR of Fmoc-Ser(3xAc GlcNAc)-OH .............................................................. 40
Figure 19. NMR of Fmoc-Thr(3xAc GlcNAc)-OH .............................................................. 41
Figure 20. NMR of Fmoc-Ser(3xAc GlcNAc)-OH .............................................................. 42
Figure 21. NMR of Fmoc-Thr(3xAc GlcNAc)-OH .............................................................. 43
Figure 22. NMR of Fmoc-Ser(3xAc GlcNAc)-OH .............................................................. 44
LIST OF ABBREVIATIONS

Ac$_2$O – Acetic Anhydride
AcOH – Acetic Acid
AD – Alzheimer’s Disease
Asn – Asparagine
Aβ – Amyloid Beta
BAANS – Bromoacetic Acid N-Hydroxysuccinimide Ester
BnNH$_2$ – Phenylamine
CCl$_3$CN – Trichloroacetonitrile
DBU – 1,8-Diazabicyclo[5.4.0]undec-7-ene
DCM – Dichloromethane
DIPEA – Diisopropylethylamine
DMF – N,N-Dimethylformamide
DNA – Deoxyribonucleic Acid
DODT – 2,2’-(Ethylenedioxy)diethanethiol
ELISA – Enzyme-Linked Immunosorbent Assay
FDA – Food and Drug Administration
Fmoc – Fluorenlymethyloxycarbonyl
Fmoc-OSu – N-(9-Fluorenylmethoxycarbonyloxy)succinimide
GlcNAc – N-Acetylglucosamine
GP – Glycopeptide
HA – Hemagglutinin
HATU – Hexafluorophosphate Azabenzotriazole Tetramethyl Uronium
HBTU – Hexafluorophosphate Benzotriazole Tetramethyl Uronium
HPLC – High Performance Liquid Chromatography
IAV – Influenza Type A
IBV – Influenza Type B
ICV – Influenza Type C
IDV – Influenza Type D
MALDI – Matrix-assisted laser desorption/ionization
mRNA – Messenger Ribonucleic Acid
NA – Neuraminidase
NaHCO₃ – Sodium Bicarbonate
NMP – N-Methylpyrrolidone
OGA – Protein O-GlcNAcase
P-tau – Hyperphosphorylated Tau
PBS – Phosphate-Buffered Saline
Pd/C – Palladium on charcoal
PTM – Post-Translational Modification
Py – Pyridine
RNA – Ribonucleic Acid
Ser – Serine
SPPS – Solid-phase peptide synthesis
TFA – Trifluoroacetic Acid
THF – Tetrahydrofuran
Thr – Threonine
TIS – Triisopropylsilane

TMB – 3,3', 5,5'-Tetramethylbenzidine

TMSOTf – Trimethylsilyl trifluoromethanesulfonate

Troc-Cl – 2,2,2-Trichloroethoxycarbonyl Chloride

UHPLC – Ultra High Performance Liquid Chromatography

Zn – Zinc


1 INTRODUCTION

Glycopeptides and glycoproteins together represent a core class of biomolecules with essential roles in diverse biological processes from energy storage to structural support and from cell signaling to cell-cell recognition.\(^1\) The complexity of cell surface glycans makes them common epitopes for immune response.\(^2,3\) At the same time, viruses take advantage of the glycan shield to subvert immune recognition.\(^4-6\) A patient’s antibodies, often against glycopeptide sequences, have considerable clinical significance. Correspondingly, the exploitation of glyco-specific antibodies for diagnostic purposes demands the elucidation of their glycopeptides epitope, that would require the development of efficient synthetic approaches to generate glycopeptides of diverse and defined structures. In addition to diagnostics, the glycopeptides have an emerging role in therapeutics as well, although in many ways, glycopeptides seem to be a poor choice of compound for therapeutic application.\(^7-9\) Nevertheless, there are significant peptide, glycan, and glycopeptide-based therapeutics with FDA approval.

1.1 Glycoconjugates of proteins and peptides

1.1.1 Nature of glycoconjugates

Protein and peptide linked glycans serve many roles. They can be important in inducing proper folding and preventing degradation of the parent protein.\(^10\) In some cases glycosylation of proteins has a regulatory/signaling role not too dissimilar to phosphorylation. There are great many specialty roles for glycan modification of peptides and proteins. For example, glycosylation of mucin gives the protein its hydrophilicity. Glycosylation (and subsequent extension of the resulting starch) of Glycogen stores glycose in starch chains that are O-linked to the protein. In viruses, oligosaccharides can serve to suppress immune response to viral epitopes that could not be so easy
mutated. At the same time, changes in oligosaccharide structure on surface of the cells that turn cancerous would flag them for clearance by the immune system.\textsuperscript{11}

Much like peptides, DNA, and RNA, glycans are built of a modest set of simple building blocks. Unlike peptides, however, glycans only derive indirectly from the genome.\textsuperscript{12} There is nothing analogous to translation, no process whereby glycan sequences are encoded explicitly in a linear sequence.\textsuperscript{13} Instead, they are built up by the actions of enzymes on simpler substrates as they are available. It is a much more statistical process. The availability of substrate and donor as well as the frequency of various enzymes and the capabilities and behavior of those enzymes in the system determines the range and frequency of glycan structures made.\textsuperscript{14}

\begin{figure}[h]
\centering
\includegraphics{glycans.png}
\caption{The monosaccharides referenced herein}
\end{figure}

All monosaccharides all have several possible points for conjugation through which they can link to another saccharide. In a linkage between one saccharide and another, the anomeric (first position) of one monosaccharide is usually linked to any other (non-anomeric) hydroxyl on the other. While chirality at the non-anomeric positions is fixed and a defining characteristic of that monosaccharide, the anomeric carbon has significant aldehyde character, and its stereochemistry
can change over the course of a reaction mechanism. As such, first position chirality is an attribute of the linkage, and not of either monosaccharide. Linkages are expressed as, “chirality, linkage position of the one saccharide – linkage position of the other saccharide.” An example would be α1-3 or β1-6.

Figure 2. Forms of N- and O-linked glycans
Glycoproteins take many different forms. The main two families are N- linked and O linked. N linked glycans are so called because the link to the protein through an amide linkage to an asparagine. These glycans begin with two GlcNAc, followed by a branching mannose structure. After this point they grow more complex, taking high mannose or complex forms, biattenary or triattenary forms. The core can also take a sialic acid. But the core GlcNAc and mannose sugars are generally seen. O-linked forms are less easy to generalize. They are linked through an ether linkage to serine, threonine, or tyrosine. The cores listed here are just an example. There are many others. O-linked monosaccharides are also important, O-GlcNAc or O-Fucose for example, serving signaling and regulatory roles.

A glycan can link to a protein or peptide in several possible ways. The two major classes are N- and O-types of linkages.\textsuperscript{15,16} N-linked glycans are joined to an asparagine through an amide bond at the anomeric carbon. O-linked glycans are joined to either a threonine or serine (and
occasionally tyrosine) through an ester bond, also at the anomeric carbon. There are a number of additional exotic linkages that have been observed. These are outside the scope of this section.

The forms of glycans in nature is complex. The number of short (6 or fewer monosaccharide units) glycans that are possible to construct by following simple assembly rules exceeds a billion. Herein, the concern lies chiefly with O-GlcNAc and truncated N-linked forms.

1.1.2 Glycoproteins in diagnosis and drug discovery

A major interest with regard to glycoproteins in diagnosis and drug discovery is as their role as immune targets. There are two proteins with significance in regards to projects discussed herein. The first is the tau family of proteins, the tangles of which are associated with neuron disfunction in Alzheimer’s disease.17

The tau proteins are a set of microtubule-associated proteins which bind to tubulin, helping to form and stabilize microtubules in axions. They are intrinsically disordered. There are six isoforms in humans, the most abundant of which is 2N4R, so-called because it bears two N-terminus inserts as well as four repeats of the 31-32 residue microtubule binding region. Human brain tau ranges from 352 residues (0N3R) to 441 residues (2N4R). Big tau, with 752 residues, appears to be mostly associated with the peripheral nervous system and is not well studied.18,19 Curiously, the additional region responsible for the dramatic size difference appears to only be very lightly phosphorylated. All the forms of tau are encoded by a single gene that would generate alternatively spliced mRNAs.

A tauopathy is any condition associated with the formation of neurofibrillary tangles (NFT) of tau. Alzheimer’s disease, chronic traumatic encephalopathy, and primary age-related tauopathy are examples of tauopathies but there are many others.20,21 In these conditions tau becomes phosphorylated above and beyond where it would be in healthy neurons. The resulting p-tau favors
aggregation with itself over binding to the tubulin protein making up the microtubules. Without tau, the tubules, which are needed for structural and transport purposes, are not stable. The result is a diseased state and a reduction in neuron function. The reason for this hyperphosphorylation is unknown. Even if the process underlying tau hyperphosphorylation and the formation of tangles is not the fundamental underlying cause behind AD, preventing the formation of the tangles could well prove beneficial from a therapeutic standpoint.

In addition to being phosphorylated, healthy tau bears O-GlcNAc. These sites are simple monosaccharides not extended any farther. The number of known O-GlcNAc sites is much smaller than the number of phosphorylation sites. That said, all known sites are also phosphorylated in AD patients. It is known that treating neuron cell cultures with Thiamet G, an inhibitor of the protein O-GlcNAcase (OGA) enzyme which cleaves O-GlcNAc, results in reduced tau phosphorylation. As such, it is speculated that a therapy which preserves or forces elevated O-GlcNAc will improve patient outcomes. Such a treatment could conceivably provide gains in other tauopathies as well.

In any case, the shift from O-GlcNAc to phosphorylation represents a considerable set of potential disease biomarkers. On one hand, it is self-evident that the number of phosphorylation sites that crop up in AD presents opportunity. But furthermore, the combination of declining response to O-GlcNAc could represent another dimension, shoring up the significance of signal in a blood test.

The second protein of interest is influenza hemagglutinin. Hemagglutinin is one of the two major viral surface glycoproteins of influenza, along with neuraminidase. Influenza A strains are named based on these two glycoproteins. For example, the virus responsible for the swine flu pandemic of 2009, is H1N1, is so-called because it contains subtype 1 (out of 18 known) of hemagglutinin and subtype one of (out of 11 known) of neuraminidase.
Hemagglutinin is a homotrimer (a protein composed of three identical units). Its function is to bind to sialic acid in surface glycans on target cells and facilitate the internalization of the virus. During the endocytosis process, the pH decreases which triggers a significant change in conformation in hemagglutinin ultimately leading to viral entry.\textsuperscript{28,29} The exposed portion of hemagglutinin is subdivided into a head region and a stem region. The head region is a primary target of antibodies.\textsuperscript{30} Neutralizing antibodies for influenza often bind the region around the sialic acid binding pocket.\textsuperscript{31} The virus is able to survive considerable mutation in the head region without loss of viability. This region varies greatly between strains and is also heavily glycosylated. It appears that oligosaccharides obscure sites that may be conserved, making the generation of antibodies that are able recognize these sites more difficult. This strategy comes at the cost of a greater innate immune recognition.\textsuperscript{32} The stem region is better conserved across strains than the head region. As such it is an appealing target for broad-spectrum therapeutics.

