INVESTIGATION OF SYNTHETIC PROTEINS PRODUCED VIA AUTOMATED FAST-FLOW PEPTIDE SYNTHESIS

by

Amanda Elizabeth Cowfer

Bachelor of Science in Molecular and Cellular Biology and BS.LAS Chemistry
University of Illinois Urbana-Champaign, 2018

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

February 2024

Signature of Author……
Department of Chemistry
November 22, 2023

Certified by……
Bradley L. Pentelute
Professor of Chemistry
Thesis Supervisor

Accepted by……...
Adam Willard
Associate Professor
Graduate Officer

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

Laura L. Kiessling, PhD.....
Thesis Committee Chair
Novartis Professor of Chemistry

Bradley L. Pentelute, PhD.....
Thesis Supervisor
Professor of Chemistry

Alex K Shalek, PhD.....
Thesis Committee Member
J. W. Kieckhefer Professor of Chemistry

INVESTIGATION OF SYNTHETIC PROTEINS PRODUCED VIA AUTOMATED FAST-FLOW PEPTIDE SYNTHESIS

by

Amanda Elizabeth Cowfer

Submitted to the Department of Chemistry on November 28, 2023
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy in Chemistry

Abstract

Flow chemistry techniques and methods have given the broad scientific community high-fidelity access to chemical compounds with minimal effort compared to traditional synthetic techniques. Since the introduction of solid phase peptide synthesis (SPPS), the peptide community has endeavored to combine the convenience of flow chemistry with the iterative steps associated with peptide elongation in SPPS. Nearly one decade ago, members of the Pentelute lab envisioned and developed a flow-based peptide synthesizer, the Automated Fast-Flow Peptide Synthesizer, or AFPS for short. This technology enabled fast, reliable access to short peptide chains, with each coupling taking less than 3 minutes in total, significantly decreasing the labor needed to produce these peptides. However, peptide chains over 50 amino acids remained challenging to produce via AFPS, microwave synthesis, or traditional SPPS batch couplings. With modern research requiring rapid and high-fidelity access to long polypeptide chains, an immediate need to develop peptide synthesis technology to produce single-domain protein polypeptides in a single shot will be critical.

Herein, I report on the arduous journey and unmatched teamwork needed to improve the AFPS systems for regular, reliable access to polypeptide chains of more than 200 amino acids in a single working day. In addition, I will highlight the workflow and knowledge needed to take a free polypeptide chain to a fully folded and biologically active protein, equivalent in form and function to its recombinant counterparts. I will discuss the iterative steps my team took to vary both chemical and mechanical and control variables to improve per-coupling yield enough to enable access to full-length single-domain proteins. On this journey, we utilized test peptides to validate synthesis quality and later synthesized a suite of full-length single-domain biologically active proteins. I will spend some time focusing on the barnase-barstar binding pair.

Next, I will dive into how I build and design each AFPS synthesizer to improve synthesis outcomes and user-friendliness while retaining the core functionality and customizability that have made the AFPS so successful in the Pentelute lab. I will highlight my role in the renovation of the first generation AFPS system, the "Automatide," and dive into the key characteristics that set our synthesizers apart from what is currently commercially available.

Finally, we report on the synthesis and characterization of several small and very interesting luciferases. Luciferases are proteins that produce bioluminescence when exposed to specific chemical substrates, and for the organisms that produce these enzymes, they play a vital role in mating, defense, and camouflage. In the research arena, luciferases have had broad applications for decades, including detection of environmental contaminants, diagnosis of pathogens, high-throughput screening for drug discovery, understanding protein-protein interactions, and more. Current efforts in the field have focused on the development of small artificial luciferases due to their many advantages over traditional larger luciferases, such as enhanced stability and increased brightness. Herein, we report on the synthesis and characterization of the copepod, Gaussia priceps, luciferase GLuc (18 kDa), and artificial luciferases picALuc (12 kDa) and LuxSit-I (14 kDa). In addition, we synthesized the mirror-image counterpart of picALuc due to its potential for broad-reaching impact in health and diagnostics; this is the first reported mirror-image bioluminescent luciferase. Finally, we will report on our efforts to develop a split-picAluc protein complement assay (PCA) using AS-MS technology, which will be the smallest and most versatile split-luciferase reported to date.

In summary, fast-flow peptide synthesis was utilized to produce and investigate several biologically relevant proteins to improve upon existing tools available to the broad chemistry and biology community.

Thesis Supervisor: Bradley L. Pentelute
Professor of Chemistry

Acknowledgments

Describing how the last half-decade of my life at MIT has affected me is extremely difficult, but I will do my best to put my endless thanks and emotions into words. I have had the unmatched opportunity to grow in ways I could not have predicted and have overcome challenges I never thought were surmountable. My success heavily relied on the unbroken support and love I have received during my time at MIT. I walked in more than a little nervous and apprehensive to start a new life thousands of miles from home, but the last five years have taught me so many new skills inside and outside the laboratory. I have discovered just as much about peptide science as I did about myself while completing my PhD.

First, a most sincere and heartfelt thank you to Prof. Bradly Pentelute for welcoming me into his group and allowing me to expand my scientific and mechanical knowledge to yet unknown heights. Thank you for believing in me when it was difficult for me to believe in myself. You helped me learn how to take a scientific project from a thought to a full-blown subgroup within the lab.

Thank you to my committee chair, Prof. Laura Kiessling, for your boundless support and encouragement. You taught me how to rise when I fell, and you never wavered in your confidence in my future success; my graduate experience truly would not have been the same without your mentorship and guidance. To Prof. Alex Shalek, thank you for being an essential and valued part of my Ph.D. journey as a member of my thesis committee. You always had a unique perspective to share with me about how I could continue to shape my research and improve as a scientist.

I could not have started my graduate journey without the steadfast mentorship of Dr. Nina Hartrampf, who taught me peptide chemistry from the ground up since I had never made a peptide in my life before coming to MIT. You always instilled a balance of hard work alongside meaningful and intentional time off. A huge thank you to the 2018 protein subgroup for giving me the deep dive tour of the Pentelute lab AFPS systems. Before my time in the lab, I was not a mechanical engineer, but I certainly feel like one now. Thank you to Dr. Joesph Brown for always offering helpful advice and getting me back on my feet after the COVID pandemic knocked me down. Without your assistance, I would have never made it to this moment and would have left MIT long ago.

To the members of my cohort, thank you for all the support and laughs through the years. Sarah Antilla, you were the absolute best partner in crime I didn't know I needed in my first few years in the lab. Too often, we were still the only two folks in the lab at 3 am our first year; perhaps not the best decision in hindsight, but a fond memory indeed. To Lord Charles (Charlie) Farquhar, we may have butted heads more than once in our early days in the lab, but you have become a true friend since that time. There is no one else I would rather gossip with or go snake shopping with. May our gay chaos reign supreme. To Alex Callahan-Callaman, thank you for listening to all my random info dumps while we were both working at the synthesizers and for all the helpful, in-depth science conversations. To Jacob Rodriquez and Xiyun Ye, I was honored to work alongside you these years, and I hope we can keep in touch as we all build our own lives after MIT.

To all the other members of the Pentelute lab, past, present, and future, thank you for building a group in which we can all take pride. We know how to work and then play hard here at MIT, and my experience is all the better for it. I will value the friendships and experiences I had for the rest of my life.

To the wonderful staff of OGE and the MSRP cohorts of 2022 and 2023, thank you for letting me end my graduate experience with a bang. Mentoring in the MSRP program has been the highlight of my summer for the past two years, and I feel I learned as much as I taught. Pod 4 and Pod 8 forever! To the wonderful folks of CADI, my time serving on the board of CADI and as a representative for QoL gave me an outlet for my activist mindset. I was honored to co-author several departmental initiatives that will far outlive my time at MIT.

To my beautiful prince, knight in shining armor, and fiancé Berenice Elizabeth Estrada Chavez, thank you for all your love, support, compassion, and kindness. Even when times were tough, you stood by me, and we thought our relationship might not survive. We both always managed to grow, learn, and improve ourselves together. I never want to stop becoming a better person with you, and I look forward to marrying you and expanding our family. I will always be on your side, just as you have been on mine.

I don't know how to properly thank my parents, Patricia and Micheal Fitzgerald, for always being there for me, but how you raised and supported me has made my success possible. I am ever grateful that you never pushed me to impress you or others but just encouraged me to pursue what made me happy. Without your support, I would never have grown into the person I am today. To my little brother, Bobby, we may have a massive age gap but know that I will always be there for

you, and I am so proud of the adult you are growing into. To my grandparents, aunts, and extended family, thank you for always welcoming me with open arms. Whether or not we were blood-related, I will always have fond memories of goofing off with the cousins, gardening with my grandma, and traveling the West with my aunt.

Finally, I give my endless thanks to my ever-loyal pupper, Stella. She may not understand every word I say and is certainly not the brightest, but she always seemed to know exactly what I need from her. Despite the scary move from Illinois to Massachusetts, I had her by my side. When I realized I was permanently disabled, she rose to the occasion and learned tasks to support me every day. Thank you to her and everyone who put up with her as my constant shadow. May my increased salary result in even better snacks for you, Stella. #labdogeforlife

Thank you all! I can never truly express how much everyone in my life has impacted how I have grown as a person.

Table of Contents

List of Figures

List of Tables

Scope of Thesis

Understanding how we can produce and utilize protein polypeptides is vital to further developing critically needed therapeutics. In this thesis, I focus on overcoming the challenges scientists face in polypeptide synthesis using the Pentelute lab's AFPS technology. I will explore applying the flow-based AFPS system to supplement traditional protein expression methods and expand the protein research space.

In **Chapter 1,** I will provide critical historical context regarding protein polypeptide production and explain how single-shot synthesis of protein polypeptides is an excellent boon to the research community. **Chapter 2** dives into the finer workings of our AFPS machines and how I rebuilt and improved our first-generation peptide synthesizer, the "Automatide." This machine can now not only produce long peptide sequences but also is a versatile testing ground for new mechanical, chemical, and software applications.

Chapter 3 describes the massive undertaking by me and a large team of talented scientists to optimize and streamline our AFPS technology to produce 200+ amino acid polypeptide chains in a matter of hours without the need for ligation techniques. This chapter dives into some of the mechanical and chemical variables investigated to improve our AFPS protocol. I will later discuss the characterization of the synthetically produced enzyme-inhibitor pair barnase-barstar.

Finally, in **Chapter 4,** I present my capstone work on the synthesis and characterization of several bioluminescent luciferase enzymes. I focus on the production of small, highly stable artificial luciferases developed to be the brightest and most versatile luciferase enzymes on the market. I will plunge into the fascinating applications of luciferases and highlight my production of the first mirror-image enzymatically active luciferase ever produced. Finally, I outline our journey toward developing the smallest split luciferase for applications in a protein complementation assay.

Chapter 1. Background and motivation

1.1 Introduction

The biochemistry of life is vast and complex,[1] with four primary classes of biomacromolecules, lipids, oligonucleotides, carbohydrates, and proteins, working in synergy to create life as we know it.[2] The focus of my thesis work will primarily be on large polypeptides, otherwise known as proteins. Proteins play an essential role in the cell and participate in cellular signaling,[3] construction,[4] and maintenance[5], and thus have been a hot area of research for pharmaceutical development for decades. Protein polypeptides are produced in the cell by the ribosome using a suite of RNA-based oligonucleotides from the 20 vital amino acids. In protein production, tRNA is used to deliver amino acids, and mRNA contains the coded sequence of the protein to be produced and is read by the ribosome.[6] After production by the ribosome, these polypeptides undergo a complex folding[7] and modification[8] process in the cell to then perform their native functions. The heart of each protein's individual role and function is the specific sequence of amino acids that comprises that polypeptide, and while codon redundancy[9] has allowed proteins to tolerate some sequence mutations, if key amino acid residues in the active site of the protein change, this can lead to a protein becoming non-functional or performing a different function altogether.[10] Researchers have long investigated and capitalized on the ability to mutate protein sequences to instill a particular function, one such method known as directed evolution,[11] with which researchers mimic the process of natural section to develop a protein with a desired trait, recently resulted in the awarding of the 2018 Nobel Prize in Chemistry to Dr. Frances Arnold.[12]

Protein structure and function are not just limited to their amino acid sequence, but they can also undergo further biochemical modification in the cell to add post-translational modifications (PTMs), which can include phosphorylation,[13] acetylation,[14] and glycosylation,[15] among others.[16] These PTMs can play critical roles in cells by turning proteins on or off,[17] labeling them for destruction via ubiquitination,[18] or signaling that the protein is to be transported elsewhere.[19] Thus, PTMs to proteins are essential in signal transduction pathways[20] and immune responses,[21] among others. Researchers have long been interested in protein polypeptides and their functions within the cell, but the limited scope of biological expression of proteins and peptides has led many researchers to investigate peptide production outside of the cell.[22] In the human genome alone, 25,000+ individual genes encode for an estimated 12,000+ proteins, many of which have yet to be fully understood or even documented.[23] Thus, we can see what appears to be an

insurmountable challenge standing before researchers today. Thankfully, research into peptides and proteins within the research communities has resulted in a plethora of knowledge and clinical applications of peptides and proteins, including hormonal therapy,[24] peptide vaccines,[25] and more.[26] Continued research in the field is vital to expand the research community's knowledge and reach further.[27] Herein, we provide an overview of some of the investigations into the proteome using solid-phase peptide synthesis, as well as our contributions to the field.

1.2 History of peptide synthesis

In the early 20th century, a pioneering researcher by the name of Emil Fisher coined the name "peptide",[28] and was the first to describe a peptide bond. Today, countless peptide chemists consider him the founding father of the entire field of peptide chemistry.[29] Much of the initial work on peptides involved complex organic synthetic conditions since few side chain protecting options existed[30], and Fisher's synthesis of short peptides from purified amino acids was considered groundbreaking and led to the Fisher-Hofmeister theory of protein structure,[31] and the now booming field of peptide and protein chemistry. Later, one of Fisher's students, Max Bergmann, introduced the first reversible protecting group for peptide synthesis, carbobenzoxy (Cbz),[32] expanding the possibilities of early solution-based peptide synthesis. Today, solution-based synthesis, while uncommon in research laboratories, is applied to synthesize peptides at a large-scale manufacturing level or for other specialized applications.[33] However, solution-based synthesis is very limited and, in practice, quite tricky and labor intensive, so other peptide synthesis modalities were desired.[34]

1.2.1 Advent of solid phase peptide synthesis (SPPS)

In the 1960s, a more tractable method of peptide synthesis was introduced by Bruce Merrifield, termed solid-phase peptide synthesis (SPPS),[35] this technique was revolutionary and replaced the complex solution phase synthesis with a solid resin scaffold. The quintessential feature of SPPS is the covalent bond between a solid support, the resin, and the growing peptide chain.[36] The process of peptide synthesis via SPPS involves cyclical coupling, wash, deprotection, and washing steps until the desired peptide chain length is achieved. The peptide tethered to the resin can then be backbone deprotected and cleaved from the solid support resin to yield the free polypeptide.[37] Unlike the ribosome, SPPS relies on taking advantage of carboxylic acid's electrophilic nature to elongate the peptide C→N whereas in the cell, the ribosome elongated

N→C; it is imperative to note, however, that peptide sequence nomenclature always writes peptides N→C.[38]

Early SPPS relied upon using t-butyloxycarbonyl (Boc) protecting group chemistry to

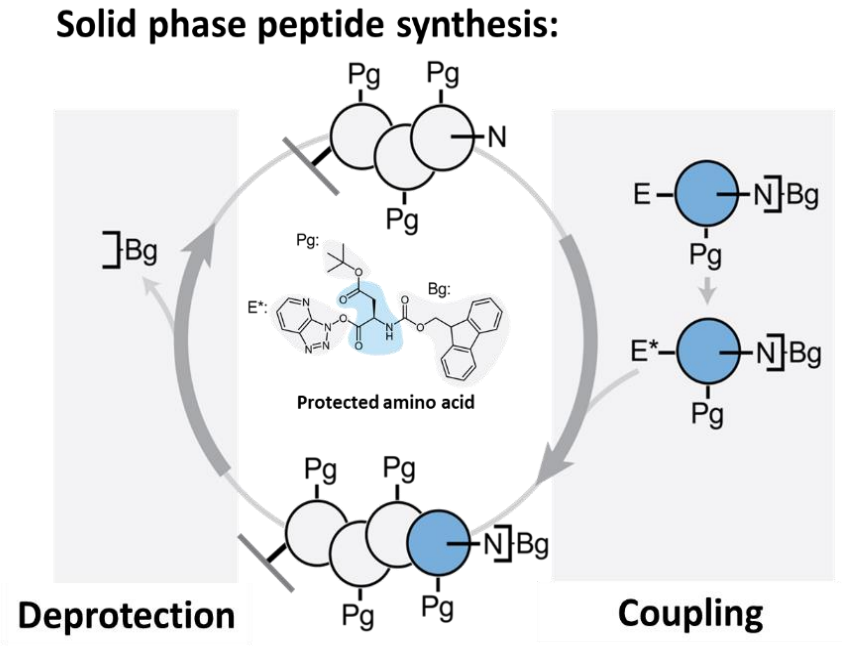


***Figure 1.1*:** Schematic showing the key steps to solid phase peptide synthesis.

A protected amino acid is first activated and then coupled to a growing peptide chain on resin. Then during deprotection the Fmoc or Boc blocking group is removed to free the terminal amine for the next cycle of synthesis. Key: Pg = amino acid backbone protecting group, N = Free amine, Bg = Fmoc or Boc blocking group, E = carboxylic acid, electrophile, E* = activated electrophile.

facilitate the iterative coupling and deprotection necessary for peptide elongation. Boc-based SPPS was utilized and improved upon for decades after its introduction,[39] but the use of the dangerous hydrofluoric acid (HF) during peptide cleavage left a desire to develop a protecting group that needed milder conditions for removal.[40] Thus, in 1970, the 9-fluorenyl methoxycarbonyl (Fmoc) protecting group was introduced[41] and proved to be beneficial for the synthesis of some peptides,[42] for instance, gramicidin, where yields improved from 5% to up to 87%.[43] By the later 1990s, nearly all peptide synthesis was being performed using Fmoc-based chemistry due to its many advantages and because of other improvements in the field, such as the introduction of different amino acid side-chain protecting groups and the optimization of cleavage protocols to minimize any size chain reactions during cleavage from resin.[44]

1.2.2 Advantages of SPPS

A massive advantage of SPPS is that, unlike the ribosome, SPPS coupling conditions allow for the coupling of completely non-natural amino acids,[45] the incorporation of specific and selective PTMs,[46] and other chemical additions to peptides (e.g., Fluorophores).[47] To that end, researchers can reliably produce peptides with PTMs by using a combination of modified amino acid building blocks containing the desired modification in the amino acids side group or by using chemistry post-peptide elongation to add the desired modification. For what may be obvious reasons, it is preferred to have the modification directly incorporated into the amino acid monomer to avoid any off-target chemical effects resulting from performing modification after elongation.[48] The main limiting factor in producing peptides with PTMs is the cost of the amino acid monomer building blocks, as many of them are difficult to obtain and expensive to make.

As a result of SPPS, researchers gained access to entirely new synthetic peptide landscapes, namely the production of peptidomimetics, which are synthetic peptides containing one or more unnatural amino acids.[49] It is well within the synthetic capabilities of SPPS to incorporate such modifications so that researchers can produce peptides containing esters in the peptide backbone, called depsipeptides,[50] inserting another carbon in the amino acid monomer, called a beta amino acid,[51] monomers containing nucleic acid moieties like peptide nucleic acids,[52] and more.[53] researchers can better develop and investigate peptide-based drugs, probes, and other durable and stable peptides for biological research with these new synthetic pathways.

Today, the use of SPPS has broadly resulted in improvements in yield and purity per coupling by performing couplings with reagents in significant excess to drive the reaction to completion.[54] Additionally, SPPS simplifies the general synthetic protocol by allowing peptide chemists to wash away unreacted material quickly in solution phase synthesis instead of using more laborious techniques, condensation, or phase separation.[55] The nature of the tethered growing peptide chain in SPPS also helps in making certain peptide chain modifications easier by limiting intramolecular reactions.[56] Over time, different solid supports or “resins” have been introduced, and each has a specific chemical makeup meant to help facilitate peptide elongation and improve synthesis outcomes.[57] For instance, for long polypeptide chains, using a resin with a lower substitution level or fewer anchor points to start a peptide chain can help avoid interchain interactions in a long growing polypeptide.[58]

1.2.3 Challenges of SPPS

However, despite SPPS addressing many synthetic challenges associated with peptide synthesis, numerous challenges remain. Among these challenges, one of the most limiting was synthesizing long complex peptides and small proteins more than 50 amino acids in length, which was nearly impossible due to extremely diminished yields per coupling.[59] Some factors contributing to diminished yields include the generation of synthetic by-products during peptide elongation, such as deletions, truncations, and peptide chain aggregation.[60] The accumulation of these by-products makes separating the desired peptide sequence, if present, from the many incorrect sequences challenging, even with advanced purification techniques.[61]

Furthermore, even though SPPS significantly improved the speed at which researchers could synthesize peptides, making sequences remained heavily time-consuming and involved researchers having to manually perform each step, with the average coupling cycle taking 20 minutes to over 2 hours.[62] However, even early in the introduction of SPPS, Merrifield, and coworkers acknowledged that SPPS was well positioned to allow for the complete automation of peptide synthesis, following further optimization of some chemical variables. To that end, one of the first ever reported automation of SPPS came from Merrifield, who used a simple two-valve and one-pump set-up to completely automate each amino acid coupling and deprotection cycle, with a single amino acid coupling taking approximately 4 hours, with no hands-on time.[63] Today, there are a plethora of SPPS instruments available for purchase, with microwave irradiation synthesis being the most common method by which these machines operate.[64]

1.2.4 Automated peptide synthesis

More recently, attention has been turning towards using flow-based chemical synthesis methods to access polypeptides quickly, with high yields and in-line monitoring.[65] In the past, flow-based chemistry SPPS approaches have primarily focused on automation and reduced reagent consumption, but in more recent times, the use of heated vessels and loops to facilitate the acceleration of Fmoc-based SPPS.[66] In 2014, the Pentelute lab introduced a manual “fast-flow” SPPS set-up that utilized a heated pre-loop, reactor containing resin, and a Varian HPLC pump to facilitate accelerated amino acid coupling, deprotection and washing steps with the entire cycle being complete in 3 minutes.[67] This manual flow set-up greatly decreased the time needed to perform amino acid couplings but still required a great deal of hands-on time for researchers operating the machine. This setup does use a significantly larger amount of raw materials than

traditional batch couplings, but the pay-off of the speed at which a researcher can make a full peptide may be worth the added cost.[68] Still, the idea of an automated peptide synthesizer remained popular with other research groups reporting similar set-ups; for example, the Gordon group built a similar system to the Pentelute's Manual Flow synthesizer and even spent some time assessing the optimal temperature, resin matrices, flow rates, and more to develop the machine to reduce material waste.[69]

To further develop the manual flow peptide synthesizer set-up, the Pentelute lab set out to build an entirely automated peptide synthesizer, and after a great deal of troubleshooting and development, the first-generation Automated Fast-Flow Peptide Synthesizer (AFPS), dupped the

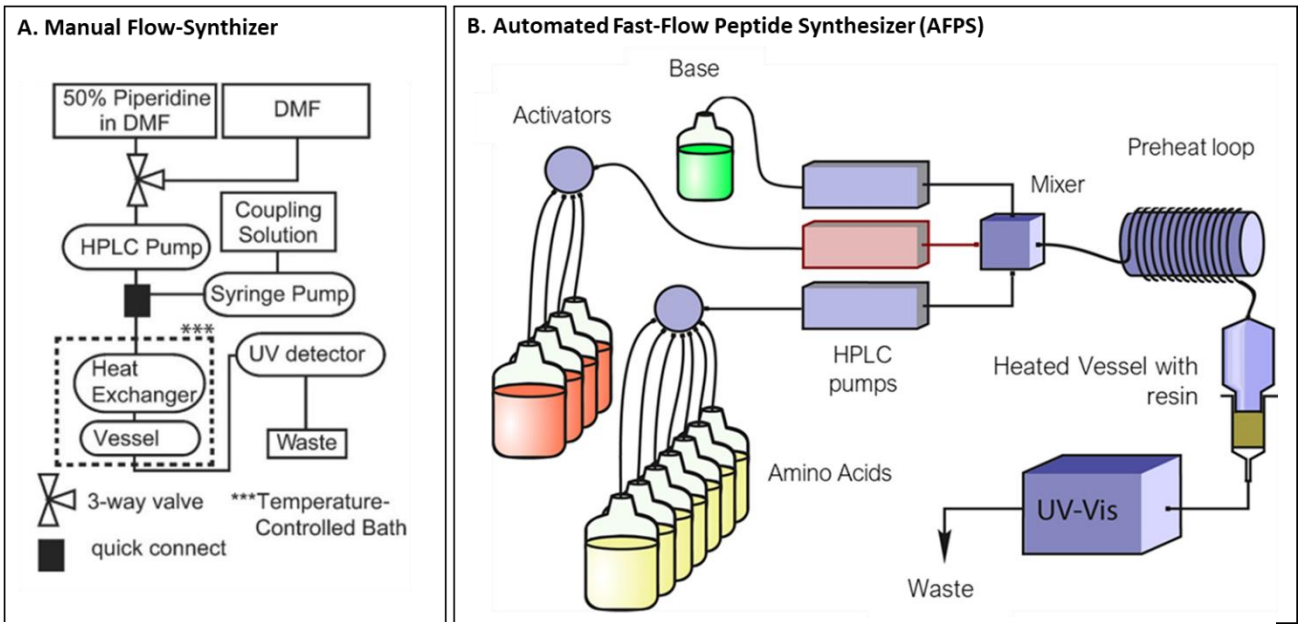


***Figure 1.2*:** Automated Flow Synthesizer examples

A. Schematic showing the set-up for the manual flow synthesizer reported by the Pentelute group in 2014, figure adapted from Simon et al. **B.** Schematic of the general AFPS system set-up used for the research performed in this thesis. Reagents are pulled from solution reservoirs by HPLC pumps and flowed through the mixer, heating loops, heated resin-filled reactor, and finally through the UV-vis module to waste. The UV-vis module allows for real-time monitoring of each amino acid coupling

"Automatide", was unveiled to the world.[70] This machine used three Varian ProStar HPLC pumps to iteratively flow reagents through heated loops into a mixing chamber, then over a resin bed encased in a heated reactor vessel. The entire machine was controlled via Laboratory Virtual Instrument Engineering Workbench (LabVIEW) software, which is a system-design platform and development environment for a visual programming language developed and distributed by National Instruments. LabVIEW is excellent at creating a user-friendly and easy-to-operate interface for data acquisition, instrument control, and other operations.[71] Further, connection to an

in-line UV-vis detector allows for real-time monitoring of both the coupling and deprotection steps.[72] Most unique about this set-up when compared to other flow-based synthesizers is that it utilized three pre-heating loops to tailor the appropriate coupling temperature to the amino acid-activator pair and the incorporation of a small disposable inexpensive reaction vessel to hold the resin and growing polypeptide chain. These steps immensely helped simplify the handling of all reagents to facilitate rapid and easy removal of reagents.[73] With this model of the AFPS, Pentelute, and co-workers were able to synthesize a 30-mer peptide in only 30 minutes, in comparable yields to commercially available peptides, vastly decreasing lead times for obtaining the desired peptide.[70] Since the introduction of the AFPS, our group has performed extensive improvements, optimizations, and complete instrument re-builds to the "Automatide", and we have also built several other generations and variants of the AFPS system. We will report on some of these efforts in *Chapters 2 and 3*.

With the introduction of the combination of the desirable traits of SPPS and flow, chemistry brings us the AFPS system. We can begin to look towards the future and aim to deliver better new peptide therapeutics. We need increased access to highly sequence-controlled chemical synthesis of long polypeptide chains and full protein chains to do this.[60] As discussed above, the synthesis of peptide chains more than 50 amino acids in length is complex due to both the time needed to perform each coupling cycle and poor yields from each cycle.[59] However, with the AFPS, we can envision a world where protein production via SPPS is a regular occurrence due to the essential need for protein-length polypeptides for the broad research community. In further sections, we will elaborate upon how the Pentelute Lab developed and optimized our AFPS technology to reliably access polypeptide chains in excess of 200 amino acids with a single-shot synthesis

1.3 Synthesis of protein polypeptides, a history

Proteins play a vital role in our body's day-to-day function and are, unfortunately, often a key part of when biological processes go wrong.[10] Some examples of disease development where proteins are directly implicated in disease progression include protein misfolding disorders, for example, Alzheimer's disease[74] and prion diseases like Mad Cow disease,[75] and genetically linked protein-linked diseases including cystic fibrosis[76] and Huntington's disease.[77] Furthermore, protein signaling cascades are critical to disease progression regarding cancer development,[78] autoimmune disorders, like rheumatoid arthritis, [79] metabolic disorders, like phenylketonuria,[80] and many more diseases and disorders than I can name in this document.[74] Thus, research into the

proteome at all levels is booming. In particular, there is a significant focus on using proteins for drug discovery and understanding disease progression.[81] With this knowledge, we can see that there is a tangible need for accessing long polypeptides and full-length proteins. This section will elaborate on how the scientific community has historically assessed these sequences.

1.3.1 Biological expression of proteins

Traditional means of accessing long peptides and full-length proteins that are longer than 50 amino acids is to use biological expression techniques.[10] Biological protein expression systems include bacterial cell expression, insect cell expression, and mammalian cell expression.[82] Each method has its advantages and disadvantages. Bacterial expression, for example, is not well suited for expressing multi-domain eukaryotic proteins since bacteria lack the necessary cell machinery to accomplish post-translational modifications (PTMs) or aid in the molecular folding of complex proteins.[83] Alternatively, mammalian cell expression is best suited for producing proteins containing the most native structure, PTMs, and functional activity; however, managing mammalian cell lines can be more time-consuming and generally difficult due to the demanding culture conditions and often yields from mammalian expression is low.[84] More recently, some researchers have been using insect cell expression to produce larger quantities of protein more quickly and with similar PTMs present as to that of mammalian cells. Insect cells are still quite demanding to culture, but the increase in gram-per-liter protein yields and generally more robust cell lines are an attractive alternative to mammalian expression.[85] Another arm of biological-based protein expression techniques includes cell-free protein expression, which uses a cocktail of all the macromolecules and cellular machinery needed for the synthesis of proteins. This method is particularly useful for incorporating some non-natural and modified amino acids and for the expression of proteins that are prone to degradation in the cell by intracellular proteases.[86] However, these systems can be quite costly to use and produce proteins in gram quantities. Only a small handful of unnatural amino acids are compatible with this method.[87] To this end, it is desired to have other paths to access proteins of interest.

Unfortunately, despite their many strengths, biological expression methods are still quite limited in scope, as discussed above. Generally, researchers are limited to only the naturally occurring amino acids, and even with advances in genetic code expansion, biological expression methods show difficulty in incorporating multiple unnatural residues in long, complex polypeptides.[88] Furthermore, many proteins are difficult to obtain from biological means due to

the protein being unstable or toxic.[89] Proteins range in size from several 1000 amino acid residues to less than 100 amino acids; additionally there are small biologically active and relevant peptides that contain only a few amino acid residues. Interestingly, nearly 30% of reported protein sequences are less than 500 amino acids in length,[90] making them interesting targets to access via chemical synthesis. To this end, a fusion of the flexibility of SPPS and the reach of biological expression is needed both to allow access to long, complex polypeptides and to incorporate multiple unnatural amino acids into these long sequences. SPPS is exceptionally well suited for this since peptide elongation is not limited by cellular machinery recognizing specific amino acids and instead can be performed with a near-unlimited number of non-conical amino acids. Furthermore, SPPS allows for specific control over proteins PTMs and allows for easy labeling of the polypeptide chain via incorporating tags and chemical handles, like biotin, during the synthesis process.

1.3.2 Ligation techniques to produce protein polypeptides

Early work on chemical protein synthesis relied heavily on the ligation of short peptide fragments of less than 50 amino acids in length due to the aforementioned weaknesses of SPPS. These fragments were produced in parallel and ligated using techniques such as native chemical ligation (NCL) or α-ketoacid-hydroxylamine ligation (KAHA) strategies.[91] NCL combines smaller peptide building blocks by reacting the C-terminal peptide thioester with an N-terminal cysteinyl peptide to produce a native peptide bond,[i] and this gives access to significantly longer polypeptide chains and synthetic proteins.[92] KAHA is an alternative ligation technique that, unlike NCL, does not rely on using cystines as participating members of the ligation process. Instead, KAHA utilizes C-C-terminal peptide α-ketoacids and N-terminal peptide hydroxylamines, which react chemoselectively in acidic aqueous conditions to form amides or esters.[93] However, NCL and KAHA development is a rather tedious and time-consuming process as developing the peptide sequences for ligation is complex, and optimizing the reaction itself is challenging. Nevertheless, ligation techniques have afforded access to many exciting protein targets.[94]

An early success in producing biologically active proteins using ligation techniques included the total synthesis of human lysozyme, a 130-mer protein with four disulfide bonds, was reported in 2007 by Durek and coworkers.[95] They accomplished the convergent synthesis of lysozyme by synthesizing four peptide fragments, each ~30 amino acids in length, and using NCL

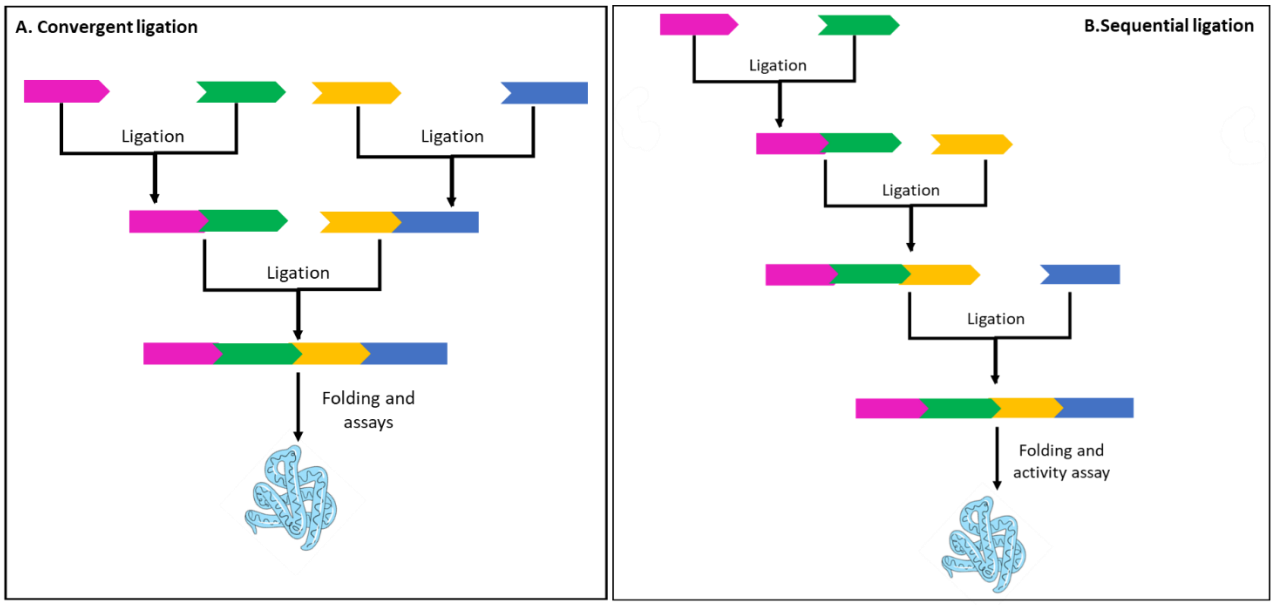


***Figure 1.3*:** Scheme for chemical ligation strategies

NCL and KAHA both can be used to produce full-length protein polypeptides, the figure above demonstrates the two general synthetic strategies of convergent ligation **(A),** and sequential ligation **(B)** used to assemble peptide fragments into the final desired polypeptide.

to facilitate linkage between all four fragments to yield full-length lysozyme. Following synthesis, this synthetically produced lysozyme demonstrated full native catalytic activity and allowed researchers to use X-ray crystallography to report a high-resolution lysozyme structure. Since the advent of NCL, many other full-length proteins have been reported, including human immunodeficiency virus-1 protease (HIV-1), a 99-mer protein prepared by ligating two smaller 51-mer and 48-mer peptide fragments.[96] The glycoprotein erythropoietin (EPO) was also reported and provides an exciting example of how SPPS and ligation techniques can be used to understand PTMs in a controlled manner, unlike the unpredictability of cellular expression.[97]

Outside of successful NCL usage, KAHA saw no small number of victories in producing full-length synthetic proteins; one example is the recent report of interferon-induced transmembrane protein 3 (IFITM3), a difficult-to-produce transmembrane protein that prevents pathogenic viruses from replicating and furthering infection.[98] Harmand and coworkers produced the 133-mer IFITM3 and chemically modified variants using KAHA facilitated by 5-oxproline

and homoserine. Three peptide fragments were synthesized and then ligated with KAHA, using a previously reported alanine scan to select the ideal ligation sites that would not inhibit the native function of IFITM3. This work highlights how KAHA can be an ideal synthesis route to produce highly hydrophobic proteins due to its tolerance to organic co-solvents. Other successes with KAHA include the production of ubiquitin-fold modifier 1 (UFM1) from three peptides sequentially ligated together,[99] and the large 184-mer ferric heme-binding protein nitrophorin 4 from five unprotected peptide fragments.[100] KAHA has presented as an excellent alternative to NCL for protein production.

Thus, despite the challenges presented when using ligation techniques to synthesize protein polypeptides, the SPPS community has cleverly overcome them to produce dozens of highly relevant and interesting proteins for further research efforts. Production of proteins via SPSP will open therapeutic doors and facilitate the development of better treatments and understanding of diseases plaguing humanity.

1.4 Summary and outlook

In this section, I have set up the historical context and apparent need for protein polypeptides produced via SPPS. Proteins play a critical role in cellular functions and are key players in disease progression, but also are key to understanding how to treat these disorders better. SPPS has been successfully used to synthesize biologically active protein polypeptides, and even in the early days of the Pentelute group, it was clear to members that we needed ready access to full-length biologically active proteins to provide to the broad research community. Past members have produced several synthetic proteins utilizing NCL techniques in the past, including native wild-type barnase and its mirror image D-barnase.[101]

With the addition of the AFPS to the Pentelute lab toolkit, I and others began to focus on optimizing our AFPS protocol to produce full-length proteins over 200 amino acids, which will be elaborated on further in *Chapter 3*. With ready access to the single-shot synthesis of proteins, I believe the peptide and greater scientific community are poised to accelerate proteome research and drug discovery.

1.5 References

1. Pace, N. R. The Universal Nature of Biochemistry. *Proc. Natl. Acad. Sci.* **2001**, *98* (3), 805–808. https://doi.org/10.1073/pnas.98.3.805.

2. Li, J.; Mach, P.; Koehl, P. MEASURING THE SHAPES OF MACROMOLECULES – AND WHY IT MATTERS. *Comput. Struct. Biotechnol. J.* **2013**, *8* (12), e201309001. https://doi.org/10.5936/csbj.201309001.

3. Nair, A.; Chauhan, P.; Saha, B.; Kubatzky, K. F. Conceptual Evolution of Cell Signaling. *Int. J. Mol. Sci.* **2019**, *20* (13), 3292. https://doi.org/10.3390/ijms20133292.

4. Wen, P.; Wang, X.; Moreno, S.; Boye, S.; Voigt, D.; Voit, B.; Huang, X.; Appelhans, D. Construction of Eukaryotic Cell Biomimetics: Hierarchical Polymersomes-in-Proteinosome Multicompartment with Enzymatic Reactions Modulated Protein Transportation. *Small* **2021**, *17* (7), 2005749. https://doi.org/10.1002/smll.202005749.

5. Gallagher, P. S.; Oeser, M. L.; Abraham, A.; Kaganovich, D.; Gardner, R. G. Cellular Maintenance of Nuclear Protein Homeostasis. *Cell. Mol. Life Sci.* **2014**, *71* (10), 1865–1879. https://doi.org/10.1007/s00018-013-1530-y.

6. Lafontaine, D. L. J.; Tollervey, D. The Function and Synthesis of Ribosomes. *Nat. Rev. Mol. Cell Biol.* **2001**, *2* (7), 514–520. https://doi.org/10.1038/35080045.

7. Zhang, H.; Gong, W.; Wu, S.; Perrett, S. Studying Protein Folding in Health and Disease Using Biophysical Approaches. *Emerg. Top. Life Sci.* **2021**, *5* (1), 29–38. https://doi.org/10.1042/ETLS20200317.

8. Silva, A. M. N.; Vitorino, R.; Domingues, M. R. M.; Spickett, C. M.; Domingues, P. Post-Translational Modifications and Mass Spectrometry Detection. *Free Radic. Biol. Med.* **2013**, *65*, 925–941. https://doi.org/10.1016/j.freeradbiomed.2013.08.184.

9. Kille, S.; Acevedo-Rocha, C. G.; Parra, L. P.; Zhang, Z.-G.; Opperman, D. J.; Reetz, M. T.; Acevedo, J. P. Reducing Codon Redundancy and Screening Effort of Combinatorial Protein Libraries Created by Saturation Mutagenesis. *ACS Synth. Biol.* **2013**, *2* (2), 83–92. https://doi.org/10.1021/sb300037w.

10. Sharp, N. Mutations Matter Even If Proteins Stay the Same. *Nature* **2022**, *606* (7915), 657–659. https://doi.org/10.1038/d41586-022-01091-6.

11. Arnold, F. H. Directed Evolution: Bringing New Chemistry to Life. *Angew. Chem. Int. Ed.* **2018**, *57* (16), 4143–4148. https://doi.org/10.1002/anie.201708408.

12. Arnold, F. H. Design by Directed Evolution. *Acc. Chem. Res.* **1998**, *31* (3), 125–131. https://doi.org/10.1021/ar960017f.

13. Ubersax, J. A.; Ferrell Jr, J. E. Mechanisms of Specificity in Protein Phosphorylation. *Nat. Rev. Mol. Cell Biol.* **2007**, *8* (7), 530–541. https://doi.org/10.1038/nrm2203.

14. Eberharter, A.; Becker, P. B. Histone Acetylation: A Switch between Repressive and Permissive Chromatin: Second in Review Series on Chromatin Dynamics. *EMBO Rep.* **2002**, *3* (3), 224–229. https://doi.org/10.1093/embo-reports/kvf053.

15. Reily, C.; Stewart, T. J.; Renfrow, M. B.; Novak, J. Glycosylation in Health and Disease. *Nat. Rev. Nephrol.* **2019**, *15* (6), 346–366. https://doi.org/10.1038/s41581-019-0129-4.

16. Doll, S.; Burlingame, A. L. Mass Spectrometry-Based Detection and Assignment of Protein Posttranslational Modifications. *ACS Chem. Biol.* **2015**, *10* (1), 63–71. https://doi.org/10.1021/cb500904b.

17. Smith, P. R.; Loerch, S.; Kunder, N.; Stanowick, A. D.; Lou, T.-F.; Campbell, Z. T. Functionally Distinct Roles for EEF2K in the Control of Ribosome Availability and P-Body Abundance. *Nat. Commun.* **2021**, *12* (1), 6789. https://doi.org/10.1038/s41467-021-27160-4.

18. Chanarat, S.; Mishra, S. K. Emerging Roles of Ubiquitin-like Proteins in Pre-MRNA Splicing. *Trends Biochem. Sci.* **2018**, *43* (11), 896–907. https://doi.org/10.1016/j.tibs.2018.09.001.

19. Rapoport, T. A. Protein Transport across the Endoplasmic Reticulum Membrane: Delivered on 8 July 2007 at the 32nd FEBS Congress in Vienna, Austria. *FEBS J.* **2008**, *275* (18), 4471–4478. https://doi.org/10.1111/j.1742-4658.2008.06588.x.

20. Van De Stolpe, A.; Holtzer, L.; Van Ooijen, H.; Inda, M. A. D.; Verhaegh, W. Enabling Precision Medicine by Unravelling Disease Pathophysiology: Quantifying Signal Transduction Pathway Activity across Cell and Tissue Types. *Sci. Rep.* **2019**, *9* (1), 1603. https://doi.org/10.1038/s41598-018-38179-x.

21. Gough, C.; Sadanandom, A. Understanding and Exploiting Post-Translational Modifications for Plant Disease Resistance. *Biomolecules* **2021**, *11* (8), 1122. https://doi.org/10.3390/biom11081122.

22. Spirin, V.; Mirny, L. A. Protein Complexes and Functional Modules in Molecular Networks. *Proc. Natl. Acad. Sci.* **2003**, *100* (21), 12123–12128. https://doi.org/10.1073/pnas.2032324100.

23. Nurk, S.; Koren, S.; Rhie, A.; Rautiainen, M.; Bzikadze, A. V.; Mikheenko, A.; Vollger, M. R.; Altemose, N.; Uralsky, L.; Gershman, A.; Aganezov, S.; Hoyt, S. J.; Diekhans, M.; Logsdon, G. A.; Alonge, M.; Antonarakis, S. E.; Borchers, M.; Bouffard, G. G.; Brooks, S. Y.; Caldas, G. V.; Chen, N.-C.; Cheng, H.; Chin, C.-S.; Chow, W.; De Lima, L. G.; Dishuck, P. C.; Durbin, R.; Dvorkina, T.; Fiddes, I. T.; Formenti, G.; Fulton, R. S.; Fungtammasan, A.; Garrison, E.; Grady, P. G. S.; Graves-Lindsay, T. A.; Hall, I. M.; Hansen, N. F.; Hartley, G. A.; Haukness, M.; Howe, K.; Hunkapiller, M. W.; Jain, C.; Jain, M.; Jarvis, E. D.; Kerpedjiev, P.; Kirsche, M.; Kolmogorov, M.; Korlach, J.; Kremitzki, M.; Li, H.; Maduro, V. V.; Marschall, T.; McCartney, A. M.; McDaniel, J.; Miller, D. E.; Mullikin, J. C.; Myers, E. W.; Olson, N. D.; Paten, B.; Peluso, P.; Pevzner, P. A.; Porubsky, D.; Potapova, T.; Rogaev, E. I.; Rosenfeld, J. A.; Salzberg, S. L.; Schneider, V. A.; Sedlazeck, F. J.; Shafin, K.; Shew, C. J.; Shumate, A.; Sims, Y.; Smit, A. F. A.; Soto, D. C.; Sović, I.; Storer, J. M.; Streets, A.; Sullivan, B. A.; Thibaud-Nissen, F.; Torrance, J.; Wagner, J.; Walenz, B. P.; Wenger, A.; Wood, J. M. D.; Xiao, C.; Yan, S. M.; Young, A. C.; Zarate, S.; Surti, U.; McCoy, R. C.; Dennis, M. Y.; Alexandrov, I. A.; Gerton, J. L.; O'Neill, R. J.; Timp, W.; Zook, J. M.; Schatz, M. C.; Eichler, E. E.; Miga, K. H.; Phillippy, A. M. The Complete

Sequence of a Human Genome. *Science* **2022**, *376* (6588), 44–53. https://doi.org/10.1126/science.abj6987.

24. Ridker, P. M.; Hennekens, C. H.; Rifai, N.; Buring, J. E.; Manson, J. E. Hormone Replacement Therapy and Increased Plasma Concentration of C-Reactive Protein. *Circulation* **1999**, *100* (7), 713–716. https://doi.org/10.1161/01.CIR.100.7.713.

25. Malonis, R. J.; Lai, J. R.; Vergnolle, O. Peptide-Based Vaccines: Current Progress and Future Challenges. *Chem. Rev.* **2020**, *120* (6), 3210–3229. https://doi.org/10.1021/acs.chemrev.9b00472.

26. Wang, L.; Wang, N.; Zhang, W.; Cheng, X.; Yan, Z.; Shao, G.; Wang, X.; Wang, R.; Fu, C. Therapeutic Peptides: Current Applications and Future Directions. *Signal Transduct. Target. Ther.* **2022**, *7* (1), 48. https://doi.org/10.1038/s41392-022-00904-4.

27. Barman, P.; Joshi, S.; Sharma, S.; Preet, S.; Sharma, S.; Saini, A. Strategic Approaches to Improvise Peptide Drugs as Next Generation Therapeutics. *Int. J. Pept. Res. Ther.* **2023**, *29* (4), 61. https://doi.org/10.1007/s10989-023-10524-3.

28. Fischer, E.; Fourneau, E. Ueber Einige Derivate Des Glykocolls. *Berichte Dtsch. Chem. Ges.* **1901**, *34* (2), 2868–2877. https://doi.org/10.1002/cber.190103402249.

29. Darmstaedter, L. Emil Fischer. *J. Chem. Educ.* **1928**, *5* (1), 36. https://doi.org/10.1021/ed005p36.

30. Vigneaud, V. D.; Ressler, C.; Swan, C. J. M.; Roberts, C. W.; Katsoyannis, P. G.; Gordon, S. THE SYNTHESIS OF AN OCTAPEPTIDE AMIDE WITH THE HORMONAL ACTIVITY OF OXYTOCIN. *J. Am. Chem. Soc.* **1953**, *75* (19), 4879–4880. https://doi.org/10.1021/ja01115a553.

31. Collins, K. D.; Washabaugh, M. W. The Hofmeister Effect and the Behaviour of Water at Interfaces. *Q. Rev. Biophys.* **1985**, *18* (4), 323–422. https://doi.org/10.1017/S0033583500005369.

32. Bergmann, M.; Zervas, L. Über Ein Allgemeines Verfahren Der Peptid-Synthese. *Berichte Dtsch. Chem. Ges. B Ser.* **1932**, *65* (7), 1192–1201. https://doi.org/10.1002/cber.19320650722.

33. Tymecka, D.; Misicka, A. Solution Phase Peptide Synthesis: The Case of Biphalin. In *Peptide Synthesis*; Hussein, W. M., Skwarczynski, M., Toth, I., Eds.; Methods in Molecular Biology; Springer US: New York, NY, 2020; Vol. 2103, pp 1–11. https://doi.org/10.1007/978-1-0716-0227-0_1.

34. Bayer, E.; Mutter, M. Liquid Phase Synthesis of Peptides. *Nature* **1972**, *237* (5357), 512–513. https://doi.org/10.1038/237512a0.

35. Merrifield, R. B. **Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide**. *J. Am. Chem. Soc.* **1963**, *85* (14), 2149–2154. https://doi.org/10.1021/ja00897a025.

36. Coin, I.; Beyermann, M.; Bienert, M. Solid-Phase Peptide Synthesis: From Standard Procedures to the Synthesis of Difficult Sequences. *Nat. Protoc.* **2007**, *2* (12), 3247–3256. https://doi.org/10.1038/nprot.2007.454.

37. Collins, J. M.; Porter, K. A.; Singh, S. K.; Vanier, G. S. High-Efficiency Solid Phase Peptide Synthesis (*HE* -SPPS). *Org. Lett.* **2014**, *16* (3), 940–943. https://doi.org/10.1021/ol4036825.

38. Agouridas, V.; Diemer, V.; Melnyk, O. Strategies and Open Questions in Solid-Phase Protein Chemical Synthesis. *Curr. Opin. Chem. Biol.* **2020**, *58*, 1–9. https://doi.org/10.1016/j.cbpa.2020.02.007.

39. Merrifield, B. Solid Phase Synthesis. *Science* **1986**, *232* (4748), 341–347. https://doi.org/10.1126/science.3961484.

40. Merrifield, R. B. New Approaches to the Chemical Synthesis of Peptides. In *Proceedings of the 1966 Laurentian Hormone Conference*; Elsevier, 1967; pp 451–482. https://doi.org/10.1016/B978-1-4831-9826-2.50013-1.

41. Carpino, L. A.; Han, G. Y. 9-Fluorenylmethoxycarbonyl Function, a New Base-Sensitive Amino-Protecting Group. *J. Am. Chem. Soc.* **1970**, *92* (19), 5748–5749. https://doi.org/10.1021/ja00722a043.

42. Barany, G.; Merrifield, R. B. A Chromatographic Method for the Quantitative Analysis of the Deprotection of Dithiasuccinoyl (Dts) Amino Acids. *Anal. Biochem.* **1979**, *95* (1), 160–170. https://doi.org/10.1016/0003-2697(79)90199-4.

43. Fields, C. G.; Fields, G. B.; Noble, R. L.; Cross, T. A. Solid Phase Peptide Synthesis of 15N-Gramicidins A, B, and C and High Performance Liquid Chromatographic Purification. *Int. J. Pept. Protein Res.* **2009**, *33* (4), 298–303. https://doi.org/10.1111/j.1399-3011.1989.tb01285.x.

44. King, D. S.; Fields, C. G.; Fields, G. B. A Cleavage Method Which Minimizes Side Reactions Following Fmoc Solid Phase Peptide Synthesis. *Int. J. Pept. Protein Res.* **2009**, *36* (3), 255–266. https://doi.org/10.1111/j.1399-3011.1990.tb00976.x.

45. Kent, S. B. H. Novel Protein Science Enabled by Total Chemical Synthesis. *Protein Sci.* **2019**, *28* (2), 313–328. https://doi.org/10.1002/pro.3533.

46. Hackenberger, C. P. R.; O'Reill, M. K.; Imperiali, B. Improving Glycopeptide Synthesis: A Convenient Protocol for the Preparation of β-Glycosylamines and the Synthesis of Glycopeptides. *J. Org. Chem.* **2005**, *70* (9), 3574–3578. https://doi.org/10.1021/jo047801v.

47. Mendive-Tapia, L.; Wang, J.; Vendrell, M. Fluorescent Cyclic Peptides for Cell Imaging. *Pept. Sci.* **2021**, *113* (1). https://doi.org/10.1002/pep2.24181.

48. Lothrop, A. P.; Torres, M. P.; Fuchs, S. M. Deciphering Post-Translational Modification Codes. *FEBS Lett.* **2013**, *587* (8), 1247–1257. https://doi.org/10.1016/j.febslet.2013.01.047.

49. Li Petri, G.; Di Martino, S.; De Rosa, M. Peptidomimetics: An Overview of Recent Medicinal Chemistry Efforts toward the Discovery of Novel Small Molecule Inhibitors. *J. Med. Chem.* **2022**, *65* (11), 7438–7475. https://doi.org/10.1021/acs.jmedchem.2c00123.

50. Alonzo, D. A.; Schmeing, T. M. Biosynthesis of Depsipeptides, *or* Depsi: The Peptides with Varied Generations. *Protein Sci.* **2020**, *29* (12), 2316–2347. https://doi.org/10.1002/pro.3979.

51. Cabrele, C.; Martinek, T. A.; Reiser, O.; Berlicki, Ł. Peptides Containing β-Amino Acid Patterns: Challenges and Successes in Medicinal Chemistry. *J. Med. Chem.* **2014**, *57* (23), 9718–9739. https://doi.org/10.1021/jm5010896.

52. Pellestor, F.; Paulasova, P. The Peptide Nucleic Acids (PNAs), Powerful Tools for Molecular Genetics and Cytogenetics. *Eur. J. Hum. Genet.* **2004**, *12* (9), 694–700. https://doi.org/10.1038/sj.ejhg.5201226.

53. Wu, R. P.; Youngblood, D. S.; Hassinger, J. N.; Lovejoy, C. E.; Nelson, M. H.; Iversen, P. L.; Moulton, H. M. Cell-Penetrating Peptides as Transporters for Morpholino Oligomers: Effects of Amino Acid Composition on Intracellular Delivery and Cytotoxicity. *Nucleic Acids Res.* **2007**, *35* (15), 5182–5191. https://doi.org/10.1093/nar/gkm478.

54. Angeletti, R. H.; Bonewald, L. F.; Fields, G. B. [32] Six-Year Study of Peptide Synthesis. In *Methods in Enzymology*; Elsevier, 1997; Vol. 289, pp 697–717. https://doi.org/10.1016/S0076-6879(97)89071-4.

55. Albericio, F.; Lloyd-Williams, P.; Giralt, E. [15] Convergent Solid-Phase Peptide Synthesis. In *Methods in Enzymology*; Elsevier, 1997; Vol. 289, pp 313–336. https://doi.org/10.1016/S0076-6879(97)89054-4.

56. Ludwick, A. G.; Jelinski, L. W.; Live, David.; Kintanar, Agustin.; Dumais, J. J. Association of Peptide Chains during Merrifield Solid-Phase Peptide Synthesis. A Deuterium NMR Study. *J. Am. Chem. Soc.* **1986**, *108* (21), 6493–6496. https://doi.org/10.1021/ja00281a007.

57. Castro, V.; Rodriguez, H.; Albericio, F. Wang Linker Free of Side Reactions. *Org. Lett.* **2013**, *15* (2), 246–249. https://doi.org/10.1021/ol303367s.

58. Tam, J. P.; Lu, Y.-A. Coupling Difficulty Associated with Interchain Clustering and Phase Transition in Solid Phase Peptide Synthesis. *J. Am. Chem. Soc.* **1995**, *117* (49), 12058–12063. https://doi.org/10.1021/ja00154a004.

59. Riniker, B.; Flörsheimer, A.; Fretz, H.; Sieber, P.; Kamber, B. A General Strategy for the Synthesis of Large Peptides: R1~he Combined Solid-Phase and Solution Approach. *Tetrahedron* **1993**, *49* (41), 9307–9320. https://doi.org/10.1016/0040-4020(93)80017-N.

60. Zompra, A. A.; Galanis, A. S.; Werbitzky, O.; Albericio, F. Manufacturing Peptides as Active Pharmaceutical Ingredients. *Future Med. Chem.* **2009**, *1* (2), 361–377. https://doi.org/10.4155/fmc.09.23.

61. Frearson, J. A.; Collie, I. T. HTS and Hit Finding in Academia – from Chemical Genomics to Drug Discovery. *Drug Discov. Today* **2009**, *14* (23–24), 1150–1158. https://doi.org/10.1016/j.drudis.2009.09.004.

62. Miranda, L. P.; Alewood, P. F. Accelerated Chemical Synthesis of Peptides and Small Proteins. *Proc. Natl. Acad. Sci.* **1999**, *96* (4), 1181–1186. https://doi.org/10.1073/pnas.96.4.1181.

63. Merrifield, R. B. Automated Synthesis of Peptides: Solid-Phase Peptide Synthesis, a Simple and Rapid Synthetic Method, Has Now Been Automated. *Science* **1965**, *150* (3693), 178–185. https://doi.org/10.1126/science.150.3693.178.

64. Pedersen, S. L.; Tofteng, A. P.; Malik, L.; Jensen, K. J. Microwave Heating in Solid-Phase Peptide Synthesis. *Chem Soc Rev* **2012**, *41* (5), 1826–1844. https://doi.org/10.1039/C1CS15214A.

65. Mándity, I. M.; Olasz, B.; Ötvös, S. B.; Fülöp, F. Continuous-Flow Solid-Phase Peptide Synthesis: A Revolutionary Reduction of the Amino Acid Excess. *ChemSusChem* **2014**, *7* (11), 3172–3176. https://doi.org/10.1002/cssc.201402436.

66. Jolley, K. E.; Nye, W.; González Niño, C.; Kapur, N.; Rabion, A.; Rossen, K.; Blacker, A. J. Highly Productive Continuous Flow Synthesis of Di- and Tripeptides in Water. *Org. Process Res. Dev.* **2017**, *21* (10), 1557–1565. https://doi.org/10.1021/acs.oprd.7b00214.

67. Simon, M. D.; Heider, P. L.; Adamo, A.; Vinogradov, A. A.; Mong, S. K.; Li, X.; Berger, T.; Policarpo, R. L.; Zhang, C.; Zou, Y.; Liao, X.; Spokoyny, A. M.; Jensen, K. F.; Pentelute, B. L. Rapid Flow-Based Peptide Synthesis. *ChemBioChem* **2014**, *15* (5), 713–720. https://doi.org/10.1002/cbic.201300796.

68. Gordon, C. P. The Renascence of Continuous-Flow Peptide Synthesis – an Abridged Account of Solid and Solution-Based Approaches. *Org. Biomol. Chem.* **2018**, *16* (2), 180–196. https://doi.org/10.1039/C7OB02759A.

69. Spare, L. K.; Laude, V.; Harman, D. G.; Aldrich-Wright, J. R.; Gordon, C. P. An Optimised Approach for Continuous-Flow Solid-Phase Peptide Synthesis Utilising a Rudimentary Flow Reactor. *React. Chem. Eng.* **2018**, *3* (6), 875–882. https://doi.org/10.1039/C8RE00190A.

70. Mijalis, A. J.; Thomas, D. A.; Simon, M. D.; Adamo, A.; Beaumont, R.; Jensen, K. F.; Pentelute, B. L. A Fully Automated Flow-Based Approach for Accelerated Peptide Synthesis. *Nat. Chem. Biol.* **2017**, *13* (5), 464–466. https://doi.org/10.1038/nchembio.2318.

71. Elliott, C.; Vijayakumar, V.; Zink, W.; Hansen, R. National Instruments LabVIEW: A Programming Environment for Laboratory Automation and Measurement. *JALA J. Assoc. Lab. Autom.* **2007**, *12* (1), 17–24. https://doi.org/10.1016/j.jala.2006.07.012.

72. Henkel, B.; Bayer, E. Monitoring of Solid Phase Peptide Synthesis by FT-IR Spectroscopy. *J. Pept. Sci.* **1998**, *4* (8), 461–470. https://doi.org/10.1002/(SICI)1099-1387(199812)4:8<461::AID-PSC165>3.0.CO;2-C.

73. Pollard, D.; Kistler, C. Disposable Bioreactors. In *Current Developments in Biotechnology and Bioengineering*; Elsevier, 2017; pp 353–379. https://doi.org/10.1016/B978-0-444-63663-8.00012-4.

74. Ashraf, G.; Greig, N.; Khan, T.; Hassan, I.; Tabrez, S.; Shakil, S.; Sheikh, I.; Zaidi, S.; Akram, M.; Jabir, N.; Firoz, C.; Naeem, A.; Alhazza, I.; Damanhouri, G.; Kamal, M. Protein Misfolding and Aggregation in Alzheimer's Disease and Type 2 Diabetes Mellitus. *CNS Neurol. Disord. - Drug Targets* **2014**, *13* (7), 1280–1293. https://doi.org/10.2174/18715273136661409170955l4.

75. Costandi, M. Proteins behind Mad-Cow Disease Also Help Brain to Develop. *Nature* **2013**, nature.2013.12428. https://doi.org/10.1038/nature.2013.12428.

76. Liou, T. G. The Clinical Biology of Cystic Fibrosis Transmembrane Regulator Protein. *Chest* **2019**, *155* (3), 605–616. https://doi.org/10.1016/j.chest.2018.10.006.

77. Schulte, J.; Littleton, J. T. The Biological Function of the Huntingtin Protein and Its Relevance to Huntington's Disease Pathology. *Curr. Trends Neurol.* **2011**, *5*, 65–78.

78. Sever, R.; Brugge, J. S. Signal Transduction in Cancer. *Cold Spring Harb. Perspect. Med.* **2015**, *5* (4), a006098–a006098. https://doi.org/10.1101/cshperspect.a006098.

79. Kim, C. W.; Cho, E. H.; Lee, Y. J.; Kim, Y. H.; Hah, Y. S.; Kim, D. R. Disease-Specific Proteins from Rheumatoid Arthritis Patients. *J. Korean Med. Sci.* **2006**, *21* (3), 478–484. https://doi.org/10.3346/jkms.2006.21.3.478.

80. Sarodaya, N.; Suresh, B.; Kim, K.-S.; Ramakrishna, S. Protein Degradation and the Pathologic Basis of Phenylketonuria and Hereditary Tyrosinemia. *Int. J. Mol. Sci.* **2020**, *21* (14), 4996. https://doi.org/10.3390/ijms21144996.

81. Bull, S. C.; Doig, A. J. Properties of Protein Drug Target Classes. *PLOS ONE* **2015**, *10* (3), e0117955. https://doi.org/10.1371/journal.pone.0117955.

82. Tripathi, N. K.; Shrivastava, A. Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development. *Front. Bioeng. Biotechnol.* **2019**, *7*, 420. https://doi.org/10.3389/fbioe.2019.00420.

