Automated Flow Peptide Synthesis

by

Alexander James Mijalis

B.S., Texas AM University (2012)

Submitted to the Department of Chemistry
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Chemistry

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2018

Signature redacted

Author .....
Department of Chemistry
May 11, 2018

Signature redacted

Certified by .....
Bradley L. Pentelute
Associate Professor
Thesis Supervisor

Signature redacted

Accepted by .....
Professor Robert W. Field
Chairman, Department Committee on Graduate Theses

This doctoral thesis has been examined by a Committee of the Department of Chemistry as follows:

Signature redacted

Professor Timothy F. Jamison ....

Chairman, Thesis Committee
Department Head and Robert R. Taylor Professor of Chemistry

Signature redacted

Professor Bradley L. Pentelute .....

Thesis Supervisor
Associate Professor of Chemistry

Signature redacted

Professor Alexander M. Klibanov ...

Member, Thesis Committee
Novartis Professor of Biological Engineering and Chemistry

Automated Flow Peptide Synthesis

by

Alexander James Mijalis

Submitted to the Department of Chemistry
on June 11, 2018, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Chemistry

Abstract

Though reported by Merrifield nearly sixty years ago, batch solid phase peptide synthesis remains slow at minutes to hours per residue. Here we report a fully automated, flow based approach to solid phase polypeptide synthesis with amide bond formation in seven seconds and total synthesis times of forty seconds per amino acid residue. Crude peptide purities and isolated yields were comparable to standard batch solid phase peptide synthesis. Process monitoring with absorbance spectroscopy allows for the immediate detection and rapid optimization of difficult-to-synthesize peptides. This instrument is flexible and allows for synthesis of peptide nucleic acids, glycopeptides, removal of orthogonal amine protecting groups, and click chemistry on the solid phase. At full capacity, this approach to peptide synthesis can yield tens of thousands of individual 30-mer peptides per year.

Thesis Supervisor: Bradley L. Pentelute
Title: Associate Professor

Acknowledgments

It is difficult to understate the contributions of my friends, family, labmates, mentors and collaborators toward finishing my graduate work here at MIT.

First, I am grateful to my advisor, Dr. Bradley Pentelute, for giving me the freedom to pursue this project when I joined the lab and for his help along the way, particularly in honing my presentation skills. I am also indebted to Dale Thomas, whose close collaboration resulted in the peptide synthesizer that is the subject of this thesis, and Mark Simon, who helped train me when I first arrived in the Pentelute lab, and with whom I worked to develop the earliest versions of the Automated Flow Peptide Synthesizer.

I am grateful for the opportunity I've had to work with such incredible colleagues in the Pentelute group, particularly Justin Wolfe, Ethan Evans, Colin Fadzen, Rebecca Holden, Nina Hartrampf, Alex Vinogradov, Amy Rabideau, Surin Mong, Dan Cohen, Dan Dunkelmann, Zak Gates, Kathy Sweeney and Rachael Fuller. During my fourth and fifth years in graduate school, I was most fortunate to mentor Vanessa Züger from ETH and Akihiro Manbo from Osaka University; I am grateful for both their patience and their enthusiasm about their projects during their time here.

A special shout-out goes to my friends Angela Phillips, Kenny Kang, Chet Berman, Wade Wang, Justin Wolfe, Colin Fadzen, Ethan Evans and Yisu Han for their encouragement and the memories we made over the past five years. Second- and third-year exams were made much more bearable by our collective struggle.

For their words of wisdom over many hours of phone conversations, I want to thank my parents Angela and Jimmy Mijalis, my sister Eleni Mijalis, and my friend Dr. Taylor Smith, M.D. I am also grateful to my mentors Judi Segall, Prof. Christopher Kevil, Prof. David Bergbreiter and Prof. Klavs Jensen for their advice along the way, both science and life related. Many thanks go to my Thesis Committee Chair, Prof. Tim Jamison, and Thesis Committee Member, Prof. Alex Klibanov for their helpful feedback on my thesis.

I also want to thank Prof. Nicola Pohl for her guidance and a fruitful collaboration

during her Radcliffe Fellowship, along with our mutual collaborator Benjamin Lee, a computer science student at Harvard, with whom we have been working to develop the open-source flow chemistry Python package MechWolf.

Finally, I want to thank the National Science Foundation's Graduate Research Fellowship Program for funding three years of my graduate work, and Dr. Reddy's laboratories and the Deshpande Center for their funding and support while developing the peptide synthesizer.

Contents

List of Figures

List of Tables

Chapter 1

Pushing the Limits of Solid Phase Peptide Synthesis with Continuous Flow

This chapter was adapted from the publication "Pushing the Limits of Solid Phase Peptide Synthesis with Continuous Flow" with the following authors:

Alexander J. Mijalis[1], Angela Steinauer[2], Alanna Schepartz[2], Bradley L. Pentelute[1]

1: Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

2: Departments of Chemistry and Molecular, Cellular and Developmental Biology, Yale University, New Haven, Connecticut 06520-8107, USA

1.1 Introduction

Synthetic polypeptides are omnipresent tools in biological chemistry. Peptides can be used as drugs to treat conditions such as diabetes, chronic pain, and cancer. They also play a role in delivering other therapeutic molecules, helping target those molecules to certain cellular compartments. Peptides can serve as affinity tags and functional handles for performing selective chemistry on biomolecules under mild conditions. These applications are possible in part due to the large sequence space and

concomitant structural and functional landscape accessible to folded polypeptides; however, due to difficulties with the chemistry, accessing a peptide of arbitrary length and sequence cannot be achieved with complete generality.

While sequence diversity is a challenge for peptide synthesis, the chemical manipulations are highly repetitive, and automation tools can save a significant amount of manual labor. Merrifield himself recognized this advantage early on, employing automation in the form of a wheel that operated much like that of a music box, dispensing reagents as it rotated and its teeth contacted levers on reservoirs. Modern liquid handling robots, operated by microcontrollers, can perform batch peptide synthesis in parallel, reducing the task of synthesis to typing in a sequence and filling up bottles. When these systems work well, they generate the desired polypeptide in high yield and purity, as indicated by a single peak in a reverse-phase HPLC chromatogram.

However, Merrifield noted that "While the mechanical and electronic capability for automated solid-phase peptide synthesis is at hand, it must be realized that the limiting factor has been and continues to be the chemistry of the process.... Until all of the assorted chemical difficulties are under control, automation cannot reach its full potential." [1] Indeed, while many peptide sequences are straightforward to synthesize with batch solid phase methods, others fail. [2] These so-called "difficult peptides" have led to a proliferation of specialized carboxylic acid activating agents, solvent systems, solid supports, and synthesis methods. Additionally, because the objective of peptide synthesis is almost always to obtain a product for another use, these problems are often solved with brute force techniques such as extreme heat and multiple couplings, leading to syntheses that are difficult to reproduce and scale.

Continuous flow peptide synthesis methods have emerged from a confluence of advancements in automation technologies, polymer chemistry, and flow chemistry tools. These methods offer significant advantages compared to batch methods, including time-resolved spectroscopic monitoring, greater efficiency of washing solid supports under continuous flow, and the ability to strategically apply heat and handle reactive intermediates on short timescales. However, there is not yet a modern, comprehensive

review comparing flow-based approaches for peptide synthesis. This review will focus on the application of flow chemistry to improve the three chemical transformations in peptide synthesis: amino acid activation, aminoacylation, and protecting group removal, and also discuss some of the outstanding chemical challenges in the field of peptide chemistry.

1.2 Continuous Flow Solid-Phase Peptide Synthesis: Historical Context

In the early 1970s, several research groups realized that peptide synthesis automation would be much easier if the solid support were contained in a column and operated as a fixed bed. [3, 4] Instead of using complex liquid handling apparatus and vacuum aspiration to dispense and wash the resins, as was the case with most SPPS procedures, widely-available HPLC pumps and valves could instead be used to deliver reagents to solid supports contained in HPLC columns. Such a system could reduce the amount of moving parts required, increase the efficiency of washing, and allow for reaction monitoring with in-line absorbance spectroscopy. As proof of principle, in 1981, Erickson, et al., described a method for continuous flow solid phase peptide synthesis using traditional polystyrene solid supports and HPLC equipment. [5]

However, the solid support materials available at the time were difficult to use with this approach. Crosslinked polystyrene supports and polyamide supports used by Merrifield and others are gelatinous, polydisperse, and difficult to pump fluid through. [6] Moreover, polystyrene resins can swell dramatically as peptide synthesis progresses: Kent showed that in N,N-dimethylformamide at room temperature, polystyrene resin can swell to over 25x its dry volume as repeat units of Leu-Ala-Gly-Val are added. [7] The challenges here become chiefly mechanical: confined in a fixed-bed column, resin swelling leads to large and unpredictable backpressure spikes and can cause resin extrusion and instrument failure unless highly robust equipment is used.

Breakthroughs in flow solid phase peptide synthesis came with the development

of new solid supports. Most notably, RC Sheppard, et al. described a low-pressure continuous-flow Fmoc peptide synthesis system using "Pepsyn K", a macroporous polyamide gel supported on kieselguhr, a type of diatomaceous earth (Figure 1-1-B). [6, 8–10] Sheppard's flow synthesis platform was one of the first to use N-α-Fmoc-protected amino acids, enhancing process safety by retiring trifluoroacetic acid as a routine reagent and introducing reaction monitoring using the Fmoc group as a chromophore. Additionally, Sheppard configured the pumps and valves to recirculate expensive pre-activated amino acid anhydrides and pentafluorophenyl esters for improved efficiency. Recirculation also ensured that reagents were well-mixed and distributed over the solid support.

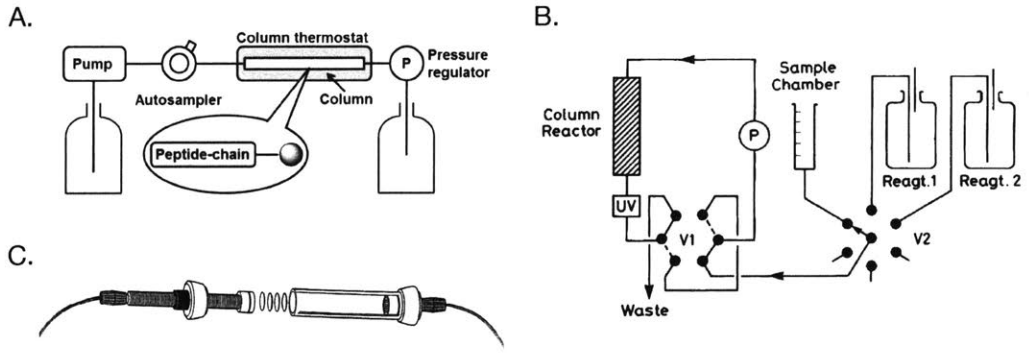

Figure 1-1: Early mechanical designs for flow synthesis of polypeptides. **A.** An HPLC-based peptide synthesizer where reagent delivery valving is replaced by an autosampler. **B.** The first example of a low-pressure flow peptide synthesis apparatus developed by Dryland and Sheppard. This device operated at room temperature, recirculating pre-formed amino acid pentafluorophenyl esters over peptidyl resin. It allowed for the UV monitoring of Fmoc removal. **C.** A modified version of the column reactor developed by Wikberg, et al., packed with cotton disks for split-pool synthesis of peptide libraries.

LKB Biochrom and Cambridge Research Biochemicals (CRB) commercialized Sheppard's flow method for solid phase peptide synthesis, releasing the Model 4175 peptide synthesizer and the PEPSYNthesizer 1, respectively. [11] Milligen acquired the rights to CRB's instrument and released the PepSynthesizer 9050, of the most popular continuous flow peptide synthesizers; however, these instruments are no longer produced. New solid supports were adapted for use in these commercial systems,

including functionalized cellulose discs for split-pool synthesis applications (Figure 1-1-C). [12, 13] In order to improve flow in a packed bed and diffusion of reagents into polymer supports, Rapp and Bayer developed TentaGel, a monodisperse, PEG-functionalized polystyrene resin. Using this technology, they demonstrated reduction of synthesis times to around 8 minutes per synthetic cycle at 45 °C. [14] Recently, Mándity, et al., described advances in flow peptide synthesis methodology using Tentagel housed in HPLC equipment at 70 °C (Figure 1-1-A). [15, 16]

We developed an alternative approach to flow-based SPPS using low-pressure flow reactors and in-situ amino acid activation (Figure 1-2). In these reactors, the solid support is loosely packed in a tube reactor or at the bottom of a syringe with extra volume for expansion. Because of high flow rates across the resin bed, turbulent flow in this head space ensures even distribution of synthesis reagents over the resin. To further reduce the backpressure, we experimented with a number of solid supports and found that ChemMatrix, a crosslinked poly(ethylene glycol), is ideal for continuous flow processes, providing minimal backpressure even at flow rates of 80 mL^{-1} min^{-1} cm^{-2}. The high swelling of ChemMatrix is favorable for peptide synthesis, but during the syntheses of long polypeptides, the resin can expand to such a large volume that it fills the reactor and causes a spike in backpressure. This can be avoided by reducing the amount of resin in the reactor. Additionally, instead of using pre-activated amino acids at room temperature, or perfoming the activation in batch before a flow coupling, the automated flow peptide synthesizer mixes amino acids with activating agents in continuous flow at 90 °C, delivering them directly to the column of pre-heated peptidyl resin. This important development results in coupling times of less than 10 seconds, and full coupling, deprotection, and wash cycles of around 40 seconds.

In a departure from solid phase methods, Livingston, et al., demonstrated "Membrane-Enhanced Polypeptide Synthesis," another clever method for separating a growing synthetic polypeptide chain from synthesis reagents (Figure 1-3-B). [17] They used a ceramic ultrafiltration membrane chamber to retain a peptide bound to a soluble, amine-functionalized polymer support, while activated amino acids, solvent, and de-

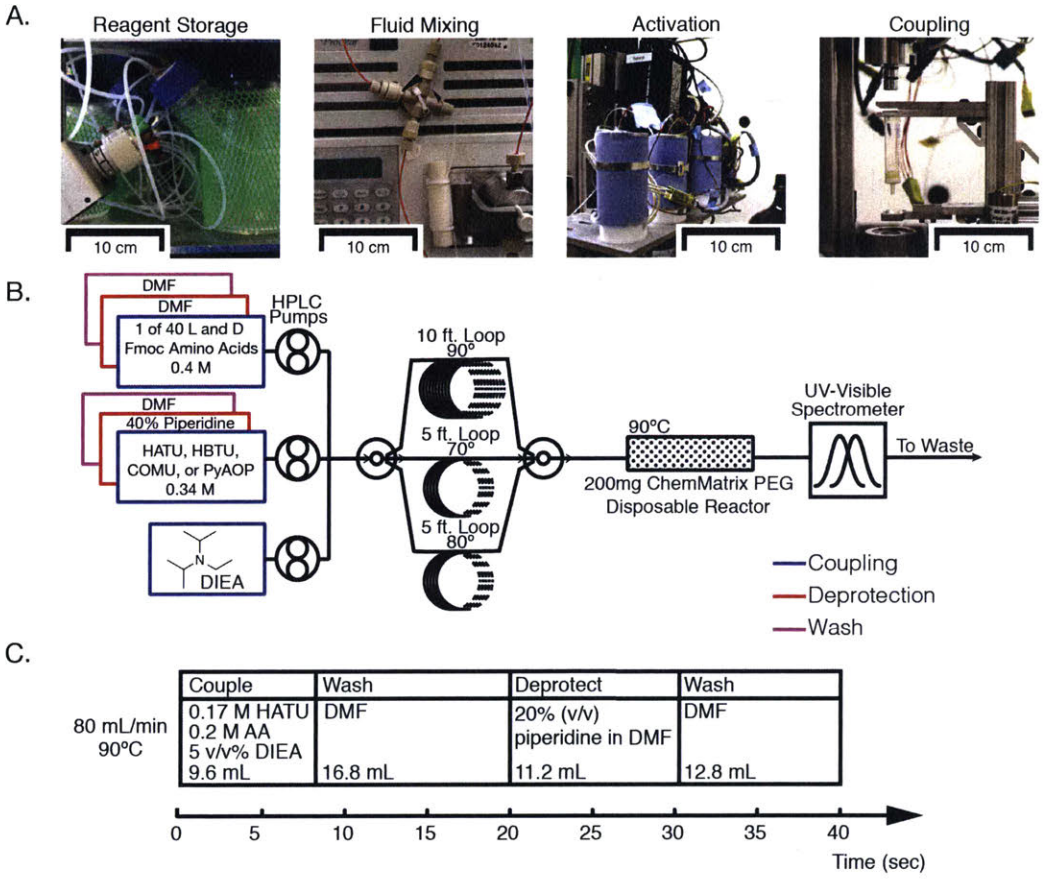

Figure 1-2: The Pentelute Lab's Automated Flow Peptide Synthesizer combines in-situ, continuous flow activation with flow monitoring of Fmoc removal. **A.** The AFPS includes modules for reagent storage, fluid handling mixing, elevated temperature activation, and elevated temperature coupling. **B.** A fluidic schematic of the AFPS. **C.** Timeline for the standard AFPS method at 90 °C with 40 second cycles for addition and deprotection of each amino acid.

protection reagents flowed freely through. This system is particularly advantageous because the final product can be transferred to downstream cleavage and purification while still in solution, facilitating automation, whereas solid supports can be difficult to handle and transfer after synthesis. Soluble polymer-bound biopolymers could also be easier to use in peptide library screening applications.

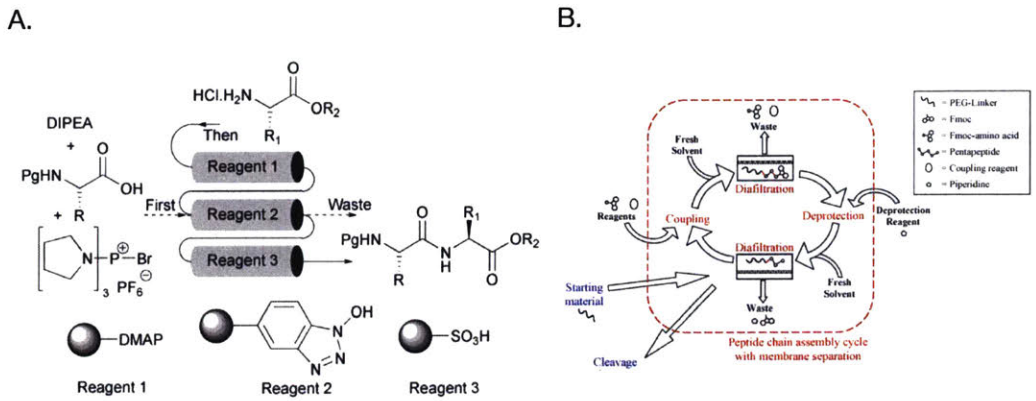

Figure 1-3: Various flow chemistry methods for polypeptide synthesis. **A.** In catch-and-release polypeptide synthesis, an activated amino acid is first immobilized on a solid substrate. In a subsequent step, a soluble amine reacts with this activated species to yield a dipeptide which is purified from solution. The system shown here was developed by developed by Ley, et al. **B.** Membrane-enhanced peptide synthesis uses a soluble, polymer-tethered amino group for chain elongation. The product is isolated from coupling and deprotection reagents by a flow diafiltration process. The Pentelute Lab's Automated Flow Peptide Synthesizer combines in-situ, continuous flow activation with flow monitoring of Fmoc removal.

1.3 Difficult Sequences: A Chemical Justification for Activation and Aminoacylation in Flow

Some polypeptides elude synthesis at room temperature. During synthesis, particular sequences of protected amino acids will become insoluble and "aggregate" on the solid support. [2, 6, 18] As aggregation occurs, the solid support shrinks as the peptide stops associating with solvent molecules and diffusion of new reagents to the N-terminus slows. The result is that both the Fmoc removal and aminoacylation steps slow down drastically. Synthesis of an aggregating peptide often yields a complex, multicomponent HPLC chromatogram with truncated products, making it unclear where optimization is required. Heat can disrupt the aggregation process by preventing hydrogen bonding and secondary structure formation on resin. In addition to preventing peptide aggregation, heating peptide synthesis speeds up the entire synthesis cycle and assists with couplings involving sterically-hindered N-alkylated amines or bulky amino acid side chains. For a process which takes anywhere from 30

minutes to hours at room temperature, heating seems ideal.

Since the early 1990s, microwave heating has been one of the most widely-used methods for heating peptide syntheses and decreasing reaction times to 1.5 – 2 minutes per cycle. [19–21] However, while the application of heat increases coupling rates, it also increases activated monomer degradation rates. In particular, certain amino acids are sensitive to loss of stereochemistry when activated, a problem that is exacerbated by microwave irradiation. [22, 23] A consideration of the mechanism of activation and amide bond formation reveals why heat is tricky to apply in batch SPPS.

Coupling of an amino acid to a polymer-supported amine comprises two distinct chemical processes: activation and aminoacylation. During amino acid activation, an N-α-amino-protected amino acid carboxylate reacts with an "activating agent" to generate an electrophilic acylating species in a reaction cascade involving one or more highly reactive intermediates. After activation has occurred, but before aminoacylation is complete, the activated amino acid can degrade in two ways. First, the activated amino acid can react to form less active species over time. In one common mechanism, activated Fmoc-amino acids in the presence of excess tertiary amine cyclize to form a 5(4H)-oxazolone which is less reactive as an acylating agent (Figure 1-4-A). [24] Second, the activated amino acid can epimerize via this 5(4H)-oxazolone intermediate, losing stereochemistry at the alpha carbon, decreasing the chiral purity of the product incorporated on the solid phase. [25–27]

After activation, the acylating species must be transported to the amino group where it reacts to form an amide bond. However, even if the aminoacylation reaction is fast, as is the case for HATU-activated amino acids at 90 °C, if the transport step is inefficient, these degradation processes become more important. As D.S. Kemp notes in his early review on racemization in peptide synthesis, "Epimerization ... depends on a ratio of rate constants and is proportional to the rates of racemization of the vulnerable species times their average lifetimes. Thus, an acylating agent that is epimerized exceedingly rapidly may lead to chirally integral product if it is generated in the presence of amine nucleophile and aminolyzes orders of magnitude more rapidly than it racemizes."[29] If the rate of amino acid degradation is augmented by heat,

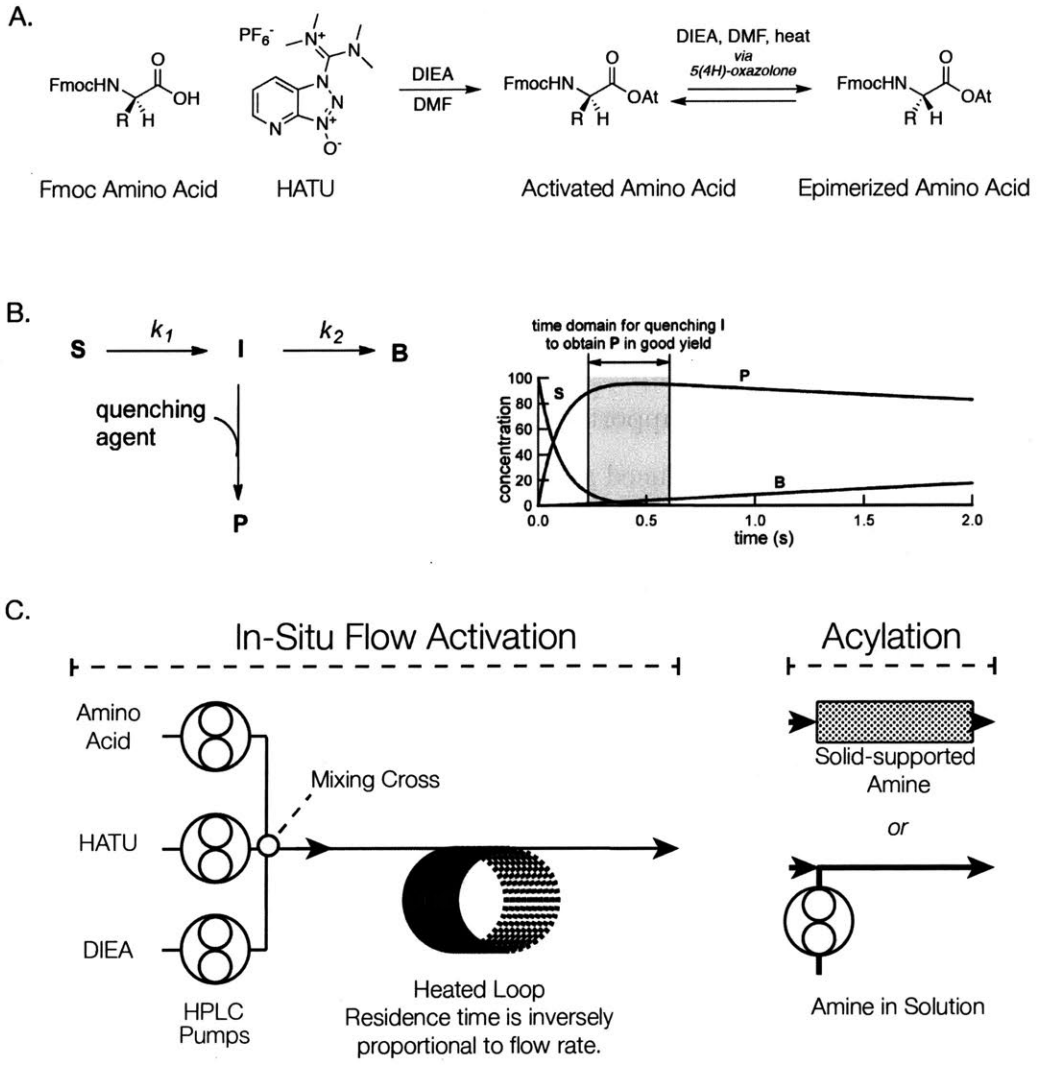

Figure 1-4: Flow activation helps control peptide synthesis at high temperature. **A.** During amino acid activation, protected amino acids react with HATU and DIEA to form an active ester. In the presence of heat and base, they degrade over time, losing stereochemistry via 5(4H)oxazolone intermediates.[28] **B.** Yoshida's generalization of the problem of handling reactive intermediates. A reactive intermediate I must be quenched after being generated from reactants but before degrading. **C.** Automated, in-situ activation of amino acids can be quenched with a solid-supported amine or an amine in solution after a pre-defined reaction time.

but the activated amino acid is not efficiently transported through the solid support, epimerized amino acid will appear in the final product, the coupling efficiency will drop, or both. Similar principles apply when more reactive acylating agents are used.

Yoshida elegantly describes the principles of handling reactive intermediates in

flow in his review on "Flash Chemistry" (Figure 1-4-B). [30] For a process where a reactive intermediate is generated and subsequently quenched, he notes that there is a finite time window during which the starting materials have been mostly consumed but the reactive intermediates have not yet degraded. The reactive intermediates should be consumed during this window. Applied to solid phase peptide synthesis under continuous flow, the amino acid activation time is selected by controlling the flow rate and the heated volume of the system between the mixing zone and the chamber containing the solid support (Figure 1-4-C). Under ideal conditions, when reagents are continuously introduced and voided from the reactor, the resin is bathed in a reagent stream where the concentration of each reactant is constant. The consequence is that activation and acylation are decoupled: activation for each amino acid can be optimized independently of any particular coupling step. Then, if a particular coupling is slow, coupling time can be extended without fear of greater epimerization by increasing the duration of the coupling step, with the caveat of increased monomer use.

We applied these principles to the in-situ activation of amino acids in our Automated Flow Peptide Synthesizer. We observed that with DMF, HATU and DIEA at 90 °C, the optimum activation time is less than 5 seconds for sensitive amino acids such as cysteine and histidine. We also observed that the extent of epimerization of the amino acid is a function of the bulk flow rate; the lower the flow rate, the more epimerization is observed (Figure 1-5). Conversely, if the flow rate is too high, unreacted HATU can remain in the fluid stream, causing truncation of the growing peptidyl chain. For comparison with other heated synthesis methods, under standard flow conditions using HATU/DIEA for activation at 90 °C, we observe ~1% Cys diastereomer incorporated into the test polypeptide GCF. Using a microwave synthesizer and the same chemistry, around 15% of the diastereomeric D-Cys GCF product is observed. Ultimately, using in-situ flow activation of amino acids with HATU with rapid delivery of the activated species to the solid support, we found that we can perform the whole peptide synthesis cycle at elevated temperature while controlling the enantiomeric purity of the final product. For difficult peptides, such as the JR 10-mer

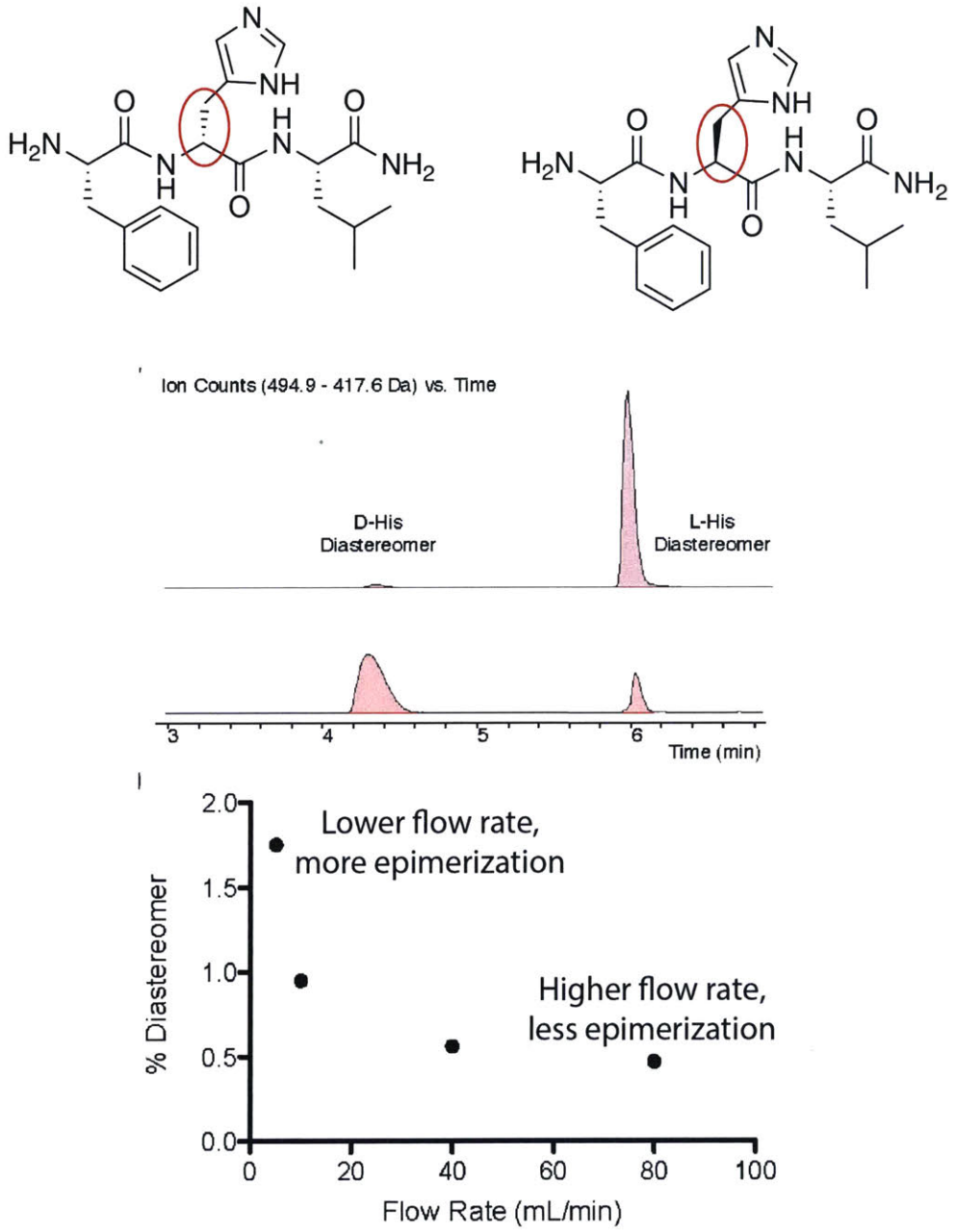

Figure 1-5: Histidine epimerization under continuous flow at elevated temperature. **A.** The two diastereomers of the model peptide FHL studied. **B.** Diastereomeric products can be chromatographically separated using HPLC. The authentic L-His and D-His variants were synthesized to confirm retention times. **C.** Histidine epimerization is a function of flow rate in the Pentelute lab's Automated Flow Peptide Synthesizer.

polypeptide, optimized flow synthesis at 90 °C substantially reduced the amount of synthesis byproducts observed compared to synthesis at room temperature (Figure

1-6).

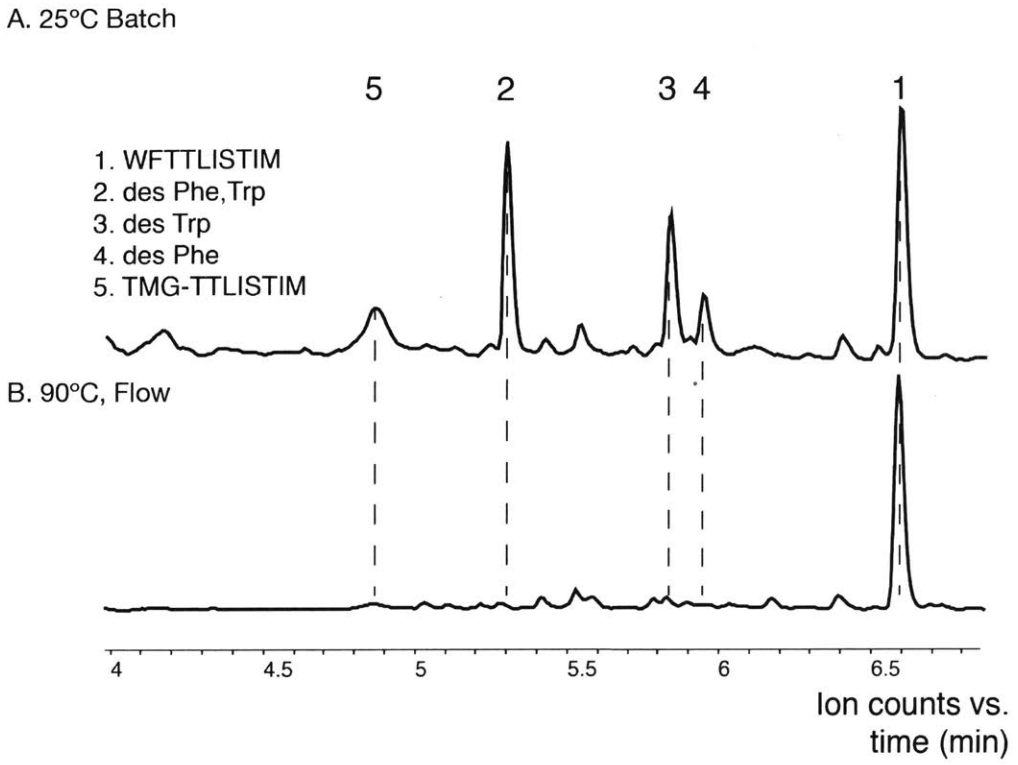

Figure 1-6: Heated flow synthesis improves the crude quality of the JR 10-mer polypeptide WFTTLISTIM-$CONH_2$. **A.** Total ion chromatogram obtained of crude JR 10-mer polypeptide synthesized in batch at room temperature. **B.** Total ion chromatogram obtained of crude JR 10-mer polypeptide synthesized in flow at 90 °C on the Pentelute Lab's Automated Flow Peptide Synthesizer.

We also investigated if heated flow peptide synthesis would be amenable for the synthesis of difficult aggregating peptides with non-standard backbone structures. We decided to attempt the synthesis of Z28, a 28-residue β_3-peptide that forms a well-defined and thermally stable tetrameric bundle in aqueous solution. [31] Z28 is an excellent starting point for the rational design of non-natural β-protein structures with enzyme-like properties. [32] This β_3-peptide is highly aggregating and difficult to synthesize even under microwave heating conditions. Under standard automated flow conditions at 90 °C, we were able to obtain the desired 28-mer in reasonably high purity in less than one minute per residue. The purified Z28 β_3-peptide exhibits concentration-dependent circular dichroism at 204 nm, consistent with the oligomer-

ization observed previously (Figure 1-7).

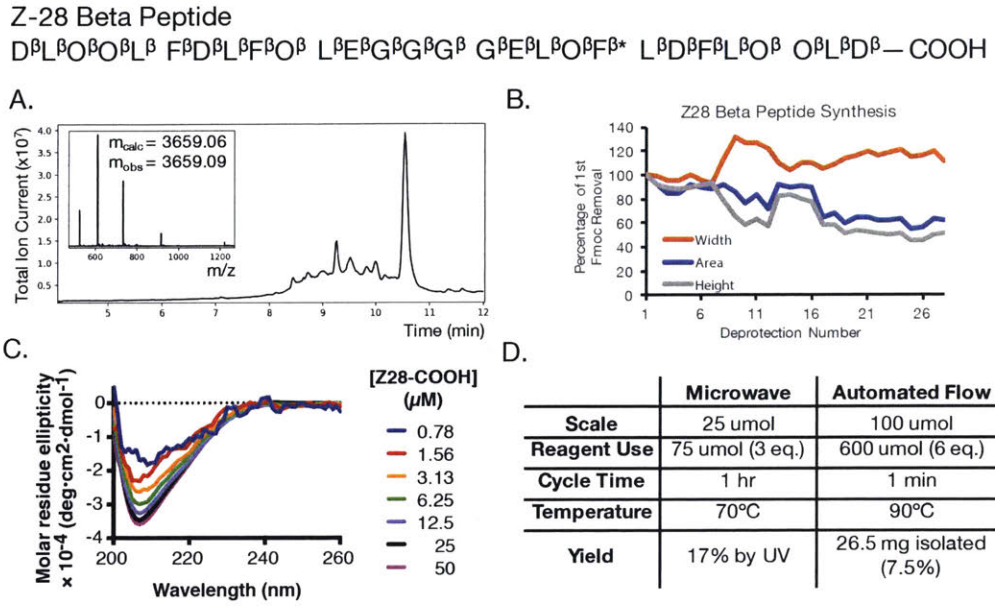

Figure 1-7: Automated flow synthesis of Z28, a bundle-forming beta peptide. **A.** Total ion chromatogram of crude Z-28 polypeptide after flow synthesis at 90 °C. **B.** Z-28 exhibits concentration-dependent circular dichroism at 204 nm. **C.** Comparison of the microwave and automated flow methods used to make Z28.

1.4 Improving Atom Economy of Peptide Synthesis with Flow Amino Acid Activation

Flow amino acid activation also has the potential to improve the atom economy of amino acid activation by allowing the routine use of amino acid chlorides as acylating agents. In the late '80s and '90s, Carpino and Benoiton explored the use of Fmoc-amino acid chlorides as acylating species for peptide synthesis. [33, 34] One motivation behind this research was that Fmoc-amino acid chlorides can be prepared cheaply and with much higher atom economy than active esters, using commodity chemicals such as thionyl chloride. Unfortunately, however, Fmoc-amino acid chlorides were found to react rapidly in solutions containing tertiary amines to yield the corresponding fluorenylmethyl oxazolones, a much less reactive acylating species that

is also prone to epimerization. [24] This undesired degradation pathway severely limited coupling yields and peptide purities when using pre-formed acid chlorides.

Although pre-formed acid chlorides are difficult to use in SPPS, acid chlorides generated in-situ yield much better results. Falb, et al., reported that bis-(trichloromethyl) carbonate, or triphosgene, generates activated amino acid species, presumably acid chlorides, in THF and in the presence of 2,4,6-trimethylcollidine. [35] Falb's method did lead to significant epimerization; however, it suggested that acid chlorides might be useful acylating agents if handled under the right conditions.

By applying flow chemistry methods to this triphosgene chemistry, the Takahashi group demonstrated a practical way to use acid chlorides in liquid phase peptide synthesis applications (Figure 1-8). [36] In this work, Fmoc-protected amino acids were activated using triphosgene and DIEA in a flow reactor with a residence time of 0.5s to form amino acid chlorides. The acid chlorides were subsequently quenched with a carboxy protected amino acid in a reactor with a residence time of less than 5 seconds. After optimization, the Fmoc-benzyl-protected Ser-Phe dipeptide was produced in 92% yield along with only 1% of the undesired diastereomer. Impressively, Fmoc-His(Trt)-OH, a monomer which is sensitive to epimerization, was activated using triphosgene in flow, yielding the dipeptide with only 2% of the undesired diastereomer in the final product. Safety considerations with triphosgene use aside, one can imagine triphosgene flow activation being used upstream of a solid support, instead of high molecular weight reagents such as HATU and BOP, for highly atom economical and rapid solid phase peptide synthesis.

1.5 Monitoring Fmoc removal in Flow SPPS

The most important innovation of Sheppard's continuous flow synthesis platform was using UV absorption spectroscopy to monitor the Fmoc removal process. [6] Though it was known that Fmoc removal could quantitated in a cuvette by UV absorption spectroscopy of the resultant dibenzofulvene/piperidine adduct, Sheppard's system added time resolution to this data. The shape of the Fmoc removal signal suggests

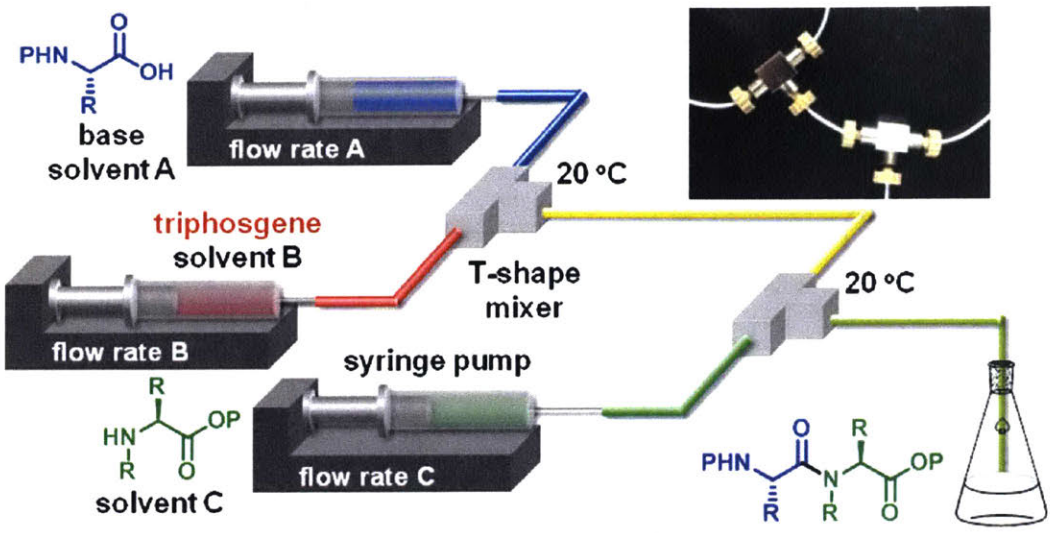

Figure 1-8: Solution-based amide bond formation in flow, developed by Takahashi, et al., uses triphosgene as an activating agent. Though this process would be expected to cause epimerization of the starting materials in batch, low residence times and heat transfer in the flow microreactor minimizes this side reaction pathway and provides acceptable yields.

the aggregation state of the polypeptide on resin. A tall, sharp peak suggests that the polypeptide N-terminus is well-solvated and that diffusion of dibenzofulvene away from the polypeptide is fast. By contrast, a low, broad curve indicates slow diffusion away from the N-terminus. Integration of the curve generates data similar to batch measurements: comparing the area under an Fmoc removal curve to the previous curve indicates the relative yield of the acylation reaction. Because the first acylation reaction is almost always quantitative, this data can be used to determine if a synthesis was successful, and if not, when it began to fail.