1.1.3 An immune role for glycopeptides in diagnosis and drug discovery

Glycoproteins are prime immune targets. However synthetic proteins, to say nothing of the additional considerations that would go into synthetic glycoproteins, are exceedingly difficult to make.\textsuperscript{33} Fortunately, truncated forms, glycopeptides, tend work well enough in most applications. Indeed, sometimes subdivision for the sake of differentiation of epitopes is exactly what is necessary for a given application. These smaller analogues can be readily producible by solid phase peptide synthesis (SPPS). And SPPS allows for automation and by extension, massive parallelization. The difficult stereo control aspects of glycochemistry are isolated to the building block preparation phase where compounds are small and stereo isomers are more separable.

Herein there are broadly two applications, both built around immune recognition of glycopeptides. In the first, the target protein is split up so that the binding specificity of existing
antibodies, both commercial and autoimmune, can be ascertained. In the final chapter of this work the opposite approach is taken. Instead, new antibodies are generated against a specific chosen glycopeptide.

1.2 Preparation of glycopeptides

1.2.1 Extraction from natural resources

If glycans linked to proteins are ubiquitous in nature, then extraction from natural sources is an obvious avenue for their acquisition. This is done in some cases, particularly for glycans that would be prohibitively complex to synthesize. However, there are difficulties associated with this approach not found with others. The isolation of any one material from a natural soup of materials necessitates great consideration in the design of the chromatographic system and the preparation of samples. The samples need to be cut down by proteases or cleaved all the way to a free glycan. There is then a need to build the compound back up if it is not of the final form needed. Ideally, for solid phase synthesis, many equivalents would be available and large amino acids typically have worse yields than smaller ones. But this issue is not unique to naturally sourced large peptides.

If a bulk source of glycoprotein can be found (egg yolk is a common one) and a reliable chromatographic method for isolating the desirable product then this strategy can be used to great effect. In any case, in this work only monosaccharides linked to a peptide chain are needed so a natural route would likely be more difficult than a synthetic one.

1.2.2 Preparation by chemical synthesis of peptides

The first synthetic peptides were produced by conventional solution-phase synthesis. This method has advantages. Namely, it is straightforward to scale preparation to arbitrary scales and there are peptides being made today by this method for this reason. Solution-phase synthesis of peptides is not fundamentally different from any other class of compound. The practical difference
is in the number of reaction steps required. The value in simplifying some aspect of purification is self-evident. As such most chemical synthesis of peptides, particularly when they need to be made quickly in a lab, is by solid-phase peptide synthesis (SPPS). SPPS was developed by R. B. Merrifield and was first published in the Journal of the American Chemical Society in 1963. In his method the first amino acid was immobilized on a solid polymer bead. The purpose of the bead is to facilitate recovery of product from leftover starting materials and reaction byproducts. The genius in this is that an arbitrary molar excess of solution-phase reagents may be used. It is routine to use 5-fold and sometimes even 20-fold excesses of solution-phase reagents. Isolating product from this magnitude of starting material would be non-trivial if the products were also in solution.

The most straightforward approach to preparing synthetic glycopeptides is to first glycosylate individual amino acids and then use these compounds as building blocks in SPPS. Glycosylating first allows greater ability to control stereo-purity. Glycosylated amino acids are far smaller than glycosylated peptides. As such separation of products and isolation of one stereoisomer from another is more practical. As an additional advantage the preparation of glycosylated amino acids makes use of cheaper starting materials (amino acids vs peptides). The effect of this is to afford greater freedom to choose glycosylation methodologies. A poor yield in a reaction between a cheap monosaccharide and a canonical amino acid is of no concern if the desired product can be easily separated from the rest of the material left behind after workup. An amino acid containing a protected monosaccharide is roughly the size of a large conventional amino acid. Compare Fmoc-Asn(GlcNAc 3xAc)-OH at 684 g/mol to Fmoc-Arg(Pbf)-OH at 649 g/mol for example. As one would expect, the overall process for adding a monosaccharide-containing amino acid is much the same as for a regular amino acid, both in process and in yield.
Disaccharide containing amino acids are a slightly different story. Fmoc-Asn(chitobiose 5xAc)-OH masses in at 971 g/mol. These still couple well enough, though solubility is poorer and yields can be reduced.

**Figure 3. A) Overall Scheme of SPPS, B) The resin used for all SPPS herein**
The resin is functionalized with an acid cleavable linker allowing the liberation of final product. The resin remains solid as it undergoes a series of reaction steps after being placed in the reactor vessel. First the Fmoc is removed in the deprotection step. Then amino acid along with amide bond forming reagents are added. Then capping by treatment with acetic anhydride may take place. The cycle can then repeat. Once all amino acids are installed the resin is removed and treated with a TFA based cocktail to cut the peptide off the resin and strip off the protecting groups.

In order to use glycosylated amino acids like any other in SPPS, it is necessary to make them at scale. Ideally, ten equivalents would be available. It is possible to get away with two if it is difficult to procure more. The scale of SPPS varies from kilograms to only a few micrograms. For most general purposes where peptides are prepared using conventional synthesizer such as a Liberty Blue, a scale on the order of dozens of milligrams is chosen. Introducing a single amino acid bearing a monosaccharide at typical scales using standard methods requires several hundred milligrams of amino acid. As such, a procedure for preparing the amino acid at gram scales in acceptable yields is desirable.

When the size of the glycan is limited and synthesis of the glycosylated amino acid is straightforward at scale, then this method for preparing glycopeptides is preferred. If the glycan is large, then attempting to install the glycan in this way is challenging. The synthetic efficiency
drops off with size. Furthermore, the larger and more complex the glycosylated amino acid is, the more precious it will be. Solid-phase synthesis chiefly shines where large excesses of reagents can be used.

The first and most obvious problem that can crop up in SPPS-based synthesis is that of deletions.\textsuperscript{41} If Fmoc deprotection is incomplete or the coupling reaction does not go to 100% completion before the resin is washed and the next cycle begins then any remaining amine will be left free to couple to the succeeding amino acid, producing a peptide with a deletion. Depending on the sequence a deletion may be difficult to remove from the desired product. For example, truncated peptides losing a charged amino acid such lysine will be fairly easy to remove because of the differing charge. There will be a significant difference in HPLC retention times in this case. Peptides losing glycine will give a much smaller change in retention time. This issue can be mitigated with capping to some extent. Exposing the resin to acetylation conditions after coupling will install a largely irremovable acetyl group at the exposed amine, leaving only desired product and capped byproduct. The capped byproduct will not grow with the rest of the chains and will likely be easier to remove from the crude product than a deletion peptide. Of course, the deletions due to incomplete Fmoc deprotection will not be capped and will be carried forward.

Another danger is that the base introduced in the coupling step may lead to a loss of Fmoc. Losing Fmoc after coupling but before the end of the coupling cycle will more often than not lead to two copies of the amino acid being installed instead of one. Generally, due to the bulkiness of the base used in this step, this does not occur often. It is possible however, when very long coupling cycles are used, particularly with excesses of base and in sequences with significant steric hindrance.\textsuperscript{42} Other common issues include racemization, aspartamide formation, fragmentation, oxidation and many others. A comprehensive discussion of side reactions is outside the scope of
this work. Following SPPS, peptides must generally be purified. This is generally by HPLC. Although peptide coupling reactions are exceedingly efficient, the extraordinary number of reactions stacked in series without opportunity to purify intermediates often results in a noticeably mixed product. To illustrate, take the example of a 20 mer where each cycle consistently reaches 99% overall yield. In this case the total yield would be $0.99^{20}$, or 82%. At a per-cycle yield of 95%, still very good compared with conventional synthetic reactions one might perform at the bench, the total yield drops to 36%. Since the side products are peptides, typically not too dissimilar from the desired peptide, HPLC or UHPLC is difficult to avoid.

1.2.3 Chemoenzymatic synthesis

If the glycan is large, the most practical option is often to build up the glycan on a simpler synthetic glycopeptide after its synthesis and purification through the use of enzymes.43-45 This option is immediately attractive from a synthetic standpoint. One can use cheap reagents in the inefficient SPPS steps and then let the enzyme bypass the usual synthetic difficulties such as stereochemistry. The major limitations of the approach are the availability of enzymes to perform the desired transformations. There are two specificity issues to be considered in enzymatic installation of a glycan: substrate specificity and donor specificity. In other words, the enzyme must recognize the section of peptide to be glycosylated (to the exclusion of all other portions) and it must be able to handle the glycan to be installed. These two variables limit the usefulness of the approach, but not as much as one might expect at a glance. As a general rule, interesting glycopeptides tend to be natural sequences and natural sequences are generally generated enzymatically in nature. Thus, we can usually expect there to be some enzyme to perform our glycosylation somewhere in nature. This does not mean that this enzyme is easy to isolate, express, and utilize in a practical fashion.
1.2.4 Challenges in the synthesis of glycopeptides

The formation of a glycosidic bond between two saccharides generally involves the first, anomeric position of one, and a free hydroxyl on another. The “donor” in this case will be a saccharide activated at the anomeric position. A monosaccharide usually has three or four hydroxyls suitable for conjugation with a properly configured donor. The first challenge in the formation of such a bond is control over which hydroxyl participates. Leaving them all or several free would result in a mix of compounds. While there are examples of this strategy being used it is better, in all but the most special circumstances, to leave exactly one hydroxy free while protecting all the others through protecting group manipulations. Indeed, these manipulations must be exceedingly clever, as all the groups to be protected are hydroxyls and, at first blush, should respond in similar ways to treatment. In the case of a monosaccharide joined to an amino acid, little concern must be paid to regioselectivity, as all functional groups are either sufficiently different, or available commercially with reasonable protection, to where a path to a single site for conjugation is intuitive.

A bigger factor in the formation of a linkage between a monosaccharide and an amino acid, one that also applies just as well to linkages between two monosaccharides, is the stereochemistry of the linkage. The chirality of a hydroxyl with respect to its parent monosaccharide generally won’t change for most conjugation strategies. With conventional methods a glucose joining at the 4th position does not risk becoming galactose as a result of the conjugation. However, the anomeric position can change, resulting in a bond that is either alpha or beta as a result of the linkage. Getting this correct is the stereospecificity problem and much of the set-up for a conjugation is built around the need to get this correct.
The choice of donor is a major factor here. The most popular donors at the moment are likely glycosyl imidates (Schmidt donors), thioglycosides, and glycosyl halides. All are straightforward to install and have different attributes.

![Figure 4. 1) A schmidt donor, 2) a thioglycoside, and 3) a glycosyl halide](image)

*These two donors are the most popular in use today. The thiol is very often a thiophenol, sometimes with further substituents, or ethanethiol. There are many other thiols in use. They are all malodorous and leave thiols as leaving groups which can destroy hydrogenation catalysts.*

The second position of the donor is a major factor. If the position is able to engage in neighboring group participation, then beta stereochemistry is favored. In the case of NAc glycans there is a reduction in the efficacy of donors due to the formation of a stable oxazoline intermediary. Depending on the details of the reaction this position might need a protecting group during the conjugation step.