83. Lozano Terol, G.; Gallego-Jara, J.; Sola Martínez, R. A.; Martínez Vivancos, A.; Cánovas Díaz, M.; de Diego Puente, T. Impact of the Expression System on Recombinant Protein Production in Escherichia Coli BL21. *Front. Microbiol.* **2021**, *12*, 682001. https://doi.org/10.3389/fmicb.2021.682001.

84. Khan, K. H. Gene Expression in Mammalian Cells and Its Applications. *Adv. Pharm. Bull.* **2013**, *3* (2), 257–263. https://doi.org/10.5681/apb.2013.042.

85. Altmann, F.; Staudacher, E.; Wilson, I. B.; März, L. Insect Cells as Hosts for the Expression of Recombinant Glycoproteins. *Glycoconj. J.* **1999**, *16* (2), 109–123. https://doi.org/10.1023/a:1026488408951.

86. Khambhati, K.; Bhattacharjee, G.; Gohil, N.; Braddick, D.; Kulkarni, V.; Singh, V. Exploring the Potential of Cell-Free Protein Synthesis for Extending the Abilities of Biological Systems. *Front. Bioeng. Biotechnol.* **2019**, *7*, 248. https://doi.org/10.3389/fbioe.2019.00248.

87. Wu, Y.; Wang, Z.; Qiao, X.; Li, J.; Shu, X.; Qi, H. Emerging Methods for Efficient and Extensive Incorporation of Non-Canonical Amino Acids Using Cell-Free Systems. *Front. Bioeng. Biotechnol.* **2020**, *8*, 863. https://doi.org/10.3389/fbioe.2020.00863.

88. Wright, T. H.; Bower, B. J.; Chalker, J. M.; Bernardes, G. J. L.; Wiewiora, R.; Ng, W.-L.; Raj, R.; Faulkner, S.; Vallée, M. R. J.; Phanumartwiwath, A.; Coleman, O. D.; Thézénas, M.-L.; Khan, M.; Galan, S. R. G.; Lercher, L.; Schombs, M. W.; Gerstberger, S.; Palm-Espling, M. E.; Baldwin, A. J.; Kessler, B. M.; Claridge, T. D. W.; Mohammed, S.; Davis, B. G. Posttranslational Mutagenesis: A Chemical Strategy for Exploring Protein Side-Chain Diversity. *Science* **2016**, *354* (6312), aag1465. https://doi.org/10.1126/science.aag1465.

89. Donnelly, M. I.; Stevens, P. W.; Stols, L.; Xiaoyin Su, S.; Tollaksen, S.; Giometti, C.; Joachimiak, A. Expression of a Highly Toxic Protein, Bax, in Escherichia Coli by Attachment of a Leader Peptide Derived from the GroES Cochaperone. *Protein Expr. Purif.* **2001**, *22* (3), 422–429. https://doi.org/10.1006/prep.2001.1442.

90. Tiessen, A.; Pérez-Rodríguez, P.; Delaye-Arredondo, L. J. Mathematical Modeling and Comparison of Protein Size Distribution in Different Plant, Animal, Fungal and Microbial Species Reveals a Negative Correlation between Protein Size and Protein Number, Thus Providing Insight into the Evolution of Proteomes. *BMC Res. Notes* **2012**, *5* (1), 85. https://doi.org/10.1186/1756-0500-5-85.

91. Saebi, A.; Brown, J. S.; Marando, V. M.; Hartrampf, N.; Chumbler, N. M.; Hanna, S.; Poskus, M.; Loas, A.; Kiessling, L. L.; Hung, D. T.; Pentelute, B. L. Rapid Single-Shot Synthesis of the 214 Amino Acid-Long N-Terminal Domain of Pyocin S2. *ACS Chem. Biol.* **2023**, *18* (3), 518–527. https://doi.org/10.1021/acschembio.2c00862.

92. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. Synthesis of Proteins by Native Chemical Ligation. *Science* **1994**, *266* (5186), 776–779. https://doi.org/10.1126/science.7973629.

93. Bode, J. W. Chemical Protein Synthesis with the α-Ketoacid–Hydroxylamine Ligation. *Acc. Chem. Res.* **2017**, *50* (9), 2104–2115. https://doi.org/10.1021/acs.accounts.7b00277.

94. Agouridas, V.; El Mahdi, O.; Diemer, V.; Cargoët, M.; Monbaliu, J.-C. M.; Melnyk, O. Native Chemical Ligation and Extended Methods: Mechanisms, Catalysis, Scope, and Limitations. *Chem. Rev.* **2019**, *119* (12), 7328–7443. https://doi.org/10.1021/acs.chemrev.8b00712.

95. Durek, T.; Torbeev, V. Yu.; Kent, S. B. H. Convergent Chemical Synthesis and High-Resolution x-Ray Structure of Human Lysozyme. *Proc. Natl. Acad. Sci.* **2007**, *104* (12), 4846–4851. https://doi.org/10.1073/pnas.0610630104.

96. Torbeev, V. Yu.; Kent, S. B. H. Convergent Chemical Synthesis and Crystal Structure of a 203 Amino Acid "Covalent Dimer" HIV-1 Protease Enzyme Molecule. *Angew. Chem. Int. Ed.* **2007**, *46* (10), 1667–1670. https://doi.org/10.1002/anie.200604087.

97. Wang, P.; Dong, S.; Shieh, J.-H.; Peguero, E.; Hendrickson, R.; Moore, M. A. S.; Danishefsky, S. J. Erythropoietin Derived by Chemical Synthesis. *Science* **2013**, *342* (6164), 1357–1360. https://doi.org/10.1126/science.1245095.

98. Harmand, T. J.; Pattabiraman, V. R.; Bode, J. W. Chemical Synthesis of the Highly Hydrophobic Antiviral Membrane-Associated Protein IFITM3 and Modified Variants. *Angew. Chem. Int. Ed.* **2017**, *56* (41), 12639–12643. https://doi.org/10.1002/anie.201707554.

99. Ogunkoya, A. O.; Pattabiraman, V. R.; Bode, J. W. Sequential α-Ketoacid-Hydroxylamine (KAHA) Ligations: Synthesis of C-Terminal Variants of the Modifier Protein UFM1. *Angew. Chem. Int. Ed.* **2012**, *51* (38), 9693–9697. https://doi.org/10.1002/anie.201204144.

100. He, C.; Kulkarni, S. S.; Thuaud, F.; Bode, J. W. Chemical Synthesis of the 20 KDa Heme Protein Nitrophorin 4 by A-Ketoacid-Hydroxylamine (KAHA) Ligation. *Angew. Chem. Int. Ed.* **2015**, *54* (44), 12996–13001. https://doi.org/10.1002/anie.201505379.

101. Vinogradov, A. A.; Evans, E. D.; Pentelute, B. L. Total Synthesis and Biochemical Characterization of Mirror Image Barnase. *Chem. Sci.* **2015**, *6* (5), 2997–3002. https://doi.org/10.1039/C4SC03877K.

Chapter 2. Renovation of First-generation Automated Fast-flow Peptide Synthesizer

2.1 Introduction and Motivation

Using flow chemistry for SPPS is an ever-expanding method of peptide synthesis due to its ability to exert control over physical parameters[1], significantly reduce by-product formation,[2] and increase rate amino acid couplings.[3] In the late 1980s, studies found that using SPPS in flow allows for automated and high-fidelity polypeptide synthesis.[4] In the Pentelute group, peptides are commonly made using SPPS in-flow on our AFPS systems, as introduced in Chapter 1.2. Implementing flow chemistry to make peptides via SPPS has allowed rapid access to peptides shorter than 50 residues, but sequences beyond that length remain elusive.

The AFPS system employs flow chemistry using robust, high-performance liquid chromatography (HPLC) pumps to deliver reagents to a fixed resin-containing reactor continuously. The AFPS system reduces each standard fluorenyl methyloxycarbonyl (Fmoc) amino acid coupling to as little as 40 seconds at temperatures up to 90 ℃. Additionally, the UV-active nature of the Fmoc protecting group allows for real-time monitoring of individual amino acid couplings, and disposable reactors allow for rapid peptide synthetic turnover.[5] The synthesis process of the AFPS consists of iterative coupling, deprotection, and amine-free DMF wash steps occurring via reagent mixing and delivery by three HPLC pumps, which flow the reagents through heating loops, over the resin and then through a UV detector to a waste receptacle.[6]

The first generation AFPS, known as the "Automatide," was the original testing and optimization ground for the talented team of researchers who first reported the AFPS technology.[7] However, over time, the Automatide fell into an unusable state of disrepair and would require a massive overhaul to return to working order. Seeing this, I decided to update, reconfigure, and improve the Automatide to allow for continued use for years to come. In this process, I desired to re-introduce the Automatide as the primary machine for testing new hardware, software, and chemistry by creating a modular, easy-to-adjust, and repair AFPS machine. On this journey, the HPLC pump refill time and flow rate became clearly important factors in the synthetic success of peptides made via AFPS, particularly long polypeptide chains.

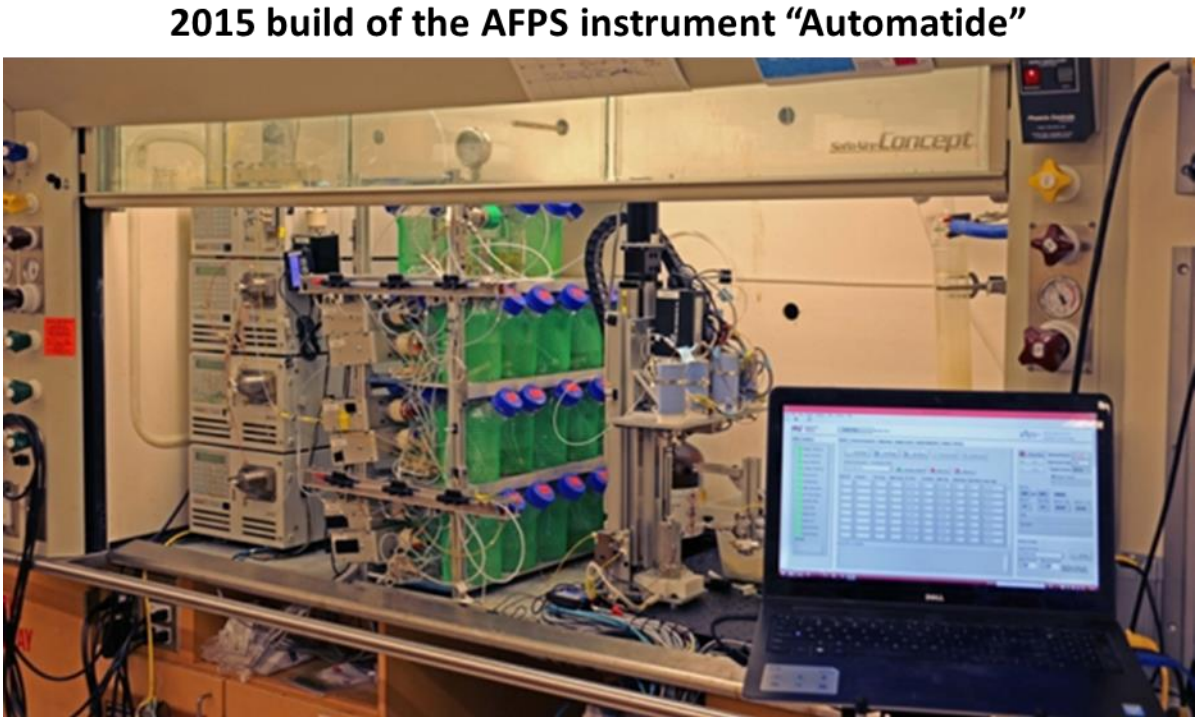


***Figure 2.1*:** Automatide design prior to rebuild

Original Automatide setup from 2017 publication Mijalis et al. with HPLC pumps stacked on the far left, reagent bottles (500 mL) and valves on the middle rack. The reactor, heating loop and the robot arm on the far right of the hood. Outside of the hood is the computer and UV-Vis module, not in frame, but stored below the computer.

In this chapter, I will take you, dear reader, through the process of deconstructing and reconstructing the Automatide and provide insight into why we made the changes and adjustments we did. Additionally, there will be sections highlighting other software and hardware changes that have been implemented for the past five years, including the newest AFPS instrument build dubbed the “Robosome” and a brief look into the complete remodel of the “Amidator” which was unfortunately lost due to extensive flood damage. We hope this section will shed light on how these machines are still being developed and improved upon as we learn more about what makes an easy-to-maintain yet high-performing peptide synthesizer.

2.2 Physical and Mechanical Changes to the Automatide

Nearly an entire decade ago, the first AFPS system, known as the “Automatide,” was built by a dedicated team of Pentelute lab members. Since its inception, this first-generation instrument has been used by countless researchers to make hundreds of peptides.[8] During the initial trials and tribulations associated with creating an entirely unique, never-before-seen, and high-performing peptide synthesizer, the hardware and software associated with the Automatide underwent many changes, large and small. However, over time, the machine began to have a series of mechanical

and software failures that became increasingly difficult to diagnose and fix. Eventually, it became clear that the whole set-up would have to be massively overhauled in order to return the Automatide to full working order. Since the pumps and other essential hardware components were individually in working order, two other young graduate students and I decided to deconstruct and then reconstruct the Automatide to allow it to continue to be useful to the lab for years to come. It was vital that we re-vamped the Automatide to make it easier to maintain and repair in the future. During the rebuild, we updated the hardware, software, and the physical layout of the Automatide to make it more user-friendly and to allow for shorter repair times in the future.[9]

To begin the reconstruction, it was essential that we took stock of the current state of the machine and determined what pieces could be re-used and what would need to be replaced or removed entirely. It was essential that we establish whether some hardware and software components needed to be removed since, during the original construction of the Automatide, the team went through many physical and software designs, and not all the resulting code and wiring had been removed from the machine with each iteration of tests. To aid us in this process, we consulted engineer Ryan Beaumont of the R.M. Beaumont Corporation to identify problem areas with the AFPS machinery and strategize a plan to address these issues. Ryan was brought on to help us with the delicate electrical and mechanical work that would have been unsafe for us to accomplish on our own without training. So, with Ryan's supervision and guidance, we undertook mechanical and electrical engineer training to prepare ourselves for the rebuild and to have future independence with instrument repair.

2.2.1 Electrical Changes to the Automatide

Following a thorough cleaning of the hood containing the Automatide, we were left with a clear understanding of the failures and successes of this first AFPS system. Regarding the instrument wiring, it was clear that we needed to both elevate the wiring inside of the hood. We discovered in our investigations that much of the original wiring had suffered extensive damage from solvent exposure due to the writing being run on the chemical fume hood's floor and drip tray, thus exposing it to large amounts of spilled corrosive solvent, figure 2.2.[10] It was obvious that this method would only lead to further repairs in the future and presented an unneeded electrical hazard.

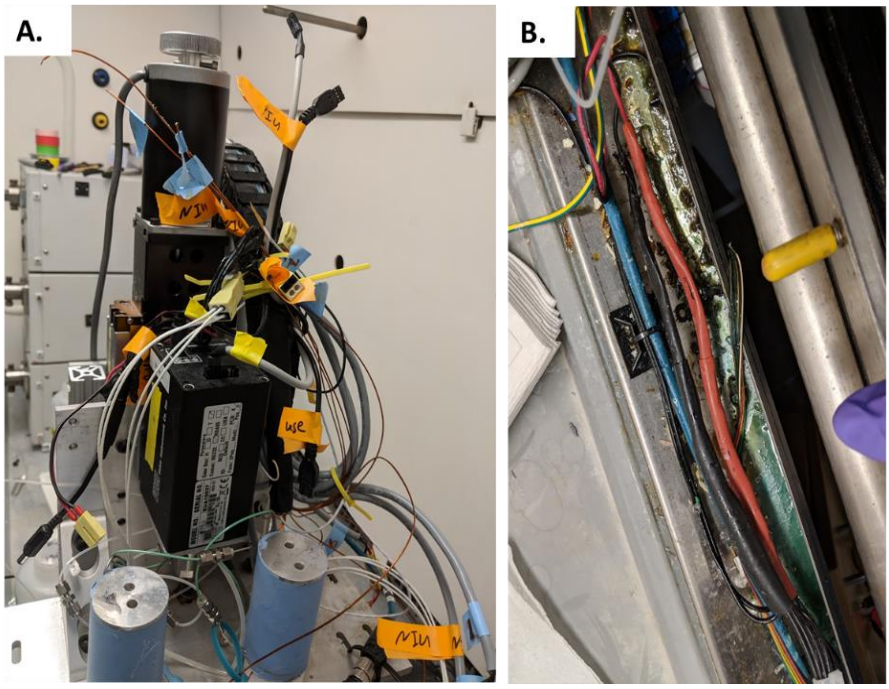


***Figure 2.2*:** Pictures showing extraneous wiring on the Automatide

A. Image of the original robotic arm set-up, all wires labeled in orange where not currently connected to anything, and due to the sheer amount of old wiring repairs were difficult and even hazardous to perform. It was clear that we needed to remove the old wiring entirely and re-wire all major connections. **B.** Image of machine wring in the chemical fume hood drip tray, all wires had melted together and/or had exposed copper.

We decided to replace all electrical connections and wiring with chemically resistant polytetrafluoroethylene (PTFE) wiring to address this.[11] PTFE is a highly chemically resistant and unreactive polymer; PTFE is not water or oil-soluble due to the high electronegativity of the

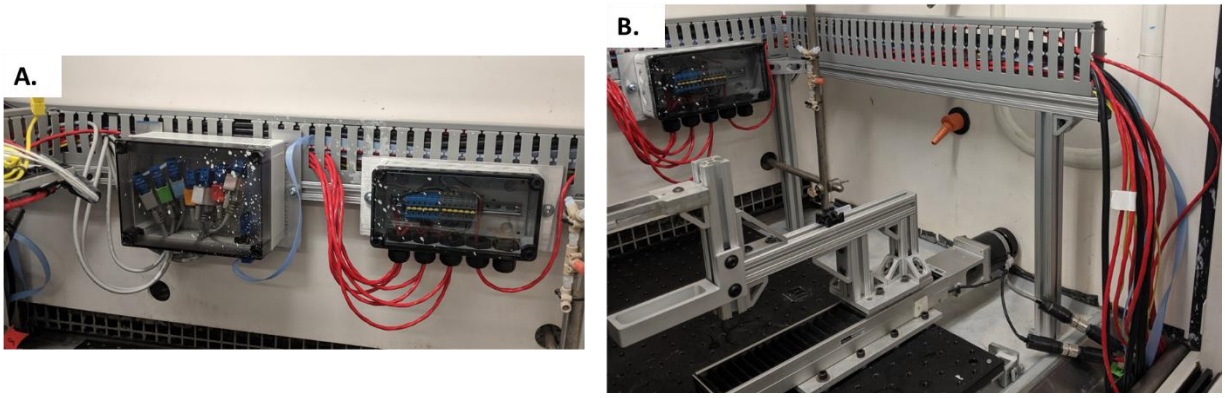


***Figure 2.3*:** Pictures showing improved wiring set-up

A. Image of wiring "boxes" containing critical connections for easy access and improved protection. Note the white dots on the plastic from DMF spills, excitingly there was no damage to the wires within the box. **B.** Image of elevated PTFE wires, protecting the hardware from solvent spills and other chemical or physical damage.

fluorine atoms, and the strong C-F bonds that make up the structure's core are extraordinarily stable and difficult to break.[12] Replacing traditional PVC wires with PTFE has prevented the degradation and damage we observed in the previous wring. As an additional protective measure, we built an

elevated wiring organizing system and included easily accessible covered boxes to contain critical connections to the pump and valve systems, figure 2.3. This change has made future additions and adjustments to these components much easier and more streamlined.

2.2.2 Physical machine layout changes to the Automatide

Next, we focused on redesigning the entire layout of the individual Automatide components in the hood. This included the pumps, UV-vis module, reagent rack, and robot arm orientation. For the reagent rack, we wanted to make the amino acid, activator, and other stock solutions more readily accessible and visible to the user. This was to help improve reagent tracking and improve safety when accessing reagents. To accomplish this goal, we rotated the amino acid storage rack 90 degrees, and all of the original 500 mL amino acid bottles were replaced with 250 mL bottles. At the same time, the activators and piperidine solution remained in 500 mL bottles. This new rack set-up mimics that of the Amidator (2016), Peptidator (2018), and Robosome(2019) and enables better and safer access to all of the necessary AFPS reagents. Furthermore, the initial Automatide set-up had no dedicated additional reagent reservoirs that could be used for non-conical amino acid incorporation. As such, we built an additional reagent rack that could be

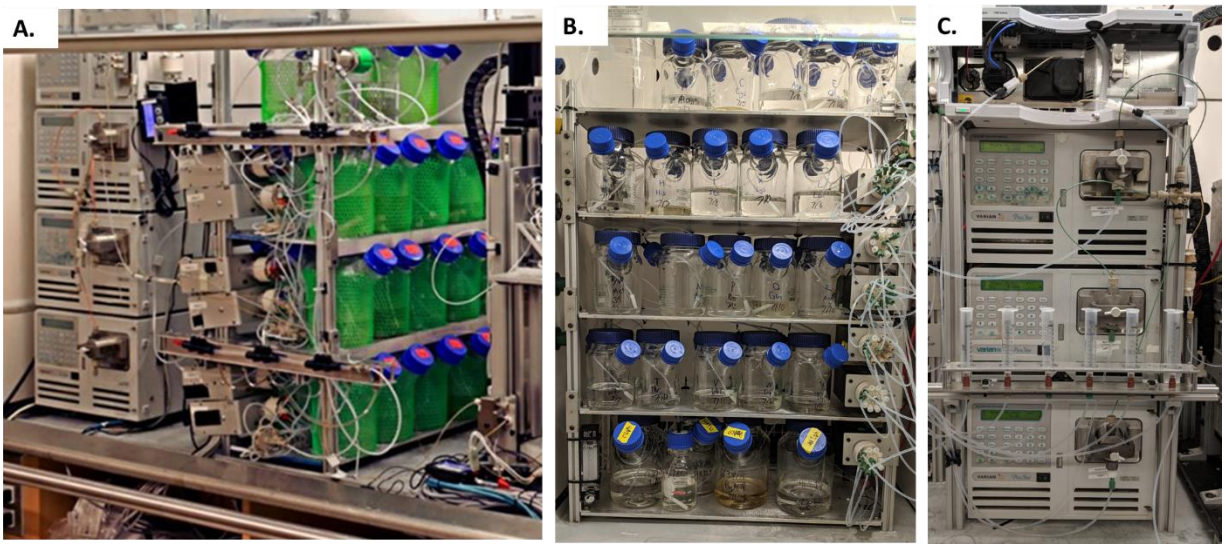


***Figure 2.4*:** Pictures of old vs new pump and reagent rack set-up.

A. Image of the old multi-pump design, with the reagent rack containing large 500 mL bottles that were difficult to access, observe, and re-fill. It is also worth noting that the reagents had to travel the width of the fume hood twice, a waste of space and a hazard to users. **B.** Image of new front facing reagent rack and valves, set up for easy access, observation, and repair. **C.** New UV-vis and pump stack, all components are rotated for easy access. Additionally, the syringe manifold allows for unnatural amino acids to be primed on the machine.

mounted on the HPLC pump stack. This rack provided nine positions for non-canonical amino acids or other chemical monomers. Furthermore, we built this rack to mount amino acid solutions

using lure-lock syringes. This minimizes reagent loss and provides a disposable reservoir to prevent excessive cleaning at these positions.

Finally, to protect the HPLC pumps further, which we refer to as the "heart of the machine," and the UV-vis module, we rotated and placed them in the center of the hood directly next to the heated reactor. As previously mentioned, the overhauled wiring connecting the pumps and valves to the computer was installed along the rear of the fume hood. By rotating the pumps to have writing in the back, we protected these , figure 2.4, vital components further. Since implementing these physical changes, we have noticed an improvement in instrument repair time and positive signs that our new elevated wires are protected from solvent spills and line breaks.

2.3 Software changes to the Automatide

The AFPS machines use Laboratory Virtual Instrument Engineering Workbench (LabVIEW) software, which is a system-design platform and development environment for a visual programming language developed and distributed by National Instruments. LabVIEW is excellent at creating a user-friendly and easy-to-operate interface for data acquisition, instrument control, and other operations.[13] LabView software is most often used for the following purposes: automated manufacturing tests and automated product design validation of systems or system components, control and monitoring of machines and pieces of equipment, and finally, monitoring the condition and status of industrial equipment.[14] LabView is unique in that all coding is done using an entirely visual interface, making LabView an ideal system to use for researchers without advanced coding knowledge.[15]

Along with physical design changes, we also saw a dire need to revamp the Automatide synthesis software in LabView. During the repair process, it was clear that while LabView was extraordinarily beneficial for the original AFPS design team, it had a massive amount of redundant or unused code. Furthermore, the improvements made to the newer generations of the AFPS had not been translated to the older versions. We saw this as an opportunity to generalize the LabView software to be the same across all generations of the AFPS while incorporating new features.

Our first vital step in performing these updates was to consult our contract engineer, Ryan Beaumont. Together, we made a detailed task list of code to be updated, removed, and added. We

Figure 2.5: General Labview VI for the AFPS

A. Depicts the LabVIEW UV-vis and pressure tracking software users see during synthesis. Further, users can see synthesis steps on the right-hand side along with the live comport connections. **B.** Shows the temperature tracking software which controls for accurate and consistent temperature readings throughout the synthesis. On the left side you can see the graphical output of temperature over the course of 5 minutes. On the left side you can see the active thermocouples with the set temperature and current temperature displayed. This panel is a helpful diagnostic tool and additionally contains a VI that will alert the instrument user to low or high temperatures. **C.** Exhibits the synthesis step software VI, in this window during an active synthesis run you can see what amino acid is to be coupled with what activator and for how long. Users can change the settings manually at this step if desired, and different synthesis .syn files can allow for other flow-based chemistries to be applied.

aimed to update software such that all old or irrelevant code was removed or rewritten so that it would not impede the future performance of the instrument.

2.3.1 Reagent tracking software

For the renovation, we focused on three main goals. The first goal was to improve the coding for our solvent monitoring software,[16] such that users would receive warnings for low

reagent levels at 20 mL left instead of at zero, figure 2.6. With the initial code, users would receive low solution warnings too late and run out of reagents. When this happens, air is pulled into the

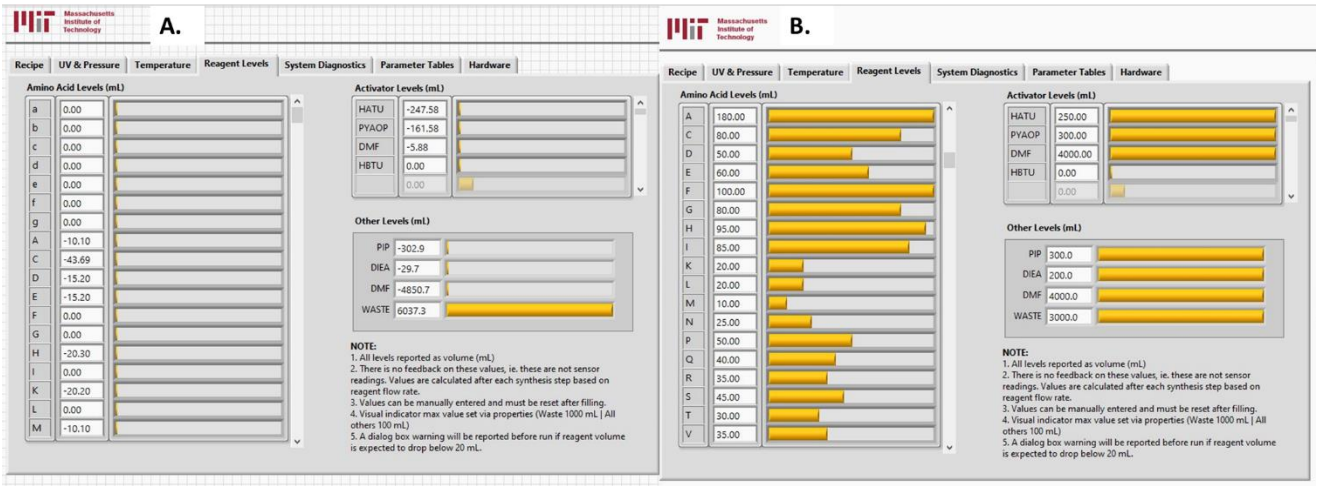


Figure 2.6: Reagent Tracking Software VI

A. Show work in progress VI with negative reagent levels, exactly what we are trying to avoid with the new code. **B.** Shows the updated reagent tracking VI with reagent levels shown, as synthesis progresses the reagent levels will drop and give the user a warning before reagents run out. Additionally at the start of the run the software calculates the amount of each reagent that will be used in the synthesis and lets the user know if they will run out during the run so the user can prepare any reagents promptly.

solvent lines, through the pumps, and into the in-line heating loops. This severely diminishes synthesis quality and outcomes and damages critical hardware on the machine. Further, once air entered the solvent lines and pumps, the entire synthesis had to be stopped, and the lab and AFPS managers needed to be brought in to repair the damage.

2.3.2 Synthesis sequence updates

Next, we overhauled the synthesis sequence text file, SYNTHESIS_SEQUENCE_BASELINE.syn, which contains the line-by-line commands to control peptide synthesis. This central file contains all the necessary controls and calls for additional files and variables needed to facilitate instrument usage. All LabView code connects back to this file. If this file is out of date or incorrect, peptide elongation will not proceed as desired, but keeping this code malleable would enable easy chemistry tests and changes for future users. Based on the data collected from the synthesis of some test peptides, we concluded that the synthesis sequence file needed to be overhauled. With Ryan Beaumont's expertise with coding and our chemical knowledge, we could not only update the syn file but change it so that future users would not have difficulty changing the files as needed.[6] Of the other files that SYNTHESIS_SEQUENCE_BASELINE.syn references, we also updated the AMINO_ACID_PRAMS.syn and HEATER_PRAMS.syn to allow for changing synthesis

conditions when coupling non-canonical amino acids or even small molecule additions. Finally, we prepared the instrument for the synthesis of longer peptides by changing internal software and hardware settings to match those of the other instruments that were already making longer peptides with success. These improvements have resulted in a safer and easier user experience on the machine. Synthesis data obtained during this process is provided in section 2.5.

A. SYNTHESIS_SEQUENCE_MACHINE_01

STEP ID	STAGE	DESC	DWELL TIME (MSEC)	DEVICE TYPE	DEVICE ID	ACTION	VALIDATE	ARG1	ARG2	ARG3
1	PREWASH	Lock all pumps	#DELAY_RTN_PUMP	PUMP	#PUMP_ID_AMINO;#PUMP_ID_ACT;#PUMP_ID_DIEA	LOCK		0 NULL	NULL	NULL
2	PREWASH	All pumps off	#DELAY_RTN_PUMP	PUMP	#PUMP_ID_AMINO;#PUMP_ID_ACT;#PUMP_ID_DIEA	SET_FLOW_ZERO		0 NULL	NULL	NULL
3	PREWASH	Position HX valve	2000	VALVE	#VALVE_ID_HX	SET_POSITION		0 #VALVE_POS_HX01	NULL	NULL
4	PREWASH	Query TinyG (wake it up)	10	TINYG	NULL	STATUS		0 NULL	NULL	NULL
5	PREWASH	Home linear actuators	15000	TINYG	NULL	HOME		0 NULL	NULL	NULL
6	PREWASH	Close reactor- stage 1	22000	TINYG	NULL	CLOSE_STAGE1		0 NULL	NULL	NULL
7	PREWASH	Master valve to DMF	#DELAY_POS_VALVE	VALVE	#VALVE_ID_MASTER	SET_POSITION		0 #VALVE_POS_DMF_ALL	NULL	NULL
8	PREWASH	Act valve to DMF	#DELAY_POS_VALVE	VALVE	#VALVE_ID_ACT	SET_POSITION		0 #VALVE_POS_DMF_ALL	NULL	NULL
9	PREWASH	Pumps on (Fill syringe with DMF)	6000	PUMP	#PUMP_ID_AMINO;#PUMP_ID_ACT	SET_FLOW_RATE		0 40000;40000	NULL	4;4
10	PREWASH	Pumps off	50	PUMP	#PUMP_ID_AMINO;#PUMP_ID_ACT	SET_FLOW_ZERO		0 NULL	NULL	NULL
11	PREWASH	Master valve to blank	#DELAY_POS_VALVE	VALVE	#VALVE_ID_MASTER	SET_POSITION		0 #VALVE_POS_MASTER_PLUG	NULL	NULL
12	PREWASH	Close reactor- stage 2	45000	TINYG	NULL	CLOSE_STAGE2		0 NULL	NULL	NULL
13	PREWASH	Act valve to DMF	#DELAY_POS_VALVE	VALVE	#VALVE_ID_ACT	SET_POSITION		0 #VALVE_POS_DMF_ALL	NULL	NULL
14	PREWASH	Master valve to DMF	#DELAY_POS_VALVE	VALVE	#VALVE_ID_MASTER	SET_POSITION		0 #VALVE_POS_DMF_ALL	NULL	NULL
15	PREWASH	AA and Act pumps on (Prewash)	60000	PUMP	#PUMP_ID_AMINO;#PUMP_ID_ACT	SET_FLOW_RATE		0 40000;40000	NULL	4;4
16	PREWASH	AA and Act pumps off	#DELAY_RTN_PUMP	PUMP	#PUMP_ID_AMINO;#PUMP_ID_ACT	SET_FLOW_ZERO		0 NULL	NULL	NULL
17	PRIME_0	Position HX valve	2000	VALVE	#VALVE_ID_HX	SET_POSITION		0 #VALVE_POS_HX_CPL	NULL	NULL
18	PRIME_0	Position AA bank valve	0	VALVE	#VALVE_ID_AMINO	SET_POSITION		0 #VALVE_POS_AMINO	NULL	NULL
19	PRIME_0	Position master valve	0	VALVE	#VALVE_ID_MASTER	SET_POSITION		0 #VALVE_POS_MASTER	NULL	NULL
20	PRIME_0	Position act valve	#DELAY_POS_VALVE	VALVE	#VALVE_ID_ACT	SET_POSITION		0 #VALVE_POS_ACT	NULL	NULL
21	PRIME_0	AA and Act pumps on (Prime)	#DWELL_TIME_AMINO	PUMP	#PUMP_ID_AMINO;#PUMP_ID_ACT	SET_FLOW_RATE		0 #FLOW_RATE_AMINO;#FLOW_RATE_AMINO	NULL	1;2
22	PRIME_0	AA and Act pumps off	#DELAY_RTN_PUMP	PUMP	#PUMP_ID_AMINO;#PUMP_ID_ACT	SET_FLOW_ZERO		0 NULL	NULL	NULL
23	PRIME_0	DIEA AA and Act pumps on (Couple)*	#DWELL_TIME_DIEA_COUPLING	PUMP	#PUMP_ID_DIEA;#PUMP_ID_AMINO;#PUMP_ID_ACT	SET_FLOW_RATE		0 #FLOW_RATE_DIEA;#FLOW_RATE_AMINO;#FLOW_RATE_AMINO	NULL	5;1;2
24	PRIME_0	All pumps off	#DELAY_RTN_PUMP	PUMP	#PUMP_ID_AMINO;#PUMP_ID_ACT;#PUMP_ID_DIEA	SET_FLOW_ZERO		0 NULL	NULL	NULL
25	PRIME_0	Act valve to DMF	0	VALVE	#VALVE_ID_ACT	SET_POSITION		0 #VALVE_POS_DMF_ALL	NULL	NULL
26	PRIME_0	AA valve to DMF	#DELAY_POS_VALVE	VALVE	#VALVE_ID_AMINO	SET_POSITION		0 #VALVE_POS_DMF_ALL	NULL	NULL
27	PRIME_0	All pumps on (Wash valves)	#DWELL_TIME_AMINO	PUMP	#PUMP_ID_DIEA;#PUMP_ID_AMINO;#PUMP_ID_ACT	SET_FLOW_RATE		0 #FLOW_RATE_AMINO;#FLOW_RATE_AMINO;#FLOW_RATE_AMINO	NULL	4;4;4
28	PRIME_0	All pumps off	#DELAY_RTN_PUMP	PUMP	#PUMP_ID_AMINO;#PUMP_ID_ACT;#PUMP_ID_DIEA	SET_FLOW_ZERO		0 NULL	NULL	NULL
29	PRIME_0	AA and Act pumps on (Wash resin)	#DWELL_TIME_WASH	PUMP	#PUMP_ID_AMINO;#PUMP_ID_ACT	SET_FLOW_RATE		0 #FLOW_RATE_AMINO;#FLOW_RATE_AMINO	NULL	4;4
30	PRIME_0	AA and Act pumps off	#DELAY_RTN_PUMP	PUMP	#PUMP_ID_AMINO;#PUMP_ID_ACT	SET_FLOW_ZERO		0 NULL	NULL	NULL
73	DEPRO_0	Position HX valve	0	VALVE	#VALVE_ID_HX	SET_POSITION		0 #VALVE_POS_HX_DEPRO	NULL	NULL
74	DEPRO_0	Master valve to dmf	0	VALVE	#VALVE_ID_MASTER	SET_POSITION		0 #VALVE_POS_DMF_ALL	NULL	NULL
75	DEPRO_0	Pip (act) valve to pip	2000	VALVE	#VALVE_ID_PIP	SET_POSITION		0 #VALVE_POS_PIP	NULL	NULL
76	DEPRO_0	AA and Act pumps on (Deprotect)	#DWELL_TIME_DEPRO	PUMP	#PUMP_ID_AMINO;#PUMP_ID_ACT	SET_FLOW_RATE		0 #FLOW_RATE_DEPRO;#FLOW_RATE_DEPRO	NULL	4;3
77	DEPRO_0	AA and Act pumps off	#DELAY_RTN_PUMP	PUMP	#PUMP_ID_AMINO;#PUMP_ID_ACT	SET_FLOW_ZERO		0 NULL	NULL	NULL
78	DEPRO_0	Pip (act) valve to dmf	#DELAY_POS_VALVE	VALVE	#VALVE_ID_PIP	SET_POSITION		0 #VALVE_POS_DMF_ALL	NULL	NULL
79	DEPRO_0	AA and Act pumps on (Wash)	#DWELL_TIME_WASH	PUMP	#PUMP_ID_AMINO;#PUMP_ID_ACT	SET_FLOW_RATE		0 #FLOW_RATE_DEPRO;#FLOW_RATE_DEPRO	NULL	4;4
80	DEPRO_0	AA and Act pumps off	#DELAY_RTN_PUMP	PUMP	#PUMP_ID_AMINO;#PUMP_ID_ACT	SET_FLOW_ZERO		0 NULL	NULL	NULL
105	POSTWASH	AA and Act pumps off (Postwash)	1000	PUMP	#PUMP_ID_AMINO;#PUMP_ID_ACT;#PUMP_ID_DIEA	SET_FLOW_ZERO		0 NULL	NULL	NULL
106	POSTWASH	Position HX valve	2000	VALVE	#VALVE_ID_HX	SET_POSITION		0 #VALVE_POS_HX01	NULL	NULL
107	POSTWASH	AA and Act pumps on (Postwash)	40000	PUMP	#PUMP_ID_AMINO;#PUMP_ID_ACT	SET_FLOW_RATE		0 40000;40000	NULL	4;4
108	POSTWASH	AA and Act pumps off	#DELAY_RTN_PUMP	PUMP	#PUMP_ID_AMINO;#PUMP_ID_ACT;#PUMP_ID_DIEA	SET_FLOW_ZERO		0 NULL	NULL	NULL
109	POSTWASH	Position HX valve	2000	VALVE	#VALVE_ID_HX	SET_POSITION		0 #VALVE_POS_HX02	NULL	NULL
110	POSTWASH	AA and Act pumps on (Postwash)	40000	PUMP	#PUMP_ID_AMINO;#PUMP_ID_ACT	SET_FLOW_RATE		0 40000;40000	NULL	4;4
111	POSTWASH	AA and Act pumps off	#DELAY_RTN_PUMP	PUMP	#PUMP_ID_AMINO;#PUMP_ID_ACT;#PUMP_ID_DIEA	SET_FLOW_ZERO		0 NULL	NULL	NULL
112	POSTWASH	Set pump display	#DELAY_RTN_PUMP	PUMP	#PUMP_ID_AMINO;#PUMP_ID_ACT;#PUMP_ID_DIEA	SET_DISPLAY		0 NULL	SEQUEN	NULL
113	POSTWASH	Master valve to plug	#DELAY_POS_VALVE	VALVE	#VALVE_ID_MASTER	SET_POSITION		0 #VALVE_POS_MASTER_PLUG	NULL	NULL
114	POSTWASH	ACT valve to plug	#DELAY_POS_VALVE	VALVE	#VALVE_ID_ACT	SET_POSITION		0 #VALVE_POS_ACT_PLUG	NULL	NULL
115	POSTWASH	Open reactor	10000	TINYG	NULL	OPEN		0 NULL	NULL	NULL

B. PRIME_SEQEUNCE_MACHINE_01

STEP ID	STAGE	DESC	DWELL TIME (MSEC)	DEVICE TYPE	DEVICE ID	ACTION	VALIDATE	ARG1	ARG2	ARG3
1	PREWASH	LOCK ALL PUMPS	#DELAY_RTN_PUMP	PUMP	#PUMP_ID_AMINO;#PUMP_ID_ACT;#PUMP_ID_DIEA	LOCK		0 NULL	NULL	NULL
19	PRIME_0	CHANGE HX VALVE POS	2000	VALVE	#VALVE_ID_HX	SET_POSITION		0 #VALVE_POS_HX_CPL	NULL	NULL
21	PRIME_0	POSITION AA VALVE	0	VALVE	#VALVE_ID_AMINO	SET_POSITION		0 #VALVE_POS_AMINO	NULL	NULL
22	PRIME_0	POSITION MASTER VALVE	0	VALVE	#VALVE_ID_MASTER	SET_POSITION		0 #VALVE_POS_MASTER	NULL	NULL
24	PRIME_0	PRIME PUMP - AA TO MASTER VALVE	20000	PUMP	#PUMP_ID_AMINO	SET_FLOW_RATE		0 40000	NULL	1
25	PRIME_0	ZERO FLOW AA PUMP	#DELAY_RTN_PUMP	PUMP	#PUMP_ID_AMINO	SET_FLOW_ZERO		0 NULL	NULL	NULL
200	PRIME_0	POSITION AA VALVE	#DELAY_POS_VALVE	VALVE	#VALVE_ID_AMINO	SET_POSITION		0 #VALVE_POS_DMF_ALL	NULL	NULL
24	PRIME_0	PRIME PUMP - DMF TO MASTER VALVE	5000	PUMP	#PUMP_ID_AMINO	SET_FLOW_RATE		0 40000	NULL	4
25	PRIME_0	ZERO FLOW AA PUMP	#DELAY_RTN_PUMP	PUMP	#PUMP_ID_AMINO	SET_FLOW_ZERO		0 NULL	NULL	NULL
26	PRIME_1	CHANGE HX VALVE POS	2000	VALVE	#VALVE_ID_HX	SET_POSITION		0 #VALVE_POS_HX_CPL	NULL	NULL
37	DEPRO_0	HOME HANDLER	10	TINYG	NULL	STATUS		0 NULL	NULL	NULL
134	POSTWASH	ZERO FLOW ALL PUMPS	#DELAY_RTN_PUMP	PUMP	#PUMP_ID_AMINO;#PUMP_ID_ACT;#PUMP_ID_DIEA	SET_FLOW_ZERO		0 NULL	NULL	NULL
19	POSTWASH	CHANGE HX VALVE POS	2000	VALVE	#VALVE_ID_HX	SET_POSITION		0 #VALVE_POS_HX01	NULL	NULL
21	POSTWASH	POSITION AA VALVE	#DELAY_POS_VALVE	VALVE		SET_POSITION	7	SET #VALVE_POS_DMF_ALL	NULL	NULL
22	POSTWASH	POSITION MASTER VALVE	0	VALVE	#VALVE_ID_MASTER	SET_POSITION		0 7	NULL	NULL
34	POSTWASH	PUMP DMF VALVE 7	10000	PUMP	#PUMP_ID_AMINO	SET_FLOW_RATE		0 40000	NULL	4
35	POSTWASH	ZERO FLOW AA PUMP	#DELAY_RTN_PUMP	PUMP	#PUMP_ID_AMINO	SET_FLOW_ZERO		0 NULL	NULL	NULL
21	POSTWASH	POSITION AA VALVE	#DELAY_POS_VALVE	VALVE		SET_POSITION	5	SET #VALVE_POS_DMF_ALL	NULL	NULL
22	POSTWASH	POSITION MASTER VALVE	0	VALVE	#VALVE_ID_MASTER	SET_POSITION		0 3	NULL	NULL
34	POSTWASH	PUMP DMF VALVE 5	10000	PUMP	#PUMP_ID_AMINO	SET_FLOW_RATE		0 40000	NULL	4
35	POSTWASH	ZERO FLOW AA PUMP	#DELAY_RTN_PUMP	PUMP	#PUMP_ID_AMINO	SET_FLOW_ZERO		0 NULL	NULL	NULL
21	POSTWASH	POSITION AA VALVE	#DELAY_POS_VALVE	VALVE		SET_POSITION	3	SET #VALVE_POS_DMF_ALL	NULL	NULL
22	POSTWASH	POSITION MASTER VALVE	0	VALVE	#VALVE_ID_MASTER	SET_POSITION		0 3	NULL	NULL
34	POSTWASH	PUMP DMF VALVE 3	10000	PUMP	#PUMP_ID_AMINO	SET_FLOW_RATE		0 40000	NULL	4
35	POSTWASH	ZERO FLOW AA PUMP	#DELAY_RTN_PUMP	PUMP	#PUMP_ID_AMINO	SET_FLOW_ZERO		0 NULL	NULL	NULL
22	POSTWASH	POSITION MASTER VALVE	#DELAY_POS_VALVE	VALVE	#VALVE_ID_MASTER	SET_POSITION		0 #VALVE_POS_DMF_ALL	NULL	NULL
34	POSTWASH	PUMP DMF MASTER HX01	60000	PUMP	#PUMP_ID_AMINO	SET_FLOW_RATE		0 40000	NULL	4
35	POSTWASH	ZERO FLOW AA PUMP	#DELAY_RTN_PUMP	PUMP	#PUMP_ID_AMINO	SET_FLOW_ZERO		0 NULL	NULL	NULL
19	POSTWASH	CHANGE HX VALVE POS	2000	VALVE	#VALVE_ID_HX	SET_POSITION		0 #VALVE_POS_HX02	NULL	NULL
34	POSTWASH	PUMP DMF MASTER HX02	60000	PUMP	#PUMP_ID_AMINO	SET_FLOW_RATE		0 40000	NULL	4
35	POSTWASH	ZERO FLOW AA PUMP	#DELAY_RTN_PUMP	PUMP	#PUMP_ID_AMINO	SET_FLOW_ZERO		0 NULL	NULL	NULL
19	POSTWASH	CHANGE HX VALVE POS	2000	VALVE	#VALVE_ID_HX	SET_POSITION		0 #VALVE_POS_HX01	NULL	NULL

C. AMINO_PRAMS_MACHINE_01

ID	AA	VALVE AA POS	VALV	ACT COUP	Q PRIME A	NSTROKES	HX CPL	HX DEPRO	UNUSED	POS VALV	NSTROKES	Q DEPRO	NSTROKES	Q WASH	NSTROKES	DEPRO TY	PRIME TYPE
a		3	3	HATU	40000	5	90::10	90::10	unused	3	21	40000	13	40000	40	0	0
b		3	4	PYAOP	40000	5	25::5	90::10	unused	3	21	40000	13	40000	40	0	0
c		3	5	HATU	40000	5	90::10	90::10	unused	3	8	40000	13	40000	40	0	0
d		3	6	HATU	40000	5	90::10	90::10	unused	3	8	40000	13	40000	40	0	0
e		3	7	HATU	40000	5	90::10	90::10	unused	3	8	40000	13	40000	40	0	0
f		3	8	HATU	40000	5	90::10	90::10	unused	3	8	40000	13	40000	40	0	0
g		3	9	HATU	40000	5	90::10	90::10	unused	3	8	40000	13	40000	40	0	0
A		7	1	HATU	40000	5	90::10	90::10	unused	7	21	40000	13	40000	40	0	0
C		7	2	PYAOP	40000	5	60::5	90::10	unused	7	21	40000	13	40000	40	0	0
D		7	3	HATU	40000	5	90::10	90::10	unused	7	8	40000	13	40000	40	0	0
E		7	4	HATU	40000	5	90::10	90::10	unused	7	8	40000	13	40000	40	0	0
F		7	5	HATU	40000	5	90::10	90::10	unused	7	8	40000	13	40000	40	0	0
G		7	6	HATU	40000	5	90::10	90::10	unused	7	8	40000	13	40000	40	0	0
H		7	7	PYAOP	40000	5	60::5	90::10	unused	7	21	40000	13	40000	40	0	0
I		7	8	HATU	40000	5	90::10	90::10	unused	7	8	40000	13	40000	40	0	0
K		7	9	HATU	40000	5	90::10	90::10	unused	7	8	40000	13	40000	40	0	0
L		5	1	HATU	40000	5	90::10	90::10	unused	5	8	40000	13	40000	40	0	0
M		5	2	HATU	40000	5	90::10	90::10	unused	5	8	40000	13	40000	40	0	0
N		5	3	HATU	40000	5	90::10	90::10	unused	5	21	40000	13	40000	40	0	0
P		5	4	HATU	40000	5	90::10	90::10	unused	5	8	40000	13	40000	40	0	0
Q		5	5	PYAOP	40000	5	90::10	90::10	unused	5	21	40000	13	40000	40	0	0
R		5	6	PYAOP	40000	5	90::10	90::10	unused	5	21	40000	13	40000	40	0	0
S		5	7	HATU	40000	5	90::10	90::10	unused	5	21	40000	13	40000	40	0	0
T		5	8	PYAOP	40000	5	90::10	90::10	unused	5	21	40000	13	40000	40	0	0
V		5	9	PYAOP	40000	5	90::10	90::10	unused	5	21	40000	13	40000	40	0	0
W		3	4	HATU	40000	5	90::10	90::10	unused	3	8	40000	13	40000	40	0	0
Y		3	2	HATU	40000	5	90::10	90::10	unused	3	8	40000	13	40000	40	0	0
		3	10	DMF	40000	5	90::10	90::10	unused	3	0	40000	13	40000	19	0	0

Figure 2.7: Important synthesis files updated during renovation

A. Displays updated synthesis sequence with all steps to complete an amino acid coupling cycle, plus pre- and post- wash steps. This file is called from the main LabVIEW VI and drives all the hardware and software needed for a synthesis. **B.** Shows the priming software sequence discussed in section 2.3.3, similar to the synthesis sequence it contains all the steps needed to prime the lines of reagent bottles automatically. **C.** Shows the amino parameters that the synthesis sequence and LabVIEW VI pull the amino acid coupling conditions for individual amino acids from. Each of these files can be adjusted for custom synthesis.

2.3.3 Post-renovation LabView updates

Following the renovation in 2019, we continued to make improvements and adjustments to the hardware and software of all the AFPS instruments. Most notably, we condensed all of the

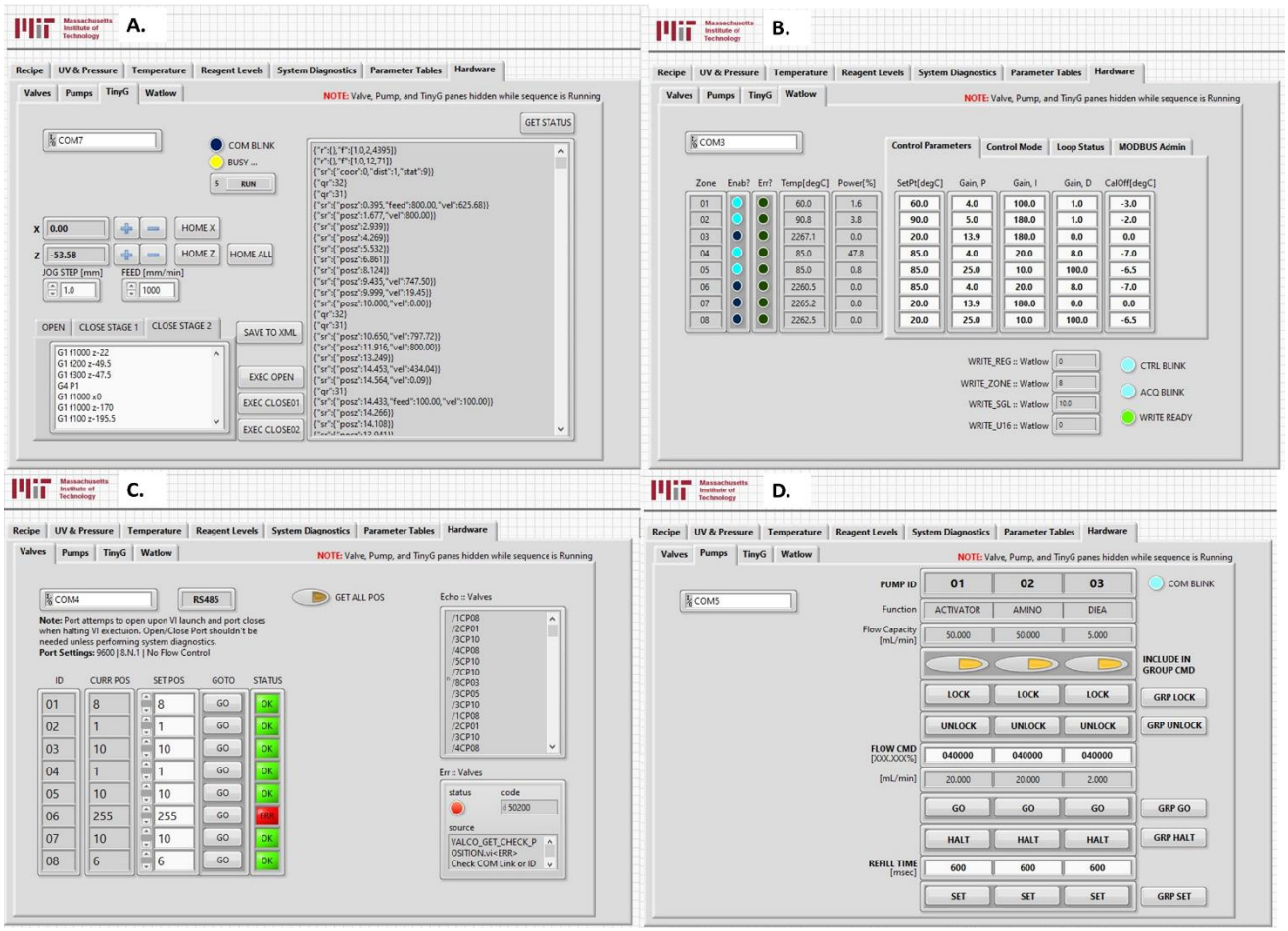


Figure 2.8: LabVIEW VI hardware controller

This figure shows the now combined hardware controllers in the main LabVIEW VI for the Automatide. Previously all controllers were separate instances of LabVIEW and could cause software failures if accidently opened or used. **A.** The TinyG Controller, robot arm movement can be controlled and changed here. **B.** The Watlow controller, Watlow controls set heater temperatures, values can be changed here if desired. **C.** Rotary Valve Controller, this window allows users to manually change valve positions outside of a synthesis. **D.** Pump Controller, this window can be used to turn pumps on and off to assess for pump health outside of a run.

individual valves, pump, robot arm (TinyG), and heater controls into one tab accessible from the main LabView window used for synthesis, figure 2.8.[17] This streamlined instrument repair and diagnostics and enabled us to train users to perform minor repairs, elevating the burden on the AFPS instrument managers. Finally, we added two new syn files, one that contains a capping protocol,[18] see Chapter 3 for more details, and one that includes a priming protocol. The priming protocol is intended to eliminate the need to hand-prime all of the reagent lines on the synthesizers. Instead, users can type the one-letter abbreviation of the amino acid ports that need to be primed, and the machine will automatically prime each line and then wash each valve with DMF.[19] Currently, this software only operates on the natural amino acid wells but could also be translated to the unnatural wells.

2.4 Investigation of pump refill time and its effect on synthesis outcomes

The AFPS instruments rely heavily on the balanced blend of mechanical, software, and chemical components that all combine to create the peptide synthesizer our group uses daily. At the heart of the AFPS instruments lie the HPLC pumps; these pumps transport all the activators, amino acids, deprotection agents, and base to a mixer and then into a reactor containing resin. It is vital for synthetic success that all the necessary components for an amino acid coupling arrive at the reactor at the same instant and are thoroughly homogenized.[20] In order for this to happen, it is necessary that the Varian HPLC pumps deliver reagents in perfect synchrony.[21] To the naked eye, it can appear that all reagents are being delivered correctly, but to truly understand the delivery to the microsecond, we must take a closer look. In the group, we have historically used a slow-motion camera, capable of capturing up to 10 seconds of footage at 960 frames per second to capture reagent delivery to the microsecond. To assist with visualization, we employed water-soluble colored dyes that, when they meet, mix to a different color. Early studies used blue and yellow dye, which would blend to produce a green-colored solution. For the studies in this report, we used green dye for activators and pink dye for amino acids that would mix to a light brown color. Due to the discontinuous nature of the HPLC pumps, it is first essential to ensure that each pump stroke occurs at the exact moment as the other pumps on the machine. If this perfect synchronization is not achieved, cavitation, the formation of air pockets during pump stroke, or other solvent delivery issues will occur.[22]

2.4.1 Initial dye test

The first series of tests resulted in the activator (green) being delivered before the amino acids (pink). Upon investigation, we noticed that the activator solvent line leading from the pump was approximately two inches longer than that of the amino acid; thus, the activator was being delivered before the amino acid. In regards to peptide elongation, this would have the chemical effect of the free amine on the growing peptide chain being capped by the excess activator. Secondly, we would observe very low per-coupling yields, and consequently, we would see very low yields of the desired peptide from a synthesis.[23] We rectified the issue by re-making the affected lines, and the following tests showed better pump synchronization. However, we still observed slightly delayed delivery of solvent solutions. However, the reagent being delayed was

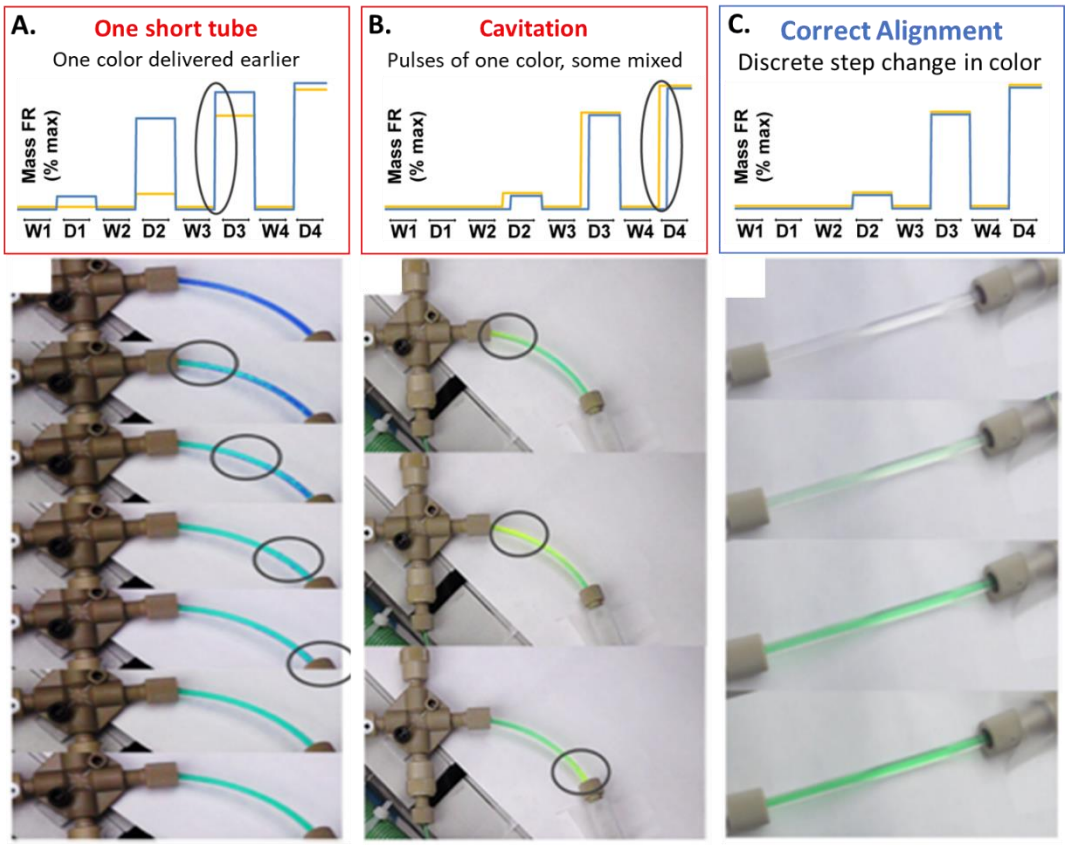


Figure 2.9: Dye tests possible results and demonstrations of incorrect alignment

A. Illustration of the effects of correct and incorrect pump synchronization on solution mixing. Incorrect pump synchronization will greatly diminish peptide yields and increase by-product production due to poor reagent delivery and timing. The following images show what we see when an AFPS system has one short tube, cavitation, and correct alignment. Blue dye corresponds to the activator HPLC pump and yellow dye to the amino acid HPLC pump, while the green color results from the mixing of the two dyes. **B.** Shows the effects of one short tube on dye mixing on the AFPS and as a cartoon illustration for clarity. Notice that the blue dye appears and disappears while the yellow dye has not yet arrived at the mixing tee, no green coloration indicating proper mixing is observed. **C.** Shows the effects of cavitation (due to pump desynchronization) on dye mixing on the AFPS and as a cartoon illustration for clarity. Notice the pulse of yellow followed by a pulse of green, indicating early delivery of yellow dye ahead of blue dye and improper mixing. **D.** Shows the effects of correct pump alignment and tube length on dye mixing on the AFPS and as a cartoon illustration for clarity. Notice that the color change from clear to green is discrete with no noticeable pulses of blue or yellow individually, and this indicates proper pump synchronization. This figure is adapted from Simon et al.

inconsistent. We saw a need to re-visit pump synchronization with a focus on the pumps themselves and the pump refill time.