Because of the large sequence space for polypeptides (20^n for an n-residue polypeptide with canonical amino acids), aggregating peptides are almost impossible to predict from the peptide's primary sequence. Absent this advance warning, UV monitoring tools let the chemist know immediately if and at which reaction step aggregation occurs, far before time-consuming peptide cleavage and chromatography steps. As an example, the Fmoc removal curves for the polypeptide $Ala_6Asn_2Ala_3Val$ suggest that aggregation occurs after addition of the 10th residue (Figure 4C). [6] This result, originally reported by Sheppard, et al., was corroborated on our instrumentation using

a PEG solid support and in-situ amino acid activation with HATU at 90 °C (Figure 1-9).

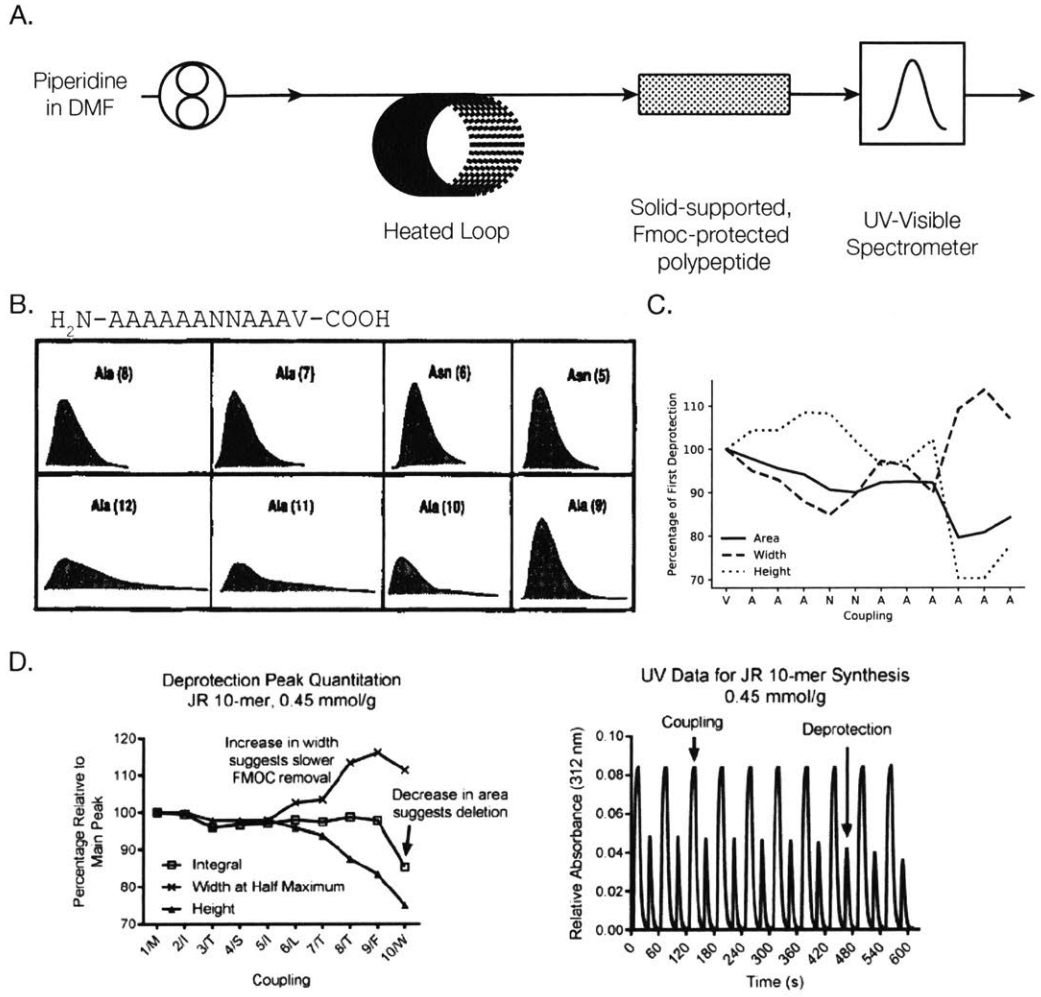

Figure 1-9: Time-resolved monitoring of Fmoc removal in flow. **A.** The general flow schematic for removal of Fmoc on a solid-supported polymer. **B.** Fmoc removal traces from the flow peptide synthesizer of Sheppard, et al. **C.** The aggregating Sheppard polypeptide was synthesized on our automated flow peptide synthesizer at 90 °C. Similar peak shape profiles were observed. **D.** Fmoc removal trace for optimization of the JR 10-mer polypeptide sequence. At elevated temperature, peak width increases and peak height decreases starting at the sixth deprotection, but it isn't until the tenth residue that aggregation prevents coupling.

In-line UV monitoring has become an indispensable tool on our Automated Flow Peptide Synthesizer. We use it as a way to optimize wash times and ensure process reproducibility – if a reagent fails to be infused or a valve malfunctions, this is immediately apparent in the UV trace. We have also used the Fmoc removal signals

to quantify and optimize deletion products in polypeptides without relying on HPLC analysis. The JR 10-mer is an aggregating peptide whose synthesis is improved at elevated temperature, but even at 90 °C, synthesis results in the deletion of the final Trp residue. By comparing the final Trp deprotection signal to the penultimate signal for a series of syntheses using resins with reduced amine substitution, we were able to determine an optimal resin substitution that does not result in this deletion (Figure 9D). Fmoc removal data is also useful in determining if low levels of chain termination or cleavage from the solid support are happening at every coupling.

1.6 Side Reactions of Resin-Bound Peptides at Elevated Temperature

One key consideration when performing solid phase chemistry at elevated temperature is that the resin-bound polymer will remain at high temperature during the entire synthesis and be subject to repeated treatment with strongly electrophilic acylating agents and base for Fmoc removal. While most peptides with standard acid-labile protecting groups are stable under these conditions, aspartic acid-containing polypeptides in particular are problematic to handle at elevated temperature. Repeated treatment with base can cause the adjacent backbone amide to become deprotonated and attack the protected aspartic acid residue, forming a 5-membered cyclic imide. [37] This product is typically observed with a slightly shifted retention time on HPLC with a mass 18 Da. less than the desired product. While at room temperature aspartimide preferentially forms in particular sequence contexts, at 90 °C, we have found that it forms in most sequence contexts.

When presented with an aspartimide-forming peptide, we find that reducing the temperature from 90 °C to 70 °C and reducing the piperidine concentration from 20% to 10% decreases the amount of aspartimide. Though these syntheses are slower, we see little effect on the crude peptide quality. If this does not solve the problem, which is often the case in a -DG- sequence context, addition of 0.1M HOBt to the

deprotection cocktail can further reduce aspartimide. However, because HOBt has a high extinction coefficient in the UV range, Fmoc removal data can no longer be obtained. We have had success with aspartic acid monomers with bulky protecting groups, such as Asp(OMpe) and Asp(OBno), as well as Hmb-protected dipeptides, but these monomers are expensive and difficult to obtain at scale, limiting their broader applicability. In our view, aspartimide formation is one major challenge for a universal fast flow SPPS method, and solving this chemical challenge would be a boon to the field.

Cleavage of the growing polypeptide chain from the solid support during synthesis is another side reaction that occurs more prominently during SPPS at elevated temperature. We have found that when preparing C-terminal peptide amides the Rink amide linker is stable, but when preparing peptide carboxylates, trityl esters are cleaved slowly during the course of each deprotection cycle. A constant, linear decrease in the Fmoc removal signals is observed when using trityl esters. We recommend the use of other linkers, such as the HMPB linker, for peptide carboxylate synthesis at elevated temperature.

We have also observed that many commercially-available amino acids contain residual ethyl acetate used in their purification and precipitation. Particularly at elevated temperature, we have observed that low levels of ethyl acetate result in gradual decrease in the Fmoc removal signal after each coupling step and limit synthesis efficiency for each step. Rigorous quality control of all chemical feedstocks is therefore crucial to maximize yields. Other troublesome contaminants include formic acid and dimethylamine in DMF, and acetic acid in both amino acids and DIEA.

1.7 Engineering Advancements

There has been a proliferation of accessible automation tools in the past 5 to 10 years, and this technology frees chemists from the constraints of commercial flow chemistry platforms. Using inexpensive microcontrollers such as Arduino or Raspberry Pi, chemists can integrate a wide variety of components, including HPLC pumps,

valves, PID temperature controllers, absorbance detectors, photodiodes, and stepper motors, into new machines that do sophisticated chemistry. These microprocessors have millisecond-level control, and most now have on-board WiFi for uploading process data to the cloud for storage and analysis. Labview, another popular platform for instrument development, makes it easy to sample data from many sources at once to generate interactive control interfaces. At the same time, 3D-printing and machining services such as ProtoLabs now allow anybody to design, prototype and manufacture highly complex, yet chemically resistant, chemical reactors.

These 'plug 'n' play' advancements were crucial our own efforts in the development of Automated Flow Peptide Synthesis. The ease of serial communication allowed us to orchestrate pumps, valves, and motors with electronics that cost less than $50. On a chemical level, the ability to control multiple HPLC pumps with precise timing ensured proper stoichiometry during the entire coupling process, a crucial advance over our previous automation attempts that eliminated truncations caused by excess HATU.

1.8 Conclusions

Continuous flow techniques on the solid phase offer significant advantages over batch methods when applied to peptide synthesis, particularly in enhancing process reproducibility, efficacy, and speed. Atherton, Dryland, and Sheppard first demonstrated that using continuous flow deprotection with Fmoc amino acids permits detection of difficult sequences during synthesis, and now, in-situ continuous flow activation and heat improves the synthesis of these aggregating peptides. While peptide chemists have struggled for decades with manual handling of highly-reactive intermediates that degrade over time and with the application of heat, flow methods are capable of handling even protected amino acid chloride intermediates that degrade after just 0.5s at room temperature. We are excited to see how continuous flow peptide synthesis evolves, and envision a future where every lab has their own "digital ribosome".

1.9 Acknowledgements

We would like to thank Justin Wolfe, Ethan Evans, Colin Fadzen, Dr. Zachary Gates, Dr. Fayçal Touti, Dr. Evan Styduhar, and Prof. JoAnne Stubbe for their helpful comments and discussions during the preparation of this manuscript.

References

1. Barany, G. & Merrifield, R. B. in *The Peptides, Analysis, Synthesis, Biology, Volume 2: Special Methods in Peptide Synthesis, Part A* (eds Gross, E. & Meienhofer, J.) 1–284 (Academic Press, New York, 1980). ISBN: 0-12-304202-X.

2. Kent, S. B. H. Chemical Synthesis of Peptides and Proteins. *Annual Review of Biochemistry* **57,** 957–989 (1988).

3. Bayer, E., Jung, G., Halász, I. & Sebestian, I. A New Support for Polypeptide Synthesis in Columns. *Tetrahedron Letters* **11,** 4503–4505 (1970).

4. Scott, R. P. W., Chan, K. K., Kucera, P. & Zolty, S. The Use of Resin Coated Glass Beads in the Form of a Packed Bed for the Solid Phase Synthesis of Peptides. *Journal of Chromatographic Science* **9,** 577–591 (1971).

5. Lukas, T. J., Prystowsky, M. B. & Erickson, B. W. Solid-Phase Peptide Synthesis under Continuous-Flow Conditions. *Proceedings of the National Academy of Sciences* **78,** 2791–2795 (1981).

6. Dryland, A. & Sheppard, R. C. Peptide Synthesis. Part 8. A System for Solid-Phase Synthesis under Low Pressure Continuous Flow Conditions. en. *Journal of the Chemical Society, Perkin Transactions 1,* 125–137. ISSN: 1364-5463 (Jan. 1986).

7. Sarin, V. K., Kent, S. B. H. & Merrifield, R. B. Properties of Swollen Polymer Networks. Solvation and Swelling of Peptide-Containing Resins in Solid-Phase Peptide Synthesis. *Journal of the American Chemical Society* **102,** 5463–5470 (1980).

8. Atherton, E., Brown, E., Sheppard, R. C. & Rosevear, A. A Physically Supported Gel Polymer for Low Pressure, Continuous Flow Solid Phase Reactions. Application to Solid Phase Peptide Synthesis. *Journal of the Chemical Society, Chemical Communications,* 1151 (1981).

9. Atherton, E. & Sheppard, R. C. Solid Phase Peptide Synthesis Using N,-Fluorenylmethoxy Carbonylamino Acid Pentafluorophenyl Esters. *J. CHEM. SOC., CHEM. COMMUN,* 165–166 (1985).

10. Dryland, A. & Sheppard, R. C. Peptide Synthesis. Part 11. A System for Continuous Flow Solid Phase Peptide Synthesis Using Fluorenylmethoxycarbonyl-Amino Acid Pentafluorophenyl Esters. *Tetrahedron* **44,** 859–876 (1988).

11. Goodman, M., Moroder, L., Felix, A., Fields, G. B. & Hodges, R. S. *Houben-Weyl Methods of Organic Chemistry Vol. E 22d, 4th Edition Supplement: Synthesis of Peptides and Peptidomimetics* ISBN: 978-3-13-182334-2 (Thieme, 2014).

12. Frank, R. & Döring, R. Simultaneous Multiple Peptide Synthesis under Continuous Flow Conditions on Cellulose Paper Discs as Segmental Solid Supports. *Tetrahedron* **44,** 6031–6040 (1988).

13. Mutulis, F., Tysk, M., Mutule, I. & Wikberg*, J. E. S. A Simple and Effective Method for Producing Nonrandom Peptide Libraries Using Cotton as a Carrier in Continuous Flow Peptide Synthesizers. doi:10.1021/CC0200078 (2002).

14. Rapp, W. & Bayer, E. *Uniform Microspheres in Peptide Synthesis: Ultrashort Cycles and Synthesis Documentation by on-Line Monitoring as an Alternative to Multiple Peptide Synthesis* in (1993), 25.

15. Mándity, I. M., Olasz, B., Ötvös, S. B. & Fülöp, F. Continuous-Flow Solid-Phase Peptide Synthesis: A Revolutionary Reduction of the Amino Acid Excess. en. *ChemSusChem* **7,** 3172–3176. ISSN: 1864-564X (Nov. 2014).

16. Mándity, I. M., Ötvös, S. B., Szőllősi, G. & Fülöp, F. Harnessing the Versatility of Continuous-Flow Processes: Selective and Efficient Reactions. *The Chemical Record* **16,** 1018–1033 (2016).

17. So, S., Peeva, L. G., Tate, E. W., Leatherbarrow, R. J. & Livingston, A. G. Membrane Enhanced Peptide Synthesis. *Chemical Communications* **46,** 2808 (2010).

18. Hyde, C., Johnson, T. & Sheppard, R. C. Internal Aggregation during Solid Phase Peptide Synthesis. Dimethyl Sulfoxide as a Powerful Dissociating Solvent. *Journal of the Chemical Society, Chemical Communications,* 1573 (1992).

19. Yu, H. M., Chen, S. T. & Wang, K. T. Enhanced Coupling Efficiency in Solid-Phase Peptide Synthesis by Microwave Irradiation. *The Journal of Organic Chemistry* **57,** 4781–4784 (1992).

20. Erdélyi, M. & Gogoll, A. Rapid Microwave-Assisted Solid Phase Peptide Synthesis. *Synthesis* **2002,** 1592–1596 (2002).

21. Collins, J. M., Porter, K. A., Singh, S. K. & Vanier, G. S. High-Efficiency Solid Phase Peptide Synthesis (HE-SPPS). *Organic Letters* **16,** 940–943. ISSN: 1523-7060 (Feb. 2014).

22. Collins, J. M., Porter, K. A., Singh, S. K. & Vanier, G. S. High-Efficiency Solid Phase Peptide Synthesis (HE-SPPS). *Organic Letters* **16,** 940–943. ISSN: 1523-7060 (Feb. 2014).

23. Palasek, S. A., Cox, Z. J. & Collins, J. M. Limiting Racemization and Aspartimide Formation in Microwave-Enhanced Fmoc Solid Phase Peptide Synthesis. *Journal of Peptide Science* **13,** 143–148 (2007).

24. Carpino, L. A., Chao, H. G., Beyermann, M. & Bienert, M. Oxazolone Formation as a Rationale for Anoma. **56,** 2635–2642 (1991).

25. Carpino, L. A., El-Faham, A. & Albericio, F. Racemization Studies during Solid-Phase Peptide Synthesis Using Azabenzotriazole-Based Coupling Reagents. *Tetrahedron Letters* **35,** 2279–2282 (1994).

26. Griehl, C., Kolbe, A. & Merkel, S. Quantitative Description of Epimerization Pathways Using the Carbodiimide Method in the Synthesis of Peptides. en.

Journal of the Chemical Society, Perkin Transactions 2, 2525. ISSN: 0300-9580, 1364-5471 (1996).

27. Goodman, M., Levine, L., 4y, V., 4nd, G. & Levine ', L. Peptide Synthesis Uia Active Esters. IV. Racemization and Ring-Opening Reactions of Optically Active Oxazolones. *Journal of the American Chemical Society* **86,** 2918–2922 (1964).

28. Jones, J. H. & Witty, M. J. An Oxazol-5(4H)-One Derived from a Benzyloxycarbonylamino-Acid. en. *Journal of the Chemical Society, Chemical Communications,* 281–282. ISSN: 0022-4936 (Jan. 1977).

29. Kemp, D. S. in *The Peptides, Analysis, Synthesis, Biology, Volume 1: Major Methods of Peptide Bond Formation* (eds Gross, E. & Meienhofer, J.) 315–383 (Academic Press, New York, 1979).

30. Yoshida, J.-i., Takahashi, Y. & Nagaki, A. Flash Chemistry: Flow Chemistry That Cannot Be Done in Batch. en. *Chem. Commun.* **49,** 9896–9904. ISSN: 1359-7345, 1364-548X (2013).

31. Petersson, E. J. & Schepartz, A. Toward β-Amino Acid Proteins: Design, Synthesis, and Characterization of a Fifteen Kilodalton β-Peptide Tetramer. *Journal of the American Chemical Society* **130,** 821–823 (2008).

32. Wang, P. S. P. & Schepartz, A. β-Peptide Bundles: Design. Build. Analyze. Biosynthesize. *Chemical Communications* **52,** 7420–7432 (2016).

33. Carpino, L. A., Chao, H. G., Beyermann, M. & Bienert, M. [(9-Fluorenylmethyl)Oxy]Carbonyl (Fmoc) Amino Acid Chlorides in Solid-Phase Peptide Synthesis. *The Journal of Organic Chemistry* **56,** 2635–2642 (1991).

34. Chen, F., Lee, Y. C. & Benoiton, N. L. Preparation of N-9-Fluorenylmethoxy Carbonylamino Acid Chlorides from Mixed Anhydrides by the Action of Hydrogen Chloride. *International Journal of Peptide and Protein Research* **38,** 97–102 (2009).

35. Falb, E., Yechezkel, T., Salitra, Y. & Gilon, C. In Situ Generation of Fmoc-Amino Acid Chlorides Using Bis-(Trichloromethyl)Carbonate and Its Utilization for Difficult Couplings in Solid-Phase Peptide Synthesis. *Journal of Peptide Research* **53,** 507–517 (1999).

36. Fuse, S., Mifune, Y. & Takahashi, T. Efficient Amide Bond Formation through a Rapid and Strong Activation of Carboxylic Acids in a Microflow Reactor. *Angewandte Chemie International Edition* **53,** 851–855 (2014).

37. Lauer, J. L., Fields, C. G. & Fields, G. B. Sequence Dependence of Aspartimide Formation during 9-Fluorenylmethoxycarbonyl Solid-Phase Peptide Synthesis. *Letters in Peptide Science* **1,** 197–205 (1995).

Chapter 2

A fully automated flow-based approach for accelerated peptide synthesis

This chapter was adapted from the publication "A fully automated flow-based approach for accelerated peptide synthesis" with the following authors:

Alexander J. Mijalis[1%], Dale A. Thomas III[2%], Mark D. Simon[1], Andrea Adamo[2], Ryan Beaumont, Klavs F. Jensen[2], Bradley L. Pentelute[1]

1: Department of Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA, 02139, USA.

2: Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA, 02139, USA. Corresponding author. Email: blp@mit.edu (B.L.P.)

% denotes Equal Authors

D.T. and A.M. designed, coded, and assembled the AFPS and designed experiments to characterize it. R.B., A.M., and D.T. designed and implemented the LabView graphical interface and revised control electronics. M.S., A.M., D.T., and A.A. designed the valving and earlier prototype versions of the AFPS. A.M., D.T., M.S., K.F.J., and B.L.P. wrote the manuscript. K.F.J. and B.L.P. supervised the work.

2.1 Introduction

Peptides and proteins are important in the search for new therapeutic molecules. [1] Underpinning peptide and protein research is the need to design new functional variants and quickly iterate on these designs to optimize potency, selectivity, and stability. Biological expression of peptides can be fast and scalable, but becomes difficult outside of the twenty genetically encoded amino acids. On the other hand, despite the expanded number of protected amino acid monomers, chemical peptide synthesis is limited by synthesis speeds of minutes to hours per amide bond. In this communication, we report our development of Automated Flow Peptide Synthesis (AFPS), a method with the flexibility of solid phase chemical synthesis in a vastly accelerated format. AFPS reduces the amide bond forming step for standard Fmoc peptide synthesis to seven seconds and the entire cycle for each amino acid addition to 40 seconds, while providing a higher level of control over the chemistry than existing manual or automated methods. UV monitoring allows for in-process assessment of Fmoc removal, and automated exchange of disposable reactors lets the user move quickly to the next peptide synthesis.

When developing this method, we were inspired by numerous successes performing complex multi-step flow reactions. [2–7] We also drew from previous advancements in the peptide field. In an early example of flow peptide synthesis, pre-activated amino acids recirculated over a bed of solid support for minutes to hours,[8, 9] using resins that were tolerant of the high pressures in this method. [10] However, amino acid recirculation, high pressure, and use of sensitive pre-activated amino acids limited these approaches.

2.2 Results and Discussion

In our earlier work [11, 12], we addressed some of these limitations with manual flow peptide synthesis; however, the approach lacked sufficient control over temperature, activation, fluid handling, and in-line monitoring. We were limited by human perfor-

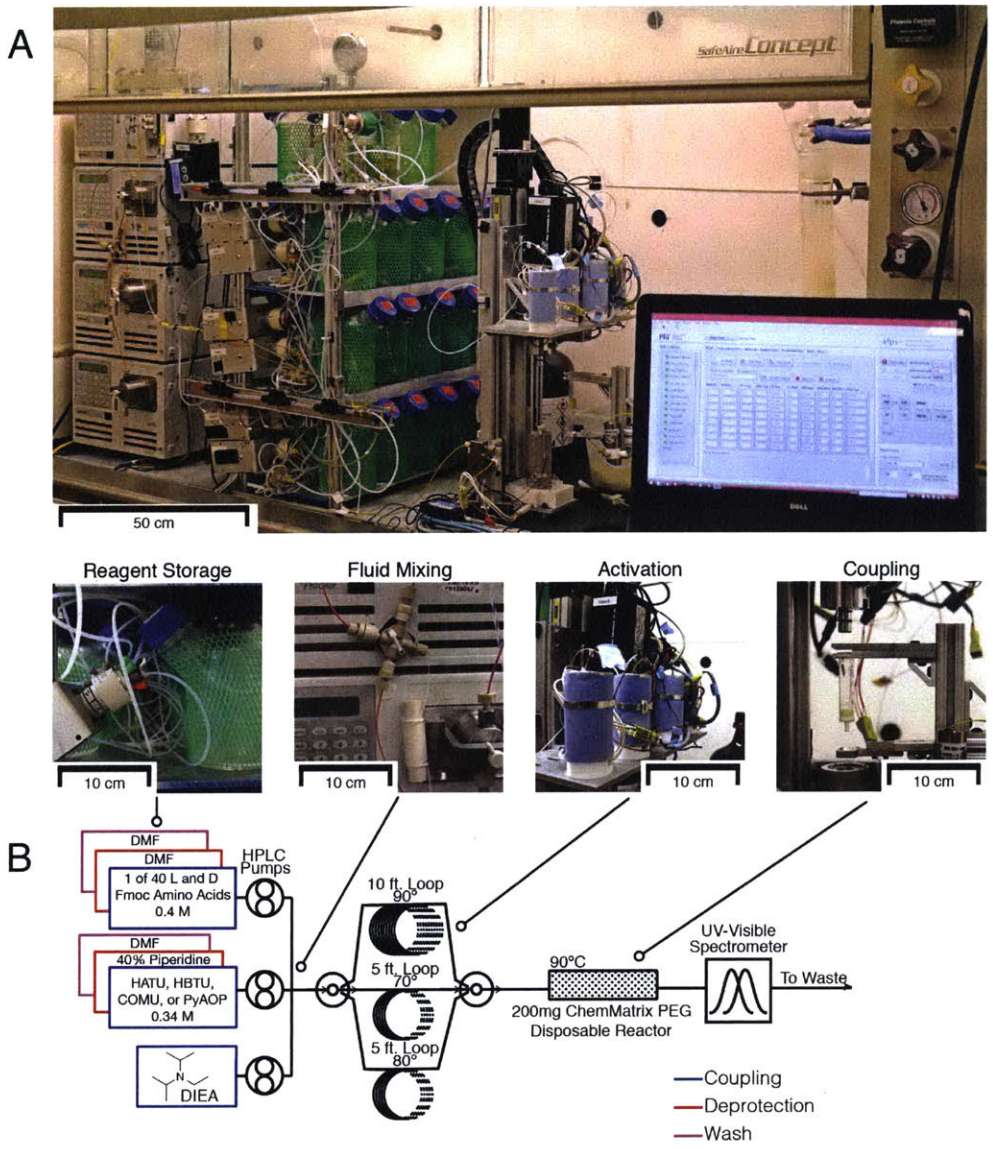

Figure 2-1: Automated flow peptide synthesis enables seven second amide bond formation and complete solid phase peptide synthesis cycles in 40 seconds. **A.** Photographs of the automated flow solid phase synthesizer modules. **B.** Process flow diagram. Amino acid, activating agent and DIEA are merged together by three HPLC pumps. Multiposition valves control the selection of the amino acid and activating agent. Amino acids are activated in one of several heated flow paths, determined by a column selector valve, then flowed over a resin bed containing peptidyl resin housed in a 6-mL fritted syringe in a heated jacket. The effluent is passed through a UV-visible spectrometer to waste.

mance associated with manual handling and delivery of hazardous reagents. AFPS ameliorates these problems by combining sophisticated automation with advance-

ments from the peptide synthesis and continuous flow pharmaceutical manufacturing fields.

The Automated Flow Peptide Synthesizer consists of five modules, depicted in Figure 2-1A-B. During a coupling reaction, the pumps draw reagents from the storage module, and mix the desired amino acid with diisopropylethylamine (DIEA), and an activating agent such as HATU or PyAOP in the mixing module. This mixture flows through the activation module, an electrically heated tubular reactor, where it quickly heats to 90 °C and forms an active ester. Within two seconds of activation, the activated amino acid flows through the coupling module, a packed bed of peptide synthesis resin maintained at 90 °C, where amide bond formation is complete within seven seconds. Figure 2-2A shows a detailed timeline for the standard synthesis methods.

Peptide synthesis resin is contained in a 6-mL disposable fritted syringe barrel for easy removal. The AFPS monitors Fmoc removal for each cycle by recording the absorbance of the reactor effluent as a function of time. The Fmoc removal absorbance chromatogram allows us to infer the deprotection efficiency, the coupling yield, and the rate of mass transfer through the peptidyl resin, aiding in identification of on-resin peptide aggregation. [9]

We initially validated AFPS by synthesizing test peptides ALFALFA (Figure 2-11) and a fragment of acyl carrier protein (ACP65-74, Figure 2-2B). [13] These peptides were synthesized in high yield with low levels of side products. Because L-amino acids are inexpensive, our routine coupling method uses 20 equivalents of activated amino acid. For precious substrates, we found that by halving both the concentration of activated amino acid (from 170 mM to 85 mM) and the volume delivered, we could achieve quantitative coupling with only 6 equivalents of activated amino acid (Figure 2-13).

We then performed a comparative study between longer peptides produced by AFPS, batch synthesis, and reputable custom peptide vendors (Figure 2-3A-B). In all cases, AFPS yielded material that was comparable to or better than the other methods in a fraction of the time. Additionally, as seen in 2-3C, in-process UV

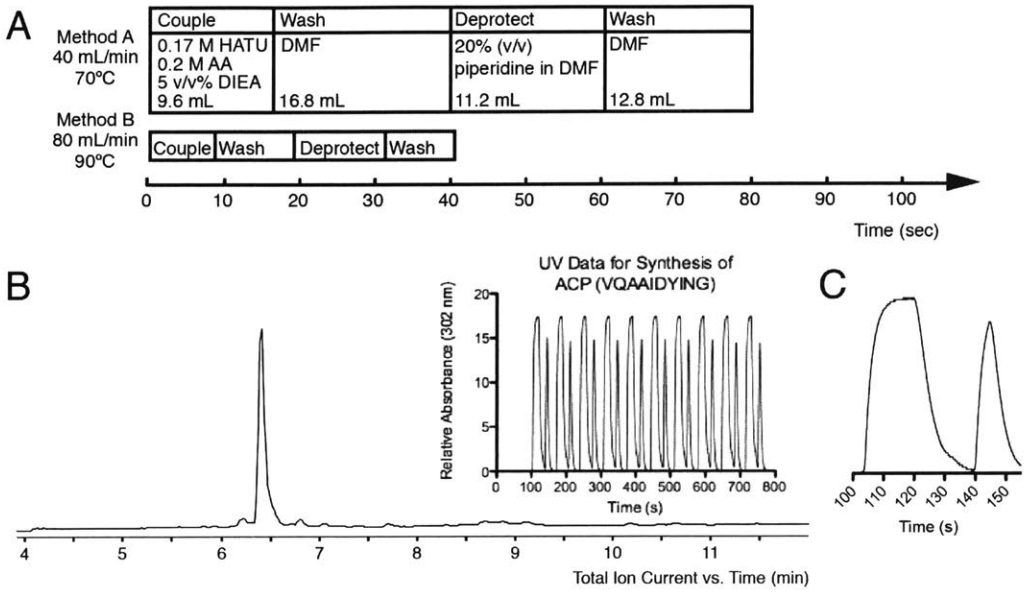

Figure 2-2: Automated flow peptide synthesis method. **A.** Cycle diagrams showing the duration of each step, the solution composition during each step after mixing, and the total volume of reagent used at each step. **B.** LC-MS data for the crude product of acyl carrier protein (65-74) synthesized on 200mg of resin (0.5mmol/g) using Method B. 314mg of peptidyl resin was obtained, yielding 42 mg (44%) crude ACP(65-74). Throughout this work, isolated crude peptide yields are based on the nominal loading of resin. **C.** Example of UV absorbance data for one coupling and deprotection cycle. See Figures 2-15, 2-16, and 2-19 for photos of the individual machine parts.

monitoring provided information about the synthetic yields of each step. The steady decrease in peak area observed for the insulin B chain resulted from chain-terminating side reactions. These byproducts appear as a series of impurities around the main peak in the LC-MS chromatogram.

Next, we assessed epimerization of Cys [14] and His [15] with high-temperature flow activation (Figure 2-5). When activated, Cys and His can lose stereochemistry at the C position, especially at elevated temperature. With a batch microwave synthesizer, coupling Fmoc-L-Cys(Trt) with HBTU and DIEA for 1.5 minutes at 90 °C under microwave irradiation causes 16.7% of the undesired D-Cys containing product to form. [16] In contrast, we found that we can suppress epimerization with AFPS by increasing the flow rate, and therefore decreasing the residence time at temperature for activated histidine and cysteine monomers. Using two model peptides, FHL and GCF, whose diastereomers are separable by RP-HPLC, we found that at 80 ml/min

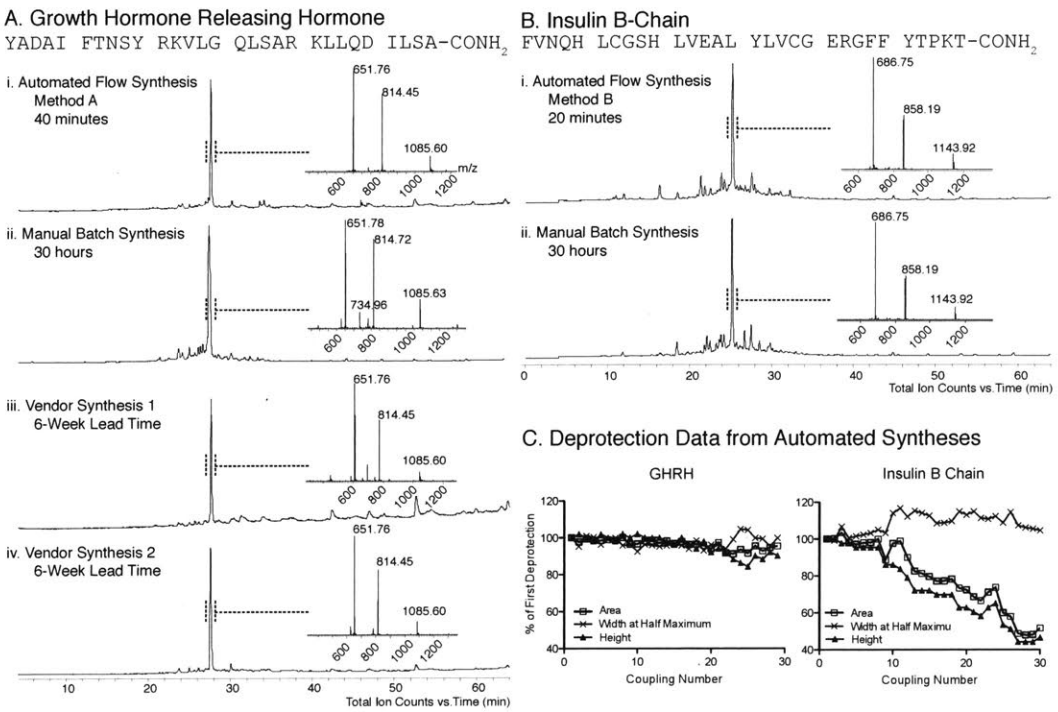

Figure 2-3: Rapid synthesis of long polypeptides and comparison of flow and batch synthesis methods. **A.** Growth hormone releasing hormone (GHRH) was synthesized in (i) 40 minutes with method A (Figure 2-1) (198 mg, 58%), and (ii) 30 hours using manual batch techniques (225 mg, 60%). GHRH was also purchased from two vendors (iii, iv) with a 6-week lead time. Cleavage of 200 mg of each of these peptidyl resins yielded 76 mg (49%, assuming 1 mmol/g loading) and 90 mg (assuming 1 mmol/g loading) respectively. **B.** The insulin B-chain was synthesized in (i) 20 minutes using Method B (163 mg, 53%), and (ii) 30 hours (153 mg, 45%) with manual batch techniques. See Figure 2-4 for synthesis byproduct identification.

0.5% D-His and 3% D-Cys were incorporated. This level of diastereomer formation is consistent with optimized room temperature batch synthesis protocols. [14]

Finally, we investigated the synthesis of "difficult" peptides. [17, 18] With such sequences, aggregation of the growing peptidyl chain causes the kinetics of coupling and deprotection to slow down, resulting in deletion and truncation products. [19, 20] However, the identification of difficult sequences requires time-intensive procedures such as LC-MS analysis, quantitative ninhydrin tests,[21] or Kaiser tests. [22] By contrast, analysis of Fmoc removal by UV absorbance under continuous flow conditions offers a quick, quantitative measure of the success of peptide assembly and the aggregation state of the resin-bound peptide.

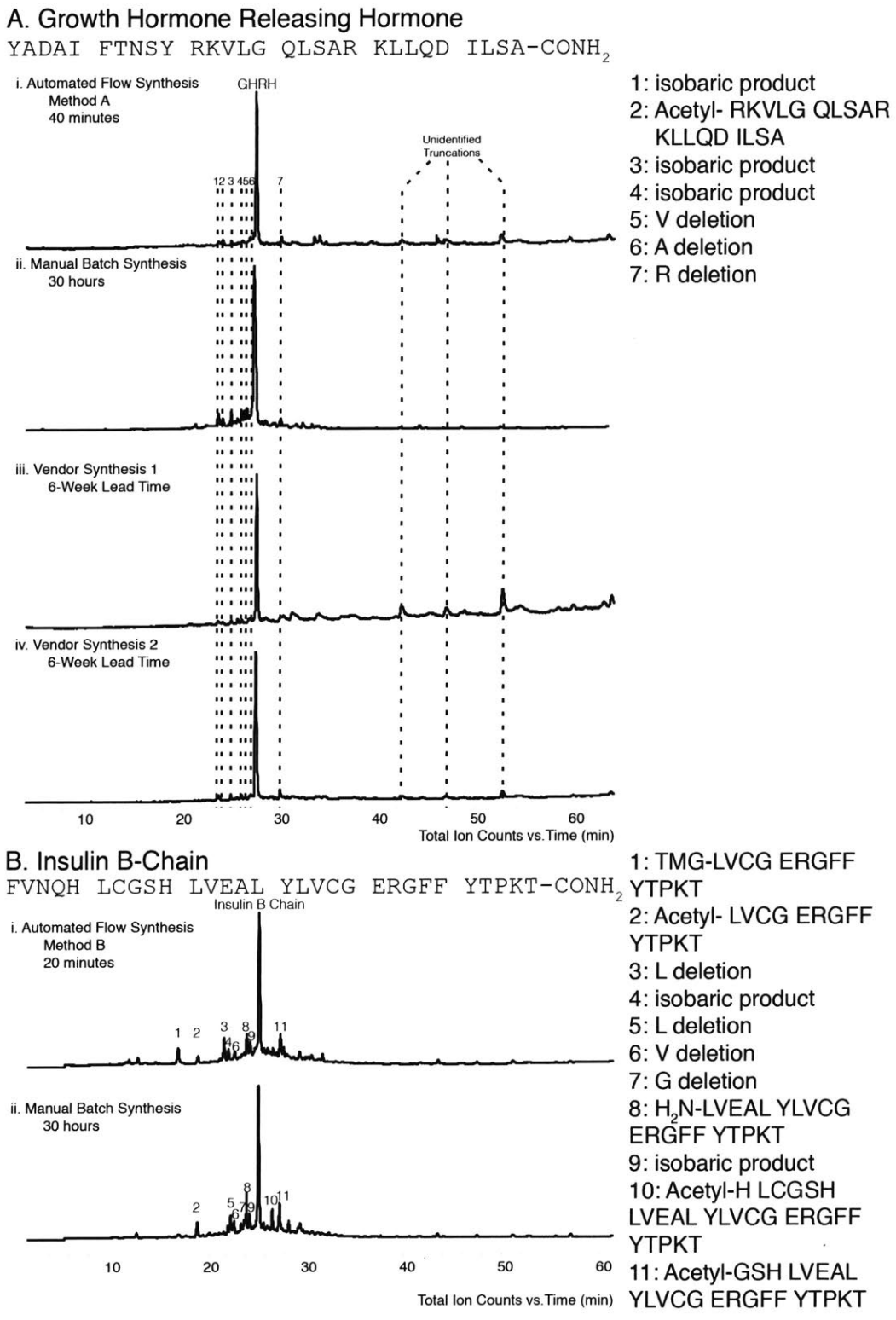

Figure 2-4: Identification of side products from **A.** GHRH and **B.** Insulin B-Chain syntheses. TMG: N-terminal tetramethylguanidyl truncation from HATU (see Figure 2-12).

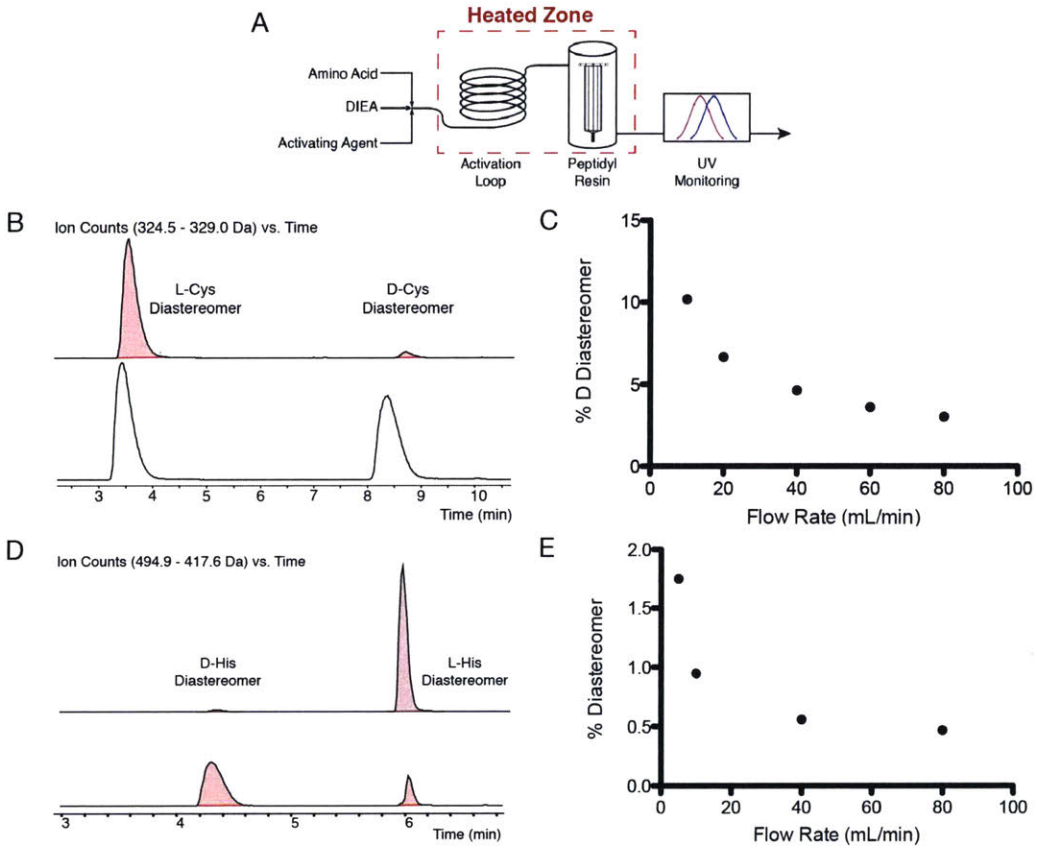

Figure 2-5: Rapid flow activation suppresses cysteine and histidine epimerization. **A.** Diagram of the heated portion of the automated flow peptide synthesizer. **B.** Top: Analysis of model peptide GCF synthesized using AFPS method B and Bottom: a 50/50 mixture of the authentic Cys diastereomers. **C.** Percentage of Cys diastereomer formation as a function of flow rate (ml/min) at 90 °C. **D.** Top: Analysis of model peptide FHL synthesized using AFPS method B, and Bottom: a 50/50 mixture of the authentic His diastereomers. E. Percentage of histidine diastereomer formation as a function of flow rate (ml/min) at 90 °C.

To study the effectiveness of UV monitoring in AFPS, we used the Jung-Redmann (JR) 10-mer, a well-known difficult peptide. [23, 24] To start, we performed batch SPPS of this peptide at room temperature, and found the synthesis begins to fail at the sixth coupling, yielding large amounts of Trp, Phe, and Thr deletion (Figure 2-6-A). Next, using AFPS, we synthesized this polypeptide at 90 °C (Figure 2-6-B) and examined the curves for Fmoc removal after each amino acid addition. The area under each deprotection curve was constant until a sharp decrease during the final deprotection. LC-MS analysis of the product revealed that the major byproduct was a Trp deletion, highlighting that higher temperature eliminated the Phe and Thr deletions observed in the batch case. The full width at half maximum of the deprotection peaks increased by almost 20% during the latter couplings, suggesting a slowing of Fmoc removal or a reduced rate of diffusion through the aggregated peptide, and predicting the final deletion.

We next used the UV readout on the Automated Flow Peptide Synthesizer to monitor and minimize Trp deletion. A common step to mitigate peptide aggregation during synthesis is to lower resin substitution;[25] therefore, we prepared a set of amine reduced resins and monitored the Fmoc removal signals during chain assembly. As seen in Figure 2-6C and Figure 2-6D, at reduced loading, the observed decrease in deprotection peak area for the final Trp coupling became less pronounced and pointed toward an optimal loading of 0.3 mmol/g, agreeing with previous observations. [25] Equally important, we found that for this deletion product, the UV readout corroborated the LC-MS data (Figure 2-7).