### 1.3 Approach of this work

#### 1.3.1 Automatic parallel synthesis of glycopeptides of large quantities

In SPPS the resin remains stationary in the system. Treatment by successive coupling and deprotection steps builds a peptide on the stationary resin. The closest thing to a purification step is filtering and washing. There is no stopping to characterize the peptide before completion usually, although there are special cases where this is done. The necessary steps are simple and require no conscious choice after synthesis begins. This facilitated the development of automated synthesis of peptides early on, and manual SPPS is usually only performed by labs without access to a machine. Multichannel machines, synthesizers that can prepare more than one peptide at a time are still fairly rare but greatly ease the preparation of arrays. These machines use much the same
hardware as some types of single channel machines but support many reaction vessels being used in parallel. A single channel machine will spend the vast majority of its time during a synthesis waiting for reactions to complete. The time needed to dispense reagents is modest. Designing the equipment so that it can be used to set up multiple reactions in parallel is very sensible. The limitation is in the design of the reactor vessels. In a single channel machine like a Liberty Blue it is straightforward to design around a microwave reactor. Introducing racks or trays of reaction vessels makes this much less easy. In this work both approaches are used. For the library produced in Chapter 2 a multichannel machine, a CEM Miltipep2 synthesizer was used. This machine allows up to several hundred peptides to be prepared in parallel. For the glycopeptide representing a region of the influenza A hemagglutinin a single channel machine with a microwave synthesizer was used.

1.3.2 Glycopeptide microarrays

In the process of peptide synthesis, there are many opportunities for iteration to generate a peptide library of alternative sequences or building blocks. One may have a need to mutate individual amino acids, such as with an alanine scan. Identifying an antibody binding site can require a huge set of peptides. To be rigorous, it would be necessary to screen the peptide epitope of an antibody with a set of peptides with overlapping sequences derived from the antigenic protein. If the sequences chosen are not overlapping there is a risk that the binding site will be cut between two peptides. An average-size protein could require well in excess of a hundred sequences. This is before we consider post translational modifications (PTMs). If we wished to examine the case where a single site on a single sequence could bear O-GlcNAclyation, phosphorylation, or left as a free residue there are but three peptides to study. If, however, one wished to examine a single sequence bearing five such sites, the number of possible valid peptides
in the set where each of these positions is varied freely blossoms to $3^5 = 243$. Numbers on this scale call for both clever choice of targets and methods allowing for massively parallel screening.

Microarrays are a popular form of high-throughput screening where compounds to be screened are linked covalently to a slide in separate, almost microscopic, spots and surveyed typically through optical means. In a microarray, hundreds or even thousands of compounds can be screened and directly compared in parallel in a single assay with mere micrograms of compound being needed for each assay.

![Figure 5. Overall scheme of arrays for antibody binding analysis](image)

The peptides are printed onto a glass slide functionalized with an NHS activated ester linker. The result is a covalent link to the surface. The antibody containing material, either serum or a sample of commercial antibody is added to the surface and allowed time, an hour typically, to bind. The plate is then washed to remove excess. Another antibody is introduced, this time an anti-IgG antibody with a fluorescence tag. The plate is washed and images made under excitation wavelength illumination.

There are methods for synthesizing peptides directly on a microarray. These methods have the advantage that they readily scale to thousands of peptides. The major downside here is the absence of opportunities for quality control. Peptide synthesis can be more problematic than one
might expect. When iterating across a large number of sequences, it is not difficult to find more than one that does not come out well on the first attempt. Often times difficult sequences require significant troubleshooting to work around. Depending on the method used for such on-site synthesis it may not be possible to determine which have been made correctly or in what purity. However, when working on these scales conventional purification is not practical, as it would require one-at-a-time HPLC purification, so there is little alternative either way.

In Chapter 2 a library of peptides containing a mixture of O-GlcNAcylation and phosphorylation is prepared. This library represents biomarker sites of considerable importance on the tau proteins. These proteins have a central role in the biochemistry of Alzheimer’s disease (AD) and these biomarkers are known to see a sharp increase in phosphorylation as the disease progresses.

In Chapter 3 the library is printed to an array and used to examine the binding of several antibodies used in AD research. This work validates the array, demonstrates unusual and perhaps misleading binding in one of the major antibodies used in AD research and validates the concept of a mixed array in this type of work. It also lays the groundwork for much broader use of the library in the study of research antibodies and patent serum.

1.3.3 Bioconjugate vaccines

Immunization is the process of training a patient’s immune system ahead of any potential exposure to a pathogen. The most ancient form of this is likely a process called variolation. In this process, a healthy individual’s skin would be exposed to the fluid from the pustules from a smallpox infected individual. The result was a mild infection, very different from the more severe infections which typically result from respiratory exposure. Still, antibodies were generated and
the healthy individual gains some resistance to severe smallpox infection for life. By the early 1800’s it was known that the use of cowpox had much the same effect and was much safer.⁵⁰

For most viruses there is no such convenient cousin that can be dropped in to yield a safe and effective immunization. An approach that has proven effective is to make the equivalent of such a virus. The live attenuated influenza vaccine is such an example.⁵¹ Alternatively one could culture the strain of interest at scale, or one bearing the same glycoproteins, and inject an inactivated sample of the virus. Inactivated virus vaccines do not generate an infection at all, and instead rely on injecting enough viral glycoprotein to where an actual infection in the patient is not needed. There are issues with all these approaches. There are cases of attenuated vaccines resulting in transmissible virus.⁵² The antibodies generated may not be ideal. Adjuvants can help manage the magnitude of the immune response, but ultimately the viability of this sort of vaccine depends on the nature of the pathogen.⁵³ The rational design of vaccines demands a very different strategy. One such approach is to program a less pathogenic virus to express a target antigen. This is what we see with the Ebola vaccine.⁵⁴,⁵⁵ Again, the viability of the approach depends on the nature of the pathogen. The ability to target epitopes such as oligosaccharides is not likely practical with this method.

Simply introducing a foreign molecule into a patient’s bloodstream in isolation is rarely sufficient to provoke an adequate immune response for immunization. An avenue for making a molecule antigenic is to conjugate it to a carrier molecule (generally a protein or polypeptide).⁵⁶ This carrier should not be immunogenic itself, but should be known to facilitate the generation of antibodies against conjugated species when taken up by immune cells.
A compound such as a glycopeptide is unlikely to yield much of a response if injected in isolation. But joined to a carrier protein, robust generation of antibodies can be observed.

In Chapter 4, a glycopeptide representing a conserved region of the influenza A virus hemagglutinin stem is prepared. This glycopeptide is then conjugated to a carrier, CRM197. The resulting bioconjugate vaccine is then evaluated in mice as a model for a universal vaccine for influenza.
2 PREPARATION OF A TAU PEPTIDE LIBRARY

2.1 Introduction

The study of AD biomarkers is built on the use of site-specific antibodies. These antibodies allow for the quick and straightforward characterization of the frequency of a particular biomarker in a sample. However, their usefulness in this role depends on the extent to which they are actually bind their sites well and with specificity. An antibody which binds well to three or four sites may still be useful if those interactions are known and unproblematic for whatever experiment is being run. Alternatively, if the purpose of an experiment is to precisely characterize a single site, as is often the case, it may be best to choose an alternative. Either way, it is best to know so that the decision can be made rationally. As a result, work characterizing commercial antibodies has considerable utility.

2.1.1 A library to support diagnostics and antibody screening

A paper of significance in this area is Ercan et al. (2017). They prepare an array of peptides representing tau sites targeted by commercial antibodies, along with the inclusion of a number of PTM modifications. The work is fairly comprehensive but lacks the inclusion of O-GlcNAc. O-GlcNAcylation and phosphorylation are tightly intertwined, with the two processes competing for the same residues. But study of glycosylated tau is immature with no study of tau glycopeptide microarrays in the literature. Given how tightly linked the two processes are, there is a strong case to study the two in unison. Here a library of tau peptides including both phosphorylation and O-GlcNAcylation in all possible combinations is prepared.

Eventually the goal will be to cover all sites known to be phosphorylated in AD. However, for the preliminary project two of the most significant known biomarker sites were chosen. The first site T181, is more or less the gold standard AD p-tau biomarker. T181 is known to become
strongly phosphorylated early in the course of AD.\textsuperscript{59} There are many commercial antibodies for its detection. The second site, S202/T205 is well correlated with Ab-pE plaques.\textsuperscript{60} It behaves a little differently from T181, with frequency growing through AD as opposed to T181 which begins to decline to some degree after symptom onset.\textsuperscript{59} Further, the most widely used antibody in p-tau research, AT8, targets this site. If limited to just two sites then these two are a reasonable choice.

\subsection*{2.1.2 Design considerations}

The choice of amino acid sequence is paramount in any project involving peptides. Longer peptides tend to represent natural proteins more faithfully while also having worse yields than the equivalent shorter peptides.\textsuperscript{61} As such some degree of optimization and iteration can be difficult to avoid. In the case of a large array where iteration is performed in the SPPS step, yield at this step is of extreme importance due to the tedious and exceedingly laborious nature of HPLC purification of mixtures of peptides. High-throughput screening in this manner demands the use of a standardized HPLC method, not one customized for each peptide. The sheer number of peptides combined with the practical limitations imposed by the realities of the available equipment can make the process cumbersome. In short, time invested in sequence optimization is typically well-spent.

Further, the production of peptides bearing PTMs also calls for a strategy regarding their introduction. Here, only phosphorylation and simple O-GlcNAc is needed, so the prudent course of action is to prepare or purchase phosphorylated and O-GlcNAc bearing amino acids and incorporate these building blocks into the growing peptide chain much like all the other amino acids. In the case of larger oligosaccharides alternatives can make sense.

There are a few other considerations. The peptides are to be conjugated to a solid surface via an NHS-activated ester linker. Thus, the N-terminus should be left free and not capped. The
choice of c-terminus is more discretionary. Since these peptides are meant as a representation of a natural system, it was decided to use a rink amide resin. Such a resin will leave a primary amide in place of a carboxylic acid. The natural protein has a secondary amide here, of course (since the polypeptide chain continues), so the use of a primary amide in place of the charged acid that is more commonly expected at the C-terminus is defensible.

2.2 Plan of synthesis

Synthesis must begin with the preparation or purchase of building blocks and other starting materials. These materials are, first, the resin, the amino acids, and the reagents needed to support to facilitate coupling (HATU and DIPEA) and Fmoc deprotection (piperdine). Fortunately, phosphorylated amino acid is available commercially at modest prices. This is not usually the case for glycosylated amino acids. Synthesis of O-GlcNAc-bearing peptides typically must begin with the synthesis of O-GlcNAc-bearing amino acids. Realistically, for the synthesis of a sizable array, the amino acid must be made in gram scales. Here, 10 g of the threonine compound and 4 g of the serine compound was made.

The next step is the operation of the synthesizer. To produce peptides in large numbers a multichannel machine supporting synthesis at small scales (<10 mmol) is needed. After the machine finishes its run, peptides are cleaved by the action of TFA with scavengers. This step also removes sidechain protecting groups. Treatment by aqueous hydrazine removes the hydroxy protecting groups from the O-GlcNAc-bearing residues leaving the final compound, albeit in an impure state.
The peptides were synthesized by SPPS as discussed in the introduction. Glycosylated amino acid was prepared in-house. After cleavage, the acetyl groups on the sugars are removed. This is by hydrazine, 5%. Then HPLC purification can be performed, and the product lyophilized and massed. Once a mass is had for the peptides, they are diluted to a stock concentration for further preparation for array printing.

To purify the peptide HPLC is required. A standardized method collecting the same time slices is called for so that most purification can be automated and performed in a timely fashion. The same sequence with different PTMs will have a different retention time. Since only sub-milligram quantities are needed for the printing, a single run for each target is sufficient. Attempting to establish an optimal retention time and collection range for each combination of sequence and PTM configuration will dramatically extend the time needed for purification over mass collection of small slices in a broad range. At the scale of scores or hundreds of targets even the later approach can take many weeks when the time for characterization is included.