2.4.2 Pump refill time changes

The essential factor we decided to investigate was the pump refill time, which is defined as the time it takes for the pump piston to pull back (withdraw, W) and then push forward (dispense, D) solvent. When pump refill times are faster, we can see a phenomenon known as cavitation that can occur at random. Cavitation is a state in which the static pressure of a liquid reduces to below the liquid's vapor pressure, leading to the formation of small vapor-filled cavities

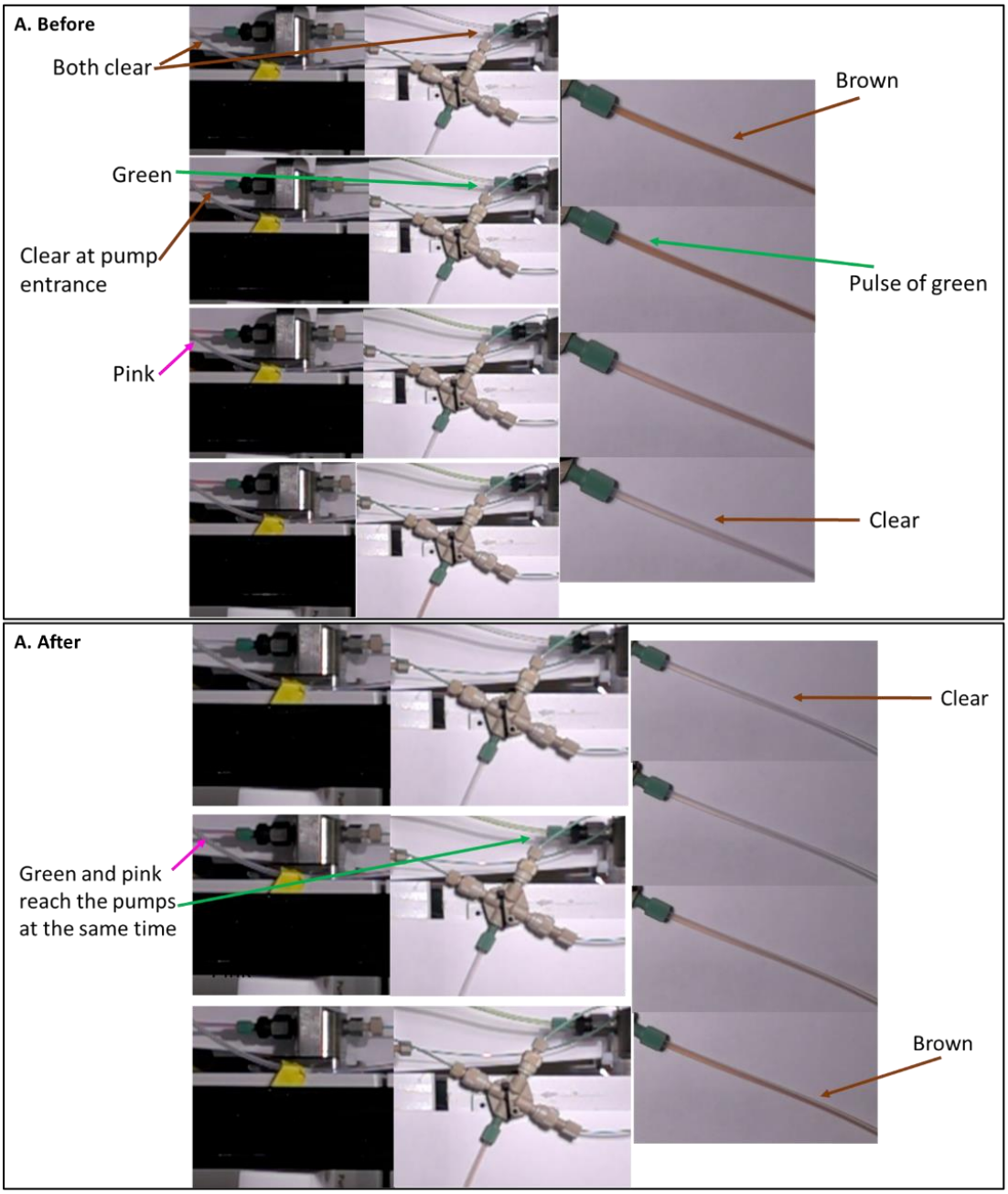


Figure 2.10: Automatide pump synchronization test results

A. Pump dye test results at 200 ms, notice the early delivery of green and the pulses of green throughout the test, this suggests that the pump is cavitating at the faster pump refill time. **B.** This shows the results of the now fully synchronized HPLC pumps, notice that both reagent lines deliver at the same time and mix to a uniform brown color.

in the liquid; in short, small air bubbles are forming, resulting in the delay in reagent delivery. The Varian ProStar HPLC pumps we use are capable of a range of refill times, and during the investigation, we noticed that each instrument's pumps were sent to a different refill time.[24] It went

as far as each pump at a single instrument was delivering reagents at vastly different rates, likely resulting in the poor reagent delivery we were observing via our slow-motion capture camera.

Interestingly, one AFPS instrument in the laboratory, known as the "Amidator," routinely outperformed the other AFPS instruments in synthesis outcomes and quality. At the time of the renovation, it was the only AFPS instrument capable of the synthesis of protein-length polypeptide chains. On investigation, we noted that the pump refill times were set to 600 milliseconds, so, with all other factors held constant, we hypothesized that the set refill time of 600ms on the Amidator pumps was the factor having a significant effect on synthesis quality and outcomes.[25] To assess the potential differential performance of the Amidator when compared to the Automatide, we performed the previously described dye tests on the Automatide again. For this test, we changed the refill time from 200 ms to 600 ms and observed that the previous flashes of unmixed green or pink dye had disappeared. Thus, we could conclude that the set pump refill time was critical for synthetic success. We further replicated these tests with the other AFPS instruments currently in the laboratory, the "Peptidator," with similar results.

2.4.3 Validation of pump refill time changes

As a final validation step, we synthesized the protein polypeptide for the ribonuclease barnase and its inhibitor barstar, which will be discussed in further detail in chapter 3 of this thesis, on the Automatide at the faster pump refill time of 200 ms and flow rate of 80 mL/min, and at the new, slower refill time of 600 ms with a flow rate of 40 mL/min. It is important to note that when the refill time on a pump is set to a relatively slow value of 600 ms, the overall flow rate must also be lowered. This is due to the added time needed to fill the pump head with each pump withdrawal; the volume within the pump head is set and cannot deliver solvent at the volumes required to maintain the 80 mL/min rate without causing severe damage to the pumps and increasing the risk of pump cavitation. The relationship between pump refill time and flow rate is an inverse relationship, so by halving the flow rate, we can maintain the delicate balance between protecting the pumps and maintaining rapid synthesis times. It is important to note that a 30-mer peptide can be synthesized in a mere 45 minutes at the faster flow rate but would take 1.5 hours at the slower flow rates. While we do see the disadvantage in the longer synthesis times, we believe that improving synthesis outcomes outweighs the slower production rate. Furthermore, even with increased synthesis times, we maintain a consistent waste output and the

AFPS technology is still the fastest and most reliable peptide synthesizer that has been reported, to our knowledge.

The results of the synthesis of barnase before and after the change in flow rate and refill time are stark. At the faster refill time and flow rate of 200 ms and 80 mL/min, we saw barely detectible levels of the barnase polypeptide via LCMS and HPLC analysis and a massive amount of side products. It was clear that the synthetic quality of this polypeptide was poor, and we could

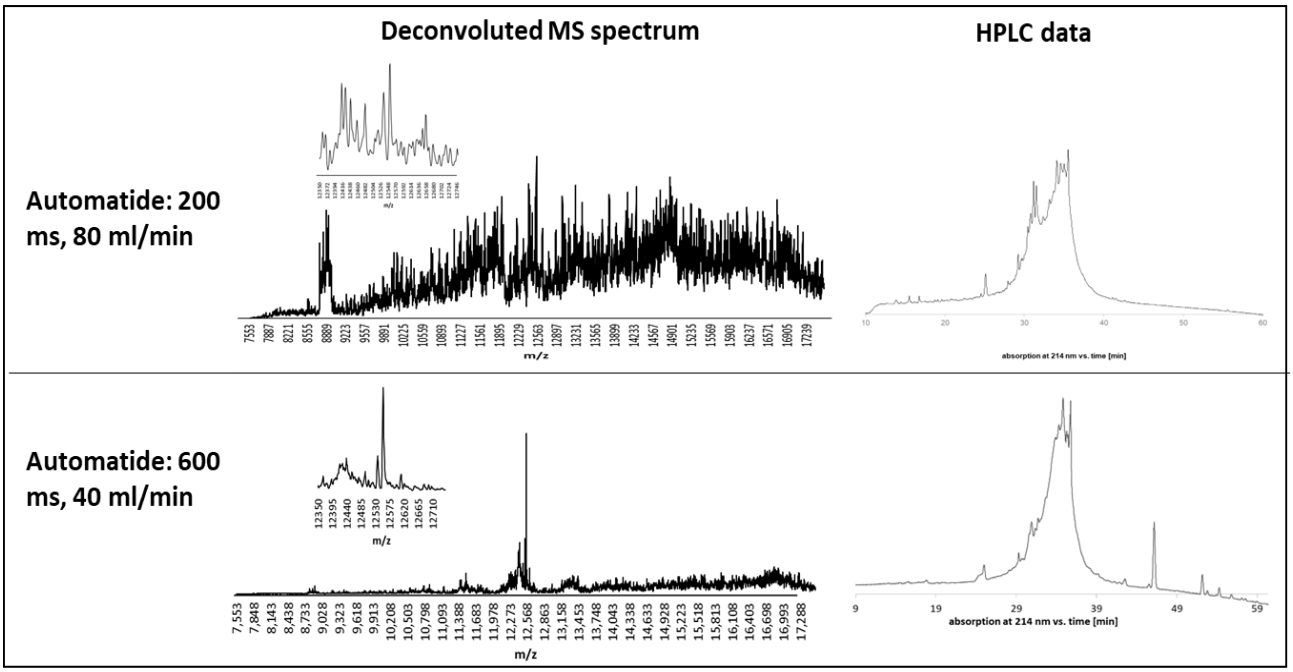


Figure 2.11: Barnase synthesis data at different pump refill times

LCMS and analytical HPLC data showing vast improvement of synthesis quality resulting from changing the refill time from 200 m/s at 80 mL/min to 600 m/s at 40 mL/min. The flow rate needs to be lowered when using a longer refill time to prevent damage to the HPLC pump heads. Note that after changing the refill time we can now see a distinct main peak for the barnase polypeptide, as well as, an increase in crude HPLC purity. The Automatide is now regularly used for single domain protein-length polypeptide synthesis and recently made a 225-mer with great success.

not separate the many truncations, deletions, and other peptide side products to yield a measurable quantity of the desired barnase polypeptide. However, once we slowed down the pump flow rate and refill time to 600 ms and 40 mL/min, we could now easily observe the full-length barnase polypeptide, and additionally when we examined the LCMS and HPLC traces, the synthetic by-products, while present, did not overwhelm the desired peptide, figure 2.11. We could now efficiently purify the desired barnase polypeptide away from any by-products. With these results in hand, it was clear to us that using a slower refill time would be critical to the synthetic success of long polypeptide chains. However, it is worth noting that the yield and purity of shorter peptide chains are significantly less affected by pump refill times. With this known, we can increase the

refill times and flow rates of AFPS instruments that only make peptides less than 30 amino acids in length with minimal effect on the yield per synthesis.

2.5 Synthetic results

The final step of our renovation was to synthesize some test peptide and protein sequences; for this, we selected a handful of short (<30 aa) and long peptide (>50 aa) sequences to synthesize and compare performance across three of our AFPS systems, the Automatide, the Peptidator, and the Amidator. A cross-instrument performance check not only gave us a chance to assess the success of our Automatide renovations but also a chance to assess the performance of the other AFPS machines.

2.5.1 Evaluation of short peptide sequences (<30 aa)

The first peptide we synthesized was known only as NN152, a peptide of interest to our pharmaceutical collaborators that had very poor synthesis outcomes on the Automatide before the rebuild.

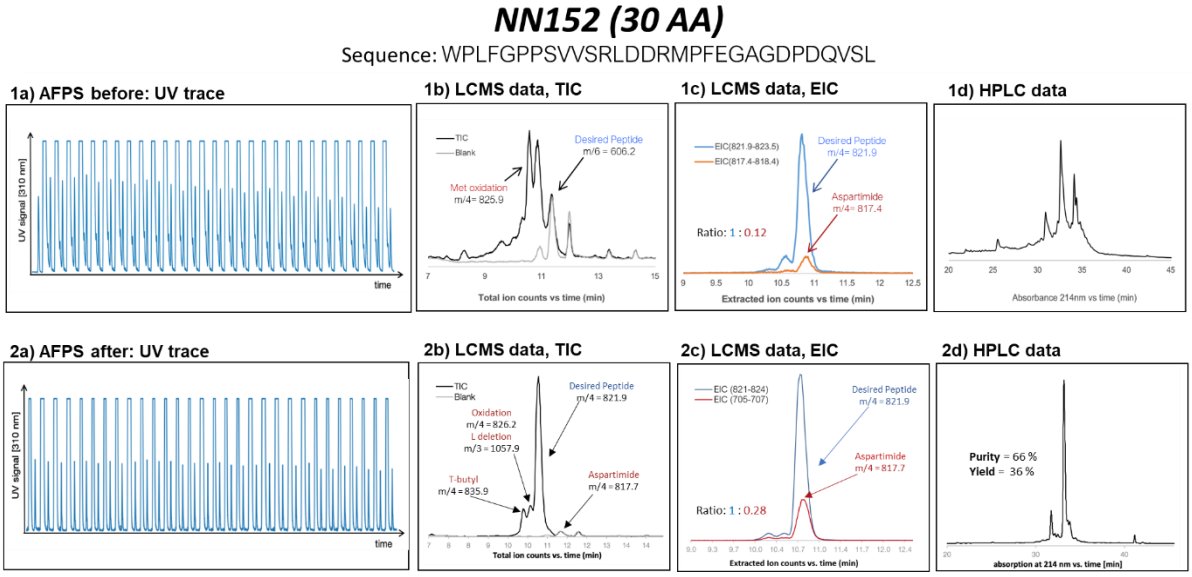


Figure 2.12: Synthetic data from NN152
Prior to the renovation synthetic data collected from the Automatide was of poor quality demonstrating clearly that the machine was not operating properly. Following renovation, we see a clear improvement in synthetic outcomes. Notice the sharp decrease in peptide by-products and much cleaner HPLC traces.

Next, we investigated the synthesis quality of a few other test peptides, including glucagon-like peptide-1 (GLP-1) and growth hormone-releasing hormone (GHRH). Both of these peptides are classically difficult to synthesize peptide with high therapeutic interest. GLP-1 plays an essential role in the secretion of insulin appetite control and can decrease inflammation and the

chance of cell apoptosis in certain circumstances.[26] GLP-1 and its derivatives are currently being successfully used to treat type-2 diabetes and as a weight loss agent.[27]

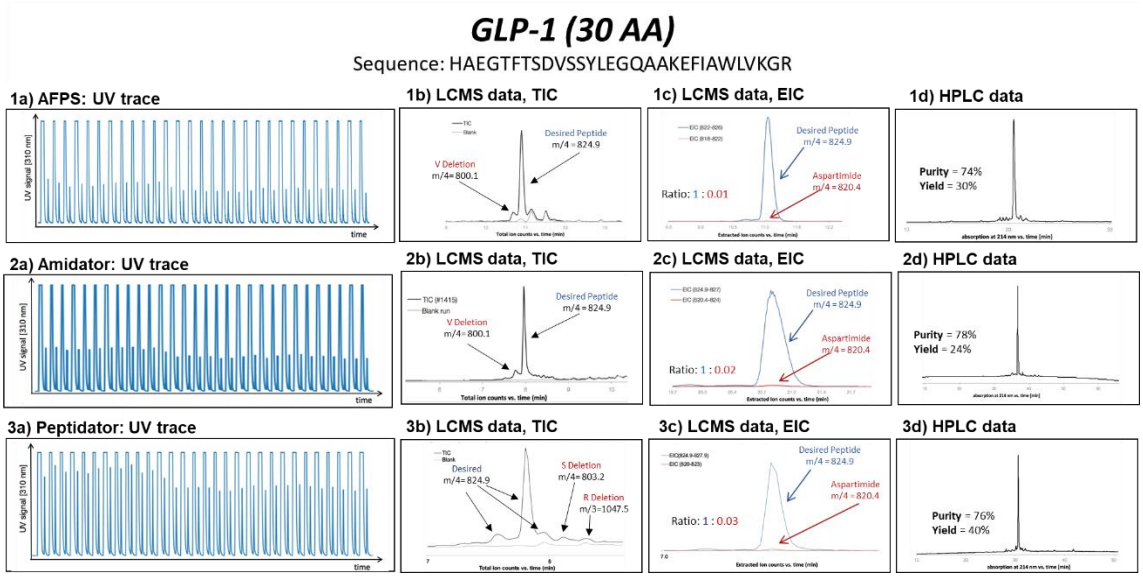


Figure 2.13: GLP-1 synthetic data from instrument comparisons

GHRH, on the other hand, is released from the hypothalamus and, in turn, stimulates the release of growth hormones from the pituitary gland.[28] Interestingly, GHRH was first reported in cancer tissues and has been investigated as a therapeutic target in cancer research.[29]

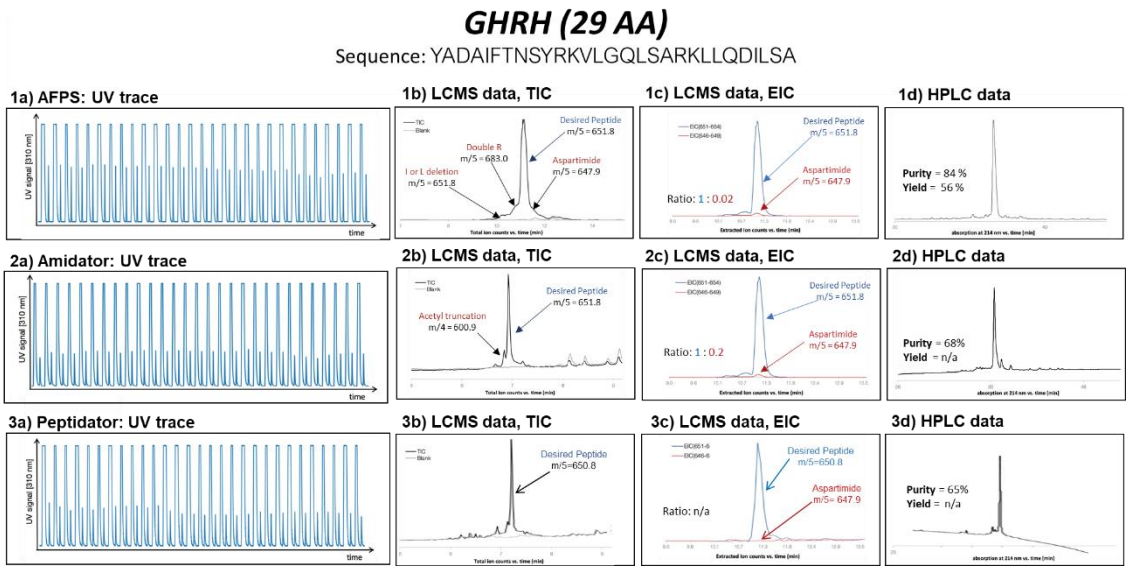


Figure 2.14: GHRH synthetic data from instrument comparisons

Finally, we also produced the peptide NN92, another peptide we produced for a pharmaceutical collaborator, that shows high levels of aspartimide formation. These three peptides will also be discussed in greater detail in Chapter 3, where we will provide further context for why we are investigating these peptides and what we have learned from their iterative synthesis.[30]

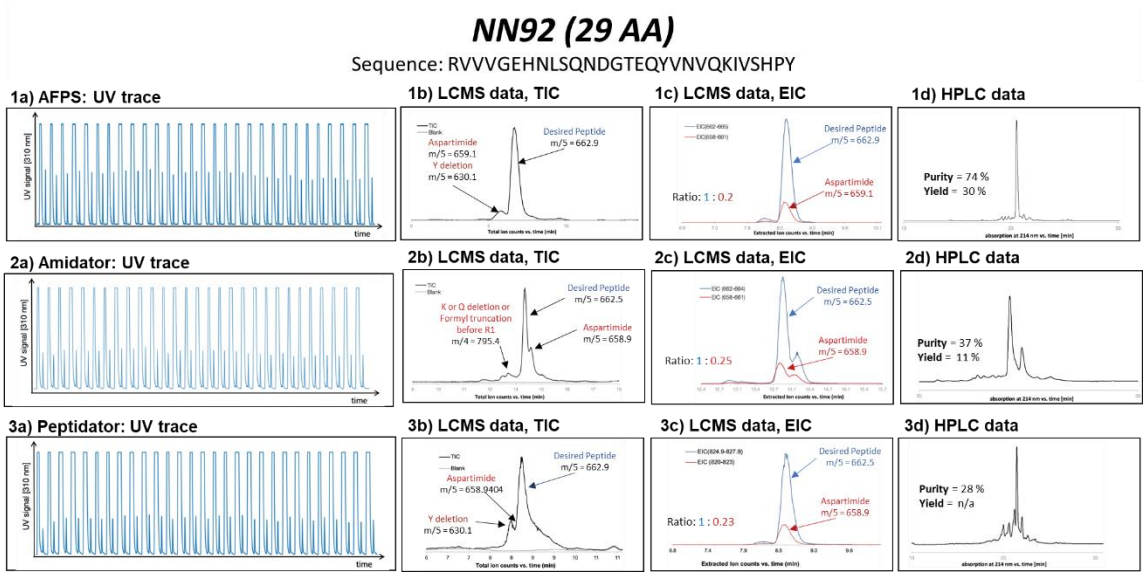


Figure 2.15: NN92 synthetic results for instrument comparison

This data shows that all our AFPS instruments perform similarly when synthesizing short

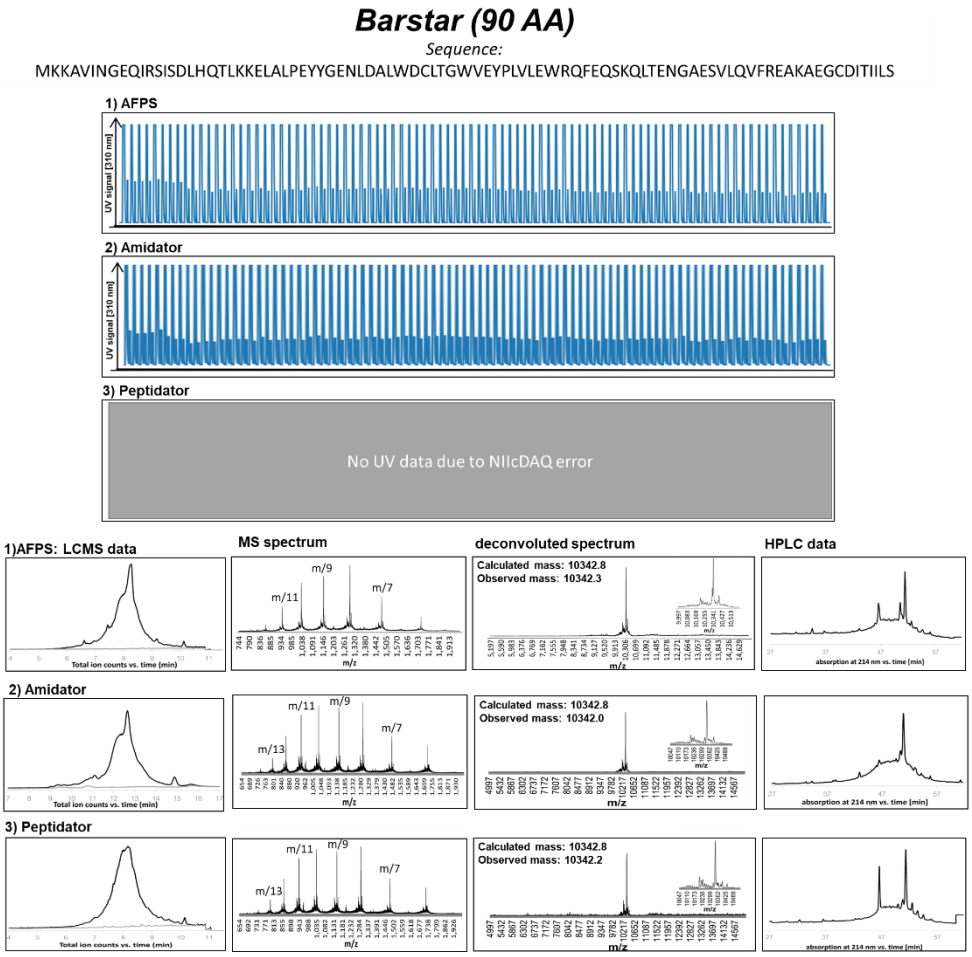


Figure 2.16: Barstar Synthetic data from instrument comparison

peptide sequences.

2.5.2 Evaluation of long peptide sequences (>80 aa)

After the successes we observed with our test peptides, we turned to full-length protein sequences, each over 80 amino acids in length. The three proteins we focused on were barnase, barstar, and HIV-1 protease. Barnase, the longest sequence of the three at 113 aa (12.5 kDa), is a ribonuclease and cleaves RNA strands at GpN sites.[31] Barnase has an enzymatic inhibitor in the form of the protein barstar, a 90 aa (10kDa) small protein.[32] This pair has been intensely studied, and we will dive into the in-depth characterization of this protein in Chapter 3.

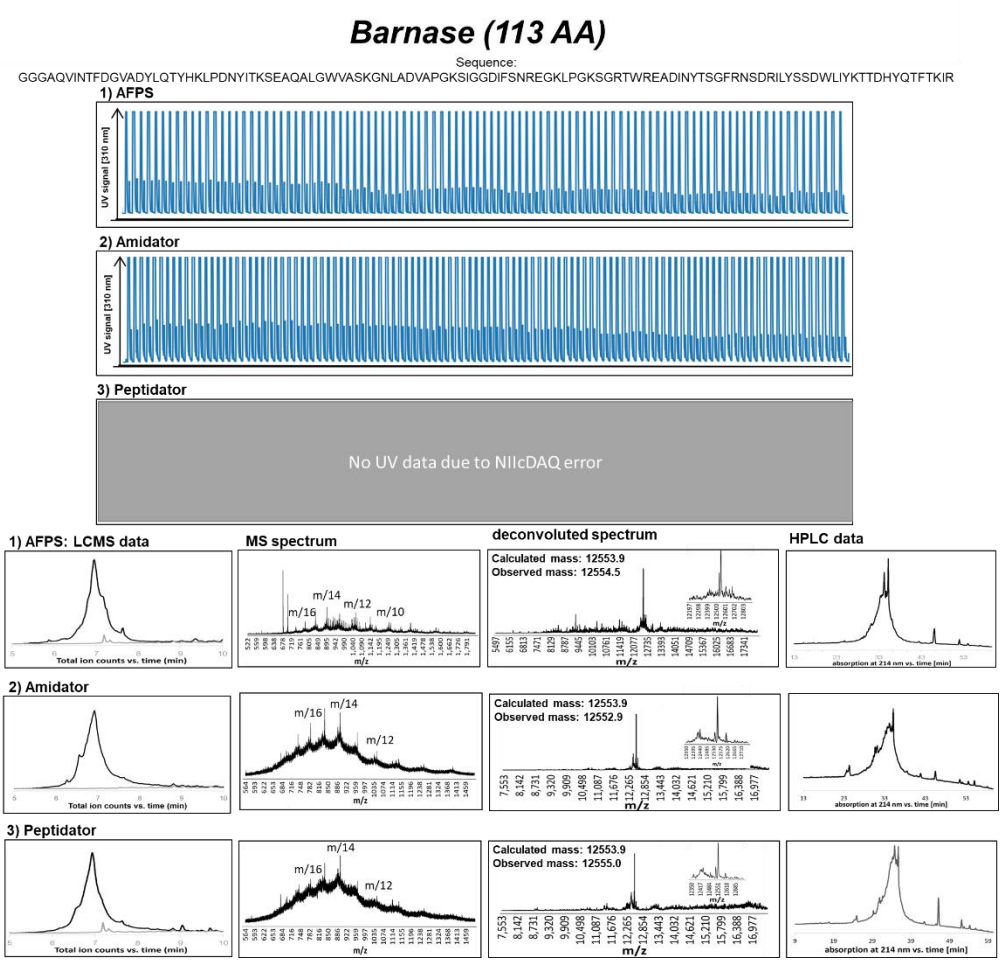


Figure 2.17: Barnase Synthetic data form instrument comparison

Finally, we synthesized the protein polypeptide HIV-1 protease, a very popular synthetic target in the SPPS community.[33] HIV-1 protease is a prominent player in viral maturation for producing new infectious virus particles, an essential step for the life and infection cycle of HIV.[34] Thus, HIV-1 protease is a popular target for drug discovery, with several currently available

protease inhibitors available to the public. These inhibitors help prevent HIV infection from occurring and can be used to treat current infections.[35]

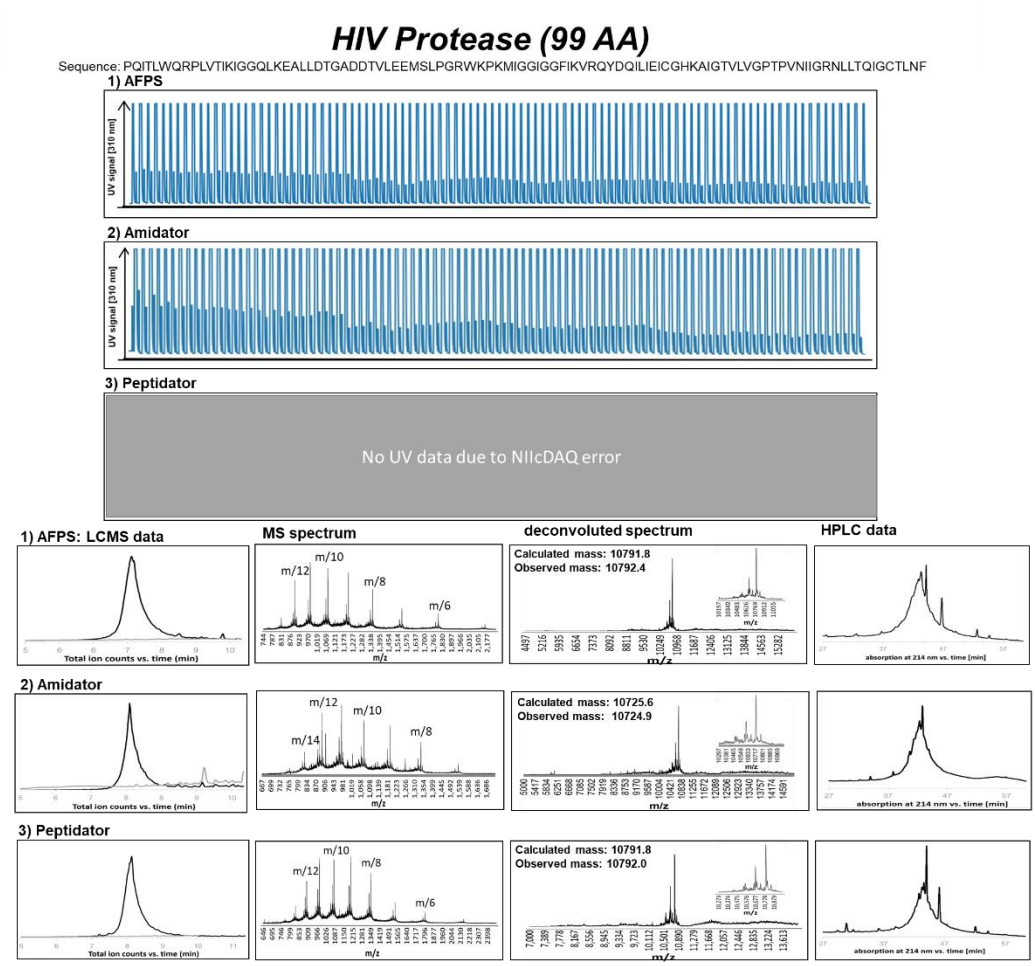


Figure 2.18: HIV-1 protease synthetic data from instrument comparisons

This data demonstrates that our renovation and critical analysis of the Automatide resulted in improvements in synthesis quality and pulled ahead of the Amidator and Peptidator in terms of per-synthesis purity and yield. Moreover, it was now apparent that each AFPS machine currently built in the group is capable of long polypeptide synthesis in excess of 150 amino acids and performs analogously across all instruments. This opens the door for future synthetic pursuits in protein chemistry.

2.6 Discussion and outlook

In this chapter, I delved into the importance of maintaining and designing our AFPS systems to be robust, easy to repair, and user-friendly with consistent performance from machine to machine. It is clear that extensive thought is needed to design these machines and that we carry

Robosome (2019)

Figure 2.19: Newest AFPS systems

Top: Image of the ribosome, installed in 2019 after the Automatide renovation. **Bottom:** Image of the currently unnamed AFPS instrument replacing the damaged Amidator. Both instruments were designed based off of what we learned from the AFPS renovation.

over software, hardware, and chemical settings that work on one machine to the next while at the same time removing and/or replacing components that are prone to failure or altogether not working to improve machine performance. As the laboratory has grown and changed over the five years following the Automatide renovation, we have continued to improve the AFPS instruments as we install new generations of the machine, and the vital information we obtained from this renovation has inextricably influenced how these new generations are designed and built. For instance, not long after the Automatide renovation, I installed a brand new AFPS machine, the Robosome, and we translated our key design changes to this new machine. Moreover, after the unfortunate loss of the 2nd generation machine, the Amidator, to water damage in early 2023, the

new machine being installed in its place was entirely designed by the Automatide renovation team in a concise effort to translate further the findings from the renovation to our newest and hopefully best AFPS instrument. Furthermore, during the time of the Automatide renovation, we were also hard at work optimizing the Amidator for protein synthesis. The process by which this was achieved will be enumerated in Chapter 3.

2.7 Materials and methods

This section outlines the general methods and materials used for the work reported in chapter 2. Methods and/or materials in this section may not apply to future sections.

2.7.1 Reagents and solvents

All reagents were purchased and used as received. Fmoc-protected amino acids (Fmoc-Ala-OHxH2O, Fmoc-Arg(Pbf)-OH; Fmoc-Asn(Trt)-OH; Fmoc-Asp-(Ot-Bu)-OH; Fmoc-Cys(Trt)-OH; Fmoc-Gln(Trt)-OH; Fmoc-Glu(Ot-Bu)-OH; Fmoc-Gly-OH; Fmoc-His(Trt)-OH; Fmoc-Ile-OH; Fmoc-Leu-OH; Fmoc-Lys(Boc)-OH; Fmoc-Met-OH; Fmoc-Phe-OH; Fmoc-Pro-OH; Fmoc-Ser(But)-OH; Fmoc-Thr(t-Bu)-OH; Fmoc-Trp(Boc)-OH; Fmoc-Tyr(t-Bu)-OH; Fmoc-Val-OH)were purchased from the Novabiochem-line from Sigma Millipore. O-(7-azabenzotriazol-1-yl)-N,N,N’,N’-tetramethyluronium hexafluorophosphate (HATU, ≥97.0%), and (7-azabenzotriazol-1- yloxy)tripyrrolidinophospho-nium hexa-fluorophosphate (PyAOP, ≥97.0%) were purchased from P3 Biosystems. Biosynthesis OmniSolv® grade N,N-dimethylformamide (DMF) was purchased from EMD Millipore (DX1732-1). AldraAmine trapping agents (for 1000 – 4000 mL DMF, catalog number Z511706), Diisopropylethylamine (DIEA; 99.5%, biotech grade, catalog number 387649), piperidine (ACS reagent, ≥99.0%), trifluoroacetic acid (HPLC grade, ≥99.0%), triisopropylsilane (≥98.0%), acetonitrile (HPLC grade), formic acid (FA, ≥95.0%), dimethyl sulfoxide (DMSO, HPLC grade, ≥99.7%), piperazine (≥99.0%) and 1,2- ethanedithiol (EDT, GC grade, ≥98.0%) were purchased from Sigma-Aldrich. Tetrafluoroethylene (TFE, extra pure, ≥99.8%) was purchased from Acros. Anisole (purum, GC grade, ≥99.0%) and 1,8-diazabicyclo(5.4.0)undec-7-ene (DBU, GC grade, ≥99.0%) were purchased from Fluka. H-Rink Amide (0.49 mmol/g and 0.18 mmol/g loading) and HMPB ChemMatrix polyethylene glycol (0.45 mmol/g loading) resin were purchased from PCAS Biomatrix. Water was deionized using a Milli-Q Reference water purification system (Millipore). Nylon 0.22 µm syringe filters were TISCH brand SPEC17984.

2.7.2 AFPS Synthesis method before renovation

All peptide data from prior to the renovation were synthesized under the following conditions on the Automatide, this system set-up is published.[6]

Unless otherwise noted, the following settings were used for peptide synthesis: flow-rate = 80 mL/min, temperature = 80 °C (loop) and 90 °C (reactor). The 50 mL/min pump head pumps 400 µL of liquid per pump stroke; the 5 mL/min pump head pumps 40 µL of liquid per pump stroke. All pumps were at a pump refill rate of 200 ms.

To start the machine a filtered syringe cartridge loaded with resin is mounted on an automated robot arm which moves the syringe into the reactor. Prior to reactor closing DMF is pumped into the syringe for 6 pumping strokes, to swell the resin. standard synthetic cycle involves a first step of prewashing the resin at elevated temperatures for 30 s at 80 mL/min. During the coupling step, three HPLC pumps are used: a 50 mL/min pump head pumps the activating agent, a second 50 mL/min pump head pumps the amino acid and a 5 mL/min pump head pumps 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 deprotection stock solution and DMF. The pumps are activated for 14 pump strokes. Both solutions are mixed in a 1:1 ratio. 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. The final synthetic step is a postwash of the resin at elevated temperatures for 30 s at 80 mL/min. A full coupling cycle is complete in 40 seconds.

2.7.3 AFPS synthesis method following renovation

All peptide data from after the renovation were synthesized under the following conditions on the Automatide, this system set-up is published.[30]

Unless otherwise noted, the following settings were used for peptide synthesis: flow-rate = 40 mL/min, temperature =70-90 °C (loop) and 90 °C (reactor). The 50 mL/min pump head

pumps 400 µL of liquid per pump stroke; the 5 mL/min pump head pumps 40 µL of liquid per pump stroke. All pumps were at a pump refill rate of 600 ms.

To start the machine a filtered syringe cartridge loaded with resin is mounted on an automated robot arm which moves the syringe into the reactor. Prior to reactor closing DMF is pumped into the syringe for 8 pumping strokes, to swell the resin. standard synthetic cycle involves a first step of prewashing the resin at elevated temperatures for 60 s at 40 mL/min. During the coupling step, three HPLC pumps are used: a 50 mL/min pump head pumps the activating agent, a second 50 mL/min pump head pumps the amino acid and a 5 mL/min pump head pumps DIEA. The first two pumps are activated for 8 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 8 pumping strokes, after which the DIEA pump is shut off and the other two pumps continue to wash the resin for another 40 pump strokes. During the deprotection step, two HPLC pumps are used. Using a rotary valve, one HPLC pump selects deprotection stock solution and DMF. The pumps are activated for 13 pump strokes. Both solutions are mixed in a 1:1 ratio. Next, the rotary valves select DMF for both HPLC pumps, and the resin is washed for an additional 40 pump strokes. The coupling–deprotection cycle is repeated for all additional monomers. The final synthetic step is a postwash of the resin at elevated temperatures for 60 s at 40 mL/min. A full synthetic cycle is complete in 120 seconds.

2.7.4 Peptide cleavage protocol

Directly after synthesis, the resin was washed with dichloromethane (3 x 5 mL), dried in a vacuum chamber, and weighed. 50% of the resin was transferred into a 50 mL conical polypropylene tube. For cleavage of peptides without cystine residues and peptides/protein sequences with cystine residues we used two different protocols.

A) Peptides[36]: Approximately 3 mL of cleavage solution (94% TFA, 1% TIPS, 2.5% EDT, 2.5% water) was added to the tube. If needed, more cleavage solution was added to ensure complete submersion. The tube was kept at room temperature for 2 hours.

B) Proteins and Cys-containing peptides:[37] Approximately 5 mL of cleavage solution (82.5% TFA, 5% water, 5% phenol, 5% thioanisole, 2.5% EDT) was added to the tube. If needed,

more cleavage solution was added to ensure complete submersion. The tube was kept on a nutating mixer at room temperature for 4 h.

After incubation ice cold diethyl ether (45 mL) was added to the cleavage mixture and the precipitate was collected by centrifugation and triturated twice more with cold diethyl ether (45 mL). The supernatant was discarded at each step. Residual ether was allowed to evaporate for 20 minutes after which the peptide was dissolved in 50% acetonitrile in water with 0.1% TFA (long peptides were dissolved 70% acetonitrile in water with 0.1% TFA). The peptide solution was filtrated with a Nylon 0.22 µm syringe filter and frozen, lyophilized until dry, and weighed.

2.7.5 Liquid chromatography-mass spectrometry (LC-MS)

For mass analysis of peptides, a filtered peptide solution (10 µL of a 1mg/mL solution) was diluted in 50% acetonitrile in water with 0.1% TFA (90 µL) to a final concentration approximately 0.1 mg/mL. LC-MS chromatograms and associated high resolution mass spectra were acquired using an Agilent 6550 iFunnel Q-TOF LC-MS system (abbreviated as 6550). Solvent compositions used in the LC-MS are 0.1% formic acid in H2O (solvent A) and 0.1% formic acid in acetonitrile (solvent B).

The following LC-MS methods were used:

A) 1-61% B over 15 min, Phenomenex Jupiter C4 column (6550)
   a. LC conditions: Phenomenex Jupiter C4 column: 1.0 × 150 mm, 5 µm, column temperature: 40 °C, gradient: 0-2 min 1% B, 2-12 min 1-61% B, 12-16 min 61-90% B; flow rate: 0.1 mL/min. A final 4-min hold was performed at a flow rate of 0.1 mL/min. The total method time was 20 min. MS is on from 4 to 12 min.
   b. MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 100–1700 m/z.

B) 1-61% B over 33 min, Phenomenex Jupiter C4 column (6550)
   a. LC conditions: Phenomenex Jupiter C4 column: 1.0 × 150 mm, 5 µm, column temperature: 40 °C, gradient: 0-2 min 1% B, 2-30 min 1-91% B, 30-34 min 61-90% B; flow rate: 0.1 mL/min. A final 4-min hold was performed at a flow rate of 0.1 mL/min. The total method time was 38 min. MS is on from 4 to 30 min.
   b. MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 100–1700 m/z.

C) 1-61% B over 15 min, Zorbax C3 column (6550)
    a. LC conditions: Zorbax 300SB-C3 column: 2.1 × 150 mm, 5 µm, column temperature: 40 °C, gradient: 0-2 min 1% B, 2-11 min 1-61% B, 11-12 min 61-95% B, flow rate: 0.8 mL/min. A final 3-min hold was performed at a flow rate of 0.8 mL/min. The total method time was 15 min. MS is on from 4 to 12 min.
    b. MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 100–1700 m/z.

D) 1-91% B over 20 min, Phenomenex Jupiter C4 column (6550)
    a. LC conditions: Phenomenex Jupiter C4 column: 1.0 × 150 mm, 5 µm, column temperature: 40 °C, gradient: 0-2 min 1% B, 2-18 min 1-91% B, 18-21 min 91% B; flow rate: 0.1 mL/min. A final 4-min hold was performed at a flow rate of 0.1 mL/min. The total method time was 25 min. MS is on from 4 to 18 min.
    b. MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 100–1700 m/z.

E) 1-61% B over 17 min, Luna C18 column (6550)
    a. LC conditions: Phenomenex Luna C18 column: 0.5 × 150 mm, 5 µm, column temperature: 40 °C, gradient: 0-2 min 1% B, 2-14 min 1-61% B, 14-18 min 61-91% B; flow rate: 0.1 mL/min. A final 5-min hold was performed at a flow rate of 0.1 mL/min. The total method time was 23 min. MS is on from 4 to 14 min.
    b. MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 100–1700 m/z.

Data were processed using Agilent MassHunter Workstation Qualitative Analysis Version B.06.00 with BioConfirm Software.

2.7.6 Analytical high-performance liquid chromatography (HPLC) protocol

For determination of purity by HPLC, the filtered peptide solution was diluted in 50% acetonitrile in water with 0.1% TFA (100 µL) to a final concentration of approximately 1.0 mg/mL.

For standard analysis of all peptide and protein sequence samples, analytical HPLC spectra were recorded on an analytical Agilent Zorbax 300SB-C3 column (2.1 mm × 150 mm, 5-µm particle size). A linear gradient of acetonitrile with a 0.08% TFA additive (solvent B) in water with a 0.1% TFA additive (solvent A) was used. After a 3-min hold, gradients of 1% B per minute ramped up over 60 min at a flow rate of 0.4 mL/min. Gradients either started at 1% B (annotated

as “1–61% B over 60 min”) or 5% B (annotated as “5–65% B over 60 min”). A final 3-min hold was performed. The total method time was 66 min. Crude HPLC purities were determined by manual integration of all signals in the area of 5−60 min.

2.7.7 Treatment of individual amino acids and other stock solutions before renovations

Each amino acid is coupled with a select coupling agent, initial conditions are as follows:

Amino acid stocks: 0.2 M in NMP

Activator stocks: 0.19 M in DMF

Base: DIEA (Neat)

Wash steps: amine-free DMF

HATU coupled amino acids: A, D, E, F, G, H, I, K, L, M, N, P, Q, S, T, V, W, Y

PyAOP coupled amino acids: R

Deprotection solution: 40% piperidine in DMF

Tempurature: 90 ℃ in reactor, 80 ℃ for all amino acids

Flow rate: 80 mL/min

2.7.8 Treatment of individual amino acids and other stock solutions after renovations

Each amino acid is coupled with a select coupling agent, more information about why these changes is available in chapter 3. Initial conditions are as follows:

Amino acid stocks: 0.4 M in amine-free DMF

Activator stocks: 0.38 M in amin free DMF

Base: DIEA (Neat)

Wash steps: amine-free DMF

HATU coupled amino acids: D, E, F, G, H, I, K, L, M, P, S, W, Y

PyAOP coupled amino acids: A, N, Q, R, T, V

Deprotection solution: 40% piperidine in DMF

Tempurature: 90 ℃ in reactor, 80 ℃ for all amino acids

Flow rate: 80 mL/min

2.7.6 Analytical high-performance liquid chromatography (HPLC) protocol

For determination of purity by HPLC, the filtered peptide solution was diluted in 50% acetonitrile in water with 0.1% TFA (100 μL) to a final concentration of approximately 1.0 mg/mL.

For standard analysis of all peptide and protein sequence samples, analytical HPLC spectra were recorded on an analytical Agilent Zorbax 300SB-C3 column (2.1 mm × 150 mm, 5-µm particle size). A linear gradient of acetonitrile with a 0.08% TFA additive (solvent B) in water with a 0.1% TFA additive (solvent A) was used. After a 3-min hold, gradients of 1% B per minute ramped up over 60 min at a flow rate of 0.4 mL/min. Gradients either started at 1% B (annotated as "1–61% B over 60 min") or 5% B (annotated as "5–65% B over 60 min"). A final 3-min hold was performed. The total method time was 66 min. Crude HPLC purities were determined by manual integration of all signals in the area of 5−60 min.

2.7.7 Treatment of individual amino acids and other stock solutions before renovations

Each amino acid is coupled with a select coupling agent, initial conditions are as follows:

Amino acid stocks: 0.2 M in NMP

Activator stocks: 0.19 M in DMF

Base: DIEA (Neat)

Wash steps: amine-free DMF

HATU coupled amino acids: A, D, E, F, G, H, I, K, L, M, N, P, Q, S, T, V, W, Y

PyAOP coupled amino acids: R

Deprotection solution: 40% piperidine in DMF

Tempurature: 90 ℃ in reactor, 80 ℃ for all amino acids

Flow rate: 80 mL/min

2.7.8 Treatment of individual amino acids and other stock solutions after renovations

Each amino acid is coupled with a select coupling agent, more information about why these changes is available in chapter 3. Initial conditions are as follows:

Amino acid stocks: 0.4 M in amine-free DMF

Activator stocks: 0.38 M in amin free DMF

Base: DIEA (Neat)

Wash steps: amine-free DMF

HATU coupled amino acids: D, E, F, G, H, I, K, L, M, P, S, W, Y

PyAOP coupled amino acids: A, N, Q, R, T, V

2.8 Acknowledgements

Individual Contributions: Renovation goals and peptide synthesis were performed and conceptualized by Amanda E Cowfer and Sarah Antilla who equally contributed to data collection and processing. Alex E Callahan also assisted with mechanical design and installation ideas and changes. Nina Hartrampf provided invaluable mentorship and guidance in the renovation process and was key to our initial understanding of the challenges to be undertaken. Ryan M. Beaumont was responsible for the final integration of all software and hardware changes into LabView, and further trained Amanda E Cowfer in LabView coding and troubleshooting. A massive thank you to Rebecca L Holden, Stephanine Hanna, and Makenzie Poskus for helpful conversations and assistance throughout this process. Finally, we thank Bradley L Pentelute for allowing us to work in his research laboratory.

Funding: Financial support for this project was provided by Novo Nordisk. Amanda E Cowfer gratefully acknowledges support from the National Science Foundation Graduate Research Fellowship under grant no. 1122374, and is additionally supported by an MIT Dean of Science Fellowship.

2.9 References

1. Ahmed, N. Peptide Bond Formations through Flow Chemistry. *Chem. Biol. Drug Des.* **2018**, *91* (2), 647–650. https://doi.org/10.1111/cbdd.13115.

2. *Flow Chemistry in Organic Synthesis*, 1st ed.; Jamison, T. F., Koch, G., Eds.; Georg Thieme Verlag KG: Stuttgart, 2018; p b-006-161272. https://doi.org/10.1055/b-006-161272.

3. Al-Warhi, T. I.; Al-Hazimi, H. M. A.; El-Faham, A. Recent Development in Peptide Coupling Reagents. *J. Saudi Chem. Soc.* **2012**, *16* (2), 97–116. https://doi.org/10.1016/j.jscs.2010.12.006.

4. Lukas, T. J.; Prystowsky, M. B.; Erickson, B. W. Solid-Phase Peptide Synthesis under Continuous-Flow Conditions. *Proc. Natl. Acad. Sci. U. S. A.* **1981**, *78* (5), 2791–2795. https://doi.org/10.1073/pnas.78.5.2791.

5. Eissler, S.; Kley, M.; Bächle, D.; Loidl, G.; Meier, T.; Samson, D. Substitution Determination of Fmoc-Substituted Resins at Different Wavelengths: Substitution Determination of Fmoc-Substituted Resins. *J. Pept. Sci.* **2017**, *23* (10), 757–762. https://doi.org/10.1002/psc.3021.

6. Mijalis, A. J.; Thomas, D. A.; Simon, M. D.; Adamo, A.; Beaumont, R.; Jensen, K. F.; Pentelute, B. L. A Fully Automated Flow-Based Approach for Accelerated Peptide Synthesis. *Nat. Chem. Biol.* **2017**, *13* (5), 464–466. https://doi.org/10.1038/nchembio.2318.

7. Truex, N. L.; Holden, R. L.; Wang, B.-Y.; Chen, P.-G.; Hanna, S.; Hu, Z.; Shetty, K.; Olive, O.; Neuberg, D.; Hacohen, N.; Keskin, D. B.; Ott, P. A.; Wu, C. J.; Pentelute, B. L. Automated Flow Synthesis of Tumor Neoantigen Peptides for Personalized Immunotherapy. *Sci. Rep.* **2020**, *10* (1), 723. https://doi.org/10.1038/s41598-019-56943-5.

8. Pomplun, S.; Jbara, M.; Schissel, C. K.; Wilson Hawken, S.; Boija, A.; Li, C.; Klein, I.; Pentelute, B. L. Parallel Automated Flow Synthesis of Covalent Protein Complexes That Can Inhibit MYC-Driven Transcription. *ACS Cent. Sci.* **2021**, *7* (8), 1408–1418. https://doi.org/10.1021/acscentsci.1c00663.

9. Assadi, A. A.; Fries, C.; Fechter, M.; Maschler, B.; Ewert, D.; Schnauffer, H.-G.; Zürn, M.; Reichenbach, M. User-Friendly, Requirement Based Assistance for Production Workforce Using an Asset Administration Shell Design. *Procedia CIRP* **2020**, *91*, 402–406. https://doi.org/10.1016/j.procir.2020.02.192.

10. Chamas, A.; Moon, H.; Zheng, J.; Qiu, Y.; Tabassum, T.; Jang, J. H.; Abu-Omar, M.; Scott, S. L.; Suh, S. Degradation Rates of Plastics in the Environment. *ACS Sustain. Chem. Eng.* **2020**, *8* (9), 3494–3511. https://doi.org/10.1021/acssuschemeng.9b06635.

11. Liang, X.; Wu, P.; Lan, L.; Wang, Y.; Ning, Y.; Wang, Y.; Qin, Y. Effect of Polytetrafluoroethylene (PTFE) Content on the Properties of Ni-Cu-P-PTFE Composite Coatings. *Materials* **2023**, *16* (5), 1966. https://doi.org/10.3390/ma16051966.

12. Lojen, D.; Zaplotnik, R.; Primc, G.; Mozetič, M.; Vesel, A. Optimization of Surface Wettability of Polytetrafluoroethylene (PTFE) by Precise Dosing of Oxygen Atoms. *Appl. Surf. Sci.* **2022**, *598*, 153817. https://doi.org/10.1016/j.apsusc.2022.153817.

13. Kalkman, C. J. LabVIEW: A Software System for Data Acquisition, Data Analysis, and Instrument Control. *J. Clin. Monit.* **1995**, *11* (1), 51–58. https://doi.org/10.1007/BF01627421.

14. Korgin, A.; Ermakov, V.; Kilani, L. Z. Automation and Processing Test Data with LabVIEW Software. *IOP Conf. Ser. Mater. Sci. Eng.* **2019**, *661* (1), 012073. https://doi.org/10.1088/1757-899X/661/1/012073.

15. Ramasamy, P.; Tharanyaa Jothimani Palanivelu, S.; Sathesan, A. Certain Applications of LabVIEW in the Field of Electronics and Communication. In *LabVIEW - A Flexible Environment for Modeling and Daily Laboratory Use*; De Asmundis, R., Ed.; IntechOpen, 2021. https://doi.org/10.5772/intechopen.96301.

16. Gorman, M. Resource and Reagent Management in a Clinical Diagnostic Laboratory Setting. *Am. J. Clin. Pathol.* **2015**, *144* (suppl 2), A191–A191. https://doi.org/10.1093/ajcp/144.suppl2.191.

17. Pietrobon, R.; Guller, U.; Martins, H.; Menezes, A. P.; Higgins, L. D.; Jacobs, D. O. A Suite of Web Applications to Streamline the Interdisciplinary Collaboration in Secondary Data Analyses. *BMC Med. Res. Methodol.* **2004**, *4* (1), 29. https://doi.org/10.1186/1471-2288-4-29.

18. Amblard, M.; Fehrentz, J.-A.; Martinez, J.; Subra, G. Methods and Protocols of Modern Solid Phase Peptide Synthesis. *Mol. Biotechnol.* **2006**, *33* (3), 239–254. https://doi.org/10.1385/MB:33:3:239.

19. Plutschack, M. B.; Pieber, B.; Gilmore, K.; Seeberger, P. H. The Hitchhiker's Guide to Flow Chemistry. *Chem. Rev.* **2017**, *117* (18), 11796–11893. https://doi.org/10.1021/acs.chemrev.7b00183.

20. Grillo, G.; Cintas, P.; Colia, M.; Calcio Gaudino, E.; Cravotto, G. Process Intensification in Continuous Flow Organic Synthesis with Enabling and Hybrid Technologies. *Front. Chem. Eng.* **2022**, *4*, 966451. https://doi.org/10.3389/fceng.2022.966451.

21. Ruzicka, Jaromir.; Marshall, G. D.; Christian, G. D. Variable Flow Rates and a Sinusoidal Flow Pump for Flow Injection Analysis. *Anal. Chem.* **1990**, *62* (17), 1861–1866. https://doi.org/10.1021/ac00216a024.

22. Gugulothu, S. K.; Kumar, P. R. V.; Deekshith, P. Exploring Cavitating Phenomenon With and Without Ultrasonic Transducer. *Procedia Eng.* **2012**, *38*, 154–164. https://doi.org/10.1016/j.proeng.2012.06.022.

23. Arbour, C. A.; Stockdill, J. L. A Mild Capping Method for SPPS on the N-Methyl Diaminobenzoyl Linker: Synthesis of an N-Acyl Urea Appended C. Elegans Neuropeptide. *Tetrahedron Lett.* **2018**, *59* (44), 3903–3906. https://doi.org/10.1016/j.tetlet.2018.08.061.

24. Bertini, S.; Saccomanni, G.; Carlo, S. D.; Digiacomo, M.; Gargini, C.; Piano, I.; Campisi, G. M.; Ghidoni, R.; Macchia, M.; Manera, C. Application of An Improved HPLC-FL Method to Screen Serine Palmitoyl Transferase Inhibitors. *Mol. Basel Switz.* **2017**, *22* (7), 1198. https://doi.org/10.3390/molecules22071198.

25. Richter, M. J.; Harutyunova, S.; Bollmann, T.; Classen, S.; Fuge, J.; Gall, H.; Gerhardt, F.; Ghofrani, H. A.; Gunkel, H.; Grünig, E.; Halank, M.; Heine, A.; Klose, H.; Lange, T. J.; Neurohr, C.; Nickolaus, K.; Opitz, C. F.; Rosenkranz, S.; Seyfarth, H.; Tello, K.; Ewert, R.; Olsson, K. M. Flow Rate Variance of a Fully Implantable Pump for the Delivery of Intravenous Treprostinil in Pulmonary Arterial Hypertension. *Pulm. Circ.* **2020**, *10* (1), 1–5. https://doi.org/10.1177/2045894020910136.

26. Müller, T. D.; Finan, B.; Bloom, S. R.; D'Alessio, D.; Drucker, D. J.; Flatt, P. R.; Fritsche, A.; Gribble, F.; Grill, H. J.; Habener, J. F.; Holst, J. J.; Langhans, W.; Meier, J. J.; Nauck, M. A.; Perez-Tilve, D.; Pocai, A.; Reimann, F.; Sandoval, D. A.; Schwartz, T. W.; Seeley, R. J.; Stemmer, K.; Tang-Christensen, M.; Woods, S. C.; DiMarchi, R. D.; Tschöp, M. H. Glucagon-like Peptide 1 (GLP-1). *Mol. Metab.* **2019**, *30*, 72–130. https://doi.org/10.1016/j.molmet.2019.09.010.

27. Wilding, J. P. H.; Batterham, R. L.; Calanna, S.; Davies, M.; Van Gaal, L. F.; Lingvay, I.; McGowan, B. M.; Rosenstock, J.; Tran, M. T. D.; Wadden, T. A.; Wharton, S.; Yokote, K.; Zeuthen, N.; Kushner, R. F. Once-Weekly Semaglutide in Adults with Overweight or Obesity. *N. Engl. J. Med.* **2021**, *384* (11), 989–1002. https://doi.org/10.1056/NEJMoa2032183.

28. Rekasi, Z.; Varga, J. L.; Schally, A. V.; Halmos, G.; Groot, K.; Czompoly, T. Antagonistic Actions of Analogs Related to Growth Hormone-Releasing Hormone (GHRH) on Receptors for GHRH and Vasoactive Intestinal Peptide on Rat Pituitary and Pineal Cells *in Vitro*. *Proc. Natl. Acad. Sci.* **2000**, *97* (3), 1218–1223. https://doi.org/10.1073/pnas.97.3.1218.

29. Cai, R.; Zhang, X.; Wang, H.; Cui, T.; Halmos, G.; Sha, W.; He, J.; Popovics, P.; Vidaurre, I.; Zhang, C.; Mirsaeidi, M.; Schally, A. V. Synthesis of Potent Antagonists of Receptors for Growth Hormone-Releasing Hormone with Antitumor and Anti-Inflammatory Activity. *Peptides* **2022**, *150*, 170716. https://doi.org/10.1016/j.peptides.2021.170716.

30. Hartrampf, N.; Saebi, A.; Poskus, M.; Gates, Z. P.; Callahan, A. J.; Cowfer, A. E.; Hanna, S.; Antilla, S.; Schissel, C. K.; Quartararo, A. J.; Ye, X.; Mijalis, A. J.; Simon, M. D.; Loas, A.; Liu, S.; Jessen, C.; Nielsen, T. E.; Pentelute, B. L. Synthesis of Proteins by Automated Flow Chemistry. *Science* **2020**, *368* (6494), 980–987. https://doi.org/10.1126/science.abb2491.

31. Shilova, O.; Kotelnikova, P.; Proshkina, G.; Shramova, E.; Deyev, S. Barnase-Barstar Pair: Contemporary Application in Cancer Research and Nanotechnology. *Molecules* **2021**, *26* (22), 6785. https://doi.org/10.3390/molecules26226785.

32. Szeluga, N.; Baldrich, P.; DelPercio, R.; Meyers, B. C.; Frank, M. H. Introduction of Barnase/Barstar in Soybean Produces a Rescuable Male Sterility System for Hybrid Breeding. *Plant Biotechnol. J.* **2023**, pbi.14155. https://doi.org/10.1111/pbi.14155.

33. Torbeev, V. Yu.; Kent, S. B. H. Convergent Chemical Synthesis and Crystal Structure of a 203 Amino Acid "Covalent Dimer" HIV-1 Protease Enzyme Molecule. *Angew. Chem. Int. Ed.* **2007**, *46* (10), 1667–1670. https://doi.org/10.1002/anie.200604087.

34. Brik, A.; Wong, C.-H. HIV-1 Protease: Mechanism and Drug Discovery. *Org. Biomol. Chem.* **2003**, *1* (1), 5–14. https://doi.org/10.1039/b208248a.

35. Weber, I. T.; Wang, Y.-F.; Harrison, R. W. HIV Protease: Historical Perspective and Current Research. *Viruses* **2021**, *13* (5), 839. https://doi.org/10.3390/v13050839.

36. Ye, X.; Lee, Y.-C.; Gates, Z. P.; Ling, Y.; Mortensen, J. C.; Yang, F.-S.; Lin, Y.-S.; Pentelute, B. L. Binary Combinatorial Scanning Reveals Potent Poly-Alanine-Substituted Inhibitors of Protein-Protein Interactions. *Commun. Chem.* **2022**, *5* (1), 128. https://doi.org/10.1038/s42004-022-00737-w.

37. Thapa, P.; Cabalteja, C. C.; Philips, E. E.; Espiritu, M. J.; Peigneur, S.; Mille, B. G.; Tytgat, J.; Cummins, T. R.; Bingham, J.-P. T-Boc Synthesis of Huwentoxin-i through Native Chemical Ligation Incorporating a Trifluoromethanesulfonic Acid Cleavage Strategy: T-Boc Synthesis of Huwentoxin-I. *Biopolymers* **2016**, *106* (5), 737–745. https://doi.org/10.1002/bip.22887.

Chapter 3: Synthesis of the Proteins by Automated Flow Chemistry

The work presented in this chapter has been reproduced from the following manuscript:

Hartrampf, N.; Saebi, A.; Poskus, M.; Gates, Z. P.; Callahan, A. J.; Cowfer, A. E.; Hanna, S.; Antilla, S.; Schissel, C. K.; Quartararo, A. J.; Ye, X.; Mijalis, A. J.; Simon, M. D.; Loas, A.; Liu, S.; Jessen, C.; Nielsen, T. E.; Pentelute, B. L. Synthesis of Proteins by Automated Flow Chemistry. *Science*, **2020**, 368 (6494), 980–987. https://doi.org/10.1126/science.abb2491.

A significant portion of this chapter was co-written with N. H. and A. S. Further, this chapter reports on most of the contents of this publication, with a particular emphasis on the work with barnase-barstar, aspartimide formation, and epimerization studies, which A.E.C. contributed to extensively.

3.1 Introduction and project motivation

Mechanical pumps, valves, solid supports, and computers have transformed the way we perform chemical reactions in the form of modern flow chemistry. Researchers in small laboratories and large pharmaceutical giants utilize continuous multi-step flow technology to access small molecules ranging from pharmaceutical ingredients to natural products and bulk commodities.[1] One example of the use and development of flow chemistry to provide access to small molecules in the pharmaceutical field is developing and optimizing a flow-based reactor for the Matteson reaction to access a key intermediate to the β-lactamase inhibitor vaborbactam.[2] In this process, the reaction must be kept at a very low temperature (-100 °C), a difficult task in a batch synthesis. This development highlights some of the advantages of flow synthesis over batch methods. Others include integrating in-line spectroscopic monitoring,[3] efficient reagent mixing, and precise control over the reaction parameters, like temperature.[4] In the past, in the Pentelute research group, we saw an opportunity to translate these flow-chemistry capabilities to the total chemical synthesis of peptides and, hopefully, in time, full-length biologically active proteins.[5] To this end, in 2017, Mijalis and coworkers reported the first-generation Automated Fast-Flow Peptide Synthesizer (AFPS) in an effort to provide further rapid access to an expanded chemical peptide space with speed and ease.[6]

The AFPS and similar flow-based peptide synthesizers,[7] are gaining traction due to their advantageous features, such as control over physical parameters and significantly reduced side-product formation.[8] As introduced in Chapter 1, after the introduction of SPPS, researchers found, as early as the 1970s,[9] that automation and high fidelity of peptide synthesis could be achieved by containing the solid support, resin, in a reactor and flowing the reagents for peptide elongation over the fixed resin bed.[10] Some of these early flow systems were quite complex and had long synthesis times. Conversely, the AFPS system,147, uses high-performance liquid chromatography (HPLC) pumps to deliver and then wash away reagents continuously and incorporates amino acid residues in as little as 40 seconds at temperatures up to 90 °C.[6] However, even with these advances, the potential of flow chemistry[11] to enable the synthesis of peptide chains in the range of single-domain proteins has not been fully realized.[12,13] In this chapter, I report on the efforts of a large team of researchers, including myself, in the Pentelute group that set out to optimize our AFPS technology to meet this challenge.