AFPS offers numerous advantages over manual fast flow synthesis,[11, 12] thermally-accelerated batch synthesis,[16, 26] and earlier continuous flow peptide synthesis methods. [9] First, heating, in-line mixing of reagents, and activation of amino acids in a modular format enables rapidly tuning chemistry on a residue-by-residue basis. Second, automation of the entire process with precise pump and valve actuation eliminates human error and variability, making the synthesis highly reproducible. Third, in-flow UV monitoring and data collection allow relative quantitation of Fmoc removal for each cycle, which correlates with amide coupling efficiency. Fourth, maintaining

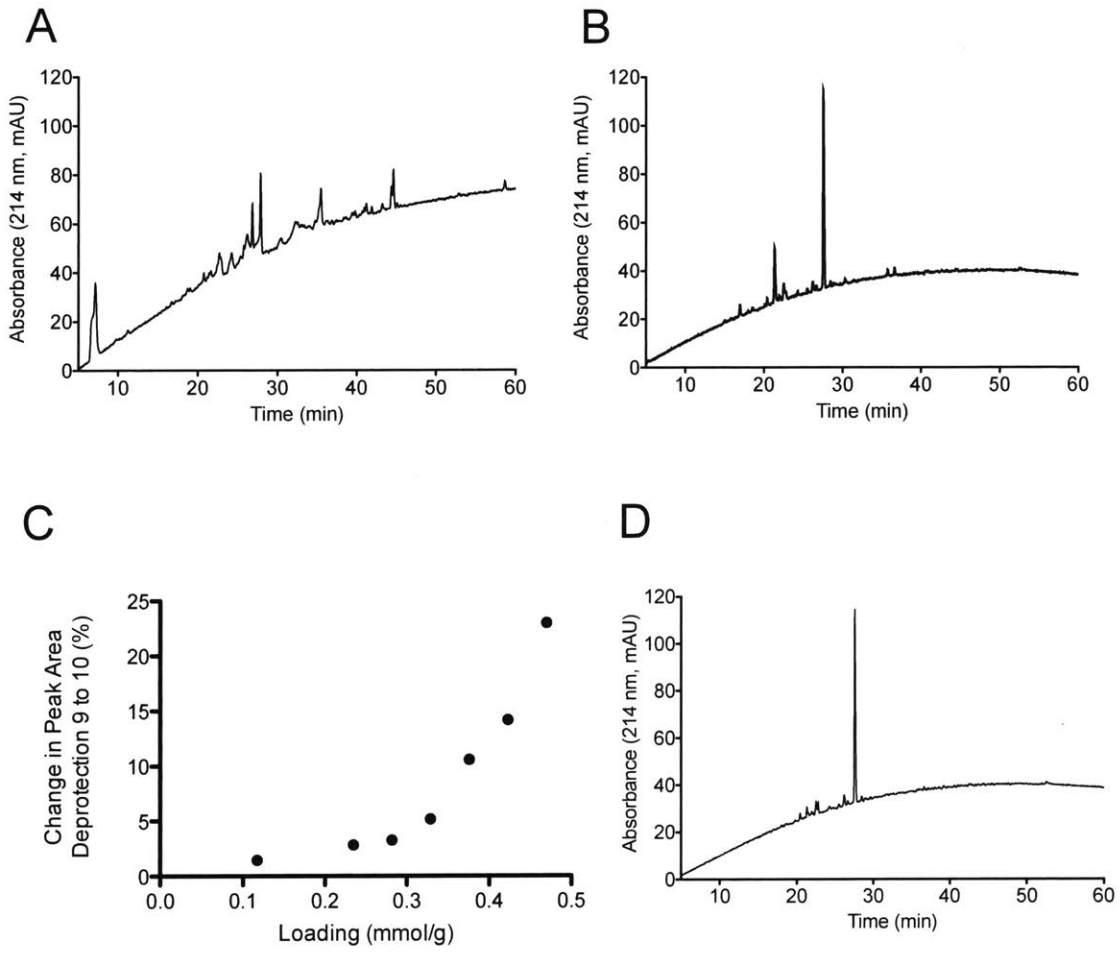

Figure 2-6: Rapid optimization of polypeptide synthesis by in-line monitoring of Fmoc removal. **A.** Crude material made using manual batch synthesis with 0.45 mmol/g resin loading. **B.** Plot of change of deprotection peak area from residue 9 to 10 from the UV absorbance data as a function of resin loading. **C-D.** LC-MS chromatogram of the final crude products for JR 10-mer prepared by **C:** automated flow with 0.45 mmol/g resin loading; and **D:** automated flow with 0.27 mmol/g resin loading. See Figure 2-7 for additional information.

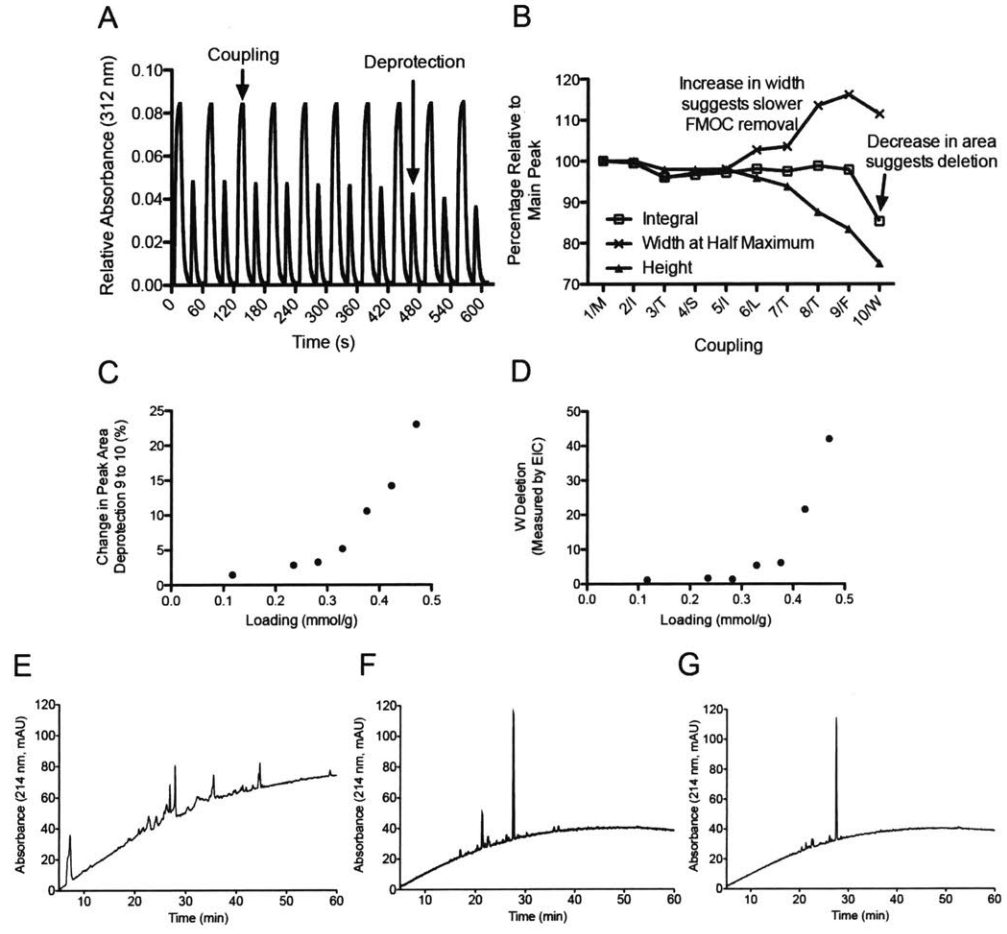

Figure 2-7: Rapid optimization of polypeptide synthesis by monitoring Fmoc removal in-line. **A.** UV absorbance data recorded during the initial synthesis of the JR 10-mer (WFTTLISTIM) using method B. **B.** Analysis of the Fmoc removal UV trace reveals peak width gradually increases after the 4th coupling, while peak area stays constant until the last coupling. **C.** Plot of change of deprotection peak area from residue 9 to 10 from the UV absorbance data as a function of resin loading. **D.** Plot of Trp deletion as a function of resin loading, determined from LC-MS. The amount of deletion measured by LC-MS corroborates the amount of deletion predicted by the UV in-process measurement. **E-G.** LC-MS chromatograms of the final crude products for JR 10-mer prepared by manual batch with 0.45 mmol/g resin loading (**E**), automated flow with 0.45 mmol/g resin loading (**F**), and automated flow with 0.27 mmol/g resin loading (**G**).

a high flux of wash solvent, deprotection agent, and activated amino acid over the resin bead significantly accelerated peptide synthesis.

Automated Flow Peptide Synthesis is a robust method to rapidly access peptides, removing synthesis as a barrier to biological research and drug discovery. Moving toward these applications, we have begun to synthesize longer, functional proteins and libraries of therapeutic candidates (Figures 2-8, 2-9, and 2-10). We believe that, by challenging the accepted speed limit for peptide synthesis, this method will facilitate new chemistry and engineering designs to approach the efficiency of Nature's ribosome.

MQKGN FRNQR KTVKC FNCGK EGHIA KNCRA PRKKG CWKCG KEGHQ MKDCT ERQAN-$CONH_2$

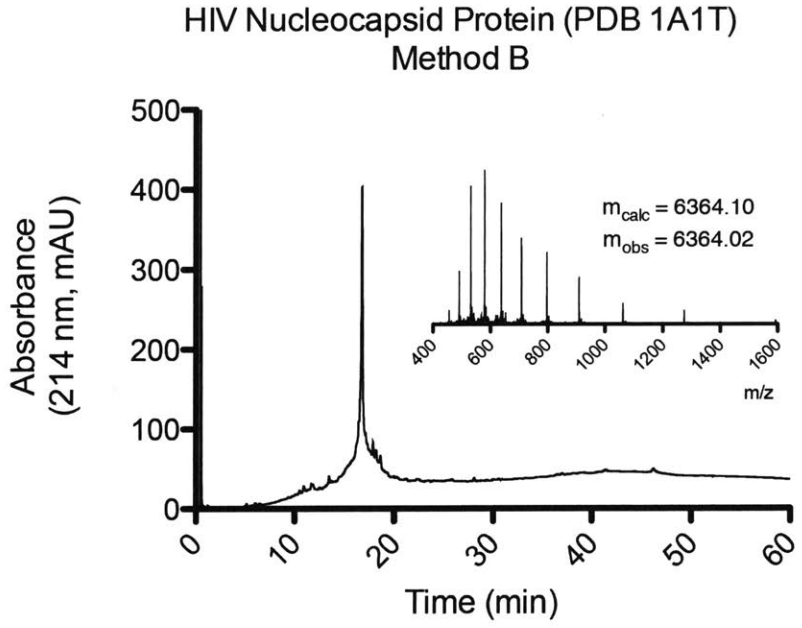

Figure 2-8: Chromatogram of crude HIV Nucleocapsid Protein, a 55-residue polypeptide containing a Zn-finger binding domain, synthesized in 40 minutes using AFPS Method B. Resin loading: 50 mg, 0.023 mmol. Crude yield: 65 mg (44%)

GGHGG NGGNG NPGCA GGVGG AGGAS GGTGV GGRGG KGGSG TPKGA DGAPG AP-NH_2
HMB-Protected

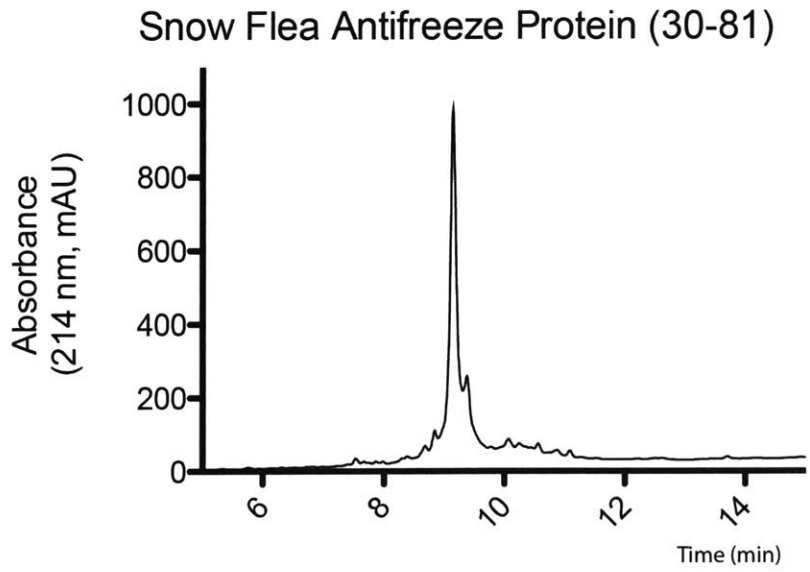

Figure 2-9: Chromatogram of crude Snow Flea Antifreeze Protein, a 52-residue polypeptide, synthesized in 35 minutes using AFPS Method B. An ortho-hydroxy-para-methoxybenzyl (HMB) protecting group was installed at the C-terminus to reduce aspartimide formation during elevated temperature SPPS.

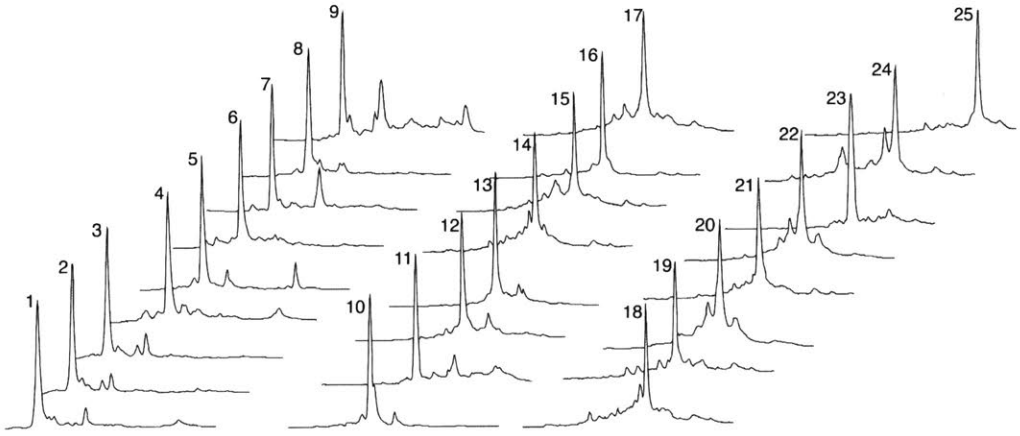

Figure 2-10: In order to demonstrate the ability of the AFPS to quickly synthesize a large number of arbitrary peptides, we prepared a set of 25 conotoxins selected the literature with no synthesis optimization in less than two days. These molecules are rich in biological function, but must be extracted from live creatures. We believe that AFPS can be a way to get these rare materials into the hands of biologists and chemists for characterization. The chromatograms shown above are crude, unfolded conopeptides. Sequences obtained from Bhatia, S. et al. Constrained De Novo Sequencing of Conotoxins. J. Proteome Res. 11, 4191–4200 (2012)

Table 2.1: Conotoxin sequences synthesized on AFPS.

Entry	Sequence	Expected Mass	Observed Mass
1	CCSIHDNSCCGI	1252.435	1252.44
2	KPCCSIHDNSCCGI	1477.583	1477.57
3	KPCCSIHDSSCCGI	1450.572	1450.58
4	CCPPAACNMGCKPCC	1498.503	1498.52
5	NLQLLCCKHTPKCCT	1702.803	1702.8
6	GCCGAPSCMAGCRPCC	1516.489	1516.5
7	TCFGCTPCC	932.29	932.28
8	TSDCCFYHNCCC	1396.402	1396.41
9	CCRTCFGCTPCC	1294.41	1294.41
10	NLQLLCCKHTPACCT	1645.745	1645.77
11	GCPWQPYC	951.362	951.34
12	CCSWDWCDHPSCTCC	1746.507	1746.52
13	CCKFPCPDSCRYLCC	1738.647	1738.66
14	KFCCDSNWCHISDCECCY	2156.723	2156.53
15	GCCHLLACRMGCSPCCW	1840.684	1840.7
16	VCCPFGGCHELCQCCE	1728.59	1728.56
17	GCCHLLACRMGCTPCCW	1854.699	1854.71
18	FCCDSNWCHISDCECCY	2028.628	2028.61
19	GCCGAFACRFGCTPCC	1596.548	1596.56
20	EIILHALGTRCCSWDVCDHPSCTCCG	2820.162	2820.02
21	SCCNAGFCRFGCTPCCY	1832.627	1832.59
22	EIILHALGTRCCSWDVCDHPSCTCC	2763.141	2763
23	CCPPALWCC	993.358	993.38
24	NCPYCVVYCCPPAYCQASGCRPP	2491.992	2491.97
25	IKIGPPCCSGCFFACA	1800.786	1800.79

2.3 Instrument Design and Experimental Methods

2.3.1 Materials

All reagents were purchased and used as received. N-α-Fmoc amino acids were purchased from Creo Salus. Fmoc-His(Boc)-OH and O-(7-Azabenzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate (HATU) was purchased from ChemImpex. Omnisolv grade N,N-dimethylformamide (DMF) was purchased from EMD Millipore (DX1726-1). Diisopropylethylamine (DIEA, catalog number 387649), piperidine, trifluoroacetic acid, triisopropylsilane, acetonitrile and 1,2-ethanedithiol (EDT) were purchased from Sigma Aldrich. H-Rink Amide ChemMatrix polyethylene glycol resin with a loading of 0.5 mmol/g was purchased from Pcas Biomatrix (catalog number 1744). (7-Azabenzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate (PyAOP) was purchased from P3 Biosystems.

2.3.2 Initial Synthesis Conditions and AFPS Characterization

The conditions for Fmoc removal and DMF washing chosen at the outset were based on our previously published work, and our starting point is shown in Figure 2-11. At 20 mL/min total system flow rate and at 70 °C, treatment with 20% piperidine was chosen to be 20s, conditions that were previously sufficient for complete Fmoc removal. The DMF washes were chosen to be 30s. We verified the washout time by introducing Fmoc amino acid into the reactor and using the UV detector to ensure that the system was cleared of any UV active material after the DMF wash.

The scheme for in-line mixing the fluid streams of activating agent and the amino acid allows for versatility in the conditions used for coupling. However, it requires a departure from the conditions traditionally used for aminoacylation in Fmoc synthesis. Typically, reagents are used at their solubility limits, around 0.4 M for Fmoc amino acids and uronium coupling agents. However, because these reagents were stored separately on the AFPS and coupling involved mixing two concentrated solutions, the final solution used for aminoacylation at the outset was composed of 0.2 M amino

acid and activating agent. For the typical coupling, a total of 9.6 mL of this coupling solution was used to ensure complete coupling. We initially tested these conditions for the synthesis of a short polypeptide, ALFALFA (Figure 2-11.

2.3.3 Optimization and Development of a High-Efficiency Synthesis Method

Once we demonstrated quantitative aminoacylation using 0.2M activated amino acid (final solution after three way mixing) at 70 °C and a 20 mL/min flow rate – a 15s coupling time – we naturally wanted to attempt to make the coupling conditions more aggressive. We synthesized ACP, a 10-residue peptide that is typically used as a diagnostic "difficult" sequence, at 70 °C, using the same volume of coupling reagent in each experiment, at 20, 40, and 60 mL/min total flow rate (Figure 2-12). At higher flow rates, we observed the increasing formation of a chain termination side product – a tetramethylguanidyl truncation during the glutamine coupling. We hypothesized that this was due to incomplete activation at elevated flow rates: when the amounts of activating agent and amino acid are nearly equal, there could be residual HATU present which can guanidinylate the N-terminus of the growing peptidyl chain. Reducing the concentration of activating agent to 0.34M, as well as ensuring full synchronization of the pump heads eliminated this side reaction in most cases, allowing us to synthesize ACP at 80 mL/min with method B in quantitative yield (Figure 2-1). For Fmoc-Arg couplings in other peptides, these truncations were still observed, so we used PyAOP as the activating agent for these couplings.

For coupling of expensive carboxylic acid substrates, we reduced the amount of reagent delivered to the resin by roughly a factor of 4. This was accomplished by halving the concentration of both the amino acid stock solution (to 0.2 M from 0.4M) and the coupling agent stock solution (from 0.34 M to 0.17M), as well as reducing the volume of each stock solution delivered from 4.8 mL to 2.8 mL. The DIEA amount (5 v/v%), temperature (90 °C), and flow rate (80 mL/min) were left unchanged. Therefore, for each coupling, a total of 5.6 mmol of reagent (roughly 6 eq. with 200

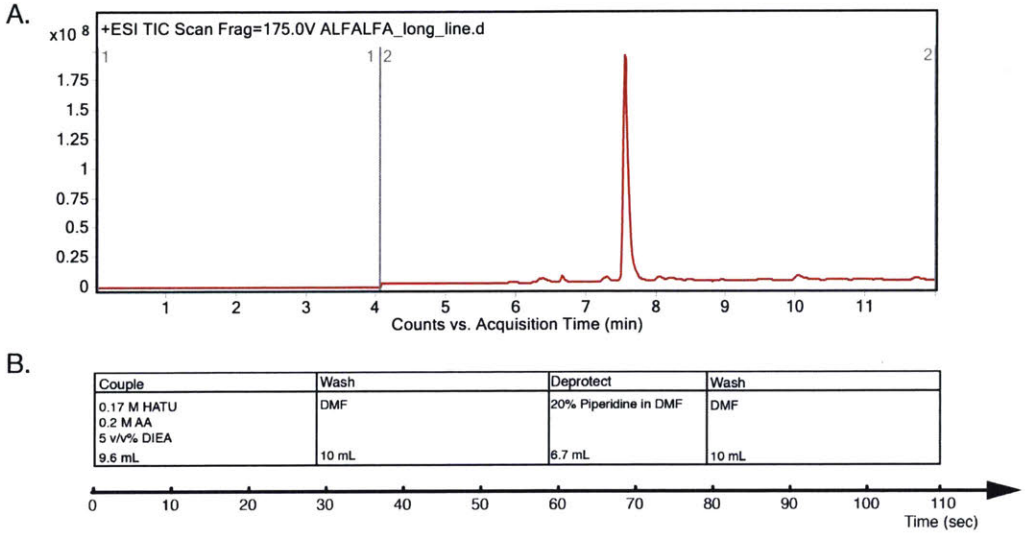

Figure 2-11: **A.** LC-MS chromatogram for a synthesis of H_{2N}-ALFALFA-$CONH_2$ (calculated mass: 750.43 Da, observed mass: 750.43 Da) using automated flow peptide synthesis with inline activation. The scheme for in-line mixing of the fluid streams of activating agent and amino acid allows for versatility in the conditions used for coupling. However, it requires a departure from the conditions traditionally used for aminoacylation in Fmoc synthesis. Typically, reagents are used at their solubility limits, around 0.4M for Fmoc amino acids and uronium coupling agents. However, because these reagents were stored separately on the AFPS and coupling involved mixing two concentrated solutions, the final solution used for aminoacylation at the outset was composed of 0.2M amino acid and activating agent. For the typical coupling, a total of 9.6 mL of this coupling solution was used to ensure complete coupling. We initially tested these conditions for the synthesis of a short polypeptide, ALFALFA. **B.** Cycle diagram for automated flow peptide synthesis at 20 mL/min. The conditions for Fmoc removal and DMF washing chosen at the outset were based on our previously published work. At 20 mL/min total system flow rate and at 70 °C, treatment with 20% piperidine was chosen to be 20s, conditions that were previously sufficient for complete Fmoc removal. The DMF washes were chosen to be 30s. We verified the washout time by introducing Fmoc.amino acid into the reactor and using the UV detector to ensure that the system was cleared of any UV active material after the DMF wash.

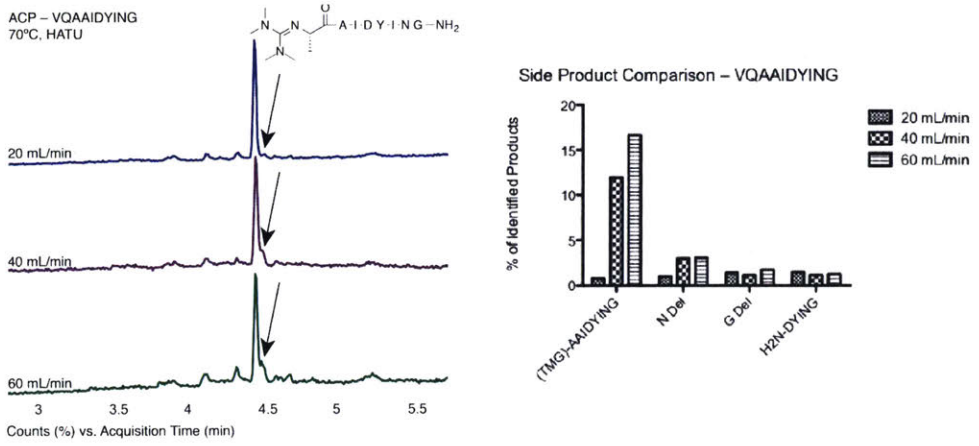

Figure 2-12: Appearance of tetramethylguanidyl adduct as a function of increasing flow rate during the test syntheses of a fragment of the acyl carrier protein. Once we demonstrated quantitative aminoacylation using 0.2M activated amino acid (final solution after three way mixing) at 70 °C and a 20 mL/min flow rate – a 15s coupling time – we naturally wanted to attempt to make the coupling conditions more aggressive. We synthesized ACP, a 10-residue peptide that is typically used as a diagnostic "difficult" sequence, at 70 °C, using the same volume of coupling reagent in each experiment, at 20, 40, and 60 mL/min total flow rate (Supplementary Figure 17). At higher flow rates, we observed the increasing formation of a chain termination side product – a tetramethylguanidyl truncation during the glutamine coupling. We hypothesized that this was due to incomplete activation at elevated flow rates: when the amounts of activating agent and amino acid are nearly equal, there could be residual HATU present which can guanidinylate the N-terminus of the growing peptidyl chain. Reducing the concentration of activating agent to 0.34M, as well as ensuring full synchronization of the pump heads eliminated this side reaction in most cases, allowing us to synthesize ACP at 80 mL/min with method B in quantitative yield (Figure 2-1). For Fmoc-Arg couplings in other peptides, these truncations were still observed, so we used PyAOP as the activating agent for these couplings.

mg of ChemMatrix Rink Amide 0.45 mmol/g resin) was used. The chromatogram for ACP(65-74) synthesized with these conditions is shown below.

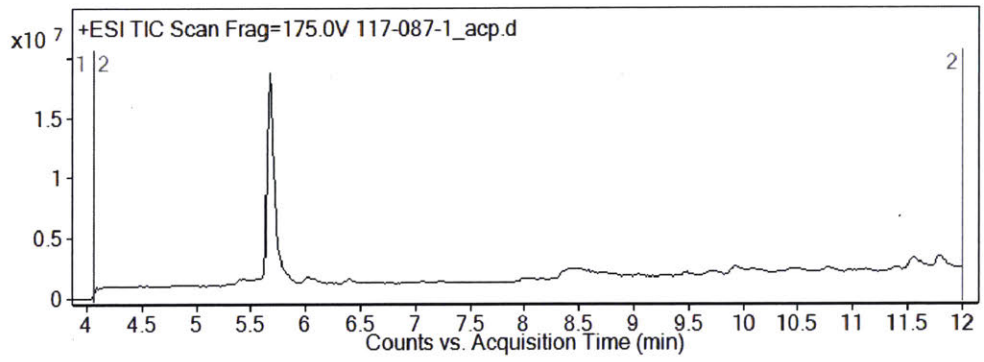

Figure 2-13: Peptide chromatogram of ACP(65-74) synthesized using the high-efficiency flow synthesis method. For coupling of expensive carboxylic acid substrates, we reduced the amount of reagent delivered to the resin by roughly a factor of 4. This was accomplished by halving the concentration of both the amino acid stock solution (to 0.2 M from 0.4M) and the coupling agent stock solution (from 0.34 M to 0.17M), as well as reducing the volume of each stock solution delivered from 4.8 mL to 2.8 mL. The DIEA amount (5 v/v%), temperature (90 °C), and flow rate (80 mL/min) were left unchanged. Therefore, for each coupling, a total of 5.6 mmol of reagent (roughly 6 eq. with 200 mg of ChemMatrix Rink Amide 0.45 mmol/g resin) was used. The chromatogram for ACP(65-74) synthesized with these conditions is shown below.

We also examined the deprotection of Fmoc-Glycine-functionalized peptidyl resin with 20% piperidine at 70, 80, and 90 °C (Figure 2-14). In all three cases, Fmoc-Gly was coupled to 200 mg of ChemMatrix Rink Amide resin at room temperature using batch coupling methods reported under the "Manual Synthesis" subheading in the Supplemental Information. The resins were then transferred to the automated flow synthesizer, where a single treatment of 20% piperidine was performed at either 70, 80, or 90 °C at 80 mL/min. In all three cases, the integrated area of the Fmoc removal peaks was the same, suggesting complete Fmoc removal. However, at higher temperatures, the peak maximum occurs earlier, suggesting either faster deprotection, faster diffusion of the Fmoc-dibenzofulvene adduct out of the resin, or both.

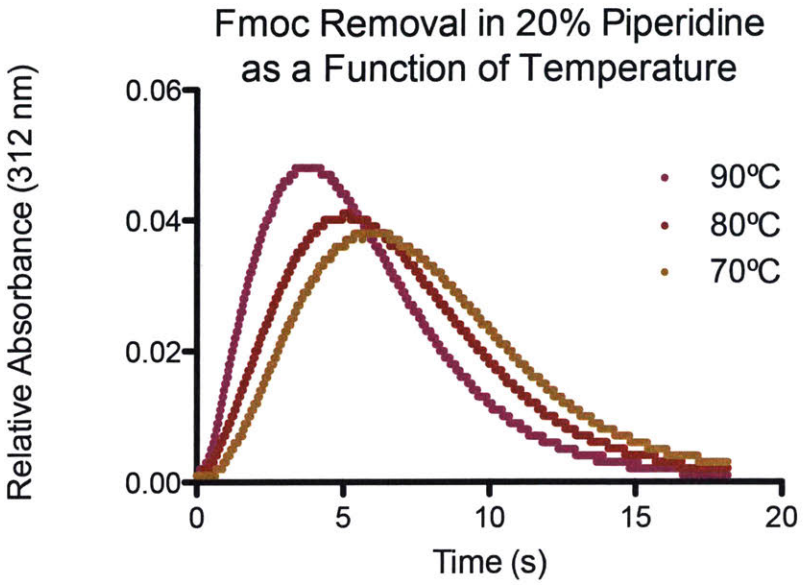

Figure 2-14: Elution of the dibenzofulvene-piperidine adduct as a result of treatment with 20% piperidine at 70, 80, and 90 °C.

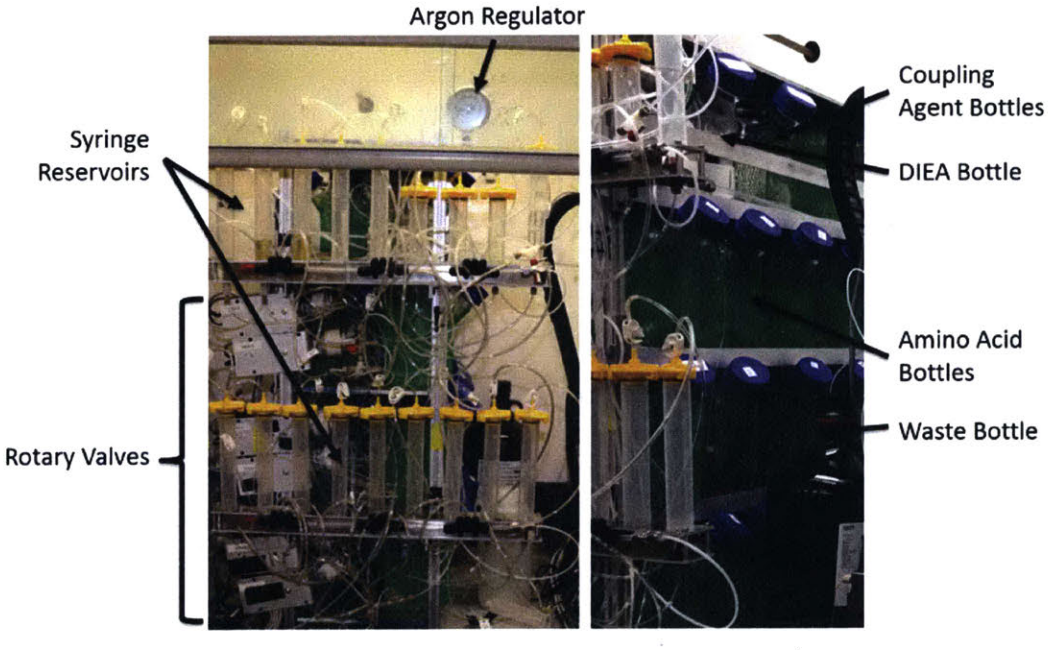

Figure 2-15: Photograph of the AFPS reagent storage system

2.3.4 Reagent Storage and Fluidic Manifold

The reagent storage system used two different vessels to contain reagents: a Chemglass three-neck 500 mL spinner flask for large volumes (CLS-1401-500), and 50 mL polypropylene syringe tubes for smaller volumes (parts # AD930-N, AD955-N). All of the glass bottles were painted with a UV-resistant matte spray paint (Krylon 1309) to reduce UV degradation of the reagents and have a green protective safety net for operation under argon pressure. The argon pressure was maintained at 5 psi pressure with a Swagelok pressure regulator (part# KCP1CFB2B7P60000). The reagent withdraw lines were outfitted with a 20 um polypropylene filter (Valco part# JR-32178) to prevent clogging of pumps, check valves, and lines from any reagent crystallization or impurities.

Each of the three rows of 9 amino acid bottles and two rows of syringes fed into a separate VICI Valco 10-position Auxillary Valve (Vici part# C25-3180EUHA) where the tenth position was connected to a DMF reservoir. Those valves each fed into a VICI Valco 10 position valve, the Main Valve. The Main Valve was connected to the Amino Acid Pump. Bottles containing HATU, other coupling agents, 40% piperidine and DMF fed into a separate 10-position valve, the Coupling Agent Valve. This valve was connected to the Coupling Agent Pump. DIEA fed directly into the DIEA Pump. The tubing lengths were selected so that the volume between the Coupling Agent Pump and the Coupling Agent Valve was equal to the volumes between the Amino Acid Pump and each Auxillary Valve. This is crucial to avoid truncations of the peptide chain, ensuring that, at the start of each coupling cycle, amino acid arrived at the mixer at the same time as activating agent.

2.3.5 Reagent Pumping and Mixing

The AFPS operates with three Varian Prostar 210 pumps. The first pump delivers either an amino acid or DMF. The second pump delivers either a coupling agent, 40% piperidine solution, or DMF. The third pump delivers DIEA. The coupling agent and amino acid pumps have a 50 mL/min stainless steel pump head (Agilent part# 50-

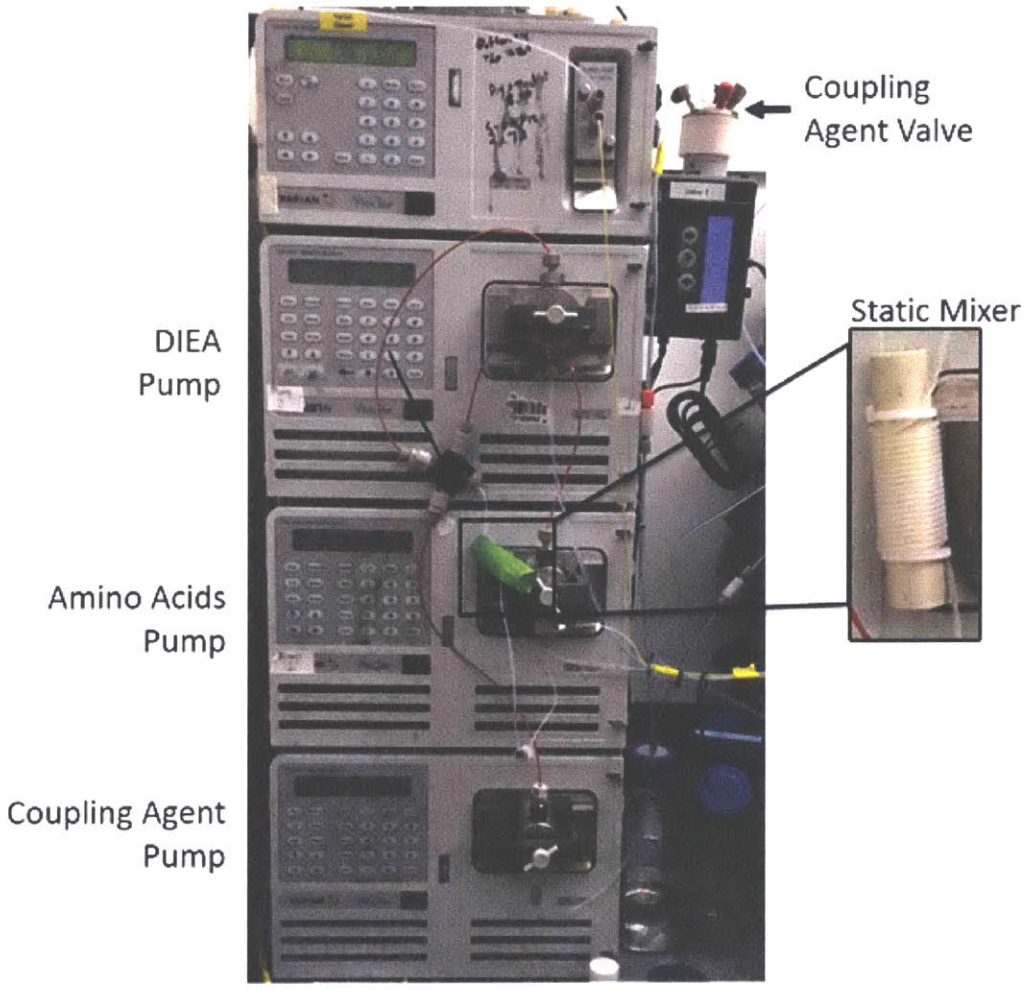

Figure 2-16: Photograph of the AFPS pumping and mixing modules

WSS). The DIEA pump has a 5 mL/min pump head (Agilent part# 5-Ti). The three pumps outlets merge at a cross (IDEX part# P-722) with three inlet check valves (IDEX part# CV-3320) to prevent diffusion between the cross and pump head. The lengths of PEEK tubing (1/16" OD, 0.020" ID) between the PEEK cross and all of the pumps have matched volumes. After the cross, a length FEP tubing (1/16" OD, 0.030"ID) was coiled 22 times around a $\frac{1}{2}$ inch cylinder to form a high dean number (> 3000) static mixer to facilitate reagent mixing. [27]

Activation and Coupling Flow Reactors After mixing, the reagent stream proceeds to a heat exchanger that is selected using a VICI Valco six-position column selector valve (Vici part# ACST6UW-EUTA). These heat exchangers consist of a length of stainless steel tubing wrapped around an aluminum spool and coated with silicone for insulation. The spools are heated with two resistive cartridge heaters (Omega part# CSS-10250/120V). For peptide synthesis method B, a 3m (10 ft, 1.37 mL) heat exchanger loop at 90 °C was used; for peptide synthesis method A, a 1.5m (5 ft, 0.68 mL) heat exchanger loop at 70 °C was used.

The coupling reactor is a two-piece, heated aluminum syringe holder. The top, or the reactor inlet, inserts into a 6 mL HSW NormJect syringe purchased from Torviq. The syringe slides down into and is encased inside of a second heated chuck, with a female luer slip fitting at the bottom. At the bottom of the syringe is a coarse polyethylene frit (9 mm x 2.3 mm, Scientific Commodities part number GP-9094) layered with a Porex polyethylene membrane with a 7-12 um pore size (part# POR-7744). The membrane filters contain the resin inside of the syringe during synthesis.

2.3.6 Process Data Collection

The software records temperature, mass flow rate, pressure, and UV absorbance during each synthesis. The Watlow PID control unit described above was used to acquire temperature data. A Bronkhorst Coriolis mass flow meter (part# M14-XAD-11-0-5) was used to acquire mass flow data and also allowed monitoring of fluid density. The differential pressure across the reactor was monitored using two DJ instrument HPLC through-bore titanium pressure sensors (part# DF2-01-TI-500-5V-41"). These

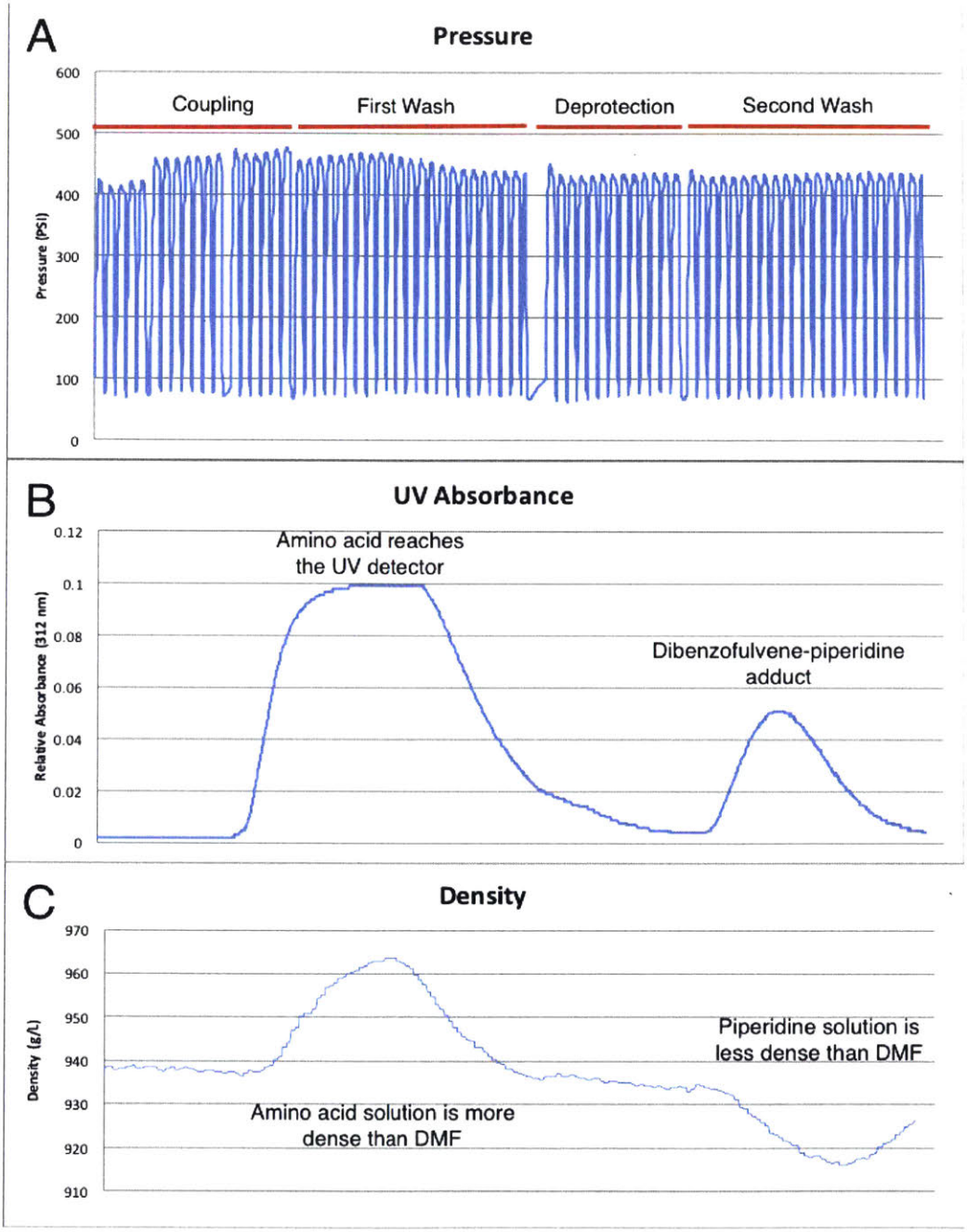

Figure 2-17: Data obtained during a single amino acid coupling cycle. **A.** The pressure trace is oscillatory because the AFPS uses reciprocating HPLC pumps for reagent delivery. The pressure maxima remains mostly constant except for the spike during the increase in flow rate during ue to DIEA infusion. **B.** UV absorbance for amino acid infusion and deprotection are slightly delayed relative to the pressure trace due to system volume. **C.** The Bronkhorst CoriFlow allows for monitoring of density of the fluid stream, serving as a verification of amino acid or piperidine infusion.

sensors were single point calibrated at 90 degrees Celsius and 100 psi.