The fractions representing each target can then be lyophilized and massed on a milligram balance. Once targets are massed samples in buffer can be prepared and array printing can be performed.
2.2.1 **Sequence choice considerations**

Here there are three possibilities to put at each PTM site. A site that can bear a PTM can either be phosphorylated, glycosylated with O-GlcNAc, or left unmodified. For a single site there are three possible peptides. For two, the number should be $3^2$, or nine, and this is easy verifiable. The first site can either be phosphorylated, glycosylated with O-GlcNAc, or unmodified. The same is true of the second site. This results in five, not six, peptides because the simple chain of amino acids resulting from the absence of a PTM is already represented in the first three. Of course it is also possible to have both sites bearing O-GlcNAc, or both being phosphorylated. This brings the total to seven. It is also possible for the first site to bear O-GlcNAc and for the second to be phosphorylated and vice versa. This brings the total to the nine peptides that were expected. Reducing the number of included PTMs increases the number of sites that can be studied for the given resources.

It is not possible to examine all possible influences on a site at once. It is possible, for conformation reasons, that binding at the site will be hugely impacted by a change made dozens of residues away. Or, at least, it is not rigorous to claim there could be no such significant impact. Only a narrow subset of interaction effects can ever realistically be looked at. It is fundamental to the design of the experiment to define reasonable boundaries. Necessarily, the scope of the modifications made in parallel must be limited.

It was decided as a part of experimental design that immediate neighboring PTMs of sites of interest would bear the full set of possible modifications, so long as the total set being varied in concert was no more than three. If more are needed then two overlapping sets of three are used. Two sets of three, in the case where the amino acid sequence is the same and the set of PTMs overlaps by two work out to $2(3^3) - 3^2 = 45$. The subtraction term results from redundancy occurring
where the differing PTM is unmodified in the second set. This is more than a 44% reduction over varying all four sites together.

2.3 Experimental

2.3.1 Building block synthesis

Figure 8. Synthesis scheme of O-GlcNAc Ser/Thr building blocks.

The first choice that must be made in designing a glycosylation is the choice of donor. All other steps will derive from the need to support that reaction. Here a Schmidt donor approach was chosen. The two most popular approaches at the moment to forming glycosidic bonds are Schmidt and thiophenol-based donors. Schmidt has a number of advantages over the thiophenol-based alternative. First there is fine downstream control of the stereochemistry of the resulting linkage. Choosing between alpha and beta can be a simple matter of temperature and solvent choice. Preparation of the trichloroacetimidate is quick and straightforward, involving no malodorous reagents. The thiophenol approach leaves thiophenol as a leaving group on roughly the same mass-scale as the monosaccharide. In addition to being malodorous, any residual thiol at all, even only a trace, will poison a palladium catalyst. Schmidt donors also have the advantage of being more
stable. TLC standards of compound 4 remained usable despite being in solution at room temperature for months. While the desired product is acetylated at the second position amine, glycosylation with donors bearing second position NAc can be difficult. The intermediate, oxazoline, is much more stable than the acyloxonium seen in the case of second position -OAc resulting in a less reactive donor.\textsuperscript{47} Since there is also an Fmoc group that must survive the entire synthesis, any protecting group installed at the second position must be cleavable by something other than base. Troc was chosen here as it has neighboring group participation and is cleaved by treatment with zinc dust and mild acid.\textsuperscript{63,64}

Synthesis begins by dissolving glucosamine in water. The system is brought to 0°C and then 2,2,2-trichloroethyl chloroformate (Troc-Cl) diluted into ether is added followed immediately by sodium bicarbonate. Consistency of the reaction system is improved if the sodium bicarbonate is added second. The volume of ether should be around 10% the volume of the water. After an hour the ice bath is removed and the reaction system is allowed to come to room temperature with continued stirring. Within another two or three hours visible solid should be present in the flask. The reaction is allowed to run for another 12-24 hours at rt, after which point the fine white solid is collected by vacuum filtration. The next step is peracetylation. The solid from the previous step is dissolved in equal volumes of pyridine and acetic anhydride and stirred, typically for at least 8 hours, yielding compound 2. It is important to allow the reaction to complete here, as the last acetyl group to be added will not be the same as the one removed in the following step and the two will be inseparable. Workup consists of concentrating on the high vac, diluting with DCM and washing with 1M HCl and saturated sodium bicarbonate before final evaporation to a syrup.

Compound 2 is dissolved in THF to which BnNH\textsubscript{2} is added. This reaction is complete after 12 h, however in practice is it usually better to halt it sooner. While conditions favor removal of
the protecting group at position 1, position 6 will also deprotect very slowly. Removal of the di-
hydroxy compound is more difficult than removal of the starting material and its carry over to
following steps is more problematic. After the reaction is halted through the addition of an
equivalent of acetic acid, the system is evaporated, dissolved in DCM and dried onto silica. A
column must then be performed with an average solvent ratio of 2:1 hexanes to ethyl acetate
yielding compound 3 as a solid foam.

Compound 3 is dissolved in anhydrous DCM and brought to 0°C with an ice bath under an
inert atmosphere. DBU and trichloroacetonitrile are added with continued stirring. After TLC
shows completion concentrate onto silica and column with an average solvent ratio of 1:1
hexanes:ethyl acetate. The resulting foam solid is compound 4.

Acceptor, either Fmoc-Ser(OH)-OBn or Fmoc-Thr(OH)-OBn, is dissolved along with
compound 4 into anhydrous DCM. The system is concentrated to dryness and dissolved a second
time, also into anhydrous DCM. Molecular sieves are added and an inert atmosphere established.
The system is brought down to -18°C or lower. This is achieved using a salt/ice bath. It can be
beneficial to go colder using acetonitrile and liquid nitrogen for this type of glycosylation,
however, in practice here salt/ice was found to be sufficient. Once temperature is established,
TMSOTf is added. Completion takes several hours. After completion the sieves are filtered off
and the system is dried onto silica. Column with an average solvent ratio of 1:1 hexanes:ethyl
acetate. The resulting white powder is compound 4. This reaction was attempted with Fmoc-
Tyr(OH)-OBn several times, only to very poor results.

Compound 4 is dissolved in a 1:1 mix of DCM and acetic acid and stirred in the presence
of zinc dust for 3-4 hours. The zinc is filtered off, the system evaporated to dryness, dissolved in
toluene and dried again, with this step being repeated until no odor of acetic acid can be detected.
The system is then dissolved in DCM. Acetic anhydride and pyridine are added. This step is quick, an hour or two is sufficient. The system is then concentrated, dissolved in DCM, washed with 1M HCl and saturated sodium bicarbonate before being concentrated onto silica and columned average solvent ratio of 1:1 hexanes:ethyl acetate yielding a white solid, compound 5.

The final step is the removal of the benzyl protection on the carboxylic acid through hydrogenation with a Pd catalyst in anhydrous THF. Usually, when this is performed in the presence of an Fmoc, the result is the loss of the Fmoc as well. This is seen here, however, the loss of the Fmoc was generally slow enough relative to where it was possible to catch and isolate the It seemed loss of Fmoc was faster with the threonine compound than with the serine, but this was not examined beyond a few TLC observations incident to production of the needed compound. After roughly two hours the reaction is complete and the catalyst filtered off. The system is concentrated onto silica and purified with a DCM:MeOH:AcOH 90:8:2 solvent system. Final product is a coarse, off-white solid. **N-α-Fmoc-O-(2-Acetamido-2-deoxy-trio-acetyl-β-D-glucopyranosyl)-L-serine.** Yield: 75%. 1H-NMR (400 MHz, DMSO-d6) 1.74 (s, 3 H, NAc), 1.92, 1.98, and 2.01 (s, 9 H total, OAc), 3.71 (q, 1 H, CH, J=9.98 Hz), 3.78 (dd, 1 H, CH, J=10.55 Hz, 3.49 Hz), 3.85 (dq, 1 H, CH, J=9.97 Hz, 1.60 Hz), 3.96-4.05 (m, 2 H), 4.17-4.33 (m, 5 H), 4.73 (d, 1 H, anomeric CH, J=8.46 Hz), 4.86 (t, 1 H, CH, J=9.71 Hz), 5.11 (t, 1 H, CH, J=9.96 Hz), 7.29 (d, 1 H, NH, J=8.32 Hz), 7.32-7.37 (m, 2 H, Fmoc ArH), 7.43 (t, 2 H, Fmoc ArH, J=7.45 Hz), 7.74 (d, 2 H, Fmoc ArH, J=7.26 Hz), 7.90 (d, 2 H, Fmoc ArH, J=7.55 Hz), 7.93 (d, 1 H, NH, J=9.01 Hz), 12.78 (s, 1 H, COOH). MALDI-TOF: [M-H]- calcd for C32H36N2O13 655.21; found: 655.75. **N-α-Fmoc-O-(2-Acetamido-2-deoxy-trio-acetyl-β-D-glucopyranosyl)-L-threonine.** Yield: 64%. 1H-NMR (400 MHz, DMSO-d6) 1.15 (d, 3 H, CH3, J=6.35 Hz), 1.79 (s, 3 H, NAc), 1.94, 1.98, and 1.99 (2, 9 H total, OAc), 3.65 (q, 1 H, CH, J=10.16 Hz), 3.77-3.85 (m, 1 H), 3.96-4.03
(m, 1 H), 4.10 (dd, 1 H, J=8.78 Hz, 3.43 Hz), 4.19-4.31 (m, 5 H), 4.72 (d, 1 H, anomeric CH, J=8.45 Hz), 4.84 (t, 1 H, CH, J=9.69 Hz), 5.14 (t, 1 H, CH, J=9.97 Hz), 6.59 (d, 1 H, NH, J=8.77 Hz), 7.33 (t, 2 H, Fmoc ArH, J=7.43 Hz), 7.43 (t, 2 H, Fmoc ArH, J=7.44 Hz), 7.75 (d, 2 H, Fmoc ArH, J=7.27 Hz), 7.90 (d, 2 H, Fmoc ArH, J=7.50 Hz), 7.98 (d, 1 H, NH, J=8.93 Hz), 12.83 (s, 1 H, COOH). MALDI-TOF: m/z [M -H] calcd for C\textsubscript{33}H\textsubscript{38}N\textsubscript{2}O\textsubscript{13} 669.23; found: 669.77.

2.3.2 Peptide synthesis

All peptides prepared in this chapter were made using a CEM Multipep 2 multichannel peptide synthesizer using a heated 96-well block. Rink amide resin with a loading of 0.68 meq/g and a mesh size of 100-200 was suspended in a 1:2 mix of DMF and DCM (23.5 mg/mL). Amino acids were dissolved in DMF to a concentration of 0.5 M. HATU at 0.5 M in DMF was used as the coupling reagent along with DIPEA at 1.33 M in NMP as the coupling-step base. Piperdine, 20% in DMF was used as the Fmoc deprotection solution. Capping solution used was 5% acetic anhydride in DMF.