Before this work, SPPS techniques were considered limited to peptide chains to no more than 50 amino acids, restricting access to much-needed full-length protein polypeptide chains by synthetic means.[14] Traditionally, to access long peptides and full-length proteins, many research and industrial groups turn to biological expression techniques, which have their own limitations. For instance, biological expression is generally limited to the naturally occurring amino acids,[15]. Even with advances in genetic code expansion, researchers have difficulty incorporating long,

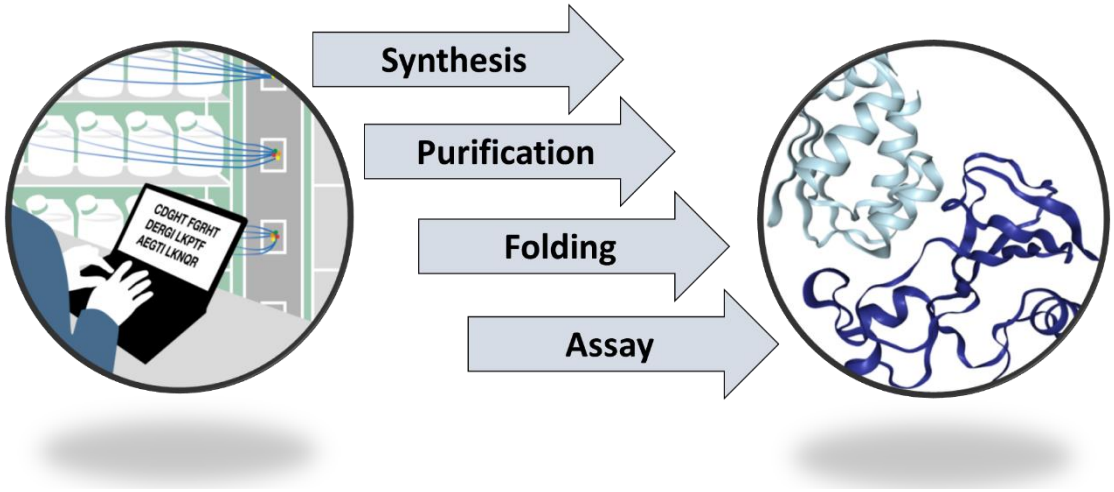


Figure 3.1: Protein production via SPPS workflow

complex polypeptides that include multiple unnatural residues.[16] With SPPS, researchers can access some long, complex polypeptides utilizing native chemical ligation (NCL) techniques. NCL combines smaller peptide building blocks by reacting the C-terminal peptide thioester with an N-terminal cysteinyl peptide to produce a native peptide bond,[17] and this gives access to significantly longer polypeptide chains and synthetic proteins. However, NCL is a rather tedious and time-consuming process as it is challenging to develop the peptide sequences for ligation and optimize the reaction itself. Nevertheless, NCL techniques continue to be a useful synthetic tool. Thus, we envisioned a fusion of the flexibility of SPPS and the reach of biological expression to allow access to long, complex polypeptides and to incorporate multiple unnatural amino acids in these long sequences.

Herein, we report a routine and rapid protocol allowing stepwise total chemical synthesis of peptide chains >50 amino acids in length at ~2.5 minutes per amino acid coupling cycle. The optimized protocol was built by collecting extensive analytical data from the AFPS system and delivering products with high fidelity and high chiral purity. Using this protocol, single-domain protein chains ranging from barstar (90 aa) to sortase $A_{59\text{-}206}$ (sortase A*, 164 aa) were synthesized in 3.5-6.5 hours. Flowing synthesis, these sequences were folded and evaluated for their

biophysical properties and enzymatic activities and compared to their recombinant counterparts. The timescale of this chemical protein synthesis is on par and, in some cases, more rapid than that of recombinant expression and, therefore, offers a practical alternative to biochemical methods while allowing access to chemical space beyond the canonical amino acids.

3.2 Instrument optimization

The first step on our journey to develop the AFPS to access single-domain protein polypeptide sequences was to access our current synthetic capabilities. To do this, we utilized a selection of "test" peptides to screen the effects of changing certain synthesis conditions on the AFPS. These peptides were also briefly mentioned in chapter 2. These peptides were glucagon-like peptide-1 (GLP-1), growth hormone-releasing hormone (GHRH), and a peptide known as NN92. GLP-1 plays an important role in the secretion of insulin appetite control and can decrease inflammation and the chance of cell apoptosis in certain circumstances.[18] GLP-1 and its derivatives are currently being successfully used to treat type-2 diabetes. As a weight loss agent.[19] GHRH, on the other hand, is released from the hypothalamus and, in turn, stimulates the release of growth hormones from the pituitary gland.[20] Interestingly, GHRH was actually first reported in cancer tissues and has been investigated as a therapeutic target in cancer research.[21]

The peptide NN92 demonstrated massive amounts of aspartimide formation when we attempted to produce it for a pharmaceutical collaborator.

Peptide sequences used for screening purposes:	
GLP-1:	HAEGTFTSDV SSYLEGQAAK EFIAWLVKGR (30 AA)
GHRH:	YADAI FTNSY RKVLG QLSAR KLLQD ILSA (29 AA)
NN92:	RVVVGEHNLS QNDGTEQYVN VQKIVSHPY (29 AA)

Table 3.1: Peptide sequences used in screening efforts

To begin the optimization process, we investigated coupling efficiency before turning to examining side reactions induced by these new coupling conditions. As mentioned in Chapter 1, the AFPS functions by mixing reagents, which are heated and delivered to a reactor containing the resin solid support using three Varian Prostar HPLC pumps. Since both the activators and Fmoc blocking group are ultraviolet-visible, we can use a UV-Vis detection module to monitor the reaction progress. Conveniently, by analyzing the data produced via UV-vis, we can better understand the efficiency of each coupling step.

As previously discussed in Chapter 2, an essential step in optimizing our general protocol involved assessing a multitude of parameters, including flow rate, reaction solvent, reagent

concentration, temperature, and coupling agents. To this end, our pump refill time experiments discussed in-depth in Chapter 2, were key in understanding the optimal flow rate needed to ensure high yields for each amino acid coupling by controlling for and optimizing the reagent mixing conditions. Other key modifications to our protocol included increasing reagent concentrations to 0.4 M,[22] using amine-free *N*, *N*-dimethylformamide (DMF), and an increase in temperature to 85–90 °C for reagent activation and coupling. To reach these conclusions, we iteratively synthesized and evaluated our test peptides' crude yield and purity, GLP-1, and GHRH.

3.2.1 Investigation of temperature

The first condition tested was temperature. We utilized the AFPS conditions currently in use on the machine, with amino acids stored in NMP at 0.2 M concentration, activators at 0.19 M concentration, and finally, our deprotection base was 40% piperidine. The temperatures in the reactor holding resin and the pre-heating loop was varied from 70-90 ℃. The resulting peptides yield and purity were analyzed via HPLC, as described in the materials and methods section. We

Entry	Peptide	T (loop)	T (reactor)	Purity by HPLC	Crude Yield
1	GLP-1	70 °C	70 °C	**46%**	17%
2	GLP-1	70 °C	90 °C	**52%**	23%
3	GLP-1	90 °C	70 °C	**53%**	25%
4	GLP-1	90 °C	90 °C	**60%**	32%
5	GHRH	70 °C	90 °C	**83%**	60%
6	GHRG	90 °C	70 °C	**78%**	43%
7	GHRH	90 °C	90 °C	**89%**	50%

Table 3.2: Synthesis results from temperature variation studies

Higher temperatures, 90 ℃, lead to increased coupling efficiency and less deletion by-products

concluded that increasing the temperature to 90 ℃ leads to a notable increase in coupling efficiency and fewer deletion by-products. At the same time, we envision further increases could lead to further improvements. However, the melting temperature of our disposable reactor cartridges limited the temperature ranges we could test.

3.2.2 Investigation of solvents

Next, it was valuable to analyze the effect of different solvents and reagent concentrations on coupling outcomes. To do this, we first assessed the maximum concentration at which the amino acids would remain stable over a 2-3 week period, and we discovered that in DMF and NMP, the amino acids could be stored at 0.4 M for weeks. Next, we compared the original NMP solvent to

DMF; interestingly, NMP performed better than DMF repeatedly. However, DMF is known for degrading into free amines[23] which can have a significant impact on synthesis quality. We saw massive improvements in synthetic outcomes when we utilized amine-free DMF (AF-DMF), produced by adding AldraAmine trapping sticks to our stock solvent bottles.

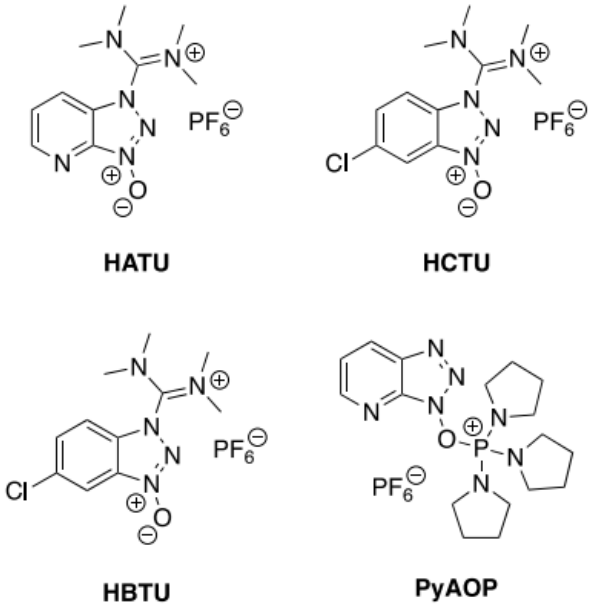
Entry	peptide	Solvent	Amino acid conc.	HPLC purity	Crude yield
1	GLP-1	DMF	0.4M	**69%**	36%
2	GLP-1	DMF	0.2M	**58%**	_[3]
3	GLP-1	NMP	0.4M	**65%**	21%
4	GLP-1	NMP	0.2M	**53%**	21%
5	GLP-1	amine-free DMF	0.4M	**71%**	13%
6	GLP-1	amine-free DMF	0.2M	**55%**	23%

Table 3.3: Solvent and reagent concentration study results

An increase of the concentration of amino acids, 0.4 M. and activators, 0.38 M, and the change to amine-free DMF as a solvent significantly improves synthesis outcomes.

3.2.3 Investigation of Activators

Finally, the performance of different activators for the coupling step was also investigated. We assessed four coupling agents, otherwise known as activators, including HATU, HBTU,



Entry	Peptide	Coupling agent	Purity by HPLC	Crude Yield
1	GLP-1	HATU [a)]	71%	24%
2	GLP-1	HBTU [a)]	66%	20%
3	GLP-1	HCTU [a)]	69%	26%
4	GLP-1	PyAOP	76%	22%

***Table 3.4**:* Results of activator screening

PyAOP and HATU as activators provide the highest purity and will be used to couple aminos

HCTU, and PyAOP. We identified the azabenzotriazol-reagents PyAOP and HATU as optimal, with PyAOP consistently giving the best synthetic outcomes. However, PyAOP is prohibitively expensive, so we also screened which of these two activators would be best for coupling each amino acid. We also learned that some amino acids would benefit from longer coupling times to improve per-coupling synthetic yields. We now had a general amino-acid-specific recipe that we developed from the iterative tests and analytical comparison of the products obtained for GLP-1.

3.2.1 Addressing aspartimide formation

Another major side-product that has inhibited our access to full-length protein polypeptide sequences, such as DG-rich collagen, is aspartimide formation, which often results in both α-Asp and the unnatural β-Asp formation. Aspartimide formation is a well-known phenomenon occurring

mainly in -DG-, -DS-, -DA-, -DD-, -DQ- rich sequences both in vitro and in vivo, and in SPSS, is particularly prevalent due to the presence of a piperidine, a strong base, during the deprotection steps of SPSS.[24] To mitigate aspartimide formation, other research groups have utilized bulky protecting groups for aspartic acid and backbone-protected glycine residues to prevent the nucleophilic attack of Asp and subsequent formation of aspartimide.[25] As increased temperature leads to more aspartimide formation, various deprotection bases, additives[26] and aspartic acid protecting groups were screened to minimize this unwanted side reaction.[27]

To do this, we utilized the test peptide NN92, which produced nearly 85% aspartimide by-product under the previous AFPS conditions. In our investigation, we found that milder deprotection bases (i.e., piperazine and HOBt/piperidine) and bulky aspartic acid protecting groups

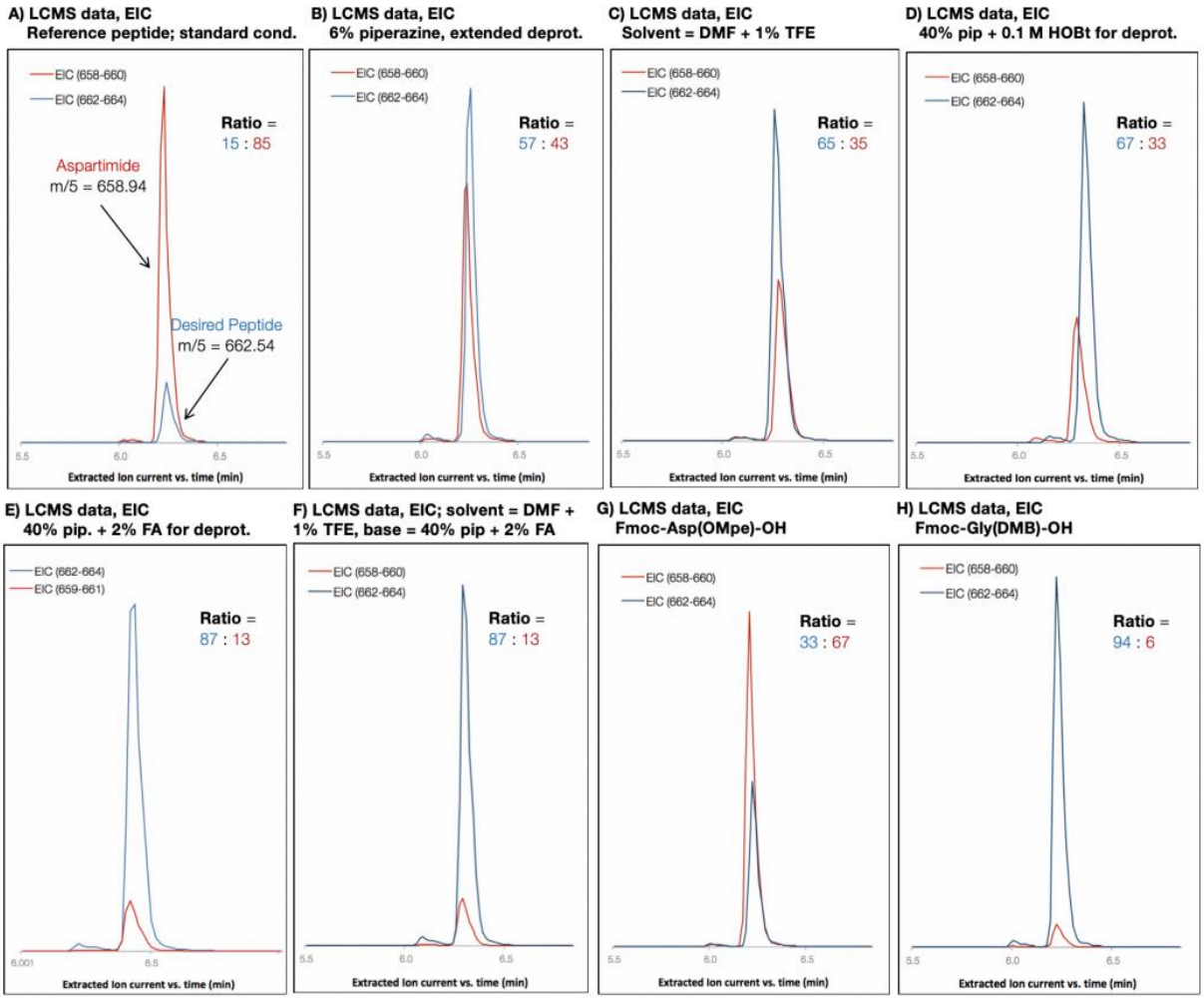


Figure 3.2: Results of solvent and protect group screens to suppress aspartimide formation
With all tested conditions we saw suppression of aspartimide formation, with the best results achieved with formic acid additive and Fmoc-Gly(DMB)-OH. We now used 2% formic acid in 40% piperidine for all deprotections.

(i.e., 3-methyl-pent-3-yl esters, OMpe) decreased the level of aspartimide formation. However, the most effective strategies were the addition of formic acid as a piperidine additive and backbone protection with dimethoxybenzyl glycine (Dmb-Gly). Formic acid (2% in 20:80 v/v piperidine: DMF) is now the standard additive for deprotection, and backbone protection can be applied on an as-needed basis. In section 3.6, we will introduce some further ideas on how we could integrate new chemistries into the AFPS synthesis protocol to suppress aspartimide formation further.

3.2.1 Addressing cystine and histidine epimerization

Finally, we needed to address the fact that certain amino acids, namely cysteine and histidine, are at a high risk of epimerization, particularly at higher temperatures. In previous reports about the AFPS, researchers assessed the effect of flow rate on epimerization.[6] For our efforts, we

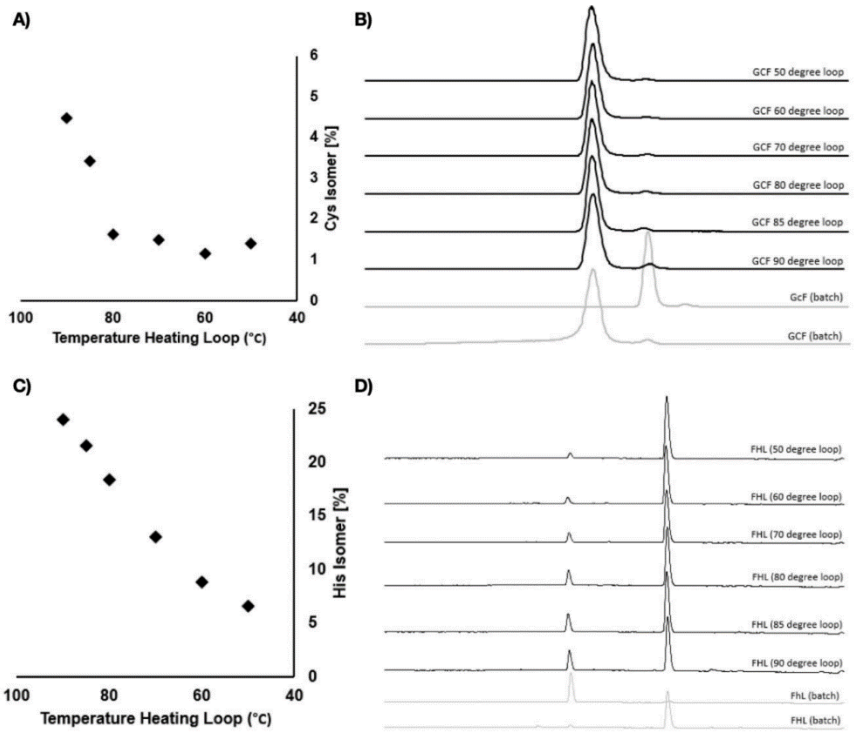


Figure 3.3: Results of 10-foot preheating loop screens with Cys(Trt) and His(Trt)

Cysteine and Histidine both demonstrated high levels of epimerization at higher temperatures, in the case of cysteine we begin to see a leveling off after 60 °C.

evaluated the effect of varied temperatures, activation agents, and side-chain protecting groups on epimerization.[28] To perform these experiments, we used the test peptides GCF, GcF, FHL, and FhL test peptides with lowercase letters representing D amino acids. As a baseline to compare, we prepared all these peptides using standard batch SPPS and then analyzed them via analytical HPLC, as described in the materials and methods section. These peptides were then iteratively

synthesized on the AFPS under different conditions and analyzed via HPCL to compare to our batch prepared samples.

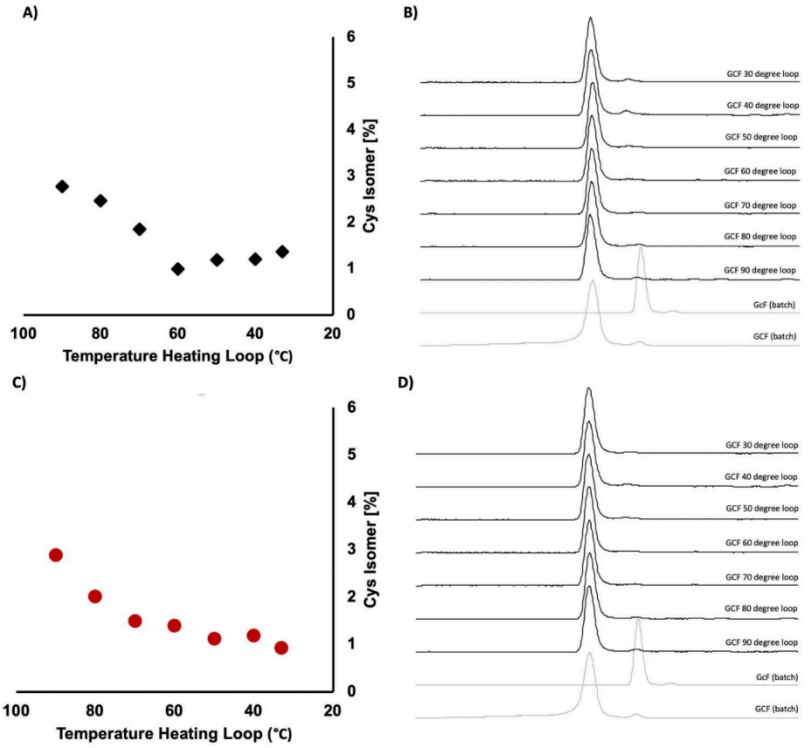


Figure 3.4: Effect of different activators on epimerization of cysteine in a 5-foot pre-heating loop

HATU and PyAOP as an activator for cysteine coupling resulted in similar levels of epimerization, with consistent amount of the D-isomer observed at coupling below 60 ℃. Red = PyAOP, black = HATU

For both amino acids, epimerization increases with activation time and temperature. Our first tests utilized the longer 10-foot pre-heating loop, and we observed that cysteine experienced a steady rate of isomerization after pre-heating loop temperatures dropped below 80 ℃. Histidine demonstrated continued drops in isomerization as the temperature decreased below 40 ℃. Approximately 4.5–5.0% D-isomer of cysteine and ~25% D-isomer of histidine was observed under standard coupling conditions at 90 °C.

With this understanding in hand, we decided to move cysteine and histidine couplings to a shorter 5-foot heating loop. We once again screened a range of temperatures, this time coupling these amino acids at temperatures as low as 33 ℃. Furthermore, we continued to couple all other amino acids other than histidine and cysteine in our 10-foot pre-heating loop set to 90 ℃. We also assessed the effect of different activators on epimerization at varying temperatures. In the case of cysteine, HATU, and PyAOP usage resulted in a similar amount of observed epimerization, and

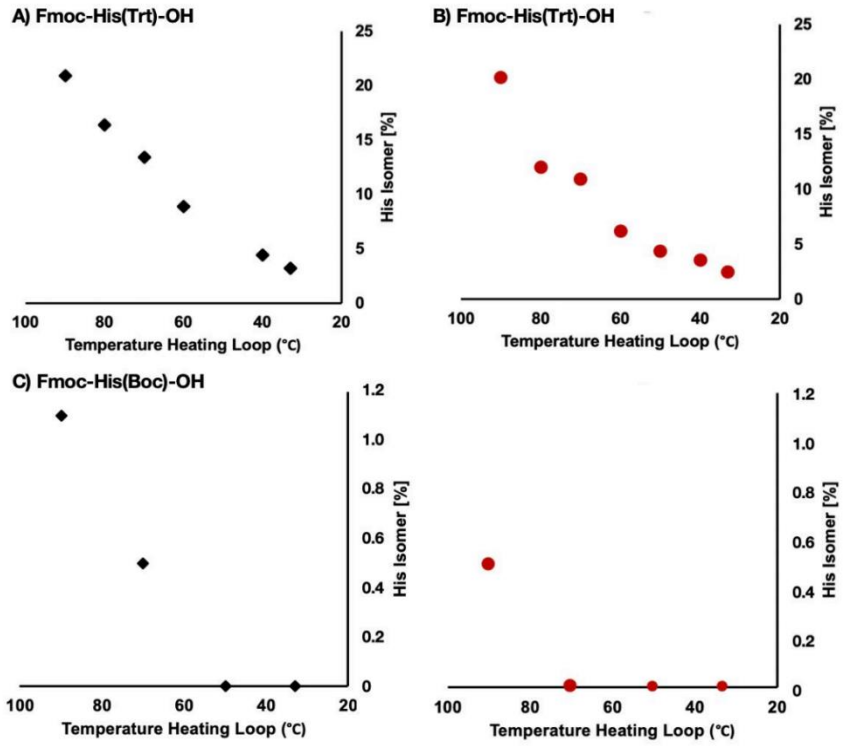


Figure 3.5: Effect of activators on histidine couplings with different protecting groups

HATU and PyAOP as an activator for histidine resulted in similar levels of epimerization with His(Trt), but with His(boc) PyAOP resulted in nearly no detectible isomerization below 60 °C. Red =- PyAOP, black = HATU

levels of epimerization did not further decrease below a coupling temperature of 60 °C, with 1.5 % D-isomer for cysteine observed.

Next, we turned our attention back to histidine; due to the extremely high levels of epimerization we observed with Fmoc-His(Trt)-OH in our previous experiments, we wanted to not only assess the relationship between activator and temperature used to couple these amino acids but assess using Fmoc-His(Boc)-OH as an alternative. Literature suggests that His(Boc) should offer some protection against epimerization, and that is what we observed.[29] Excitingly, with Boc as the protecting group for histidine, we notice nearly a complete elimination of detectible D-isomer, <0.5% observed at temperatures below 60 °C.

Thus, we decided that coupling cystine and histidine at 60 °C would provide the best balance of high per-coupling yields with low isomerization. Moreover, the amount of D-isomer did not change significantly over 100 amino acid couplings, indicating that epimerization of cysteine and histidine only occurs during the activation step. Implementation of these conditions allowed us to solidify the general AFPS protocol, which was then applied to the production of sequences exceeding 50 amino acids.

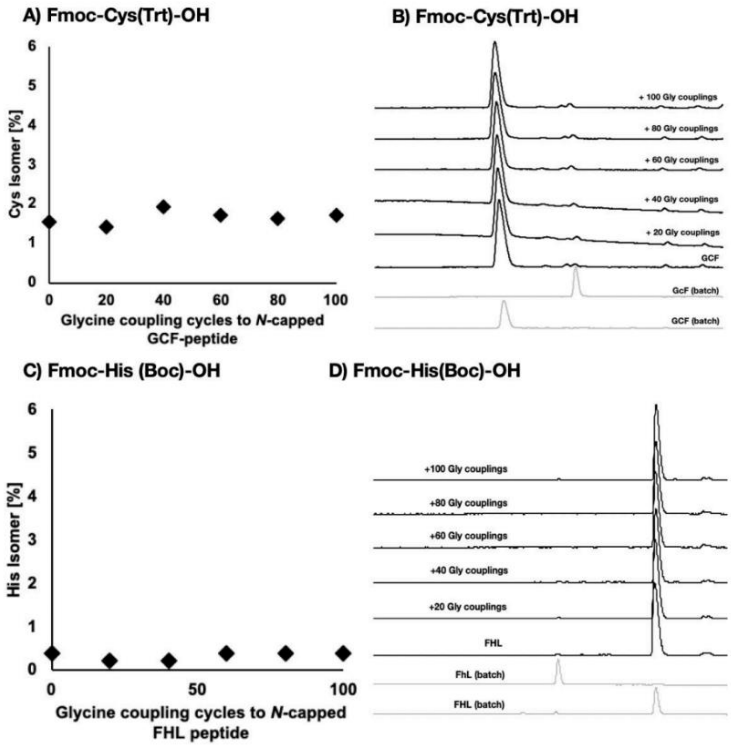


Figure 3.6: Results of continued coupling cycles on epimerization

Exposure of capped test peptides did not result in any notable change in epimerization of Cys(Trt) or His(Boc), meaning coupling both of these aminos at 60 ℃ is best moving forward.

3.3 Synthetic outcomes form the optimized protocol

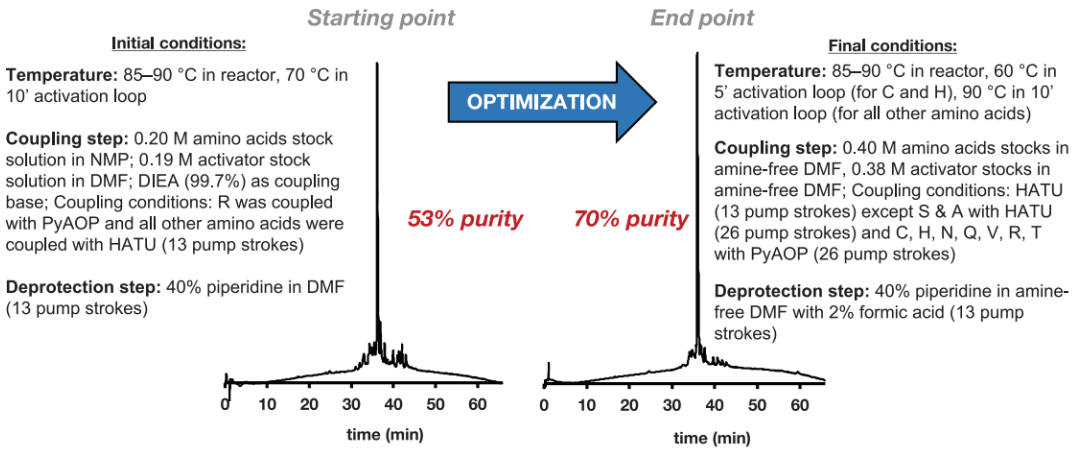


Figure 3.7: Summary of optimized AFPS protocol

With the arduous task of optimizing the AFPS protocol, we can now focus on the production of long polypeptide chains. We have seen vast improvements in the synthetic quality of our test peptides, GLP-1, GHRH, and NN92, and we applied the techniques used to produce high-quality pure short peptides to longer sequences.

3.3.1 Synthetic results from test peptides

As discussed above, we used key test peptides to assess and optimize the AFPS protocol

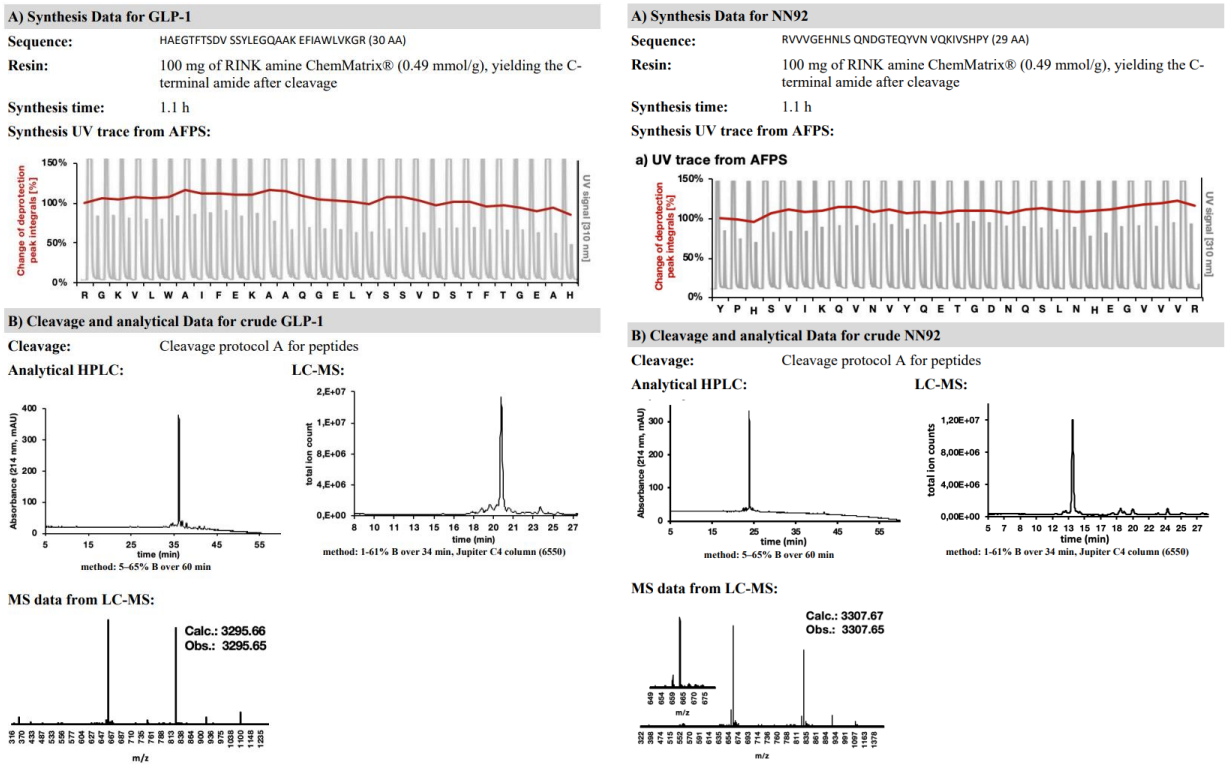


Figure 3.8: Synthetic results for GLP-1 (right) and NN92 (Left) following AFPS optimization

iteratively. The synthetic results of two test peptides, GLP-1 and NN92, are provided below. With this data, we provide the UV-vis trace with the change of deprotection peak integrals, and these integrals indicate the synthetic success of each coupling, and steep drops indicate possible peptide aggregation during synthesis. Both of these peptides were produced in good yields and purity in under 1.5 hours.

3.3.2 Synthetic results from AFPS vs traditional SPPS

Another critical step for us to demonstrate that the optimized AFPS protocol outperforms traditional synthesis methods was to compare the synthetic results of our AFPS system to commercially available microwave synthesis. Thanks to the generosity of our pharmaceutical collaborators, Novo Nordisk, we obtained samples of two proteins produced via microwave synthesis. In this comparison, AFPS and standard batch SPPS protein polypeptide sequences were synthesized at room temperature, 70°C, and 90°C on a commercial synthesizer and our optimized AFPS system.[30] The resulting resin was cleaved and analyzed in our laboratory, and findings indicated substantially improved synthetic outcomes for the optimized AFPS protocol. Machine-specific, optimized conditions were used on each instrument to achieve the best synthesis outcome.

The two proteins used in this evaluation were proinsulin (86 amino acids) and human

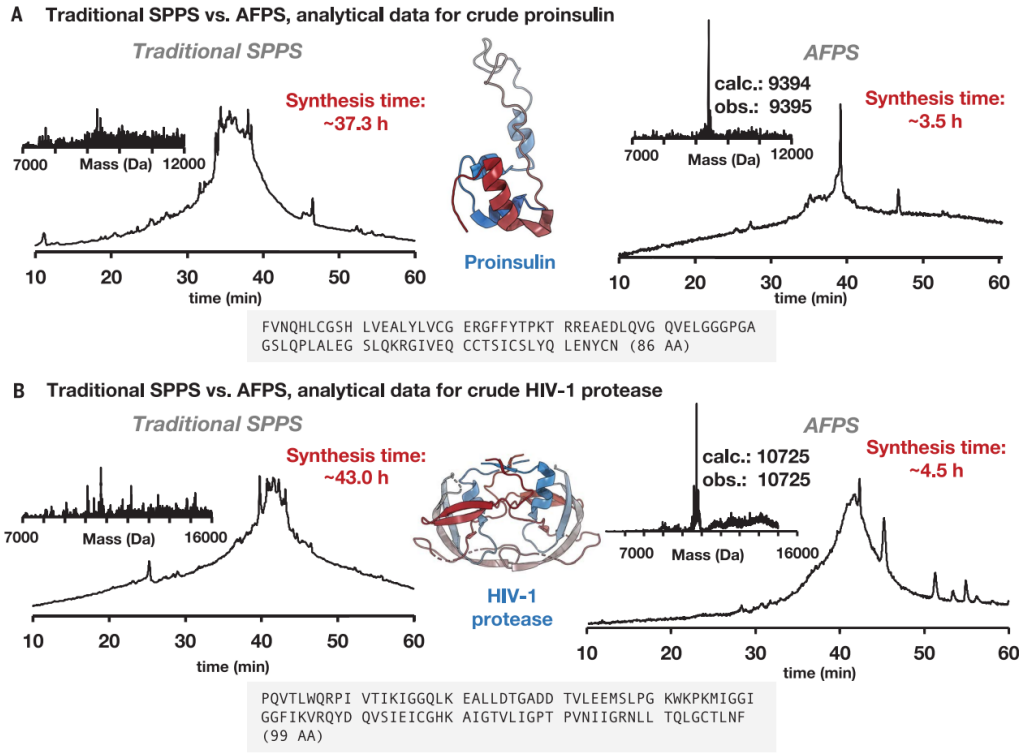


Figure 3.8: Comparison of Traditional SPPS with the AFPS for proinsulin and HIV-1 protease

Analytical HPLC data of the crude proinsulin (A) and HIV-1 protease (B) are presented as the main chromatographic traces with absorbance detection at 214 nm (additional details in the SM). Deconvoluted masses are displayed in the insets. Analytical data for the synthesis of crude protein chain using SPPS on a commercially available synthesizer at 70°C with total cycle times of 26 min per amino acid and 40 equivalents of amino acid for each coupling are displayed on the left; analytical data for the synthesis of crude protein chain using AFPS at 90°C with 60 equivalents of amino acid for each coupling are displayed on the right. PDB 2KQP (proinsulin) (58) and 2JE4 (HIV-1 protease dimer with inhibitor) were used.

immunodeficiency virus-1 (HIV-1) protease (99 amino acids). Both proteins have been produced via SPPS using native chemical ligation (NCL) techniques. Proinsulin was assembled from three

peptide fragments individually prepared by SPPS and ligated under mild conditions.[31] HIV-1 protease, on the other hand, was previously prepared using stepwise and chemical ligation routes under Boc-SPPS conditions.[32] Both of these proteins are of high therapeutic interest and are extensively studied.[33] Using our standard AFPS protocol, the syntheses of proinsulin and HIV-1 protease were completed in 3.5 and 4.5 hours, respectively. HPLC purification yielded 2.2 mg (1% yield) of purified proinsulin and 5.3 mg (1% yield) of purified HIV-1 protease.

For HIV-1 protease and proinsulin, the AFPS yielded the desired product as the major species along with minor by-products of similar weight, as determined by analytical HPLC and liquid chromatography–mass spectrometry (LC-MS). Conversely, synthesis on commercially available peptide synthesizers took approximately five times longer, 37.3 hours and 43 hours for proinsulin and HIV-1 protease, respectively, resulting in a complex compound mixture. The desired peptide sequence was nearly impossible to detect via LC-MS. We could reasonably conclude that the AFPS substantially improves synthetic outcomes, not only in performance and yield but also in the speed of production compared to traditional SPPS.

3.3.3 Synthetic results of all single domain proteins synthesized

Finally, we began to synthesize several protein polypeptide chains of varying lengths and

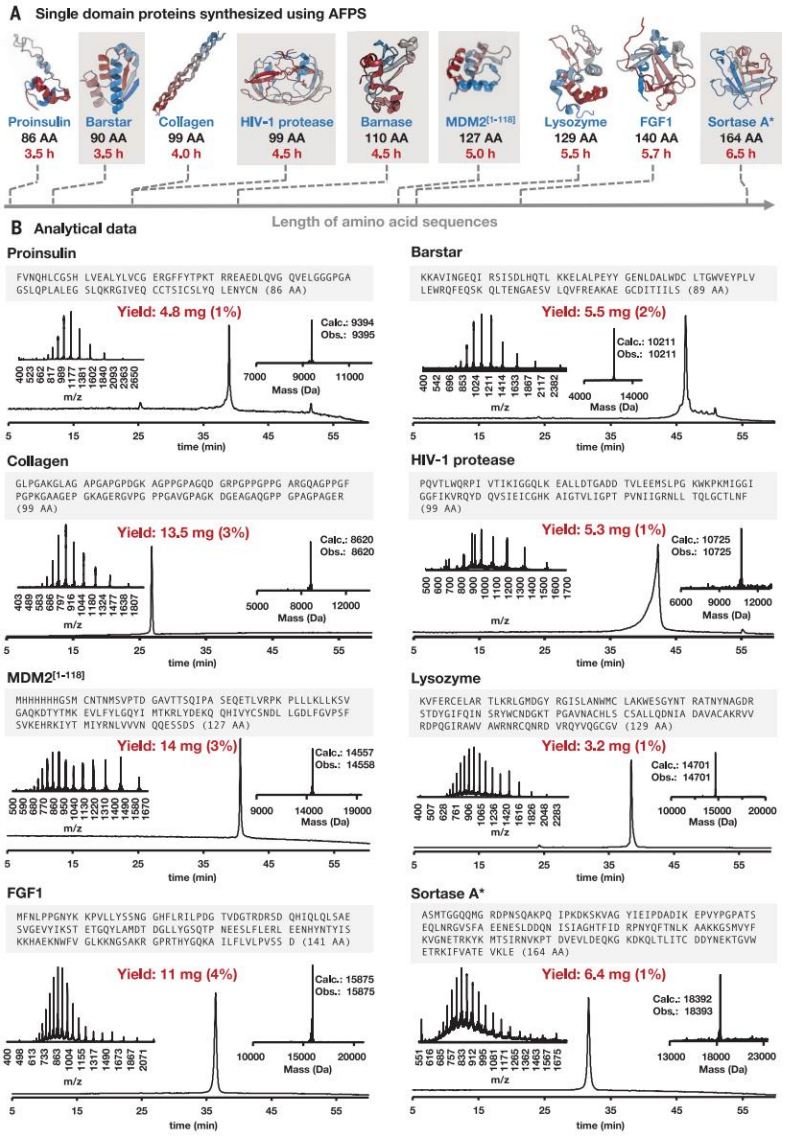


Figure 3.9: Summarized synthesis data of single-domain proteins produced via AFPS

(A) Sequences produced using an AFPS instrument. Sequences highlighted in gray were folded and purified, and their structure and biological activity were evaluated. All sequences were synthesized using the same standard recipe. PDB 1AY7 (barstar) (59), 2KQP (proinsulin) (58), 1CGD (collagen) (60), 2JE4 (HIV-1 protease dimer with inhibitor) (30), 1BRS (barnase) (57), 3G03 (MDM2) (61), 2NWD (lysozyme) (62), 4Q9G (FGF1) (63), and 2KID (sortase A) (64) were used. **(B)** Analytical data for the purified sequences of proinsulin, barstar, collagen, HIV-1 protease, MDM2[1–118], lysozyme, FGF1, and sortase A*. For all cases, analytical HPLC data of the purified protein chains are presented as the main chromatographic trace with absorbance detection at 214 nm. Electrospray ionization (ESI) mass spectrum (upper left) and deconvoluted mass spectrum (upper right) are also shown in each case. Both spectra were obtained by summation of the entire LC peak

compositions to demonstrate the general applicability of our AFPS protocol. We synthesized and

evaluated protein polypeptide chains ranging from ~70 to ~170 amino acids. We selected these sequences carefully and focused on producing proteins that have been heavily evaluated and are well understood. These sequences were chosen to enable comparison with literature data and/or commercially available recombinant protein samples. We chose not only historically relevant targets for drug discovery, such as HIV-1 protease[33] and murine double minute 2 (MDM2)[34] but also proteins that serve as therapeutics themselves, such as FGF1[35] and proinsulin.[36] Barstar,[37] barnase,[38] lysozyme,[39] MDM2,[40] and sortase A*[41] allowed for a direct comparison of recombinant and synthetic proteins either through purchase or in-house bacterial expression.

Furthermore, we wanted to demonstrate the ability of AFPS technology to incorporate noncanonical amino acids rapidly and simultaneously in more significant numbers and of greater diversity than biological methods were tested by synthesizing derivatives of barnase and HIV-1 protease containing site-directed mutations.

In the case of barnase, we incorporated p-bromophenylalanine at a site previously investigated for mutational tolerance.[42] Then, we produced synthetic HIV-1 protease in which two methionine and one cysteine residues were replaced, as previously described in the literature, with norleucine and aminobutyric acid, respectively, to avoid potential oxidation side products and increase synthetic efficiency.[43] All sequences were successfully synthesized in 3.5 to 6.5 hours. The desired protein was the main product in every synthesis, and HPLC purification yielded milligram quantities of all protein polypeptides. Isolated yields after HPLC purification ranged from 2.2 to 19.0 mg (1 to 5%), an appropriate amount of material to perform biological activity and biophysical assays. Here, we demonstrated that the optimized AFPS allows for the routine stepwise chemical synthesis of peptide chains of up to ~170 amino acids, substantially decreasing the time and labor associated with the chemical production of single-domain proteins.

3.3.4 Example: Synthesis and purification of native barnase and barstar

Here, I provide a specific example of the synthetic pipeline in the synthesis and purification of the barnase-barstar binding pair; the relevancy of these proteins will be further enumerated in section 3.3.4.

Barstar (89 aa) was synthesized on the AFPS instrument, the Amidator via SPPS in flow on RINK amine ChemMatrix® LL (0.18 mmol/g), which will yield a C-terminal amide after cleavage. Low-loading resin was used since previous studies comparing the loading of resins used in SPPS unveiled that using lower-loading resins often results in higher synthetic yields due to

limiting the effects of interchain interactions, such as aggregation, for the growing polypeptide.[24] Barnase (113 aa) was synthesized on the Amidator as well using 50 mg of HMPB ChemMatrix (0.44 mmol/g) pre-coupled to arginine, yielding the C-terminal carboxylic acid after cleavage. Utilizing this resin allows us to produce the native peptide sequence as a cell would express. Furthermore, the synthesis with HMPB demonstrates that SPPS does not limit us to having an unnatural C-terminal amide. Still, with an extra step, we can yield the native carboxylic acid.

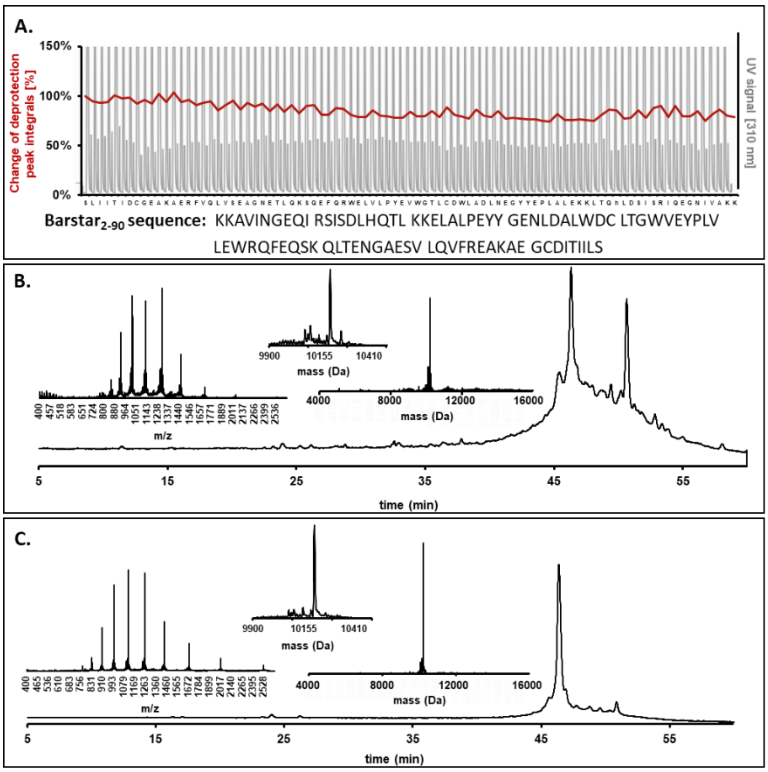


Figure 3.10: Barstar synthesis data

A. UV trace from the synthesis of barstar with overlaid graphed change in percent of deprotection peak integrals. Notice early aggregation in -IDCG- region and generally steady deprotection peak height thereafter. Barstar was synthesized on RINK Amide resin which will yield the C-terminal amide after cleavage. **B.** HPLC traces (absorbance at 214 nm in mAU), LCMS data (m/z in Da), and deconvoluted mass spectra (obtained from integration over all LCMS signals) of folding synthetic barstar following purification by anion exchange chromatography. Calculated mass: 10211 Da, Observed mass: 10211 Da. **C.** Analytical HPLC traces (absorbance at 214 nm in mAU), LCMS data (m/z in Da), and deconvoluted mass spectra (obtained from integration over all LCMS signals) of folding synthetic barstar following purification by anion exchange chromatography. Calculated mass: 10211 Da, Observed mass: 10211 Da.

Synthesis of barstar was completed in approximately 3.5 hours. The synthesis of barnase was done in 4.5 hours, after which the polypeptide was cleaved from the resin, and the crude material was analyzed by LCMS and analytical HPLC. Following cleavage, the dry lyophilized powder was dissolved in a solution of 6M guanidinium hydrochloride (Gn·HCl), 100 mM DTT, and 55 mM $NaHPO_4$ at pH 7.8 and loaded onto a semipreparative C3 column at 60 °C for

purification by RP-HPLC. DTT was added to prevent the formation of disulfide bonds during the purification process to increase post-purification yields,[44] and Gn·HCl was added to help solvate the long hydrophobic barstar polypeptide. An optimized gradient of 5-43% B with 1.5% B/min, 43-63% B with 0.2% B/min, and 63-73% B with 1% B/min was used for the purification with mobile phase A being water with 0.1% TFA and mobile phase B being acetonitrile with 0.1% TFA. Purifying long peptide chains in a heated column and a slow gradient for the expected elution time of barstar allows for enhanced separation of byproducts from desired products and noticeably increased yields of clean peptides.[45] Following fractional analysis, 5.5 mg (2% yield) of purified barstar peptide was recovered and moved forward into folding and biological assays. This presents the first time barstar was able to be synthesized via SPPS. Previously, due to its highly hydrophobic nature, it was not an appropriate candidate for study in the Pentelute group;[46] however, we have now shown that synthesis of the barstar polypeptide in flow on the AFPS is possible. Similarly,

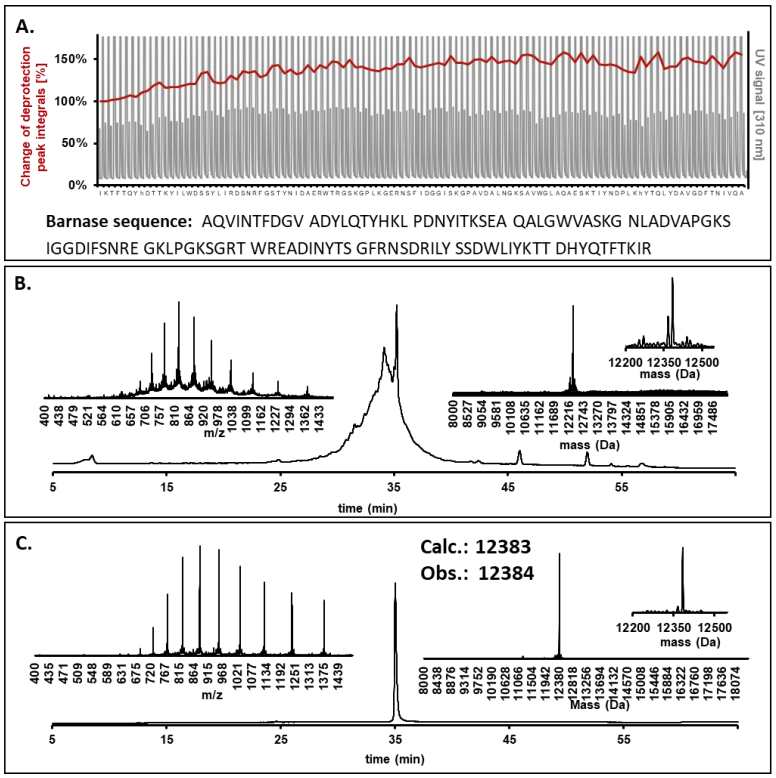


Figure 3.11: Barnase synthesis data

A. UV trace from the synthesis of barnase with overlaid graphed change in percent of deprotection peak integrals. **B.** HPLC traces (absorbance at 214 nm in mAU), LCMS data (m/z in Da), and deconvoluted mass spectra (obtained from integration over all LCMS signals) of folding synthetic barstar following purification by anion exchange chromatography. Calculated mass: 12383 Da, Observed mass: 12384 Da. **C.** Analytical HPLC traces (absorbance at 214 nm in mAU), LCMS data (m/z in Da), and deconvoluted mass spectra (obtained from integration over all LCMS signals) of folding synthetic barstar following purification by anion exchange chromatography. Calculated mass: 12383 Da, Observed mass: 12384

barnase was purified via HPLC with an optimized gradient of 5–25% B with 1% B/min, 25–45%

B with 0.2% B/min, 45–65% B with 1% B/min, with mobile phase A being water with 0.1% TFA and mobile phase B being acetonitrile with 0.1% TFA. Following fractional analysis, 8.4 mg (2% yield) of purified barnase peptide was recovered and moved forward into folding and biological assays.

3.4 Biological evaluation of synthetic proteins

We had to perform extensive biological evaluations of the proteins to fully evaluate and demonstrate the power of single-domain protein polypeptides synthesized via the AFPS. Moving into this step, one challenge is that determining the purity of long synthetic polypeptides similar to the ones we have produced can have difficulties associated with identifying and quantifying by-products using standard analytical techniques. This is because the differences in our produced sequences can be as minor as a one or two amino acid deletion, which is extraordinarily difficult to purify away from the full-length polypeptide.[47] In the cell, a protein polypeptide is quickly folded into its native globular structure. This folding process is critical to the protein's ability to perform its native biological activity. Thus, we can use the tertiary structure of a protein as a measure of the chemical integrity of the primary amino acid sequence.[48]

To this end, we selected appropriate folding conditions for all our produced proteins and then applied biological purification methods to purify any misfolded protein further. These misfolded proteins were likely not the desired protein polypeptide sequence. We were then posed to characterize these proteins' tertiary structure with biophysical and functional assays alongside recombinant protein standards. Our overall goal was to demonstrate the fidelity of our AFPS protocol in producing full-length and biologically active synthetic proteins of defined covalent structure and high chiral integrity.

3.4.1 Brief overview of the folding and characterization of HIV-1 protease, sortase A*, and MDM2

Following purification, we were able to fold and perform functional assays on sortase A, MDM2, and HIV-1 protease. Folding of the synthetic proteins was case-specific and was achieved either by following a literature protocol or by screening for optimal conditions. In our explorations, in-house enzymatic assays show comparable activity of synthetic proteins obtained by AFPS and their recombinant equivalents obtained via biological expression in our group or purchased from a vendor. Enzymatic catalysis is highly sensitive[49] to minor changes in the enzyme's tertiary structure, and even single point mutations can significantly impact function.[50]

First, we looked at and confirmed the primary structure of HIV-1 protease via LC-MS and HPLC methods, and these data are available in section 3.3.3. HIV-1 protease's native function is to hydrolyze the peptides of HIV and plays an important role in HIV infection progression and the later development of AIDS.[51] Thus, HIV-1 protease is a popular drug target and has been extensively studied. We can characterize HIV-1 protease's proteolytic activity by using a fluorogenic peptide; this will allow for quantification of its proteolytic activity by observing the rate of peptide degradation and the resulting fluorescence increase.[52] Our synthetic HIV-1 protease demonstrated a Michaelis constant of K_M = 20.9 ± 1.0 mM (mean ± SE) and a turnover of k_{cat} = 29.6 ± 4.1 s^{-1} (mean ± SE). These results are very similar to the literature values reported for this sequence.[53] Further, incubation of the synthetic protease with a model substrate peptide results in wild-type-like specificity with exclusive cleavage at a single Phe/Pro site.[54]

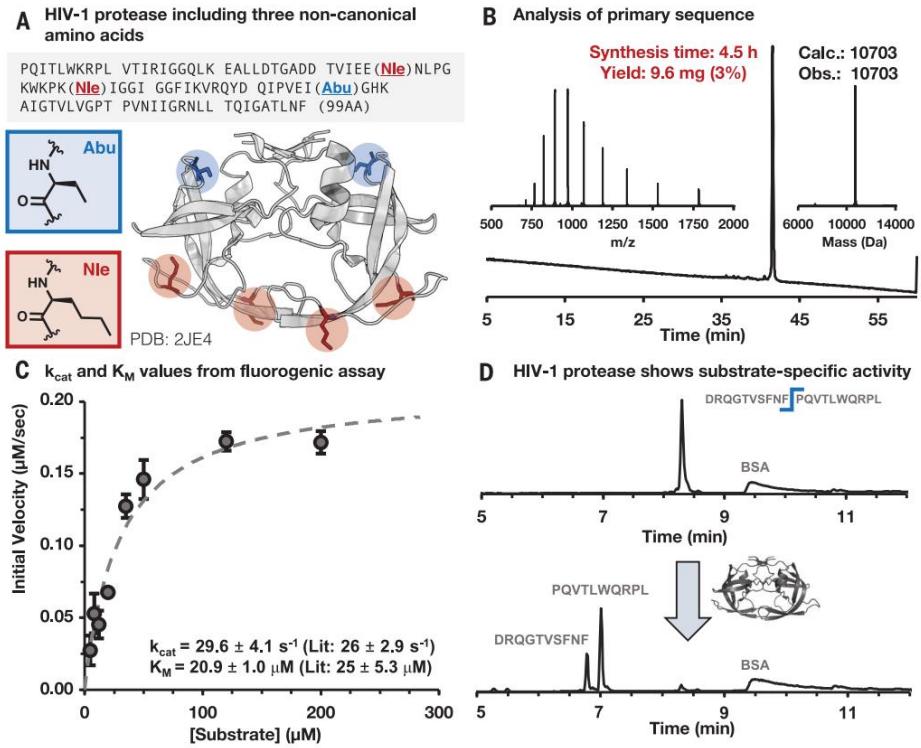
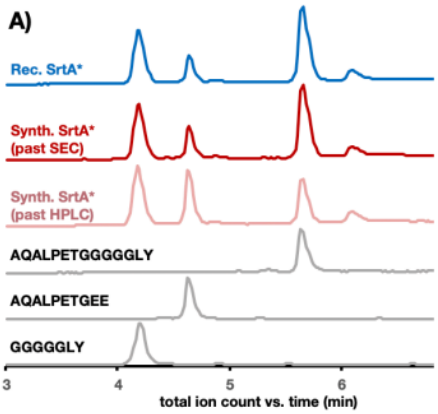


Figure 3.12: Synthetic HIV-1 protease assay outcomes

A. Crystal structure of HIV-1 protease dimer with highlighted noncanonical amino acids aminobutyric acid (Abu, blue) and norleucine (Nle, red). PDB 2JE4 **B.** Primary structure obtained from AFPS. **C.** Quantitative enzymatic activity assay performed in triplicate for the determination of k_{cat} and K_M values. **D.** Qualitative substrate specificity assay with model substrate p12nt, in which HIV-1 protease exclusively cleaves at a single Phe-Pro site whereas bovine serum albumin (BSA) stays intact.

Next, we investigated Sortase $A_{59\text{-}206}$; Sortase A* is a transpeptidase produced by Gram-positive bacteria that catalyzes a cell wall sorting reaction at a threonine-glycine bond in the LPXTG motif.[55] We synthesized the 164-amino acid long sortase A* variant



B)

Protein sample	GGGGGLY	AQALPETGEE	AQALPETGGGGGLY
Recombinant SrtA*	37%	13%	**50%**
Synthetic SrtA* (past SEC)	38%	15%	**47%**
Synthetic SrtA* (past HPLC)	43%	30%	**27%**

Figure 3.13: Assay results for sortase A*

A. LCMS traces (TIC) of the three assay reactions and the three peptide standards. Ligation product AQALPETGGGGGLY was formed for all three assay conditions indicating active sortase A in each reaction. B) Area under each peak in the TIC, manually integrated in MassHunter software. The SEC purified synthetic sortase A has similar values to the recombinant sample, with improved activity over synthetic folded sample which was not purified by SEC.

[P94S/D160N/K196T] to allow for direct comparison to a recombinant standard.[56] At a concentration of 0.01 mg/mL, synthetic sortase A* led to 47% product formation by LC-MS within 24 h (starting from 0.2 mg/mL GGGGGLY and AQALPETGEE as test substrates). This conversion value is comparable to that determined for the recombinant protein (50% product formation within 24 h). Enzymatic activity assays of these two synthetic proteins produced by AFPS, therefore, confirmed both the high substrate specificity and comparable activity to recombinant enzymes and literature values.[57]

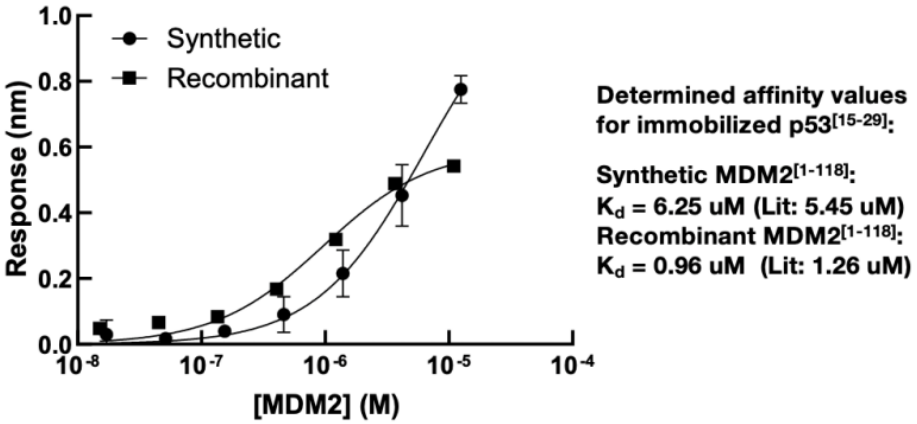


Figure 3.14: Assay results for MDM2

Determination of Kd from concentration-dependent equilibrium binding responses. All experiments were carried out in triplicates and Kd values were determined from equilibrium binding responses of immobilized p53 to MDM2 at various concentrations.

Finally, we assessed the binding of synthetic MDM2 and compared it to a recombinantly produced sample. To further quantify the binding of a synthetic protein to a known ligand, we also characterized the *N*-terminal binding domain of MDM2$^{[1-118]}$.[58] The binding of MDM2 to p53 is a

key interaction in multiple pathways upregulated in cancer,[59] and is highly studied and a major target for drug discovery efforts.[60] We folded milligram quantities of synthetic MDM2$^{[1\text{-}118]}$ and characterized its binding to immobilized p53$^{[14\text{-}29]}$ using biolayer interferometry. Synthetic MDM2$^{[1\text{-}118]}$ displayed an affinity toward p53 (K_d = 6.25 μM) comparable to the literature value (K_d = 5.45 μM) obtained under the same folding conditions. Thus, we have demonstrated that our synthetic protein samples perform comparably to recombinant samples and literature values. In the next section, we will dive deep into the folding and characterization of the barnase-barstar binding pair.