UV monitoring at 312 nm was accomplished by using a Varian Prostar 230 UV-Vis detector fitted with a super prep dual path length flow cell (nominal path lengths of 4 mm and 0.15 mm). This dual path length flow cell setup allows for high dynamic range absorbance measurements – whenever the absorbance increases past the linear range for the large flow cell, the instrument switches to recording the absorbance through the smaller flow cell. In order to assure accurate measurements during the flow cell switchover, the ratio of path lengths was calculated using a standard solution of dibenzofulvene prepared as described in [28] and programmed into the detector via the front panel.

Temperature and mass flow data were acquired through serial communication with the Watlow PID and Bronkhorst flow meter. Electronic voltage measurements for pressure and UV data were obtained from the instrument using a National Instruments NI cDAQ-9184 (part number 782069-01) with a NI 9205 32-channel analog input card (part number 779357-01). Data points were recorded with averaging every 50 ms. On the UV detector, the signal response time was set to 10 ms and the full voltage scale was 100 mv.

2.3.7 Serial Communication with Pumps and Valves and Arduino Prototyping

We initially prototyped the control system on an Arduino Mega. The pumps and valves were daisy chained and connected to separate TTL serial ports on the Arduino using the RS232 MAX3232 SparkFun Transceiver Breakout (SparkFun part BOB-11189) (Figure 2-18). Standard RS-485 serial protocols were used for communication with the Varian ProStar 210 pumps and VICI Valco valves. Pump communication was at 19200 baud, 8 bit, even parity, with 1 stop bit. Valve communication was at 9600 baud with no parity and one stop bit. Example serial drivers for the pumps and valves written for Arduino are included as part of the supplementary material.

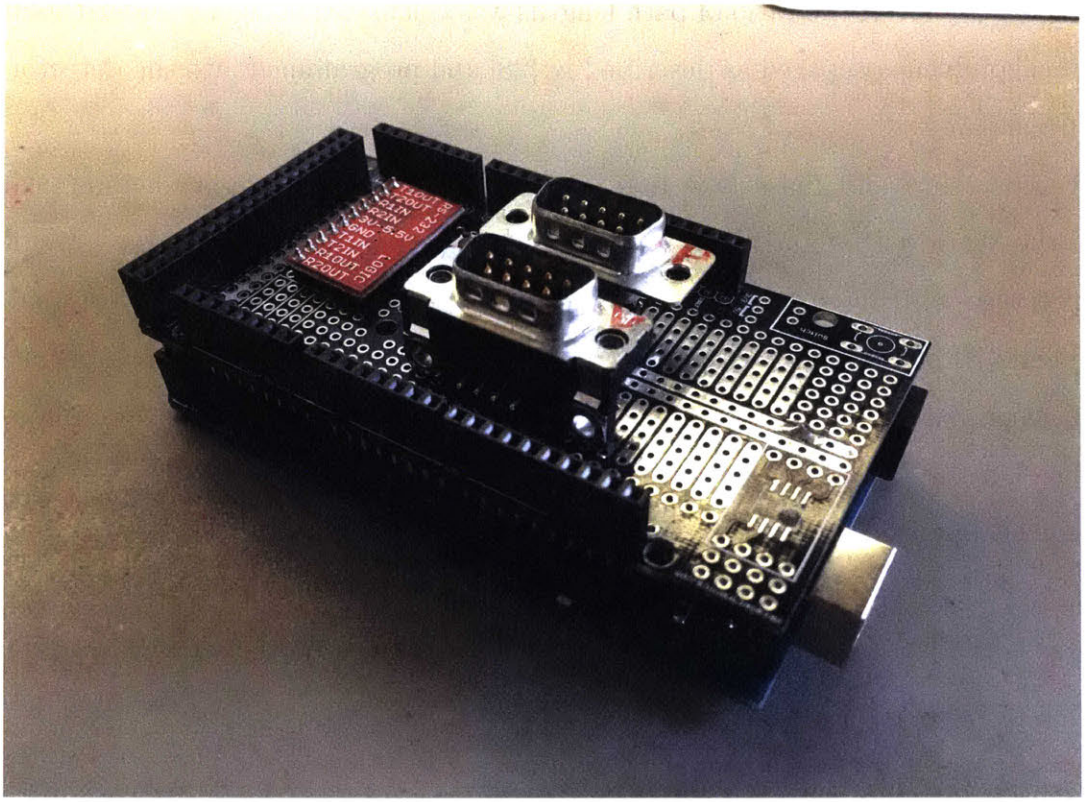

Figure 2-18: Photograph of the arduino with serial adapter board used for the first generation AFPS.

2.3.8 Heating and Temperature Control

Activation reactors and the syringe holder were heated using $\frac{1}{4}$" diameter, 120V/220W cartridge heaters from Omega Engineering (CSH-103220/120V), and the syringe inlet was heated using $\frac{1}{4}$" diameter, 120V/150W (CSS-10120/120V) cartridge heaters from Omega Engineering. All heaters were connected to a standard AC power supply with relays controlled using an 8-channel Watlow EZ-Zone RM controller (part number RMHF-1122-A1AA) with on-board PID. Type-K thermocouples attached to each device were connected to the Watlow controller and read into the software through the RS-232 serial port using software provided by Watlow. All thermocouples were calibrated using a single point calibration at 0 °C.

2.3.9 Software Control and Sequencing.

The synthesizer is controlled over Ethernet and USB on a Windows computer with a LabView VI (Figure 2-20). The VI has a graphical interface to allow a user to create a recipe for the desired peptide. Recipes allow users to control the flow rate, the amount of amino acid used, the activating agent, the temperature and residence time of activation, the deprotection residence time, and the amount of deprotection reagent for each step of the synthesis. Once, the user has created the desired recipe, he or she submits it to the machine queue and presses "Run." If during the synthesis the user notices a change in the synthesis quality by monitoring the UV trace, he or she can modify the recipe for any subsequent coupling step on the fly. When "Run" is pressed, the software populates the predefined routine for each amino acid with the users selected amino acid, flow rates, temperatures, amount of reagents, and type of activating reagent.

The code consists of operations performed on either pumps, valves, or motors. Each operation consists of a set of inputs and a dwell time. Valves accept a valve ID and valve position; pumps accept a pump ID and pump flow rate; motors accept a motor ID and motor position. After a step is complete, the program waits until completion of the dwell time before executing the next step. Dwell times represented

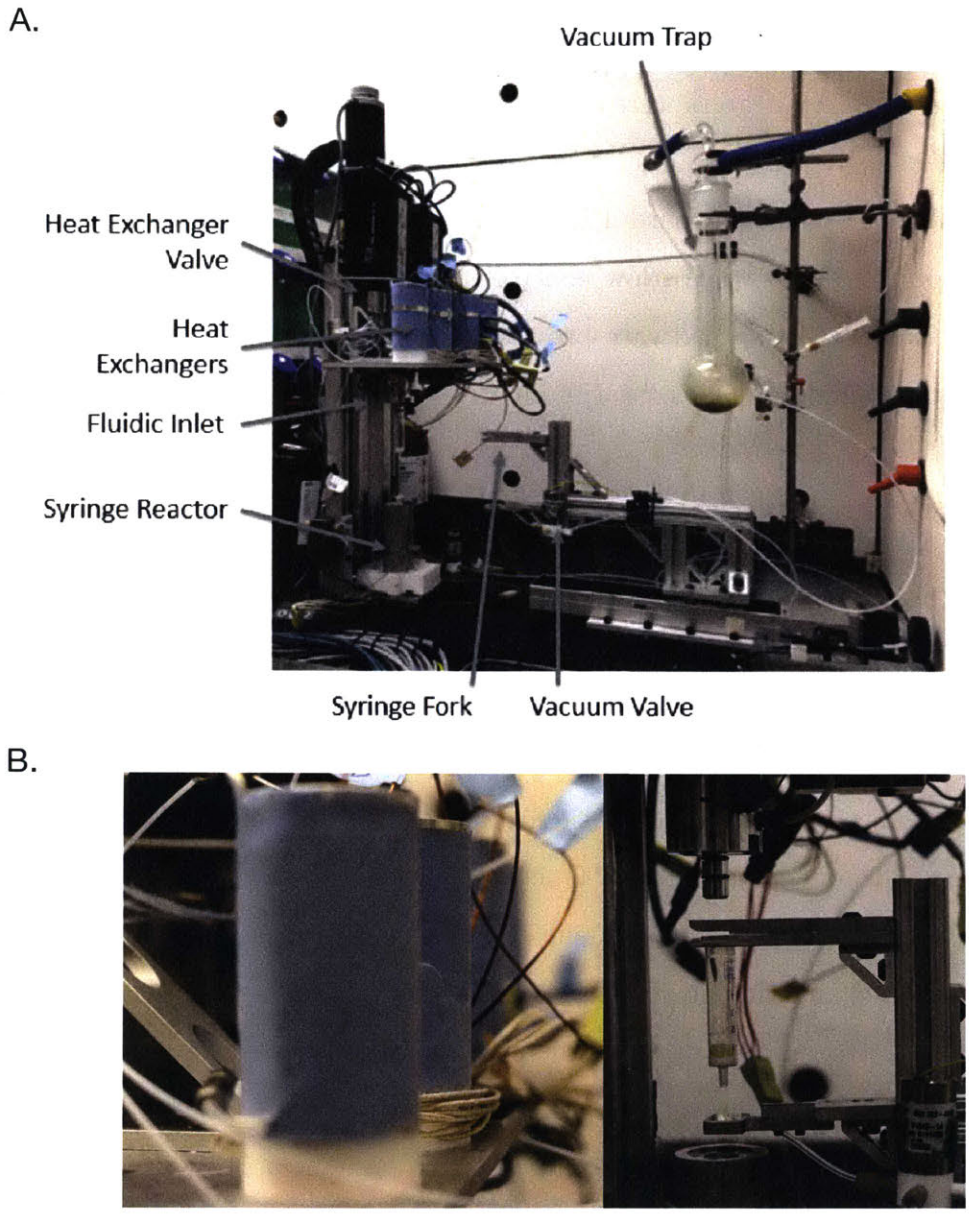

Figure 2-19: **A.** Overview of the reactors on the AFPS. **B.** Photograph of the heated activation reactors (left) and the disposable syringe cartridge with fluidic inlet and jacket (right).

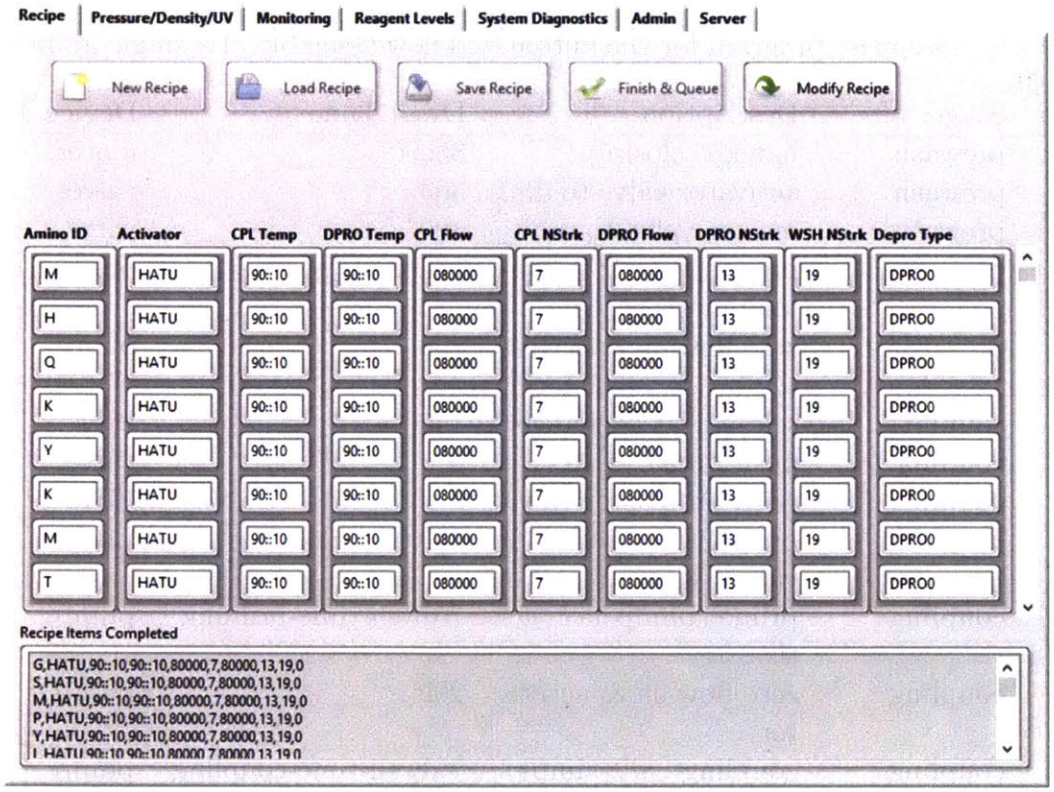

Figure 2-20: LabView interface for the automated flow peptide synthesizer

by #variables are computed on the fly using the recipe input. For instance, the dwell time after actuation of the pumps in step 12 is determined by the "CPL NStrk" (number of coupling strokes) parameter in the recipe and is computed by dividing the number of strokes by the frequency of the pump head oscillations. Table 2.2 shows a program for the assembly of a single amino acid peptide. This sequence is recapitulated in the Arduino code included as supplementary information.

2.3.10 Protocol for Peptide Synthesis on the Automated Flow Peptide Synthesizer

Supplementary Table 2 outlines the standard AFPS method. 200 mg of ChemMatrix PEG Rink Amide resin with a nominal loading of 0.5 mmol/g is loaded into a 6 mL Torviq fritted syringe fitted with an additional 7-12 µm Porex UHMWPE (XS-POR-7474) membrane on top of the frit. The resin is preswollen with DMF for 5 minutes,

Table 2.2: Stepwise program for the automated flow assembly of a single amino acid on resin.

Step	Stage	Description	Dwell Time	Device Type
1	prewash	handler closing	35000	motors
2	prewash	activator valve to dmf	300	valve
3	prewash	master valve to dmf	300	valve
4	prewash	prewash– aa & activator pumps	10000	pump
5	prewash	zero flow aa & activator	200	pump
6	coupling	select heat exchanger	1500	valve
7	coupling	position aa valve	0	valve
8	coupling	position master valve	0	valve
9	coupling	position activator valve	300	valve
10	coupling	prime– pump aa & activator	#dwell-time-priming	pump
11	coupling	zero flow aa & activator	200	pump
12	coupling	coupling– all pumps	#dwell-time-coupling	pump
13	coupling	zero flow all pumps	200	pump
14	coupling	activator valve to dmf	0	valve
15	coupling	aa valve to dmf	300	valve
16	coupling	flush– aa & activator pumps	#dwell-time-priming	pump
17	coupling	zero flow all pumps	200	pump
18	coupling	post coupling wash-aa & activator pumps	#dwell-time-wash	pump
19	coupling	zero flow all pumps	200	pump
20	deprotection	change hx valve pos	0	valve
21	deprotection	master valve to dmf	0	valve
22	deprotection	pip (activator) valve to pip	1500	valve
23	deprotection	depro pump aa & activator	#dwell-time-depro	pump
24	deprotection	zero flow aa & activator	200	pump
25	deprotection	pip (activator) valve to dmf	300	valve
26	deprotection	depro wash pump aa & activator	#dwell-time-wash	pump
27	deprotection	zero flow aa & activator	200	pump
28	postwash	postwash aa & activator	10000	pump
29	postwash	open syringe handler	10000	motors

after which large resin aggregates are manually broken up by inserting the syringe plunger. The syringe is filled with DMF, loaded onto the fluidic inlet, and loaded into a 90 °C heated chamber. The synthesizer is set up as shown in Figure 2-1, with all reagents pumped at a total flow rate of 80 $ml \cdot min^{-1}$ though a cross manifold, a mixer, and a 10 ft stainless steel heated loop at 90 °C before being pumped over the resin. Three Varian Prostar 210 HPLC pumps are used, two with 50 $ml \cdot min^{-1}$ pump heads for amino acid and activating agent, and one with a 5 $ml \cdot min^{-1}$ pump head, for diisopropylethylamine (DIEA). The 50 $ml \cdot min^{-1}$ pump head pumps 400 µL of liquid per pump stroke; the 5 $ml \cdot min^{-1}$ pump head pumps 40 µL of liquid per pump stroke.

The standard synthetic cycle used involves a first step of prewashing the resin at elevated temperatures for 20s at 80 mL/min. During the coupling step, three HPLC pumps are used: a 50 $ml \cdot min^{-1}$ pump head pumps the activating agent (typically 0.34 M HATU), a second 50 $ml \cdot min^{-1}$ pump head pumps the amino acid (0.4M) and a 5 $ml \cdot min^{-1}$ pump head pumps diisopropylethylamine (DIEA). The first two pumps are activated for 5 pumping strokes in order to prime the coupling agent and amino acid before the DIEA pump is activated. The three pumps are then actuated together for a period of 7 pumping strokes, after which the activating agent pump and amino acid pump are switched using a rotary valve to select DMF. The three pumps are actuated together for a final 5 pumping strokes, after which the DIEA pump is shut off and the other two pumps continue to wash the resin for another 16 pump strokes.

During the deprotection step, two HPLC pumps are used. Using a rotary valve, one HPLC pump selects 40% piperidine and the other selects DMF. The pumps are activated for 13 pump strokes. After mixing, the final concentration of piperidine is 20%. Next, the rotary valves select DMF for both HPLC pumps, and the resin is washed for an additional 16 pump strokes. The coupling/deprotection cycle is repeated for all additional monomers.

2.3.11 Analytical Peptide Cleavage and Side Chain Protecting Group Removal

Approximately 10 mg of peptidyl resin was added to a 1.5 mL Eppendorf tube. 200 µL of cleavage solution (94% TFA, 1% TIPS, 2.5% EDT, 2.5% water) was added to the tube and incubated at 60 °C for 5 minutes. After completion of cleavage, 200 µL TFA was added to the tube to rinse the resin, and as much liquid as possible was transferred into another tube using a pipet tip, avoiding resin. To the tube of cleavage solution, 800 µL cold diethyl ether was added. The tube was shaken – a visible waxy precipitate formed and was collected by centrifugation. The supernatant ether was poured off and two more ether washes were performed.

Finally, the waxy solid was allowed to dry briefly under a stream of nitrogen gas. 500 µL of 50% acetonitrile in water was added to the tube and mixed thoroughly. This solution was filtered through a centrifugal basket filter and diluted 1:10 in 50% acetonitrile in water with 0.1% TFA for the liquid chromatographic analysis.

2.3.12 Preparative Peptide Cleavage

After synthesis, peptidyl resin was washed with dichloromethane, dried in a vacuum chamber, and weighed. The resin was transferred into a 15 mL conical polypropylene tube. Approximately 7 mL of cleavage solution (94% TFA, 1% TIPS, 2.5% EDT, 2.5% water) was added to the tube. More cleavage solution was added to ensure complete submersion. The tube was capped, inverted to mix every half hour, and was allowed to proceed at room temperature for 2 hours.

Then, the resin slurry was filtered through a 10 µm polyethylene membrane disk fitted into a 10 mL Torviq syringe. The resin was rinsed twice more with 1 mL TFA, and the filtrate was transferred into a 50 mL polypropylene conical tube. 35 mL ice cold diethyl ether were added to the filtrate and left to stand for 30 minutes to precipitate the peptide. The precipitate was collected by centrifugation and triturated twice more with 35 mL cold diethyl ether. The supernatant was discarded.

Finally, residual ether was allowed to evaporate and the peptide was dissolved in

50% acetonitrile in water. The peptide solution was frozen, lyophilized until dry, and weighed.

2.3.13 Analytical Liquid Chromatographic Analysis of Peptide Samples

Figure 2-1 and Figure 2-11: 1µL of the diluted peptide sample with a concentration approximately 200 mg/mL was analyzed on an Agilent 6520 LC-MS with a Zorbax 300SB-C3 column (2.1 mm x 150 mm, 5µm particle size). A gradient of acetonitrile in water with a 0.1% formic acid additive was used. Gradients started at 5% acetonitrile, and after a two-minute hold, ramped to 65% acetonitrile at a rate of 5% acetonitrile per minute. A final two-minute hold was performed. The total method time was 14 minutes.

Figures 2-3 and 2-6: 1µL of the peptide sample with a concentration approximately 200 mg/mL was analyzed on an Agilent 6520 LC-MS with a Zorbax 300SB-C3 column (2.1 mm x 150 mm, 5µm particle size). A gradient of acetonitrile in water with a 0.1% formic acid additive was used. Gradients started at 5% acetonitrile, and after a two-minute hold, ramped to 65% acetonitrile at a rate of 1% acetonitrile per minute. A final two-minute hold was performed. The total method time was 64 minutes.

2.3.14 Synthesis of Reduced Loading Resin and JR 10-mer Loading Study

Reduced loading resins were prepared by coupling the first amino acid, Fmoc-Met-OH, to 200 mg of ChemMatrix Rink Amide HR resin, which has a nominal loading of 0.45 mmol/g. For one batch of resin, 1 mmol of Fmoc-Met-OH was mixed with 2.5 mL of a 0.4M HBTU, 0.4M HOBT solution and dissolved. 500 µL of DIEA was added, mixed thoroughly, and then added to the resin. Coupling was allowed to proceed for 1 hr. For the other four batches of resin, acetic acid was substituted stoichiometrically for Fmoc-Met-OH to reach the desired loadings: 90%, 80%, 70%, 60%, 50%, and 25% of the original loading. The relative loadings used in Figure 2-6 were computed using

the first deprotection integral.

Then, on the automated flow peptide synthesizer, the sequence WFTTLISTIM was synthesized on each of these resins. The conditions were: 90 °C for the heat exchanger, reactor inlet, and reactor body. 80 mL/min coupling/deprotection, 7 coupling strokes, 13 deprotection strokes, 24 washing strokes. The first coupling – M– added nothing to the resin, as the resin was still fully protected. The first deprotection gave a baseline UV absorbance and allowed estimation of the loading obtained during the capping step.

2.3.15 Manual Synthesis of JR 10-mer, Insulin B chain, and GHRH

These peptides were synthesized following protocols from the literature. [29] ChemMatrix Rink-amide resin (0.1 mmol; 0.45 mmol/g) was used. Amino acids were activated for 30 seconds by first dissolving 0.55 mmol of the amino acid to be coupled in 1.25 mL 0.4 M HBTU/0.4 M HOBT, and then adding 122 µL (0.7 mmol) of DIEA. After 30 seconds, the solution was added to the resin. The couplings were allowed to proceed for 30 minutes with intermittent stirring.

After each coupling step, a 45 mL DMF flow wash was performed. Then, 3 mL of 20% (v/v) piperidine was added to the resin, stirred, and allowed to incubate for 5 minutes. This process was repeated once. After each deprotection, a 45 mL flow wash was performed, followed by a 1-minute batch treatment with DMF.

2.4 Acknowledgements

This research was supported by startup funds from the MIT Department of Chemistry for B.L.P., the MIT Deshpande Center for Technological Innovation, and Dr. Reddy's Laboratories Limited and a National Science Foundation Graduate Student Fellowship for A.M. We also thank Dr. Andrew Teixeria, Ethan Evans, Dr. Guillaume Lautrette, Justin Wolfe, Sonam Jain, Sung Kim, and Adam J. West for expert

technical assistance, suggestions, and as early adopters of the technology.

References

1. Albericio, F. & Kruger, H. G. Therapeutic Peptides. *Future Medicinal Chemistry* **4,** 1527–1531. ISSN: 1756-8919 (Aug. 2012).

2. Carter, C. F. *et al.* ReactIR Flow Cell: A New Analytical Tool for Continuous Flow Chemical Processing. *Organic Process Research & Development* **14,** 393–404. ISSN: 1083-6160 (Mar. 2010).

3. Baxendale, I. R. The Integration of Flow Reactors into Synthetic Organic Chemistry. en. *Journal of Chemical Technology & Biotechnology* **88,** 519–552. ISSN: 1097-4660 (Apr. 2013).

4. McQuade, D. T. & Seeberger, P. H. Applying Flow Chemistry: Methods, Materials, and Multistep Synthesis. *The Journal of Organic Chemistry* **78,** 6384–6389 (2013).

5. Glasnov, T. N. & Kappe, C. O. The Microwave-to-Flow Paradigm: Translating High-Temperature Batch Microwave Chemistry to Scalable Continuous-Flow Processes. *Chemistry - A European Journal* **17,** 11956–11968 (2011).

6. Yoshida, J.-i., Takahashi, Y. & Nagaki, A. Flash Chemistry: Flow Chemistry That Cannot Be Done in Batch. en. *Chem. Commun.* **49,** 9896–9904. ISSN: 1359-7345, 1364-548X (2013).

7. Adamo, A. *et al.* On-Demand Continuous-Flow Production of Pharmaceuticals in a Compact, Reconfigurable System. eng. *Science (New York, N.Y.)* **352,** 61–67. ISSN: 1095-9203 (Apr. 2016).

8. Dryland, A. & Sheppard, R. C. Peptide Synthesis. Part 8. A System for Solid-Phase Synthesis under Low Pressure Continuous Flow Conditions. en. *Journal of the Chemical Society, Perkin Transactions 1,* 125–137. ISSN: 1364-5463 (Jan. 1986).

9. Dryland, A. & Sheppard, R. C. Peptide Synthesis. Part 11. A System for Continuous Flow Solid Phase Peptide Synthesis Using Fluorenylmethoxycarbonyl-Amino Acid Pentafluorophenyl Esters. *Tetrahedron* **44,** 859–876 (1988).

10. Rapp, W. & Bayer, E. *Uniform Microspheres in Peptide Synthesis: Ultrashort Cycles and Synthesis Documentation by on-Line Monitoring as an Alternative to Multiple Peptide Synthesis* in (1993), 25.

11. Simon, M. D. *et al.* Rapid Flow-Based Peptide Synthesis. en. *ChemBioChem* **15,** 713–720. ISSN: 1439-7633 (Mar. 2014).

12. Mong, S. K., Vinogradov, A. A., Simon, M. D. & Pentelute, B. L. Rapid Total Synthesis of DARPin pE59 and Barnase. en. *ChemBioChem* **15,** 721–733. ISSN: 1439-7633 (Mar. 2014).

13. Yu, H. M., Chen, S. T. & Wang, K. T. Enhanced Coupling Efficiency in Solid-Phase Peptide Synthesis by Microwave Irradiation. *The Journal of Organic Chemistry* **57,** 4781–4784 (1992).

14. El-Faham, A. & Albericio, F. Peptide Coupling Reagents, More than a Letter Soup. *Chemical Reviews* **111,** 6557–6602 (2011).

15. Chan, W. C. & White, P. D. *Fmoc Solid Phase Peptide Synthesis* English. ISBN: 978-0-19-963725-6 (Oxford University Press, 2000).

16. Collins, J. M., Porter, K. A., Singh, S. K. & Vanier, G. S. High-Efficiency Solid Phase Peptide Synthesis (HE-SPPS). *Organic Letters* **16,** 940–943. ISSN: 1523-7060 (Feb. 2014).

17. Johnson, T., Quibell, M., Owen, D. & C. Sheppard, R. A Reversible Protecting Group for the Amide Bond in Peptides. Use in the Synthesis of 'Difficult Se-

quences'. en. *Journal of the Chemical Society, Chemical Communications*, 369–372 (1993).

18. Wöhr, T. & Mutter, M. Pseudo-Prolines in Peptide Synthesis: Direct Insertion of Serine and Threonine Derived Oxazolidines in Dipeptides. *Tetrahedron Letters* **36,** 3847–3848. ISSN: 0040-4039 (May 1995).

19. Kent, S. B. H. Chemical Synthesis of Peptides and Proteins. *Annual Review of Biochemistry* **57,** 957–989 (1988).

20. Schnölzer, M., ALEWOOD, P., JONES, A., ALEWOOD, D. & KENT, S. B. H. In Situ Neutralization in Boc-Chemistry Solid Phase Peptide Synthesis. *International Journal of Peptide and Protein Research* **40,** 180–193 (2009).

21. Sarin, V. K., Kent, S. B. H., Tam, J. P. & Merrifield, R. B. Quantitative Monitoring of Solid-Phase Peptide Synthesis by the Ninhydrin Reaction. *Analytical Biochemistry* **117,** 147–157 (1981).

22. Kaiser, E., Colescott, R. L., Bossinger, C. D. & Cook, P. I. Color Test for Detection of Free Terminal Amino Groups in the Solid-Phase Synthesis of Peptides. *Analytical Biochemistry* **34,** 595–598 (1970).

23. Carpino, L. A. *et al.* Synthesis of 'Difficult' Peptide Sequences: Application of a Depsipeptide Technique to the Jung–Redemann 10- and 26-Mers and the Amyloid Peptide Aβ(1–42). *Tetrahedron Letters* **45,** 7519–7523. ISSN: 0040-4039 (Sept. 2004).

24. Hjørringgaard, C. U., Brust, A. & Alewood, P. F. Evaluation of COMU as a Coupling Reagent for in Situ Neutralization Boc Solid Phase Peptide Synthesis. en. *Journal of Peptide Science* **18,** 199–207. ISSN: 1099-1387 (Mar. 2012).

25. Barany, G. & Merrifield, R. B. in *The Peptides, Analysis, Synthesis, Biology, Volume 2: Special Methods in Peptide Synthesis, Part A* (eds Gross, E. & Meienhofer, J.) 1–284 (Academic Press, New York, 1980). ISBN: 0-12-304202-X.

26. Varanda Leandro M. & Miranda M. Terêsa M. Solid-phase Peptide Synthesis at Elevated Temperatures: A Search for an Optimized Synthesis Condition of Unsulfated Cholecystokinin-12. *The Journal of Peptide Research* **50,** 102–108. ISSN: 1397-002X (Jan. 2009).

27. Jiang F., Drese K. S., Hardt S., Küpper M. & Schönfeld F. Helical Flows and Chaotic Mixing in Curved Micro Channels. *AIChE Journal* **50,** 2297–2305. ISSN: 0001-1541 (Aug. 2004).

28. Gude, M., Ryf, J. & White, P. D. An Accurate Method for the Quantitation of Fmoc-Derivatized Solid Phase Supports. en. *Letters in Peptide Science* **9,** 203–206. ISSN: 0929-5666, 1573-496X (July 2002).

29. Dang, B., Dhayalan, B. & Kent, S. B. H. Enhanced Solvation of Peptides Attached to "Solid-Phase" Resins: Straightforward Syntheses of the Elastin Sequence Pro-Gly-Val-Gly-Val-Pro-Gly-Val-Gly-Val. *Organic Letters* **17,** 3521–3523. ISSN: 1523-7060 (July 2015).

Chapter 3

Synthesis of S-glycosylated cysteine monomers and a di-glycosylated, 45-mer antibacterial protein

3.1 Abstract

Cysteine S-glycosylation is a naturally-occurring modification important for protein function in not only prokaryotes, but also eukaryotes. Because S-glycosylated proteins are generally resistant to enzymatic cleavage and hydrolysis, this sulfur linkage is also of interest in the development of new therapeutics. However, the cost and difficulty of accessing peptides and proteins with S-glycosylated amino acids has limited the exploration of the S-glycome and potential therapeutic targets. Here we report a single-step flow synthesis of N-acetylhexosamine cysteine from low-cost, commercially-available starting materials and conditions to incorporate these monomers into polypeptides in less than one minute while avoiding elimination reactions. These methods are robust and permit the complete synthesis of the difficult-to-synthesize antimicrobial peptide glycocin-F without intermediate ligation steps. We used these methods to generate a number of variants of glycocin-F to probe the effect of S-linked carbohydrate modifications on the bioactivity of the parent peptide.

3.2 Introduction

Cysteine S-glycosylation is a naturally-occurring, post-translational modification first discovered in bacterial antimicrobial peptides.[1] S-GlcNAcylated proteins were recently isolated from mouse brain tissue, indicating that S-glycosylated proteins may also play an important role in eukaryotic cellular function. [2] While more than 500 putative S-glycosylated sequences were found in a metabolic labeling study in human tissues using peracetylated N-azidoglucosamine [3], the magnitude of S-linked glycosylation in humans is still unclear since the finding that the peracetylated sugars used in such studies react in a nonenzymatic fashion with cysteine thiols in cells. [4] These proteomics experiments could be validated by the preparation of authentic S-GlcNAcylated polypeptide standards; however, these peptides are still expensive to make, requiring multi-step synthesis to prepare the monomers even before a potentially difficult peptide coupling. Even with the monomer in hand, these methods yield short polypeptides at best; synthesis of the 45-mer Glycocin-F, for instance, requires two or three peptide starting materials and one or two convergent ligation steps with purifications between each step. [5–7]

The S-glycosyl linkage has notable chemistry compared to the O-glycosyl linkage, suggesting different biological roles for these molecules. First, S-glycosylation is stable to a wide array of O-glycosidase enzymes, a counterpoint to the much more dynamic O-glycosylation which is frequently remodeled in vivo.[8, 9] In addition to enzymatic stability, S-linked glycans are also stable to general acid and base conditions, making this linkage a potentially attractive option for delivering bioactive, glycosylated cargo into harsh environments, such as the stomach. [10, 11] And despite the slight difference in bond angles at the sulfur and oxygen atoms, S-glycosylated peptides have similar activity to their O-glycosylated counterparts. For example, variants of Glycocin-F that were prepared with two S-linked GlcNAc residues were more active than the O/S-glycosylated natural product.[5] Furthermore, in a study on α-synuclein, substitutiton of the O-GlcNAc serine for S-GlcNAc cysteine did not affect the secondary structure of the protein or the inhibition of protein aggregation

conferred by the GlcNAc. [8]

Two examples of naturally-occurring S-glycosylated polypeptides are the disulfide-linked bacteriocidins Glycocin-F (GcF) and Thurandacin-B (Thd)[1, 12]. These molecules are unique in that they feature both an O- and an S-linked GlcNAc in the same molecule. In the case of Thurandancin-B, a single glycosyl transferase enzyme was reported to catalyze both the O- and the S-GlcNacylation, with the S-glycosylation reaction occurring faster than the O-glycosylation. The existence of promiscuous S-glycosyl transferrases, combined with the more nucleophilic nature of the cysteine thiol, suggests that this protein modification may be present in other organisms and proteins. However, the S-glycome is still poorly understood.

Here, we report a fast and efficient method for the preparation of S-glycosylated polypeptides and miniproteins comprising two advances in S-glycosylation chemistry. First, we demonstrate a single-step synthesis of Fmoc-protected, S-HexNAcylated monomers in both batch and continuous flow modes, avoiding protecting group shuffling and using inexpensive, commercially-available starting materials. Second, we report a standard, automated method for incorporation of these glycosylated monomers onto polypeptides in less than one minute per synthetic cycle in flow. Crucially, we found that the glycosyl linkage is stable to the conditions of heated solid phase peptide synthesis. Finally, using both of these methods, we demonstrate the synthesis of glycosylated analogues of the 45-amino acid protein Glycocin-F without ligation steps and discuss the associated synthetic challenges.

3.3 Results and Discussion

Generally, $InBr_3$ catalysts are used for thiol glycosylation with glycosyl donors containing a variety of leaving groups at the anomeric center, including halogens [13] and acetate [14]. Previous reported syntheses of Fmoc-protected, S-glycosylated monomers vary in strategy but generally are around 4 steps from starting materials. One strategy involves formation of the acyl-protected thioglycoside which is then reacted with a protected β-bromoalanine derivative to form the cysteine. [10] An-

other strategy involves direct glycosylation of Fmoc-Cys-OH with N-Troc-protected glucosamine tetraacetate in the presence of an $InBr_3$ catalyst. [8] The free acid is then converted into a pentafluorophenyl ester before Troc removal to avoid formation of a cyclic product. A similar route by Hojo, et al., affords the glycosylated cysteine beginning with N-α-allyl cysteine which is glycosylated with glucosamine pentaacetate. [15]

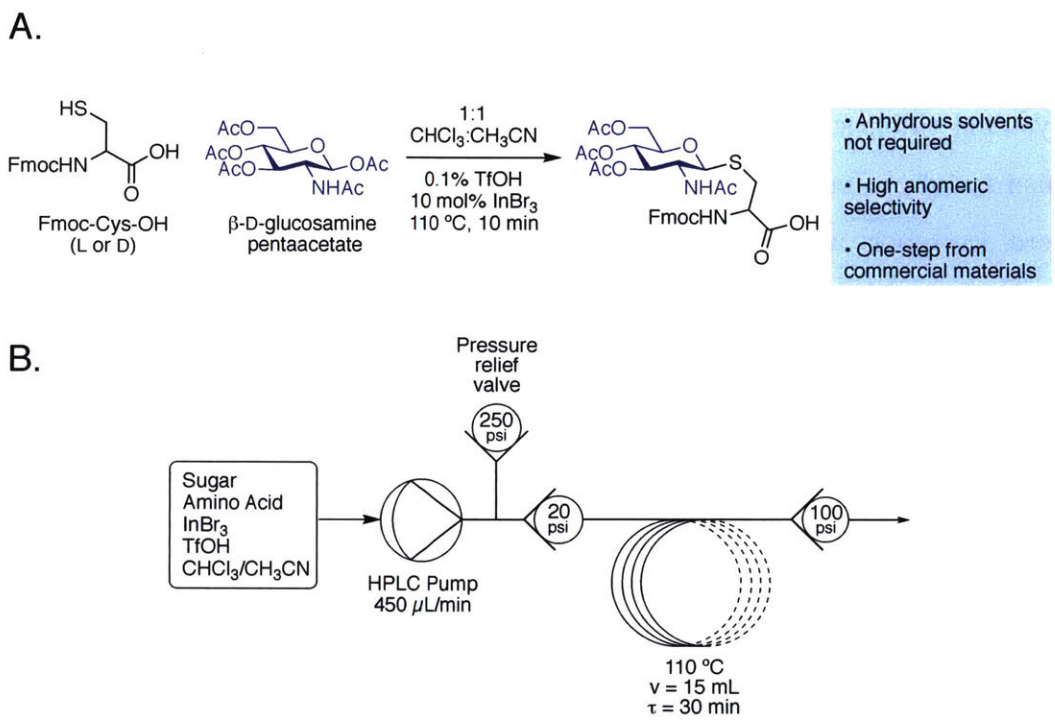

Figure 3-1: Synthesis of S-glycosylated cysteine hexosamine monomers for use in solid phase peptide synthesis. **A.** Single-step synthesis of Fmoc-protected glycosylated amino acids without an acid protecting group. Products are generated in >99% anomeric selectivity without special considerations for keeping the system anhydrous. **B.** Continuous flow apparatus used for the synthesis.

Given that typical SPPS couplings generally require ~0.5 mmol of monomer per coupling, our main considerations in developing a synthesis for glycosylated cysteine polypeptides were monomer cost and ease-of-production. We avoided the thioglycoside route because of its length and complexity. We also drew inspiration from the Pratt and Hojo syntheses, opting for direct glycosylation of cysteine while trying to preserve the free acid, instead of making a pre-formed pentafluorophenyl ester, for performing HATU-mediated couplings on the automated flow peptide synthesizer.

Furthermore, we wanted to start from glucosamine pentaacetate for cost purposes and avoid acid protecting groups. We found that refluxing Fmoc-Cys-OH in chloroform for 48 hours in the presence of peracetylated GlcNAc, ManNAc, or GalNAC with catalytic $InBr_3$ and TfOH yields the desired product in a single step in high anomeric excess (Figure 3-1).

We also explored continuous flow techniques in order to find conditions to speed up this reaction and provide a possible route for automated synthesis and purification. Continuous flow reactors make superheating reagents and performing reactions under pressure much safer and easier than in batch, and we hypothesized that superheating the reactants would greatly speed up this chemistry. The desired product Fmoc-L-Cys-GlcNAc$(OAc)_3$ was obtained in a single step by dissolving glucosamine pentaacetate, Fmoc-Cys-OH, 15 mol% $InBr_3$ and 0.1% TfOH in a 50/50 mixture of chloroform and acetonitrile and pumping the reagents through a 110 °C heated loop reactor with 30 minute residence time at 100 psi. The crude product was collected at the end of the reactor and analyzed by reverse-phase LC-MS, indicating formation of a product with the calculated mass and with starting materials as the major byproducts (Figure 3-2). The crude material was purified by flash chromatography and obtained in 35% yield. Both TfOH and $InBr_3$ are required for this reaction to proceed. Using acetonitrile as a cosolvent maintains the product in solution, while in neat chloroform it otherwise gels up in the tubing. In continuous flow mode, a bench-scale 12-mL loop reactor with HPLC pump can generate around 7 grams of purified product in 24 hours.

After developing a scalable process for S-glycosylated monomer production, we began incorporating these monomers into polypeptides using automated flow solid phase methodology at elevated temperatures. [16] Under these conditions, peptide-to-peptide conditions are highly reproducible and amino acid couplings are usually complete in under 10 seconds. Using a test peptide FGC*GLLKNK, where C* is Fmoc-Cys(GlcNAc)-OH, we scanned a range of temperatures and flow rates, observing that coupling is robust at 70 °C and complete with 4.5 eq. of the glycosylated monomer. We did not observe epimerized products which have been reported during

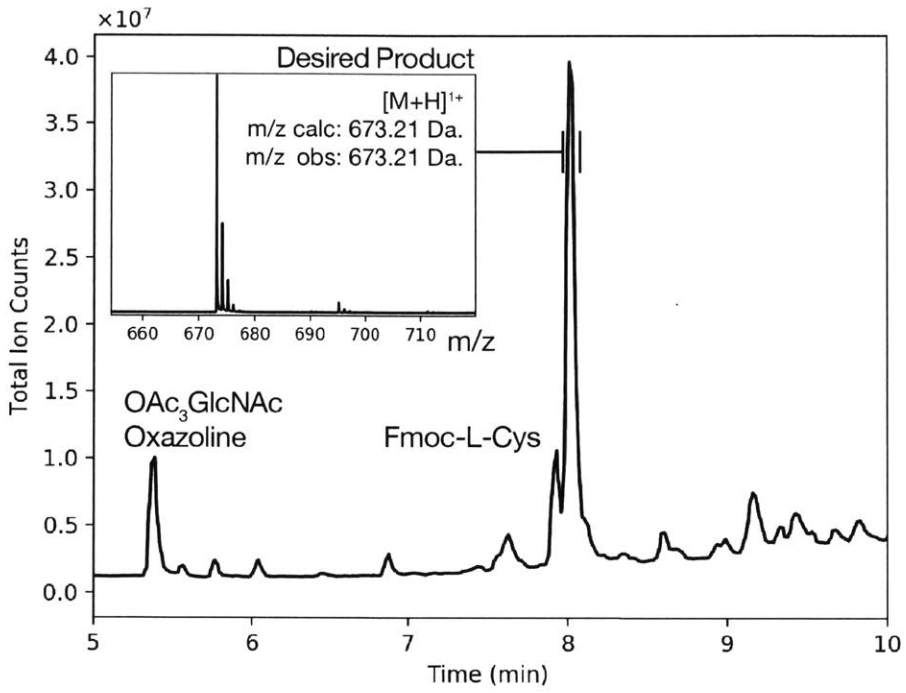

Figure 3-2: LC-MS chromatogram of crude Fmoc-L-Cys(β-$OAc_3GlcNAc$)-OH

coupling of O-glycosylated serine. Concerned about the potential for elimination of the sugar during extended treatment with piperidine, we treated this test peptide with 20% piperidine in DMF for 5 minutes at 90 °C under continuous flow. We observed no degradation of the polypeptide under these conditions. We achieved cleavage of the acyl protecting groups on the sugar by treatment with 10% (v/v) hydrazine hydrate at room temperature in DMF for 30 minutes (Figure 3-3.