Resin suspension (0.5 mL) is added by pipette to a fritted reaction-well in a 96-well heated plate which is then immediately drained. The machine method begins with swelling the resin in DMF (400ul) and bringing the temperature to 50 °C. After draining the coupling cycle begins. 1: Resin is deprotected by treatment with deprotection solution (180 ul) for 5 minutes twice. 2: The resin is washed with DMF. 3: HATU solution, DIPEA solution, and amino acid solution are added to a mixing vial in an 80/24/84 ratio and a total of 188 ul of this mix is added to the resin. Resin is allowed to react for 30 minutes with the temperature maintained at 50 °C. This coupling step is performed twice, both for 30 minutes each. 4: The resin is then washed. 5: Capping is performed by treatment with capping solution (400 ul) for 5 minutes. 6: The resin is then washed with DMF. The steps 1-5 are repeated for each amino acid to be installed. After the final amino acid is installed
deprotection is again performed by treatment with 20% piperdine in DMF (180 ul) for 5 minutes twice.

After the machine is finished, the resin is removed to a fritted tube and treated with TFA:Water:TIS:DODT in a 37:1:1:1 ratio for 4 h for cleavage of the peptide from the resin and the removal of the side chain protecting groups. Acetylation on the GlcNAc is then removed by treatment with 5% aqueous hydrazine for 3 hours.

2.3.3 First attempt

In the interest of representing the epitope as well as possible, >20mers were chosen.

![Peptide Structures](image)

*Figure 9. The initial peptides chosen*

*Positions in red are the PTM sites of greater interest. The sites in cyan will be varied as well.*

The set of all possible PTM combinations for these peptides is 90, a manageable number for the first array. Unfortunately, the yield of these peptides was not great. The first batch consisted of the bare peptides as well as all the singly modified peptides. For many peptides, no product could be seen in the MALDIs of the crude material. The MALDIs were dominated by peaks of mass corresponding to the sequence truncated after proline and capped with an acetyl group. Such truncated peptides would result if Fmoc deprotection at the proline goes well and coupling of the following amino acid goes poorly. The sequence is capped using acetic anhydride in the final step of each cycle. If the peptide were not capped here, the result would be a point deletion, which is typically harder to remove.
Figure 10. Example from the first set of peptides.

Capping after P200 is represented by the peak of 1594 m/z. Several other proline capped can be seen among the lower masses. No product can be observed in the second spectrum.
Curiously, yields of phosphorylated amino acids were observed to be worse than their glycosylated equivalents, with messier corresponding crude MALDIs.

2.3.4 Final sequences

Although a number of methods were tried to improve the overall yield of the sequences, these ultimately did not bear fruit. The only thing that improved the yield by enough to make a meaningful difference was shortening the peptides and in some cases shifting the sequence one or two positions.

![Figure 11. The final sequences used](image)

The improvement in the appearance in the crude MALDI of the peptides of reduced length compared with their longer equivalents is stark. In the test batch of the full length bare and mono-PTM peptides had little if any product. The equivalent test batch of the shorter peptides was entirely usable.

Another issue that arose was the incomplete deprotection of the phosphate. The commercial phosphorylated threonine and serine have a benzyl protecting group. While on a hydroxyl or carboxylic acid this group would require hydrogenation, on a phosphate treatment by standard TFA-based cocktail is reportedly sufficient. In practice the protecting group could be seen in most MALDIs. However, increasing the cleavage reaction time to five hours seemed to help considerably. Considering only a very small yield (~5%) would leave enough final product to work with, no further consideration of the issue was made.
Figure 12. Example of shorter peptides

These 14 mer peptides are the analogs of the ones seen in Figure 10. Capping after P200 (709 m/z) is minimal. A different issue is seen here, with the protecting group failing to come off the phosphate (1475 m/z). This can be managed with protracted treatment with cleavage cocktail.
2.3.5 Purification and finalization

Purification took place on a Shimadzu preparatory HPLC with a semi-preparatory C18 column. The solvent system was water and acetonitrile, both with 0.1% TFA, on a gradient of 5%-20% acetonitrile over 20 minutes. Fractions were collected starting from 5 minutes going to 15 minutes. Afterwards fractions were sent for MALDI, the MALDI results were compared with the peaks seen in the HPLC, and reasonable fractions combined and lyophilized. After the first round of lyophilization product was transferred to 1.5 mL Eppendorf tubes that had been massed on a milligram balance using the smallest amount of water possible. Lyophilization was performed again, and final masses measured. To the samples was added a volume of millipore water needed to result in a standard concentration as would be needed for the array printer. Final analysis, such as MALDI and HPLC was performed at this time.

2.4 Discussion

A mixed tau glycopeptide/phosphopeptide library was successfully prepared. Currently the library does not yet contain all sequences planned. Discussion on the scope of the current library can be found in Chapter 3. Preliminary arrays have been constructed using a relevant subset of the library as it existed at the time of printing and the library will continue to be expanded in the interest of representing the sites as well as possible and elucidating more interactions. The preparation of arrays from the library is not technically challenging given the tools available and arrays will continue to be printed and tested as the library expands to reach its original design parameters.

The main outstanding challenge in the expansion of the array is in the continued low yields of the (199) SPGSPGTPGSRSRT series. Attempts to prepare the set have yielded only the mono-PTM peptides and several of the multi-O-GlcNAc peptides. It may well be the case that whatever
difficult subsequence caused the problems in the first batch of peptides has survived to some extent in B despite its smaller size. To address this additional shift sequences will be made. If this is insufficient, differing protecting groups will be tried for some of the residues with large groups. If aggregation is causing the difficulties with the synthesis then any change which results in a change in the conformation a possible avenue. A possible last resort would be protection of the backbone nitrogens or, in the absolute worst case scenario, the use of a microwave synthesizer to prepare a particularly relevant subset of the total.

The need to prepare amino acid to support the synthesis of the library has been addressed. Gram-scale preparation over a few weeks is routine at this point in the project. As such, this will not be an issue in a follow up project.

The follow up project will cover all PTM sites observed in AD. This is too many sites to allow for much in the way of individual tweaking of the sequences. As such it should probably target a length of 15 residues or less from the beginning. More than that and synthetic challenges will burden the project.
2.5 Supporting Spectra

Figure 13. GlcNTroc (3x Ac)
$^1$H NMR, 400 MHz, CDCl$_3$. 
Figure 14. GlcNTroc Trichloroacetimidate Donor
$^1H$ NMR, 400 MHz, CDCl$_3$. 
Figure 15. Fmoc-Thr(3xAc-GlcNTroc)-OBn

1H NMR, 400 MHz, CDCl₃.
Figure 16. Fmoc-Ser(3xAc GlcNAc)-OBn
$^1$H NMR, 400 MHz, CDCl$_3$. 
Figure 17. Fmoc-Thr(3xAc GlcNAc)-OBn

H NMR, 400 MHz, CDCl₃.
Figure 18. NMR of Fmoc-Ser(3xAc GlcNAc)-OH
$^1$H NMR, 400 MHz, DMSO-d6
Figure 19. NMR of FmocThrGlcNAcOH.

$^{1}H$ NMR, 400 MHz, DMSO-d$_6$. 

FmocHN-COOH
Figure 20. NMR of Fmoc-Ser(3xAc GlcNAc)-OH
$^{13}$C DEPT135 NMR, 400MHz, DMSO-d6
Figure 21. NMR of Fmoc-Thr(3xAc GlcNAc)-OH
$^{13}$C DEPT135 NMR, 400MHz, DMSO-$d_6$
Figure 22. NMR of Fmoc-Ser(3xAc GlcNAc)-OH
$^1H/^13C$ HSQCED NMR, 400MHz, DMSO-d6
Figure 23. NMR of Fmoc-Thr(3xAc GlcNAc)-OH
$^1$H/$^1$C HSQCED NMR, 400MHz, DMSO-d6
Figure 24. MALDI-TOF of serine building block
Figure 25. MALDI-TOF of threonine building block
3 A TAU ARRAY FOR ALZHEIMER’S DIAGNOSTICS

3.1 Introduction

The diagnosis of Alzheimer’s Disease (AD) is a difficult thing at present. There is no blood test, no quick and easy screen that could be applied to the population at large. This is unfortunate, as biochemical changes can be seen many years before symptoms become noticeable.\(^5\)\(^9\) Indeed, the disease is often not diagnosed until very late, if it is diagnosed at all. Existing treatments for AD are poor to put it charitably, but if the disease can be identified through mass screens a decade before symptoms, a treatment that merely slows its progress could well prove up to the task of alleviating much or most of the suffering in patients and families.\(^6\)\(^5\),\(^6\)\(^6\) Indeed, the restoration of badly damaged brain tissues may well turn out to be outside the scope of what therapeutic agents can achieve. In any case, preventing damage before it occurs is incontrovertibly superior to attempting to correct damage after the fact. Either way, the path forward for AD lays in early detection by some screening method which can be cheaply and easily applied to many millions of people after a certain age as a part of routine health checks.

One of the two main biochemical changes occurring in AD is the formation of hyperphosphorylated tau tangles.\(^1\)\(^7\) Tau normally serves to stabilize microtubules in axions.\(^6\)\(^7\) These tubules serve vital transport purposes.\(^6\)\(^8\) When tau becomes hyperphosphorylated it prefers to associate with itself and forms tangles.\(^6\)\(^9\) There is some debate as to the toxicity of these tangles.\(^7\)\(^0\),\(^7\)\(^1\) Some forms appear worse than others but in any case the result of their formation is a failure to stabilize the tubules and a diseased state in the neuron. The diversity of phosphorylation sites has proven a fruitful source of biomarkers of AD, though no commercial tests have yet resulted from its study.\(^7\)\(^2\),\(^7\)\(^3\)
3.1.1 P-tau

Tau becomes phosphorylated over the course of many conditions. Alzheimer’s Disease is one of these tauopathies.\textsuperscript{74} Some of these conditions are the result of injury, possible suggesting some sort of defensive mechanism behind increasing phosphorylation levels. That said, tau pathologies can behave much like prions, propagating once introduced from a foreign source.\textsuperscript{75} Currently, this is not well understood.

Phosphorylation and O-GlcNACylation compete with each other for the modification of Ser, Thr, and Tyr residues of the tau protein.\textsuperscript{58} Since hyperphosphorylation of tau promotes its aggregation, O-GlcNACylation of tau has been hypothesized to have a protective effect by blocking tau phosphorylation and decreasing its deposition in tangles.

Despite the relative isolation of neurological tissues, tau in healthy individuals is known to exist in the blood at pg/mL concentrations. Further, healthy individuals are known to have an autoimmune response to tau. As tau loses O-GlcNAc as it becomes hyperphosphorylated, it stands to reason that any immune response to O-GlcNAc sites on tau may well become diminished over the progression of AD.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{tau_2N4R.png}
\caption{Tau (2N4R) Known AD phosphorylation sites in blue and green. Green sites have also been observed bearing O-GlcNAc. Regions underlined are studied in the current work.}
\end{figure}
3.1.2  $pT181$, $pS202$, and $pT205$ Antibodies

The two sites chosen for study in the initial project are $pT181$ and $pS202/pT205$. Eleven antibodies have been purchased for this project. For the $pT181$ phosphorylation site antibodies 5H9L11, AT270, MABN388, EPR22064, EPADXR4, EPR23506, EPADXR4 were purchased. For the $pS202/pT205$ region antibodies AT8, EPR20390, EPR2402, EPR2403(2) were purchased. Antibodies AT8 and EPR20390 bind when both sites are phosphorylated. EPR2402 binds phosphorylated $pS202$ and EPR2403(2) binds phosphorylated $pT205$. The most widely cited antibodies in this set are AT8 and AT270 with 621 and 108 references respectively per the manufacturer.