3.4.2 Example: Folding and characterization of the barnase-barstar pair

Two proteins we did extensive work with to understand their biological and biophysical properties were barnase (110 aa), a ribonuclease, and barstar (90 aa), barnase's inhibitor, which are both natively expressed in the bacterium *Bacillus Amyloliqefaciens.*[61] This pair has also been subjected to extensive kinetic protein folding studies that revealed their biological folding pattern[62] and structure,[63]. The biological assays to study barnase and barstar interactions are well-documented, even within the Pentelute group itself.[64] Barnase, a bacterial RNase that cleves at

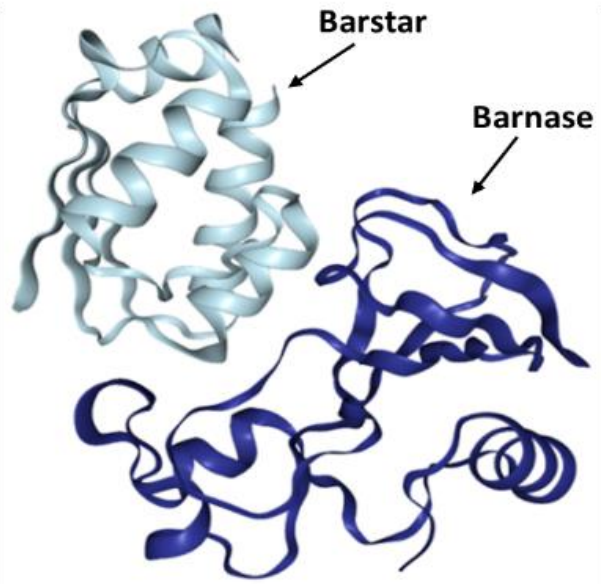


Figure 3.15: Barnase-barstar binding interaction based on X-ray crystallography data (2ZA4**).**

GpN sites in RNA, is an excellent model for investigating protein folding, denaturation, and binding to its inhibitor protein barstar.[65] The barnase-barstar complex[66] is especially interesting because of the extraordinarily high binding affinity (K_d ~10^{-14}M) of the pair, which has even led some to consider using the barnase-barstar binding pair in place of the traditional streptavidin:

biotin system.[67] In the past, members of the Pentelute group were able to synthesize barnase and its mirror image D-barnase using NCL techniques. In this publication, barnase's activity was assessed

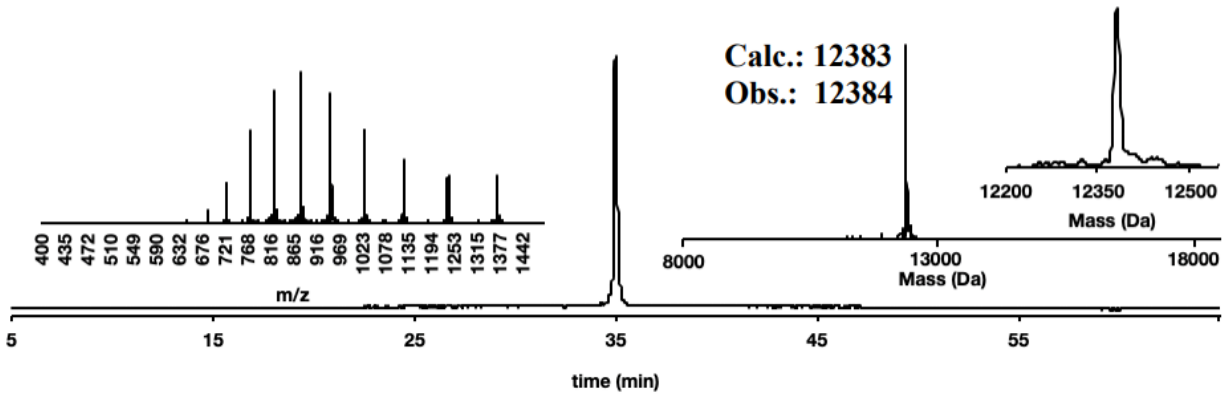


Figure 3.17: LCMS and HPLC data for SEC purified barnase

using a modified version of the supersensitive fluorogenic RNase assay developed by Raines and coworkers;[68] however, inhibition of barnase activity by barstar was not reported. Synthetically produced barstar has not been reported before due to its hydrophobic nature, which results in extensive inter-chain interactions between the growing polypeptide anchored to resin. This results in extensive deletions and truncations early on in the synthesis of barstar, and using traditional batch SPPS, researchers could not easily resolve this issue. However, using flow-based SPPS via the AFPS, we have successfully produced and folded barstar as well. The barnase-barstar binding pair presented an opportunity to demonstrate the performance of our improved AFPS conditions that have resulted in increased synthesis quality and yield of long, complex peptide chains. This improvement has permitted the synthesis of barnase, barnase variants, and barstar in flow, allowing us to not only assay synthetic barnase's activity but also assay for inhibition of synthetic barnase by synthetic barstar. Herein, I describe the synthesis and subsequent biological assays of native barnase and barstar.

As shown earlier in this chapter, section 3.3.3, the primary structures of synthetic and recombinant barnase and barstar were indistinguishable by LC-MS and HPLC methods. With this in mind, we first turned to folding and biophysical assays of barnase. To fold the native barnase polypeptide, we used a method adapted from Mong et al. 2.5 mg of wild-type barnase was dissolved in 150 µL of 6 M Gn·HCl solution in 50 mM NaH2PO4 RNase free1 buffer, pH = 7.8. The mixture was serially diluted in four steps to 0.3 M Gn·HCl using a buffered RNase-free solution (150 mM Tris buffer pH 7.5, 150 mM NaCl). The final solution was then filtered and subjected to size exclusion chromatography (SuperdexTM 75 Increase 10/300 GL, 0.25 mL/min),

using isocratic eluent 150 mM Tris buffer pH 7.5, 150 mM NaCl, RNase free. Fractions containing SEC-purified barnase were pooled and concentrated with a 3K molecular weight cut-off spin filter

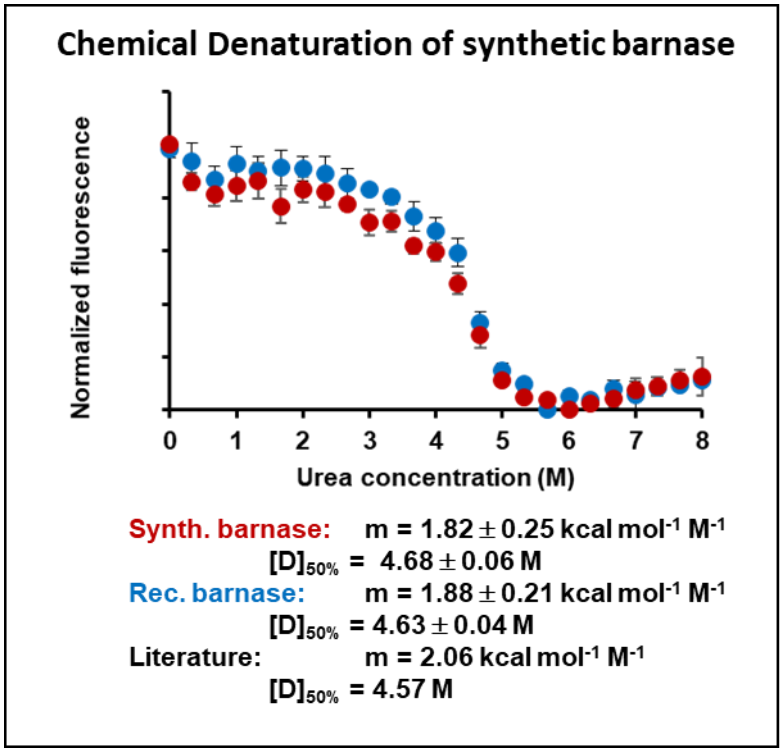


Figure 3.18: Results of chemical denaturation studies of barnase

Structural evaluation of barnase in a chemical denaturation assay using urea as denaturant performed in triplicate, results are reported as mean ± standard error (SE). Error bars on the graph indicate standard deviation.

to yield a total of 0.8 mg of native barnase, corresponding to a 33% isolated yield.

With folded and biologically purified barnase in hand, we turned to using a chemical denaturation fluorometric assay as a read-out for the integrity of the tertiary structure.[69] In this assay, tyrosine fluorescence was used to monitor the folding equilibrium, as the urea concentration was varied. In brief, the protocol for this assay involved loading a 384-well plate using a mixture of native barnase in 25 different urea concentrations ranging from 0 – 8 M (at 0.33 M steps). The solutions were equilibrated at room temperature for two hours, and then the emission spectrum was recorded. The transition midpoint ([D]50% - the concentration of urea at which half of the sample is unfolded) and m-values (the slope of the unfolding transition) were determined by fitting the normalized fluorescence readings. In these studies, synthetic native barnase exhibited a transition midpoint that compared well to both the authentic recombinant sample and literature value ($[D]_{50\%,\ synthetic} = 4.68 \pm 0.06$ M; $[D]_{50\%,\ recombinant} = 4.63 \pm 0.04$ M (mean ± SE); $[D]_{50\%,\ literature} = 4.57$ M). More importantly, the *m*-values obtained in the experiment, which describe the slope of the unfolding transition and are a sensitive measure of structural homogeneity, were similar

($m_{synthetic}$ = 1.82 ± 0.25 kcal mol^{-1} M^{-1}; $m_{recombinant}$ = 1.88 ± 0.21 kcal mol^{-1} M^{-1} (mean ± SE); $m_{literature}$ = 2.06 kcal mol^{-1} M^{-1}).

$$I = I_f - (I_f - I_0)e^{-\frac{k_{cat}}{K_M}[E](t-t_0)} \quad (1)$$

Next, we used a fluorogenic substrate to perform a fluorogenic RNase activity assay. Barnase catalyzes hydrolysis at diribonucleotide GpN sites, so we utilized the substrate 6-FAM-dA^{D}rG^{D}dA^{D}dA^{D}-6-TAMRA). Barnase's specific activity can be measured by monitoring the hydrolysis of this DNA/RNA hybrid that contains a Förster Resonance Energy Transfer (FRET) fluorophore pair.[70] In short, the protocol involved diluting the fluorogenic substrate in assay buffer and monitoring fluorescence at 515 nm upon excitation at 495 nm. Then, the fluorescence was observed every 10 seconds for at least 500 seconds to measure the starting fluorescence, I0, and to ensure no background cleavage took place prior to the addition of enzymes. Next, an aliquot of folded barnase was added to the substrate to start the enzymatic cleavage reaction, and the solution

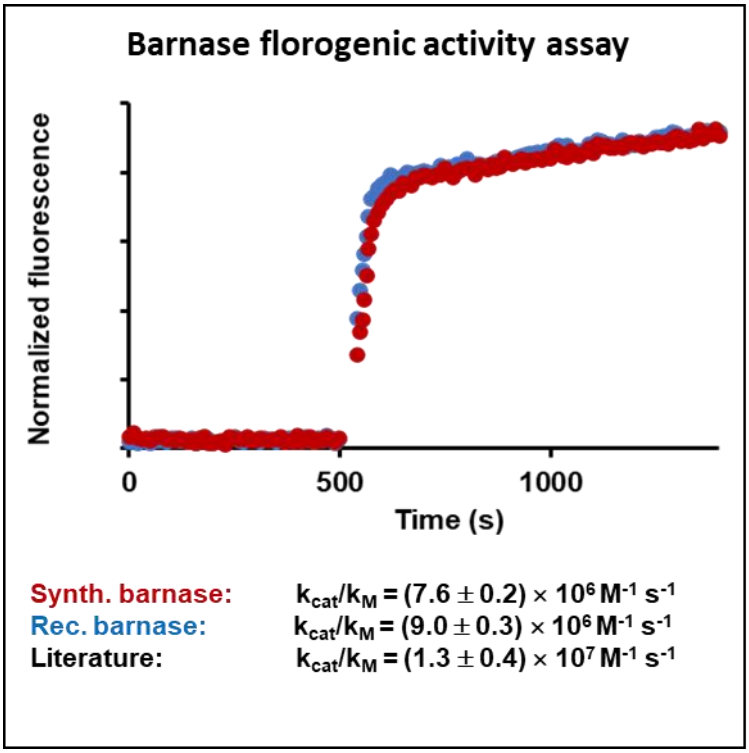


Figure 3.19: Results of barnase fluorogenic activity assay

Quantitative enzymatic activity assay performed in triplicate, error bars not displayed for clarity, details outlined in SM. k_{cat}/K_M values are reported as mean ± SE

was rapidly mixed. The increase in fluorescence was then monitored for 1000 seconds. The results of this assay showed that the enzymatic efficiency of synthetic barnase was $k_{cat}/K_M = (7.6 \pm 0.2) \times 10^6\ M^{-1}\ s^{-1}$ (mean ± SE), which is comparable to that of recombinant barnase ($k_{cat}/K_M = (9.0 \pm 0.3) \times 10^6\ M^{-1}\ s^{-1}$) (mean ± SE) as determined using the same assay.

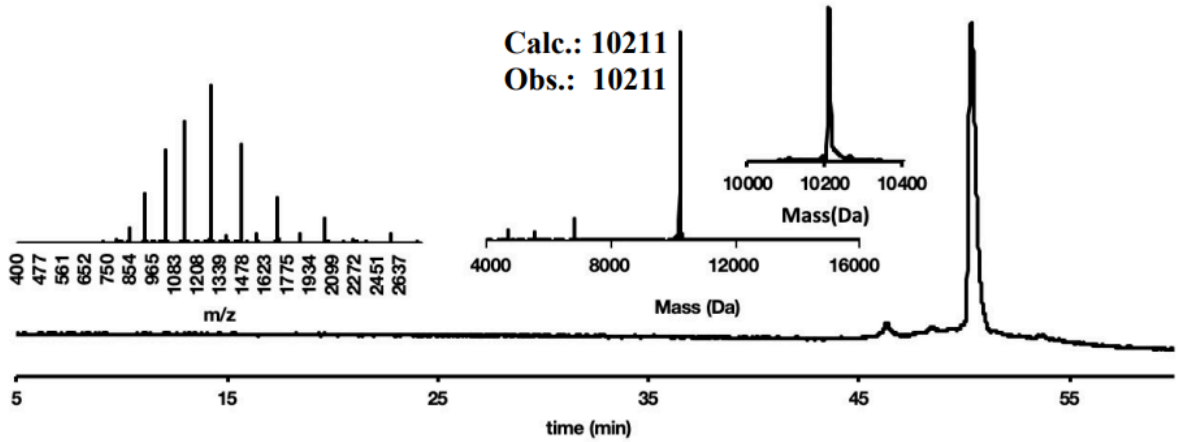


Figure 3.20: LCMS and HPLC data for folded and IEC purified barstar

Finally, we assess the ability of barnase to bind selectively and with high affinity to its inhibitor barstar. First, however, we need to fold and biologically purify synthetically produced barstar. To do this, we turned to a critical player in the study of barstar's biophysical characteristics is, Jayant B. Udgaonkar, who has multiple barstar folding and unfolding studies published using either a high concentration of guanidium hydrochloride or urea as the unfolding buffer.[71] To stay in line with the folding of barstar's binding partner barnase, we utilized Gn·HCl in our unfolding buffer. The final folding conditions after optimization consisted of dissolving lyophilized barstar in an RNase-free solution of 6 M Gn·HCl and native buffer (5 mM NaH_2PO_4, 250µM EDTA, 250 µM DTT) as the unfolding buffer, which was 10-fold serially diluted to a concentration of 0.6 M Gn·HCl, using the native buffer. EDTA and DTT were added to prevent disulfide bond formation,[72] and serial dilution prevents aggregation of the folding barstar intermediates and increases the yield of properly folded barstar.[73] The resulting solution was then desalted and immediately subjected to anion exchange chromatography using gradient elution with RNase-free mobile phases 20 mM Tris buffer pH 7.4 and 20 mM Tris NaCl pH 7.4. Fractions containing the desired protein eluted at a concentration of approximately 300 mM NaCl; as previously reported,[74] and clean fractions were pooled and concentrated with a final yield of 5% or 0.14 mg of folded barstar.

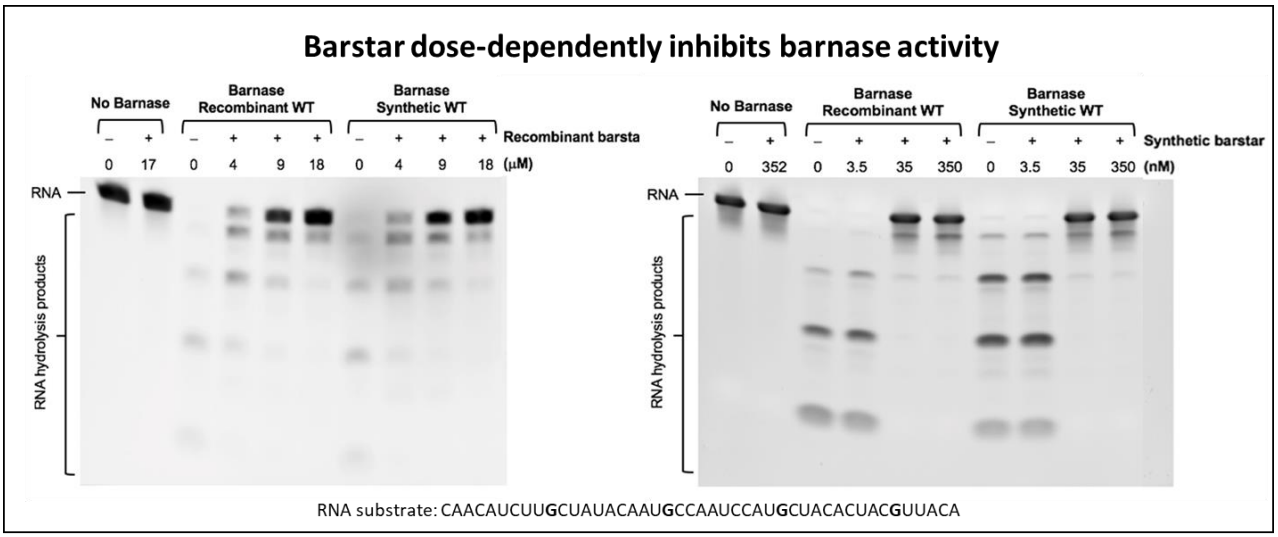


Figure 3.21: Results of gel-based barnase inhibition assay with barstar

These RNA hydrolysis assay gel shows that RNase activity inhibition by barstar is dose dependent for recombinant WT barnase and synthetic WT barnase. Additionally, both synthetic barstar and recombinant barstar inhibit barnase comparably and in a dose-dependent manner.

Now, we can assess the ability of synthetic barstar to effectively inhibit wild-type (WT) barnase by utilizing an optimized RNA hydrolysis assay previously used in the group to study D-barnase. For this study recombinant barstar was purchased and used as a baseline comparison of the performance of the synthetic barstar. Folded barnase and recombinant barnase were obtained for the above-outlined studies. Since barnase catalyzes hydrolysis at diribonucleotide GpN sites, we selected an RNA substrate containing four GpN sites, CAACAUCUU**G**CUAUACAAU**G**CCAAUCCAU**G**CUACACUAC**G**UUACA, to allow for easy analysis of the performance of barstar inhibiting barnase. Assay results showed that synthetic barstar selectively inhibits RNase activity of both synthetic and recombinant WT barnase comparably to recombinant barstar in a dose-dependent manner. As the relative concentration of barstar to barnase increases, more of the RNA substrate is intact, indicating that barnase is significantly less active due to increased levels of inhibition by barstar.[75]

These results from the synthesis of barstar and barnase and subsequent confirmation of native biological and biophysical activity, combined with other protein work by Pentelute lab members, shows us that our AFPS technology is now able to reliably produce polypeptides of sufficient chiral purity and quality to facilitate the synthesis of biologically active single domain proteins. This proof-of-concept has created the framework through which we can focus more on practical applications of synthetic proteins.

3.5 Discussion

This chapter reports on how we have optimized the AFPS protocol to produce full-length protein polypeptide sequences over three times longer than previously accessible by routine standard SPPS methods, an advantage of flow chemistry over common batch chemistry methods for peptide synthesis.[14] An improvement to existing flow protocols was achieved by a rapid screening of variables in a reproducible reaction setup. Even though in this study, AFPS yields superior results over traditional synthesis methods in terms of total synthesis time and crude product quality, general challenges associated with peptide synthesis, such as low atom economy and the use of DMF as a solvent, remain unsolved. Another potentially limiting feature of our setup is the synthesis scale. The reactor capacity used in our study allows up to 200 mg of resin with a loading of 0.49 mmol/g. However, we have already made strides in integrating a larger reactor into the current system. Since implementing the first generation AFPS, we have produced over 5000 peptides. Using machine learning and other computational methods, we have been using the resulting synthesis data to improve peptide synthesis in flow further. Ultimately, we intend this report to serve as a blueprint for the automated flow synthesis of other biopolymers of interest.[76]

A robust, widely available routine method for chemical production of proteins is poised to impact chemical biology and the development of new therapeutics strongly. Our advances provide a viable solution to reliably assemble long linear peptide chains, shifting the focus in the field of chemical protein synthesis to improving folding protocols and, most importantly, applications. Combined with chemical ligation, rapid stepwise production of single-domain proteins by AFPS technology will extend the practical applications of total chemical synthesis to the majority of human proteins (~30 kDa).[77] In this respect, we envision adapting the incorporation of peptide hydrazides for thioester-based ligation to our AFPS protocol, an approach previously achieved with manual flow instrumentation.[78] Additional research avenues opened by our method include rapid access to mirror-image proteins, post-translationally modified proteins, and *de novo*-designed, abiotic proteins.[79] Introduction of non-canonical amino acids as point mutations in native proteins will make accessible variants with significantly altered biological function, e.g., catalytic activity.[80] Finally, AFPS has the potential to enable on-demand production of time-sensitive and potentially life-saving personalized medicine, such as for enzyme replacement therapy or neoantigen cancer vaccines.[81] To this end, in the following chapter, we will report on the synthesis

of several bioluminescent proteins, including an active mirror-image protein containing multiple disulfide bonds.

3.6 Future directions

Peptide name	Sequence	Length
Scorpion Toxin II	VKDGYI	6 AA
HA2-HIVTAT	GLFGAIAGFIENGWEGMIDGGRKKRRQR	28 AA
GLP-2	HGDGSFSDEMNTILDNLAARDFINWLIQTKITD	33 AA
Miraculin	DSAPNPVLDIDGEKLRTGTN	20 AA

Table 3.5: Sequences used in aspartimide studies

As discussed above, the AFPS system can still undergo further optimization to enable access to additional interesting biopolymers, both native and abiotic. One important optimization area we could further explore is that of aspartimide formation. As mentioned above, one major side-product that has inhibited our access to many unique sequences, such as DG-rich collagen, in aspartimide formation, often resulting in both α-Asp and the unnatural β-Asp formation. Aspartimide formation is a well-known phenomenon occurring mainly in -DG-, -DS-, -DA-, -DD-, -DQ- rich sequences both in vitro and in vivo, and in SPSS, is particularly prevalent due to the presence of a piperidine, a strong base, during the deprotection steps of SPSS.[25] To mitigate aspartimide formation, other research groups have utilized bulky protecting groups for aspartic acid and have backbone-protected glycine residues as well to prevent the nucleophilic attack of Asp and the formation of aspartimide.[82] Past work to address aspartimide formation has resulted in the development of several backbone-protecting groups that had varying levels of success at preventing aspartimide formation. For application on the AFPS, we intend to explore installing a backbone protecting group on Glycine to provide an alternative method to prevent aspartimide formation, with a focus on 2,4-dimethoxy benzyl (Dmb), since we saw such success with Dmb-Gly in our initial aspartimide formation studies.[83]

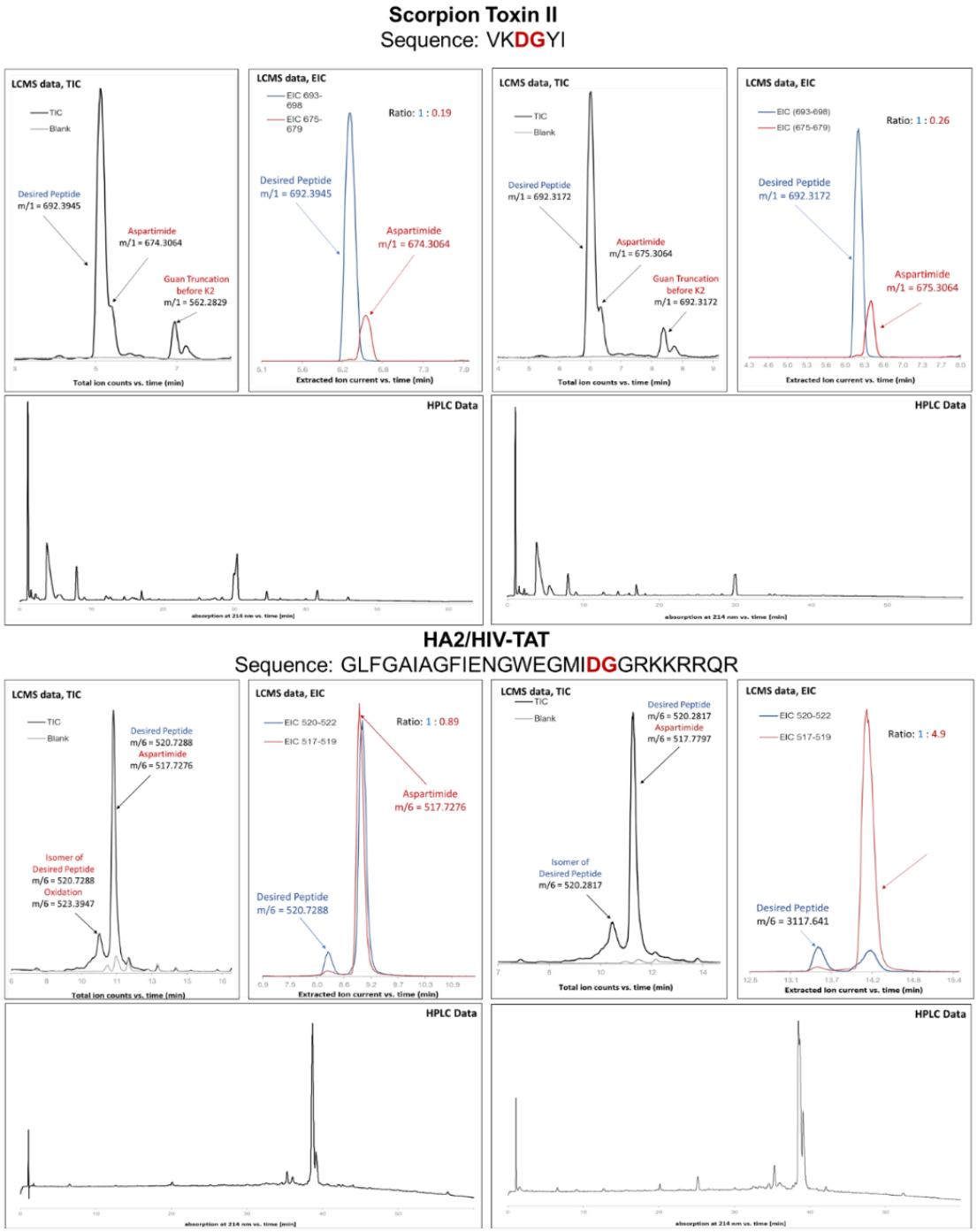


Figure 3.22: Synthetic data for Scorpion toxin II and HA2/HIV-TAT with aspartimide formation analysis Comparison of screened peptidse Scorpion Toxin II (top) and HA2/HIV-TAT (bottom) in both current conditions of 40% piperidine + 2% Formic Acid (left) and pre-optimization conditions of 40% piperidine (right). For STII notice similar by-products for each synthesis, but a noticeable increase in aspartimide formation without FA present. Unfortunately, the HPLC trace for neither condition resulted in useful analytical data, making STII a non-ideal peptide for further study. For HA2/HIV-TAT we also notice similar by-products, and noticeable increase in aspartimide formation without FA present, nearly 85% aspartimide. Further, there is still nearly 50% aspartimide formation under optimized conditions, making HA2/HIV-TAT an optimal test peptide for reducing aspartimide formation.

To this end, we needed to select a suite of aspartimide-prone peptide sequences to add to

our current test sequence NN92. After a dive into the literature, we selected four potential peptides of interest. The four peptides that were screened are as follows: a fragment of Scorpion Toxin II[84] (STII), a classic aspartimide-prone peptide used in many other studies,[85] an HA2 and HIV conjugate (referred to herein as HA2/HIV-TAT) which has been heavily studied for use as a cell-penetrating peptide,[86] Glucagon-like peptide-2 (GLP-2) a human-derived peptide that is closely related to GLP-1,[87] and Miraculin 1-20 a fragment of the taste-modifying protein that makes bitter foods taste sweet.[88] The final two peptides contain multiple DX (X= G, S, I, L, N, etc.) sites and are being considered to address the aspartimide problem in peptides with numerous aspartimide-prone sites by other research groups. Overall, our preliminary studies show that each peptide's synthesis and analysis performance varies. In an ideal test peptide, we would observe limited by-

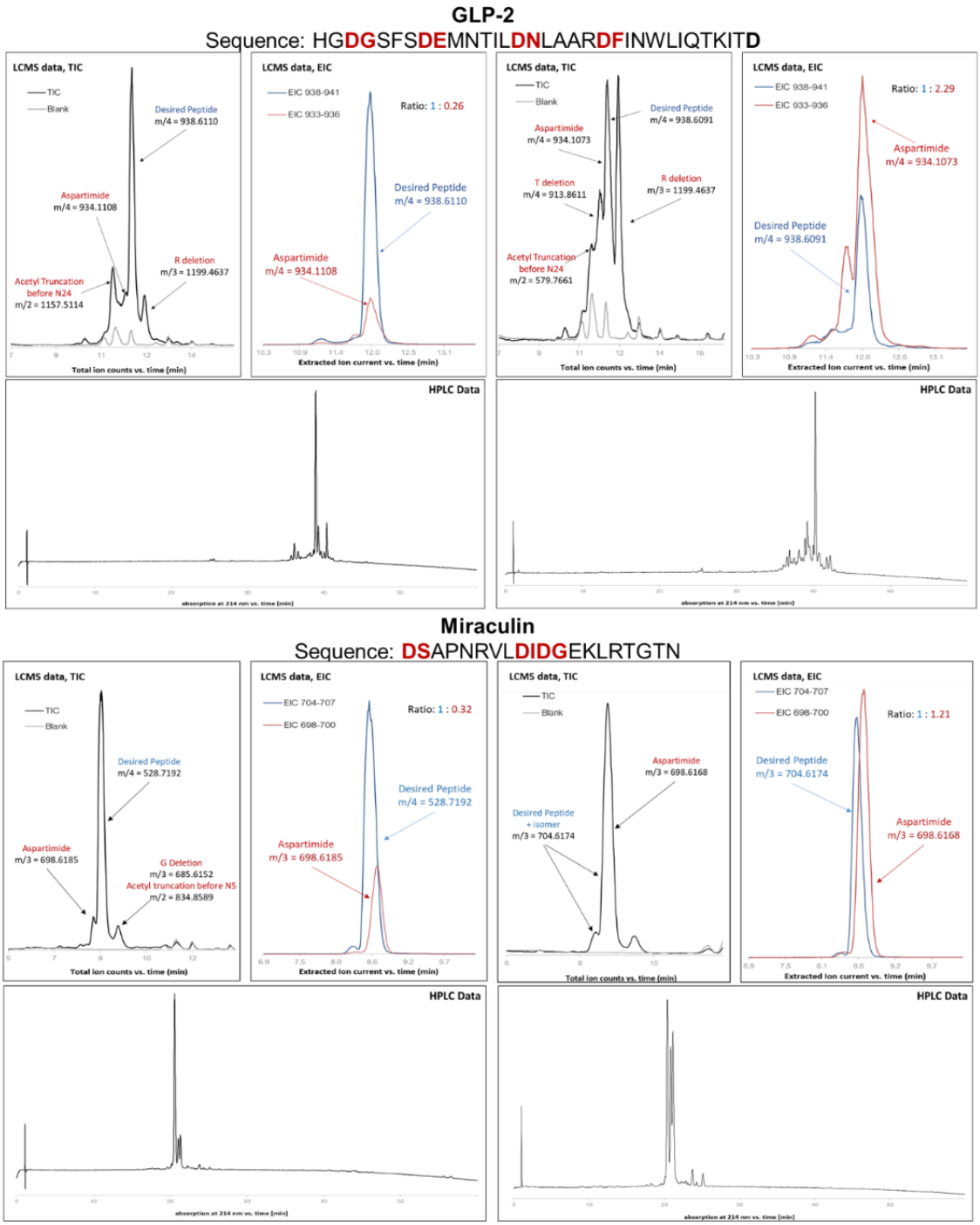


Figure 3.23: Synthesis data for GLP-2 and Miraculin with aspartimide formation analysis

Comparison of screened peptides GLP-2 (top) and Miraculin 1-20 (bottom) in both current conditions of 40% piperidine + 2% Formic Acid (left) and pre-optimization conditions of 40% piperidine (right). For GLP-2 notice increased by-products for synthesis without FA, as well as a noticeable increase in aspartimide formation and the presence of what appears to be an isomer for aspartimide as well. The HPLC UV trace appears to give good resolution of the various by-products observed in the LCMS trace. For Miraculin notice increase presence of by-products for optimized synthesis conditions and presence of an isomer as well. There is, however, a noticeable increase in aspartimide formation without FA present and the HPLC UV trace also appears to reflect and support the observations from the LCMS data.

products aside from aspartimide, high aspartimide formation overall, and easy-to-collect and

analyze HPLC and LCMS data. After analysis, the longer peptides assessed appeared appropriate to submit to future studies.

With test peptides selected, we are poised to adjust the AFPS Synthesis_Sequence_ Baseline to include the necessary steps required to incorporate different backbone-protecting groups on glycine's next to aspartic acids to prevent aspartimide formation. To do this, we will utilize peptoid chemistry and an exciting new avenue to explore and integrate into our AFPS systems. Incorporating these protecting groups via chemical strategies is more cost-effective in the long term than using pre-coupled amino acids purchased from vendors. We envision testing a series of amine-based backbone protecting groups, including Dmb, to screen for the group most effective at reducing aspartimide formation. Studies to this end are paused at this time, but hopefully, this project can be seen through to the end in the future. Nonetheless, this demonstrates

how we can continue to improve the AFPS to access longer, more complex sequences at lower costs.

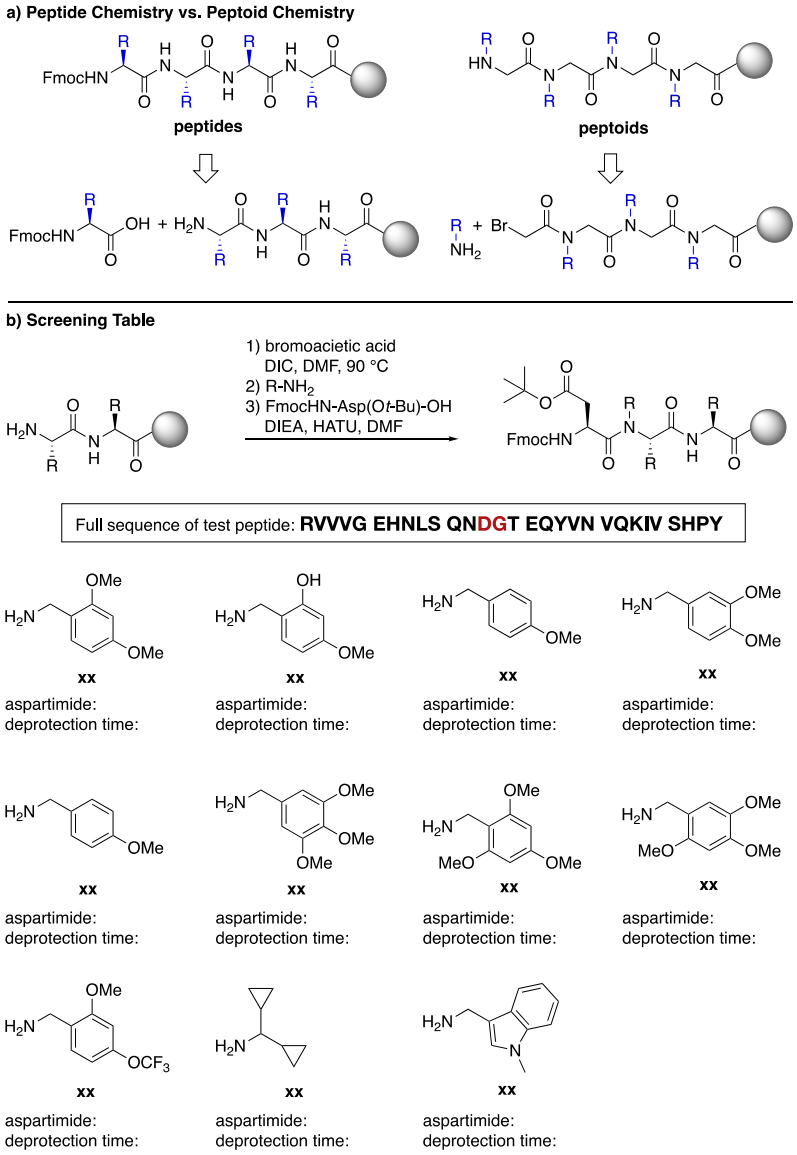


Figure 3.24: Proposed synthetic scheme for in-flow backbone protection with peptoid chemistry

3.7 Materials and Methods

This section contains the materials and methods used in this chapters work. Please note that these methods may not apply to other chapters

3.7.1 Reagents and solvents

All reagents were purchased and used as received. Fmoc-protected amino acids (Fmoc-Ala-OHxH_2O, Fmoc-Arg(Pbf)-OH; Fmoc-Asn(Trt)-OH; Fmoc-Asp-(Ot-Bu)-OH; Fmoc-Cys(Trt)-OH; Fmoc-Gln(Trt)-OH; Fmoc-Glu(Ot-Bu)-OH; Fmoc-Gly-OH; Fmoc-His(Trt)-OH;

Fmoc-Ile-OH; Fmoc-Leu-OH; Fmoc-Lys(Boc)-OH; Fmoc-Met-OH; Fmoc-Phe-OH; Fmoc-Pro-OH; Fmoc-Ser(But)-OH; Fmoc-Thr(t-Bu)-OH; Fmoc-Trp(Boc)-OH; Fmoc-Tyr(t-Bu)-OH; Fmoc-Val-OH), N-α-Fmoc-L-α-aminobutyric acid (Fmoc-Abu-OH) and N-α-Fmoc-L-norleucine (Fmoc-Nle-OH) were purchased from the Novobiochem-line from Sigma Millipore; FmocHis(Boc)-OH was purchased from ChemPep, Inc.; Fmoc-p-bromo-Phe-OH was purchased from Bachem (product number 4042637); O-(7-azabenzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate (HATU, ≥97.0%), N,N,N′,N′-tetramethyl-O-(1H-benzotriazol-1-yl)uronium hexafluoro-phosphate (HBTU, ≥97.0%), O-(1H-6-Chlorobenzotriazole-1-yl)-1,1,3,3- tetramethyluronium hexafluorophosphate (HCTU, ≥97.0%) and (7-azabenzotriazol-1- yloxy)tripyrrolidinophospho-nium hexa-fluorophosphate (PyAOP, ≥97.0%) were purchased from P3 Biosystems. Biosynthesis OmniSolv® grade N,N-dimethylformamide (DMF) was purchased from EMD Millipore (DX1732-1). N-Methyl-2-pyrrolidone (NMP, ≥99.0%) was purchased from Sigma-Aldrich and dried over PPT Pure Process Technology solvent system. AldraAmine trapping agents (for 1000 – 4000 mL DMF, catalog number Z511706), Diisopropylethylamine (DIEA; 99.5%, biotech grade, catalog number 387649), piperidine (ACS reagent, ≥99.0%), trifluoroacetic acid (HPLC grade, ≥99.0%), triisopropylsilane (≥98.0%), acetonitrile (HPLC grade), formic acid (FA, ≥95.0%), dimethyl sulfoxide (DMSO, HPLC grade, ≥99.7%), piperazine (≥99.0%) and 1,2- ethanedithiol (EDT, GC grade, ≥98.0%) were purchased from Sigma-Aldrich. Tetrafluoroethylene (TFE, extra pure, ≥99.8%) was purchased from Acros. Anisole (purum, GC grade, ≥99.0%) and 1,8-diazabicyclo(5.4.0)undec-7-ene (DBU, GC grade, ≥99.0%) were purchased from Fluka. 1- Hydroxybenzotriazole hydrate (HOBt, ≥97.0%) was purchased from Peptides International. HRink Amide (0.49 mmol/g and 0.18 mmol/g loading) and HMPB ChemMatrix polyethylene glycol (0.45 mmol/g loading) resin were purchased from PCAS Biomatrix. Water was deionized using a Milli-Q Reference water purification system (Millipore). Nylon 0.22 µm syringe filters were TISCH brand SPEC17984.

3.7.2 Automated fast-flow peptide synthesizer set-up

All peptides were synthesized on an automated-flow system, which was built in the Pentelute lab ("Amidator") (19), which is similar to the published AFPS system.(17) Capitalized letters refer to L-amino acids, uncapitalized letters refer to D-amino acids.

Unless otherwise noted, the following settings were used for peptide synthesis: flow-rate = 40 mL/min, temperature = 90 °C (loop) and 85–90 °C (reactor). The 50 mL/min pump head

pumps 400 µL of liquid per pump stroke; the 5 mL/min pump head pumps 40 µL of liquid per pump stroke. The pump refill time is 600 ms. The standard synthetic cycle involves a first step of prewashing the resin at elevated temperatures for 60 s at 40 mL/min. During the coupling step, three HPLC pumps are used: a 50 mL/min pump head pumps the activating agent, a second 50 mL/min pump head pumps the amino acid and a 5 mL/min pump head pumps DIEA.

To start the machine a filtered syringe cartridge loaded with resin is mounted on an automated robot arm which moves the syringe into the reactor. Prior to reactor closing DMF is pumped into the syringe for 8 pumping strokes, to swell the resin. standard synthetic cycle involves a first step of prewashing the resin at elevated temperatures for 60 s at 40 mL/min. The first two pumps are activated for 8 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 8 pumping strokes, after which the DIEA pump is shut off and the other two pumps continue to wash the resin for another 40 pump strokes.

During the deprotection step, two HPLC pumps are used. Using a rotary valve, one HPLC pump selects deprotection stock solution and DMF. The pumps are activated for 13 pump strokes. Both solutions are mixed in a 1:1 ratio. Next, the rotary valves select DMF for both HPLC pumps, and the resin is washed for an additional 40 pump strokes. The coupling–deprotection cycle is repeated for all additional monomers. The final synthetic step is a postwash of the resin at elevated temperatures for 60 s at 40 mL/min. A full synthetic cycle is complete in 120 seconds.

3.7.3 Peptide cleavage protocol

Directly after synthesis, the resin was washed with dichloromethane (3 x 5 mL), dried in a vacuum chamber, and weighed. 50% of the resin was transferred into a 50 mL conical polypropylene tube. For cleavage of peptides without cystine residues and peptides/protein sequences with cystine residues we used two different protocols.

A) Peptides[89]: Approximately 3 mL of cleavage solution (94% TFA, 1% TIPS, 2.5% EDT, 2.5% water) was added to the tube. If needed, more cleavage solution was added to ensure complete submersion. The tube was kept at room temperature for 2 hours.

B) Proteins and Cys-containing peptides:[90] Approximately 5 mL of cleavage solution (82.5% TFA, 5% water, 5% phenol, 5% thioanisole, 2.5% EDT) was added to the tube. If needed,

more cleavage solution was added to ensure complete submersion. The tube was kept on a nutating mixer at room temperature for 4 h.

After incubation ice cold diethyl ether (45 mL) was added to the cleavage mixture and the precipitate was collected by centrifugation and triturated twice more with cold diethyl ether (45 mL). The supernatant was discarded at each step. Residual ether was allowed to evaporate for 20 minutes after which the peptide was dissolved in 50% acetonitrile in water with 0.1% TFA (long peptides were dissolved 70% acetonitrile in water with 0.1% TFA). The peptide solution was filtrated with a Nylon 0.22 µm syringe filter and frozen, lyophilized until dry, and weighed.

3.7.4 Liquid chromatography-mass spectrometry (LC-MS)

For mass analysis of peptides, a filtered peptide solution (10 µL of a 1mg/mL solution) was diluted in 50% acetonitrile in water with 0.1% TFA (90 µL) to a final concentration approximately 0.1 mg/mL. LC-MS chromatograms and associated high resolution mass spectra were acquired using an Agilent 6550 iFunnel Q-TOF LC-MS system (abbreviated as 6550). Solvent compositions used in the LC-MS are 0.1% formic acid in H2O (solvent A) and 0.1% formic acid in acetonitrile (solvent B).

The following LC-MS methods were used:

A) 1-61% B over 15 min, Phenomenex Jupiter C4 column (6550)
   a. LC conditions: Phenomenex Jupiter C4 column: 1.0 × 150 mm, 5 µm, column temperature: 40 °C, gradient: 0-2 min 1% B, 2-12 min 1-61% B, 12-16 min 61-90% B; flow rate: 0.1 mL/min. A final 4-min hold was performed at a flow rate of 0.1 mL/min. The total method time was 20 min. MS is on from 4 to 12 min.
   b. MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 100–1700 m/z.

B) 1-61% B over 33 min, Phenomenex Jupiter C4 column (6550)
   a. LC conditions: Phenomenex Jupiter C4 column: 1.0 × 150 mm, 5 µm, column temperature: 40 °C, gradient: 0-2 min 1% B, 2-30 min 1-91% B, 30-34 min 61-90% B; flow rate: 0.1 mL/min. A final 4-min hold was performed at a flow rate of 0.1 mL/min. The total method time was 38 min. MS is on from 4 to 30 min.
   b. MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 100–1700 m/z.

C) 1-61% B over 15 min, Zorbax C3 column (6550)

a. LC conditions: Zorbax 300SB-C3 column: 2.1 × 150 mm, 5 μm, column temperature: 40 °C, gradient: 0-2 min 1% B, 2-11 min 1-61% B, 11-12 min 61-95% B, flow rate: 0.8 mL/min. A final 3-min hold was performed at a flow rate of 0.8 mL/min. The total method time was 15 min. MS is on from 4 to 12 min.
b. MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 100–1700 m/z.

D) 1-91% B over 20 min, Phenomenex Jupiter C4 column (6550)
a. LC conditions: Phenomenex Jupiter C4 column: 1.0 × 150 mm, 5 μm, column temperature: 40 °C, gradient: 0-2 min 1% B, 2-18 min 1-91% B, 18-21 min 91% B; flow rate: 0.1 mL/min. A final 4-min hold was performed at a flow rate of 0.1 mL/min. The total method time was 25 min. MS is on from 4 to 18 min.
b. MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 100–1700 m/z.

E) 1-61% B over 17 min, Luna C18 column (6550)
a. LC conditions: Phenomenex Luna C18 column: 0.5 × 150 mm, 5 μm, column temperature: 40 °C, gradient: 0-2 min 1% B, 2-14 min 1-61% B, 14-18 min 61-91% B; flow rate: 0.1 mL/min. A final 5-min hold was performed at a flow rate of 0.1 mL/min. The total method time was 23 min. MS is on from 4 to 14 min.
b. MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 100–1700 m/z.

F) 1-61% B over 9 min, Zorbax C3 column (6520)
a. LC conditions: Zorbax 300SB-C3 column: 2.1 × 150 mm, 5 μm, column temperature: 40 °C, gradient: 0-2 min 1% B, 2-11 min 1-61% B, 11-12 min 61-95% B, flow rate: 0.8 mL/min. A final 3-min hold was performed at a flow rate of 0.8 mL/min. The total method time was 15 min. MS is on from 4 to 12 min.
b. MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 300–3000 m/z.

G) 1-91% B over 9 min, Zorbax C3 column (6520)
a. LC conditions: Zorbax 300SB-C3 column: 2.1 × 150 mm, 5 μm, column temperature: 40 °C, gradient: 0-2 min 1% B, 2-11 min 1-91% B, 11-12 min 91-95%

B; flow rate: 0.8 mL/min. A final 3-min hold was performed at a flow rate of 0.8 mL/min. The total method time was 15 min. MS is on from 4 to 12 min.

b. MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 300–3000 m/z.

Data were processed using Agilent MassHunter Workstation Qualitative Analysis Version B.06.00 with BioConfirm Software.

3.7.5 Analytical high-performance liquid chromatography (HPLC) protocol

For determination of purity by HPLC, the filtered peptide solution was diluted in 50% acetonitrile in water with 0.1% TFA (100 µL) to a final concentration of approximately 1.0 mg/mL.

For standard analysis of all peptide and protein sequence samples, analytical HPLC spectra were recorded on an analytical Agilent Zorbax 300SB-C3 column (2.1 mm × 150 mm, 5-µm particle size). A linear gradient of acetonitrile with a 0.08% TFA additive (solvent B) in water with a 0.1% TFA additive (solvent A) was used. After a 3-min hold, gradients of 1% B per minute ramped up over 60 min at a flow rate of 0.4 mL/min. Gradients either started at 1% B (annotated as "1–61% B over 60 min") or 5% B (annotated as "5–65% B over 60 min"). A final 3-min hold was performed. The total method time was 66 min. Crude HPLC purities were determined by manual integration of all signals in the area of 5−60 min.

For analysis of epimerization, analytical HPLC spectra were recorded on an analytical Agilent Zorbax 300SB-C18 column (2.1 mm × 150 mm, 5-µm particle size). A linear gradient of acetonitrile with a 0.08% TFA additive (solvent B) in water with a 0.1% TFA additive (solvent A) was used. After a 3-min hold, a gradient of 2% B per minute ramped up over 20 min at a flow rate of 0.4 mL/min (annotated as "1–41% B over 20 min"). A final 3-min hold was performed. The total method time was 30 min.

3.7.6 Mass-directed reversed-phase high performance liquid chromatography (RP-HPLC)

For RP-HPLC purification, the lyophilized peptide sample was dissolved in the gradient starting concentration (e.g., 5% acetonitrile in water with 0.1% TFA) or 6 M Guanidinium chloride containing 100 mM DTT. All samples filtrated with a Nylon 0.22 µm syringe filter prior to purification.

For all HPLC purifications, a gradient of acetonitrile with 0.1% TFA additive (solvent B) and water with a 0.1% TFA additive (solvent A) was used. All samples were purified on a

semipreparative Agilent Zorbax 300SB-C3 column (9.4 mm × 250 mm, 5-µm particle size) at a flow rate of 4 mL/min unless otherwise noted.

Specific purification conditions such as column temperature and gradient are specified for each case. In order to optimize the purification for the longer peptide chains the procedure below was followed:

A 60-minute analytical run was performed prior to the purification. 1–2 mg of crude material from 900 µL of solvent was injected on a semipreparative Zorbax 300SB-C3 column at 60 °C (gradient: 5–65% B with 1% B/min). The method was then adjusted according to the elution profile.

The gradient was changed to 0.2% B around the B concentration at which the product eluted (e.g., if the compound eluted at 39% B we adjusted the method to: < 50 mg of crude on a semipreparative Zorbax 300SB-C3 column at 60 °C (gradient: 5-29% B with 1% B/min, 29-49% B with 0.2% B/min, 49-65% B with 1% B/min). These conditions were used unless otherwise noted.

3.7.7 Protein concentration determination protocol

The concentration of proteins in solution was determined by absorbance at 280 nm (A280) or bicinchoninic acid (BCA) assay. The detailed protocol is as followed:

- A280: The concentration of proteins was calculated using the Beer-Lambert law by measuring the absorbance of the protein sample at 280 nm. The absorbance was measured by averaging at least two independent readings of the same sample on a BioTek Synergy HT plate reader outfitted with a BioTek Take 3 micro-volume plate. Unless indicated otherwise, the molar extinction coefficient of the protein of interest was estimated based on the sequence of the protein via ExPASy Swiss Institute of Bioinformatics - Bioinformatics Resource Portal.
- BCA assay: The concentration of proteins was determined by the Pierce™ Rapid Gold BCA Protein Assay Kit from ThermoFisher Scientific (catalog number A53226) following the manufacturer's instructions. Briefly, nine standards of Albumin protein (BSA) at different concentrations was prepared in addition to at least two dilutions of the desired protein, in triplicates. In a 96-well plate, 10 µL of each of the protein samples were added. 200 µL of the Pierce Rapid Gold BCA Working Reagent was then added to the wells using a multichannel pipette. The plate was incubated at room temperature for 5 minutes. The

absorbance at 480 nm was then measured using a BioTek Synergy HT plate reader. Finally, the concentration of the desired protein was calculated based on the standard curve generated with the BSA standards, averaged and reported.

3.7.8 Determination of peptide and protein yield

Molecular weight of peptide sequences was determined via ChemDraw, accounting for the weight of a TFA counter-ion for each basic residue (K, R, H) in addition to the N-terminal amine.

For example, for a peptide with sequence "KALE" the molecular weight of the peptide as TFA salt is calculated as 916 g/mol (= 688 + 2 × 114).

The weight of lyophilized powders of the peptides was directly measured using analytical scales (XS205DU Analytical Balance, Mettler-Toledo) [note: use of deionizers such as SPI Westek Workstation Still Air Ionizer helps with measurements]. Following folding, protein concentration was measured based on the outlined procedures above.

Theoretical yield was determined based on weight of the resin, resin loading, and the molecular weight (with TFA) of each peptide.

For example, for the KALE sequence synthesized on 50 mg resin with 0.44 mmol/g loading, theoretical yield is:

theoretical yield = 0.44 mmol/g × 50 mg × 916 g mol = 20 mg

Yield of crude peptide was determined based on the ratio of weight of lyophilized crude peptide (as TFA salt) to theoretical yield multiplied by the purity determined by UV absorption at 280 nm (analytical HPLC).

In the example above, if 10 mg of crude KALE peptide is produced and the purity by analytical HPLC is 50%,

synthesis yield is: yield = 10 mg/20 mg × 0.50 × 100 = 25%

Yield past HPLC was calculated based on weight ratio of HPLC-purified peptides to theoretical yield (both as TFA salts).

In the example above, if 4.0 mg of KALE peptide is isolated from HPLC, the yield of the synthesis past HPLC is:

yield = 4 mg/20 mg × 100 = 20%

Folding and chromatography (size exclusion, ion exchange, etc.) yield was calculated based on concentration of protein in solution and the mass of lyophilized purified peptide (as TFA salt).

In the example above, if 1 mL of 1 mg/mL solution of KALE peptide is obtained from folding & size exclusion chromatography of 4 mg of HPLC-purified KALE, the purification yield is:

purification yield = (1 mg⁄mL × 1 mL)/4 mg × 100 = 25%

3.7.9 Pre-optimization AFPS conditions

At the beginning of the optimization process the AFPS system, Amidator had the following conditions:

- Temperature 85–90 °C in reactor, 70 °C in 10’ activation loop
- Flow Rate 40 mL/min
- Coupling step
    - 0.20 M amino acids stock solution in NMP
    - 0.19 M activator stock solution in DMF
    - DIEA (99.7%) as coupling base
- Coupling conditions:
- R was coupled with PyAOP
- all other amino acids were coupled with HATU (13 pump strokes)
- Deprotection step
    - 40% piperidine in DMF (13 pump strokes)
- Washing steps
    - DMF (40 pump strokes)

In the synthesis process amino acid, activator and base are mixed in a 10:10:1 ratio in the coupling step and deprotection solution and DMF are mixed in a 1:1 ratio during the deprotection step.

3.7.10 Final AFPS conditions

Following the optimization process discussed in section 3.2 the AFPS system the Amdiator was used under the following conditions.

- Temperature
    - 85–90 °C in reactor,
    - 60 °C in 5’ activation loop (for C and H),
    - 90 °C in 10’ activation loop (for all other amino acids)
- Flow Rate 40 mL/min

- Coupling step
    - 0.40 M amino acids stocks in amine-free DMF
    - 0.38 M activator stocks in amine-free DMF
- Coupling conditions:
    - HATU (13 pump strokes) except S&A w/ HATU (26 pump strokes)
    - C, H, N, Q, V, R, T w/ PyAOP (26 pump strokes)
- Deprotection step
    - 40% pip in amine-free DMF with 2% formic acid (13 pump strokes)
- Washing steps
    - Amine-free DMF (40 pump strokes

In the synthesis process amino acid, activator and base are mixed in a 10:10:1 ratio in the coupling step and deprotection solution and DMF are mixed in a 1:1 ratio during the deprotection step.

3.7.11 Synthesis conditions for commercially available synthesizers

All proteins were synthesized at Novo Nordisk using PurePep™ amino acids (Gyros Protein Technologies) on a commercially available peptide synthesizer at room temperature and at elevated temperatures. A standard peptide synthesis protocol used at Novo Nordisk was applied for all syntheses. For coupling, ethyl cyanohydroxyiminoacetate (Oxyma) (0.3 M) and amino acid (0.3 M) were premixed, before N,N′-diisopropylcarbodiimide (DIC, 3.0 M) was added. For the room temperature synthesis Collidine (3.0 M) was added as activator. 0.1 M Oxyma in 20% Piperidine in DMF was used for deprotection.

For room temperature synthesis (including capping), the general protocol included the following steps with a total 3 hours and 37 minutes needs to incorporate one amino acid:

step	Volume (uL)	Mixing time (min)	Heated?	Repeat
Deprotection	6000	10	-	2
DCM wash	1500	0	-	1
DMF wash	6000	0	-	6
Amino acid/Oxyma	5000	0	-	1
DIC	1000	0	-	1
Collidine	1000	0	-	1
Mix and coupling	-	120	-	1
Drain/Dry	-	0	-	1
DCM wash	300	0	-	1
Mix	-	9	-	1
Cap	5000	0	-	1
Collidine	1000	0	-	1
Mix	-	20	-	1
Drain/Dry	-	0	-	1
DMF	6000	0	-	3
DCM	1500	0	-	1
DMF	6000	0	-	1

Table 3.7: Conditions for room temperature synthesis with commercially available synthesizers

For 70 and 90 °C syntheses, the general protocol included the following steps, with each amino acid incorporation take 26 minutes:

step	Volume (uL)	Mixing time (min)	Heated?	Repeat
Deprotection	3000	1	-	1
DMF wash	1000	0	-	1
Deprotection	3500	2	x	1
DMF wash	1000	0	-	4
DCM wash	3000	0	-	1
DMF wash	3000	0.3	-	3
Amino acid/Oxyma	3000	0	-	1
DIC	300	0	-	1
DCM wash	300	0	-	1
Mix	-	9	x	1
Drain/Dry	-	0	-	1
DCM wash	3000	0	-	1
DMF wash	1000	0	-	4
DMF wash	3000	0.3	-	2

Table 3.8: Conditions for heated synthesis with commercially available synthesizers

3.7.12 Expression and purification of recombinant barnase

Barnase was expressed based on previous report from Okorokov et al.(65) In short, E. coli strain XL-1 Blue was transformed with plasmid pMT1002 (Addgene plasmid 8621). Next, 1 mL of 8-h culture of XL-1 Blue cells carrying the plasmid grown in LB medium with carbenicillin (100 µg/mL), was diluted 1:1000 into the same medium (1 L) and grown at 28 °C, shaking at 200 rpm. The culture had a density of ~OD600 = 0.6 in approximately 24 h, at which point a further 500 mL of preheated LB medium (85 °C) was poured into the cell culture, and the shaker was incubated at 42 °C for 30 minutes. Next, the temperature was adjusted to 37 °C and cells were cultured for another 18 hours. Acetic acid was added to the culture until a final pH of ~4.3 as determined by pH paper. The cells were centrifuged for 30 minutes at 8000 rpm. The pellet was discarded while the supernatant was divided into 4000 mL portions, flash frozen with liquid nitrogen and kept at -80 °C until purification. For purification, 400 mL of the supernatant was thawed on ice and buffer-exchanged into buffer SPA (50 mM MES, pH 6.3) using Amicon 3K concentrator (15 mL, EMD Millipore). The resulting solution was loaded onto two Capto S

columns attached in series (5 mL, GE Healthcare) and the desired protein was eluted with a gradient of SPB (50 mM MES, pH 6.3, 1 M NaCl) in SPA. The fractions containing the desired protein were combined and concentrated using Amicon 3K concentrator (15 mL, EMD Millipore). The protein was then filtered using a 0.22 µm filter and purified via size exclusion chromatography with buffer P (20 mM Tris, pH 7.5, 150 mM NaCl). The fractions containing pure barnase were combined and concentrated using Amicon 3K concentrators and the molecular weight of the final protein was determined by LC-MS. Overall, 3 mg of recombinant barnase was isolated from 400 mL of the culture supernatant.

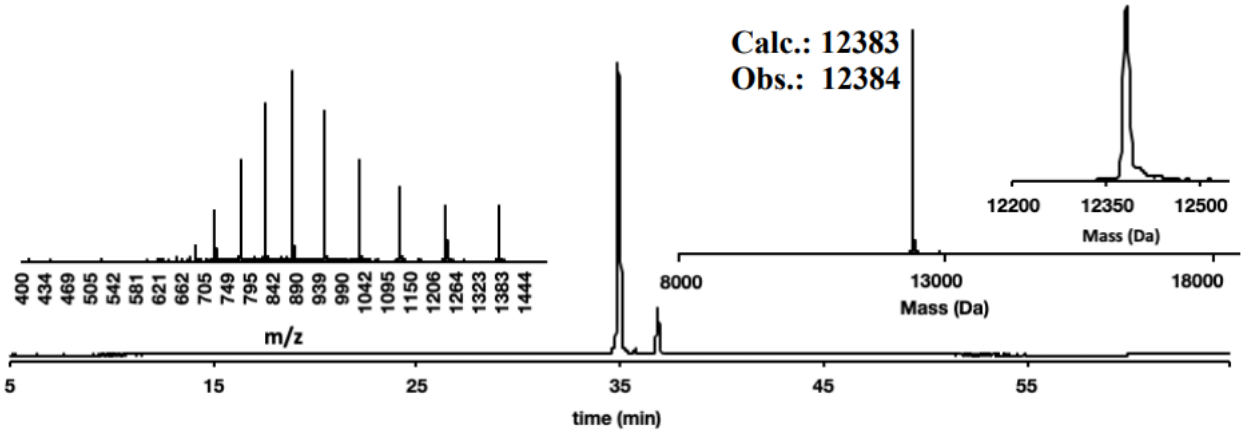


Figure 3.25: Recombinant barnase HPLC and LCMS traces

3.7.13 Comparison of timeline for expression vs. synthesis

Timeline for recombinant expression of barnase: Transformation of XL-1 Blue cells with barnase plasmid (1 h), growth of starter culture (8 h), growth of bacterial cultures (24 h), heat shock and initiation of protein expression (19 h), acidification and centrifugation (1–3 h based on the scale), buffer exchange (3–6 h based on the scale), ion exchange chromatography (4 h), fraction analysis and concentration (6 h), size exclusion chromatography (5 h), fraction analysis and concentration (6 h). Total 77–82 h. Timeline for production of synthetic barnase: Fast flow synthesis (5 h), cleavage from resin, global deprotection and ether precipitation (5 h), lyophilization (10–16 h), HPLC purification (3 h), lyophilization and fraction analysis (done in parallel 10–16 h), folding and size exclusion chromatography (5 h), fraction analysis and concentration (6 h). Total 44–56 h.

3.7.14 RNA hydrolysis assay for barnase and barstar

Hydrolysis of RNA by synthetic barnase proteins was used to assess the activity of these proteins. Substrate RNA* was purchased from Millipore-Sigma as PAGE-purified dry solid. The RNA was dissolved in a solution of 10 mM Tris, 1 mM EDTA, pH 7.7 according to the

manufacturer's recommendation and stored in 0.5 µg/µL aliquots at -80 °C. RNA hydrolysis was initiated by addition of the RNA to a solution of barnase variants (recombinant wild type, synthetic wild type, synthetic R110FBr) in 10 mM Tris, 1 mM EDTA, pH 7.7 buffer, and left at room temperature for 20 minutes. The total reaction volume was 10 µL, with 7 µM final concentration of RNA and 3 nM concentration of barnase. The reactions were quenched at the time points by addition of 10 µl of TBE-Urea gel loading dye (LC6876, Thermofisher Scientific) and rapidly frozen on liquid nitrogen.

Barnase activity was inhibited by its known inhibitor barstar to ensure any RNase activity observed in the assay was barnase-specific. Prior to the introduction of RNA substrate, the barnase variants (recombinant wild type, synthetic whild type, synthetic R110FBr) were incubated with either recombinant barstar (Fig. S14), or synthetic barstar, for 10 minutes on ice. The final concentration of RNA in the reaction was 7 µM and the concentration of barnase was 3 nM. The concentration of recombinant barstar incubated in each reaction was either 0 µM, 4 µM, 9 µM or 18 µM. The concentration of synthetic barstar incubated in each reaction was either 0 nM, 3.5 nM, 35 nM or 350 nM.

Immediately prior to gel analysis, the quenched reaction mixtures were heated at 95 °C for 5 minutes. Analysis was done on 15% denaturing polyacrylamide gel (Novex™ TBE-Urea Gels, EC68855BOX, Thermofisher Scientific) using 180 V for 60 minutes. The gel was stained with ethidium bromide at 1 µg/mL for 30 minutes and then washed with water three times. The gel was visualized on the ChemiDoc gel imager (Bio-Rad).