These procedures are extensible to synthesis of longer glycosylated polypeptides. As a proof-of-concept, we attempted the one-shot synthesis of variants of the 43-residue glycoprotein Glycocin-F to understand the role of the sugar on its antimicrobial activity. Previous reported syntheses of this glycopeptide required three peptide fragments, multiple backbone protecting groups to prevent peptide aggregation in the fragments, and problematic epimerization of a C-terminal histidine residue of one of the fragment thioesters during ligation. [5, 6] We thought that the automated flow peptide synthesizer, with heat and in-process UV monitoring, could help overcome challenges in the synthesis of this and similar glycoproteins.

We began by synthesizing the non-glycosylated variant of Glycocin-F (with Ser-

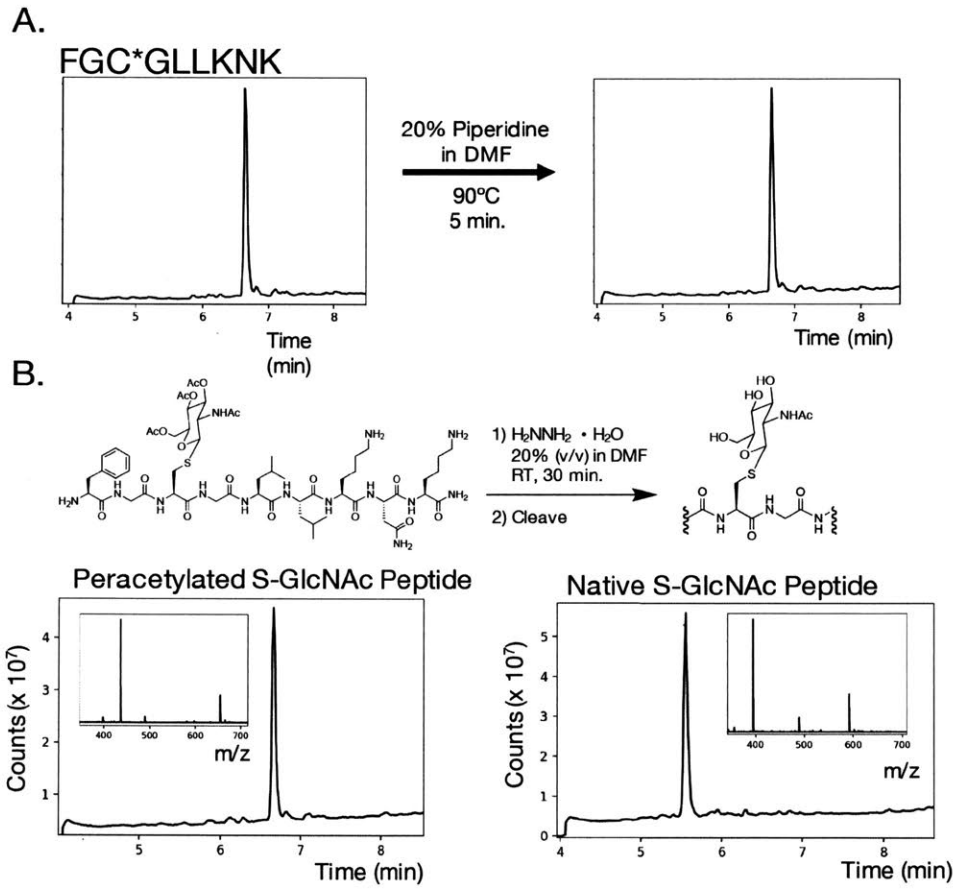

Figure 3-3: Glycosylated amino acids are compatible with high-temperature, flow SPPS conditions. **A.** The polypeptide FGC*GLLKNK was synthesized using a standard flow method for incorporating the glycosylated amino acid. This polypeptide is stable to repeated piperidine treatments at 90 °C without deglycosylation or cleavage of acyl protecting groups. **B.** Acyl protecting groups are quantitatively removed with hydrazine treatment at room temperature. C* = L-Cys(GlcNAc-$(OAc)_3$)

ine instead of glycosylated cysteine residues) to understand its synthesis. Immediately, UV monitoring of the Fmoc removal steps revealed severe peptide aggregation around residue 30 leading to N-terminal truncations and deletion products during the subsequent couplings (Figure 3-4-A). Guided by this information, we incorporated a dimethoxybenzyl (DMB)-protected glycine immediately before the aggregating region. This single backbone protecting group eliminated this aggregation and drastically improved the purity of the crude polypeptide (Figure 3-4-B).

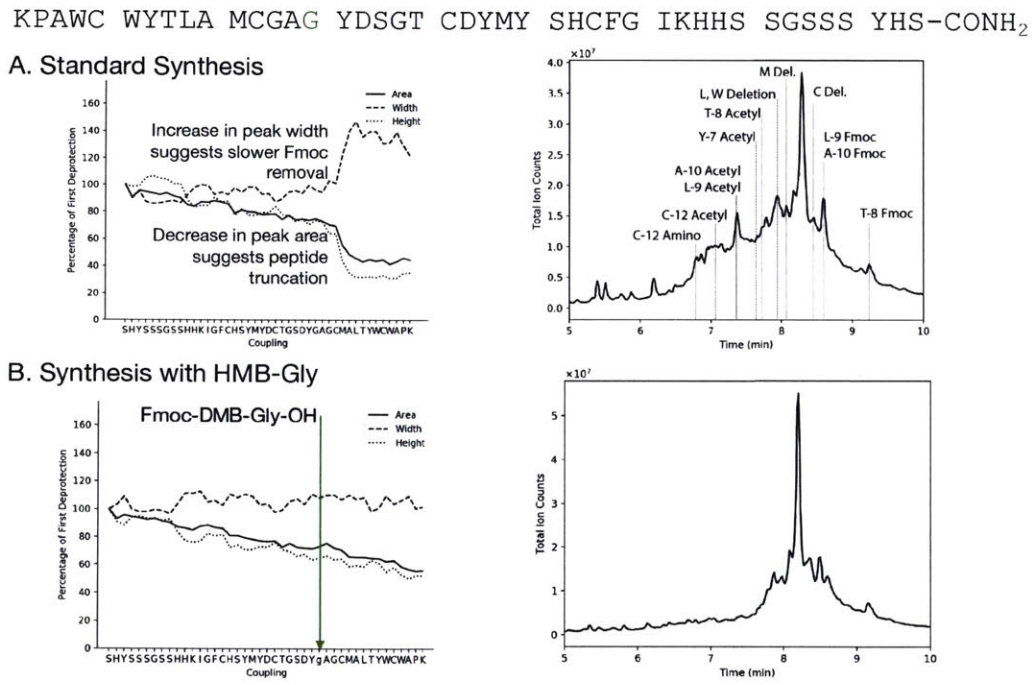

Figure 3-4: Data-driven installation of DMB backbone protecting group enables the one-shot synthesis of Glycocin-F Aglycone. **A.** Glycocin-F-Aglycone was synthesized in one-shot using standard automated flow methods at 90 °C. Significant aggregation was observed during the last 14 residues that corresponded to significant deletions and truncations in the LC-MS chromatogram. **B.** A single backbone-protected glycine residue was installed before this aggregating region that substantially improved the quality of the crude polypeptide.

After optimization of the crude polypeptide synthesis, we began incorporating Fmoc-protected S-glycosylated amino acids. After hydrazine treatment, the C-terminal GlcNAc-containing peptide was cleaved and isolated. Next, both the non-glycosylated and glycosylated variants were purified and subjected to oxidative folding conditions.

In both cases, one predominant folded isomer with two disulfide bonds was obtained; however, the presence of the C-terminal glycosylation site drastically increased the rate of disulfide bond formation. For the non-glycosylated polypeptide, some amount of product with just one disulfide bond was obtained.

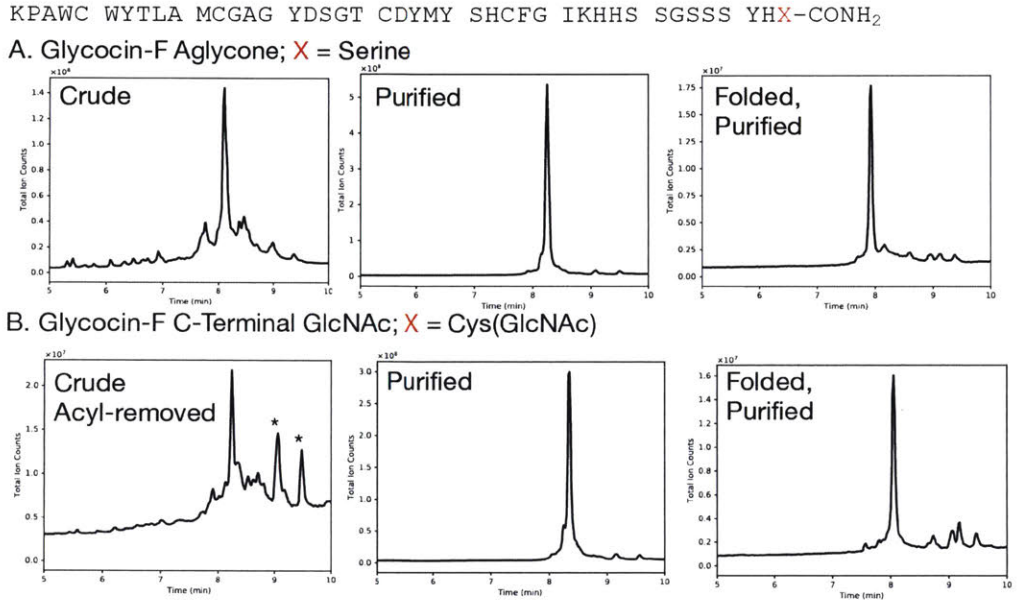

Figure 3-5: Synthesis of glycosylated variants of an antimicrobial polypeptide Glycocin-F. A. Synthesis of Glycocin-F aglycone. This synthesis can be achieved using elevated temperature and resin amino substitution of 0.05 mmol/g. B. HPLC-ESI-MS chromatograms of mono-S-glycosylated Glycocin-F with GlcNAc. Before cleavage and analysis, this peptide was treated with 10% (v/v) hydrazine hydrate in DMF to remove the acyl protecting groups on the sugar.

Cleavage and purification of the di-GlcNAc sequence was substantially more difficult. In this sequence the second glycosylated monomer is positioned next to an aspartic acid. While the non-glycosylated and singly glycosylated material tolerated on-resin hydrazine treatment without byproduct formation, treatment of the di-glycosylated material led to equal formation of a +14 Da. product, indicating the presence of an acid hydrazide. In order to avoid this side reaction, for the di-ManNAc peptide, we first cleaved the peptide from the resin, yielding the crude peracetylated di-ManNAc product along with a singly deacetylated byproduct. The crude peptide was dissolved in 6M guanidinium chloride to which 5% hydrazine was added, and

after an hour at 4 °C diluted with a cysteine/cystine buffer at pH 8.2 to afford the folded product. We are currently preparing the rest of the constructs and biological studies are underway.

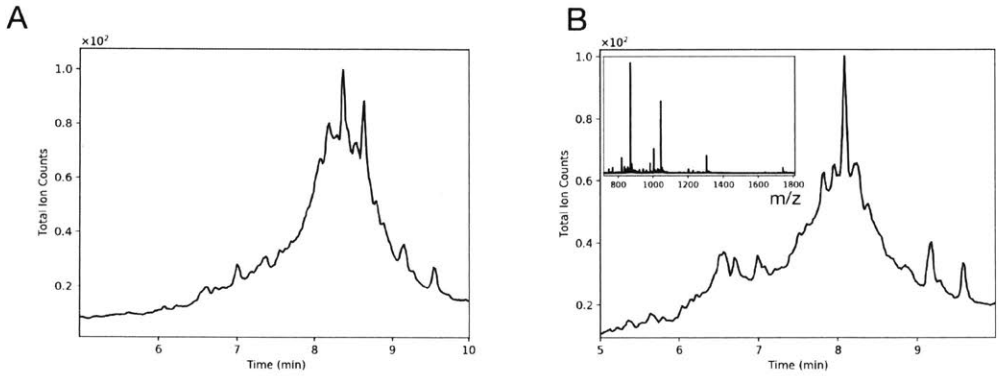

Figure 3-6: Synthesis of di-ManNAc Glycocin-F. **A.** Crude, reduced, peracetylated di-ManNAc Glycocin-F was obtained following global cleavage and deprotection to yield the desired product along with the singly de-acetylated material. **B.** Acyl groups of peracetylated di-ManNAc Glycocin-F were removed before oxidative folding to yield the folded di-ManNAc Glycocin F. m/z calc: 4,345.05 Da; m/z obs: 4,305.05 Da.

In conclusion, we have demonstrated continuous flow chemistry to quickly and scalably generate S-glycosylated monomers for incorporation into polypeptides. We used these monomers to prepare a difficult-to-synthesize 45-mer glycopeptide in a single synthetic procedure, paving the way for exploration of the unique properties of the S-glycosyl linkage.

3.4 Materials and Methods

3.4.1 Flow Synthesis of Fmoc-L-Cys(β-GlcNAc(OAc)$_3$)-OH

Fmoc-L-Cysteine (1.87g, 5.45 mmol, 1.2 eq.) and glucosamine pentaacetate (1.73g 4.43 mmol, 1 eq.) were dissolved in 90 mL of a 1:1 mixture of chloroform and acetonitrile with gentle heating. To this mixture, InBr3 (420 mg, 1.18 mmol, 0.2 eq.) and 90 µL methanesulfonic acid were added. The homogenous reaction mixture was infused at 450 µL/min into a 600 cm tube reactor (perfluoroalkane (PFA); OD =

3.8 mm; ID = 1.6 mm) immersed in a 110 ^{o}C oil bath and fitted with a 100 PSI backpressure regulator at the outlet and a 250 psi safety release valve at the inlet. The crude reaction mixture was collected at the end as a light brown solution.

The reaction mixture was evaporated to dryness, dissolved in 20% acetonitrile in dichloromethane with 1% acetic acid, and purified by column chromatography using a gradient of 20%-60% acetonitrile in dichloromethane as the mobile phase with 1% acetic acid an additive. Yield: 1.05g (35%)

^{1}H NMR (400 MHz, DMF-d7): δ 8.04 (d, J = 9.9 Hz, 1H), 7.94 (d, J = 7.5 Hz, 2H), 7.79 (d, J = 7.5 Hz, 2H), 7.60 (d, J = 8.2 Hz, 1H), 7.49 – 7.41 (m, 2H), 7.41 – 7.32 (m, 2H), 5.29 (t, J = 9.8 Hz, 1H), 5.06 – 4.97 (m, 2H), 4.48 (td, J = 8.8, 4.3 Hz, 1H), 4.37 – 4.28 (m, 3H), 4.25 (dd, J = 12.2, 5.2 Hz, 1H), 4.18 – 4.05 (m, 2H), 3.96 (ddd, J = 10.2, 5.2, 2.4 Hz, 1H), 3.37 (dd, J = 13.9, 4.4 Hz, 1H), 3.02 (dd, J = 13.9, 9.2 Hz, 1H), 2.10 – 2.01 (m, 6H), 1.99 (s, 3H), 1.85 (s, 3H).

^{13}C NMR (101 MHz, DMF): δ 173.36, 171.30, 170.93, 170.76, 170.61, 157.37, 145.30, 145.28, 142.22, 128.81, 128.27, 128.25, 126.53, 121.17, 84.94, 76.46, 74.98, 70.17, 67.58, 63.46, 55.83, 53.77, 48.11, 32.48, 21.03, 21.02, 21.01, 20.97.

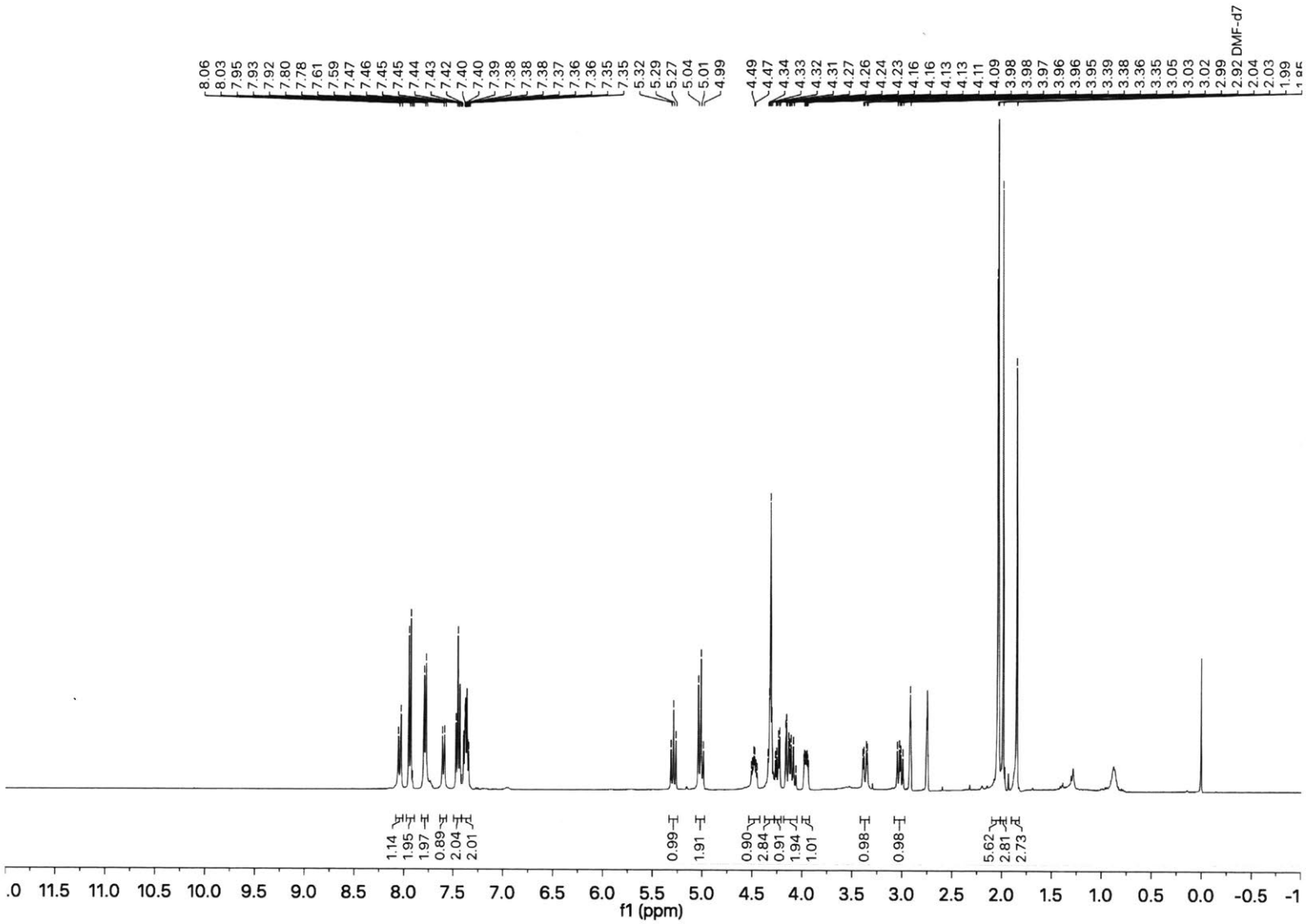

Figure 3-7: Fmoc-L-Cys(β-OAc$_3$GlcNAc)-OH ^1H NMR

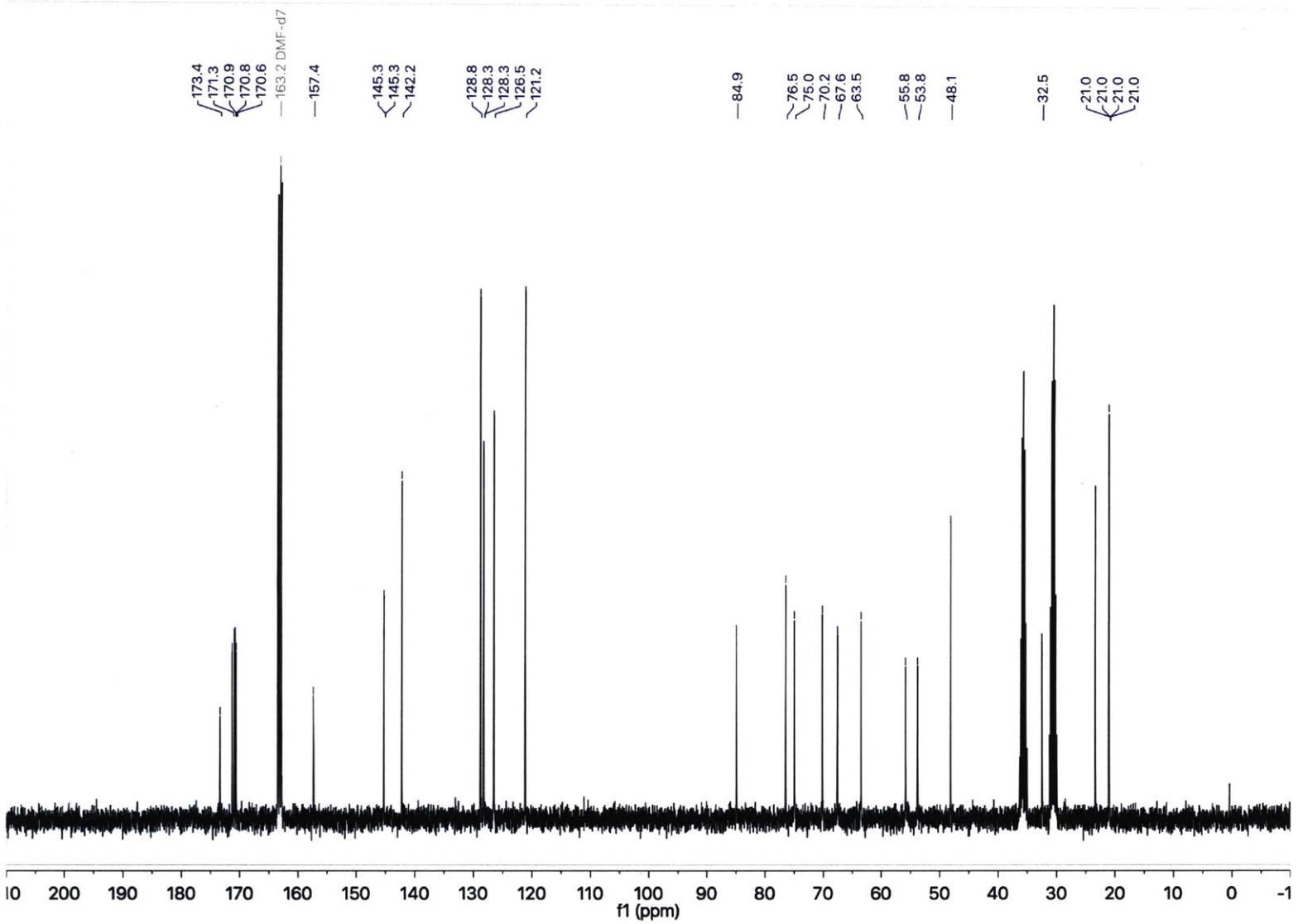

Figure 3-8: Fmoc-L-Cys(β-OAc$_3$GlcNAc)-OH ^{13}C NMR

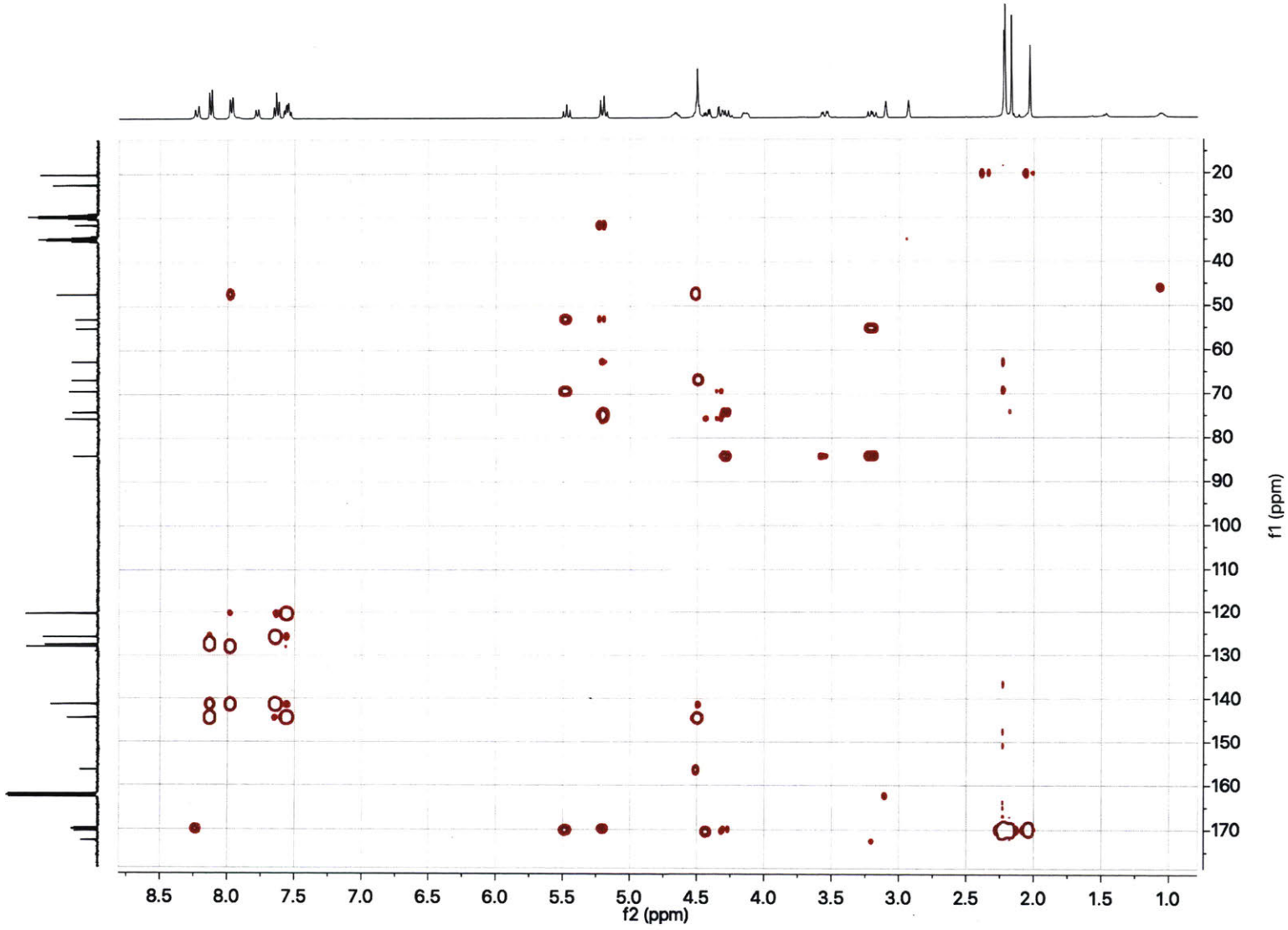

Figure 3-9: Fmoc-L-Cys(β-OAc$_3$GlcNAc)-OH HMBC NMR

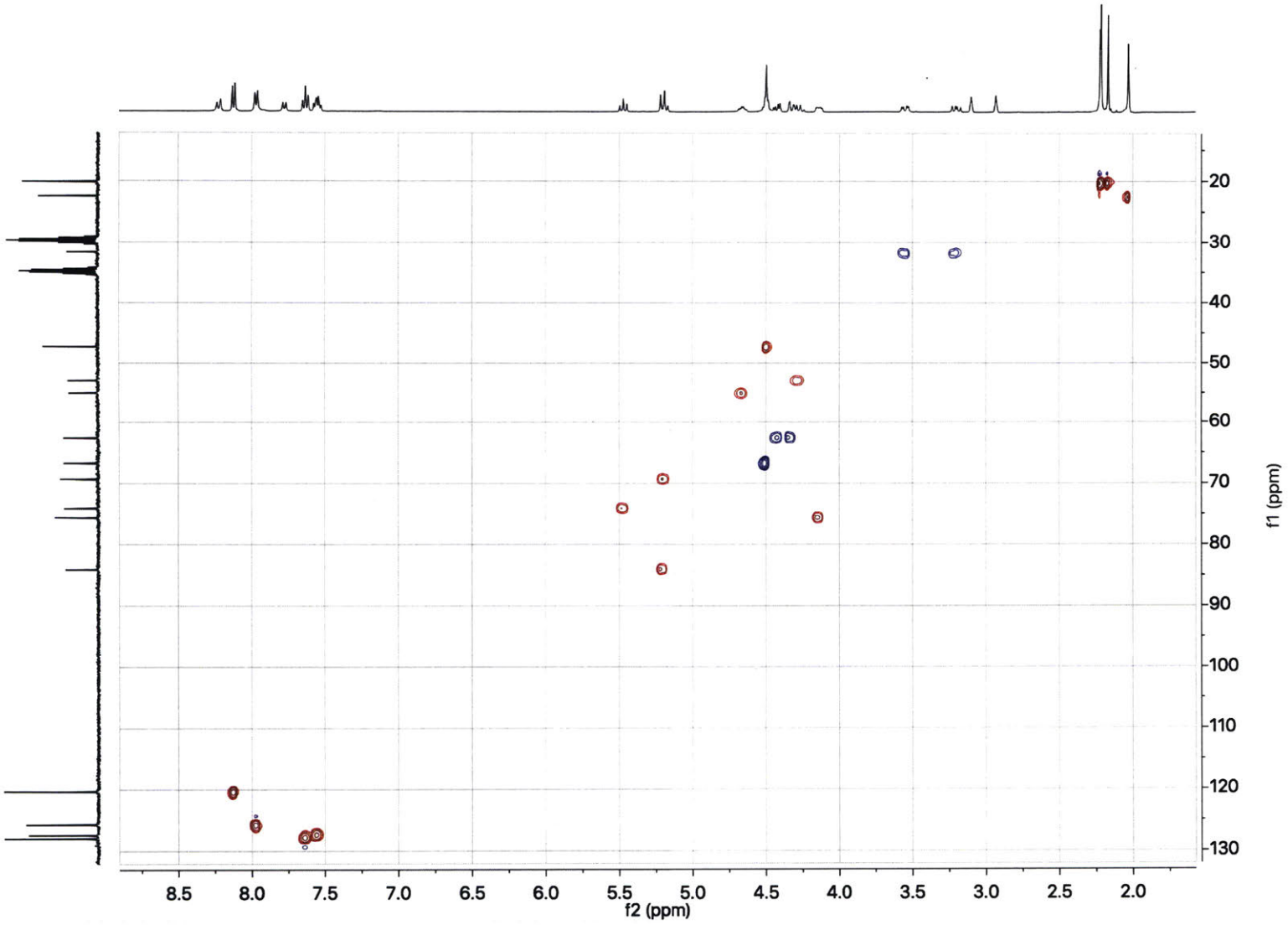

Figure 3-10: Fmoc-L-Cys(β-OAc$_3$GlcNAc)-OH HSQC NMR

References

1. Stepper, J. *et al.* Cysteine S-Glycosylation, a New Post-Translational Modification Found in Glycopeptide Bacteriocins. en. *FEBS Letters* **585,** 645–650. ISSN: 1873-3468 (Feb. 2011).

2. Maynard, J. C., Burlingame, A. L. & Medzihradszky, K. F. Cysteine S-Linked N-Acetylglucosamine (S-GlcNAcylation), A New Post-Translational Modification in Mammals. en. *Molecular & Cellular Proteomics* **15,** 3405–3411. ISSN: 1535-9476, 1535-9484 (Jan. 2016).

3. Xiao, H. & Wu, R. Global and Site-Specific Analysis Revealing Unexpected and Extensive Protein S-GlcNAcylation in Human Cells. *Analytical Chemistry* **89,** 3656–3663. ISSN: 0003-2700 (Mar. 2017).

4. Qin, W. *et al.* Artificial Cysteine S-Glycosylation Induced by Per-O-Acetylated Unnatural Monosaccharides during Metabolic Glycan Labeling. en. *Angewandte Chemie International Edition* **57,** 1817–1820. ISSN: 1521-3773 (Feb. 2018).

5. Amso, Z. *et al.* Total Chemical Synthesis of Glycocin F and Analogues: S -Glycosylation Confers Improved Antimicrobial Activity. en. *Chemical Science* **9,** 1686–1691 (2018).

6. Brimble, M. A. *et al.* Synthesis of the Antimicrobial S-Linked Glycopeptide, Glycocin F. en. *Chemistry - A European Journal* **21,** 3556–3561. ISSN: 1521-3765 (Feb. 2015).

7. Bisset, S. W. *et al.* Using Chemical Synthesis to Probe Structure–Activity Relationships of the Glycoactive Bacteriocin Glycocin F. *ACS Chemical Biology.* ISSN: 1554-8929. doi:10.1021/acschembio.8b00055 (Apr. 2018).

8. De Leon, C. A., Levine, P. M., Craven, T. W. & Pratt, M. R. The Sulfur-Linked Analogue of O-GlcNAc (S-GlcNAc) Is an Enzymatically Stable and Reasonable Structural Surrogate for O-GlcNAc at the Peptide and Protein Levels. *Biochemistry* **56,** 3507–3517. ISSN: 0006-2960 (July 2017).

9. Fiore, M., Lo Conte, M., Pacifico, S., Marra, A. & Dondoni, A. Synthesis of S-Glycosyl Amino Acids and S-Glycopeptides via Photoinduced Click Thiol–ene Coupling. *Tetrahedron Letters* **52,** 444–447. ISSN: 0040-4039 (Jan. 2011).

10. Thayer Desiree A., Yu Henry N., Galan M. Carmen & Wong Chi-Huey. A General Strategy toward S-Linked Glycopeptides. *Angewandte Chemie International Edition* **44,** 4596–4599. ISSN: 1433-7851 (June 2005).

11. Capon, B. Mechanism in Carbohydrate Chemistry. *Chemical Reviews* **69,** 407–498. ISSN: 0009-2665 (Aug. 1969).

12. Wang, H. *et al.* The Glycosyltransferase Involved in Thurandacin Biosynthesis Catalyzes Both O- and S-Glycosylation. *Journal of the American Chemical Society* **136,** 84–87. ISSN: 0002-7863 (Jan. 2014).

13. Chandra, S., Yadav, R. N., Paniagua, A. & Banik, B. K. Indium Salts-Catalyzed O and S-Glycosylation of Bromo Sugar with Benzyl Glycolate: An Unprecedented Hydrogenolysis. *Tetrahedron Letters* **57,** 1425–1429. ISSN: 0040-4039 (Mar. 2016).

14. Szabó, L. Z. *et al.* Preparation of S-Glycoside Surfactants and Cysteine Thioglycosides Using Minimally Competent Lewis Acid Catalysis. *Carbohydrate Research* **422,** 1–4. ISSN: 0008-6215 (Mar. 2016).

15. Hojo, K. *et al.* Aqueous Microwave-Assisted Solid-Phase Peptide Synthesis Using Fmoc Strategy. III: Racemization Studies and Water-Based Synthesis of Histidine-Containing Peptides. en. *Amino Acids* **46,** 2347–2354. ISSN: 0939-4451, 1438-2199 (June 2014).

16. Mijalis, A. J. *et al.* A Fully Automated Flow-Based Approach for Accelerated Peptide Synthesis. *Nature Chemical Biology* **13,** 464–466 (2017).

Chapter 4

Automated flow synthesis of peptide nucleic acids (PNA)

After optimizing the synthesis of polypeptides and recognizing the effectiveness of heat on the synthesis of aggregating peptides, we turned to the synthesis of peptide nucleic acids (PNA). Peptide nucleic acids are uncharged antisense oligonucleotides, sequence-specific binders of DNA and RNA, featuring an aminoethylglycyl backbone with pendant nucleobases.[1] When first discovered, PNA synthesis was achieved by protecting the primary amine of the aminoethylglycine backbone with a Boc group and protecting the exocyclic amino groups of cytosine, adenine, and guanine with carboxybenzoyl (Cbz) groups. However, much like with Boc peptide synthesis, Boc PNA synthesis has largely been phased out in favor of Fmoc chemistry due to issues with safety – each Boc removal requires treatment with trifluoroacetic acid, and global deprotection requires HF treatment – and yield.

Fmoc-protected PNA monomers feature an N-terminal protecting group and can feature a variety of protecting groups on the nucleobase amines, though the benzhydryloxycarbonyl (Bhoc) group is most common and is the only commercially-available Fmoc PNA monomer. [2, 3] However, peptide nucleic acids are known in the literature to be particularly difficult to synthesize with Fmoc chemistry, both because they aggregate and because they undergo a number of side reactions during synthesis (Scheme 4-1).[4] There is still no generic method for accessing PNA of ap-

preciable length using the Fmoc-Bhoc protection scheme. Here we report the rapid, Fmoc-based synthesis of PNA using the automated flow peptide synthesizer. Using the precise control over fluid flow and temperature offered by our synthesis technique, we can perform the synthesis at 70 °C, helping to prevent PNA aggregation, improving coupling yield and reducing coupling times, while minimizing undesirable deletions, isomerization, and adducts. We demonstrate the effect of varying coupling conditions on the synthesis outcome for a test substrate and extend this method to the synthesis of a bioactive 18-mer peptide nucleic acid.

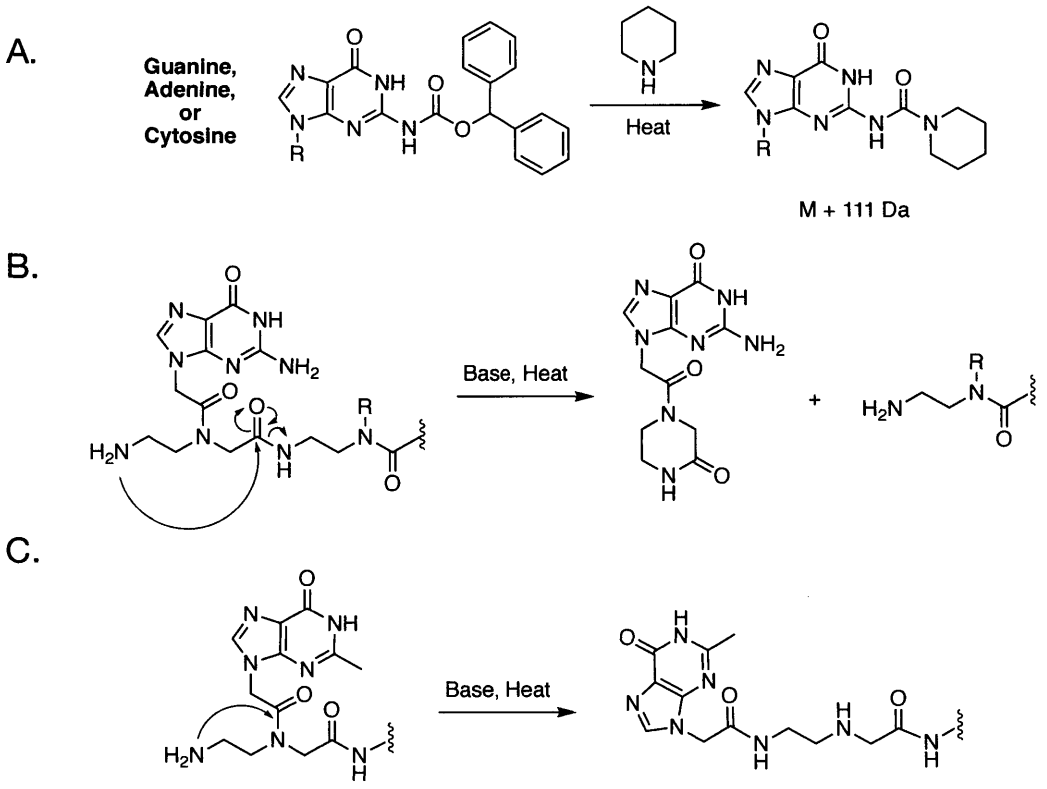

Figure 4-1: Side reactions during Fmoc PNA synthesis. **A.** The acid-labile Bhoc protecting group is commonly used to protect nucleobase exocyclic amino groups for Fmoc PNA synthesis. At elevated temperatures, it is cleaved by piperidine to yield a byproduct with an apparent +111 Da. mass. **B.** Backbiting of the newly deprotected amino group can lead to formation of a ketopiperazine that gets washed away, causing deletion of the N-terminal PNA unit.[5] **C.** Base and heat can promote transamidation of the nucleobase to the N-terminus, resulting in an isomeric, secondary amine-containing product.

Besides monomer non-incorporation, The reported side reactions that occur dur-

ing PNA synthesis occur during Fmoc removal. Upon treatment with piperidine, the N-terminal amine can be deleted via "backbiting" that forms a cyclic ketopiperazine moiety which flows away from the solid support.[5] The N-terminal amine can potentially attack the tertiary amide which links the nucleobase to the backbone and is 5 atoms away, transferring the nucleobase to the N-terminus and truncating the synthesis. It is worth noting that this product was not observed when Fmoc-Cbz monomers were used.[5] Finally, in a previously unreported side reaction that we observe at elevated temperature, piperidine can react directly with the benzhydryl carbamate (Bhoc) protecting group of the exocyclic amines used for Guanine, Cytosine, and Adenine. The resulting piperidinyl urea is stable to the acidic conditions of cleavage and leads to an apparent +111 Da. side product.

Before embarking on this project, we naïvely attempted to synthesize the 18-mer exon-skipping EGFP sequence using Fmoc/Bhoc protected PNA monomers and standard coupling procedures at 90 °C on the automated flow peptide synthesizer. We observed abundant deletion products using this method and a byproduct with mass M+111 Da. after even the first six residues of the synthesis. After 18 coupling and Fmoc removal steps, the desired product could not be separated or identified in the field of byproducts (Figure 4-2.

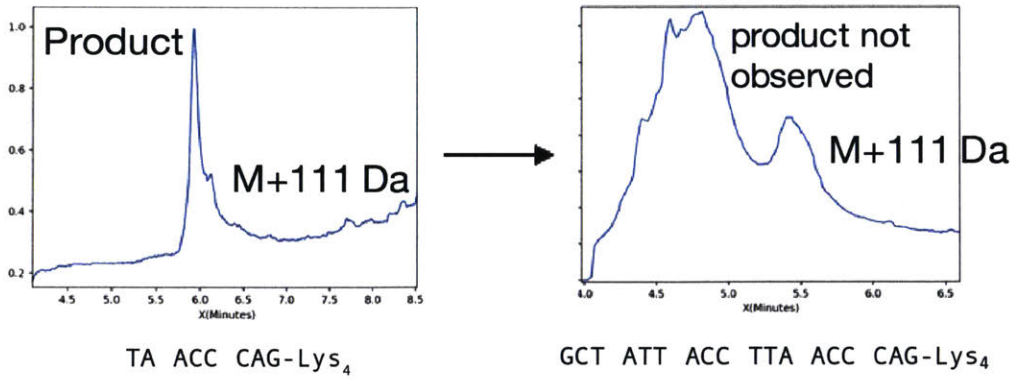

Figure 4-2: PNA synthesis attempt at 90 °C. The desired product was not observed after 18 coupling cycles.

Using the automated flow peptide synthesizer, we have independent control over the temperatures and flow rates for the coupling, washes, and Fmoc removal steps.