At the time of the printing of the array the library contained a much more complete set of peptides representing the region around $pT181$. As such the two best-cited antibodies from the set representing this region.

3.2  Methods

A mixed glyco/phospho-peptide array was prepared and used to screen binding against tau epitopes. All sites known to be phosphorylated in AD will eventually be included as part of the broader study. For the initial study, two sites were chosen, $pT181$ and $pS202/pT205$. These sites have commercial antibodies that are well-studied and are a fixture in the AD biomarker literature. Every marked PTM site will be varied, and will either be unmodified, phosphorylated, or O-glycosylated in all possible combinations. A library of peptides for these sites was prepared in Chapter 2.
Figure 27. Illustration of the nomenclature system

Sequence labeled C covers site 181 and A and B cover 202/205. All sites in red and cyan will bear either be left unmodified, bear phosphorous, or O-GlcNAc. In every possible combination so that we can see the effect of including both PTMs on binding. Each peptide is named starting with the letter of the sequence. Either C or A or B. Then there is a number of additional characters equal in number to the number of possible sites. So for C there are four more characters since there are four PTM sites in the sequence. A and B have three so they are followed each by three characters. Zero for unmodified. G for glycosylation. P for phosphorylation.

Due to the number of peptides a consistent and descriptive nomenclature is needed to refer to each member of the library. Here, each sequence has a letter, A, B, or C, for the current set. The letter is then followed by a string containing zero, “p,” or “g,” representing an unmodified PTM site, or sites bearing O-GlcNAc or phosphate respectively. The length of the string is equal to the number of PTMs and the order matches the order of the PTM sites in the peptide. So A000 and B000 are the unmodified peptide sequences. C0000 is the unmodified C peptide, having four positions since it has four PTM sites. Apg0 would be the sequence with a phosphate at the first PTM site, an O-GlcNAc at the second, and no PTM at the third site.

At the time of the printing of the second array, only the bare peptides, the single PTM peptides of A and B, and the full C series, save for the triply phosphorylated peptides were available. Additionally, some unusual binding for antibody AT270 was observed in the first array, so a number of additional short sequences were made. The array was printed with all the peptides available at the time.
3.3 Results

Two arrays have been printed so far. The first array is a subset of the second showing nothing different and so will not be discussed here. The second array consists of all the singly modified A and B series peptides as well as all C series peptides save for the triply phosphorylated peptides. Ultimately the yield of the phosphorylated peptides was low enough to where production of the triply phosphorylated form was not practical, even in the truncated form. Binding studies are ongoing.
Figure 28. Antibody 5H9L11 binding on Array 2
Figure 29. Antibody AT270 binding on Array 2
3.3.1 Preliminary observations

Much of the signal in the array 2 data makes immediate sense. Both antibodies are observed binding C0p00, which contains the binding site for both 5H9L11 and AT270. Both are observed binding Cp00 to a much lesser extent. The site differs from the formal binding site by having an alanine in place of a proline two residues removed from the phosphorylation site. In terms of bulk and hydrophobicity the two are close enough to where this is not a surprise. The phosphorylation site is also much closer to the terminus than in C0p00. Binding is observed for both wherever a C-series peptide is phosphorylated in either the first or second position, regardless of other PTMs. Very little binding is observed elsewhere, save for an anomaly discussed in the next section.

Figure 30. Select peptides and their binding with respect to 5H9L11 and AT270

The binding ratios were observed to vary in response to neighboring PTMs. In both antibodies Cgp00 has worse binding compared with C0p00. This is reasonable. C0p00 is the native site. Crowding from a glycan nearby is unlikely to help. Also, Cpp00 binds much more poorly than C0p00. It may well be the case that the ideal antibody binding site includes the bare threonine...
at position 175 so including either a glycan or a phosphorylation site there does serious harm to the binding.

![Figure 31. Overview of the relationships seen](image)

The ideal binding site for both antibodies, as per the manufacturer’s claims, shows the strongest binding. Inclusion of either an additional phosphorylation or glycosylation six residues toward the n-terminus does damage to the binding. Adding an O-GlcNAc improves the mediocre binding of the peptide with a phosphorylation site on the PTM site near the n-terminus even though the glycosylation in isolation results in no binding.

Something very different is seen when an O-GlcNAc is added to Cp000. Cpg00 does substantially better than Cp000. This is seen in both antibodies but the difference in AT270 is a full 5-fold increase over Cp000. This difference is difficult to explain. An interaction with residues on the antibody is a possibility. The glycan may be driving a more favorable conformation. Arriving at a conclusive answer as to why this pattern is seen here would require extensive follow up studies.

### 3.3.2 AT270 binding site

Early in the study unusual binding was observed for antibody AT270. It was already known that this antibody exhibits some degree of non-specific binding. But a willingness to bind both -AKpTPP- (Cp000) and -PGpSRS- (B00p) was seen, suggesting that the antibody was able to bind
a phosphorylated site next to a positive charge. To follow up eight additional peptides were added to the array.

**Figure 32. Unusual binding of AT270**

A series of peptides was made in response to the observation that AT270 will bind B00p. Starting from a sequence known to bind the positive charge is dropped, showing that binding depends on the neighboring positive charge. Then the system is simplified to little beyond the combination of a positively charged residue and a phosphorylation site. AT270 is willing to bind all these forms of charge plus phosphorylation while 5H9L11 shows greater specificity.

The first peptide sequence is AKpTPP. This was chosen because it is known to bind both AT270 and 5H9L11. The next in the series is AQpTPP. Glutamine is similar in size to lysine without bearing a charge. This is meant to see if the charge is needed for binding. The third peptide continues in much the same vein. Here a negatively charged residue is used to see if this is enough to kill off any residual binding. The fourth, GGpTGG, is meant to simplify the environment down to little beyond a phosphorylation site. Then, with the fifth through eighth peptides, the positive charge is brought back in the form of arginine or lysine neighboring the phosphorylation site. Removal of positive charge immediately eliminates detectable binding in AT270. 5H9L11 retains
binding until the switch to a negative charge. The subset of the remaining peptides with a phosphorylated site next to a positively charged residue on either N or C terminus sides do show some degree of binding with AT270. Binding for 5H9L11 is only seen when lysine is on the n-terminus side of the phosphorylation site, which matches the configuration at pT181.

Further expansion of this panel is warranted. It would be interesting to see if binding would remain if histidine were used in place of lysine or arginine. The glycines could also be replaced with bulkier amino acids, both hydrophilic and hydrophobic.

3.4 Discussion

3.4.1 Binding constants

Currently, there is no way of establishing $K_D$ in the system. Binding is relative from one peptide to another, but there is no indication of order of magnitude. The binding could be too poor to be of interest across the board. In order to address this issue and establish a point of comparison, a secondary assay is needed. To resolve this issue Biacore will be used to establish binding constants for several combinations of peptide and antibody known to have differing signal intensity across the array. Biacore works using a principle called surface plasmon resonance. In short, one species is covalently bonded to the surface of a gold-coated slide and polarized light is reflected off the backside of the gold film. When a second species binds to the first there is a change in resonance resulting in a change in the absorption angle in the light reflected off the surface. Usually in Biacore a ligand is bound to the chip and the protein is passed over. Here, due to its precious nature, the antibody will be immobilized and the peptide will be allowed to flow over it.
3.4.2 Future work

The biggest current limitation on the expansion of the work is the size of the available library. While the C series is complete, B is not. A is nearly complete, with only two peptides missing at present. Once the expanded library is complete arrays will be printed.

Here, two of the eleven antibodies have been assayed against the array. Once the expanded array is finished all eleven antibodies will be assayed against it. This expanded set includes AT8, the single most widely used antibody in p-tau research. AD and healthy patient serum has been acquired from Lee Biosolutions. Once the final version of the array has been printed and validated the serum will be run. Further follow-up work will be performed as needed to validate binding of note seen at this step. Binding seen at this step would represent results that could directly apply to the development of blood test based AD detection systems.

3.4.3 Conclusions

A tau peptide library prepared as per the previous section was printed as an array. This array was successfully used to examine relative binding of commonly used p-tau antibodies. The array shows strong signal, pushing passed the detection threshold in some cases. The signal varies from peptide to peptide. In most cases the variation was within the scope of what one might expect. In a few cases the results were well outside simple explanation, warranting further examination and follow up. In any case, the diversity of results suggests the overall approach is sound. Including O-GlcNAc did result in complex signal which may nor may not be borne out in other antibodies and may or may not help improve recognition when applied to serum.

Further, in the course of this work some explanation of the nature of the nonspecific binding of AT270 was found. Given the wide use of AT270 in the study of the pT181 biomarker site, these results advance our understanding of the works built on this antibody.
4 A PATH TO A UNIVERSEAL INFLUENZA VACCINE

4.1 Introduction

The influenza vaccine is one of the most well-known. Every year the vaccine is updated to target the strains of influenza expected to be prevalent in the upcoming year. No vaccine currently targets more than a handful of cherry-picked strains. Developing a universal vaccine for influenza – one that would target whole families of strains, possibly including those that have not yet evolved – has been a goal for decades. This has proven difficult as the regions of the virus most readily accessible to the immune system are not conserved. When a patient fights off one strain more often than not their immune system preferentially generates antibodies for the non-conserved epitopes. When a company prepares a flu vaccine they encounter the same issue and only manage to produce vaccines that confer resistance to the targeted strains. In some years these companies do a fairly good job of predicting what strains will be common. In other years they mass produce vaccines for less optimal strains. And of course, every year a new series of strains emerge. Occasionally a one of these has pandemic potential.

4.1.1 The influenza viruses

The influenza viruses are a group of RNA viruses in the Orthomyxoviridae family. They are negative sense, meaning translation cannot occur directly off the viral RNA. Instead, transcription is performed by an RNA polymerase working from the 3′-to-5′ end. The resulting intermediary is then translated. This contrasts with positive sense viruses where translation occurs directly off the viral RNA or ambisense viruses where both are seen. Influenza is able to undergo reassortment, a process whereby genes are shared between strains. The IAV genome contains 8 strands of RNA, which encode for total of ten proteins. A cell infected with more than one strain may generate virions bearing a novel mix of RNA strands. So a host infected with an H1N1
virus and a H2N2 virus could pass on a H1N2 virus. This ability, combined with the diversity of reservoirs of IAV, is a driving force in the emergence of pandemic strains. No new mutations are strictly necessary for a new strain to emerge. It can be sufficient for a human-adapted strain and an animal strain to cross.

There are four genres of influenza, type A (IAV), type B (IBV), type C (ICV), and type D (IDV). Clinically, IAV tends to present with cough, sneezing, headache, malaise, as well a range of other symptoms. Sometimes serious complications such as pneumonia or myocarditis can result. Asymptomatic cases are known, but there is some ambiguity as to the importance of such cases on the management of the spread of the virus. IAV is the biggest driver of mortality in humans. One reason for this is its extensive animal reservoirs, including both avian and mammalian hosts. Usually IAB infection results in less severe symptoms though this will vary from case to case. The primary reservoir of IBV is humans, so there is much less potential for sudden crossover events resulting in pandemic-potential strains, although it is a frequent cause of illness in humans. ICV is an underappreciated childhood virus, with perhaps 90% of children producing antibodies for it. Symptoms tend to be mild. IDV was only discovered in 2011. The extent to which type D infects humans remains somewhat unknown. Antibodies against type D have been found widely in human serum and workers with direct interaction with cattle (a primary reservoir), were shown to have a higher seropositivity compared with non-cattle exposed populations. While the type has not been studied as extensively as the others, there is good reason to believe the type has only modest pandemic potential. The biggest concern remains IAV.