(*) sequence of substrate RNA – bolded are the potential sites of cleavage by barnase.
CAACAUCUU**G**CUAUACAAU**G**CCAAUCCAU**G**CUACACUAC**G**UUACA

3.7.15 Extended folding and SEC purification of barnase

The folding procedure of wild type barnase was adapted from Mong et al.(18) 2.5 mg of wild type barnase was dissolved in 150 µL of 6 M Gn·HCl solution in 50 mM NaH2PO4 RNase free1 buffer, pH = 7.8. The mixture was serially diluted in four steps to 0.3 M Gn·HCl using a buffered RNase free solution (150 mM Tris buffer pH 7.5, 150 mM NaCl). The final solution was then filtered and subjected to size exclusion chromatography (SuperdexTM 75 Increase 10/300 GL, 0.25 mL/min), using isocratic eluent 150 mM Tris buffer pH 7.5, 150 mM NaCl, RNase free. The elution profile is depicted below.

Fractions containing the protein were concentrated using a 3K molecular weight cut off spin filter, flash frozen using liquid nitrogen, and stored at −80 °C. A total of 0.8 mg of wild type barnase was isolated, corresponding to a 33% isolated yield. The purity of the final product was assessed by LCMS and HPLC.

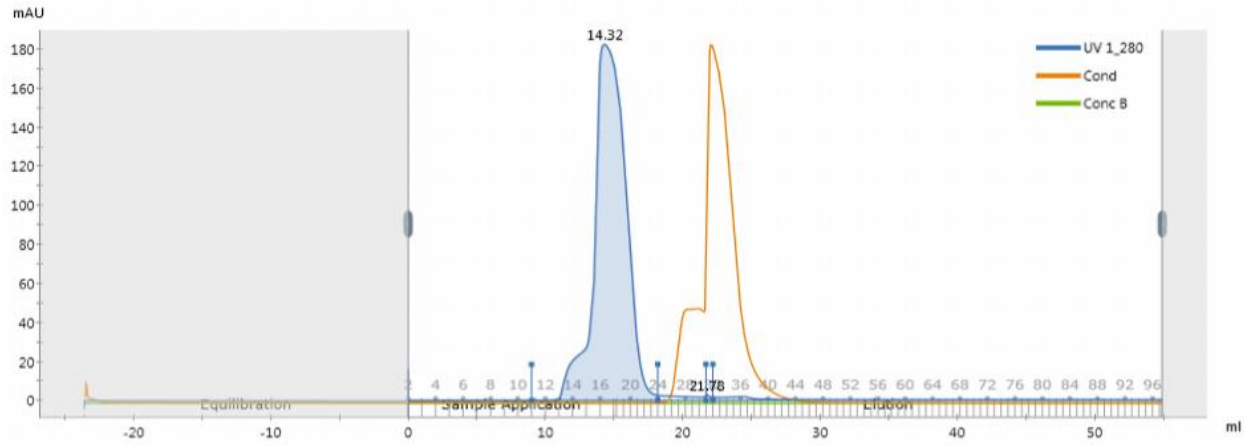


Figure 3.36: SEC purification trace of barnase

3.7.16 Extended folding and IEC purification of barstar

The folding procedure of wild type Barstar was adapted from Shastry et al.(66) 3.0 mg of HPLC-purified wild type barstar was dissolved in 1050 µL of a 6 M Gn·HCl solution in 5 mM NaH2PO4, 250µM EDTA, 250 µM DTT RNase free buffer, pH 8. The mixture was serially diluted in ten steps to 0.6 M Gn·HCl using a buffered RNase free solution (5 mM NaH2PO4, 250µM EDTA, 250 µM DTT, pH 8). The resulting solution then desalted using a 7 K Zeba™ Spin Desalting Column and filtered, and subjected to anion exchange chromatography (GE Healthcare HiTrap™ Q Sepharose Fast Flow IEX Column), using gradient elution with RNase free mobile phases 20 mM Tris buffer pH 7.4 and 20 mM Tris, 1 M NaCl pH 7.4 (gradient: 0-45% at 0.5% B/min). The elution profile is depicted below.

Fractions containing the protein were concentrated using a 3K molecular weight cut off spin filter, flash frozen using liquid nitrogen, and stored at −80 °C. A total of 0.14 mg of wild type barstar was isolated corresponding to 5% isolated yield. The purity of the final product was assessed by LC-MS and HPLC.

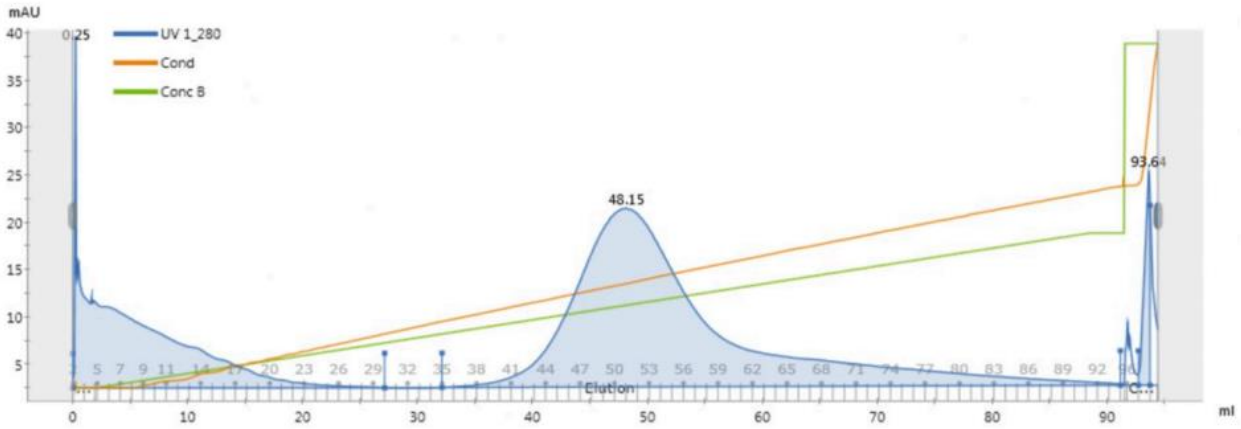


Figure 3.37: IEC purification trace of barstar

3.7.17 Fluorogenic RNase activity assay

Hydrolysis of a fluorogenic substrate (6-FAM-dA^{D}rG^{D}dA^{D}dA^{D}-6-TAMRA) by the recombinant and synthesized barnases can be used to calculate the kcat/KM of the proteins based on the first order rate equation (1).

$$I = I_f - (I_f - I_0)e^{-\frac{k_{cat}}{K_M}[E](t-t_0)} \quad (1)$$

In this equation, I is the fluorescence at time t, I0 is fluorescence of the intact substrate, If is the fluorescence of the hydrolyzed substrate, [E] is the total enzyme concentration, kcat, KM are steady state enzyme kinetic parameters.(50, 55) The fluorogenic substrate for the assay was purchased from ChemGenes as HPLC-purified solid. The substrate was dissolved in assay buffer solution (100 mM MES, 100 mM NaCl, pH 6.0) and kept in 200 µM aliquots at –80 °C until use.

In each assay, the substrate was diluted in the assay buffer and the fluorescence of the substrate solution was monitored at 515 nm upon excitation at 495 nm using Tecan plate reader M1000. The fluorescence was monitored every 10 seconds, after 3 second shake, for at least 500 seconds to measure the starting fluorescence, I0, and to ensure no background cleavage took place prior to the addition of enzymes. To start the cleavage reaction, an aliquot of enzyme was added

to the substrate and the solution was rapidly mixed. The increase in fluorescence was then monitored for at least 1000 seconds. The concentration of the substrate in the reaction mixture at the time of addition of barnase was 200 nM and barnase was approximately at 2 nM. Total reaction volume was 250 µL.

The experiment was performed in triplicates. Blank runs were performed in parallel with the experimental runs. For the blank runs, assay buffer was added to wells instead of barnase. Blank runs were otherwise set up identical to the experimental runs. The fluorescence readings from experimental runs were first subtracted from the fluorescence readings of their blank. The data was normalized and then analyzed using the nonlinear regression, exponential, plateau followed by one phase association analysis using Prism 8 software from Graphpad. The K generated from this analysis corresponds to kcat [E]/KM based on equation 1 above. The exact concentration of protein in each case was determined via absorbance at 280 nm using reported molar extinction coefficient values for barnase, 27411 M-1 cm-1 . (67)

Accordingly, kcat/KM (mean ± SE) of the synthetic and recombinant wild-type barnase were determined to be (7.6 ± 0.2) × 106 M-1 s-1 and (9.0 ± 0.3) × 106 M-1 s-1 , respectively. The values match the literature reported kcat/KM of wild-type recombinant barnase (1.3 ± 0.4) × 107 . (65) Furthermore, the kcat/KM ± SE of the R110FBr synthetic variant barnase was calculated to be (2.3 ± 0.8) × 106 M-1 s-1 .

3.7.18 Chemical denaturation assay for barnase

The chemical denaturation profile of barnase was obtained by fluorescence spectroscopy. In this assay, a solution of each barnase variant (recombinant wild type, synthetic wild type, synthetic R110FBr) was prepared at 1 µM in either buffer A (50 mM MES, pH 6.3) or buffer B (8 M urea, 50 mM MES, pH 6.3). In a 384-well plate, using the two stock solutions, mixtures of each barnase variant was prepared at 25 urea concentrations ranging 0 – 8 M (at 0.33 M steps). The final concentration of barnase in each case was 1 µM while the buffer was 50 mM MES at pH 6.3. The mixtures were prepared in triplicates. The solutions were equilibrated at room temperature for two hours. Then the emission at 315 nm upon excitation at 290 nm was recorded. The transition midpoint ([D]50% - the concentration of urea at which half of the sample is unfolded) and m-values (the slope of the unfolding transition) were determined by fitting of the normalized fluorescence readings as described (47, 67), using the Prism 8 software from Graphpad.

3.7.19 Folding HIV-1 protease

The procedure for folding HIV Protease was taken from Johnson et al.(31) Lyophilized HIV-1 protease (2.5 mg) was dissolved in 2 mL of denaturing buffer (6 M guanidine hydrochloride, 200mM NaPi, pH 7.4). A dialysis cassette (Slide-A-Lyzer™ 3.5MWCO, Sigma Aldrich), was pretreated for 2 min. with dialysis buffer A (50 mM NaOAc pH 5.6), before the denatured HIV-1 protease sample was injected. The dialysis cassette was placed in a vessel containing 350 mL dialysis buffer A and left to stir at room temperature for 3 h. The cassette was then transferred to dialysis buffer B (350 mL of 10 mM NaOAc pH 5.6 prechilled to 4 °C) and left at 4 °C for 14 h. The protein containing buffer was removed from the cassette, filtered by a Pall™ 0.22 μm PTFE syringe filter, and stored in 100 uL aliquots at −80 °C. For all further assays, frozen fractions were used within 1 hour after defrosting and excess from the aliquot was discarded after use. Concentration of the resulting sample was determined using a Pierce™ Rapid Gold BCA Protein Assay Kit (supplier) of the undiluted stock HIV-1 Protease in triplicate compared to the BSA standard. Stock was found to contain 0.41 mg/mL protein which corresponds to a refolding yield of 50%. The concentration of active enzyme within the stock was determined by active site titration in the method of Windsor et al.(68) using darunavir and was found to be 76%.

3.7.20 Fluorogenic protease activity assay for HIV-1 protease

Assay conditions were taken from Johnson et al with slight adaptions taken from the original report for the assay from Toth et al.(31, 51) The fluorogenic substrate Abz-Thr-Ile-Nle-p-nitro-Phe-Gln-Arg-NH_2 (Abz = 2, aminobenzoic acid, Nle= norleucine, p-nitro-Phe = p-nitro phenylalanine) was purchased from Bachem (product no. 4030748.0005) and was used as a 4 mM stock in DMSO (concentration determined gravimetrically). Fluorescence time-courses were measured at 37 °C on a Tecan™ M1000 Pro plate reader (ex: 355 nm, em: 430 nm, bandwidths 5 nm) with a gain setting of 124 in a Costar™ black chimney-well 96-well plate. Assays were conducted in a final volume of 200 μL of 50 mM NaOAc pH 5.6 maximum 4.5% DMSO, substrate (5−200 μM), and protease (10 nM active dimer) with 3 replicates. Calibration of fluorophore was performed as described in Toth et al.(42), and quantitation of the initial rates were made within the linear range of the detector (40 μM cleaved substrate at a gain setting of 124) (3 replicates). Initial velocity data was used with less than 20% of total substrate cleavage. Kinetic constants were derived from fitting the initial velocities to the Michaelis-Menten equation using Graphpad™ Prism's nonlinear least-squares regression.

3.7.21 Substrate specificity assay for HIV-1 protease

The qualitative assay with model substrate p12nt was previously reported by Schneider et al.(30) The 20-mer substrate was synthesized on the Amidator using standard optimized peptide synthesis conditions. 21 mg of the peptide was purified on a C3 Zorbax semi-prep column (9.4 x 250 mm) with a gradient from 15% to 75% ACN over 62 minutes with a flow-rate of 4 mL/min (Fig. S21). The cleanest fraction affording 1 mg was used for all further assays, and analytical LC-MS data collected. p12NT (10 µg) was incubated with HIV-1 Protease (300 nM) in a total volume of 30 µL of 50 mM NaOAc pH 5.6 with 0.5 mg/mL BSA. Reactions were left at 37 °C for 14 h and quenched with snap-freezing in liquid N2 followed by storage at –80 °C. The crude reaction mixture was analyzed by LC-MS. Fragments of the substrate peptide corresponding to cleavage at the conserved site were observed with loss of full-length peptide. No peptides corresponding to non-specific cleavage of the substrate peptide nor the native BSA in solution were observed even after prolonged incubation of 16 h.

3.7.22 Folding and SEC purification of synthetic sortase A[59-206];P94S/D160N/K196T

Synthetic sortase A[59-206];P94S/D160N/K196T (1.8 mg) was dissolved in 150 µL of 6 M Gn·HCl denaturing solution in 50 mM Tris, 20 mM DTT, pH 7.5. Concentration was determined by A280 reading on plate reader and adjusted by the addition of Gn·HCl solution to a concentration of 8 mg/mL sortase A. (extinction coefficient of sortase A is 14440 M-1 cm-1).

The mixture was then serially diluted twenty-fold in nine steps utilizing 50 mM Tris, 150 mM NaCl, 20 mM DTT, pH 7.5 resulting in 0.3 M Gn·HCl and 0.4 mg/mL sortase A. The solution was left overnight (18 h) at room temperature. The solution was then concentrated and filtered and subjected to size exclusion chromatography (SuperdexTM 75 Increase 10/300 GL, 0.4 mL/min), using isocratic eluent 20 mM Tris, 150 mM NaCl, pH 7.5. The elution profile is depicted below. Fractions were analyzed by LCMS and the cleanest fractions were pooled and concentrated utilizing a 3K molecular weight cut off spin filter and stored at –80 °C. A total of 0.11 mg of sortase A was isolated corresponding to a 6% isolated yield. The purity of the final product was assessed by LCMS and HPLC.

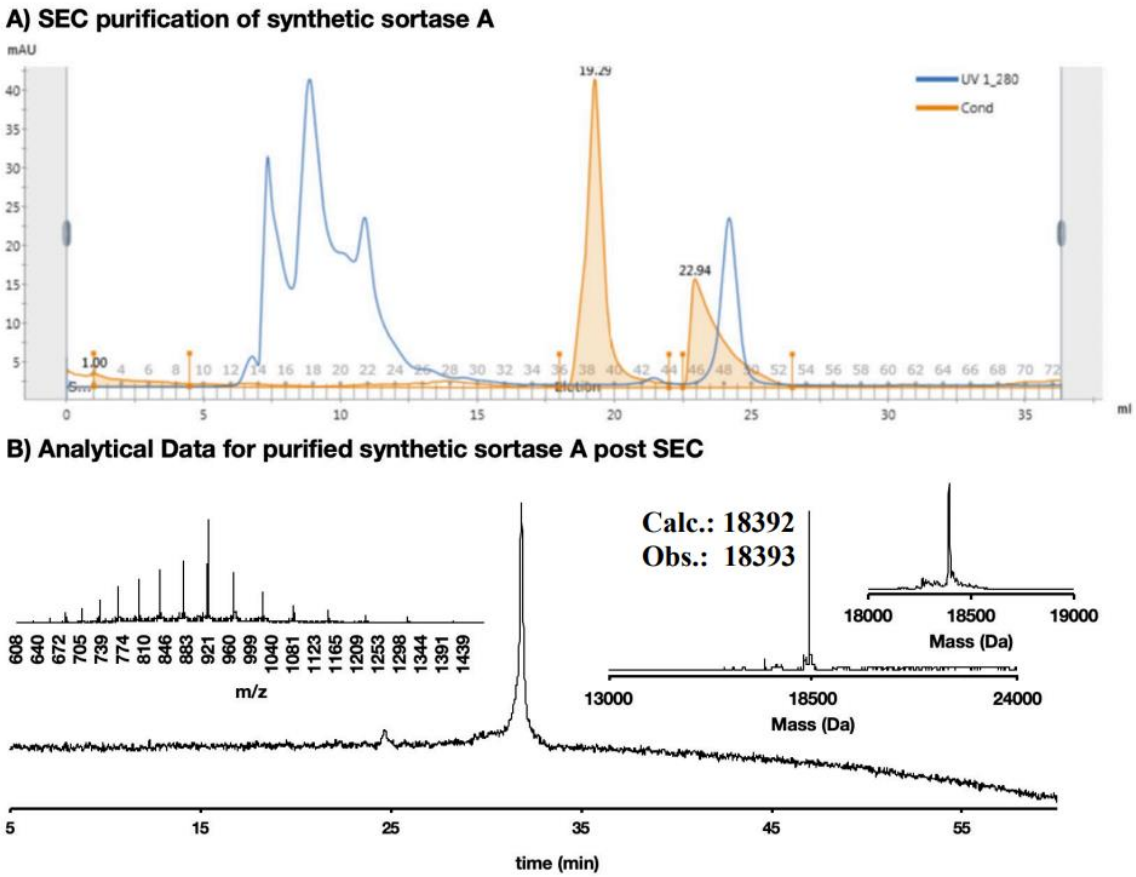


Figure 3.38: SEC elution profile and analytical data for sortase A*

3.7.23 Semiquantitative activity assay for sortase A*

Sortase A performs a transpeptidation reaction where the active enzyme cleaves the threonine-glycine bond in the LPXTG motif and ligates to a polyglycine. Therefore, an active sample of sortase A in the presence of peptides AQALPETGEE and GGGGGLY should generate the ligation product AQALPETGGGGGLY. Samples of folded recombinant sortase A, folded synthetic sortase A, and folded synthetic sortase A post-SEC were diluted to a concentration of 0.02 mg/mL in 50 mM Tris, 150 mM NaCl, 20 mM DTT, pH 7.5 as determined by A280 reading on a plate reader. The sortase A samples (15 µl) were then added to a reaction mixture containing 6 µL AQALPETGEE_(am) (1 mg/mL), 6 µl GGGGGLY_(am) (1 mg/mL) and 3 µL calcium chloride (50 mM). Mixtures were left at room temperature for 20 hours and quenched with 30 µL of a 2% trifluoroacetic acid solution with 50:50 (v/v) water: acetonitrile. Samples were analyzed by LCMS.

3.7.23 Folding of synthetic MDM2

Synthetic MDM2[1-118] was refolded according to the procedure from Zhan et al.(69) Lyophilized MDM2[1-118] (0.6 mg; 35 nmol) was dissolved in phosphate buffered saline (PBS) containing 6 M Guanidine hydrochloride (70 µL) and 20 mM DTT at pH 7.11. MDM2[1-118] concentration was determined by UV280 and adjusted to 150 µM (extinction coefficient of MDM2: 10430 m-1 cm-1). 50 uL of the resulting solution was serial diluted into 250 uL of folding buffer containing PBS and 20 mM DTT at pH 7.31, to give final conditions of 25 µM MDM2 in PBS containing 1 M guanidine hydrochloride and 20 mM. DTT. The resulting solution was kept at room temperature for 1 h before performing biolayer interferometry.

3.7.24 Preparation of biotinylated p53(15-29)

15-29P53-like peptide (SQETFSDLWKLLPEN) was synthesized by Fmoc-based SPPS, with H-Rink Amide-ChemMatrix resin (130 mg/synthesis; 0.06 mmol). After the Fmoc deprotection, peptide labeling with N-terminal D-biotin was performed on the resin bound protected peptides by treating the resin with a solution of Biotin-PEG4-propionic acid (ChemPep Inc., 0.75 mmol), HATU (0.38M in DMF; 1.87 mL; 0.71 mmol) and DIEA (1.5 mmol) for 1.5 h at room temperature. Upon completion, the resin was washed with DMF (5x) and DCM (5x) and dried under reduced pressure. The peptide was cleaved and processed using the standard protocol for peptide cleavage.

3.7.25 Assay of MDM2 binding activity with BLI

Synthetic MDM2 in refolding buffer (see section 6.4.1) and commercial MDM2 (Abcam 167941) were brought to 1 mg/mL bovine serum albumin (BSA), 0.02% Tween 20 by addition of 10 mg/mL BSA, 0.2% Tween 20 in PBS. The resulting solutions were diluted serially into 1 mg/mL BSA, 0.02% Tween 20 in PBS for BLI assay.

Biolayer interferometry was performed using an Octet Red96 system (ForteBio; Menlo Park, CA) and black, polypropylene, chimney well, flat-bottom 96 well plates (Greiner Bio-One, Kremsmünster, Austria). Wells were filled with 200 µL of the appropriate solution. Streptavidin biosensors (ForteBio) were equilibrated in 1 mg/mL BSA, 0.02% Tween 20 PBS buffer for at least 10 min prior to use. Sample plates were equilibrated at 30 ºC for 5 minutes before the start of an experiment, and kept at 30 ºC throughout. Sample plates were agitated at either 1000 or 1500 rpm throughout the assay.

The assay protocol was as follows:

1) 60 sec ‘baseline’ in 1 mg/mL BSA, 0.02% Tween 20, PBS buffer;

2) 120 sec ‘p53 immobilization’ in 1 mg/mL BSA, 0.02% Tween 20 PBS buffer containing ~400 nM biotin-p53(15-29);

3) 120 sec ‘baseline’ in 1 mg/mL BSA, 0.02% Tween 20 PBS buffer;

4) 300 sec ‘association’ in MDM2. BLI assay was run in triplicates. Equilibrium response (nm) was plotted against MDM2 concentration to determine Kd of synthetic and recombinant MDM2.

3.8 Acknowledgements

Acknowledgments: We thank Prof. T.F. Jamison, Dr. H.U. Stilz, Dr. L.F. Iversen, Dr. K. Little and Dr. D. Lundsgaard for productive discussions and administrative support. We acknowledge the participants of the 26th American Peptide Symposium and the 8th Chemical Protein Synthesis Meeting, especially Prof. P.D. Dawson, Prof. R.T. Raines and Prof. S.B.H. Kent, for helpful discussions and for suggesting additional experiments. We also thank Dr. E.D. Evans, Dr. F.W.W. Hartrampf and R.L. Holden for careful proofreading of the manuscript. We are grateful to Dr. N. Truex for providing recombinant sortase A*. Finally, we acknowledge C.M.T. Hartrampf for designing Fig. 1A;

Funding: Financial support for this project was provided by Novo Nordisk. A.S., A.E.C., C.K.S. gratefully acknowledge support from the National Science Foundation Graduate Research Fellowship under Grant No. 1122374; A.E.C. is additionally supported by an MIT Dean of Science Fellowship.

Author contributions: N.H., T.E.N. and B.L.P. conceptualized the research; N.H., M.P. and S.L. optimized synthesis conditions; A.J.M. provided the software for UV data analysis and M.S. built the AFPS used in this report; M.P., A.J.C. and C.J. performed comparison of AFPS with traditional SPPS methods; N.H, A.S., M.P., A.J.C., A.E.C., S.H., S.A., C.K.S., and A.J.Q. synthesized, purified and analyzed protein samples; N.H., Z.P.G. and B.L.P. conceptualized folding and biological evaluation of the synthetic proteins; N.H., A.S., M.P., Z.P.G., A.J.C. and X.Y. performed biological evaluation and expression of recombinant proteins; N.H., Z.P.G, A.L. and B.L.P wrote the manuscript with input of all co-authors

3.9 References

1. Britton, J.; Raston, C. L. Multi-Step Continuous-Flow Synthesis. *Chem. Soc. Rev.* **2017**, *46* (5), 1250–1271. https://doi.org/10.1039/C6CS00830E.

2. Li, J.; Ballmer, S. G.; Gillis, E. P.; Fujii, S.; Schmidt, M. J.; Palazzolo, A. M. E.; Lehmann, J. W.; Morehouse, G. F.; Burke, M. D. Synthesis of Many Different Types of Organic Small Molecules Using One Automated Process. *Science* **2015**, *347* (6227), 1221–1226. https://doi.org/10.1126/science.aaa5414.

3. Sletten, E. T.; Nuño, M.; Guthrie, D.; Seeberger, P. H. Real-Time Monitoring of Solid-Phase Peptide Synthesis Using a Variable Bed Flow Reactor. *Chem. Commun.* **2019**, *55* (97), 14598–14601. https://doi.org/10.1039/C9CC08421E.

4. Webb, D.; Jamison, T. F. Continuous Flow Multi-Step Organic Synthesis. *Chem. Sci.* **2010**, *1* (6), 675. https://doi.org/10.1039/c0sc00381f.

5. *Flow Chemistry in Organic Synthesis*, 1st ed.; Jamison, T. F., Koch, G., Eds.; Georg Thieme Verlag KG: Stuttgart, 2018; p b-006-161272. https://doi.org/10.1055/b-006-161272.

6. Mijalis, A. J.; Thomas, D. A.; Simon, M. D.; Adamo, A.; Beaumont, R.; Jensen, K. F.; Pentelute, B. L. A Fully Automated Flow-Based Approach for Accelerated Peptide Synthesis. *Nat. Chem. Biol.* **2017**, *13* (5), 464–466. https://doi.org/10.1038/nchembio.2318.

7. Gordon, C. P. The Renascence of Continuous-Flow Peptide Synthesis – an Abridged Account of Solid and Solution-Based Approaches. *Org. Biomol. Chem.* **2018**, *16* (2), 180–196. https://doi.org/10.1039/C7OB02759A.

8. Ahmed, N. Peptide Bond Formations through Flow Chemistry. *Chem. Biol. Drug Des.* **2018**, *92* (1), 1398–1398. https://doi.org/10.1111/cbdd.13204.

9. Bayer, E.; Jung, G.; Halász, I.; Sebestian, I. A New Support for Polypeptide Synthesis in Columns. *Tetrahedron Lett.* **1970**, *11* (51), 4503–4505. https://doi.org/10.1016/S0040-4039(01)83961-7.

10. Lukas, T. J.; Prystowsky, M. B.; Erickson, B. W. Solid-Phase Peptide Synthesis under Continuous-Flow Conditions. *Proc. Natl. Acad. Sci.* **1981**, *78* (5), 2791–2795. https://doi.org/10.1073/pnas.78.5.2791.

11. Spare, L. K.; Laude, V.; Harman, D. G.; Aldrich-Wright, J. R.; Gordon, C. P. An Optimised Approach for Continuous-Flow Solid-Phase Peptide Synthesis Utilising a Rudimentary Flow Reactor. *React. Chem. Eng.* **2018**, *3* (6), 875–882. https://doi.org/10.1039/C8RE00190A.

12. Fuse, S.; Otake, Y.; Nakamura, H. Peptide Synthesis Utilizing Micro-Flow Technology. *Chem. - Asian J.* **2018**, *13* (24), 3818–3832. https://doi.org/10.1002/asia.201801488.

13. Fuse, S.; Mifune, Y.; Takahashi, T. Efficient Amide Bond Formation through a Rapid and Strong Activation of Carboxylic Acids in a Microflow Reactor. *Angew. Chem. Int. Ed.* **2014**, *53* (3), 851–855. https://doi.org/10.1002/anie.201307987.

14. Kent, S. B. H. Total Chemical Synthesis of Proteins. *Chem Soc Rev* **2009**, *38* (2), 338–351. https://doi.org/10.1039/B700141J.

15. Wright, T. H.; Bower, B. J.; Chalker, J. M.; Bernardes, G. J. L.; Wiewiora, R.; Ng, W.-L.; Raj, R.; Faulkner, S.; Vallée, M. R. J.; Phanumartwiwath, A.; Coleman, O. D.; Thézénas, M.-L.; Khan, M.; Galan, S. R. G.; Lercher, L.; Schombs, M. W.; Gerstberger, S.; Palm-Espling, M. E.; Baldwin, A. J.; Kessler, B. M.; Claridge, T. D. W.; Mohammed, S.; Davis, B. G. Posttranslational Mutagenesis: A Chemical Strategy for Exploring Protein Side-Chain Diversity. *Science* **2016**, *354* (6312), aag1465. https://doi.org/10.1126/science.aag1465.

16. Agouridas, V.; El Mahdi, O.; Diemer, V.; Cargoët, M.; Monbaliu, J.-C. M.; Melnyk, O. Native Chemical Ligation and Extended Methods: Mechanisms, Catalysis, Scope, and Limitations. *Chem. Rev.* **2019**, *119* (12), 7328–7443. https://doi.org/10.1021/acs.chemrev.8b00712.

17. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. Synthesis of Proteins by Native Chemical Ligation. *Science* **1994**, *266* (5186), 776–779. https://doi.org/10.1126/science.7973629.

18. Müller, T. D.; Finan, B.; Bloom, S. R.; D'Alessio, D.; Drucker, D. J.; Flatt, P. R.; Fritsche, A.; Gribble, F.; Grill, H. J.; Habener, J. F.; Holst, J. J.; Langhans, W.; Meier, J. J.; Nauck, M. A.; Perez-Tilve, D.; Pocai, A.; Reimann, F.; Sandoval, D. A.; Schwartz, T. W.; Seeley, R. J.; Stemmer, K.; Tang-Christensen, M.; Woods, S. C.; DiMarchi, R. D.; Tschöp, M. H. Glucagon-like Peptide 1 (GLP-1). *Mol. Metab.* **2019**, *30*, 72–130. https://doi.org/10.1016/j.molmet.2019.09.010.

19. Wilding, J. P. H.; Batterham, R. L.; Calanna, S.; Davies, M.; Van Gaal, L. F.; Lingvay, I.; McGowan, B. M.; Rosenstock, J.; Tran, M. T. D.; Wadden, T. A.; Wharton, S.; Yokote, K.; Zeuthen, N.; Kushner, R. F. Once-Weekly Semaglutide in Adults with Overweight or Obesity. *N. Engl. J. Med.* **2021**, *384* (11), 989–1002. https://doi.org/10.1056/NEJMoa2032183.

20. Rekasi, Z.; Varga, J. L.; Schally, A. V.; Halmos, G.; Groot, K.; Czompoly, T. Antagonistic Actions of Analogs Related to Growth Hormone-Releasing Hormone (GHRH) on Receptors for GHRH and Vasoactive Intestinal Peptide on Rat Pituitary and Pineal Cells *in Vitro*. *Proc. Natl. Acad. Sci.* **2000**, *97* (3), 1218–1223. https://doi.org/10.1073/pnas.97.3.1218.

21. Cai, R.; Zhang, X.; Wang, H.; Cui, T.; Halmos, G.; Sha, W.; He, J.; Popovics, P.; Vidaurre, I.; Zhang, C.; Mirsaeidi, M.; Schally, A. V. Synthesis of Potent Antagonists of Receptors for Growth Hormone-Releasing Hormone with Antitumor and Anti-Inflammatory Activity. *Peptides* **2022**, *150*, 170716. https://doi.org/10.1016/j.peptides.2021.170716.

22. SCHNölzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B. H. *In Situ* Neutralization in Boc-chemistry Solid Phase Peptide Synthesis: Rapid, High Yield Assembly of Difficult Sequences. *Int. J. Pept. Protein Res.* **1992**, *40* (3–4), 180–193. https://doi.org/10.1111/j.1399-3011.1992.tb00291.x.

23. Lu, X.; Wang, W.; Zhang, L.; Hu, H.; Xu, P.; Wei, T.; Tang, H. Molecular Mechanism of *N* , *N* -Dimethylformamide Degradation in *Methylobacterium* Sp. Strain DM1. *Appl. Environ. Microbiol.* **2019**, *85* (12), e00275-19. https://doi.org/10.1128/AEM.00275-19.

24. Tam, J. P.; Lu, Y.-A. Coupling Difficulty Associated with Interchain Clustering and Phase Transition in Solid Phase Peptide Synthesis. *J. Am. Chem. Soc.* **1995**, *117* (49), 12058–12063. https://doi.org/10.1021/ja00154a004.

25. Zahariev, S.; Guarnaccia, C.; Pongor, C. I.; Quaroni, L.; Čemažar, M.; Pongor, S. Synthesis of 'Difficult' Peptides Free of Aspartimide and Related Products, Using Peptoid Methodology. *Tetrahedron Lett.* **2006**, *47* (25), 4121–4124. https://doi.org/10.1016/j.tetlet.2006.04.074.

26. Michels, T.; Dölling, R.; Haberkorn, U.; Mier, W. Acid-Mediated Prevention of Aspartimide Formation in Solid Phase Peptide Synthesis. *Org. Lett.* **2012**, *14* (20), 5218–5221. https://doi.org/10.1021/ol3007925.

27. Mergler, M.; Dick, F.; Sax, B.; Weiler, P.; Vorherr, T. The Aspartimide Problem in Fmoc-Based SPPS. Part I. *J. Pept. Sci.* **2003**, *9* (1), 36–46. https://doi.org/10.1002/psc.430.

28. Palasek, S. A.; Cox, Z. J.; Collins, J. M. Limiting Racemization and Aspartimide Formation in Microwave-Enhanced Fmoc Solid Phase Peptide Synthesis. *J. Pept. Sci.* **2007**, *13* (3), 143–148. https://doi.org/10.1002/psc.804.

29. Arbour, C. A.; Kondasinghe, T. D.; Saraha, H. Y.; Vorlicek, T. L.; Stockdill, J. L. Epimerization-Free Access to C-Terminal Cysteine Peptide Acids, Carboxamides, Secondary Amides, and Esters *via* Complimentary Strategies. *Chem. Sci.* **2018**, *9* (2), 350–355. https://doi.org/10.1039/C7SC03553E.

30. Pennington, M. W.; Zell, B.; Bai, C. J. Commercial Manufacturing of Current Good Manufacturing Practice Peptides Spanning the Gamut from Neoantigen to Commercial Large-Scale Products. *Med. Drug Discov.* **2021**, *9*, 100071. https://doi.org/10.1016/j.medidd.2020.100071.

31. Yang, Y.; Hua, Q.; Liu, J.; Shimizu, E. H.; Choquette, M. H.; Mackin, R. B.; Weiss, M. A. Solution Structure of Proinsulin. *J. Biol. Chem.* **2010**, *285* (11), 7847–7851. https://doi.org/10.1074/jbc.C109.084921.

32. Johnson, E. C. B.; Malito, E.; Shen, Y.; Rich, D.; Tang, W.-J.; Kent, S. B. H. Modular Total Chemical Synthesis of a Human Immunodeficiency Virus Type 1 Protease. *J. Am. Chem. Soc.* **2007**, *129* (37), 11480–11490. https://doi.org/10.1021/ja072870n.

33. Brik, A.; Wong, C.-H. HIV-1 Protease: Mechanism and Drug Discovery. *Org. Biomol. Chem.* **2003**, *1* (1), 5–14. https://doi.org/10.1039/b208248a.

34. Oliner, J. D.; Pietenpol, J. A.; Thiagalingam, S.; Gyuris, J.; Kinzler, K. W.; Vogelstein, B. Oncoprotein MDM2 Conceals the Activation Domain of Tumour Suppressor P53. *Nature* **1993**, *362* (6423), 857–860. https://doi.org/10.1038/362857a0.

35. Gasser, E.; Moutos, C. P.; Downes, M.; Evans, R. M. FGF1 — a New Weapon to Control Type 2 Diabetes Mellitus. *Nat. Rev. Endocrinol.* **2017**, *13* (10), 599–609. https://doi.org/10.1038/nrendo.2017.78.

36. Bresson, D. Anti-CD3 and Nasal Proinsulin Combination Therapy Enhances Remission from Recent-Onset Autoimmune Diabetes by Inducing Tregs. *J. Clin. Invest.* **2006**, *116* (5), 1371–1381. https://doi.org/10.1172/JCI27191.

37. Sevcík, J.; Urbanikova, L.; Dauter, Z.; Wilson, K. S. Recognition of RNase Sa by the Inhibitor Barstar: Structure of the Complex at 1.7 Å Resolution. *Acta Crystallogr. D Biol. Crystallogr.* **1998**, *54* (5), 954–963. https://doi.org/10.1107/S0907444998004429.

38. Serrano, L.; Kellis, J. T.; Cann, P.; Matouschek, A.; Fersht, A. R. The Folding of an Enzyme. *J. Mol. Biol.* **1992**, *224* (3), 783–804. https://doi.org/10.1016/0022-2836(92)90562-X.

39. Durek, T.; Torbeev, V. Yu.; Kent, S. B. H. Convergent Chemical Synthesis and High-Resolution x-Ray Structure of Human Lysozyme. *Proc. Natl. Acad. Sci.* **2007**, *104* (12), 4846–4851. https://doi.org/10.1073/pnas.0610630104.

40. Czarna, A.; Popowicz, G. M.; Pecak, A.; Wolf, S.; Dubin, G.; Holak, T. A. High Affinity Interaction of the P53 Peptide-Analogue with Human Mdm2 and Mdmx. *Cell Cycle* **2009**, *8* (8), 1176–1184. https://doi.org/10.4161/cc.8.8.8185.

41. Xia, X.; Longo, L. M.; Blaber, M. Mutation Choice to Eliminate Buried Free Cysteines in Protein Therapeutics. *J. Pharm. Sci.* **2015**, *104* (2), 566–576. https://doi.org/10.1002/jps.24188.

42. Guo, H. H.; Choe, J.; Loeb, L. A. Protein Tolerance to Random Amino Acid Change. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (25), 9205–9210. https://doi.org/10.1073/pnas.0403255101.

43. Johnson, E. C. B.; Malito, E.; Shen, Y.; Rich, D.; Tang, W.-J.; Kent, S. B. H. Modular Total Chemical Synthesis of a Human Immunodeficiency Virus Type 1 Protease. *J. Am. Chem. Soc.* **2007**, *129* (37), 11480–11490. https://doi.org/10.1021/ja072870n.

44. Braakman, I.; Helenius, J.; Helenius, A. Manipulating Disulfide Bond Formation and Protein Folding in the Endoplasmic Reticulum. *EMBO J.* **1992**, *11* (5), 1717–1722. https://doi.org/10.1002/j.1460-2075.1992.tb05223.x.

45. *High-Performance Liquid Chromatography of Peptides and Proteins: Separation, Analysis, and Conformation*; CRC Press, 2017. https://doi.org/10.1201/9780203751947.

46. Wong, K.-B.; Daggett, V. Barstar Has a Highly Dynamic Hydrophobic Core: Evidence from Molecular Dynamics Simulations and Nuclear Magnetic Resonance Relaxation Data. *Biochemistry* **1998**, *37* (32), 11182–11192. https://doi.org/10.1021/bi980552i.

47. Anfinsen, C. B. Principles That Govern the Folding of Protein Chains. *Science* **1973**, *181* (4096), 223–230. https://doi.org/10.1126/science.181.4096.223.

48. Kent, S. B. Racemic & Quasi-Racemic Protein Crystallography Enabled by Chemical Protein Synthesis. *Curr. Opin. Chem. Biol.* **2018**, *46*, 1–9. https://doi.org/10.1016/j.cbpa.2018.03.012.

49. Knowles, J. R. Tinkering with Enzymes: What Are We Learning? *Science* **1987**, *236* (4806), 1252–1258. https://doi.org/10.1126/science.3296192.

50. Arnold, U.; Hinderaker, M. P.; Nilsson, B. L.; Huck, B. R.; Gellman, S. H.; Raines, R. T. Protein Prosthesis: A Semisynthetic Enzyme with a β-Peptide Reverse Turn. *J. Am. Chem. Soc.* **2002**, *124* (29), 8522–8523. https://doi.org/10.1021/ja026114n.

51. Weber, I. T.; Agniswamy, J. HIV-1 Protease: Structural Perspectives on Drug Resistance. *Viruses* **2009**, *1* (3), 1110–1136. https://doi.org/10.3390/v1031110.

52. Toth, M. V.; Marshall, G. R. A Simple, Continuous Fluorometric Assay for HIV Protease. *Int. J. Pept. Protein Res.* **2009**, *36* (6), 544–550. https://doi.org/10.1111/j.1399-3011.1990.tb00994.x.

53. Abboud, S. A.; Cisse, E. H.; Doudeau, M.; Bénédetti, H.; Aucagne, V. A Straightforward Methodology to Overcome Solubility Challenges for N-Terminal Cysteinyl Peptide Segments Used in Native Chemical Ligation. *Chem. Sci.* **2021**, *12* (9), 3194–3201. https://doi.org/10.1039/d0sc06001a.

54. Schneider, J.; Kent, S. B. H. Enzymatic Activity of a Synthetic 99 Residue Protein Corresponding to the Putative HIV-1 Protease. *Cell* **1988**, *54* (3), 363–368. https://doi.org/10.1016/0092-8674(88)90199-7.

55. Ton-That, H.; Liu, G.; Mazmanian, S. K.; Faull, K. F.; Schneewind, O. Purification and Characterization of Sortase, the Transpeptidase That Cleaves Surface Proteins of *Staphylococcus Aureus* at the LPXTG Motif. *Proc. Natl. Acad. Sci.* **1999**, *96* (22), 12424–12429. https://doi.org/10.1073/pnas.96.22.12424.

56. Ling, J. J.; Policarpo, R. L.; Rabideau, A. E.; Liao, X.; Pentelute, B. L. Protein Thioester Synthesis Enabled by Sortase. *J. Am. Chem. Soc.* **2012**, *134* (26), 10749–10752. https://doi.org/10.1021/ja302354v.

57. Policarpo, R. L.; Kang, H.; Liao, X.; Rabideau, A. E.; Simon, M. D.; Pentelute, B. L. Flow-Based Enzymatic Ligation by Sortase A. *Angew. Chem. Int. Ed.* **2014**, *53* (35), 9203–9208. https://doi.org/10.1002/anie.201403582.

58. Oliner, J. D.; Pietenpol, J. A.; Thiagalingam, S.; Gyuris, J.; Kinzler, K. W.; Vogelstein, B. Oncoprotein MDM2 Conceals the Activation Domain of Tumour Suppressor P53. *Nature* **1993**, *362* (6423), 857–860. https://doi.org/10.1038/362857a0.

59. Chène, P. Inhibition of the P53-MDM2 Interaction: Targeting a Protein-Protein Interface. *Mol. Cancer Res. MCR* **2004**, *2* (1), 20–28.

60. Touti, F.; Gates, Z. P.; Bandyopadhyay, A.; Lautrette, G.; Pentelute, B. L. In-Solution Enrichment Identifies Peptide Inhibitors of Protein-Protein Interactions. *Nat. Chem. Biol.* **2019**, *15* (4), 410–418. https://doi.org/10.1038/s41589-019-0245-2.

61. Hartley, R. W. Barnase and Barstar: Two Small Proteins to Fold and Fit Together. *Trends Biochem. Sci.* **1989**, *14* (11), 450–454. https://doi.org/10.1016/0968-0004(89)90104-7.

62. Bhuyan, A. K.; Udgaonkar, J. B. Observation of Multistate Kinetics during the Slow Folding and Unfolding of Barstar. *Biochemistry* **1999**, *38* (28), 9158–9168. https://doi.org/10.1021/bi990285w.

63. Golbik, R.; Fischer, G.; Fersht, A. R. Folding of Barstar C40A/C82A/P27A and Catalysis of the Peptidyl-Prolyl Cis/Trans Isomerization by Human Cytosolic Cyclophilin (Cyp18). *Protein Sci. Publ. Protein Soc.* **1999**, *8* (7), 1505–1514. https://doi.org/10.1110/ps.8.7.1505.

64. Vinogradov, A. A.; Evans, E. D.; Pentelute, B. L. Total Synthesis and Biochemical Characterization of Mirror Image Barnase. *Chem. Sci.* **2015**, *6* (5), 2997–3002. https://doi.org/10.1039/C4SC03877K.

65. Hartley, R. W. Barnase–Barstar Interaction. In *Methods in Enzymology*; Elsevier, 2001; Vol. 341, pp 599–611. https://doi.org/10.1016/S0076-6879(01)41179-7.

66. Kostyukevich, Y.; Shulga, A. A.; Kononikhin, A.; Popov, I.; Nikolaev, E.; Deyev, S. CID Fragmentation, H/D Exchange and Supermetallization of Barnase-Barstar Complex. *Sci. Rep.* **2017**, *7* (1), 6176. https://doi.org/10.1038/s41598-017-06507-2.

67. Sreenivasan, V. K. A.; Kelf, T. A.; Grebenik, E. A.; Stremovskiy, O. A.; Say, J. M.; Rabeau, J. R.; Zvyagin, A. V.; Deyev, S. M. A Modular Design of Low-Background Bioassays Based on a High-Affinity Molecular Pair Barstar:Barnase. *PROTEOMICS* **2013**, *13* (9), 1437–1443. https://doi.org/10.1002/pmic.201200491.

68. Mong, S. K.; Vinogradov, A. A.; Simon, M. D.; Pentelute, B. L. Rapid Total Synthesis of DARPin pE59 and Barnase. *ChemBioChem* **2014**, *15* (5), 721–733. https://doi.org/10.1002/cbic.201300797.

69. Kellis, J. T.; Nyberg, K.; Fersht, A. R. Energetics of Complementary Side Chain Packing in a Protein Hydrophobic Core. *Biochemistry* **1989**, *28* (11), 4914–4922. https://doi.org/10.1021/bi00437a058.

70. Kelemen, B. R.; Klink, T. A.; Behike, M. A.; Eubanks, S. R.; Leland, P. A.; Raines, R. T. Hypersensitive Substrate for Ribonucleases. *Nucleic Acids Res.* **1999**, *27* (18), 3696–3701. https://doi.org/10.1093/nar/27.18.3696.

71. Pradeep, L.; Udgaonkar, J. B. Osmolytes Induce Structure in an Early Intermediate on the Folding Pathway of Barstar. *J. Biol. Chem.* **2004**, *279* (39), 40303–40313. https://doi.org/10.1074/jbc.M406323200.

72. Trivedi, M. V.; Laurence, J. S.; Siahaan, T. J. The Role of Thiols and Disulfides on Protein Stability. *Curr. Protein Pept. Sci.* **2009**, *10* (6), 614–625. https://doi.org/10.2174/138920309789630534.

73. Yamaguchi, H.; Miyazaki, M. Refolding Techniques for Recovering Biologically Active Recombinant Proteins from Inclusion Bodies. *Biomolecules* **2014**, *4* (1), 235–251. https://doi.org/10.3390/biom4010235.

74. Schreiber, G.; Fersht, A. R. The Refolding of Cis- and Trans-Peptidylprolyl Isomers of Barstar. *Biochemistry* **1993**, *32* (41), 11195–11203. https://doi.org/10.1021/bi00092a032.

75. Wang, T.; Tomic, S.; Gabdoulline, R. R.; Wade, R. C. How Optimal Are the Binding Energetics of Barnase and Barstar? *Biophys. J.* **2004**, *87* (3), 1618–1630. https://doi.org/10.1529/biophysj.104.040964.

76. Lutz, J.-F.; Ouchi, M.; Liu, D. R.; Sawamoto, M. Sequence-Controlled Polymers. *Science* **2013**, *341* (6146), 1238149. https://doi.org/10.1126/science.1238149.

77. Wheelan, S. J.; Marchler-Bauer, A.; Bryant, S. H. Domain Size Distributions Can Predict Domain Boundaries. *Bioinformatics* **2000**, *16* (7), 613–618. https://doi.org/10.1093/bioinformatics/16.7.613.

78. Simon, M. D.; Heider, P. L.; Adamo, A.; Vinogradov, A. A.; Mong, S. K.; Li, X.; Berger, T.; Policarpo, R. L.; Zhang, C.; Zou, Y.; Liao, X.; Spokoyny, A. M.; Jensen, K. F.; Pentelute, B. L. Rapid Flow-Based Peptide Synthesis. *ChemBioChem* **2014**, *15* (5), 713–720. https://doi.org/10.1002/cbic.201300796.

79. Li, J. C.; Liu, T.; Wang, Y.; Mehta, A. P.; Schultz, P. G. Enhancing Protein Stability with Genetically Encoded Noncanonical Amino Acids. *J. Am. Chem. Soc.* **2018**, *140* (47), 15997–16000. https://doi.org/10.1021/jacs.8b07157.

80. Xiao, H.; Nasertorabi, F.; Choi, S.; Han, G. W.; Reed, S. A.; Stevens, R. C.; Schultz, P. G. Exploring the Potential Impact of an Expanded Genetic Code on Protein Function. *Proc. Natl. Acad. Sci.* **2015**, *112* (22), 6961–6966. https://doi.org/10.1073/pnas.1507741112.

81. Suree, N.; Liew, C. K.; Villareal, V. A.; Thieu, W.; Fadeev, E. A.; Clemens, J. J.; Jung, M. E.; Clubb, R. T. The Structure of the Staphylococcus Aureus Sortase-Substrate Complex Reveals How the Universally Conserved LPXTG Sorting Signal Is Recognized. *J. Biol. Chem.* **2009**, *284* (36), 24465–24477. https://doi.org/10.1074/jbc.M109.022624.

82. Cardona, V.; Eberle, I.; Barthélémy, S.; Beythien, J.; Doerner, B.; Schneeberger, P.; Keyte, J.; White, P. D. Application of Dmb-Dipeptides in the Fmoc SPPS of Difficult and Aspartimide-Prone Sequences. *Int. J. Pept. Res. Ther.* **2008**, *14* (4), 285–292. https://doi.org/10.1007/s10989-008-9154-z.

83. Abdel-Aal, A. M.; Papageorgiou, G.; Raz, R.; Quibell, M.; Burlina, F.; Offer, J. A Backbone Amide Protecting Group for Overcoming Difficult Sequences and Suppressing Aspartimide Formation. *J. Pept. Sci.* **2016**, *22* (5), 360–367. https://doi.org/10.1002/psc.2877.

84. Behrendt, R.; Huber, S.; White, P. Preventing Aspartimide Formation in Fmoc SPPS of Asp-Gly Containing Peptides - Practical Aspects of New Trialkylcarbinol Based Protecting Groups: New Protecting Groups Preventing Aspartimide Formation in Fmoc SPPS. *J. Pept. Sci.* **2016**, *22* (2), 92–97. https://doi.org/10.1002/psc.2844.

85. Behrendt, R.; Huber, S.; Martí, R.; White, P. New *t* -Butyl Based Aspartate Protecting Groups Preventing Aspartimide Formation in Fmoc SPPS: NEW *t* -BUTYL BASED ASPARTATE PROTECTING GROUPS FOR FMOC SPPS. *J. Pept. Sci.* **2015**, *21* (8), 680–687. https://doi.org/10.1002/psc.2790.

86. Umotoy, J. C.; de Taeye, S. W. Antibody Conjugates for Targeted Therapy Against HIV-1 as an Emerging Tool for HIV-1 Cure. *Front. Immunol.* **2021**, *12*, 708806. https://doi.org/10.3389/fimmu.2021.708806.

87. Wiśniewski, K.; Sueiras-Diaz, J.; Jiang, G.; Galyean, R.; Lu, M.; Thompson, D.; Wang, Y.-C.; Croston, G.; Posch, A.; Hargrove, D. M.; Wiśniewska, H.; Laporte, R.; Dwyer, J. J.; Qi, S.; Srinivasan, K.; Hartwig, J.; Ferdyan, N.; Mares, M.; Kraus, J.; Alagarsamy, S.; Rivière, P. J. M.; Schteingart, C. D. Synthesis and Pharmacological Characterization of Novel Glucagon-like Peptide-2 (GLP-2) Analogues with Low Systemic Clearance. *J. Med. Chem.* **2016**, *59* (7), 3129–3139. https://doi.org/10.1021/acs.jmedchem.5b01909.

88. Koizumi, A.; Tsuchiya, A.; Nakajima, K.; Ito, K.; Terada, T.; Shimizu-Ibuka, A.; Briand, L.; Asakura, T.; Misaka, T.; Abe, K. Human Sweet Taste Receptor Mediates Acid-Induced Sweetness of Miraculin. *Proc. Natl. Acad. Sci.* **2011**, *108* (40), 16819–16824. https://doi.org/10.1073/pnas.1016644108.

89. Ye, X.; Lee, Y.-C.; Gates, Z. P.; Ling, Y.; Mortensen, J. C.; Yang, F.-S.; Lin, Y.-S.; Pentelute, B. L. Binary Combinatorial Scanning Reveals Potent Poly-Alanine-Substituted Inhibitors of Protein-Protein Interactions. *Commun. Chem.* **2022**, *5* (1), 128. https://doi.org/10.1038/s42004-022-00737-w.

90. Thapa, P.; Cabalteja, C. C.; Philips, E. E.; Espiritu, M. J.; Peigneur, S.; Mille, B. G.; Tytgat, J.; Cummins, T. R.; Bingham, J.-P. T-Boc Synthesis of Huwentoxin-i through

Native Chemical Ligation Incorporating a Trifluoromethanesulfonic Acid Cleavage Strategy: T-Boc Synthesis of Huwentoxin-I. *Biopolymers* **2016**, *106* (5), 737–745. https://doi.org/10.1002/bip.22887.

Chapter 4. Synthesis and application of artificial luciferases

The work presented below is being adapted from the following in-progress manuscript:

Amanda E. Cowfer[1], David Sarabia-Castillo[1], Joesph S. Brown[1], Micheal M. Lee[1], Gha Y. Lee[1], Andrei Loas[1], Bradley L. Pentelute[1,2,3,4*] **Rapid single-shot synthesis of artificial luciferases and their variants.** *Manuscripts in preparation.*

4.1 Introduction and Motivation

Bioluminescence is the emission of visible light by living organisms via a biochemical reaction. It is found in a wide variety of living organisms, from deep-sea marine inhabitants to fungi and insects.[1] Evolutionarily, we can observe bioluminescence in a broad number of genera, including in bacteria, dinoflagellates, arthropods, or mollusks, and even in some vertebrate fish, however, bioluminescence, unlike fluorescence, has not been reported in mammals of any kind.[2] At its most basic, the biochemical process to produce bioluminescence relies on the interaction of a luciferin substrate with a luciferase enzyme to produce an oxyluciferin product via oxidation pathways.[3] This electronically excited oxyluciferin product emits photons as it relaxes down to a ground state.[4] This has made bioluminescent proteins quite valuable to the research community since they can provide imaging sensitive to substrate addition, whereas fluorescent proteins consistently emit light; it is prevalent in the research community for two different luciferases with different substrate specificities to be used as orthogonal reporters.[5] From here, researchers have broken down bioluminescence-based enzymes into two broad categories: luciferases, which

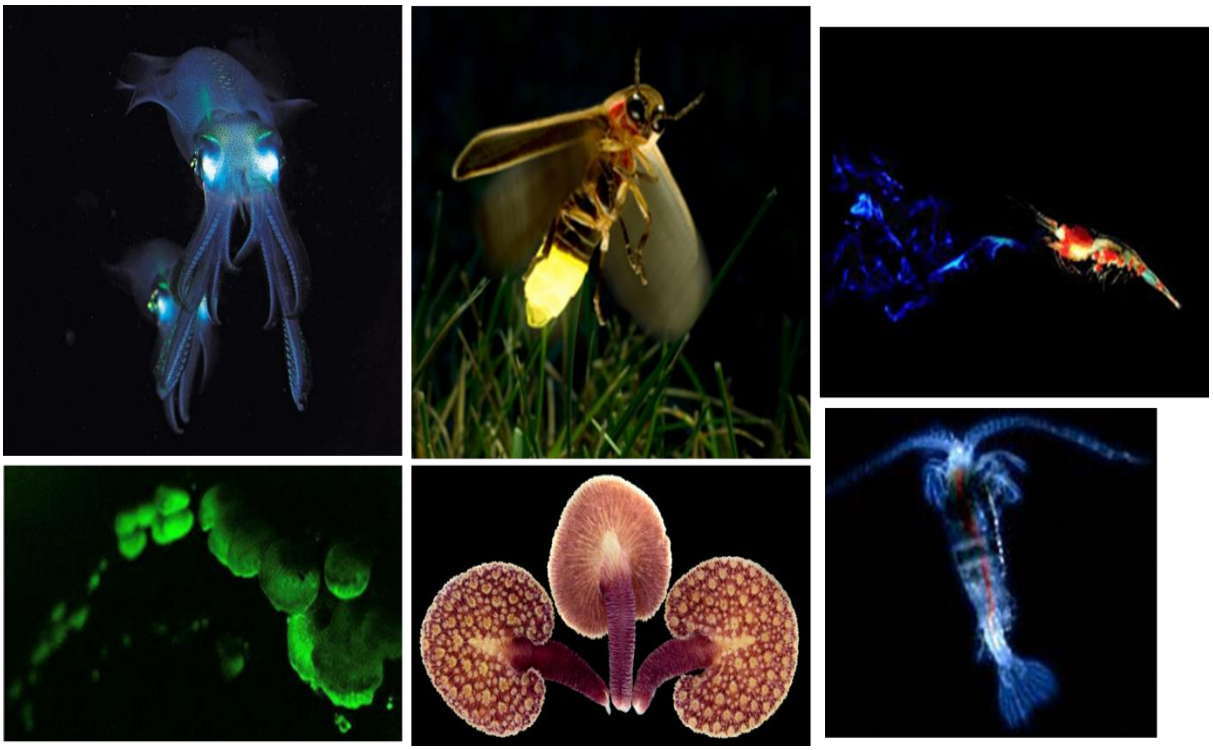


Figure 4.1: Image of organisms that produce bioluminescence

Top row left to right: Cuttlefish utilize various photoproteins to communicate and hunt for prey. Further, bioluminescence plays a crucial role in the mating patterns of Cuttlefish. Fireflies, native to the continental North Americas, are the most well-known luciferase producers in nature and use luciferases to attract mates. Interestingly, Firefly's bioluminescent display pattern is directly linked to environmental factors, like temperature or barometric pressure. The deep-sea shrimp *Oplophorus gracilirostris* uses bioluminescence as a defense to distract predators and escape being hunted unscathed. The luciferase enzyme from this shrimp was further improved into NanoLuc, a very popular new-age luciferase.
Bottom row left to right: Bioluminescent mushrooms have been reported to produce light in deep caves world-wide, however it is unknown why and is under active investigation. The sea pansy, *Renilla reniformis*, from which RLuc, the first coelenterazine based luciferases, was first discovered, and characterized. Today RLuc is one of the most popular luciferases used in research. The marine copepod *Gassuia Princpes* produces the very popular and stable luciferases GLuc. GLuc has been extensively studied due to its long half-life and bright bioluminescence.

release oxyluciferin after enzymatic processes are complete,[6] and photoproteins are bioluminescent enzymes that permanently bind to their substrate after the enzymatic process completes.[7]

In this report, I will focus on the synthetic production of proteins from the luciferase category of bioluminescent proteins. Luciferases have been reported in over 75 different species and have 11 different luciferin substrates reported in nature to date,[8] with more luciferin substrates that have been developed in research laboratories to provide a broader range of usable luciferins in different emission colors and with varying in-vivo stability.[9] In research, the most common luciferases used include those from firefly beetles (*Lampyridae*)[10], click beetles (*Elateridae*),[11]

glow worms (*Phengodidae*),[12] and various marine organisms such as the sea pansy (*Renillidae*),[13] and more new age luciferases from marine copepods (*Gassuia, Meridia).*[14] The various species that produce luciferase use them for various purposes such as mating,[15] defense,[15] camouflage,[16] and hunting.[17] Fireflies use luciferase-mediated bioluminescence to attract mates and, in more morbid circumstances, eat their competitors before they mate.[18] Another example of how organisms use luciferases is that small marine copepods, similar to krill, use bioluminescence to warn of and confuse predators.[19] Deep sea squid and jellyfish utilize their bioluminescent luciferases to help them hunt by providing a better view of the surrounding area or acting as a direct attractant for their prey.[20]

Luciferases in research are used in various applications, including the detection of environmental containments,[21] diagnosis of pathogens,[22] high-throughput screening for drug discovery,[23] understanding protein-protein interactions,[24] and more. In the world of high-throughput screening, luciferases make up approximately 20% of all detection techniques utilized in academic and industrial applications. Among luciferases used in high-throughput screening research, the most common is FireFly luciferase (FLuc) and *Rentilla* Luciferase (RLuc), and depending on whether researchers are using them for biochemical or cell-based assays will

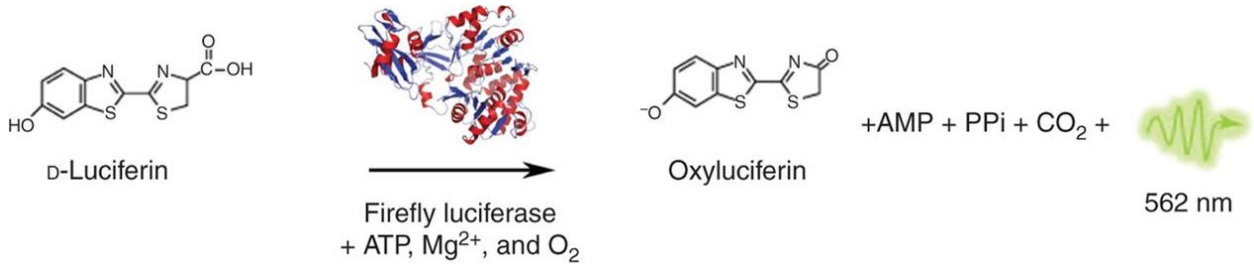


Figure 4.2: Luciferin class enzymatic pathway

The enzymatic reaction of Firefly luciferase with its substrate D-luciferin utilizes a variety of cofactors, including ATP, a chelating metal, often Mg^{2+}, and oxygen. Firefly luciferase mutants can be developed to use different metal cofactors for different applications. As the product oxyluciferin forms, a photon of light-based energy is released and can be detected at 562 nm. The need for many cofactors to perform the native enzymatic function makes firefly luciferase applications limited in many instances.

influence the choice of luciferase protein applied.[25] Luciferase-based assays are particularly useful for bioassays due to their low signal-to-noise ratio, the protection from photobleaching toxicity often found in fluorescence-based assays, and their propensity to react dynamically to reporter transcription levels in real-time due to the relatively short cellular half-life of FLuc and RLuc.[26] However, the short half-life can be a disadvantage in some assays, and to that end, researchers are slowly turning towards using more robust and longer-lasting luciferases.[27] An

example of a common application of luciferases is to use them as a reporter gene to serve as an indication of the transcription of specific proteins in cells. Unlike fluorescent-based reporters like GFP, luciferases can respond quickly to dynamic changes in the cellular environment, such as cellular export of proteins, protein degradation, and cell signaling pathways, without needing any external energy input to produce light.[28]

There are two main classes of luciferases used for research purposes, and each has

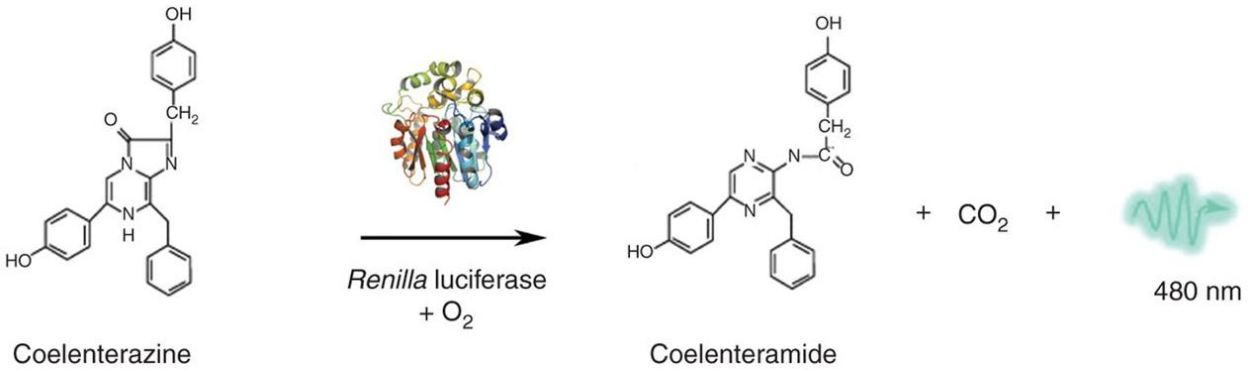


Figure 4.3: Coelenterazine class enzymatic pathway

Coelenterazine-based luciferases, such as *Renilla* luciferase, do not require any cofactors other than oxygen to perform their native enzymatic function. As the product of this reaction, coelenteramide, is formed a photon of energy is also released at 480 nm. Many coelenterazine-based luciferases can also utilize different substrates, meaning one enzyme can use two substrates with different emission spectra, a great advantage to researchers.

advantages and disadvantages, but often, researchers will use them in tandem. The first class of luciferases is the luciferin class; the most well-known luciferase in this class is firefly luciferase, the first one characterized by researchers.[29] Luciferin class luciferases utilize the substrate D-luciferin and a variety of cofactors, including ATP, Mg^{2+}, and oxygen, to oxidate D-luciferin into oxyluciferin with the byproducts of AMP, free phosphate, and carbon dioxide, and most importantly visible yellow light in the 562 nM range.[30] The second class of luciferase enzymes is known as the coelenterazine class. A commonly used member of this class is *Renilla* luciferase, from the sea pansy.[31] Coelenterazine-based luciferases are unique in that they do not require a massive amount of cofactors to oxidize coelenterazine into coelenteramide, using only oxygen to facilitate this process and producing carbon dioxide and blue-green light in the visible 480 nM range.[32] Because of these different functional modalities, RLuc and FLuc are often utilized in dual reporting assays; for example, researchers have developed a robust dual luciferase reporting assay to monitor mRNA splicing efficiency in vivo quantitatively. This assay works such that depending on the splicing site, either FLuc or RLuc is produced from the mRNA, and depending on the ratio of RLuc and FLuc, researchers can quantify the splicing pathway being utilized in the cell.[33]

Furthermore, luciferases are used in other applications like bioluminescence resonance energy transfer (BRET)[34] and bioluminescence imaging *in vivo*.[35]

Clearly, luciferases have broad and attractive applications in many areas of research, and the use of luciferases continues to increase over time. However, for most of the history of luciferase-based research, focus has been primarily put on improving and utilizing larger luciferases (>50 kDa), such as the classic FireFly Luciferase (62 kDa) and *Rentilla* luciferase discussed above. However, in the past two decades, the focus has shifted towards using, developing, and characterizing smaller (<20 kDa), brighter, and more stable luciferases. This was primarily spurred by discovering organisms that produce small coelenterazine class luciferases that are smaller and much brighter than the popular RLuc. One of these luciferases, OLuc19, is from the deep-sea shrimp *Oplophorus gracilirostris. OLuc19* is named for its unique presentation, with this luciferase consisting of a complex of two proteins, but the catalytic region is only 19kDa.[36] While this OLuc19 luciferase did demonstrate a brighter luminescence than RLuc, it has an extraordinarily short half-life. Thus, researchers engineered the much brighter and extremely stable NanoLuc (19 kDa) from OLuc19. NanoLuc is a very popular commercially available luciferase today and demonstrates nearly 150 times the intensity of light emitted than that of FLuc.[37] Another recently reported luciferase is from the marine copepod *Metridia pacifica*, which has two relatively thermostable and small (~20 kDa) luciferases, MpLuc1 and MpLuc2,[38] which are similar to OLuc19, were utilized to develop the much smaller luciferase TurboLuc (16 kDa). TurboLuc has proven quite valuable for high-throughput screening assays due to low interference with the small molecule compounds being screened in these assays.[39] Finally, one of the most used small coelenterazine luciferases, GLuc, comes from the marine copepod *Gaussia princeps*.[40] GLuc

is a very popular secreted reporter due to its stable nature, long *in vivo* half-life, and bright bioluminescence.