To start optimizing peptide nucleic acid synthesis, we first examined the effects of temperature and piperidine treatment time on the distribution of byproducts for a 4-residue substrate containing all four nucleobases. This substrate included a C-terminal peptide tag to facilitate analysis by liquid chromatography/mass spectrometry. Interestingly, we observed that, using HATU in DMF as the coupling agent, nucleobase coupling at 70 °C resulted in much less deletion products than nucleobase coupling at 90 °C (Figure 4-3). After synthesis, the resin-bound PNA was treated with 20% piperidine at 90 °C for 1, 2, 3, and 4 minutes, and no increase in deletion products was observed, leading us to conclude that deletion sequences occur during coupling, not Fmoc removal, and that in contrast to observations by Thomson, et al., the ketopiperazine deletion mechanism is probably not relevant.[5]

During the course of other experiments synthesizing peptides with Bhoc nucleobase acetic acid moieties, we also observed a M+111 byproduct. However, this byproduct was not observed in sequences containing only Thymine. Because Thymine is the only non-Bhoc protected nucleobase used in PNA synthesis, we hypothesized that the M+111 byproduct was likely related to the Bhoc protecting group. Chemical structure analysis indicates that substitution of the benzhydrol for a piperidine unit to form the N-piperidinyl urea would lead to a byproduct of the observed mass. We observed that for the test PNA TCAG-ALFK-$CONH_2$, the amount of this +111 byproduct varied linearly as a function of treatment time with piperidine at 90 °C.

The same resin-bound, Bhoc-protected PNA also forms an isobaric product which has a later elution time by reverse-phase HPLC. We believed that this isomer could be the result of nucleobase transamidation to the N-terminus, facilitated by a 5-membered transition state. We observed that this isomeric byproduct, like the piperidine adduct, varies linearly as a function of piperidine treatment at 90 °C, suggesting a base-catalyzed process. While we could not unambiguously assign the atom connectivity of this product, it is effectively suppressed by limiting the piperidine treatment temperature to 70 °C.

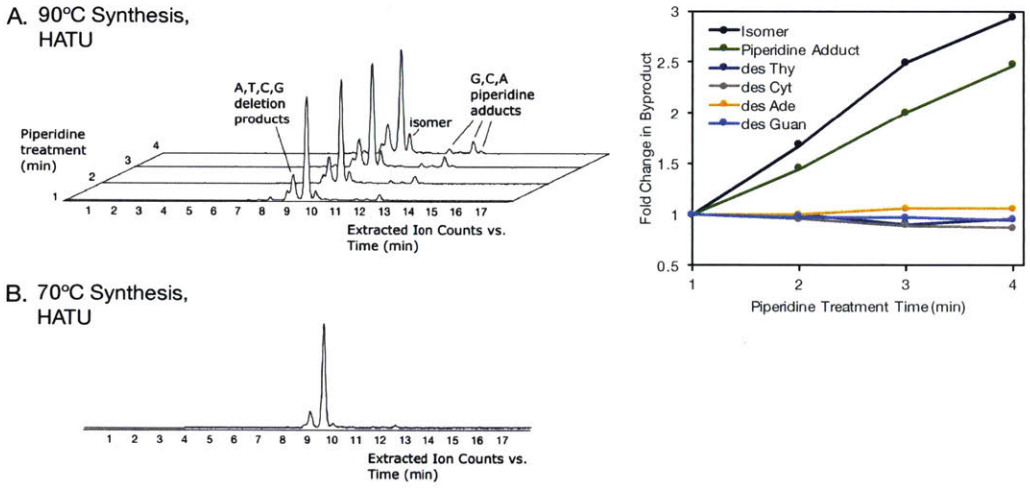

Figure 4-3: PNA synthesis quality is dependent on Fmoc removal conditions. **A.** The PNA TCAG-Lys_3 was synthesized at 90 °C under standard flow conditions. After the synthesis, the resin was treated with 20% piperidine for 4 minutes and sampled every minute. Products were quantified by comparing ion current of a characteristic +3 ion species. **B.** PNA synthesis at 70 °C exhibits lower amounts of nucleobase deletion, isomerization, and the +111 product is not observed.

References

1. Hyrup, B. & Nielsen, P. E. Peptide Nucleic Acids (PNA): Synthesis, Properties and Potential Applications. *Bioorganic & Medicinal Chemistry* **4,** 5–23. ISSN: 0968-0896 (Jan. 1996).

2. Srinivasu, P., Zbigniew, P. & Nicolas, W. Expanding the Scope and Orthogonality of PNA Synthesis. en. *European Journal of Organic Chemistry* **2008,** 3141–3148. ISSN: 1099-0690 (June 2008).

3. Breipohl, G., Knolle, J., Langner, D., O'Malley, G. & Uhlmann, E. Synthesis of Polyamide Nucleic Acids (PNAs) Using a Novel Fmoc/Mmt Protecting-Group Combination. *Bioorganic & Medicinal Chemistry Letters* **6,** 665–670. ISSN: 0960-894X (Mar. 1996).

4. Development of an Efficient and Low-Cost Protocol for the Manual PNA Synthesis by Fmoc Chemistry. en. *Tetrahedron Letters* **51,** 3716–3718. ISSN: 0040-4039 (July 2010).

5. Thomson, S. A. *et al.* Fmoc Mediated Synthesis of Peptide Nucleic Acids. *Tetrahedron* **51,** 6179–6194. ISSN: 0040-4020 (May 1995).

Chapter 5

Flow Alloc Removal and Synthesis of nucleobase-functionalized polypeptides

5.1 Flow deprotection of N-allyloxycarbamates, an orthogonal amino protecting group

Nucleobase-functionalized polymers are hybrid materials containing both amino acid and nucleobase residues that can have interesting supramolecular properties and show promise for the development of novel antisense therapeutics. Garner, et al., developed the alpha-helical peptide nucleic acid concept,[1, 2] in which nucleobases attached via serine residues would bind sequence-specificially DNA upon folding. Takahashi, et al., successfully used nucleobase-containing α-amino acids as protein engineering tools, incorporating them into analogues of HIV-1 protease, increasing affinity to target hairpin RNA strands.[3, 4] However, previous studies on nucleobase-containing polymers was fundamentally limited by synthetic techniques. In general, these polymers contained fewer than 5 nucleobases and include only thymine and cytosine, excluding purine bases adenine and guanine.

Automated Flow Peptide Synthesis, developed in our group, can achieve peptide

synthesis in times of less than 40 seconds per amino acid. With this technology in hand, we set out to expand its chemical reaction scope in the hopes of creating a method for printing nucleobase peptides of any topology with a dual protecting group strategy (Scheme 5-1. However, there is one main limitation when performing reactions on a solid support in flow: because HPLC pumps, microfluidic tubing, and narrow-bore valves are used, it is crucial that all reagents are free of particulate matter for the duration of their time in the system. Precipitates will cause immediate failure.

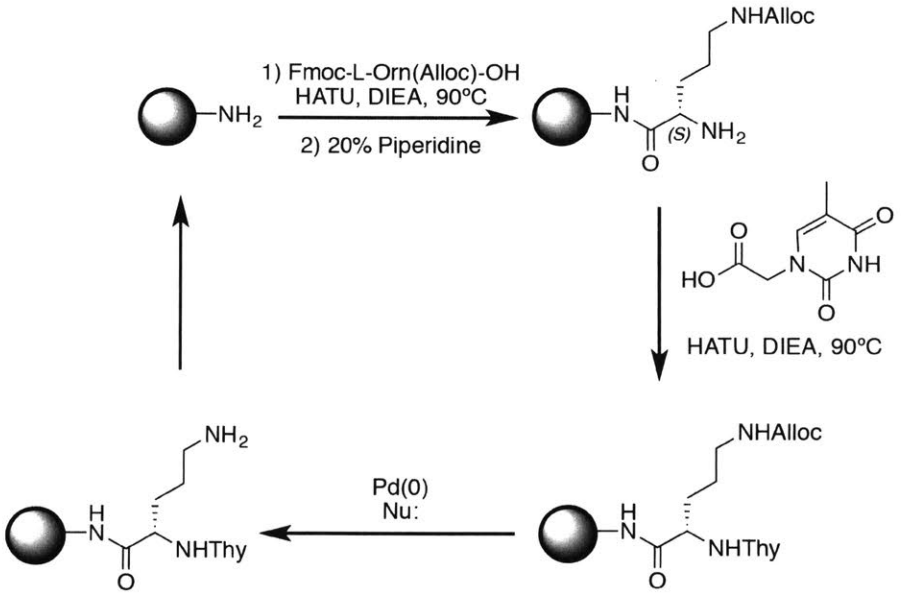

Figure 5-1: Scheme for the submonomer synthesis of nucleobase-functionalized polypeptides

With this in mind, I attempted to develop an automated flow method for orthogonal amine deprotection. I chose the allyloxy carbamate (Alloc) protecting group as a target for removal because there are many commercially-available Alloc-protected units, and it is easily removed with catalytic amounts of Pd(0).[5] However, early attempts to use $Pd(PPh_3)_4$ as a deprotection agent failed on the automated flow peptide synthesizer because when heated, it precipitated almost instantly to form palladium black, clogging the flow path. A solvent screen failed to identify a solvent compatible with high temperature synthesis. Moreover, tetrakis triphenylphosphine palladium (0) has a number of other non-ideal properties: it oxidizes readily in air

and must be prepared fresh, and it is not highly soluble in DMF, whe solvent of choice for heated flow peptide synthesis. Dichloromethane, the solvent commonly used with Alloc removal reactions, is too low-boiling to be used at 90 °C.

I hypothesized that a Buchwald Pd oxidative addition precatalyst could be ideal for Alloc protecting group removal in flow at elevated temperature.[6] The precatalysts themselves are air-stable, ideal for long-term storage on the automated flow peptide synthesizer. Importantly, these Pd(II) precatalysts can be reduced to Pd(0) species on-demand by treatment with a base. I envisioned preparing stable solutions of these precatalysts, storing them on the synthesizer until needed, at which point they would be combined on-demand with piperidine, generating the catalytic Pd(0) species and an allyl cation scavenger in solution.

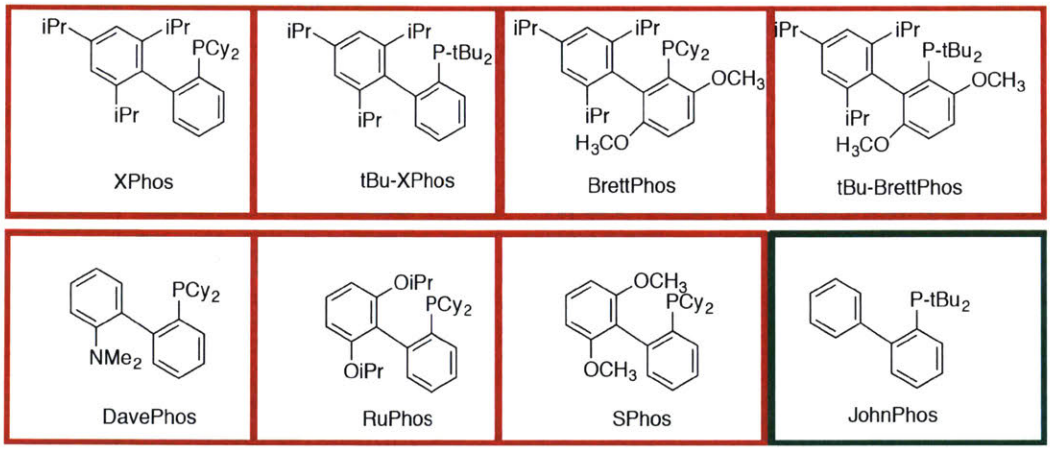

Figure 5-2: Solubility of bulky phosphine ligands in N,N-dimethylformamide.

2-aminobiphenyl Pd(II) precatalyst with BrettPhos as a ligand was synthesized and tested in batch for its suitability in performing this reaction in DMF in flow. BrettPhos precatalyst was dissolved in DMF and treated with an equal volume of 20% piperidine. The solution immediately turned brown and started to form a fine palladium black precipitate. I thought this might be because piperidine coordinating to the newly-formed Pd(0) species, so I experimented with the addition of other ligands to the solution before piperidine treatment. I screened a library of bulky phosphine ligands (Figure 5-2) and of the 8 tested, JohnPhos was the only one soluble in DMF. Addition of JohnPhos to the Buchwald catalyst mixture yielded a red solu-

tion upon treatment with piperidine that does not form precipitates even after 24h at 90 °C. Using this catalyst mixture on the AFPS, Alloc removal was quantitative in one minute with just 10 mol% catalyst (10 micromoles) with respect to the resin substitution on a test peptide substrate, Orn(Alloc)-ALFALF (Figure 5-3). With this chemistry and an appropriate monomer with both Alloc and Fmoc protecting groups, we are now quickly and easily able to create branched polypeptides in flow.

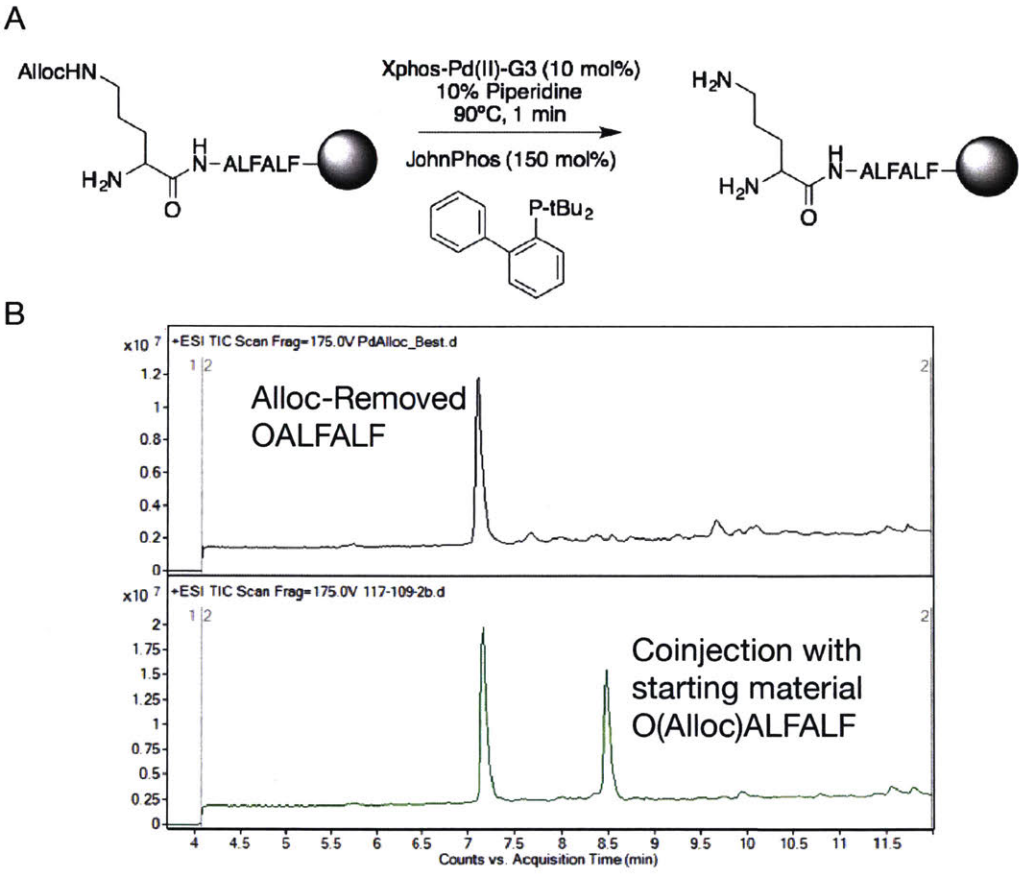

Figure 5-3: Removal of the Alloc protecting group on a test peptide substrate. **A.** Scheme for flow alloc removal on a resin-bound test substrate Orn-ALFALF.

5.2 Synthesis of nucleobase-functionalized polypeptides

We chose to attach nucleobases to the peptide backbone using standard amide bond forming chemistry and Bhoc-protected nucleobases modified with an acetic acid moiety at N-1 for pyrimidines or N-9 for purines. We chose these precursors because the Bhoc protecting group is stable to base treatment and is cleaved under acidic conditions.

We initially attempted to synthesize test polypeptides containing 5 pyrimidine residues thymine and cytosine. With thymine-containing nucleobases, we found that phosphonium-derived coupling agents, such as PyAOP, are able to activate the heterocyclic amide for attack by piperidine, causing an irreversible modification to the base. We attempted synthesis of N-boc protected thymine; however, this material is not shelf-stable and reverts to the non-protected form over time. Using HATU as the coupling agent, we synthesized polypeptides containing 5 thymine or cytosine residues in high purity (Figure 5-4).

Purine residues were more difficult to couple. We attempted to optimize the coupling of these nucleobase derivatives by varying the flow rate, temperature, and activating agent used during the nucleobase coupling. We had previously identified PyAOP as being problematic for tautomerizable nucleobases (Thymine and Guanine) so we chose HBTU and HATU as the activating agents for this study. To our surprise, using HATU as a coupling agent caused large inefficiencies in coupling for purine nucleobases. Using HBTU as the coupling agent, Guanine and Adenine monomers were incorporated with 90% or greater efficiency.

5.3 Materials and Methods

5.3.1 Synthesis of palladium precatalysts

1. Ammonium biphenyl mesylate

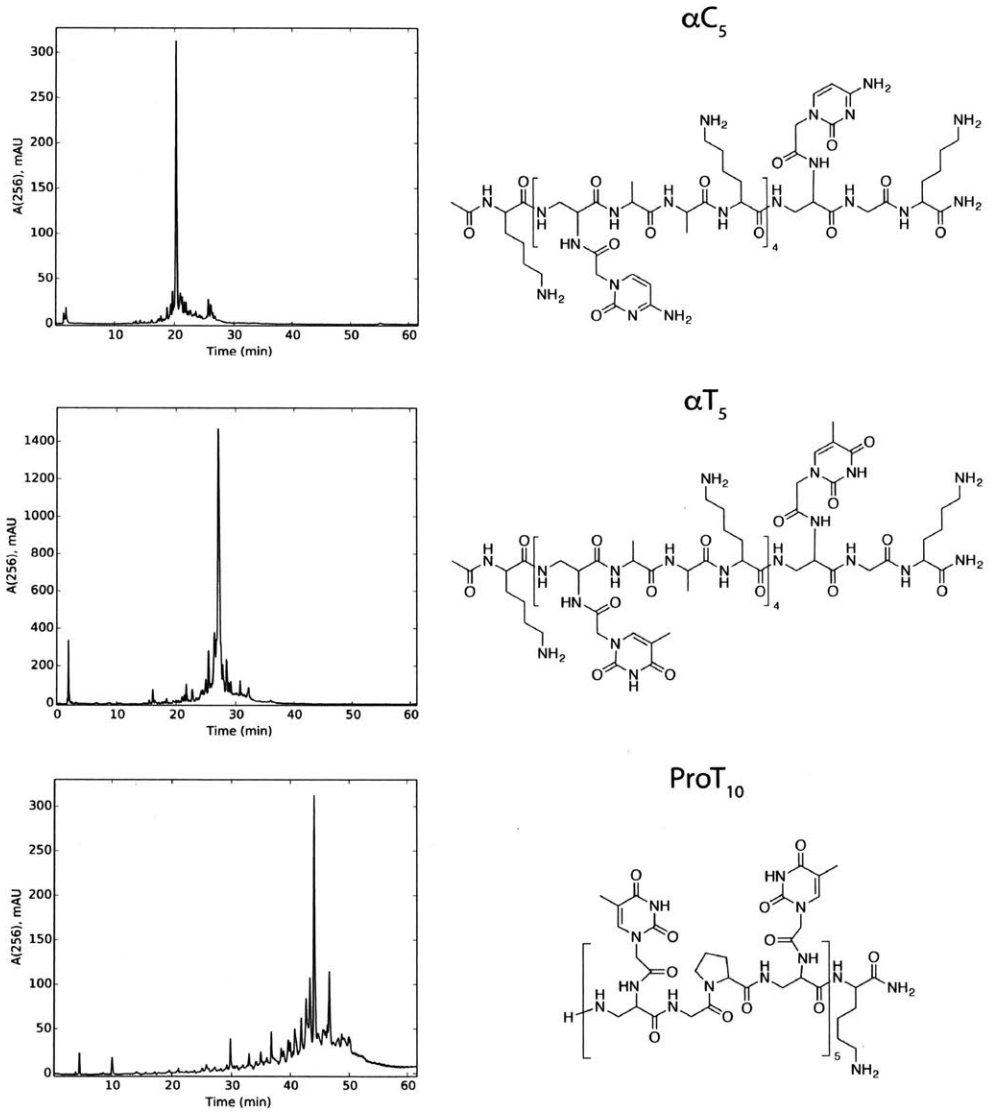

Figure 5-4: Analysis of crude nucleobase-functionalized peptides synthesized on the automated flow peptide synthesizer

Procedure adapted from Buchwald, et al.[6] 2-aminobiphenyl (3.00 g, 17.73 mmol, 1 eq.) was dissolved in 60 mL diethyl ether in a 250 mL round bottom flask. To this solution, methanesulfonic acid (1.15 mL, 1.70g, 17.73 mmol, 1 eq.) pre-dissolved in 15 mL diethyl ether was added dropwise and stirred. A light solid precipitated. After sonication, 100 methane sulfonic acid was added and stirred for an additional hour. The solid was filtered, washed with diethyl ether and dried in vacuo to yield 4.49 g product that was used directly (96%).

2. μ-OMs Pd Dimer

   Procedure adapted from Buchwald, et al.[6] Crude ammonium biphenyl mesylate (4.49g, 16.9 mmol, 1 eq.) and $Pd(OAc)_2$ (3.79 g, 16.9 mmol, 1 eq.) were dissolved in 90 mL toluene at room temperature, placed under nitrogen, then heated in a water bath to 60 °C for 1 hour. Upon heating, a tan solid precipitated from solution. The reaction mixture was allowed to cool on ice, filtered, washed with toluene, and dried on high vacuum to yield 5.31 g product (85%).

3. Brett-Phos Generation 3 Precatalyst

   Procedure adapted from Buchwald, et al.[6] To a 50 mL round bottom flask, 1.10 g BrettPhos and 756 mg (1.02 mmol, 0.5 eq.) Pd(OMs) dimer was added, dissolved in 10 mL dichloromethane, placed under nitrogen, and stirred overnight at room temperature. After reaction, the DCM was evaporated and 15 mL diethyl ether was added. AFter sonication, the dark brown oil became a light tan solid. The solid was filtered and dried in vacuo to yield 1.73 g Brett-Phos Pd G3 precatalyst (94%). ^{31}P NMR (162 MHz, $CDCl_3$) δ 42.65 ppm.

5.3.2 Removal of N-Alloxycarbonyl protecting groups in flow

Alloc removal was accomplished in three steps in flow. In the first step, 2.4 ml each of solutions containing (1) 12.5 mg/mL JohnPhos in 20% piperidine, and (2) 4.6 mg/mL BrettPhos Pd G3 and 25 mg/mL JohnPhos were delivered through two pumps at 40 mL/min each (80 mL/min total) and 90 °C. In the second step, the Pd reagents were

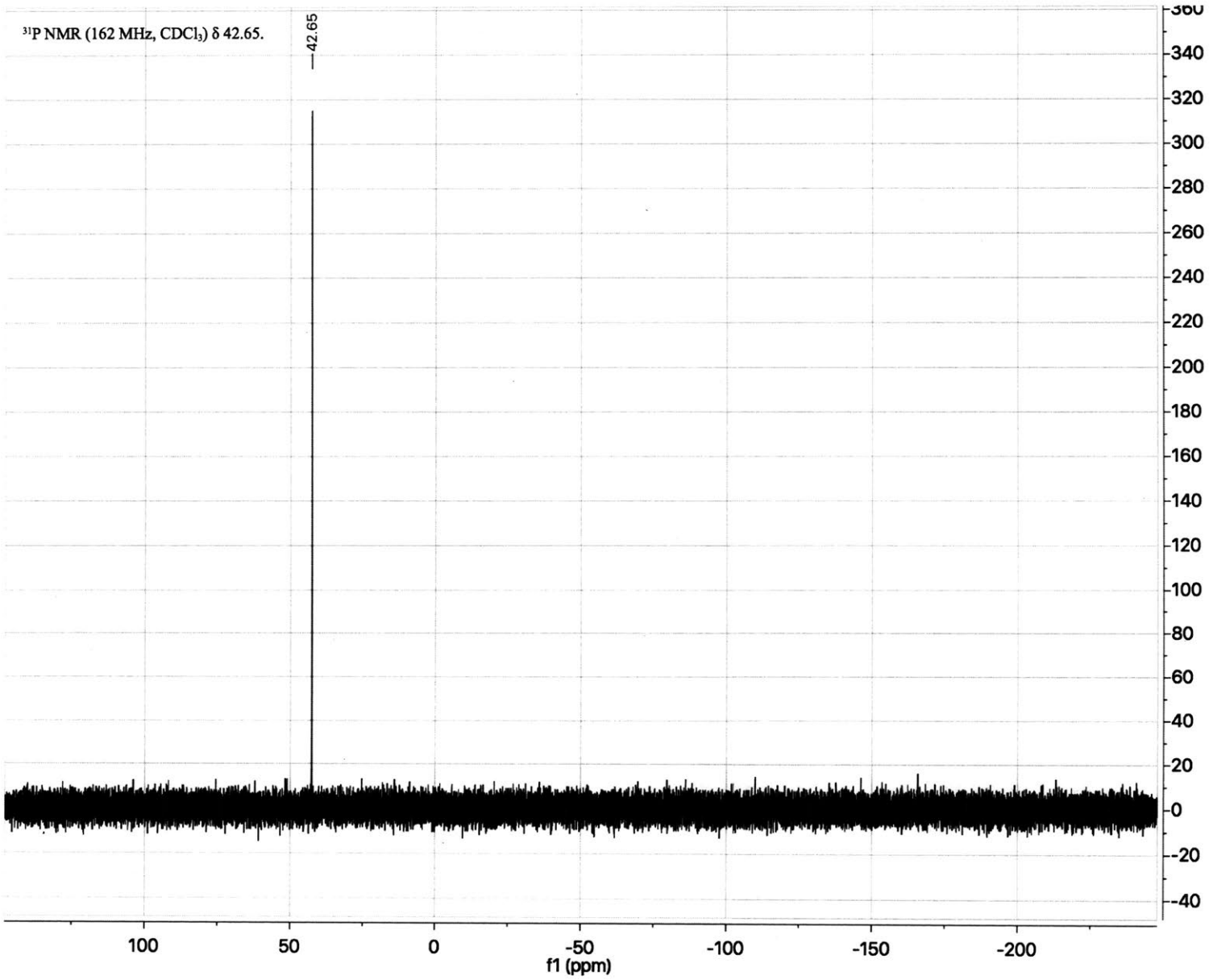

Figure 5-5: BrettPhos Pd G3 Precatalyst ^{31}P NMR

washed out by delivering 5.2 mL each (1) 12.5 mg/mL JohnPhos and (2) DMF at 40 mL/min each (80 mL/min total flow rate). Finally, a DMF wash was performed.

5.3.3 Synthesis of nucleobase acetic acids

Procedures for Bhoc-cytosine-1-acetic acid and Bhoc-adenine-1-acetic acid were adapted from Wissinger, et al.[7, 8] Procedures for Thymine-1-acetic acid were adapted from Breeger et al.[9]

1. Thymine-1-Acetic Acid

   Thymine (10g, 79.3 mmol) and potassium hydroxide (22g, 392 mmol) were dissolved in 200 mL water and heated to 95 °C. Bromoacetic acid (22g, 159 mmol) was added and the solution was refluxed for 1h. The reaction mixture was acidified to pH 2 and then cooled. Upon cooling, a white precipitate formed which was filtered, washed with cold water and ethyl acetate, and dried in vacuo to yield 7.4g of a white powder (53% yield). 1H NMR (DMSO): 1.78ppm (D, 3H, J = 1.19 Hz); 4.36 ppm (s, 2H); 7.48 ppm (m, 1H, J = 1.19 Hz); 11.33 ppm (s, 1H); 13.09 ppm (s, 1H)

2. Benzyl Thymine-1-Acetate

   Thymine (10g, 79.3 mmol) was suspended in 240 mL DMF with stirring. Potassium tert-butoxide (10.7g, 95.2 mmol) was added, and the solution heated up and became slightly more viscous. Benzyl bromoacetate (15 mL, 95 mmol) was added dropwise over 5 minutes and left to stir overnight.

   The reaction was evaporated (75 °C, 40 mbar), and once dry, 100 mL water and 3 mL concentrated hydrochloric acid were added. A cookie dough-like solid formed. The water was decanted and dried under vacuum, and the solid was triturated with methanol, sonicated, and filtered to yield 11.6g of a white powder (57% yield).

   For purification, 11g of this solid was recrystallized by dissolving in 600 mL boiling methanol and cooling slowly to room temperature. 7.8g of pure material

was recovered (85%).

1H NMR (DMSO): 11.43 ppm (s, 1H); 7.33-7.45 ppm (m, 5H); 7.54 ppm (m, 1H, J=1.18 Hz) 5.21 ppm (s, 2H); 4.57 ppm (s, 2H); 1.77 ppm (d, 3H, J = 1.18 Hz)

3. Boc Thymine-1-Acetic Acid

   Boc Benzyl Thymine-1-Acetate (2.7g, 7.2 mmol) was added to a 250 mL round-bottom flask and dissolved in a 50:50 mixture of acetone and methanol (60 mL). Palladium on carbon (40 mg) was added. After evacuating the flask and purging with nitrogen three times, the flask was evacuated once more and placed under an atmosphere of hydrogen. The reaction was stirred for two days at room temperature and filtered through a pad of celite to remove the catalyst. The solvent was removed and the product dried under high vacuum to yield 1.7g of a white foam (86% yield).

   1H NMR (DMSO): 7.65 ppm (m, 1H, J = 1.20 Hz); 4.43 ppm (s, 2H); 1.81 ppm (d, 3H, J = 1.20 Hz); 1.51 ppm (s, 9H)

4. Benzyl Cytosine-1-Acetate

   To a suspension of cytosine (25.0g, 225 mmol) in 450 mL DMF, potassium tert-butoxide (27.8g, 248 mmol) was added and stirred under an atmosphere of nitrogen. The suspension was heated up to 90 °C for 1h, and the liquid part of the suspension turned yellow. The suspension was then cooled down to 0 °C on an ice bath, and benzyl bromoacetate (37 mL, 248 mmol) was added dropwise over 45 minutes using an addition funnel. Once addition was complete, the suspension was stirred for 1h and then taken off of the ice bath and allowed to warm to room temperature. The mixture developed a red-brown hue, not unlike chocolate milk, as it stirred for the next 24h. The reaction was quenched by adding 15 mL of acetic acid and the solution was evaporated by half. Water (200 mL) was added and the mixture was allowed to stand overnight. The solid was filtered, washed with ethanol and diethyl ether, and dried in vacuo to yield

51.8 g of a white solid (88% yield).

1H NMR (DMF): 7.75 ppm (d, 2H, J = 7 Hz); 7.08 ppm (br. s, 1H); 5.88 ppm (d, 2H, J = 7Hz); 5.25 ppm (s, 2H); 4.67 ppm (s, 2H)

5. Benzyl Bhoc Cytosine-1-Acetate

   A 1L round-bottom flask equipped with a stir bar was charged with benzyl cytosine-1-acetate (51.8g, 200 mmol), carbonyl diimidazole (51.8g, 320 mmol), diisopropylethylamine (70 mL, 400 mmol). DMF (400 mL) was added and the suspension was heated to 60 °C. Within 15 minutes, all off the suspension dissolved to give a yellow-brown solution. Benzhydrol (47.9g, 260 mmol) was then added, and after 1h, two more portions of benzhydrol (4.4g) were added at 30 minute intervals. Heating was stopped after 9h, and the reaction was allowed to stir overnight.

   The next day, the DMF was evaporated (60 °C, 20 torr). Methanol (300 mL) was added to the thick oil and put on ice. A large amount of crystals formed in the flask and filtered. The mother liquor was treated with ice-cold 50% methanol in water (200 mL), and the resulting precipitate was filtered and combined with the first filtrate. The product was washed with methanol and dried in vacuo to yield 65.2g of an off-white powder (70% yield).

   1H NMR (DMSO): 11.04 ppm (s, 1H); 8.05 ppm (d, 1H, J = 7 Hz); 7.25-7.5 ppm (m, 15H); 6.98 ppm (d, 1H, J = 7 Hz); 6.80 ppm (s, 1H); 5.18 ppm (s, 2H); 4.69 ppm (s, 2H)

6. Bhoc Cytosine-1-Acetic Acid

   A 1L round-bottom flask was charged with Benzyl Bhoc Cytosine-1-Acetate (12.7g, 470 mmol). A mixture of acetonitrile, methanol, water, and ethanol (2:2:1:1, 350 mL) was added to the flask, and boiled and sonicated until all of the starting material was dissolved. The heat was turned off and stirring continued until the boiling stopped (note: if the solution cools significantly or evaporates during heating, the startang material will crystallize; if this happens,

more solvent should be added and the mixture should be boiled again). Then, an aqueous solution of lithium hydroxide (11.3g, 270 mmol in 85 mL water) was added to the stirring yellow solution. Immediately, a TLC sample was taken (30% methanol in ethanol as the eluent) and the starting material appeared to be nearly quantitatively converted to product. After 3 minutes of stirring, the reaction was quenched by the addition of an aqueous solution of citric acid (26g, 135 mmol in 125 mL water) and placed on ice to crystallize. The precipitate was filtered, washed with water, and dried under high vacuum overnight to yield 7.7g of Bhoc Cytosine-1-Acetic Acid (77% yield).

1H NMR (DMSO): 13.08 ppm (br. s, 1H); 11.0 ppm (br. s, 1H); 8.03 ppm (d, 1H, J = 7 Hz); 7.25-7.5 ppm (m, 10H); 6.95 ppm (d, 2H, J = 7 Hz); 6.81 ppm (s, 1H).

ESI-MS (Negative mode): m/z observed = 378.1097, $[M-H]^{-1}$.

7. Benzyl Adenine-9-Acetate

A dry, two-neck, 1L round-bottom flask was charged with a stir bar and adenine (15.0g, 111 mmol). The adenine was suspended in DMF (400 mL), placed under a nitrogen atmosphere, and cooled to 0 °C. Sodium hydride (5.50g, 60% dispersion in mineral oil, 133 mmol) was added to the suspension and stirred rapidly while flushing nitrogen gas in through one neck of the flask and out the other. Hydrogen gas was evolved. After the gas evolution subsided, benzyl bromoacetate (19.4 mL, 122 mmol) was added dropwise over the course of 1h. The turbid suspension first turned a light orange which deepened as the solution clarified. After stirring an additional 30 minutes at 0 °C, the flask was warmed to room temperature and allowed to stir overnight.

The reaction mixture was evaporated to remove most of the DMF and chilled water (300 mL) was added to the reddish oil while being swirled vigorously. A white solid precipitated. The mixture was placed on ice and the precipitate was filtered, washed with water, ethanol (3x100 mL) and diethyl ether (200 mL). Brown impurities were removed in the washes, yielding 22.6g of a tan solid after

drying in vacuo (72% yield).

1H NMR (DMF): 8.22 ppm (s, 1H); 8.19 ppm (s, 1H); 7.3-7.5 ppm (m, 5H); 7.31 ppm (s, 2H); 5.26 ppm (s, 2H); 5.25 ppm (s, 2H)

8. Bhoc Adenine-9-Acetic Acid

A 100 mL round-bottom flask was charged with carbonyldiimidazole (2.35g, 14.4 mmol, 1.6 eq.) and 30 mL DMF and heated to 120 °C. Benzyl adenine acetate (2.50g, 8.8 mmol, 1 eq.) was added portionwise and stirred for 1h. After 1h, the temperature was reduced to 95 °C and benzhydrol (2.50g, 13.6 mmol, 1.5 eq.) was added and the reaction mixture was stirred at 95 °C for 30 minutes. Heat was discontinued and the reaction mixture was stirred overnight.

The next day, DMF was evaporated and the oil was dissolved in 5% hexanes in ethyl acetate (40 mL). This was filtered over a silica gel pad (4cm x 4cm). The solvent was evaporated, and 25 mL 50% acetonitrile in ethanol was added. The solution was cooled to 0 °C and LiOH•H2O (1g, 23.8 mmol, 2.7 eq.) dissolved in 15 mL water was added. After stirring for 10 minutes, TLC indicated complete consumption of starting material to a product with Rf = 0.5 (30% MeOH in EtOAc). The pH was carefully adjusted to 3.5 while monitoring with a pH probe by adding 9.6g citric acid dissolved in 50 mL water. The reaction was then placed into an ice bath to crystallize.

After filtration, the white solid was washed with water, cold ethanol, and diethyl ether. The solid was dried in vacuo to yield 1.95g of a white powder (55% over two steps).

1H NMR (DMSO): 13.42 ppm (s, 1H); 10.95 ppm (s, 1H); 8.62 ppm (s, 1H); 8.42 ppm (s, 1H); 7.2-7.5 ppm (m, 10H); 6.82 ppm (s, 1H); 5.08 ppm (s, 2H). 13C NMR (DMSO): 169.11 ppm; 152.13 ppm; 151.71 ppm; 151.12 ppm; 149.36 ppm; 144.87 ppm; 140.92 ppm; 128.51 ppm; 127.73 ppm; 126.50 ppm; 122.68 ppm; 77.26 ppm; 44.30 ppm.

ESI-MS (Negative mode): m/z observed = 402.1218, $[M-H]^{-1}$.

9. Benzyl 2-Amino-6-Chloropurine-9-Acetate

   27.9 g 2-amino-6-chloropurine (165 mmol, 1 eq.) was added to a 1L round-bottom flask. 300 mL DMF was added, and the mixture heated to 85 °C in a water bath. The material was insoluble. To the suspension, potassium carbonate (34.4g, 249 mmol, 1.51 eq.) was added. The suspension was stirred briefly at 85 °C and then cooled to 0 °C on an ice bath. Benzyl bromoacetate (28 mL) was added dropwise in 2 portions over the course of 3 hours. The reaction was stirred as it warmed to room temperature, and then left to stir at room temperature overnight.

   After stirring overnight, the reaction mixture had turned orange but was still milky. The precipitates were filtered from the reaction mixture and washed with DMF to yield around 400 mL of a dark brown solution. The product was precipitated by pouring the DMF mixture into 200 mL of vigorously stirring 1M aqueous HCl. After cooling on ice, the product was filtered and washed with water. The crude, off-white powder was then recrystallized from 300 mL boiling acetonitrile to yield white, needle crystals that were washed with cold acetonitrile, methanol, and ether. The product was dried in vacuo to yield 30.0g of a fluffy white to pink cotton-like solid (57.4%).

   1H NMR (DMSO): 8.10 ppm (s, 1H); 7.3-7.4 ppm (m, 5H); 7.00 ppm (s, 2H); 5.18 ppm (s, 2H); 5.05 ppm (s, 2H).

10. Bhoc Benzyl 2-Amino-6-Chloropurine-9-Acetate

    Benzyl 2-Amino-6-Chloropurine Acetate (15.04g, 47.2 mmol, 1 eq.) was placed into an oven-dried 1L flask equipped with a stir bar. THF (250 mL) was added, and the flask placed on ice under an inert atmosphere of nitrogen. Starting material did not fully dissolve at this point. After the flask had cooled to 0 °C, triphosgene (5.12g, 17 mmol, 0.36 eq.) was added and the reaction mixture stirred for 30 minutes. DIEA (18.5 mL, 104 mmol, 2.2 eq.) was added dropwise via a syringe over 10 minutes, accompanied by the evolution of a white gas, and the reaction was stirred for an additional 30 minutes. After 30 minutes, the

gas had disappeared, and the reaction mixture was a milky yellow. Benzhydrol (11.0g, 59.7 mmol, 1.26 eq.) that had been dried under high vacuum for 1.5h was then added, and the reaction was stirred for 18h while warming to room temperature.

After stirring overnight, the reaction mixture was yellow with a white precipitate. TLC analysis (5% MeOH in DCM) indicated conversion to one major product with a few minor impurities, including some starting material. To the reaction mixture, 40 mL ethanol was added, and the reaction mixture clarified completely. The solvent was evaporated and the residues dissolved in 200 mL dichloromethane. The organic phase was extracted with 100 mL 10% aqueous citric acid, 150 mL 5% sodium bicarbonate, and then dried with magnesium sulfate. The dichloromethane was removed and the residues dissolved in 100 mL boiling methanol and left to cool slowly. The crystallized product was filtered, washed with diethyl ether, and dried in vacuo to yield 16.04g of a white powder (64% Yield).

1H NMR (DMSO): 11.00 ppm (s, 1H); 8.50 ppm (s, 1H); 7.2-7.6 ppm (m, 15H); 6.80 ppm (s, 1H); 5.18 ppm (s, 2H); 5.16 ppm (s,2H).

11. Bhoc Guanine-9-Acetic Acid

THF (100 mL) was loaded with NaH (3.0g of a 60% dispersion in mineral oil, 76.0 mmol, 5.0 eq.) and cooled to $-78\,^{\circ}C$ in dry ice/acetone under an inert atmosphere. 3-hydroxypropionitrile (4.85 mL, 76.0 mmol, 5.0 eq.) was added and the reaction warmed to $0\,^{\circ}C$ on ice. After 1h stirring, gas evolution had mostly stopped and the solution was much more viscous. Benzyl Bhoc 2-Amino-6-chloropurine acetate (8.0g, 15.2 mmol, 1 eq.) was added, and the reaction was warmed to room temperature and stirred. Immediately upon addition of starting material, the solution was a slightly translucent yellow. Over the course of the next two hours, a white precipitate formed.

After 5h, the THF was evaporated to yield a yellow solid and water (50 mL) was added. The yellow solid dissolved, and the resulting solution was adjusted to pH

3.1 with 20% citric acid (~80 mL). A large amount of white solid precipitated. The mixture was cooled on ice, filtered, and washed with water. Crude material was dissolved in ~150 mL boiling methanol and left to crystallize overnight. Residual mineral oil rose to the top and was skimmed off. After crystallization, product was filtered, washed with cold methanol (100 mL) and diethyl ether (200 mL) and dried in vacuo. Low-res ESI-MS was acquired on an Advion single quadrupole instrument in negative mode with ammonium formate buffer.