4.1.2 Public health impact

Each year around 300,000-500,000 die as a result of influenza virus infections on average. Influenza is a respiratory pathogen. While it can spread through direct contact with eyes
and other mucosal tissues, a more common transmission mode is the spread through droplets and aerosols.\textsuperscript{100,101} The relative significance of droplets vs aerosols remains a matter of debate.\textsuperscript{102,103} Transmission is greatest in colder months (“flu season”) due to a range of factors that are still a matter of debate, including humidity, protracted survival of the virus in reduced temperature air, and an coincident increase in frequency of vitamin D deficiency in human populations.\textsuperscript{104,105} The seasonal impact of influenza remains high even in the absence of a pandemic.\textsuperscript{106} Estimates of the annual cost of the seasonal flu in the United States exceed 87 billion dollars.\textsuperscript{107} Of course, the impact of major pandemics is far greater than the seasonal flu.\textsuperscript{108,109}

All five of the influenza pandemics seen in modern times are of type A (IAV). The first modern influenza pandemic was the 1918 Influenza Pandemic. Often called the “Spanish flu” this is a bit of a misnomer, as actually it is not very likely to have originated in Spain. Spain became associated with the outbreak because they were neutral in the first world war, which was being fought at that time. As such they faced less internal pressure to avoid publication of public health concerns and many early reports came out of Spain.\textsuperscript{110} It is likely the place of origin will never be known for sure. It is known that pigs were exhibiting signs of influenza-like illness in 1918 in the United States, however it is suspected this transmission was from human to swine. Similar transmission was also seen in China.\textsuperscript{111}

Once the virus began spreading in humans, the long-distance supply lines of the world war gave it easy passage overseas. Roughly a third of the world’s population was infected with this virus which ultimately claimed between 20 and 50 million lives. At the time it was incorrectly believed by many that the flu was caused by a bacteria called Haemophilus influenzae. While the pandemic was actually caused by a strain of H1N1 influenza, bacterial coinfection was a major factor in pandemic mortality.\textsuperscript{112} Attempts to develop an immunization against the virus were not
successful. Other practices, such as quarantines, lockdowns, and mask wearing were likely somewhat more beneficial.\textsuperscript{113,114} The pandemic came in waves, with three or four waves between 1918 and 1920.\textsuperscript{108,115} Ascertaining the final death toll is difficult, but most estimates hover around 50 to 100 million.\textsuperscript{116-118}

Another strain of H1N1 was responsible for a more recent pandemic, the “2009 swine flu” pandemic. This virus was unusual due to the high degree of hospitalization of younger people (32\%-45\% of hospitalized patents were under 18) paired with a significant degree of older individuals bearing antibodies active against the strain.\textsuperscript{119,120} This immunity resulted from similarities between its antigenic structure and the antigenic structure of older strains of H1N1, including the 1918 virus.\textsuperscript{121,122} In all likelihood, the 2009 H1N1 made the jump into humans in the second half of 2008.\textsuperscript{123} The 2009 H1N1 was less virulent than the 1918 H1N1 but still likely claimed several hundred thousand lives by the time it ended.\textsuperscript{124,125}

Initially there was hesitation to switch over production from the usual seasonal influenza vaccine to one dedicated to addressing the pandemic since production capacity was limited and tied up with seasonal vaccine production.\textsuperscript{126,127} Growth of the virus was also slow. The mass roll-out of a vaccine in the US did not begin until after the peak of the pandemic had passed.\textsuperscript{127,128} Which is not to say the vaccine did not help; there could easily have been more peaks that were avoided. However, the episode does illustrate the deficiencies in existing influenza response systems. Vaccines exist but they are slow to prepare and slow to deploy. The existing infrastructure is well-suited to producing an annual vaccine but is not necessarily set-up to respond quickly enough to sudden changes. Even if production were of no issue, and vaccine were available immediately, reaching herd immunity takes time, and vaccine hesitancy can be an issue even if sufficient doses are made available swiftly.
4.1.3 Current vaccines

There are three types of vaccine currently available. The most widely used is the inactivated vaccine. This is the standard “flu shot” that most people receive. The cost of these vaccines is very low. They generally made by cultivating virus in chicken eggs followed by chemical sterilization and, in some cases, processing to remove unneeded components of the virus. A major limitation of this vaccine is noted low duration of immunity, with resistance falling in half after only a few months.\textsuperscript{129}

Another vaccine that is available is the live attenuated vaccine, which is recognizable as the nasal spray version. This vaccine involves the production of a less virulent strain through reassortment of the target virus with a cold-adapted strain. These cold-adapted strains grow only poorly at body temperature, naturally limiting the scope of the infection. Much like with the classic vaccinia-based smallpox vaccine, infection by the strain does more good than harm because it causes only very mild symptoms in most recipients and confers resistance to more virulent strains. Young patients see better efficacy on average with this live attenuated vaccine, likely due in part to the generation of IgA as well as IgG antibodies.\textsuperscript{130,131} But overall efficacy is more limited in older patients due to the need of the virus to generate an infection and this may not happen in patients with an experienced immune system.\textsuperscript{132}

Finally, there is a recombinant vaccine. This vaccine is built on the baculovirus insect virus, modified to express influenza hemagglutinin. This virus infects insects, and indeed cultivation occurs in insect cell lines.\textsuperscript{133} The main advantage of this vaccine is quick time to production for new strains. A benefit of a recombinant platform is that consensus or predicted hemagglutinin could be used in principle to yield a more comprehensive target profile.\textsuperscript{134,135}
4.1.4 Stalk-targeting antibodies

Ideally, a vaccine would target all strains concurrently, even strains that had yet to arise. To accomplish such a thing, a vaccine would need to generate antibodies active against one or more conserved regions on a surface viral glycoprotein. If the region is conserved because the functionality is essential and will be broken by changes sufficient to allow immune escape, then it is plausible that evolving a breakout strain may prove difficult.

A number of antibodies have been identified which have broad cross-strain neutralization. The implications are two-fold. First, the existence of antibodies targeting influenza genera very broadly suggests a monoclonal antibody treatment against novel strains could developed as a general therapeutic. Thus in the event of a pandemic mortality could conceivably be mitigated to some degree assuming production were up to the task. More relevantly, it shows that it should be possible in principle to develop a vaccine able to generate antibodies responsive to both current and future strains.

4.1.5 A stalk-targeting vaccine

To address the threat posed by the influenza A family as a whole, both existing and hopefully future strains as well, it would be ideal to develop a vaccine that can train the immune system to recognize a highly conserved epitope. Unfortunately, the areas on influenza that are conserved, such as the stem region of hemagglutinin (HA), turn out to be remarkably resistant to the generation of antibodies. To overcome this limitation, a conserved section of HA stem bearing an N-linked glycan was identified as a target by a collaborator, Dr. Ming Luo, and was prepared in isolation and conjugated to an immunogenic carrier protein.
4.1.6 A conserved epitope

The HA stem area around Asn154 is well-conserved. Going from H1 to H3 only results in a change of a single methyl.

The conserved epitope to be targeted in this project is shown in Figure 33. Between H1 and H3 there is only a valine-to-isoleucine difference. This is a modest change, with only a hydrogen substituted for a methyl, and it is entirely plausible that antibodies for one would recognize the other. In any case, the fact that the region does not change by a great degree, and the one place it does is a mutation from one mid-sized hydrophobic residue to another, is a promising sign. The limited change suggests that these residues are needed for the protein to function in a fashion consistent with the evolutionary success of the virus.

The exact density of epitopes on the carrier needed to see a good immune response is unknown. More copies should result in a more robust response. For the initial study, three repeats per glycopeptide was chosen to increase density over single repeats. It may well be the case that fewer repeats is acceptable. Fewer repeats would be less challenging to synthesize at scale, possibly allowing for more equivalents to be used during the conjugation. More copies, extending to four or five would directly increase the loading above the theoretical limit.
4.2 Experimental

4.2.1 The target

The target is an N-linked glycopeptide. N-glycans have a consistent core structure, beginning with GlcNAc (β-1,4) GlcNAc followed by three mannose in a branching structure. However, since there is little benefit in generating antibodies against the N-glycan core, minimizing the size of the glycan should be prudent.

Generally a glycopeptide or other epitope is unlikely to yield much of a response if injected in isolation. But joined to a carrier protein, robust generation of antibodies can be observed. This is the underlying principle of bioconjugate vaccines. The peptide to be loaded must be designed, a protein chosen, and a linkage strategy devised.

Figure 34. The initial target
It was decided for the first test to go for a glycopeptide consisting of three copies of the epitope in a single peptide chain. Additionally, a cystine was included to facilitate linkage. The glycan is N-linked, and so in its natural form will continue with another GlcNAc and a branched mannose structure, however this structure is not helpful. Antibodies need to be generated against the conserved amino acid sequence, not the common N-glycan structure.

Three copies of the epitope were stung together to define the sequence of the peptide. The most obvious reason to do this is to increase the number of copies of the epitope. Ten conjugated peptides would result in 30 copies of the epitope instead of only ten. The outer copies are also farther away from the unnatural linker and the protein, which may well confer benefits. One could speculate that if the linker proves too immunogenic then having more peptide relative to linker and
having it spaced farther out would result in more antibody generation against the desired sections. Here the decision was made to directly join one epitope to the next. Antibodies may well be generated against the uninteresting set of amino acids centered on the boundary between epitopes and including some sort of linker is an option. Some sort of unnatural linker like a click-based approach would ease synthesis dramatically, since small peptides could be made and strung together quickly rather than being made less efficiently in one string on the synthesizer. In any case, the linkage between epitopes, whether it consists of amino acids or a synthetic linker, will not be representative of anything we would wish to generate antibodies against. There may be merit in exploring other options for a follow-up project. No optimization in this area has been performed. It may well be the case that a smaller peptide links more efficiently to the protein, resulting ultimately in more copies of the repeat over a longer one. A larger peptide will have a more pronounced secondary structure that may or may not be conducive to conjugation. Further, there could be crowding as well around some of the conjugation positions making the linkage of several larger peptides more difficult.

The carrier protein chosen is CRM197. CRM197 is the diphtheria toxin with a G52E mutation. This single point mutation is enough to attenuate the toxicity to the point where it is safe to use as a carrier. It is well-suited to this purpose as evidenced by the fact that there are several FDA approved vaccines built on the platform, including Pfizer’s Pneumococcal pneumonia vaccine, Prevnar and GlaxoSmithKline’s Meningitis vaccine, Menveo.

CRM197 has 41 primary amines. Any of these amines is suitable for conjugation in principle. In practice, many may be located such that they are not readily accessible for conjugation. CRM197 can undergo a conformation change following conjugation. This could result in additional amines becoming accessible for conjugation so two rounds of treatment were
performed. Linkage to the carrier protein presents a number of challenges. The linkage strategy chosen for the initial study was to preload the carrier with a maleimide linker, prepare the peptide with an extra cysteine, then link the peptide through the cysteine to the maleimide. There are other linkage options. Reductive amination is a common choice, especially for the linkage of glycans as the aldehyde form is well suited for coupling. BAANS is also popular. Since the peptide was designed and made in-house, it was a simple matter to include a cysteine. Cysteine-maleimide couplings are popular for bioconjugation. They are known to be efficient, click-like reactions.