Currently, the development of small artificial luciferases is rising to the forefront of luciferase research due to several vital advantages these small luciferases have over traditional systems. These advantages include enhanced thermostability, chemical stability, increased brightness, and, of course, their small size, dramatically improving assay outcomes.[41] Among the natural small luciferase sequences that were isolated, most are from bioluminescent marine

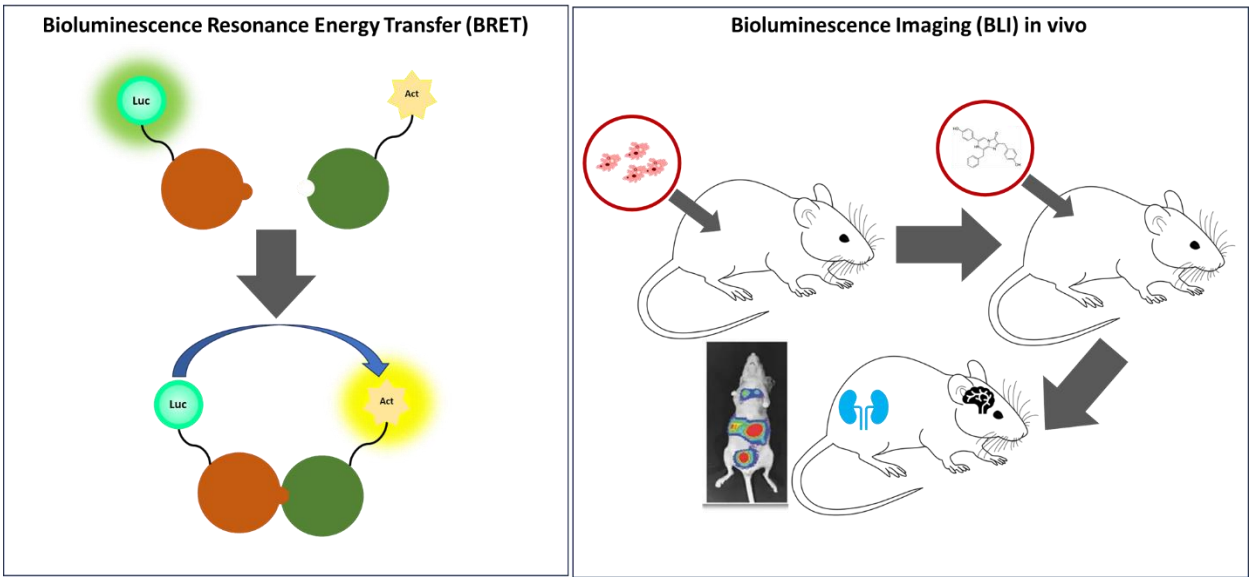


Figure 4.4: **Example of luciferase use in biomedical science via BRET and BLI**

Bioluminescence resonance energy transfer, known as BRET, utilizes the transfer of energy between a luciferase donor and fluorescence accepter and has been used to measure protein-protein interactions. Bioluminescence imaging *in vivo* relies on the transition of a cell line containing an expressed luciferase as a molecular reporter and then injecting the luciferase's substrate to image where the protein is being expressed.

copepods,[42] and have allowed for the development of small artificial luciferases by de novo enzyme design coupled with machine learning and consensus sequence-driven mutagenesis strategy (CSMS). These artificial luciferases are ideal candidates for experimental studies where larger luciferase enzymes may interfere with data collection and efficacy.[43]

In early 2022, Ohmuro-Matsuyama and coworkers reported the development of picALuc (13 kDa), the smallest coelenterazine class luciferase reported to date, from the much larger artificial luciferase ALuc30 (21 kDa).[44] I saw this report and saw an unmatched opportunity to use the Pentelute lab's AFPS systems to produce these enzymes. I envisioned that following the synthesis and characterization of these enzymes, I would be in a unique position to move towards incorporating unnatural amino acids, add facile bioconjugation handles, change the enzymatic

activity of these luciferases, and make customizable enzymes for biological applications. This project is the brainchild of my extensive chemical and biological knowledge and background. I believe this will be a great boon to the luciferase community as luciferase applications expand.

In this chapter, I report on the synthesis and characterization of the copepod, *Gaussia priceps*, luciferase GLuc (18 kDa), and artificial luciferases picALuc (12 kDa), its parent ALuc30, and LuxSit-I (14 kDa) via automated, fast flow peptide synthesis (AFPS). In addition, we synthesized the mirror-image counterpart of picALuc due to its potential for broad-reaching impact in health and diagnostics. This is the first report of a mirror-image bioluminescent luciferase. Finally, we will report on our efforts to develop a split-picAluc protein complement assay (PCA), the smallest and most versatile split-luciferase reported to date. To our knowledge, these are the first luciferase enzymes produced via solid-phase peptide synthesis (SPPS). This allows us to delve into the space of developing customizable and durable luciferases that can be used for future research applications.

4.2 Synthesis and characterization of disulfide-containing luciferases

Luciferases have long been used as sensitive reporters for biological imaging, pathogen detection, and more. As a class of proteins, they have been intensively studied and developed for decades, with RLuc and FLuc being very popular in the broad research community. In recent decades, however, smaller marine copepod coelenterazine-based luciferases have been reported and are being utilized and developed aggressively. A popular marine copepod luciferase is found in the deep-sea organism *Gassuia princips* and weighs in at only 18 kDa and 185 amino acids and utilizes coelenterazine (CTZ) as a substrate. I will refer to this luciferase as GLuc herein. GLuc is reported to be nearly 200 times brighter than the much larger FLuc and RLuc, with a high level of thermostability, which can be attributed to its five disulfide bonds.[1]

One significant way this subclass of luciferases is being studied and improved upon is through a “consensus sequence-driven mutagenesis strategy” (CSMS). CSMS compares an extensive database of RNA, DNA, or protein sequences to develop a sequence encompassing the most common biological monomers reported at each position and then produces a polymer sequence optimized to the researcher's specifications.[2] CSMS is a popular strategy in genetics to understand evolution better and create tools to help researchers better characterize the genome.[3] In the world of luciferases, a modified form of CSMS has been utilized to investigate the core enzymatic region of copepod luciferases; researchers have termed this “single-sequence

alignment" (SSA).[4] Kim and coworkers have reported a class of artificial luciferases developed from consensus sequence analysis by utilizing this strategy.[5] In this work, the sequence of 13-25 copepod luciferases was analyzed to produce these artificial luciferases. Copepod luciferases were chosen as an ideal model due to their high sequence similarity, small size, and two distinct domains.[6] ALuc30, a highly stable, bright, and small artificial luciferase, came from these reports.

A few years after the original report of ALuc30, another group of researchers discovered that ALuc30 remained stable and active when nearly 74 amino acids from both the N- and C-terminus were removed; thus, the smallest known luciferase to date was born in the form of picALuc.[7] This small, very bright, and highly thermostable luciferase has opened previously

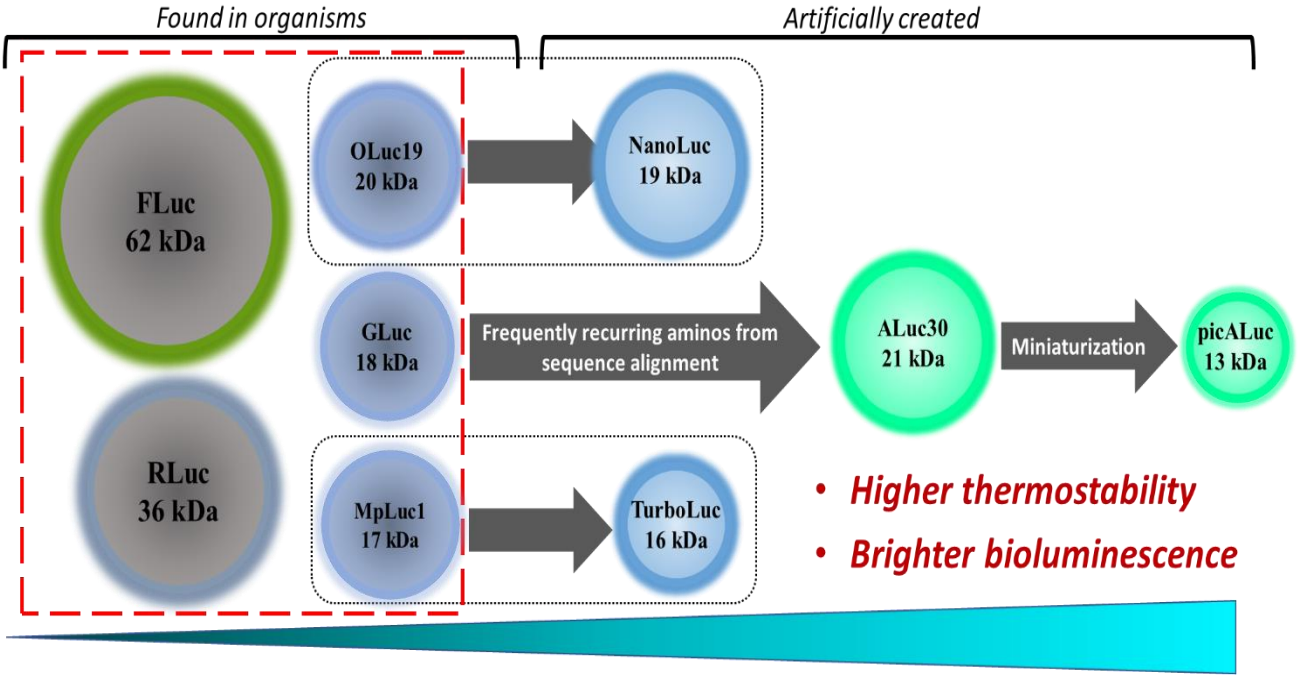


Figure 4.5: Scheme outlining the history of luciferases and the development of picAluc

Highlighted by the red box on the left are luciferases discovered in nature from organisms on the land and in the sea. FLuc and RLuc are the two most popular luciferases used on the market but are also very large luciferases. The smaller found-in-nature luciferases shown, OLuc19, MpLuc1, and GLuc, are all found in deep sea copepods, and are much smaller and brighter luciferases than RLuc and FLuc. In the black boxes I indicate the engineering of NanoLuc from OLuc19, and TurboLuc from MpLuc1. Both NanoLuc and TurboLuc are very small and stable luciferases enzymes and are commercially available. Finally, I demonstrate how sequence alignment strategies have resulted in the development of ALuc30, and other artificial luciferases. ALuc30 was further developed and miniaturized to produce picALuc, the smallest and brightest luciferase reported to date. Brightness of the luciferases shown generally increases across the image from left to right.

unavailable doors to the research community.

However, early reports of this protein have described it as challenging to express reliably in mammalian or bacterial cells. In this, I saw an exciting opportunity to use the AFPS technology

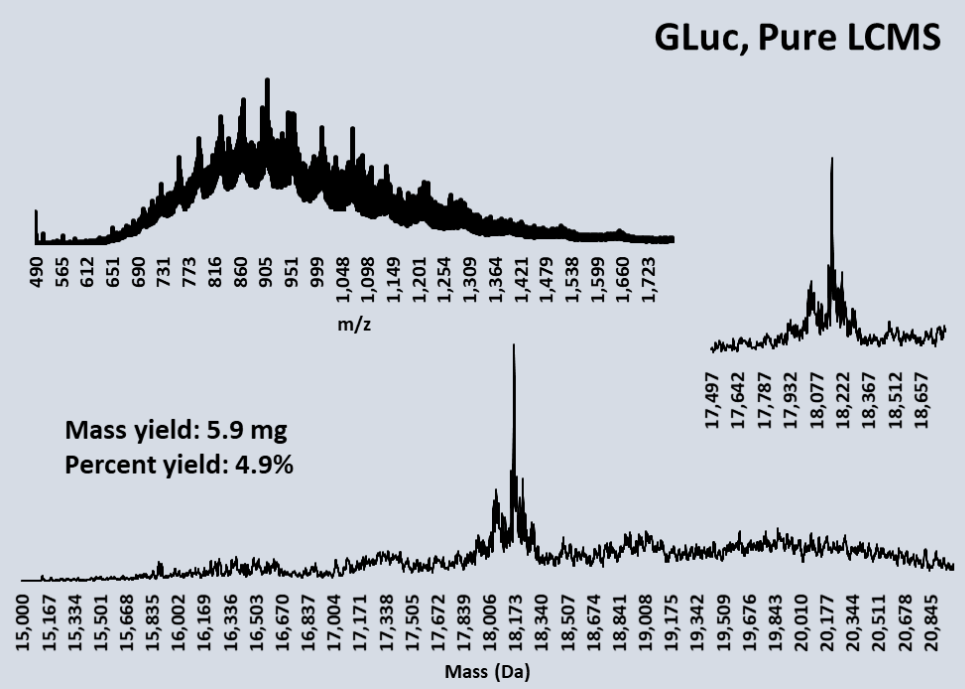
in the Pentelute lab to produce picALuc and other luciferases synthetically. By bypassing the need for cellular expression, I can not only produce picALuc in very high yields and excellent purity, but I can also easily introduce a few non-canonical amino acids into the sequence. I am opening the door to creating customizable, durable, and reliable luciferases for various applications. I can envision that having reliable access to luciferases with bioconjugation handles or even an entirely mirror image luciferase would create limitless opportunities for researchers to apply luciferases to solve more pressing scientific problems. For instance, researchers developing luciferase-based biosensors have struggled to develop bioconjugation handles to adhere their luciferases to a sensor surface due to limitations with amino acid incorporation with traditional protein expression methods, a limitation I do not face with the AFPS technology. In the following sections, I present my work developing a reliable method to produce biologically active luciferases synthetically with the AFPS, setting the groundwork for future discoveries.

4.2.1 Synthesis and folding of *Gassuia* luciferase

To begin my studies into luciferases, I needed a robust and well-understood model system to compare my synthetic luciferases. Further, due to the large number of disulfide bonds within picALuc, I wanted to use a closely related recombinant counterpart to develop our folding protocol. Historically, proteins with multiple disulfide bonds are prone to misfolding and aggregation in cellular environments. This propensity to misfolding dramatically increases when a researcher attempts to fold a disulfide-rich protein outside of a cellular environment where the various chaperones, disulfide isomerases, and other machinery cannot facilitate proper protein folding or correct improperly folded proteins. Luckily, since picALuc was created by trimming down the sequence of the much longer ALuc30, and ALuc30 was developed through intensive sequence alignment efforts of several copepod luciferase sequences, I identified where to start my search for a model luciferase.

To this end, I turned my attention to *Gauissa* luciferase, GLuc, due to the conserved disulfide bonds between GLuc, picALuc, and ALuc30. Furthermore, GLuc is a well-studied luciferase with broad applicability in the luciferase community. Gluc was initially reported in the early 2000s and is a highly stable, small, secreted luciferase. GLuc is one of an unknown number of coelenterazine-based luciferases found in deep-sea marine copepods.[8] Most importantly, GLuc has been submitted to a plethora of characterization studies, and recombinant samples, both tagged and untagged, are readily available from vendors.

Among the numerous publications reporting on GLuc is a paper focusing on the oxidative folding pathway of GLuc. In this study, researchers thoroughly characterized the oxidative folding of GLuc using a wide variety of small molecules and even biomacromolecule oxidative substrates. Upon the introduction of oxidants, unfolded GLuc was refolded with a reported recovery of >60% of native bioluminescent activity. Thus, Yu and colleagues believed GLuc could make a suitable reporter for oxidative folding-based assays with a fast, facile readout for understanding small molecule inhibitors of oxidative reporting, replacing the more traditional RNase methods that can



GLuc sequence:

KPTENNEDFNIVAVASNFATTDLDADRGKLPGKKLPLEVLKEMEANARKAG**C**TRG**C**LI**C**LS
HIK**C**TPKMKKFIPGR**C**HTYEGDKESAQGGIGEAI
VDIPEIPGFKDLEPMEQFIAQVDL**C**VD**C**TTG**C**LKGLANVQ**C**SDLLKKWLPQR**C**ATFASKIQ
GQVDKIKGAGGD

Calculated mass: 18169.034 Da., Observed mass: Da. 18168.8

Figure 4.6: GLuc purified LCMS data with deconvolution spectrum

On the top left is the MS data from purified GLuc. The bottom contains the deconvoluted MS data, with the desired mass of GLuc being the primary peak observed. In the middle is a zoomed-in version of the deconvoluted MS data indicating that there may be some minor side products that were not resolved with RP-HPLC purification. GLuc was recovered in good yield.

have much more complicated assay workups.[9] In the report, the oxidants discussed included glutathione redox buffers, hydrogen peroxide, tetrathionate, and dehydroascorbic acid, with protein disulfide isomerases utilized to help facilitate proper disulfide bond formation. This report

gave me a clear set of folding parameters to test for folding my synthetic GLuc and move on to applications with picALuc.

My first step was to synthesize GLuc using our automated fast-flow peptide synthesizers on RINK amide low-loading resin (0.18 mmol/g). Low-loading resin is generally the best choice

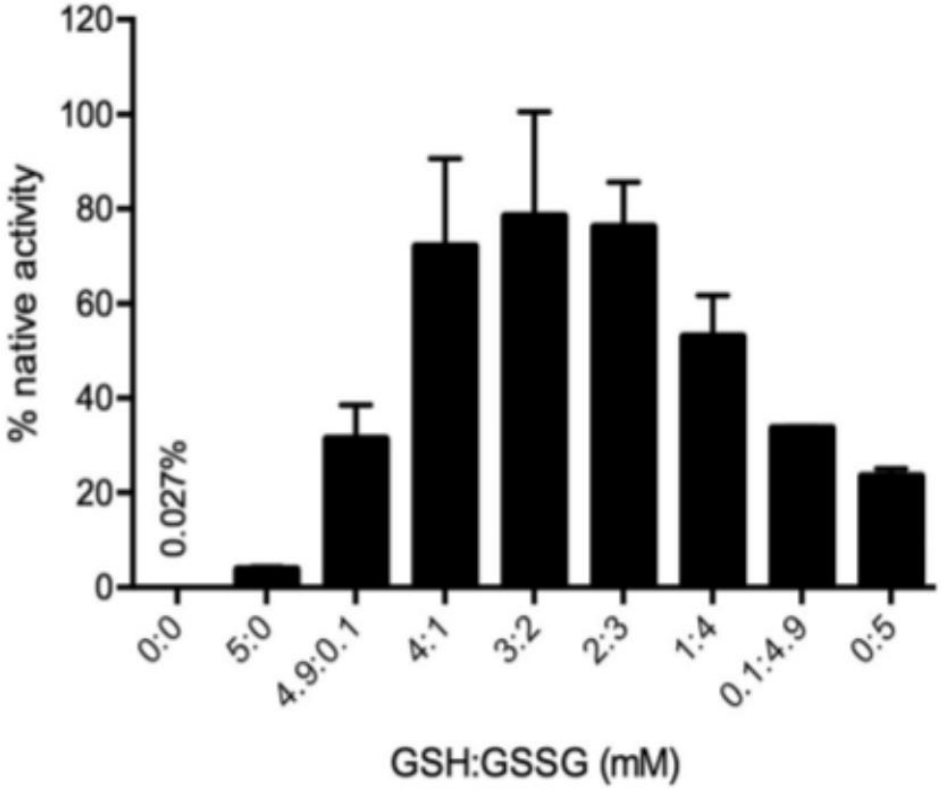


Figure 4.7: Graph adapted from Yu et al. demonstrating rGLuc refolding in the presence of a glutathione redox buffer

Notice that when no redox solution is present GLuc does not spontaneously refold, the optimal folding conditions from this comparisons appear to be a ratio of 3:2 GSH:GSSG (mM) with a 2:3 ratio having similar results. This data was collected after a 2 hour incubation with the redox buffer and normalized to recombinant GLuc.

for long polypeptide sequences since this resin allows for plenty of space for the growing peptide chains, and it prevents excessive inter-chain interactions that could result in single amino acid deletions and truncations.[10] GLuc's 168 amino acid length, after removing the native 17-mer signaling peptide, is on the longer end of the polypeptides we can successfully access via AFPS. Still, we have seen great success synthesizing proteins longer than 200 amino acids, with some sacrificing analytical data quality. The 17-mer peptide removed from the native GLuc sequence is a signal sequence located at the N-terminus of the protein polypeptide. This sequence functions as a message to cellular machinery to send the respective protein to be processed through the Golgi Apparatus and then secreted out of the cell. This signal sequence generally contains hydrophobic

amino acids and is always a minimum of 6 amino acids long. For GLuc, the secretory sequence is MGVKVLFALICIAVAEA, as reported by Yarimizu and coworkers.[11]

GLuc was synthesized via automated flow on solid-phase H-Rink ChemMatrix resin, leaving a free C-terminal amide instead of the native carboxylic acid. Each amino acid coupling cycle consists of amino acid activation, followed by amide coupling of a Fmoc-protected amino acid and deprotection of the Fmoc blocking group. The AFPS system includes a unique in-line UV-vis absorbance detector, which allows for real-time qualitative monitoring of the efficiency of each amino acid incorporation. With absorbance set to 310 nm, we can observe the exposure of resin to excess amino acid and activator and, more importantly, the chemical removal of the fluorene-containing Fmoc protecting group. With this data, we can use the width and height of each deprotection peak to identify any synthetic irregularities, including aggregation of the growing peptide chain, via observance of a consistent decrease in deprotection peak height. GLuc's 185 residue sequence was synthesized in 7.5 hours in 376 synthetic steps, with each residue incorporation taking an average of 2.5 minutes.

Immediately following synthesis, the resin-bound GLuc was stored under vacuum until peptide cleavage. Following peptide cleavage, GLuc was purified via high--performance liquid chromatography (RP-HPLC) using a standard 5-65% B over 72 minutes method (Solvent A is water + 0.1% TFA, and Solvent B is Acetonitrile in 0.1% B), and resulting fractions were analyzed

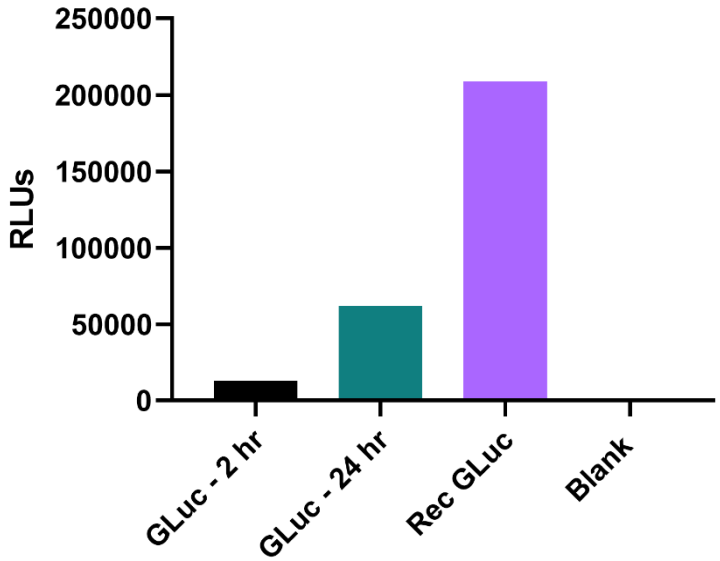


Figure 4.7: GLuc luminescence results following a 2 hour and 24 hour incubation time

These results, taken using only the theoretical protein concentration, demonstrate the overnight incubation times are better for folding outcomes and that the GSH and GSSG redox buffer is appropriate to allow for disulfide bond shuffling. However, compensation for actual protein concentration will be needed in the future.

by liquid chromatography-mass spectrometry (LCMS). Some difficulty with GLuc ionization on our LCMS was experienced, and a 1% formic acid solution helped partially resolve the ionization problem. Following analysis, clean fractions were pooled, and 5.9 mg of GLuc polypeptide was recovered at a calculated yield of 4.9%.

Next, I moved into accessing appropriate folding conditions for our synthetic GLuc. Based on the abovementioned study, GLuc should fold in various oxidative conditions. I decided that

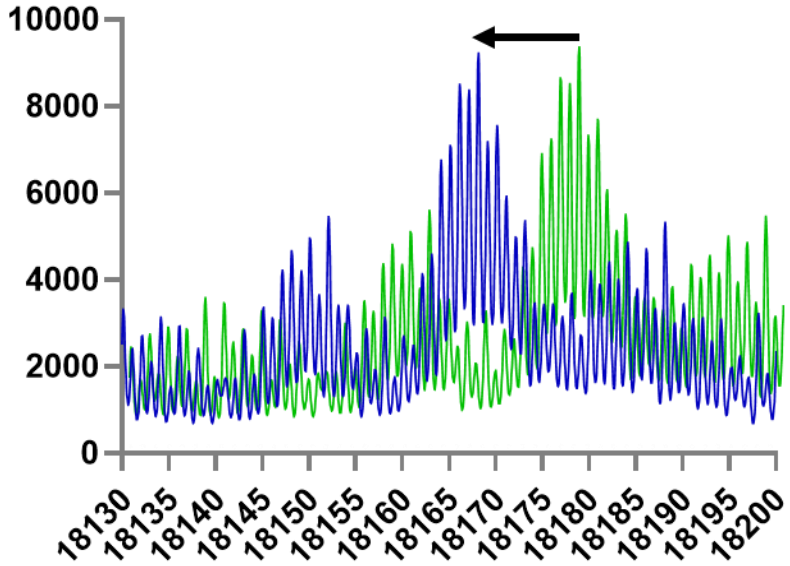


Figure 4.8: Graph of LCMS data for GLuc demonstrating the 10 Da mass shift corresponding to disulfide bond formation

Mass shift data for GLuc was showing the mass shift of folded GLuc after disulfide bond formation, it is important to note that bond formation does not guarantee proper folding for function. Data was obtained by analyzing LCMS traces with a 0.05 Da resolution.

using a glutathione redox buffer would be the best initial condition to assess. Historically, buffers containing different ratios of reduced glutathione to oxidized glutathione (GSH: GSSG) allows for some amount of disulfide bond shuffling during protein refolding, while the other oxidants mentioned, like hydrogen peroxide or tetrathionate, do not have this same propensity.[12] For our folding studies, I resolved to perform "quick and dirty" protein folding followed by a bioluminescent assay to assess folding outcomes. For this type of folding, I will only work with theoretical protein concentrations based on the lyophilized polypeptide mass weighed before folding.

The first condition I tested was a redox buffer containing GSH: GSSG in a ratio of 3:2, with a total of 5 mM glutathione redox buffer total, in a buffer of 50 mM sodium phosphate, 50 mM sodium chloride, and 1 mM Ethylenediaminetetraacetic acid (EDTA) at a pH of 8.0, this will herein be referred to as the "folding buffer." First, lyophilized GLuc was weighed and dissolved in 6M guanidine hydrochloride, with 55 mM potassium phosphate as a buffering agent, at a concentration of 2 mg/mL. Then, the folding buffer was prepared and added to the dissolved peptide and thoroughly mixed, and this solution was left to sit at room temperature on a bench top for 2 hours. After 2 hours, the GLuc folding solution was submitted to a bioluminescent assay and LCMS analysis. Via LCMS, I was looking for the formation of 5 disulfide bonds by monitoring for a mass shift of approximately 10 Da, indicating that all the cysteine residues are now oxidized and the hydrogens have been lost. For the bioluminescent assay, I diluted GLuc to 100 nM and introduced its substrate, coelenterazine, at a volume of 50 uL and a concentration of 3 uM. These values were adjusted as needed for future experiments.

My initial results showed complete disulfide formation via LCMS but low relative bioluminescence under our assay conditions. I ran my synthetic GLuc sample in parallel with a recombinant GLuc sample purchased from NanoLight Technologies. I was dismayed to see very little synthetic GLuc activity when compared to recombinant GLuc but resolved to allow the protein to remain overnight in the folding buffer in the hope that the disulfide bonds would shuffle to yield enzymatically active GLuc. To my delight, an overnight incubation (24+ hours) showed a significant increase in bioluminescent activity in my synthetic GLuc sample, and following accurate protein concentration analysis, I saw comparable activity between synthetic and recombinant GLuc. With this success, I felt I could move into synthesizing and folding the artificial luciferases ALuc30 and picALuc.

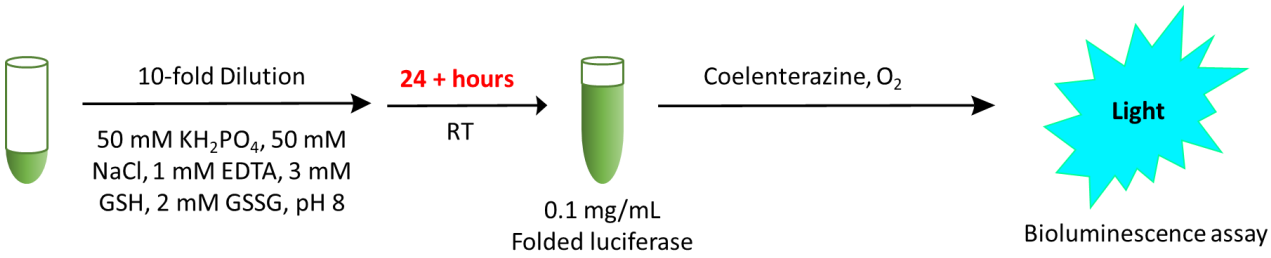


Figure 4.9: Finalized disulfide containing luciferase folding protocol workflow

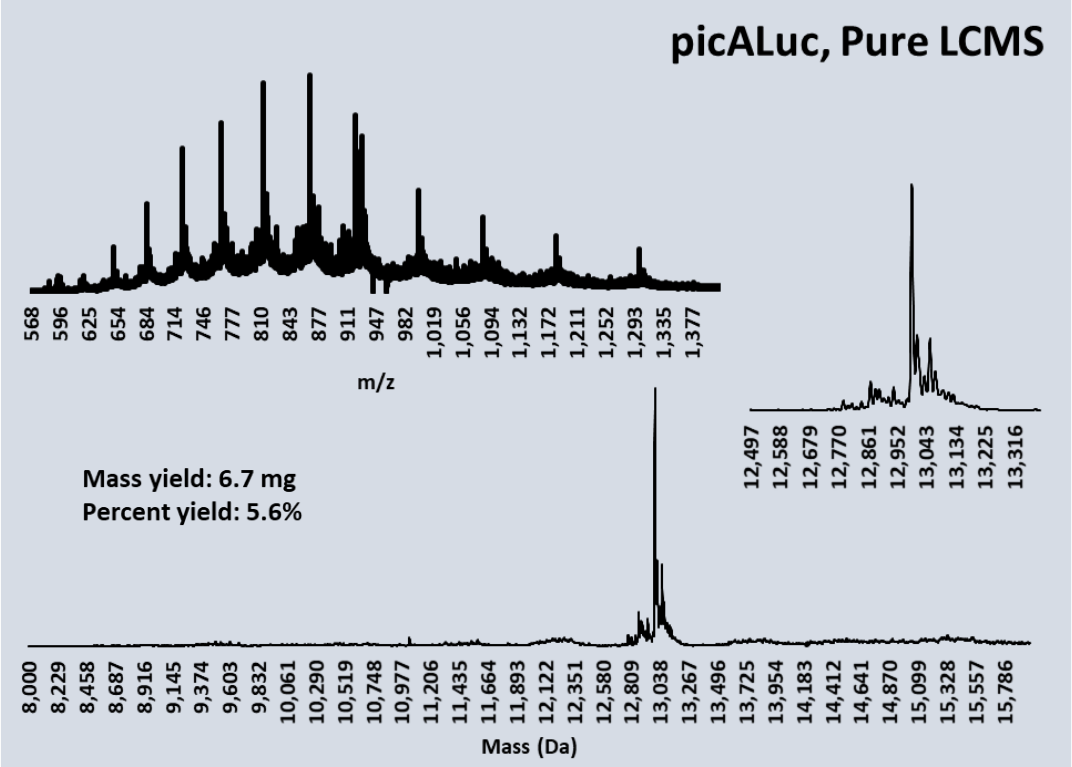
4.2.2 Synthesis and folding of the artificial luciferases picAluc and ALuc30

```
GLuc            KPTENNEDFNIVAVASNFATTDLDAD-------------------------RGKLPGKK
ALuc30          NHHHHHHHHDIVGVEGKFGTTDLETDLFTIVEDMNVISRDTDVDANRADRGRRGKLPGKK
picALuc         ------------------------------------------------------KLPGKK
                                                                       ******

GLuc            LPLEVLKEMEANARKAGCTRGCLICLSHIKCTPKMKKFIPGRCHTYEGDKESAQGGIGEA
ALuc30          LPLEVLKELEANAQKAGCTRGCLICLSHIKCTAKMKKWLPGRCESWEGDKETGQGGIGEA
picALuc         LPLEVLKELEANAQKAGCTRGCLICLSHIKCTAKMKKWLPGRCESWEGDKETGQGGIGEA
                ********:****.******************.****::**** ::******:.*******

GLuc            IVDIPEIPGFKDLEPMEQFIAQVDLCVDCTTGCLKGLANVQCSDLLKKWLPQRCATFASK
ALuc30          IVDIPEIPGFKELAPMEQFIAQVDLCADCTTGCLKGLANVKCSALLKKWLPSRCAGFADK
picALuc         IVDIPEIPGFKELAPMEQFIAQVDLCADCTTGCLKGLANVKCSALLKKWLPSRC------
                ***********:* ************.*************:** *******.**

GLuc            IQGQVDKIKGAGGD
ALuc30          IQAQVDTIKGAGGS
picALuc         --------------
```

Figure 4.10: Blast sequence alignment results for picALuc, ALuc30 and GLuc

Blast sequence alignment was performed using CLUSTAL multiple sequence alignment by MUSCLE (3.8) and the sequences used contained no tags. Notice the high level of alignment, demonstrating the high level of conservation between the three proteins, most importantly of the cystines that form disulfide bonds.

After our success with folding GLuc, the model protein for this study, I turned my efforts to the synthesis and folding of picALuc and its parent polypeptide ALuc30. Due to the high sequence alignment and conserved cysteine residues, observed between these three protein polypeptides, we hypothesized that the folding conditions that worked for GLuc should translate to ALuc30 and picALuc. The first synthesis of the full-length polypeptide for ALuc30 (194 aa, 21 kDa) and picALuc (120 aa, 13 kDa) was performed on our AFPS instrument, the Peptidator. Following synthesis, I cleaved the full-length polypeptide before RP-HPLC purification. I used a similar method as with our GLuc synthesis, with our gradient being 5-65% B over 80 minutes,



picALuc sequence:

KLPGKKLPLEVLKELEANAQKAG**C**TRG**C**LI**C**LSHIK**C**TAKMKKWLPGR**C**ESWEGDKETGQ
GGIGEAIVDIPEIPGFKELAPMEQFIAQVD
L**C**AD**C**TTG**C**LKGLANVK**C**SALLKKWLPSR**C**

Calculated mass: 12999.460 Da., Observed mass: Da. 12999.3

Figure 4.11: Purified LCMS traces for picALuc

On the top left is the MS data from purified picALuc. The bottom contains the deconvoluted MS data with the desired mass of picALuc being the primary peak observed. In the middle is a zoomed-in version of the deconvoluted MS data indicating that there is minimal side product formation after RP-HPLC purification. picALuc was recovered in good yield.

with (Solvent A is water + 0.1% TFA, and Solvent B is Acetonitrile in 0.1% B), and resulting fractions were analyzed by liquid chromatography-mass spectrometry (LCMS).

Unfortunately, ALuc30 did not ionize via LCMS, even after adding up to 5% formic acid in the solution and adjusting the electrospray ionization process. However, I chose to move ALuc30 forward into folding studies regardless. In the future, I hope to test MALDI-TOF as an alternative MS method to analyze ALuc30, and as a last resort, I could perform SDS-PAGE analysis. On the other hand, picALuc was purified, and 6.7 mg of material was recovered at a 5.6% yield.

Similar to GLuc, I used a redox buffer containing GSH: GSSG in a ratio of 3:2, with a total of 5 mM glutathione redox solution in the folding buffer. First, lyophilized HPLC purified

picALuc and crude lyophilized ALuc30 were weighed and dissolved in 6M guanidine hydrochloride, with 55 mM potassium phosphate as a buffering agent, at a concentration of 2 mg/mL. Then, the folding buffer was prepared, added to the dissolved peptide, and thoroughly mixed. This solution was then left to sit at room temperature on a bench top for 2 hours, and protein concentration was assessed via A280 using Beer's law. Then, the folding proteins were submitted to bioluminescent assay analysis. Next, the mixtures were incubated overnight (24+ hours) at room temperature and submitted to LCMS analysis. Once again, I was looking for the formation of 5 disulfide bonds by monitoring for a mass shift of approximately 10 Da, indicating that all the cysteine residues are now oxidized, and the hydrogens have been lost. For the bioluminescent assay, I diluted GLuc to 100 nM and introduced its substrate, coelenterazine, at a volume of 50 uL and a concentration of 3 uM. These values can be adjusted as needed in future experiments. I saw

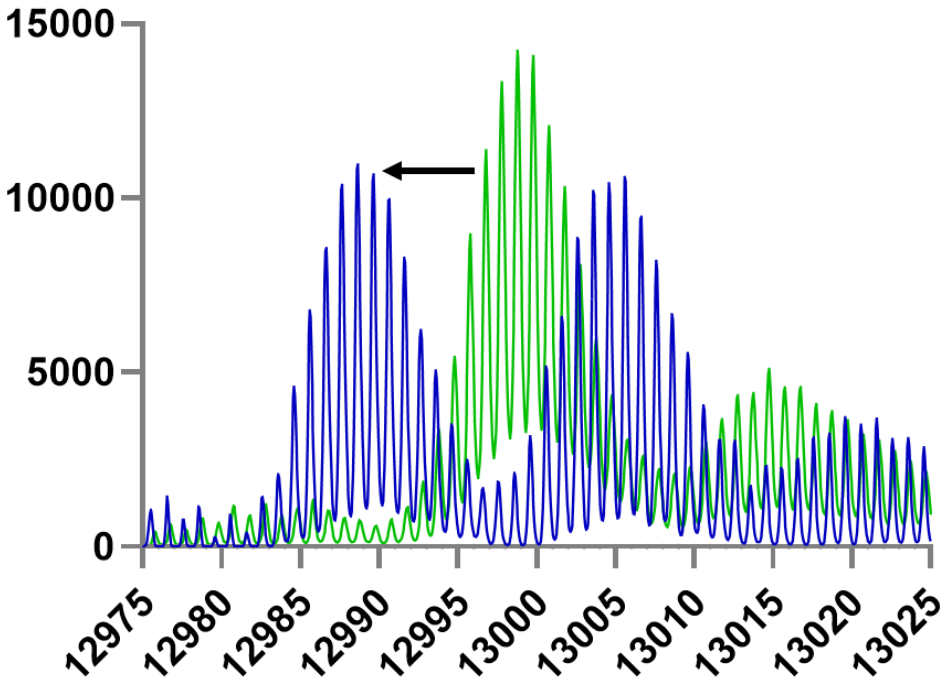


Figure 4.12: LCMS analysis of picALuc mass shift during disulfide bond formation

Mass shift data for picALuc was showing the mass shift of folded picALuc after disulfide bond formation, it is important to note that bond formation does not guarantee proper folding for function. Data was obtained by analyzing LCMS traces with a 0.05 Da resolution.

that both picALuc and ALuc30 behave similarly to GLuc and require overnight folding for optimal enzymatic activity. Furthermore, picALuc demonstrated the expected 10 Da mass shift, indicating disulfide bond formation.

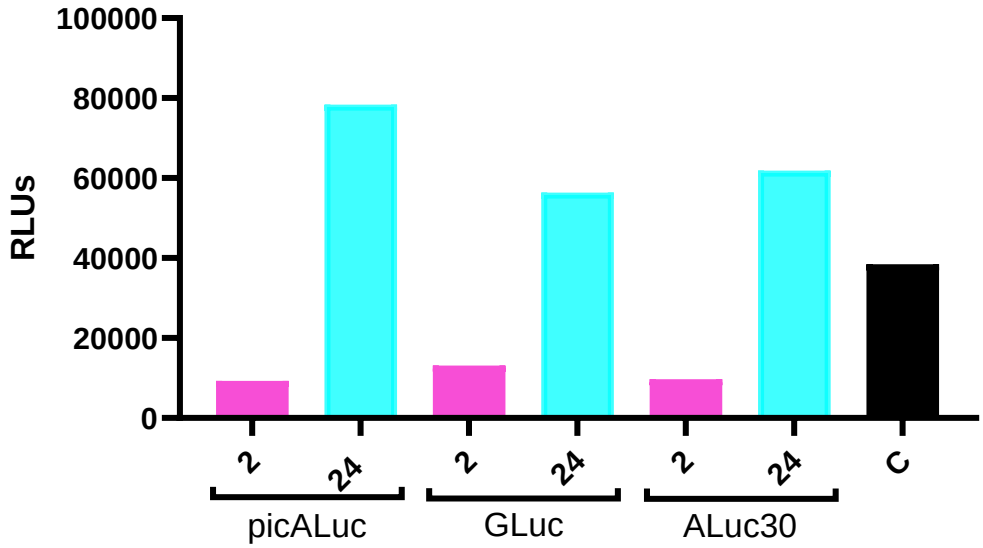


Figure 4.13: Bioluminescent data comparing 2 vs. 24 hour folding times

Bioluminescent assay results comparing outcomes for a 2 hour vs 24 hours folding time. We see a dramatic increase in luminescence after 24 hours, and when comparing values to our recombinant GLuc standard we see that are synthetic luciferase samples outperform recombinant consistently. Furthermore, we see that even crude protein will reliably fold in these conditions.

With this, I am ready to move into biological purification and further characterization of these proteins. These results were very exciting since controlled disulfide bond formation is historically difficult when there is more than one bond to be made. We were prepared to employ protein disulfide isomerases if needed, but these luciferases have proven to be extraordinarily robust.

4.2.3 Characterization of folded luciferases

With folding conditions well in hand, I investigated a suitable biopurification strategy. I first considered ion exchange chromatography; however, all three proteins have relatively neutral pI's which would make IEC optimization challenging. Next, I turned to a classic method used by the luciferase community to purify recombinant luciferases,[13] size exchange chromatography, using an isocratic elution buffer of 200 mM NaCl, 1 mM EDTA, and 20 mM Tris-HCl at pH 7.5. Following purification, I had sufficient material to move into further characterization studies.

First, I obtained kinetics traces of all purified luciferases using a TecanSpark reader with an autoinjector. Using an autoinjector helps remove human error that can make hand pipetting these luciferases non-ideal. In fact, I noticed vast differences in the data collected from run to run

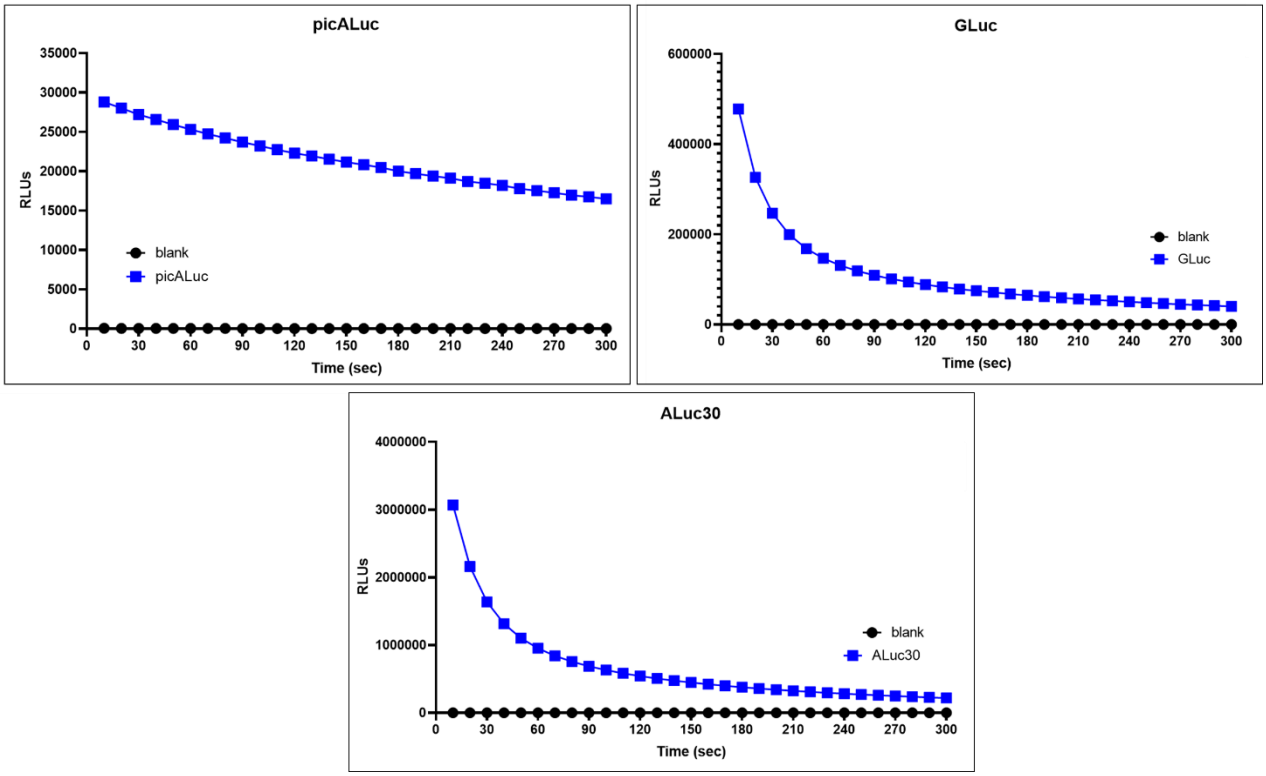


Figure 4.14: Kinetic bioluminescent traces of all luciferases

Bioluminescent assay results collected with a Tecan Spark plate reader using an autoinjector to inject CTZ substrate and mix the substrate for 5 seconds. Traces were collected over the course of 5 minutes with a reading every 10 seconds. GLuc and ALuc30 (crude folded) show distinct flash kinetics and quick decay, conversely picALuc demonstrates much more steady decay with less of a steep drop in luminescence over time.

due to our handling of the CTZ substrate and the "flash" kinetic properties of many copepod-based

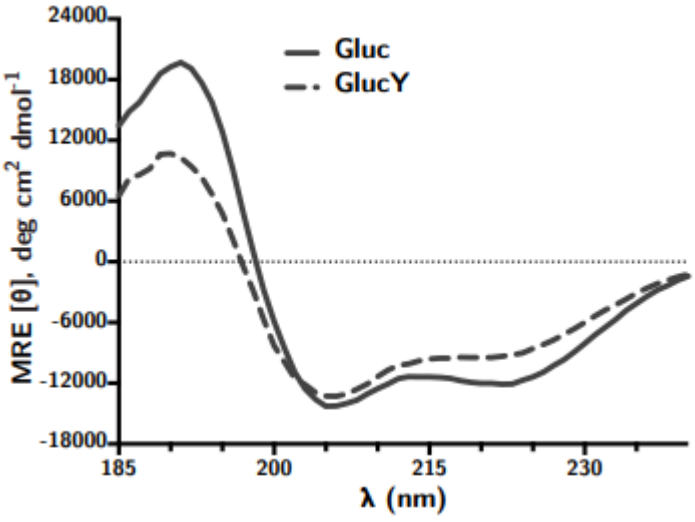


Figure 4.15: Example CD spectra data for GLuc

This CD spectra of GLuc is adapted from Hunt et al. and demonstrated the alpha-helix rich luciferase structure, I hope to collect similar data with my synthetic samples.

luciferases.[14] With our Tecan reader, I was able to observe the slow decay of luminescence over time as CTZ is oxidized to coelenteramide.

Analysis of these proteins was also attempted using Michaelis-Menten kinetic analysis, a method the original reporters of picALuc used to compare picALuc to NanoLuc and other common luciferases to demonstrate its effectiveness. However, upon closer inspection and literature review, I did not believe it appropriate to pursue kinetic analysis this way due to copepod luciferases' extraordinarily fast enzymatic turnover. This is one significant knowledge gap that plagues coelenterazine-based luciferases. Researchers do not fully understand how these enzymes function catalytically and how the so-called “flash kinetic” properties affect the assay data collected. Flash kinetics in luciferases are observed through the initial bright burst of light following the addition of coelenterazine, followed by a rapid light decay. Unfortunately, even with the introduction of more substrate, the bioluminescence readings remain low. Evidence in the literature shows that GLuc and other marine copepod luciferases may be rendered catalytically inactive after several enzymatic cycles.[14]

Next, I intend to characterize the 3D structure of these luciferases using circular dichroism spectroscopy (CD). CD analysis of these closely related luciferases should show alpha-helix-rich spectra for all luciferases reported, similar to previous studies of GLuc, see figure 4.15.[15] Further, I hope to grow crystals to submit these proteins to x-ray crystallography analysis.

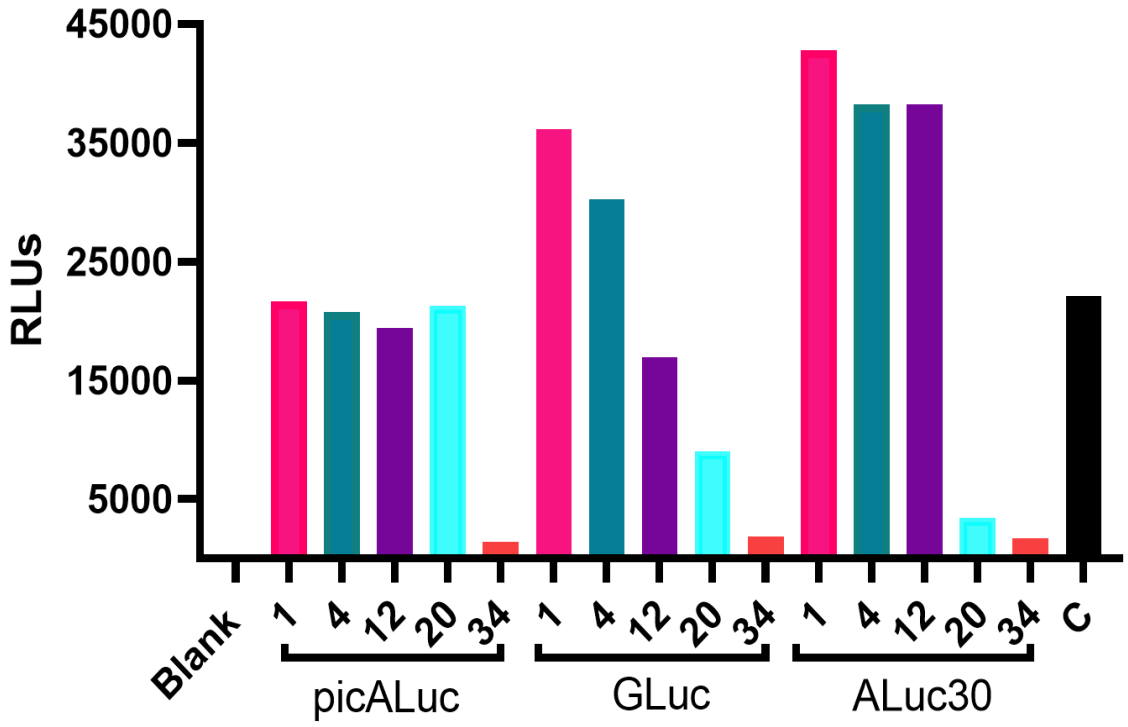


Figure 4.16: Bioluminescent data of folded luciferases over time.

This graph contains bioluminescent data time points taken over a 34-week time span, the first point displayed in pink is after 1 week at room temperature and all luciferases retain activity greater than that of a recombinant sample. The second data point in dark blue-green was taken after 4 weeks, or 1 month and little decay of function is observed. After 12 weeks, displayed in purple, I started to see a steep decay in function for GLuc, but picALuc and ALuc30 still retained near equal activity to initial data collection. At 20 weeks, light blue, GLuc and ALuc30 have experienced a significant drop in luminescence with picALuc still retaining a shocking amount of activity. Finally, at 34 weeks, 8 months, I see all luciferases losing nearly all significant bioluminescence. This demonstrates the extraordinary and unmatched stability of these luciferases.

Finally, I hoped to demonstrate the shear robustness of these luciferases over time by leaving folded samples of ALuc30, picALuc, and GLuc on a benchtop at room temperature to observe the loss of bioluminescent function over time. Shockingly, all the synthetic luciferases maintained enzymatic function even after eight months stored on a benchtop at a concentration of 0.1 mg/mL in folding buffer. This data demonstrates the power of these new-age artificial luciferases and further bolstered my desire to investigate them.

4.3 Synthesis and characterization of a luciferase with no disulfide bonds

Another exciting area of protein science luciferases that has gained traction is the de novo design of artificial enzymes using machine learning techniques. In a recent publication from the renowned Baker lab, they reported the first ever entirely artificial designed luciferases LuxSit-I

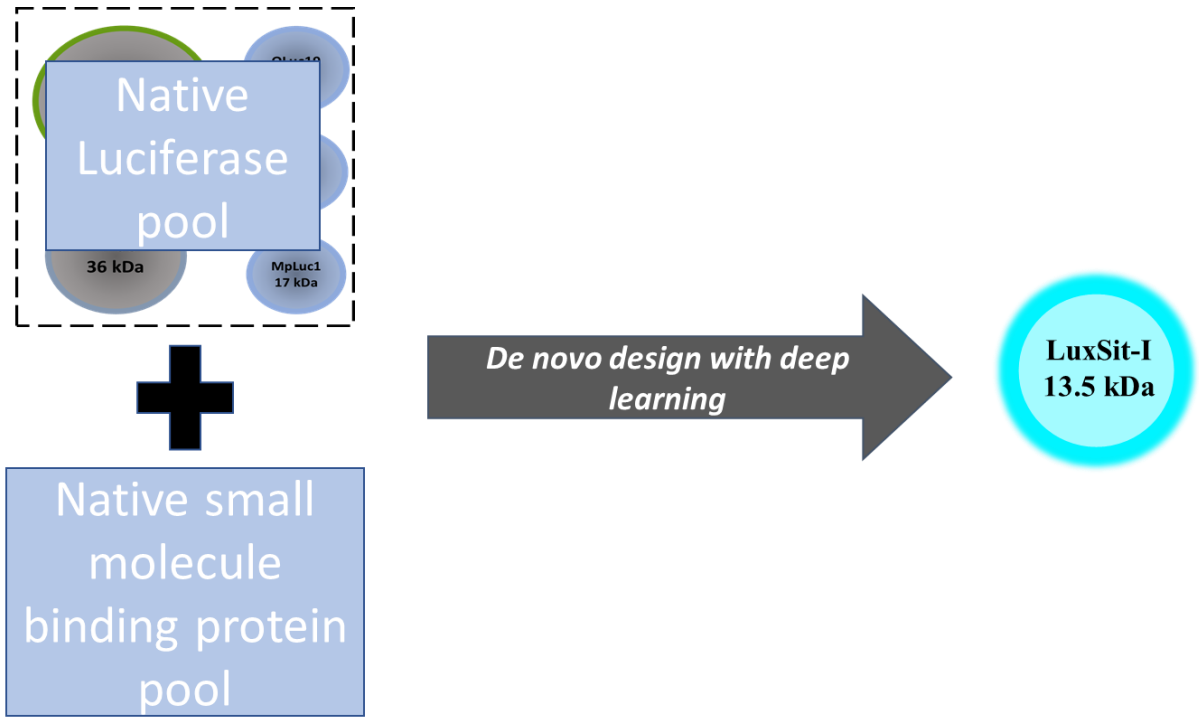
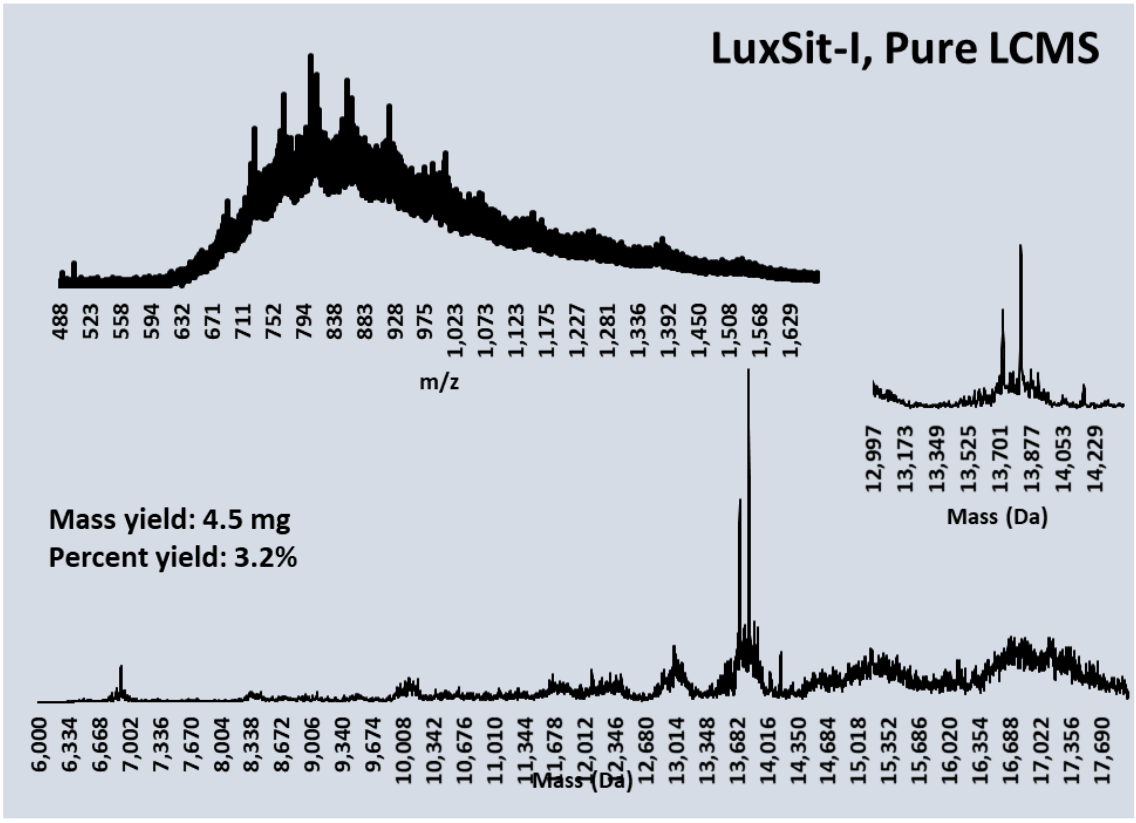


Figure 4.17: Scheme summarizing the development of LuxSit-I

Highlighted by the black dashed box on the top left are luciferases discovered in nature from organisms on the land and in the sea. On lower left is a box of native proteins that bind the small molecule diphenylterazine. Both of these pools were used to de novo design with deep learning the two artificial luciferases LuxSit-I and HTZ3-G4 (not pictured).

and HTZ3-G4.[58] De novo design of luciferases is set to be a great boon to the luciferase community since luciferases are so well suited to use in bioassays and biological imaging, similar to fluorescent proteins, however unlike fluorescent proteins luciferases have lacked the same development as molecular probes that fluorescent proteins have.[59] This is partly because of the complexity of developing luciferases that recognize non-native substrates and partly because of the relatively small number of well-understood and documented luciferases.[60] Thus, the Baker group sought to use machine learning to spearhead the development of small, highly stable, easy-to-express, and highly specific luciferases. To do this, they used a synthetic luciferin, diphenylterazine (DTZ)[61], and a novel deep-learning-based 'family-wide hallucination'[62] approach to generate a large number of protein structures containing an appropriately shaped

binding pocket for DTZ. These protein structures were then transformed into specific amino acid sequences that encode them. When I saw this publication in early 2023, I saw an exciting opportunity to expand my work with luciferases. I decided to synthesize and characterize LuxSit-I.



LuxSit-I sequence:

MSEEQIRQFLRRFYEALDSGDADTAASLFHPGVTIHLWDGVTFTSREEFREWFERLFSTSKDA
QREIKSLEVRGDTVEVHVQLHATHNGQKHTVDLTHHWHFRGNRVTEVRVHINPTG

Calculated mass: Da. 13821.269, Observed mass: Da. 13821.8

Figure 4.19: LCMS data for RP-HPLC purified LuxSit-I

On the top left is the MS data from purified LuxSit-I. The bottom contains the deconvoluted MS data, with the desired mass of LuxSit-I being the primary peak observed. In the middle is a zoomed-in version of the deconvoluted MS data indicating that there is minimal side product formation after RP-HPLC purification. LuxSit-I was recovered in good yield.