1H NMR (DMSO): 11.7 ppm (s, 1H); 11.2 ppm (s, 1H); 7.9 ppm (s, 1H); 7.2-7.6 ppm (m, 10H); 6.8 ppm (s, 1H); 4.8 ppm (s, 2H)

ESI-MS (Negative Mode): m/z calculated: 418.11 Da; m/z observed: 418.1 Da, $[M-H]^{-1}$

5.3.4 Nucleobase acetic acid spectra

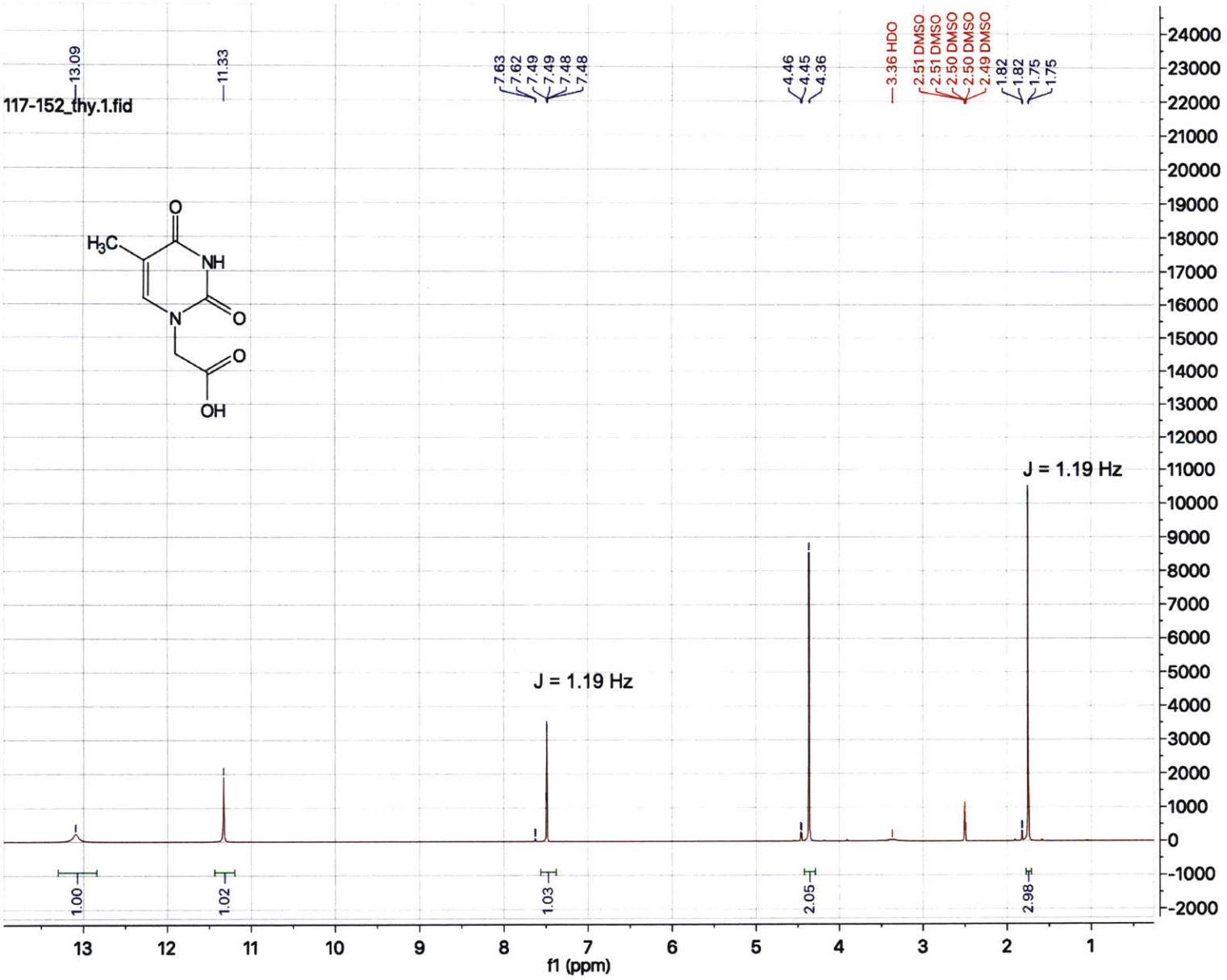

Figure 5-6: Thymine-1-Acetic Acid ^{1}H NMR

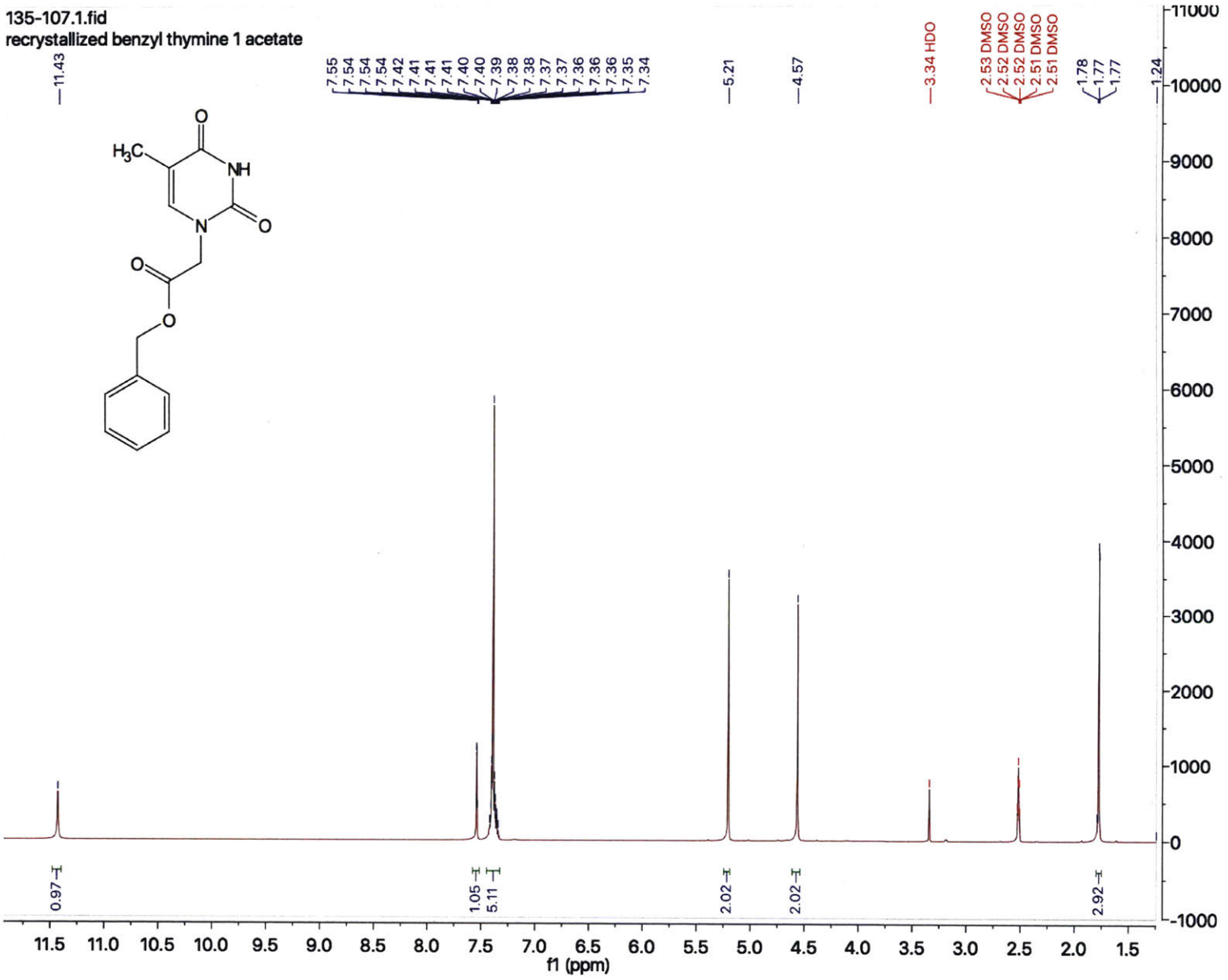

Figure 5-7: Benzyl Thymine-1-Acetate ^{1}H NMR

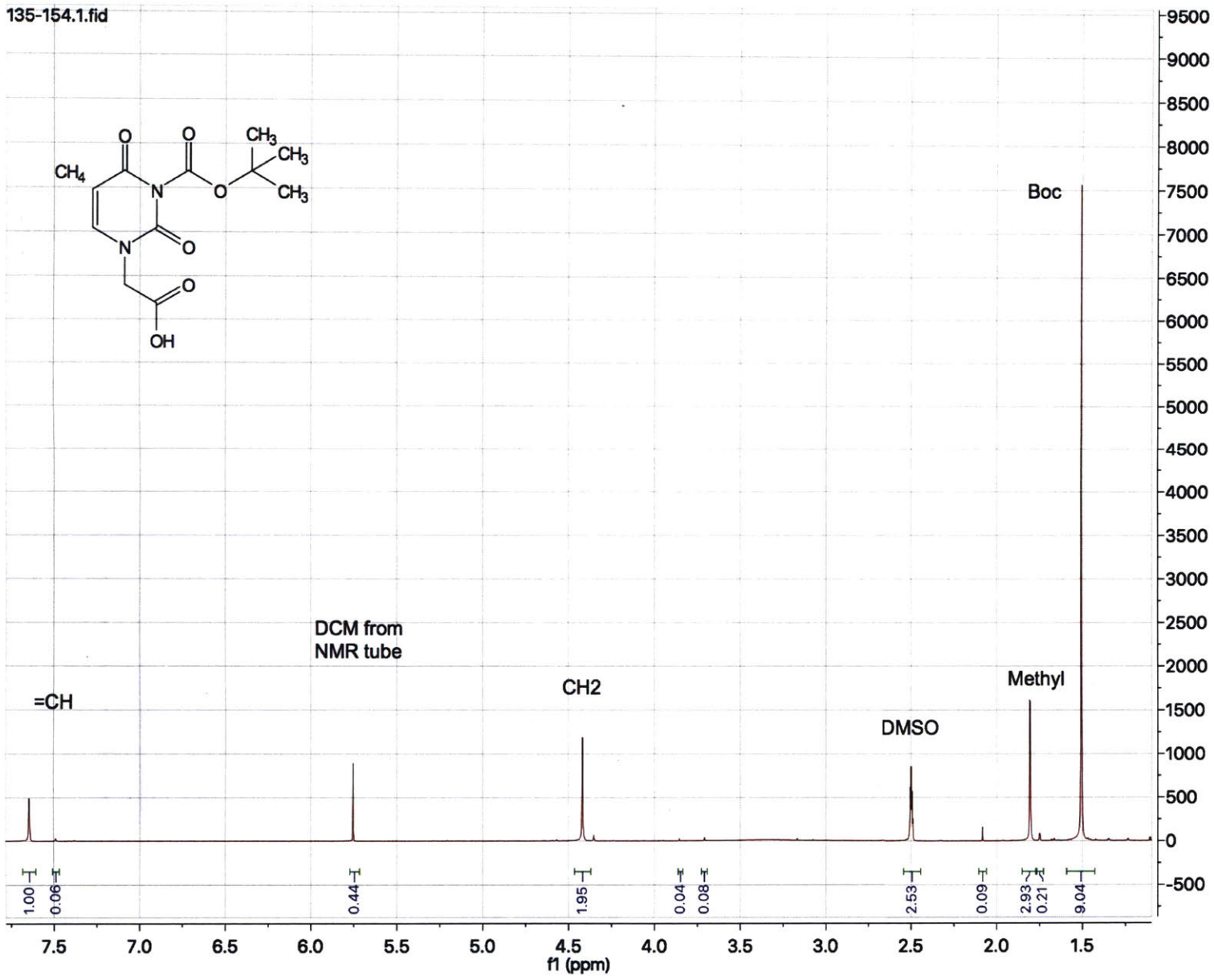

Figure 5-8: Boc Thymine-1-Acetic Acid ^{1}H NMR

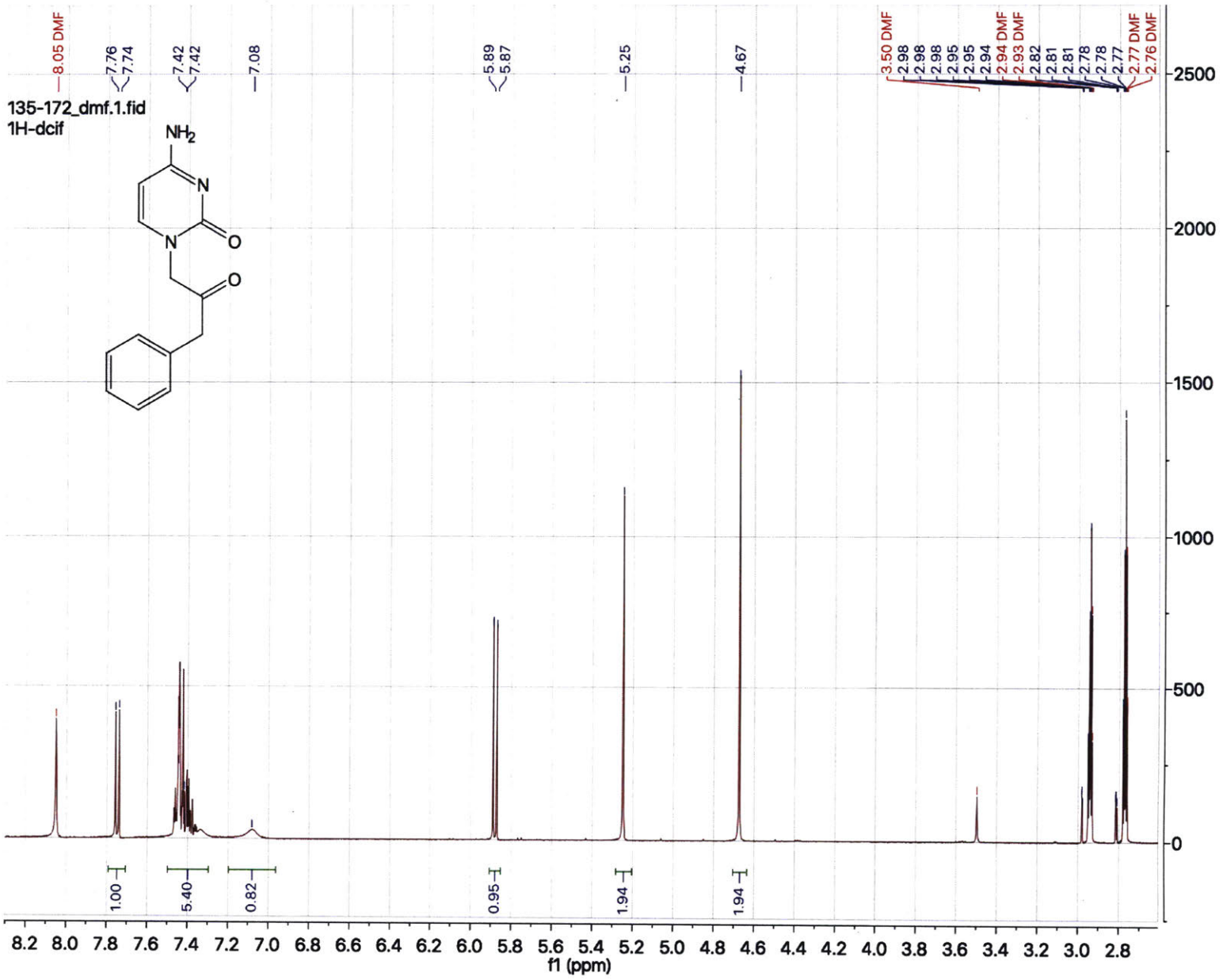

Figure 5-9: Benzyl Cytosine-1-Acetate ^{1}H NMR

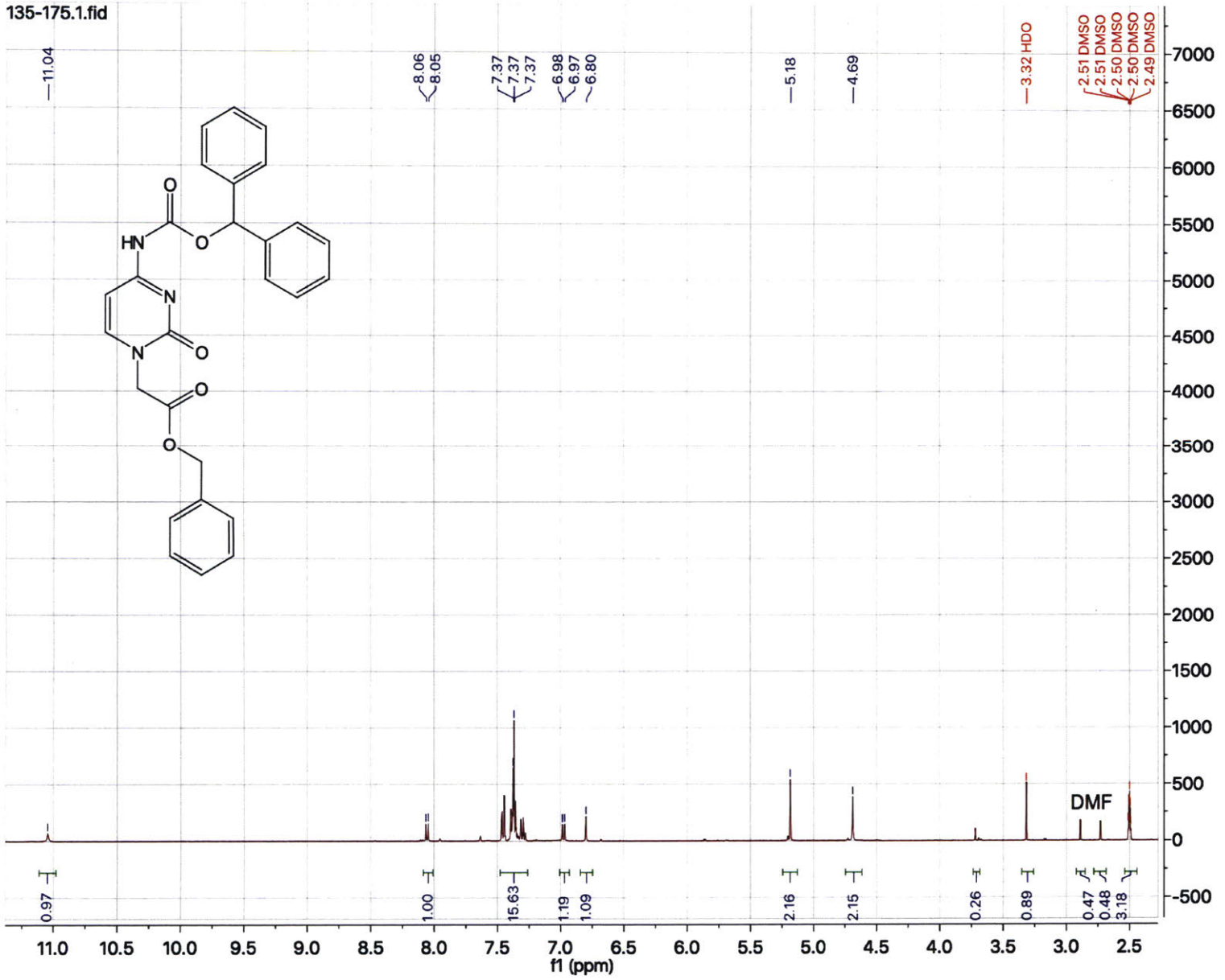

Figure 5-10: Benzyl Bhoc Cytosine-1-Acetate ^{1}H NMR

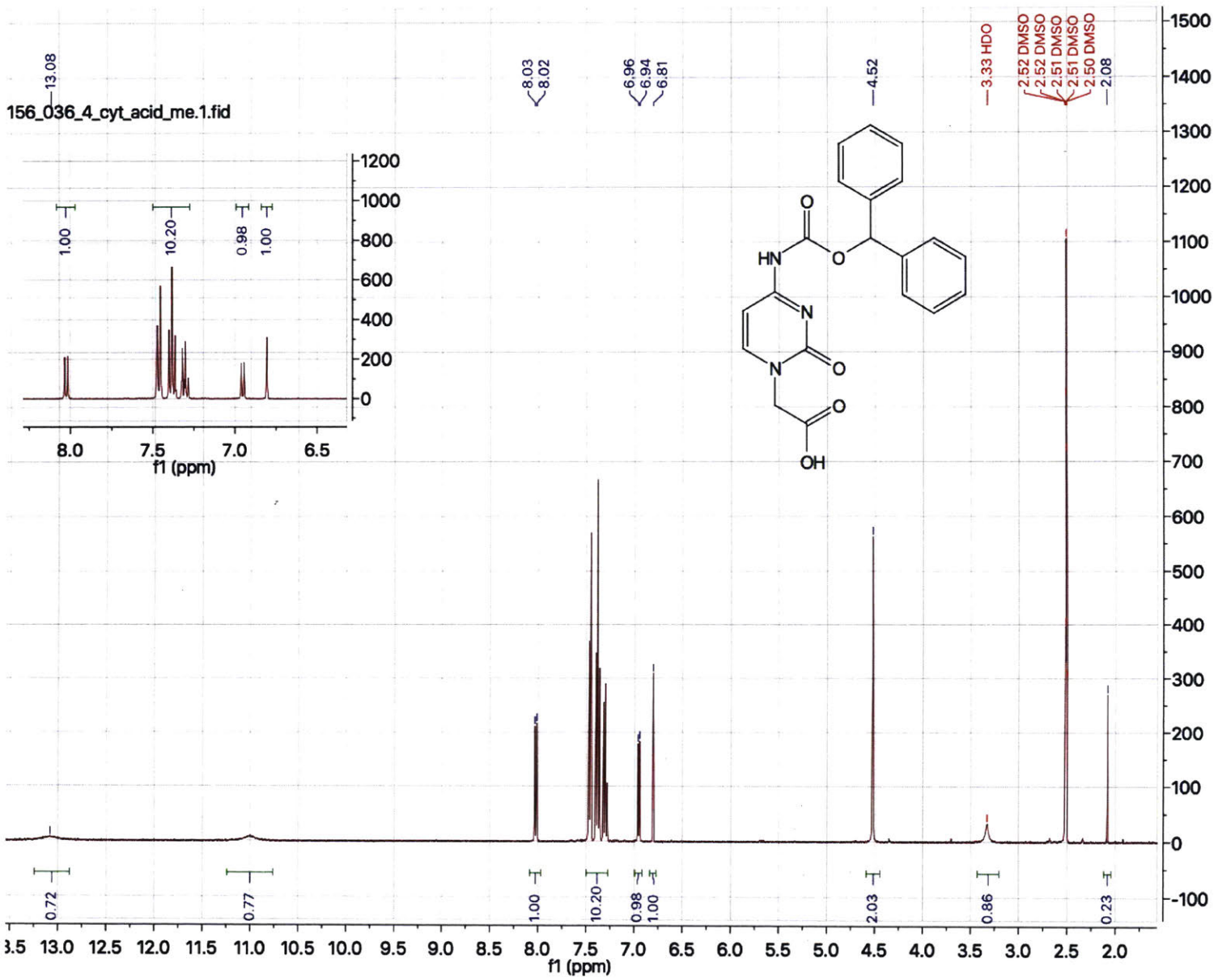

Figure 5-11: Bhoc Cytosine-1-Acetic Acid ^{1}H NMR

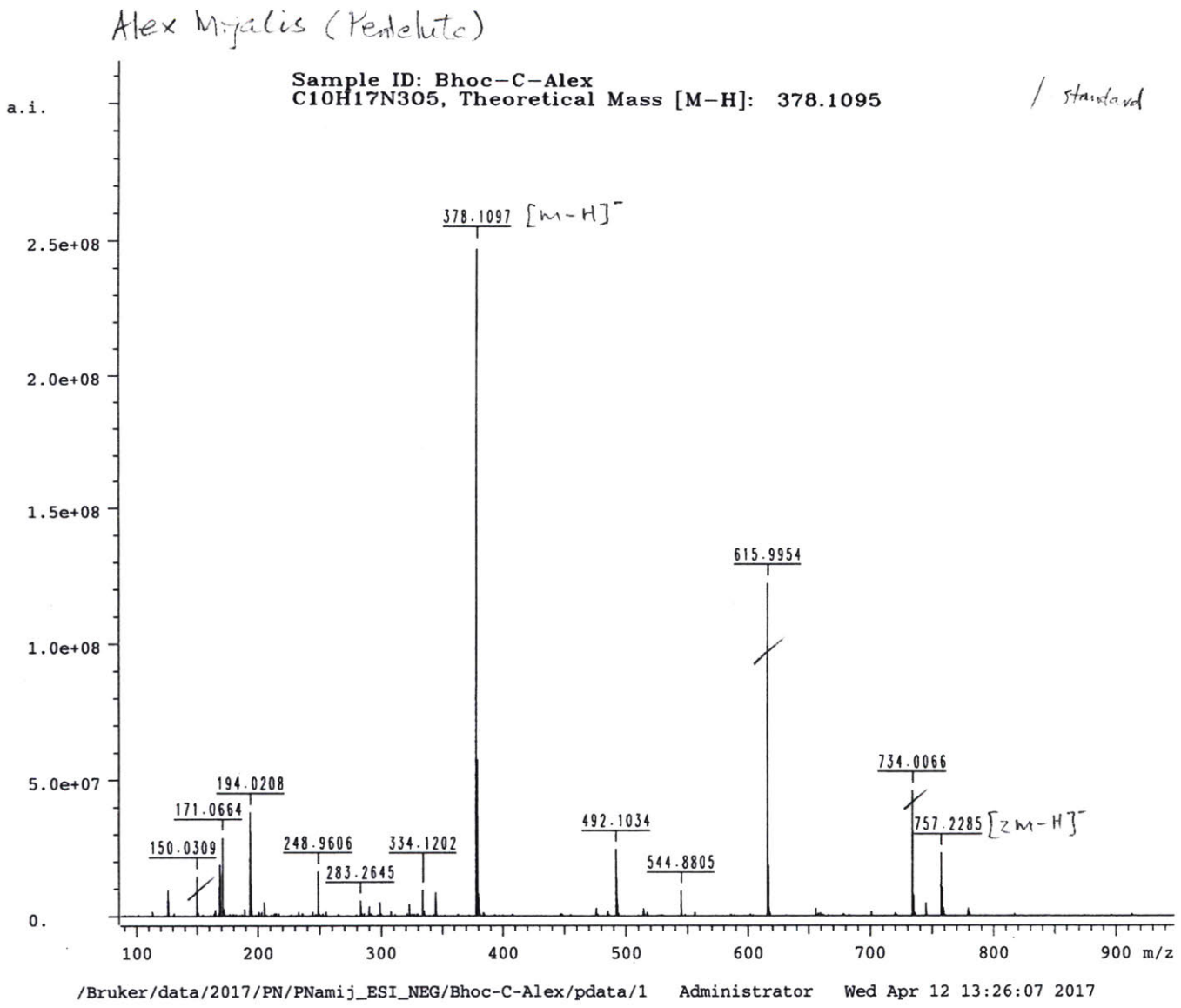

Figure 5-12: Bhoc Cytosine-1-Acetic Acid FTMS

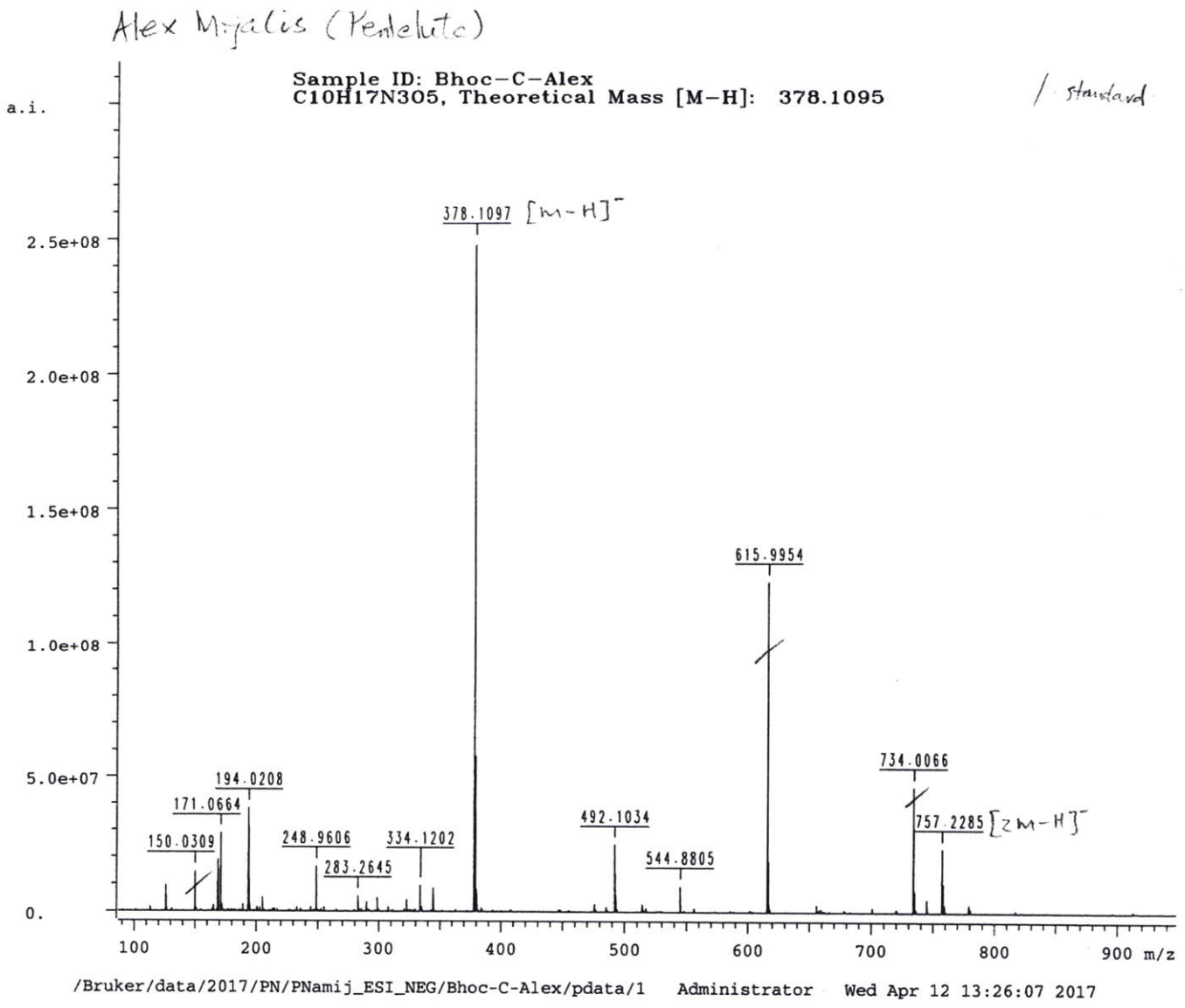

Figure 5-13: Benzyl Adenine-9-Acetate ^{1}H NMR

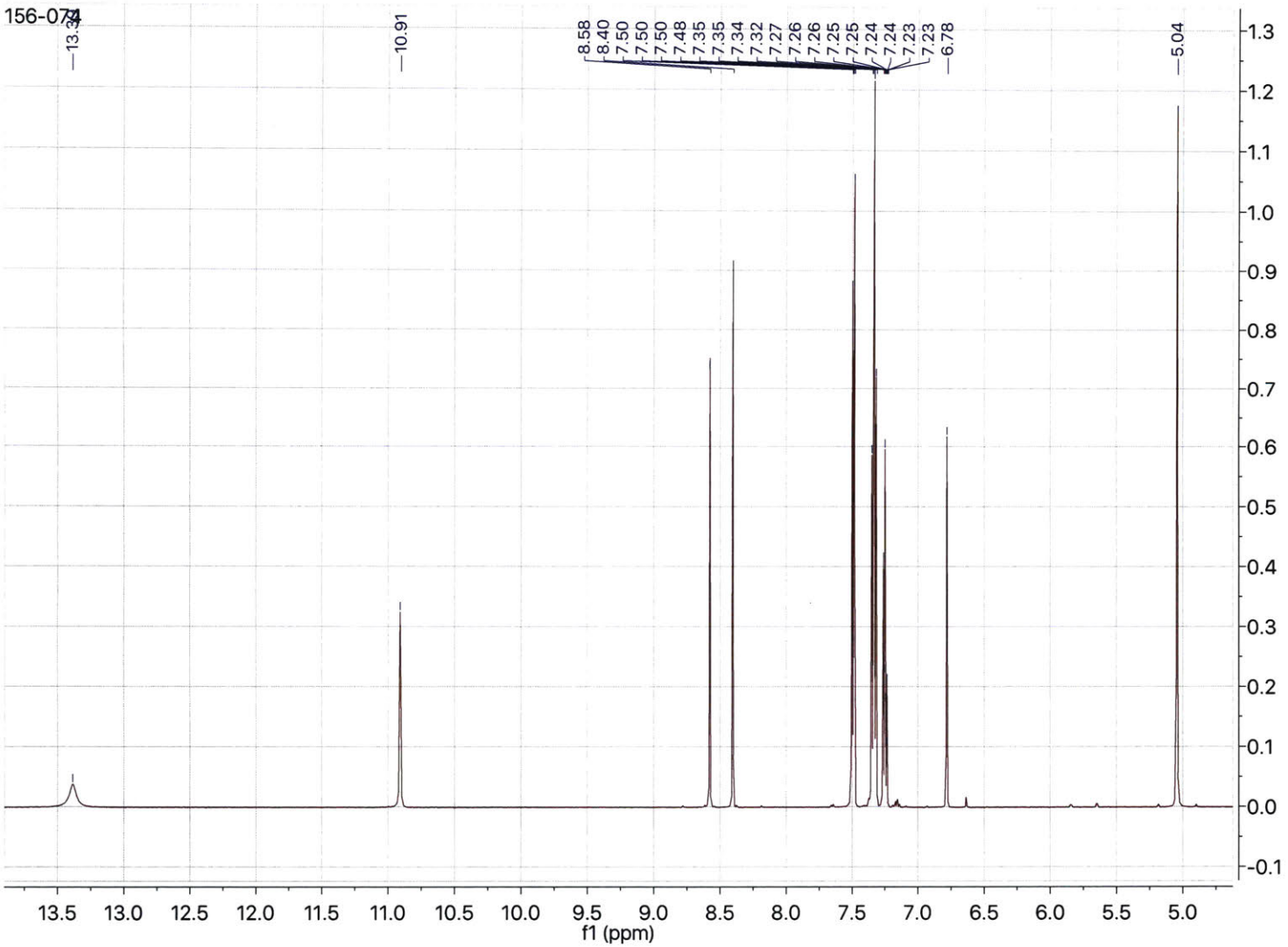

Figure 5-14: Bhoc-Adenine-9-Acetic Acid ^{1}H NMR

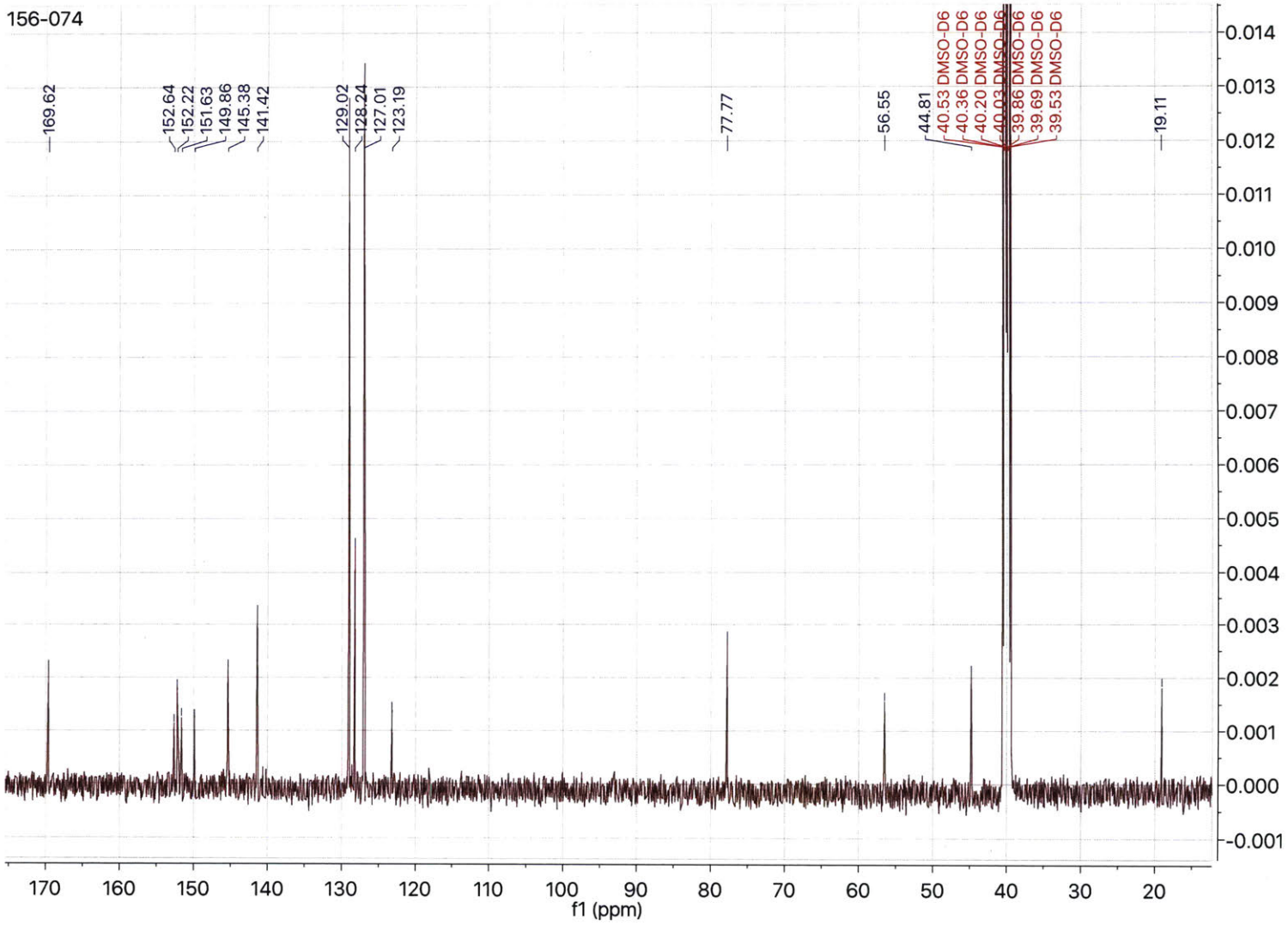

Figure 5-15: Bhoc-Adenine-9-Acetic Acid ^{13}C NMR

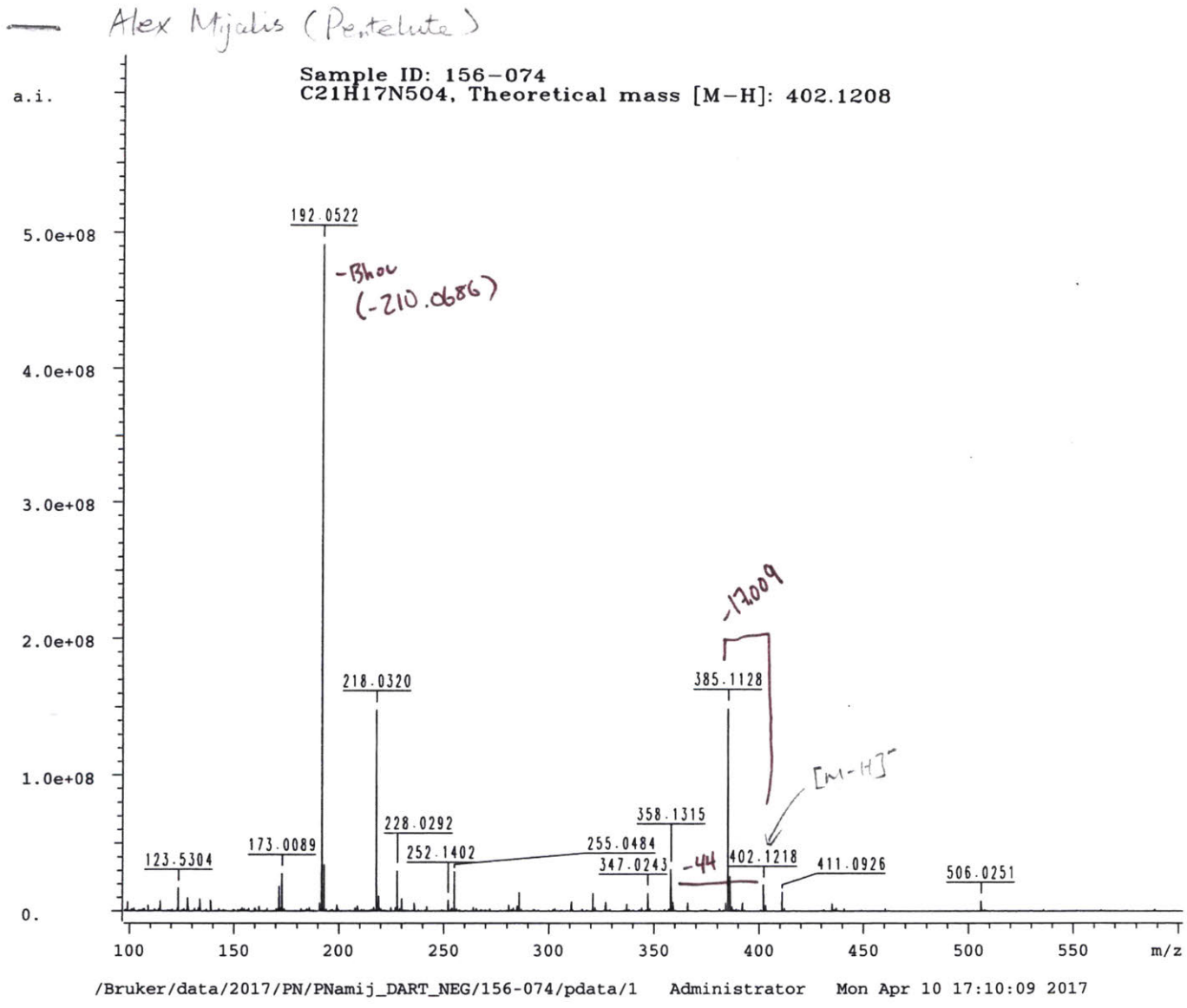

Figure 5-16: Bhoc-Adenine-9-Acetic Acid HRMS

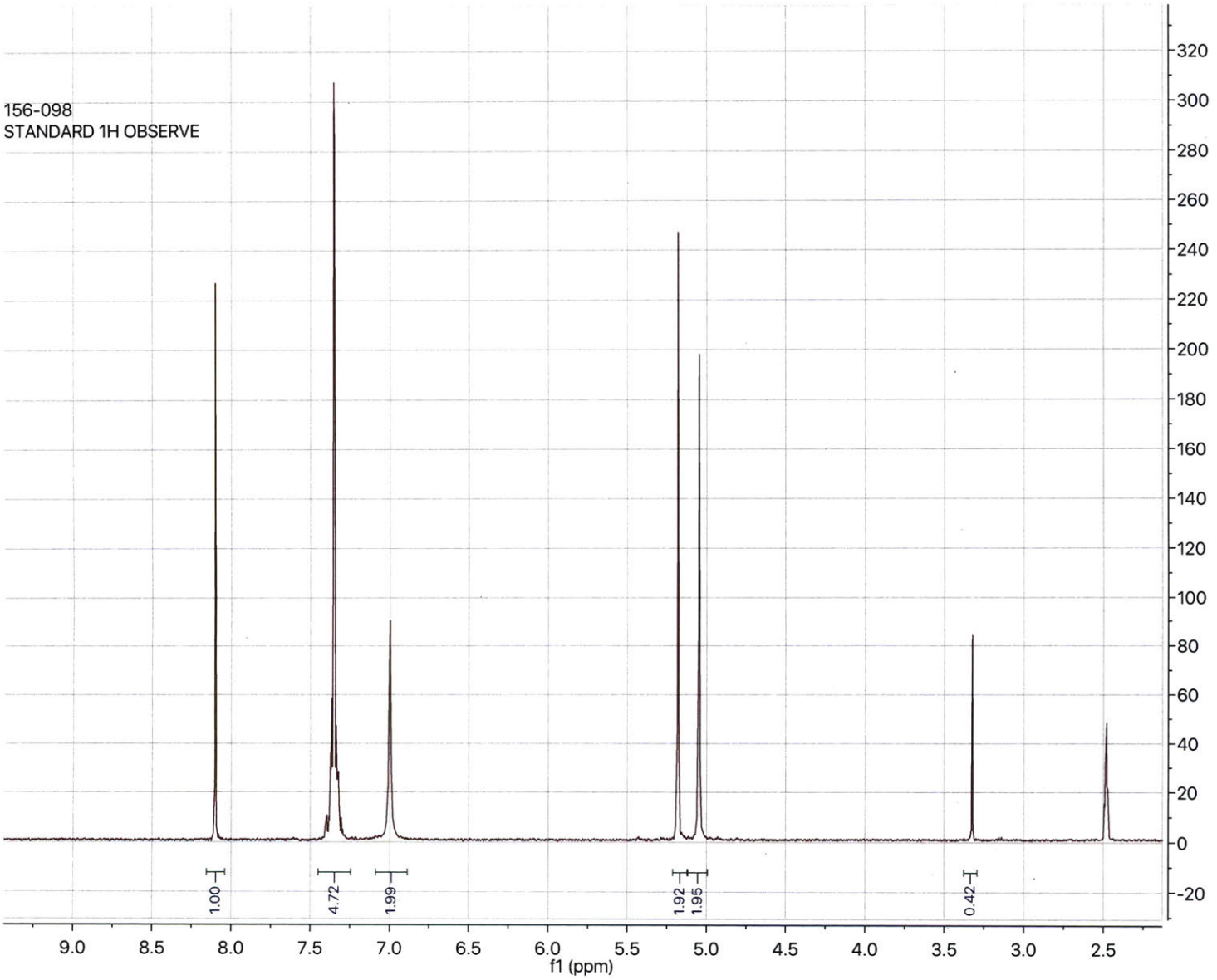

Figure 5-17: Benzyl 2-Amino-6-Chloropurine-9-Acetate ^{1}H NMR

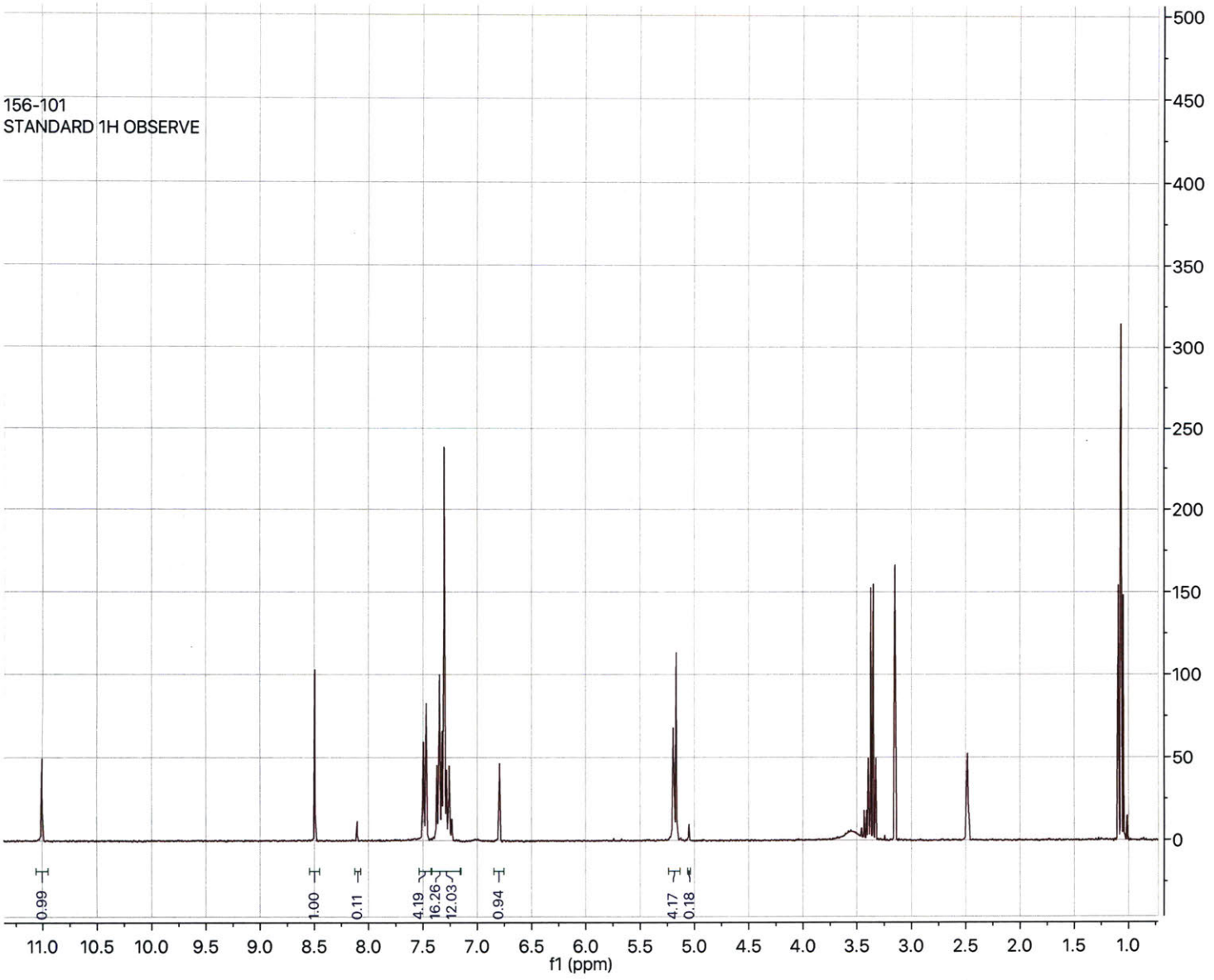

Figure 5-18: Bhoc Benzyl 2-Amino-6-Chloropurine-9-Acetate ^{1}H NMR

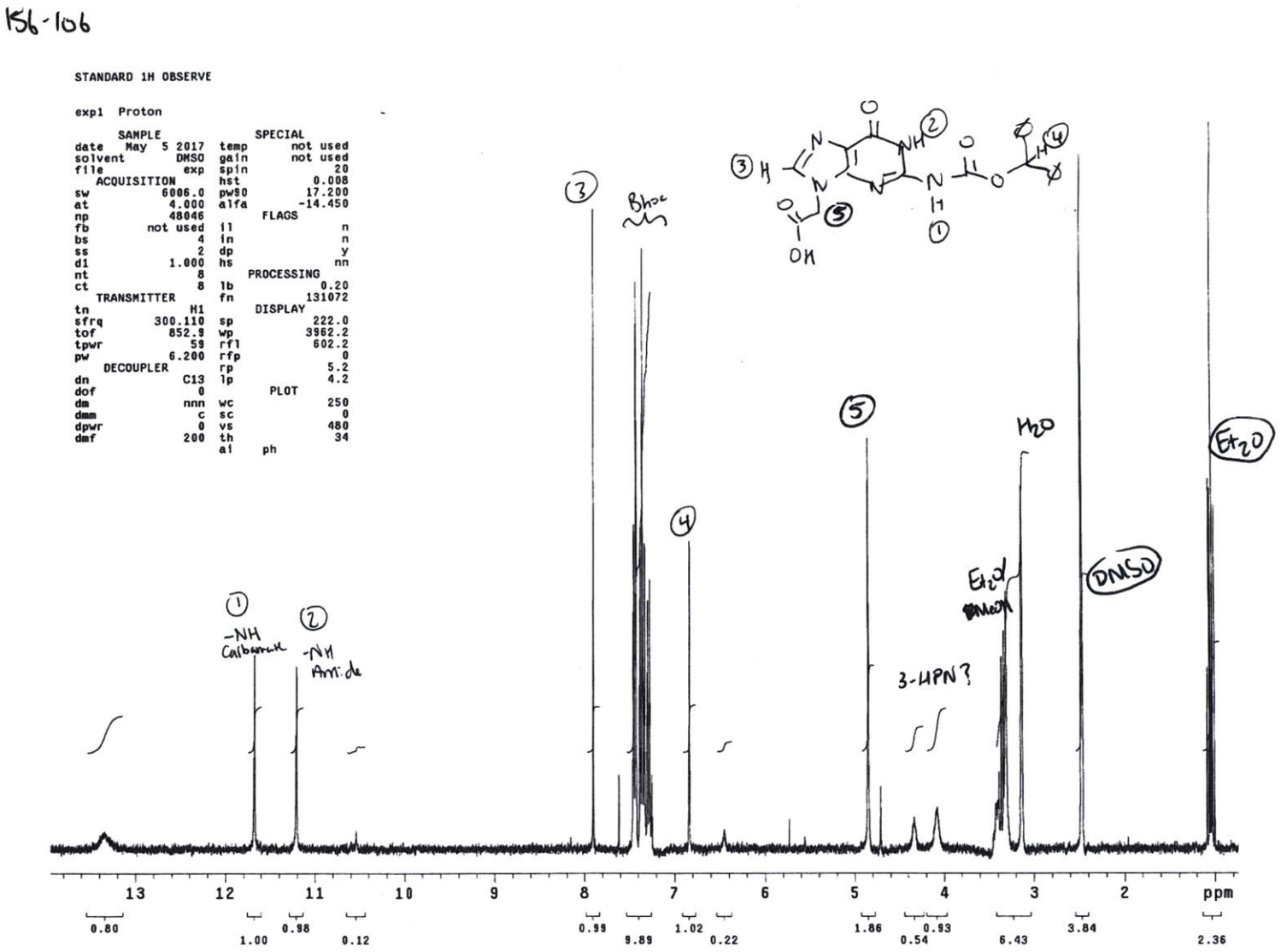

Figure 5-19: Bhoc Guanine-9-Acetic Acid ^{1}H NMR

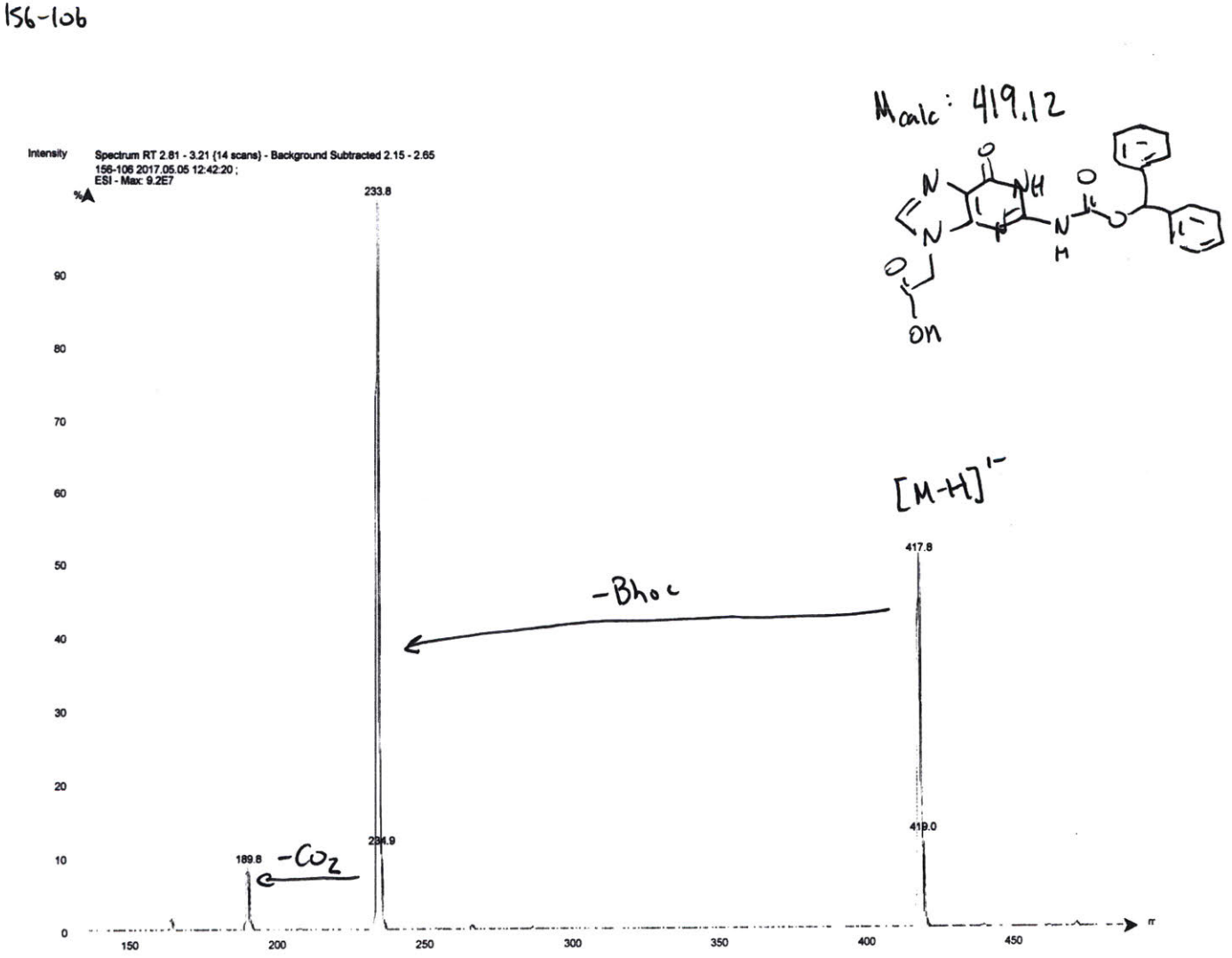

Figure 5-20: Bhoc Guanine-9-Acetic Acid ESI-MS

References

1. Huang, Y., Dey, S., Zhang, X., Sönnichsen, F. & Garner, P. The α-Helical Peptide Nucleic Acid Concept: Merger of Peptide Secondary Structure and Codified Nucleic Acid Recognition. *Journal of the American Chemical Society* **126,** 4626–4640. ISSN: 0002-7863 (Apr. 2004).

2. Garner, P., Dey, S., Huang, Y. & Zhang, X. Modular Nucleic Acid Surrogates. Solid Phase Synthesis of α-Helical Peptide Nucleic Acids (αPNAs). *Organic Letters* **1,** 403–406. ISSN: 1523-7060 (Aug. 1999).

3. Tsuyoshi, T., Akihiko, U. & Hisakazu, M. Nucleobase Amino Acids Incorporated into the HIV-1 Nucleocapsid Protein Increased the Binding Affinity and Specificity for a Hairpin RNA. en. *ChemBioChem* **3,** 543–549. ISSN: 1439-7633 (June 2002).

4. Tsuyoshi, T., Dewi, Y. & Hisakazu, M. Utilization of L-α-Nucleobase Amino Acids (NBAs) as Protein Engineering Tools: Construction of NBA-Modified HIV-1 Protease Analogues and Enhancement of Dimerization Induced by Nucleobase Interaction. *ChemBioChem* **7,** 729–732. ISSN: 1439-7633 (May 2006).

5. Gomez-Martinez, P., Dessolin, M., Guibé, F. & Albericio, F. N α -Alloc Temporary Protection in Solid-Phase Peptide Synthesis. The Use of Amine–borane Complexes as Allyl Group Scavengers. en. *Journal of the Chemical Society, Perkin Transactions 1,* 2871–2874 (1999).

6. C. Bruno, N., T. Tudge, M. & L. Buchwald, S. Design and Preparation of New Palladium Precatalysts for C–C and C–N Cross-Coupling Reactions. en. *Chemical Science* **4,** 916–920 (2013).

7. Srinivasu, P., Zbigniew, P. & Nicolas, W. Expanding the Scope and Orthogonality of PNA Synthesis. en. *European Journal of Organic Chemistry* **2008,** 3141–3148. ISSN: 1099-0690 (June 2008).

8. Debaene, F. & Winssinger, N. Azidopeptide Nucleic Acid. An Alternative Strategy for Solid-Phase Peptide Nucleic Acid (PNA) Synthesis. English. *Organic Letters* **5,** 4445–4447. ISSN: 1523-7060, 1523-7052 (Nov. 2003).

9. Sascha, B., Martin, v. M., Ulrich, H. & Thomas, C. Investigation of the Pathways of Excess Electron Transfer in DNA with Flavin-Donor and Oxetane-Acceptor Modified DNA Hairpins. en. *Chemistry – A European Journal* **12,** 6469–6477. ISSN: 1521-3765 (Aug. 2006).

Appendix A

Automated flow click chemistry on solid support

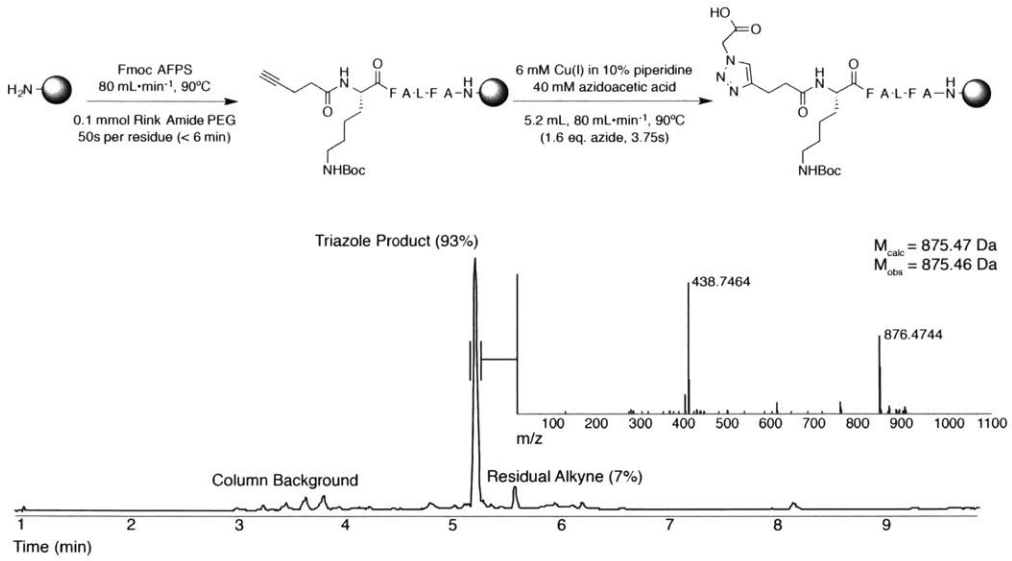

Figure A-1: Rapid click chemistry on the solid phase

In addition to performing amide bond formation in flow, we thought that the AFPS could be used to perform other types of chemistry on the solid phase. As a proof of concept, we performed click chemistry in flow by first synthesizing the polypeptide KFALFA-$CONH_2$ on resin. Next, we coupled 4-pentynoic acid to the N-terminus. We then performed a click coupling to the resin-bound alkyne by flowing 2.6 mL each of (1) 12 mM CuI in 20% piperidine and (2) 80 mM azidoacetic acid at

80 mL/min and 90 °C over the peptidyl resin. By LC-MS analysis, the reaction went to 93% conversion with just 1.6 eq. azide in under 5 seconds.

Appendix B

Monitoring Peptide Aggregation

As discussed previously (see Section 1.3) a longstanding hypothesis for why certain peptide syntheses fail is aggregation of the growing polypeptide chain on the solid support. This is characterized by shrinking of the solid support as the peptide stops associating with solvent molecules, a drastic slowing of the Fmoc removal and aminoacylation steps. Synthesis of an aggregating peptide often yields a complex, multicomponent HPLC chromatogram, as we observed in Figures 3-4 and 2-7. Time-resolved analysis of the Fmoc removal peaks with flow peptide synthesis provides a metric to observe this phenomenon (see Figure 2-14 and Figure 1-9). As Fmoc removal slows, the observed width of the Fmoc removal signal after piperidine treatment increases and the height decreases. This can be accompanied by an immediate, drastic change in area, as observed in 3-4, indicating truncation or deletion, or a delayed decrease in area, as observed in 2-7, where the Fmoc removal slows down but does not lead to immediate failure of the chemistry.

Since the installation of the first Automated Flow Peptide Synthesizer in the Pentelute Lab, we have collected data on over 3,000 peptide sequences made over the past two to three years. Can this data be mined to detect aggregating sequences, and if so, can we learn anything about the sequences that lead to this phenomenon? The answer to this question may help predict difficult-to-synthesize sequences before synthesis.

To start answering this question, I began filtering the AFPS data to find aggre-

gating sequences. The search was restricted to sequences that contained the standard 20 amino acids plus azidopentynoic acid (labeled as "Z" in these chromatograms). These "Z"-containing sequences were the result of the synthesis of a number of cell-penetrating peptides for click chemistry. The deprotection peaks were then integrated, and all files where there was a large increase in the Fmoc removal peak width and decrease in the Fmoc removal peak height during the synthesis were plotted. These values are plotted as a percentage of the first deprotection, and syntheses were selected where the difference between the normalized height and width was greater than 20%). Because some syntheses failed or had anomalous UV data, the resulting ~250 chromatograms were sorted manually. The resulting dataset contained 181 aggregating sequences. The serial numbers, sequences, and aggregating motifs are plotted in Table B.1. The aggregating motif is the sequence synthesized on resin before the onset of aggregation.

From this data, we can start to understand the aggregating nature of certain amino acid homopolymers, such as poly(Gly) (SN 2396), which aggregates after 11 residues, and poly(Val) (SN 2556, 2557), which aggregates after just 6 residues. Serine, Isoleucine, Leucine, Alanine, Tyrosine, Methionine, and Phenylalanine also aggregate after a finite number of residues are added. Copolymers of Serine and Glycine also tend to aggregate after 10-15 residues, as observed in sequences with serial numbers 1752 197, 1746, 1747, 1754, and 1813, among others.

By importing the aggregating motifs into a python script in the form of a dictionary {serial-number: sequence ... }, the sequences were aligned using the python script at the end of this section. The output of the script is shown in Figure B-1.

Table B.1: Aggregating peptides identified in the AFPS dataset. Sequences are written from the N to the C terminus.

Serial Number	Sequence	Aggregating Motif
35	AGYLL GKINL KALAA LAKKI LC	AALAK KILC
36	YADAI FTNSY RKVLG QLSAR KLLQD ILSA	SYRKV LGQLS ARKLL QDILS A
175	GSGSG SGSGS GS	GSGSG SGSGS GS
182	SSDPL VVAAS IIGIL HLILW ILDRL	PLVVA ASIIG ILHLI LWILD RL
197	GSGSG SGSGS GS	SGSGS GSGSG S
202	DICLP RWGCL WGGGG SAVGA LEGPR NQDWL GVPRQ L	ICLPR WGCLW GGGGS AVGAL EGPRN QDWLG VPRQL
212	CPFLP VVYGL LKGRR	PFLPV VYGLL KGRR
214	GSGSG SGSGS GS	SGSGS GSGSG S
258	QVTCA GAGLL KG	QVTCA GAGLL KG
272	GLFDI IKKIA ESFC	IIKKI AESFC
281	MHMHK TTSYR IRVLV GVDVY RMSHT CLTSS SG	VLVGV DVYRM SHTCL TSSSG
301	MGHLH ICMVW RVNTS GHILS VGAKS YSSHK TG	VGAKS YSSHK TG
314	FLGGS GCKGS GGS	FLGGS GCKGS GGS
351	DWGGQ HHGLR EVLAA ALFAS CLWGZ	AALFA SCLWG Z

Continued on next page

Continued from previous page

Serial Number	Sequence	Aggregating Motif
361	CILKP QMPWE LWDIL QEISP EEIQP NPPSS GMLGI IIMMT L	PQMPW ELWDI LQEIS PEEIQ PNPPS SGMLG IIIMM TL
362	GGGGG SGGSG GSLL	GGSGG SGGSL L
364	SSGNQ NLYAM YQLSH FQSIS VLZ	LYAMY QLSHF QSISV LZ
1391	CGLAF LGFLG AAGST MGAWS QPKKK RKVA	FLGFL GAAGS TMGAW SQPKK KRKVA
1481	MQIFV KTLTG KTITL EVEPS DTIEN VKAKI QDKEG IPPDQ QR	TLTGK TITLE VEPSD TIENV KAKIQ DKEGI PPDQQ R
1612	GGGRG KLLGY VWPYC L	RGKLL GYVWP YCL
1617	LGMAP ALQPT QGAMP AFASA FQRRA GGVLV ASHLQ SFLEV SYRVL RHLAQ	SYRVL RHLAQ
1648	CAGYL LGKIN LKALA ALAKK ILC	LAALA KKILC
1658	GGGGS GGSGG SGGSG GSGGS GGS	GGSGG SGGSG GSGGS
1693	CIGAV LKCLT TGLPA LISWI KRKRQ Q	TGLPA LISWI KRKRQ Q