4.2.2 Glycopeptide preparation

![Synthetic scheme for the bioconjugate vaccine](image)

**Figure 35. Synthetic scheme for the bioconjugate vaccine**

Preparation is by SPPS. The plan of synthesis of the glycopeptide follows what was seen with the array peptides. After SPPS using building blocks including GlcNAc-Asn, the glycans are deprotected in hydrazine. The protein, prepared in e. coli, is preloaded with linker and the peptide is then conjugated to the linker.

The preparation of the conjugate began with the synthesis of the glycopeptide. To prepare the peptide a number of starting materials are required. These materials, bar N-GlcNAc asparagine,
which was prepared in-house by Harmon Greenway, were purchased from commercial suppliers. Solid phase synthesis was performed on a Liberty Blue automated peptide synthesizer. After cleavage from the solid resin the peptide was treated with 5% hydrazine to remove the protecting groups on the glycans. Material was directly injected into an Agilent semi-prep HPLC with a C18 semi prep column.

As is sometimes the case with longer peptides bearing multiple PTMs, crude yield was not high. Although the correct mass dominated the MALDI, there was no dominant peak in the HPLC and more than one peak and the desired mass, suggesting racemization. This is reasonable, as cysteine is known to be particularly susceptible to racemization. After repeat passes a sample of highly pure material was obtained.

4.2.3 Bioconjugation

Before conjugation could take place conditions first had to be optimized. The peptide was precious, with only a small amount being made. To simulate the peptide, Ac-Cys-OH was chosen as surrogate. While Ac-Cys-OH is much smaller than the peptide it has a free thiol and is large enough to see a shift in the protein mass by MALDI. Fmoc-Cys-OH would have been even better from a mass standpoint but its solubility in aqueous solutions is more limited and there was concern that linking several dozen Fmoc groups to the protein would be deleterious in more ways than one.

Although CRM197 has 41 primary amines, at no point in the study was the loading of the linker pushed past 22. Even increasing the number of equivalents by an order of magnitude and adding another treatment cycle failed to push the number higher. After suitable conditions were found the target was conjugated to the CRM197.
4.2.4 Sample Preparation

Figure 36. The conjugate used in mouse studies
1: CRM197, 2: CRM+linker, 3-7: CRM+linker and 3-7 peptides. First the protein is loaded with linker. In this case a shift corresponding to an average loading of 19 linkers is seen. The peak is broader. The linker mass is too low for these to resolve as different peaks. Next loading of the peptide onto the protein. Here, the peptide is large enough to cause the signal to split cleanly into different peaks based on loading number. The first visible peak here represents 3 peptides loaded onto CRM. The second 4 and so on. Based on the intensities the average should be roughly 5 peptides per protein, or 15 copies of the epitope per protein.

To 250 ul of a CRM197 solution (2.0 mg/mL, PBS) in a 1.5 ml Eppendorf tube was added 150 ul of a Sulfo-EMCS linker solution (1.74 mg/mL, PBS). System was gently rocked at room temperature for one hour and spun down to 100 ul in a 30 kDa membrane spin column. Additional 1x PBS (300 ul) was added to the column and the system was spun down to 100 ul again. A further 300 ul of 1x PBS was added to the column and the system was spun down to 100 ul again. 90 ul was set aside on ice, and the residue was diluted to 400 ul with HPLC-grade water before being spun down to 20 ul. The resulting solution should have had a concentration of 42 pmol/ul based on starting concentration and this material was used to run a MALDI. The product gained an average of 3689.7 m/z, indicating the addition of 19.1 linkers (each increases the mass by 193).
To the 90 ul set aside was added 300 ul of a glycopeptide peptide stock solution (1.92 mg/mL, PBS) and 10 ul of the PBS buffer to maintain consistent volume. System was gently rocked at RT for one hour and spun down to 100 ul in a 30 kDa membrane spin column. 0.1x PBS (300 ul) was added to the column and the system was spun down to 100 ul again. A further 300 ul of 0.1x PBS was added to the column and the system was spun down to 100 ul again. The solution should have had a concentration of 75 pmol/ul based on starting concentration. One uL was drawn off and used to run a MALDI. The solution was then diluted to 500 ul with 0.1x PBS and evenly portioned into 5 Eppendorf tubes and frozen. Theoretical yield is 7.2 nMol, or 0.58 mg, assuming an average loading of 5 glycopeptides and 19.1 linkers per protein. Actual yield was found to be 126 ug (22%).

The loadings ranging from three-to-seven could be seen cleanly as separate peaks in the MALDI-TOF. If we assume that the different loadings do not appreciably alter the odds of being detected we can roughly estimate loading, which should be no less than five. This represents a modest loading of approximately fifteen copies of the epitope per protein in average. Ideally, a greater loading of glycopeptide would have been obtained. However, since the needed amount is not known, the decision was made to proceed. A mouse study was conducted with the compound as prepared above.

4.2.5 Initial mouse study structure

Fifteen mice, female, BALB/c, were divided into three equal cohorts. The first cohort was dosed with PBS. The second with two doses of 5 ug of vaccine spaced two weeks apart. The third received three doses of vaccine, the first two doses being 5 ug and the final being only ug per mouse due to the limited size of the batch of vaccine. Two weeks after the final dose serum was drawn and the mice were injected with 10x the LD50 of H1N1-PR08. Weight and survival were tracked.
Blue bars are glycopeptide coated plates. Red bars are virus coated plates. This assay was performed by Ross Terrell. Serum samples from control mice (“PBS”) as well as from the two vaccinated groups was run against the plates. Here the three dose mice are labeled as “vaccinated.” The plates were then treated with anti-mouse IgG, then TMB, and finally sulfuric acid before being read. Serum from mice that had previously survived infection (“Infected”) was included as well. Mice that were dosed with three rounds of the vaccine generated a robust antibody response to the glycopeptide, with nearly a thousand-fold signal increase over the serum from mice dosed with PBS. The mice that had survived infection did not appear to generate antibodies against the sequence represented by the glycopeptide. The vaccinated mice generated a roughly two-fold increase vs the PBS mice against the virus, suggesting that there was at least some degree of binding of antipeptide antibodies against the actual virus.

An ELISA was performed on the sera samples. Much like a microarray ELISA involves the immobilization of target materials against which antibody-bearing sample is analyzed. The ELISA was performed using glycopeptide-coated plates and virus coated plates. Serum samples from control mice (“PBS” in figure) as well as from the two vaccinated groups was run against the plates. The plates were then treated with anti-mouse IgG, then TMB, and finally sulfuric acid before being read. Serum from mice that had previously survived infection (“Infected” in figure) was included as well. Mice that were dosed with three rounds of the vaccine (“Vaccinated” in figure) generated a robust antibody response to the glycopeptide, with nearly a thousand-fold signal increase over the serum from mice dosed with PBS. The mice that had survived infection
did not appear to generate antibodies against the sequence represented by the glycopeptide. The vaccinated mice generated a roughly two-fold increase vs the PBS mice against the virus, suggesting that there was at least some degree of binding of antipeptide antibodies against the actual virus.

Only survival data from the two dose mice is available. Mice were euthanized after the loss of 25% of their initial body mass. Eleven days after being infected with the virus one mouse had been euthanized from the vaccine group and three mice had been euthanized from the control group.

![Figure 38. Mouse weight and survival data](image)

*GC mice are two dose group. PBS mice are the control group. Four of five two dose mice survived the challenge. Two of five control mice survived the challenge.*

4.3 Discussion

A conserved, eight residue N-glycan bearing epitope in the influenza A virus hemagglutinin was prepared in the form of a three-repeat glycopeptide using SPPS. For the initial test the epitope was repeated three times in a single peptide chain. This chain included an additional cysteine at the c-terminus to facilitate linkage. A carrier protein, CRM197, was successfully prepared in sufficient quantities for conjugation. Conditions were developed suitable
for the loading of ~20 linkers. The glycopeptide was successfully conjugated to the carrier protein at moderate (~5x) loading numbers. The ELISA results show that antibodies were generated against the glycopeptide, and these antibodies seem to have had some reactivity toward the virus as well. The results of the mouse studies were promising, but ultimately inconclusive for statistical reasons. Two groups of five are too small to draw meaningful conclusions from unless the difference is between the two groups is stark. Here, the number of surviving vaccinated mice is only two individuals greater than in the control group.

The project represents a meaningful attempt to develop a vaccine targeting a conserved epitope. While the results are suggestive, they are not significant. The biggest hurdle was the availability of peptide for conjugation. Synthesis of the glycopeptide target used here was not efficient enough to make possible large-scale studies with significant sized groups of mice.

Early follow up work not included here was attempted. This mostly centered on attempting to prepare the epitope in the form of blocks that could be joined one after another to form sequences of arbitrary length. While an interesting approach, the formation of amide bonds between large molecules is often of mediocre yield in the absence of astronomical excesses of reagent. This is a reason why native chemical ligation is usually favored for the chemical synthesis of proteins. A stronger option, if a block-based path were pursued, would be a click linkage of some sort between the blocks. An option might be to use two orthogonal click reactions and prepare two blocks each with alternating halves of the click linkage. In this way it would be possible to quickly build up polymers of arbitrary length directly on the protein through alternating treatment by one block after the other. If click linages are unsuitable, one can envision native chemical ligation-based approaches such as the use of a ros-cleavable thiodiglycolic anhydride-based protection strategy.
on the amine to allow for discrete coupling and deprotection steps with minimal damage to the protein.

In any case, the results are suggestive enough to merit a follow-up project long the same lines. The choice of CRM197 was the right one and there is no reason to expect that a move to an alternative carrier would be sensible. Further, the core principle of targeting a conserved epitope is sound. Between the damage done every year by the seasonal flu and the ever-looming specter of further pandemics, the need for a broad-spectrum preventative like a universal vaccine remains great.
REFERENCES


(34) Zhang, Q.; Li, Z.; Song, X. Preparation of Complex Glycans From Natural Sources for Functional Study. Frontiers in Chemistry 2020, 8 (508).


(41) Ralhan, K.; KrishnaKumar, V. G.; Gupta, S. Piperazine and DBU: a safer alternative for rapid and efficient Fmoc deprotection in solid phase peptide synthesis. Rsc Advances 2015, 5 (126), 104417.


Kessler, S.; Harder, T. C.; Schwemmle, M.; Ciminski, K. Influenza A Viruses and Zoonotic Events—Are We Creating Our Own Reservoirs? *Viruses* 2021, 13 (11), 2250.


Asha, K.; Kumar, B. Emerging Influenza D Virus Threat: What We Know so Far! *Journal of Clinical Medicine* 2019, 8 (2).


(93) Sederdahl, B. K.; Williams, J. V. Epidemiology and Clinical Characteristics of Influenza C Virus. Viruses 2020, 12 (1).


cells that binds to group 1 and group 2 influenza A hemagglutinins. Science 2011, 333 (6044), 850.


(150) Sabatino, D. Medicinal Chemistry and Methodological Advances in the Development of Peptide-Based Vaccines. Journal of Medicinal Chemistry 2020, 63 (23), 14184.


(152) Hall, D. J.; Van Den Berghe, H. M.; Dove, A. P. Synthesis and post-polymerization modification of maleimide-containing polymers by ‘thiol-ene’ click and Diels–Alder chemistries. Polymer International 2011, 60 (8), 1149.