1699	GGGRG LRRLG RKIAH GVKKY GPTVL RIIKK YG	KIAHG VKKYG PTVLR IIKKY G

Continued on next page

Continued from previous page

Serial Number	Sequence	Aggregating Motif
1731	QGLIG DICLP RWGCL WGDSV GGGGS	LWGDS VGGGG S
1742	DICLP RWGCL WGGGG S	WGCLW GGGGS
1743	QGLIG DICLP RWGCL WGDSV GGGGS	WGDSV GGGGS
1746	TFEDL LHYYG SGSGS GSGSG S	SGSGS GSGSG S
1747	GSGSG SGSGS GS	GSGSG SGSGS GS
1754	GSGSG SGSGS GS	GSGSG SGSGS GS
1813	GGGGS GGSGG SGGSG GSGGS V	GSGGS GGSV
1821	SSDPL VVAAS IIGIL HLILW ILDRL	PLVVA ASIIG ILHLI LWILD RL
1877	DLDLE MLAPY IPMDD DFQL	PMDDD FQL
1902	GLAFL GFLGA AGSTM GAWSQ PKKKR KV	LGFLG AAGST MGAWS QPKKK RKV
1966	GSGSG SGSGS GS	SGSGS GSGSG S
1969	GSGSG SGSGS GS	SGSGS GSGSG S
2048	CAVGA LEGPR NQDWL GVPRQ L	RNQDW LGVPR QL
2093	GSGSG SGSGS GS	SGSGS GSGSG S
2094	GSGSG SGSGS GS	GSGSG SGSGS
2101	CPFLP VVYGL LKGR	PFLPV VYGLL KGR
2108	GLAFL GFLGA AGSTM GAA	LAFLG FLGAA GSTMG AA

Continued on next page

Continued from previous page

Serial Number	Sequence	Aggregating Motif
2139	GSGSG SGSGS GS	SGSGS GSGSG S
2187	KNKGG LKLGV MS	LKLGV MS
2192	PCWLM HW	PCWLM HW
2199	KNKGG LKLGV MSCLA P	LKLGV MSCLA P
2221	LCFKE SY	LCFKE SY
2225	LCSGY TF	LCSGY TF
2229	LCMLY TY	LCMLY TY
2245	RRFFG CYLTC ILKTE EGNLP STGG	YLTCI LKTEE GNLPS TGG
2247	RRFFG CYLTC ILKTE EGN	YLTCI LKTEE GN
2295	KAGCK AGCKA GCKAG CKAGC KAGC	KAGCK AGCKA GC
2296	KAGCK AGCKA GCKAG C	KAGCK AGCKA GC
2345	KRVKA GYLLG KINLK ALAAL AKKIL	ALAAL AKKIL
2346	AGYLL GKINL KALAA LAKKI LKRVK	AALAK KILKR VK
2347	GGGGG TSFAE YWALL SP	GTSFA EYWAL LSP
2396	GGGGG GGGGG GGGGG GGGGG	GGGGG GGGGG G
2397	LLLLL LLLLL LLLLL LLLLL	LLLLL LLLL

Continued on next page

Continued from previous page

Serial Number	Sequence	Aggregating Motif
2399	AAAAA AAAAA AAAAA AAAAA	AAAAA AAA
2401	IIIII IIIII IIIII IIIII	IIIII
2427	AAAAA AAAAA	AAAAA AAA
2429	LLLLL LLLLL	LLLLL LLLL
2431	IIIII IIIII	IIIII
2433	GGGGG GGGGG	GGGGG GGGGG
2448	SSSSS SSSSS	SSSSS SSS
2449	SSSSS SSSSS SSSSS SSSSS	SSSSS SSS
2451	YYYYY YYYYY	YYYYY YYYYY
2452	YYYYY YYYYY YYYYY YYYYY	YYYYY YYYYY
2487	TTTTT TTTTT TTTTT TTTTT	TTTTT TTTTT T
2492	GLAFL GFLGA AGSTM GAWSQ PKKKR KV	LGFLG AAGST MGAWS QPKKK RKV
2501	VDYTI TVYAV TGRGD SPASS KPISI NY	TITVY AVTGR GDSPA SSKPI SINY
2509	MMMMM MMMMM MMMMM MMMMM	MMMMM MMMMM M
2511	FFFFF FFFFF	FFFFF FFFFF
2512	FFFFF FFFFF FFFFF FFFFF	FFFFF FFFFF
2536	TITVY AVTGR GDSPA SSKPI S	TITVY AVTGR GDSPA SSKPI S
2556	VVVVV VVVVV	VVVVV V

Continued on next page

Continued from previous page

Serial Number	Sequence	Aggregating Motif
2557	VVVVV VVVVV VVVVV VVVVV	VVVVV V
2561	AGYLL GKINL KALAA LAKKI L	ALAAL AKKIL
2563	GLAFL GFLGA AGSTM GAWSQ PKKKR KV	LGFLG AAGST MGAWS QPKKK RKV
2578	MCAAE GFGLL KG	MCAAE GFGLL KG
2582	KGYCA TFGLL KG	KGYCA TFGLL KG
2688	LLIIL RRRIR KQAHA HSKLL IILRR RIRKQ AHAHS K	AHSKL LIILR RRIRK QA-HAH SK
2754	GLAFL GFLGA AGSTM GAWSQ PKKKR KV	LGFLG AAGST MGAWS QPKKK RKV
2763	DVGEC THCGG ADCTG SCTCT NWSSC VCMNF SSSEE GECGC TCZ	FSSSE EGECG CTCZ
2769	DVGEC THCGG ADCTG SCTCT NWSSC VCKYF SSSGA GECGC ACYZ	FSSSG AGECG CACYZ
2776	GGGGG SGGSG GSGGS GGSGG SGGS	GGSGG SGGSG GS
2784	MHMHK TTSYR IRVLV GVDYR MSHTC LTSSS G	VLVGV DYRMS HTCLT SSSG
2797	MNGHY PCYLI TSVLV GATTS GVPVV VHLRV G	VLVGA TTSGV PVVVH LRVG

Continued on next page

Continued from previous page

Serial Number	Sequence	Aggregating Motif
2805	MRSTH QRVRR PRNLC SFKHK WLIKF LKTLT G	WLIKF LKTLT G
2925	AGYLL GKINL KASAA LAKKS L	ASAAL AKKSL
2926	AGYLL GKINL KACAA LAKKC L	KACAA LAKKC L
2935	LTFEH WWAQL CS	FEHWW AQLCS
2957	VDYTI TVYAV TG	VDYTI TVYAV TG
2984	MNGHY PCYLI TSVLV GATTS GVPVV VHLRV G	VLVGA TTSGV PVVVH LRVG
3001	DAEFR HDSGY EVHHQ KLVFF AEDVG SNKGA IIGLM VGGVV IA	LMVGG VVIA
3004	MRSTH QRVRR PRNLC SFAHK WLIKF LKTLT G	LIKFL KTLTG
3008	MGHLA ICMVW RVNTS GHILS VGHKS YSSHK TG	ILSVG HKSYS SHKTG
3009	MGALH ICMVW RVNTS GHILS VGHKS YSSHK TG	ILSVG HKSYS SHKTG
3015	MLFMR LTKKT MATKF CPFRA KRKHR ERRAL YG	LFMRL TKKTM ATKFC PFRAK RKHRE RRALY G
3038	MCPFL PVVYG GSGGS GGSMC PFLPV VY	VVYGG SGGSG GSMCP FLPVV Y

Continued on next page

Continued from previous page

Serial Number	Sequence	Aggregating Motif
3041	GLAFL GFLGA AGSTM GAWSQ PKKKR KV	LGFLG AAGST MGAWS QPKKK RKV
3042	GLAFL GFLGA AGSTM GAWSQ PKKKR KV	LGFLG AAGST MGAWS QPKKK RKV
3066	AGYLL GKINL KASAA LAKKS L	ASAAL AKKSL
3097	GCPRI LMRCK QD	PRILM RCKQD
3106	AGYLL GKINL KASAA LAKKS L	SAALA KKSL
3107	AGYLL GKINL KACAA LAKKC L	AALAK KCL
3124	GLEQL ESIIN FEKLT EWTSS	FEKLT EWTSS
3133	GGGGG SGGSG GSGGS	GGSGG SGGSG GS
3143	AGYLL GKINL KALAA LAKKI L	ALAAL AKKIL
3148	DAEFR HDSGY EVHHQ KLVFF AEDVG SNKGA IIGLM VGGVV I	LMVGG VVI
3156	AAAAA ANNAA AV	AAAAA NNAAA V
3162	CGGGS SSSSS SSRRR	SSSSS SSSRR R
3165	DAEFR HDSGY EVHHQ KLVFF AEDVG SNKGA IIGLM VGGVV I	LMVGG VVI
3166	GGGGG SGGSG GSLL	GGSGG SGGSL L
3177	FLGCS GGSGS KGS	FLGCS GGSGS KGS
3178	KGFLG GSGGS GSGCS	KGFLG GSGGS GSGCS

Continued on next page

Continued from previous page

Serial Number	Sequence	Aggregating Motif
3196	VDYTI TVYAV TGRGE SPASS KPISI NY	TITVY AVTGR GESPA SSKPI SINY
3197	TITVY AVTGR GESPA SSKPI S	ITVYA VTGRG ESPAS SKPIS
3213	GCGSS GSGCG SSGSG CG	SGSGC GSSGS GCG
3251	GSSLL SSG	GSSLL SSG
3252	GSSLL SSG	GSSLL SSG
3254	GSSLL SSG	GSSLL SSG
3271	DYVLT VEDDK	YVLTV EDDK
3272	ISTSS TIANI LAAAZ	SSTIA NILAA AZ
3276	ISTSS TIANI LAAZ	SSTIA NILAA Z
3297	SSGCS SGSGG KGS	SSGCS SGSGG KGS
3299	GGSKG GSGSS GGSGS GCS	GSGSS GGSGS GCS
3316	AGYLL GKINL KASAA LAKKS L	ASAAL AKKSL
3332	YFQIG YMISL IAFFZ	YMISL IAFFZ
3337	DVFLS TTVFL MLSTZ	FLSTT VFLML STZ
3338	TVFLM LSTTC FLKYZ	TVFLM LSTTC FLKYZ
3351	DAEFR HDSGY EVHHQ KLVFF AEDVG SNKGA IIGLM VGGVV I	IIGLM VGGVV I
3357	CILKP QMPWE LWDIL QEISP EEIQP NPPSS GMLGI IIMMT L	IIIMM TL
3366	DVFLS TTVFL MLSTZ	VFLST TVFLM LSTZ

Continued on next page

Continued from previous page

Serial Number	Sequence	Aggregating Motif
3367	TVFLM LSTTC FLKYZ	TVFLM LSTTC FLKYZ
3385	KGFLG GSGGS GSGCS	LGGSG GSGSG CS
3389	KGFLG GSGGS GSGCS	GFLGG SGGSG SGCS
3397	SSGCS SGSGG CGS	SSGCS SGSGG CGS
3398	GGSCG GSGSS GGSGS GCS	GGSGS SGGSG SGCS
3424	AGYLL GKINL KALAA LAKKI L	ALAAL AKKIL
3425	GLAFL GFLGA AGSTM GAWSQ PKKKR KVZ	GFLGA AGSTM GAWSQ PKKKR KVZ
3441	DNEPD HYILT PLTLI QRMNL LMKIS Z	PLTLI QRMNL LMKIS Z
3446	KPAWC WYTLA MCGAG YDSGT CDYMY SHCFG IKHHS SGSSS YHS	AMCGA GYDSG TCDYM YSHCF GIKHH SSGSS SYHS
3451	KIKEL LPDWG GQHHG LREVL AAALF AZ	EVLAA ALFAZ
3452	DWGGQ HHGLR EVLAA ALFAS CLWGZ	AALFA SCLWG Z
3462	DAEFR HDSGY EVHHQ KLVFF AEDVG SNKGA IIGLM VGGVV I	IIGLM VGGVV I
3482	GGGGG SGGSG GSLL	GGSGG SGGSL L
3484	DWGGQ HHGLR EVLAA ALFAS CLWGZ	AAALF ASCLW GZ

Continued on next page

Continued from previous page

Serial Number	Sequence	Aggregating Motif
3495	MNRIF HKRST YQMVF GRCSA FTSTY HVLIS YG	IFHKR STYQM VFGRC SAFTS TYHVL ISYG
3504	MNRIF HKRST YQMVF GRCSA FTSTY HVLIS YG	CSAFT STYHV LISYG
3506	MNGHY PCYLI TSVLV GATTS GVPVV VHLRV G	VLVGA TTSGV PVVVH LRVG
3512	RAELQ ASDHR PVMAI VEVEV QEVDV Z	IVEVE VQEVD VZ
3513	HRPVM AIVEV EVQEV DVGAR ERVZ	EVQEV DVGAR ERVZ
3515	KPAWC WYTLA MCGAG YDSGT CDYMY SHCFG IKHHS SGSSS YHS	LAMCG AGYDS GTCDY MYSHC FGIKH HSSGS SSYHS
3517	ACAAA AAGLL KG	ACAAA AAGLL KG
3518	ACAAW AAGLL KG	ACAAW AAGLL KG
3521	DWGGQ HHGLR EVLAA ALFAS CLWGZ	AALFA SCLWG Z
3536	KEKVI PLVTS FIEAL FMTVD KGSFG Z	LFMTV DKGSF GZ
3591	DWGGQ HHGLR EVLAA ALFAS CLWGZ	AALFA SCLWG Z
3596	KPAWC WYTLA MCGAG YDSGT CDYMY SHCFG IKHHS SGSSS YHS	LAMCG AGYDS GTCDY MYSHC FGIKH HSSGS SSYHS

Continued on next page

Continued from previous page

Serial Number	Sequence	Aggregating Motif
3598	DCAAH AAEAV WCGK	HAAEA VWCGK
3601	HAEGT FTSDV SSYLE GQAAK EFIAW LVKGR	AAKEF IAWLV KGR
3604	CILKP QMPWE LWDIL QEISP EEIQP NPPSS GMLGI IIMMT L	PQMPW ELWDI LQEIS PEEIQ PNPPS SGMLG II-IMM TL
3605	GGGGG SGGSG GSLL	GGSGG SGGSL L
3623	HRPVM AIVEV EVQEV DVGAR ERVZ	VEVQE VDVGA RERVZ
3634	HRPVM AIVEV EVQEV DVGAR ERVZ	EVEVQ EVDVG ARERV Z
3639	HRPVM AIVEV EVQEV DVGAR ERVZ	EVQEV DVGAR ERVZ
3642	SSGNQ NLYAM YQLSH FQSIS VLZ	YAMYQ LSHFQ SISVL Z
3643	HRPVM AIVEV EVQEV DVGAR ERVZ	VEVQE VDVGA RERVZ
3644	GGGGG SGGSG GSLL	GGSGG SGGSL L
3655	FNMQQ QRRFY EALHD PNLNE EQRNA KIKSI RD	FNMQQ QRRFY EALHD PNLNE EQRNA KIKSI RD
3658	GGGGG SGGSG GSLL	GSGGS LL
3661	CILKP QMPWE LWDIL QEISP EEIQP NPPSS GMLGI IIMMT L	IIMM TL

Continued on next page

Continued from previous page

Serial Number	Sequence	Aggregating Motif
3662	CILKP QMPWE LWDIL QEISP EEIQP NPPSS GMLGI IIMMT L	IIIMM TL
3671	NFQQD VGTKT TIRLM NSQLV TTEKR FLK	VGTKT TIRLM NSQLV TTEKR FLK
3674	AVVSS AGSLK SSQLG REIDD HDAVL RFNG	VVSSA GSLKS SQLGR EIDDH DAVLR FNG
3682	HRPVM AIVEV EVQEV DVGAR ERVZ	HRPVM AIVEV EVQEV DVGAR ERVZ
3687	HRPVM AIVEV EVQEV DVGAR ERVZ	VEVQE VDVGA RERVZ
3688	SSGNQ NLYAM YQLSH FQSIS VLZ	YAMYQ LSHFQ SISVL Z
3689	DWGGQ HHGLR EVLAA ALFAS CLWGZ	AAALF ASCLW GZ
3692	KPAWC WYTLA MCGAG YDSGT CDYMY SHCFG IKHHS SGSSS YHS	AMCGA GYDSG TCDYM YSHCF GIKHH SSGSS SYHS

```
alignments = { sn: { sn: 0 for sn in aggregating_motifs } \
       for sn in aggregating_motifs }

from Bio import pairwise2
import numpy
from scipy.cluster.hierarchy import linkage,dendrogram
from matplotlib import pyplot as plt
```

```
for sn,sequence in aggregating_motifs.items():
    for sn2,sequence2 in aggregating_motifs.items():
align = pairwise2.align.localms(sequence,sequence2,2,-1,-1,-0.5)
#print(sequence,sequence2,align)
if not align:
    alignments[sn][sn2] = 0
else:
    alignments[sn][sn2] = align[0][4]

#Turn the alignment {sn: {sn2:score...} ...}into a numpy arary
np_align = numpy.matrix([[value2 for key2,value2 in value.items()] \
 for key,value in alignments.items()])

x_labels = [value for key,value in aggregating_motifs.items()]

plt.figure()
linkages = linkage(np_align)
dn = dendrogram(linkages,labels=x_labels)
plt.show()
```

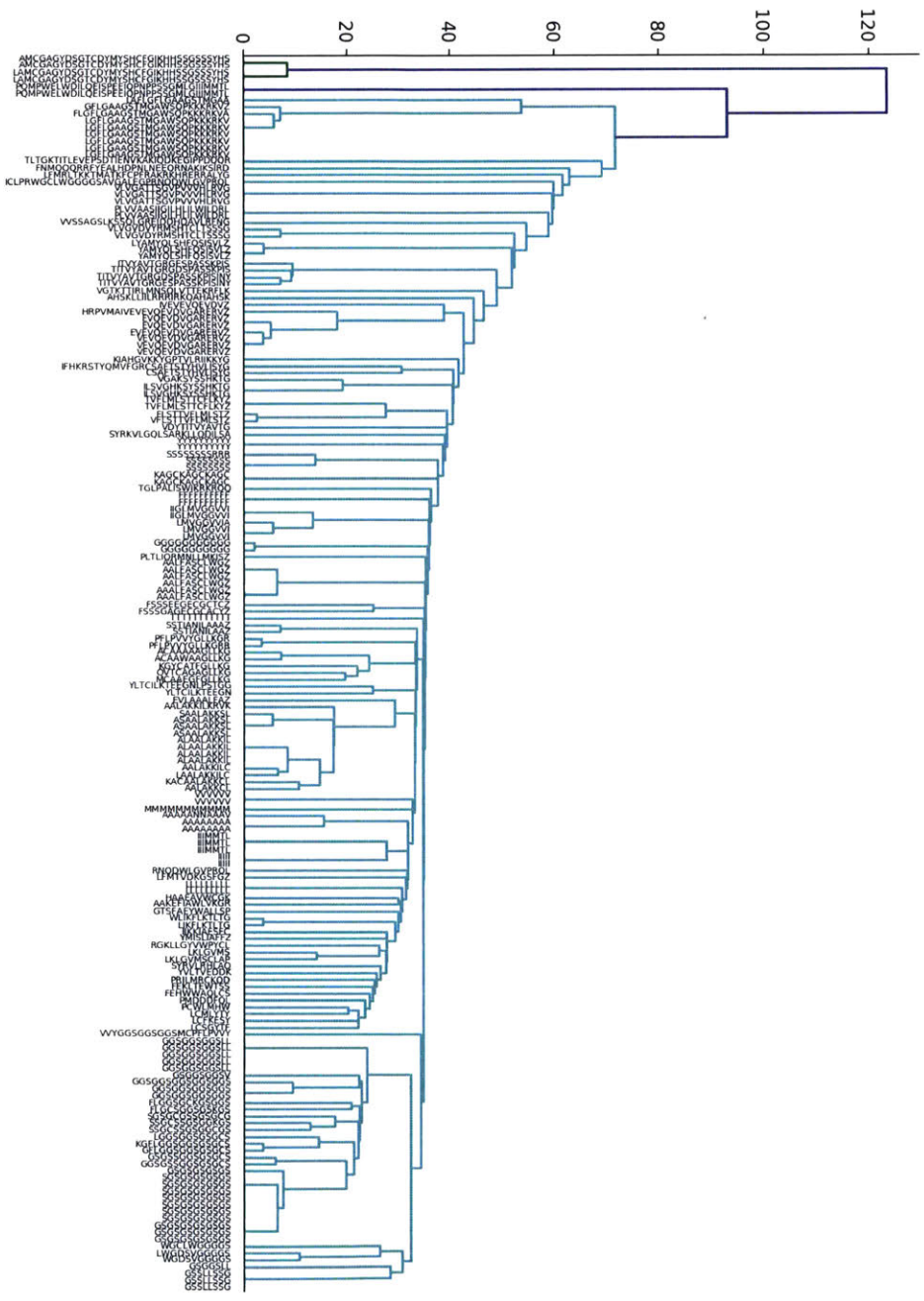

Figure B-1: Dendrogram of aligned aggregating sequences