4.3.1 Synthesis and folding and characterization of LuxSit-I

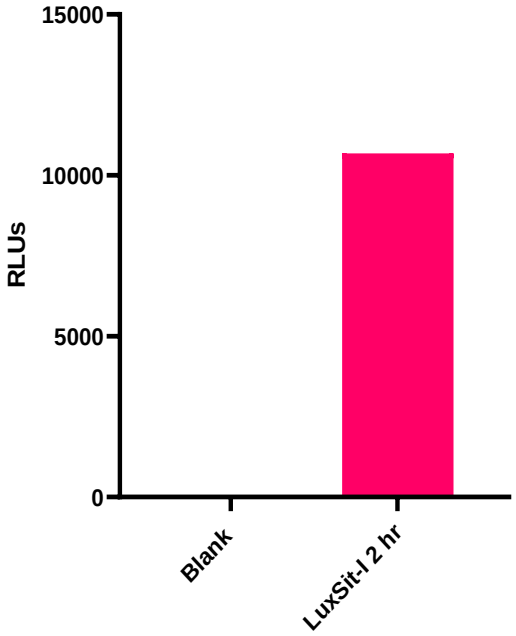


Figure 4.20: Bioluminescent assay results for initial LuxSit-I folding studies

This graph contains bioluminescent data from following the folding of LuxSit-I. LuxSit-I is not being directly compared to GLuc for our purposes due to these two proteins utilizing vastly different substrates, make direct comparison difficult. LuxSit-I was assayed after 2 hours in folding buffer and further incubation times did not yield brighter bioluminescence. Future studies could further optimize *in situ* folding conditions.

After its report, LuxSit-I became the shortest known luciferase protein polypeptide at only 117 amino acids in length. However, it is still heavier than picALuc at 13.9 kDa when compared to picALuc's 13 kDa. Nonetheless, these two miniature luciferases will be a great boon and tool for the luciferase community. Additionally, the ability to de novo design luciferases will enable a greatly expanded chemical space researchers can explore.

As always, I began my endeavors with LuxSit-I by first synthesizing it on the AFPS with RINK Amide low-loading resin, cleaving it to release the full-length protein polypeptide, and then

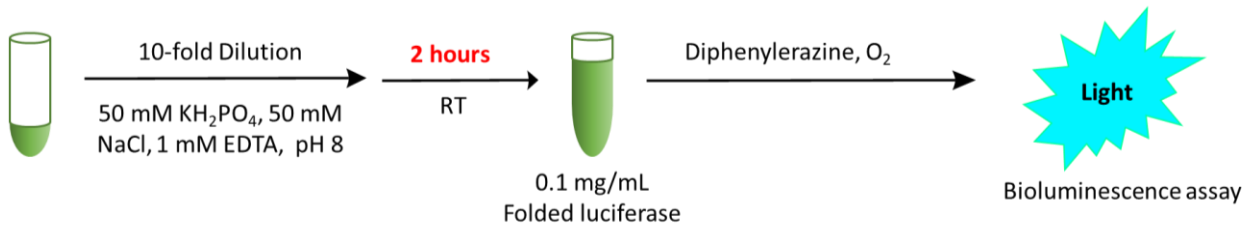


Figure 4.18: **LuxSit-I folding workflow utilized in this study**

submitting it to RP-HPLC purification using the same 5-65 purification method described above. Following LCMS analysis of the resulting fractions, I recovered 4.5 mg of purified protein

polypeptide, resulting in an overall yield of 3.2%. For protein folding I used an adapted version of our GLuc folding protocol, removing the redox pair GSH and GSSG since there is no disulfide bonds to be formed. For folding LuxSit-I was weighted from lyophilized polypeptide and dissolved in 6M guanidine hydrochloride, with 55 mM potassium phosphate as a buffering agent, at a concentration of 2 mg/mL. I then serially diluted LuxSit-I into a buffer of 50 mM sodium phosphate, 50 mM sodium chloride, and 1 mM Ethylenediaminetetraacetic acid (EDTA) at a pH of 8.0. The protein folding buffer mix was then thoroughly mixed and this solution then was left to sit at room temperature on a bench top for 2 hours. After 2 hours the LuxSit-I folding solution was submitted to a bioluminescent assay and LCMS analysis. A 24 hour timepoint was also taken and we saw no significant change in bioluminescent activity indicating LuxSit-I was fully folded after

With folding conditions in hand, I moved towards investigating a suitable bio purification strategy and, I once again turned to size exchange chromatography, using an isocratic elution

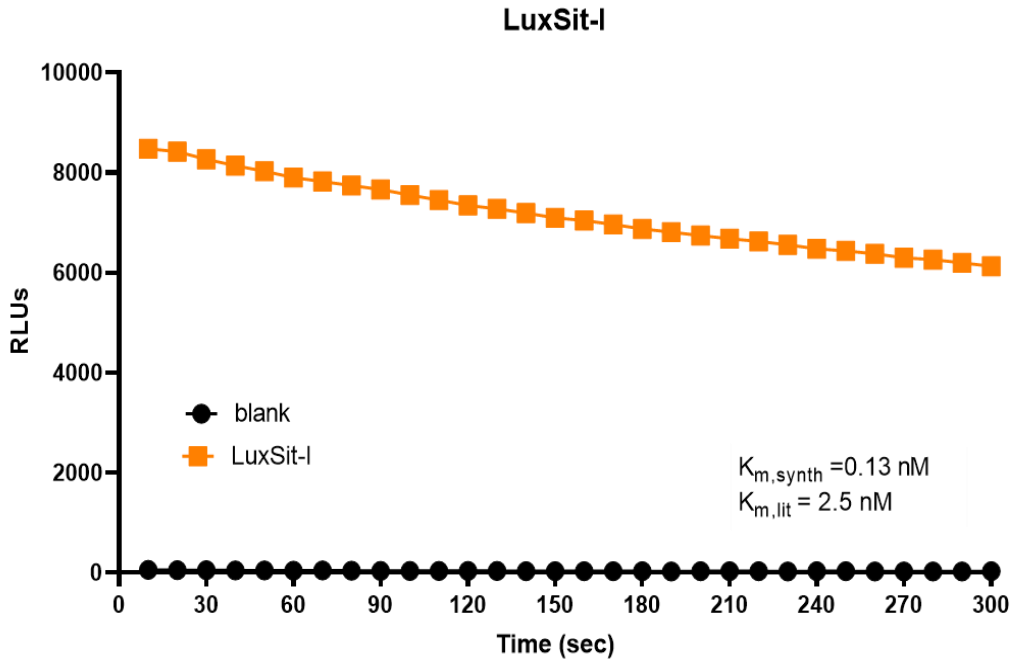


Figure 4.21: LuxSit-I luminescence decay with kinetic analysis

Bioluminescent assay results were collected with a Tecan Spark plate reader using an autoinjector to deliver the coelenterazine substrate. Upon addition of substrate, the solution was double orbital mixed for 5 seconds. Bioluminescent traces were collected over 5 minutes with a reading every 10 seconds. Kinetic analysis was accomplished using Michaelis-Menten kinetic curve fitting with Prism 10. Notice the slow bioluminescent decay, indicating this de novo artificial luciferase potentially follows a more traditional kinetics profile than the typical "flash kinetics" profiles from the other luciferases reported in this study.

buffer of 200 mM NaCl, 1 mM EDTA, 20 mM Tris-HCl pH 7.5. Following purification, I had sufficient material to move into further characterization studies.

Analysis of these proteins was also attempted using Michaelis-Menten kinetic analysis, a method the original reporters of LuxSit-I used to characterize this protein. We used varying concentrations of LuxSit-I's substrate diphenylterazine to assess kinetics, I ended up with a K_m value of 0.13 nM for the synthetic sample, and while this is lower than the literature reported value of 2.5 nM I believe this difference could be due to non-optimal folding conditions.

4.4 Synthesis and characterization of a mirror-image luciferase

In life, the critical biomacromolecules, proteins, are assembled by the ribosome from all L-amino acids. Further, the majority of the tree of life cannot produce or utilize D-amino acids in the cell. However, many therapeutics, biological reporters, and other protein-based endeavors suffer from the fact that native proteins are easily recognized by the cell and degraded.[63] This has resulted in researchers exploring the usage of D-amino acids in drug discovery.

In the Pentelute lab, our AFPS systems provide facile access to mirror-image proteins made from all D-amino acids that maintain the native biological activity of the parent L-protein [64] while imparting extraordinary protection against cellular degradation.[65] Not all enzymatic proteins are well suited for developing and applying mirror-image proteins due to substrate recognition and enzymatic turnover challenges. Nonetheless, researchers have successfully used synthetically produced mirror-image proteins for ligand binding-based drug discovery efforts.[66] To this end, I

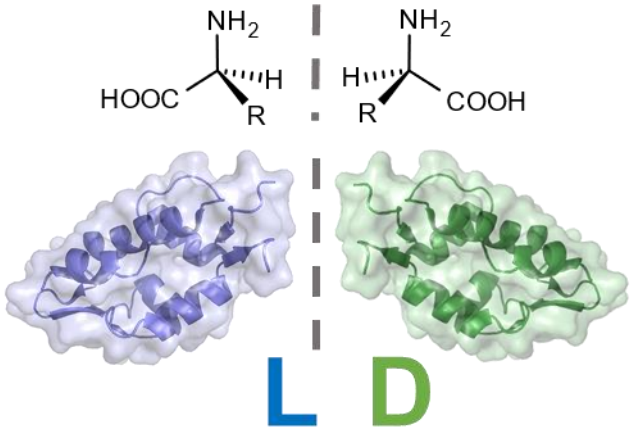


Figure 4.22: Structure depicting an L vs. D protein and small molecule pair.

The above diagram depicts an L and D protein and small molecule pair. L stands for "Levorotation", and the molecule is rotated counterclockwise when reading it. D, on the other hand stands for "Dextrorotation", and the molecule is rotated clockwise when reading it.

saw a distinctive opportunity in the family of luciferases I have been working with to synthesize the first ever reported mirror image luciferase in the form of D-picALuc.

D-picALuc is perfectly poised to be an exceptional first dive into mirror-image bioluminescent protein production due to its small size, bright luminescence, and, most importantly, its achiral substrate coelenterazine. Since coelenterazine is achiral, the mirror image luciferases I produce should display near-identical enzymatic activities as their native counterparts. In this section, I report on the single-shot synthesis of D-picALuc using AFPS technology. I envision that this groundbreaking development will lead to the development and application of more stable and robust molecular probes for biological research.

4.4.1 Synthesis and folding of D-picALuc

Synthesis of mirror image proteins is not accessible in biological systems. Traditionally, the only way to access full-length D-proteins was to undertake a complicated and time-consuming process using ligation techniques to join smaller peptide fragments together to create an entire full-length D-protein polypeptide. This endeavor is not only difficult but also very expensive and wasteful due to the high cost of D-amino acids and low yields commonly associated with chemical ligation techniques. Nonetheless, researchers have successfully produced enzymatically active full-length mirror image proteins, like the recent report of DNA-ligase, using chemical ligation

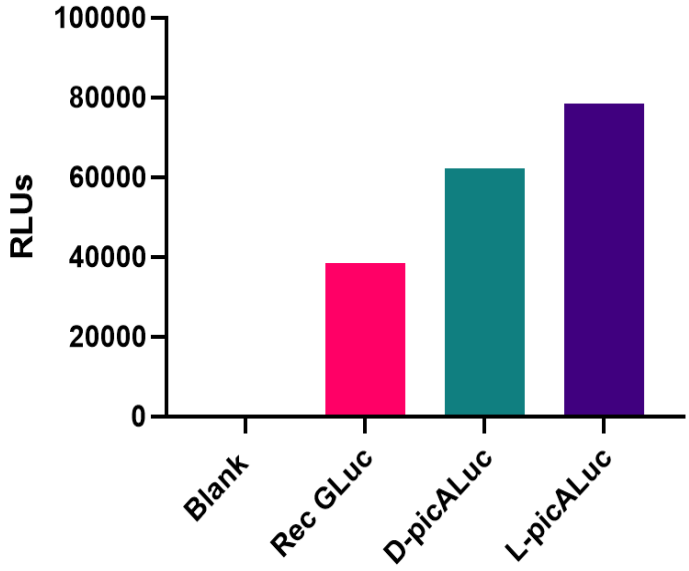
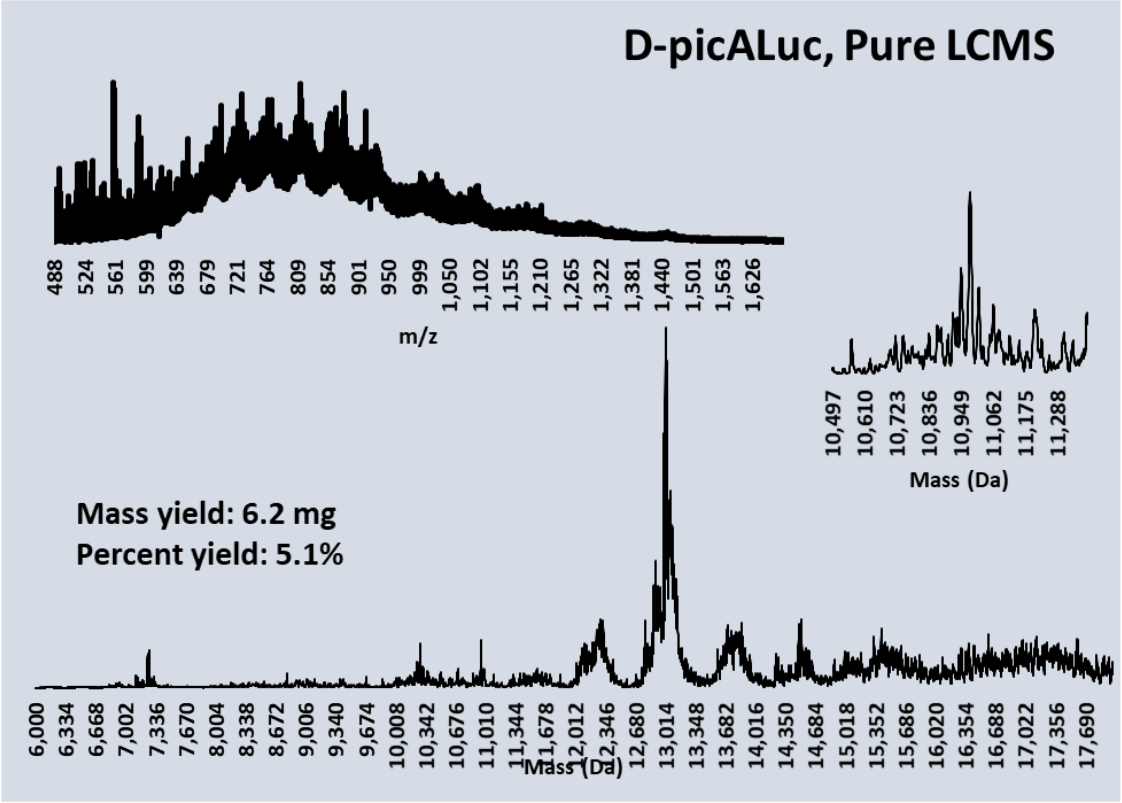


***Figure 4.23*:** Bioluminescent data from D-picALuc test folding

These results show use that our chemically synthesized D-picALuc is enzymatically active like its L-picALuc counterpart, we do notice a slightly smaller RLU value for D vs L due to using crude un-purified D-picALuc peptide and purified L-picALuc.

techniques.[67] With the AFPS, however, I can bypass the difficulties of chemical ligation and



picALuc sequence:

KLPGKKLPLEVLKELEANAQKAG**C**TRG**C**LI**C**LSHIK**C**TAKMKKWLPGR**C**ESWEGDKETGQ
GGIGEAIVDIPEIPGFKELAPMEQFIAQVD
L**C**AD**C**TTG**C**LKGLANVK**C**SALLKKWLPSR**C**

Calculated mass: 12999.460 Da., Observed mass: Da. 12999.7

***Figure 4.24*:** LCMS data of purified D-picALuc

On the top left is the MS data from purified D-picALuc. The bottom contains the deconvoluted MS data, with the desired mass of D-picALuc being the primary peak observed. In the middle is a zoomed-in version of the deconvoluted MS data, indicating that there is minimal side product formation after RP-HPLC purification. D-picALuc was recovered in good yield.

instead produce D-protein polypeptides in a single shot. While this endeavor is an expensive undertaking, it could be a massive boon to drug discovery and biological research. The Pentelute Lab has a publication by Callahan and coworkers in pre-print documenting the production of over 12 unique single-domain mirror image proteins, which compromises nearly one-third of the total D-proteins reported to date.[68] With proof of concept D-protein synthesis in hand, I turned my attention to broadening our D-protein synthesis efforts with D-picALuc. This presented a quite unique opportunity to produce the first-ever synthesized mirror image bioluminescent protein.

My first step was to synthesize D-picALuc using our automated fast-flow peptide synthesizers loaded with ultra-pure D amino acids under the same conditions described in section 4.2. I used RINK amide low-loading resin (0.18 mmol/g). I felt confident in achieving synthetic success with D-picALuc since L-picALuc was produced in good quality with high yields on the AFPS. D-picALuc was synthesized under standard AFPS conditions and then stored under vacuum until peptide cleavage. Following peptide cleavage, picALuc was purified via high--performance liquid chromatography (RP-HPLC) using a standard 5-65% B over 72 minutes method (Solvent A is water + 0.1% TFA, and Solvent B is Acetonitrile in 0.1% B), and resulting fractions were analyzed by liquid chromatography-mass spectrometry (LCMS). Following analysis, clean fractions were pooled, and 6.2 mg of D-picALuc polypeptide was recovered at a calculated yield of 5.1%. Some of the initial folding and assay efforts described in the following paragraphs involved using crude D-picALuc to optimize folding and assay conditions, whereas the L-picALuc used was RP-HPLC purified.

Following the successes with my native L-luciferases, I used my previously established folding protocol to fold D-picALuc. I used a redox buffer containing GSH: GSSG in a ratio of 3:2,

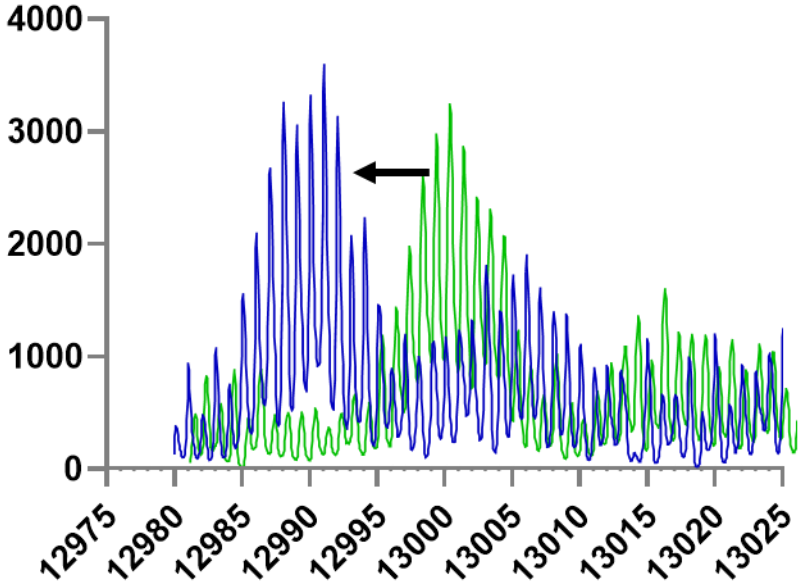


***Figure 4.25*:** Mass shift LCMS data for folded D-picALuc

Mass shift data for D-picALuc was showing the mass shift of folded picALuc after disulfide bond formation, it is important to note that bond formation does not guarantee proper folding for enzymatic function. Data was obtained by analyzing LCMS traces with a 0.05 Da resolution.

with a total of 5 mM glutathione redox buffer total, in the folding buffer. First, lyophilized HPLC purified D-picALuc was weighted and dissolved in 6M guanidine hydrochloride, with 55 mM potassium phosphate as a buffering agent, at a concentration of 2 mg/mL. Then, the folding buffer

was prepared, added to the dissolved peptide, and thoroughly mixed. This peptide buffer solution was then left to sit at room temperature on a bench top for 24 hours, and protein concentration was assessed via A280 using Beer's law. The folded proteins were then submitted to bioluminescent assay analysis.

I used LCMS analysis to monitor disulfide bond formation of the five disulfide bonds by looking for a mass shift of approximately 10 Da, which would indicate that all the cysteine residues are now in an oxidized state. For the-post folding bioluminescent assay, I diluted D-picALuc to 100 nM and introduced its substrate coelenterazine at a volume of 50 uL and a concentration of 3 uM. These values can be adjusted as needed. D-picALuc demonstrated the expected 10 Da mass shift, indicating disulfide bond formation. Further, the bioluminescent assay data shown in Figures 4.23 and 4.27 suggests that the disulfide bonds formed to produce an active protein and did not misfold. This data was collected using crude, i.e., just cleaved from the resin, D-picALuc polypeptide. I now have enzymatically active D-picALuc in hand. In the next section, I describe my efforts toward further characterization of D-picALuc and direct comparisons of D-picALuc to L-picALuc.

4.4.2 Characterization of D-picALuc

With the confirmation of the enzymatic function of my mirror image luciferase, I can move to size exchange chromatography using an isocratic elution buffer of 200 mM NaCl, 1 mM EDTA, and 20 mM Tris-HCl pH 7.5. Following purification, I had sufficient material to move into further characterization studies with D-picALuc and L-picALuc. Some of the studies described below were performed using only RP-HPLC purified peptide. In contrast, others were performed using SEC-purified material, and the traces in Figure 4.27 were collected with crude unpurified D-picALuc peptide.

First, I obtained kinetic traces of L-picALuc and D-picALuc luciferases using a newly installed TecanSpark reader with an autoinjector. With the TecanSpark, I can also observe the slow decay of luminescence over time as coelenterazine is oxidized to coelenteramide. The bioluminescent traces shown in Figure 4.27 are from SEC-purified L-picALuc and crude, freshly freed from resin, D-picALuc. I was able to document the similar behavior of our D-picALuc to L-

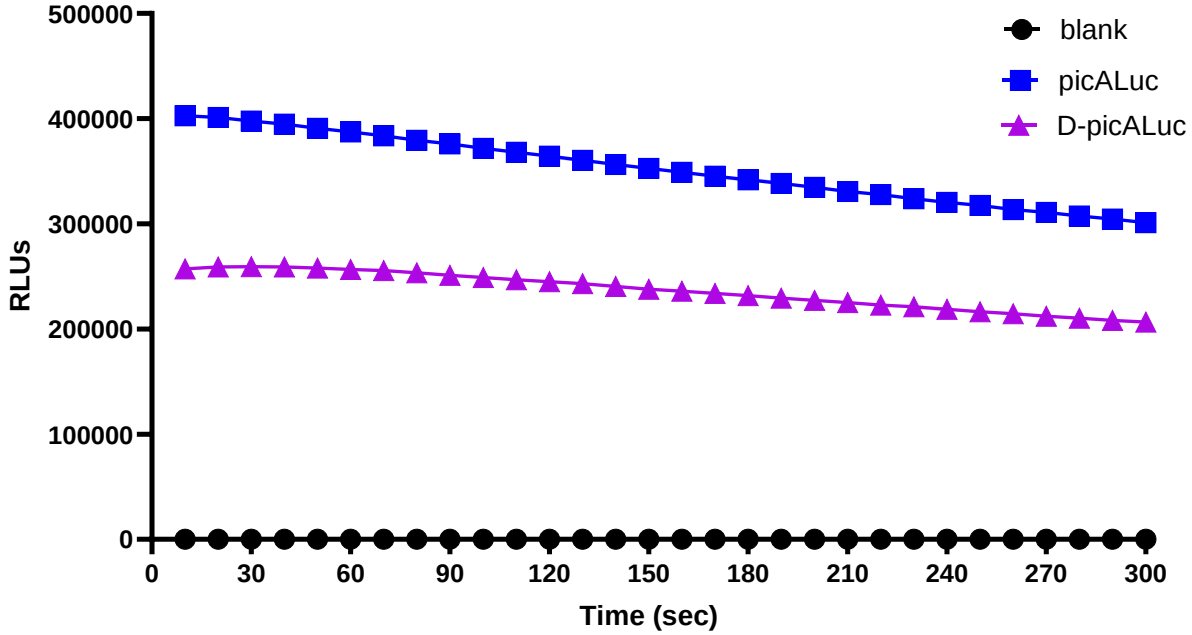


Figure 4.26: Bioluminescent traces for folded unpurified D-picALuc and purified L-picALuc

Bioluminescent assay results collected with a Tecan Spark plate reader using an autoinjector to inject CTZ substrate and mix the substrate for 5 seconds. Traces were collected over the course of 5 minutes with a reading every 10 seconds. Notice the similar decay patterns of both luciferases over time, D-picALuc does show lower bioluminescence, but this is likely linked to the unpurified nature of D-picALuc.

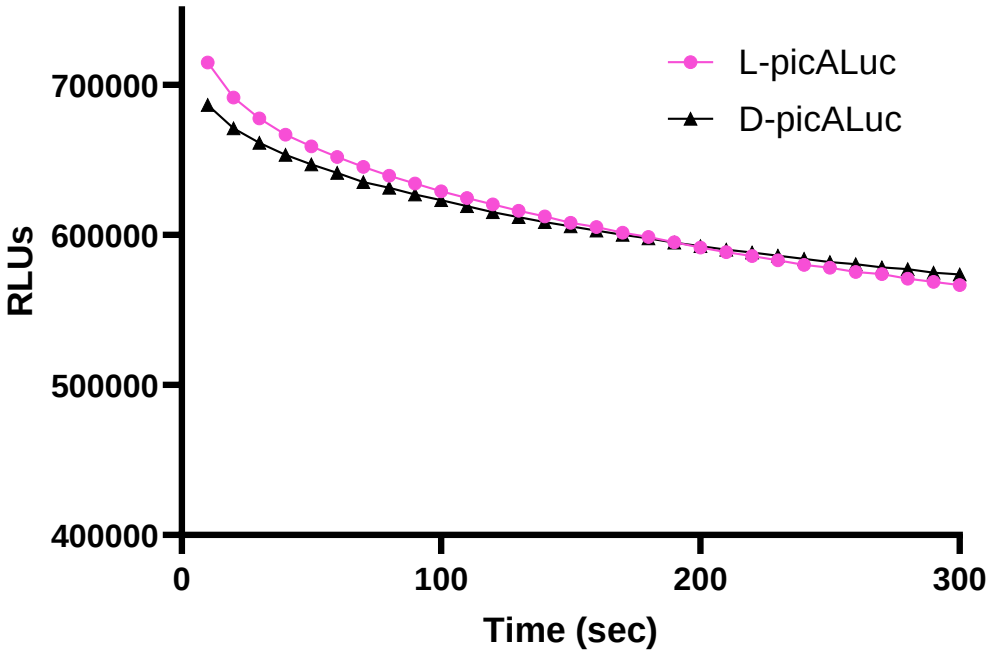


Figure 4.27: Bioluminescent traces for folded SEC purified D-picALuc and L-picALuc

Bioluminescent assay results collected with a Tecan Spark plate reader using an autoinjector to inject CTZ substrate and mix the substrate for 5 seconds. Traces were collected over the course of 5 minutes with a reading every 10 seconds. Notice the near identical decay patterns of both luciferases over time.

picALuc; however, similar to my early folding studies, I observed lower activity from D-picALuc, but this is clearly due to using un-purified crude peptide for data collection. In Figure XX, a similar bioluminescent trace was obtained on the TecanSpark with SEC-purified L-picALuc and D-picALuc, and I was able to demonstrate that both synthetically produced samples enzymatic activity is nearly indistinguishable from the other.

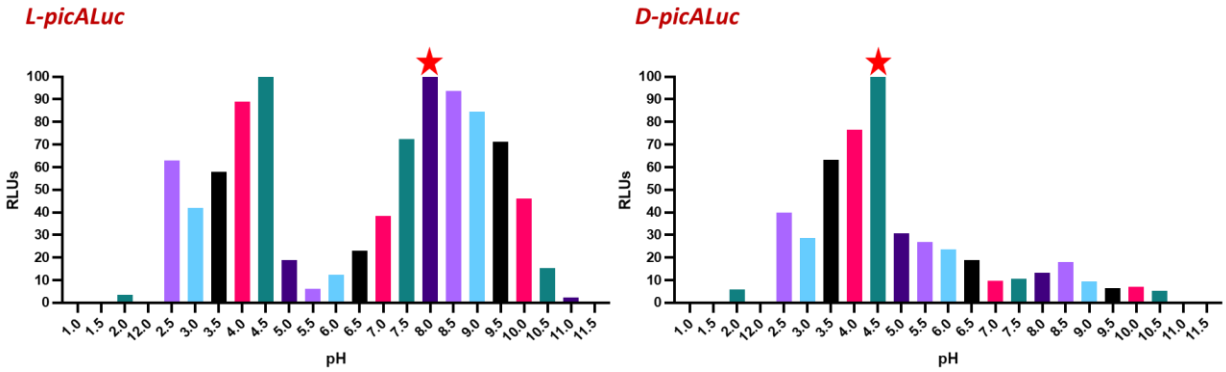


Figure 4.28: Preliminary pH study data for D and L picALuc

Luciferase activity was monitored at a variety of pH's from pH 1-12, data was normalized to the highest value in each run. I noticed a marked difference in the stability of D-picALuc at higher pH values and are investigating the cause.

Next, I moved into defining the biophysical properties of L-picALuc and D-picALuc. Similar to the original report for picALuc, I decided to study enzymatic activity at various pHs and temperatures. Some preliminary data showed, unfortunately, inconsistencies, and further investigation will be needed, or this pathway abandoned.

Finally, I hoped to demonstrate the durability of D-proteins in cellular environments by exposing them to proteases. Studying the proteolytic cleavage of L-picALuc and D-picALuc with a variety of different proteases will allow me to study enzymatic activity decay over time. I hope to see D-picALuc outlast L-picALuc by a significant margin. Three proteases, trypsin, chymotrypsin, and clostripain, were selected for these studies. Trypsin is a serine protease that cleaves at arginine and lysine. Chymotrypsin is closely related to trypsin and cleaves at tryptophan and tyrosine residues. Lastly, clostripain, a more recently discovered and characterized protease, cleaves specifically at arginine residues.

The protocol for this study involves incubating folded RP-HPLC purified D-picALuc and L-picALuc at 37 ℃, with a 1:10 by mass ratio of each protease (protease: luciferase). I took bioluminescence readings every 48 hours to measure decay over time. I predicted that the L-picALuc would retain some activity for at least two days, as most copepod luciferases are stable in cellular environments for that time, after which we should see a steep activity drop-off. Fortunately, my results fully supported this prediction, with all three L-picALuc samples retaining measurable amounts of bioluminescent activity on day two but little to no activity after four days. Only the L-picALuc sample incubated in trypsin retained enzymatic activity after day four.

With this data, I feel confident that I have successfully produced the first-ever reported mirror image bioluminescent protein that shows near-identical enzymatic activity to the native L-protein. Furthermore, D-picALuc exhibited high tolerance to degradation by proteases, leaving D-picALuc and other mirror-image proteins perfectly poised to be applied to biological assays, like

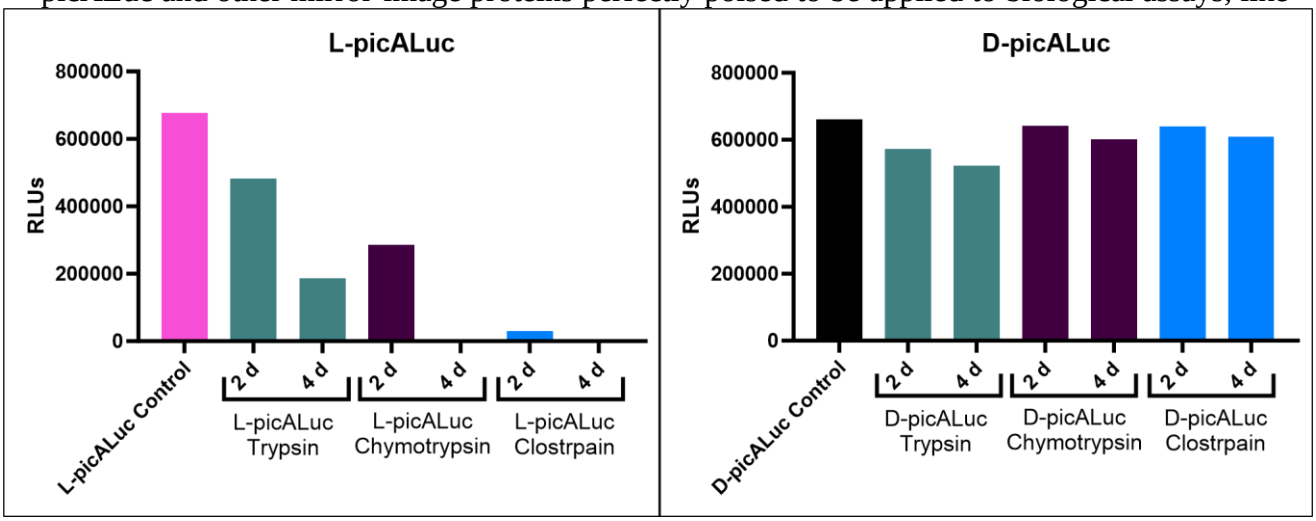


***Figure 4.29*:** L-picALuc and D-picALuc protease assay results

Luciferase activity for RP-HPLC purified L-picALuc and D-picALuc was monitored over a four-day period while included at 37°C with one of three proteases, Trypsin, Chymotrypsin, and Clostripain. As I predicted L-picALuc degraded significantly faster than D-picALuc with D-picALuc retaining near full activity after 4 days.

cellular trafficking, where a durable molecular probe is needed. I look forward to seeing further applications of D-proteins in the Pentelute Lab.

4.5 Development of a picALuc protein complementation assay

Proteins and their interactions play a pivotal role in cellular dynamics and the ability for organisms to respond to outside stimuli. For instance, protein signaling cascades stimulate the cell's response to illness and infection or changing environmental factors like heat or cold.[69] However, measuring, understanding, and investigating these vital protein-protein or protein-biomacromolecule interactions is extraordinarily difficult due to the sheer complexity of these cellular processes.[70] To peer into the complex interplay of cellular machinery in a way that does not interrupt the native cellular processes, researchers have turned to using split reporter modalities.[71]

A split report functions by taking a large protein or other molecule and splitting it in such a way that when the two halves come back together, there is a recovery of the native activity of the reporter.[72] However, this binding activity cannot disrupt the native binding affinities of the

interaction being studied. Otherwise, the information gained will not be accurate to the protein-protein interaction being studied. Historically, researchers have used split fluorescent reporters[73] and even enzymatic reporters like β-Galactosidase to study these cellular interactions. Yet, these methods can be non-ideal since many fluorescent split reporters do not reversibly associate, and using enzymes generally relies on measuring the accumulation of an enzymatic product.[74] Nonetheless, split reporters have seen great success in the research community and have illuminated many important intracellular interaction.[75]

To develop more reliable, quantitative, and dynamic reporters' scientists have historically turned to split luciferase reporter constructs to shine a light on many difficult-to-study protein-biomacromolecule interactions.[76] Split luciferase assays have a marked advantage over split fluorescent protein and enzymatic reporter assays. Not only does using a luciferase circumvent the risk of photobleaching or phototoxicity often associated with the light excitation needed for fluorescent proteins, but luciferases also have a much higher signal-to-noise ratio, making data analysis and presentation far smoother.[77]

Generally, the two most common split luciferase systems have been that of firefly luciferase, FLuc, and *Renilla* luciferase, RLuc, and both have been heavily optimized and applied to a variety of systems. For example, a FLuc split system was used by Chen and coworkers to develop and validate a system to detect a variety of protein-protein interactions in plants to aid researchers in studying disease tolerance and understanding how to engineer the best possible pesticides and disease-resistant food-producing plants.[78] In addition, split firefly luciferase systems have undergone a massive amount of development to produce the most stable and reliable split protein fragments to use for *in vivo* studies. In one study, scientists used a combinatorial library

screening approach to create a split FLuc system that is highly generalizable, has a low background signal, and a high signal after association with the protein-protein interaction being studied.[79]

However, the one massive disadvantage of split FLuc and RLuc systems is their sheer size, as each complementary fragment of the split system can be 35 kDa in size or more. The size of these fragments can result in these split complementation assays interfering with native protein-

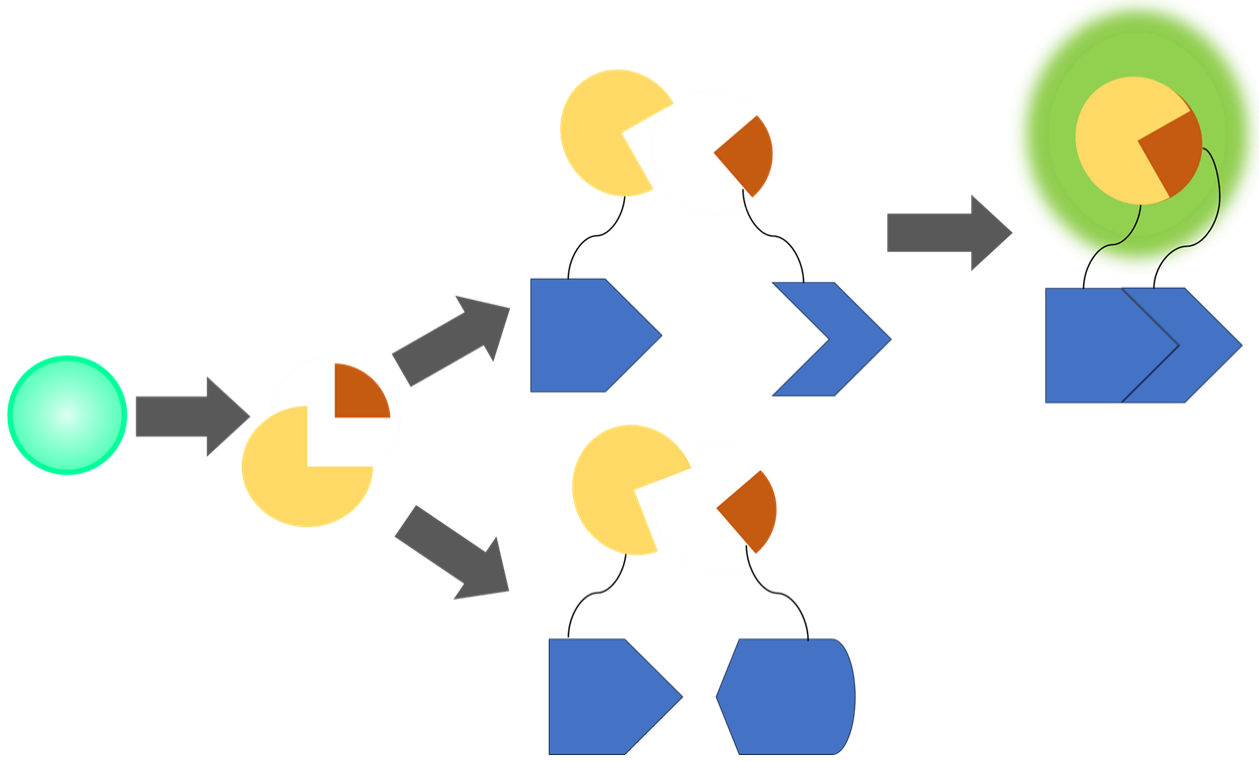


Figure 4.30: Scheme depicting the general pathway for a split luciferase system

Split luciferase systems work by taking a luciferase enzyme and selecting an appropriate split point at which neither fragment is enzymatically active on its own, but when re-associated the luciferase then performs its native function. The larger of the two fragments is generally referred to as the "predator" and the smaller of the two the "prey". Note in the scheme above that if the two proteins or other biomacromolecules the luciferase "predator" and "prey" are attached to do not associate no bioluminescence will be observed. If the two biomacromolecules associate I expect to see a measurable and quantifiable uptick in bioluminescence.

protein interactions. So, in an ideal split luciferase system, the two fragments would not only produce substantial bioluminescence once bound, but the two fragments of the luciferase would also be rather small themselves. This is so that the fragments would have minimal influence on the native affinity and association kinetics of the interacting proteins being studied.[80] With this in mind, researchers at the Promega Corporation utilized the small and very bright luciferase NanoLuc (19 kDa) to develop a split luciferase reporter system, the smallest of its kind. From this work came the NanoBiT reporter, which is now commercially available.[81] During the COVID-19

pandemic, several research groups utilized this small, versatile system to detect anti-SARS-CoV-2 antibodies to create a rapid, quantitative, low-cost, easy-to-use diagnostic tool. This is just one example of the NanoBiT system being applied, and there are countless others.

Thus, there is a clear interest in small, stable, and versatile split luciferase systems that are of crucial importance and highly sought after. My work with picALuc has inspired me to pursue my own robust, bright, and extraordinarily small split luciferase system. With picALuc being the smallest luciferases reported to date, I should be able to develop the smallest predator-prey split luciferase pair. In addition, with the library screening technology available in the Pentelute laboratory, I can envision developing and screening large compound libraries to discover completely non-natural peptidomimetic binders to expand the scope of split luciferase assays. Using non-natural amino acids in a split protein complementation assay binder will confer a great deal of *in vivo* stability and further expand the areas where these assays can be applied. I believe that by diving into the development of a split picALuc reporter, I will add a much-needed tool to

the biologist's toolkit. In the following sections, I discuss the progress toward this goal and my outlined vision for future experiments.

4.5.1 Design and evaluation of luciferase split site

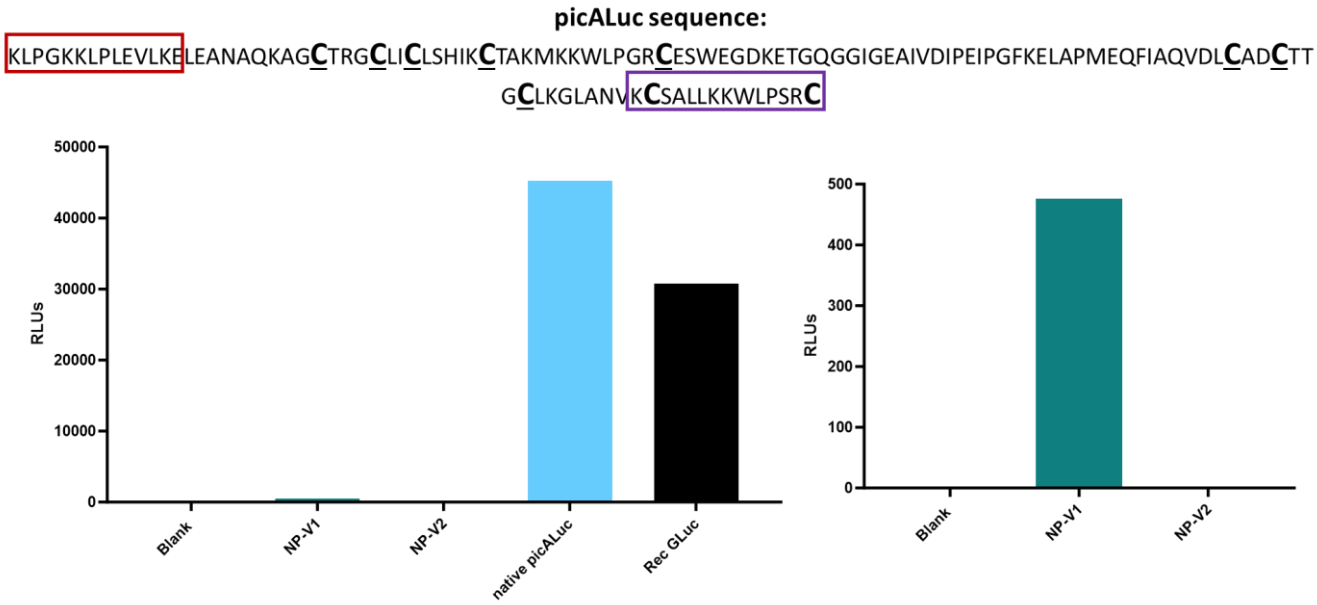


Figure 4.31: Initial assay results from our two potential split luciferase constructs

This figure shows the split sites assessed in my studies. NP-V1 is boxed in purple and NP-V2 is boxed in red. The results from a 1:1000 APO-picALuc: NP incubation for 2 hours show recovery of bioluminescent activity with the NP-V1 variant. This variant was moved into further studies, and while initial bioluminescent recovery appears low, this is actually an ideal outcome and will hopefully result in low interference with the protein-protein interaction being studied when this is moved into *in vivo* studies.

The first step in developing a split luciferase is to find an appropriate site to cleave the full-length protein such that both fragments are stable and neither display any inherent bioluminescent activity. Most importantly, the two predator and prey fragments must be able to come together and, when associated, restore native bioluminescent activity. To this end, I was blessed to have a beautiful dataset from the first report of picALuc showing that all bioluminescent activity would be lost in the protein if either 14 amino acids from the N- or C- terminus were deleted. With this in mind, I synthesized two versions of my own initial split luciferase contrast to test. I named the shorter 14-mer peptide fragment the "native peptide," NP-V1 for the C-terminal peptide and NP-V2 for the N-terminal peptide. The corresponding more giant "predator" fragments were named APO-picALuc-V1 and APO-picALuc-V2.

I synthesized via AFPS, cleaved, and RP-HPLC purified all four polypeptides using the methodology discussed in this chapter's previous sections. Next, I prepared for my first assessment of whether either NP fragment would be appropriate for further development into a split luciferase

construct. For this assessment, I incubated each NP peptide with its corresponding APO-picALuc variant at a 1:1000 APO-picALuc: NP ratio for 2 hours at room temperature. Each incubated pair was then submitted to bioluminescent assay analysis to assess for any recovery of native enzymatic activity. Excitingly, the initial screens showed promise in a C-terminal 14-mer peptide, NP-V1. NP-V1 showed partial bioluminescent enzymatic recovery and was appropriate to move forward into further studies.

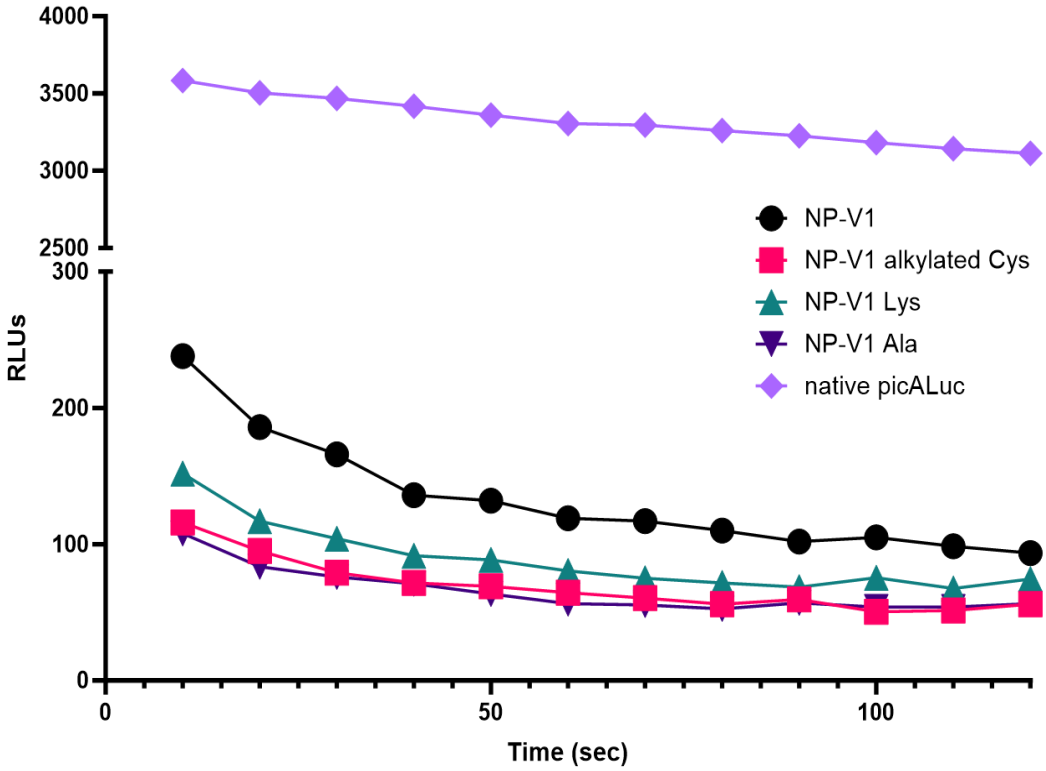


Figure 4.32: Assay results from analysis of cystine residues in NP-V1

The data presented above are the results of the binding tests to assess if the native cysteines are necessary for bioluminescent activity recovery. This data was collected over a 2-minute kinetic read on a TecanSpark plate reader. Notice that the three peptide variants of NP-V1, which while slightly less active than the original peptide, still retain good bioluminescent activity recovery. I will move forward with the NP-V1A variant for future studies to avoid any cystine cyclization.

My first test following the decision to pursue NP-V1 was to assess if the cystines within the sequence, which form a disulfide bond in the native folded protein, are necessary for activity recovery. To test this, I first alkylated both cystines to see if removing the disulfide bond would prevent enzymatic recovery. Next, I synthesized a peptide with both cysteines replaced with alanine to assess whether the cystine chemical structure was required for recovery of bioluminescent activity. Finally, I synthesized an NP-V1 variant with a C-terminal lysine since I intend to move into library screening studies, and one lysine residue is always added to aid in

peptide ionization. My main conclusion from this test was that not only would an additional lysine not interfere with peptide binding, but the cystines and the respective disulfide bond was also

KASALLKKWLPSRAK	15 AA	NP-V1A
KASALLKKWLPSRK	14 AA	NP-V1A-1C
KASALLKKWLPSK	13 AA	NP-V1A-2C
SALLKKWLPSRAK	13 AA	NP-V1A-2N
SALLKKWLPSRK	12 AA	NP-V1A-1C-2N
SALLKKWLPSK	11 AA	NP-V1A-2C-2N

Table 4.1: Table of shortened NP-V1A peptides tested

wholly unnecessary for binding as well. This will make future library design more manageable since I will not have to worry about my peptide macrocyclizing, which would greatly convolute data analysis and limit potential applications of this split luciferase construct. NP-V1A, depicted in Table 4.1, was the sequence that resulted from this study.

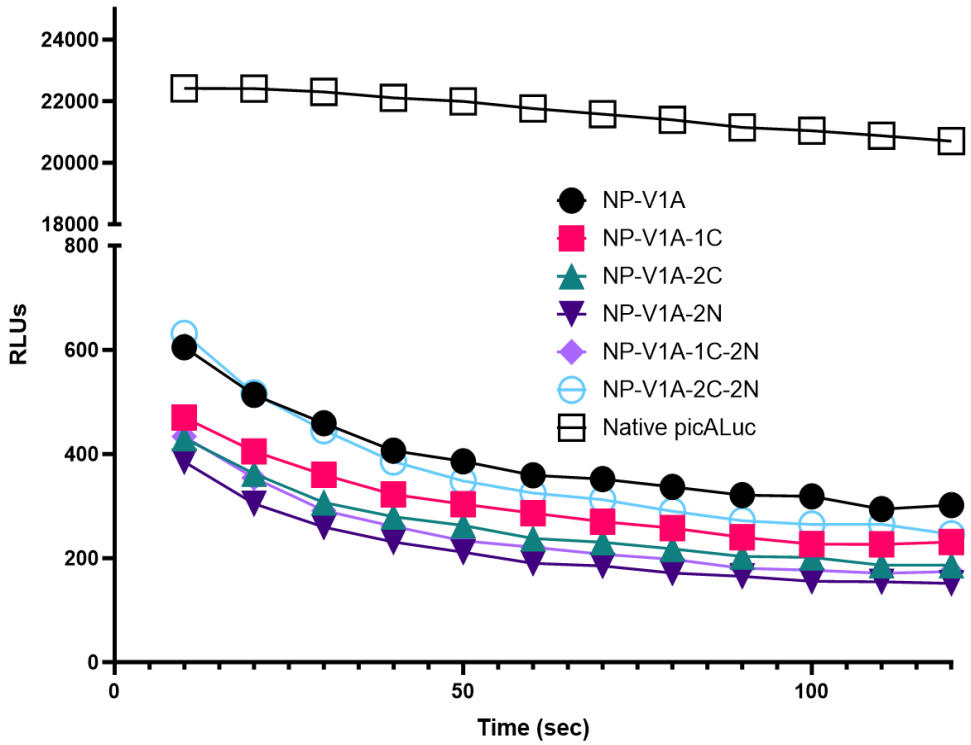


Figure 4.33: Bioluminescent assay data from shorten NP-V1A variants

The data presented above are the results of the binding tests to assess shortening the NP-V1A prey peptide will still allow for bioluminescent activity recovery. This data was collected over a 2-minute kinetic read on a TecanSpark plate reader. Notice that the shorter peptide variants of NP-V1A, retain good activity that is comparable to the full-length 14-mer peptide. I will move forward with the NP-V1A-10 variant for future studies.

Next, I was inspired and challenged by the NanoBiT system once again and desired to make my "prey" peptide, currently a 14-mer, into something shorter, similar to the NanoBiT 11-mer size. To this end, I iteratively deleted residues from the N- and C-terminus of NP-V1A and assessed bioluminescent activity recovery. Strikingly, the shortest peptide tested, only ten amino acids in length, was sufficiently capable of enzymatic recovery after a 2-hour incubation period. I decided to move forward with this peptide into further library design and screening.

The following steps of this project will involve assessing specific binding affinities, K_D, of all of the peptides tested above. This will be done by screening the predator fragment, APO-picALuc-V1, by titrating against the prey peptides at different concentrations. K_D values can then be determined by fitting the data to a one-site-specific binding model in Prism 10 or similar software.

4.5.2 Moving towards library screening for peptide binders

Chemically accessed libraries are an excellent alternative to genetically encoded libraries for new ligand discovery. Solid-phase organic synthesis offers a powerful technique for generating peptide libraries that sample broad chemical diversity. There are two primary approaches for mining peptide libraries generated from solid-phase synthesis: on-bead screening, commonly referred to as the one-bead-one-compound (OBOC) approach, and in-solution affinity selections.[82]

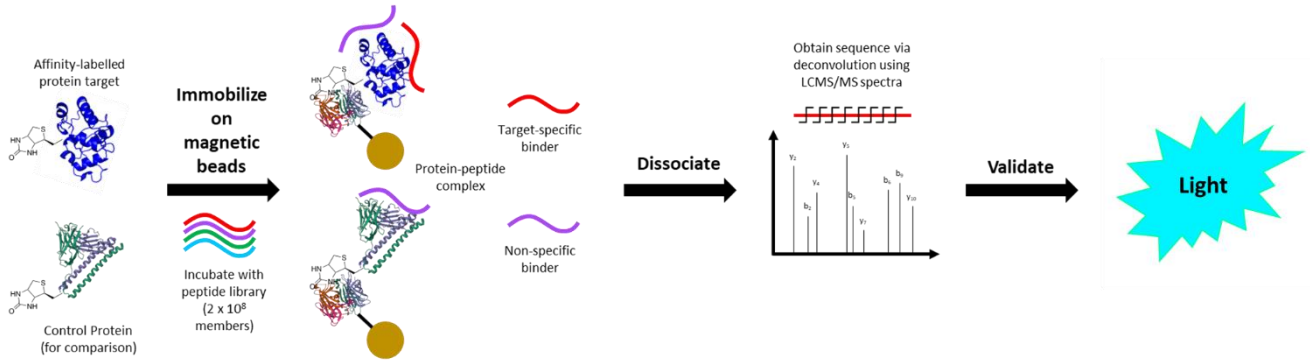


Figure 4.34: AS-MS workflow to discovery novel peptide binders for a split luciferase

This figure walks through the AS-MS workflow for de novo peptidomimetic binder discovery. Step one requires anchoring the APO-picALuc-V1 protein polypeptide to magnetic beads using a biotin tag on the polypeptide and a streptavidin tag on the magnetic beads. For this step, a control protein must be used for comparison to assess for any off-target binding effects. Once secured to the magnetic bead APO-picALuc and the control protein will both be incubated with the desired peptide library. After incubation, any bound peptides are dissociated from the protein, and the sequence is identified using LCMS/MS spectra via OrbiTrap. Any identified "hit" peptides will be synthesized via AFPS and validated by the bioluminescent assay protocol described herein.

In the Pentelute group we use affinity selection-mass spectrometry (AS-MS), using magnetic bead reagents, to identify high-affinity binders from randomized libraries of 10^8 synthetic peptides. With respect to accessible diversity, this synthetic library design is on par with molecular biology-based combinatorial libraries. Diversity is a key determinant of selection outcome, so careful handling of the spilt-pool synthesis is key. The primary benefit of synthetic peptide libraries is that I have great chemical control over the library design and can easily synthesize and screen libraries containing multiple non-canonical amino acids, something not possible with traditional library screening methods.[83]

To this end, I propose synthesizing and evaluating two $X_{10}K$ peptide libraries, one containing all D-amino acid peptides and one with non-canonical amino acids incorporated. I have the power with a bioluminescent protein to validate binders quickly and accurately for specific binding via luminescence assays instead of the typical biolayer interferometry, and exciting new angle of binding analysis for our group. I hope to be able to develop and report on three peptide binders for a split luciferase construct, one with canonical L amino acids, one entirely made of D amino acids, and one non-canonical binder. This will have a substantial impact on future split luciferase studies looking at anything from protein-protein interactions in the cell to complex drug delivery systems with my non-native peptide tags.

4.6 Discussion and Conclusion

This chapter reports on my efforts on the synthesis and characterization of the deep-sea marine copepod, *Gaussia priceps*, luciferase GLuc (18 kDa), and artificial luciferases picALuc (12 kDa), ALuc30 (21 kDa) and LuxSit-I (14 kDa) via automated fast flow peptide synthesis (AFPS). In addition, I synthesized the mirror-image counterpart of picALuc due to its potential for broad-reaching impact in health and diagnostics. This is the first known report of a mirror-image bioluminescent luciferase. Finally, I report on my efforts to develop a split-picAluc protein complement assay, which will be the smallest and most versatile split-luciferase reported to date. To my knowledge, these are the first luciferase enzymes produced via solid-phase peptide synthesis (SPPS). This will allow me and other researchers to delve into the space of developing customizable and durable luciferases that can be used for future research applications.

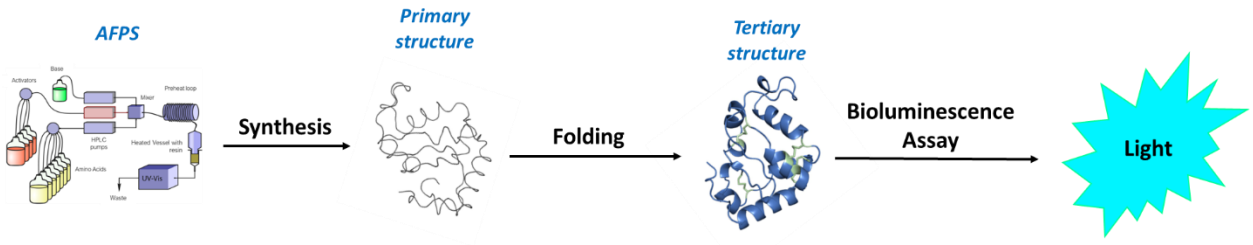


Figure 4.35: Workflow to produce and validate function for synthetic artificial luciferase

With my robust, widely available routine method for the chemical production of proteins, the Pentelute lab continues to be poised to have a substantial impact on chemical biology and the development of new therapeutics. Our advances provide a viable solution to reliably assemble long linear peptide chains, shifting the focus in the field of chemical protein synthesis to improving folding protocols and, most importantly, applications. Further applications of the AFPS technology will continue to allow researchers to enter the unknown, with the sky as the limit. With this report on the synthesis of luciferases and the design of a protein complement assay, I have made a significant contribution to the field of peptide chemistry. I have also added valuable tools into the pockets of biologists using luciferases for groundbreaking discoveries around the world.

4.7 Future directions

The synthesis of luciferases in the Pentelute laboratory has opened the door to new techniques within the Pentelute group itself. In the near future, I hope D-picALuc will be utilized in animal studies to compare its stability and potential toxicity to that of L-picALuc *in vivo*.

Furthermore, I would be so excited to be able to grow crystals of both L- and D-picALuc and report on the x-ray crystallography structure of these two proteins, another first for the field.

I also envision applications of our PCA design being utilized to have a less work-intensive way to assess drug delivery *in-vivo* using bioluminescent imaging techniques to image live animals without the need to sacrifice animals to evaluate drug delivery levels. In my own group, in fact, I can see applications to studying cell-penetrating peptide delivery into organs and cells.

Finally, one of the group's younger members is spearheading the application of NMR analysis of substrate turnover via saturated transfer difference to develop a facile, quick, and accurate method to study luciferase flash kinetics. Further, this student is also moving LuxSit-I into reusable glass bead peptide synthesis as a quick re-out method to access synthetic success as we optimize the AFPS for different reusable peptide resin substitutes. I am excited to see where else the research community can apply my luciferases and the AFPS technology. I hope to leave a lasting legacy in these small bioluminescent proteins.

4.8 Materials and Methods

Below are the materials and methods for this section, these methods may not apply to other sections.

4.8.1 Reagents and solvents

All reagents were purchased and used as received. Fmoc-protected amino acids (Fmoc-Ala-OHxH_2O, Fmoc-Arg(Pbf)-OH; Fmoc-Asn(Trt)-OH; Fmoc-Asp-(Ot-Bu)-OH; Fmoc-Cys(Trt)-OH; Fmoc-Gln(Trt)-OH; Fmoc-Glu(Ot-Bu)-OH; Fmoc-Gly-OH; Fmoc-His(Trt)-OH; Fmoc-Ile-OH; Fmoc-Leu-OH; Fmoc-Lys(Boc)-OH; Fmoc-Met-OH; Fmoc-Phe-OH; Fmoc-Pro-OH; Fmoc-Ser(But)-OH; Fmoc-Thr(t-Bu)-OH; Fmoc-Trp(Boc)-OH; Fmoc-Tyr(t-Bu)-OH; Fmoc-Val-OH), N-α-Fmoc-L-α-aminobutyric acid (Fmoc-Abu-OH) and N-α-Fmoc-L-norleucine (Fmoc-Nle-OH) were purchased from the Novobiochem-line from Sigma Millipore; FmocHis(Boc)-OH was purchased from ChemPep, Inc Fmoc-protected D-amino acids (Fmoc-Ala-OH x H2O, Fmoc-Cys(Trt)-OH, Fmoc-Asp(OtBu)-OH, Fmoc-Glu(OtBu)-OH, Fmoc-Phe-OH, Fmoc-His(Boc)-OH, Fmoc-Ile-OH, Fmoc-Lys(Boc)-OH, Fmoc-Leu-OH, Fmoc-Met-OH, Fmoc-Asn(Trt)-OH, Fmoc-Pro-OH, Fmoc-Gln(Trt)-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Ser(tBu)-OH, Fmoc-Thr(tBu), Fmoc-Val-OH, Fmoc-Trp(Boc)-OH, and Fmoc-Tyr(OtBu)-OH) were purchased from both the Novabiochem-line from Millipore-Sigma and Chem-Impex. O-(7-azabenzotriazol-1-yl)-N,N,N’,N’-tetramethyluronium hexafluorophosphate (HATU, ≥97.0% and

(7-azabenzotriazol-1- yloxy)tripyrrolidinophospho-nium hexa-fluorophosphate (PyAOP, ≥97.0%) were purchased from P3 Biosystems. Biosynthesis OmniSolv® grade N,N-dimethylformamide (DMF) was purchased from EMD Millipore (DX1732-1). AldraAmine trapping agents (for 1000 – 4000 mL DMF, catalog number Z511706), Diisopropylethylamine (DIEA; 99.5%, biotech grade, catalog number 387649), piperidine (ACS reagent, ≥99.0%), trifluoroacetic acid (HPLC grade, ≥99.0%), triisopropylsilane (≥98.0%), acetonitrile (HPLC grade), formic acid (FA, ≥95.0%), dimethyl sulfoxide (DMSO, HPLC grade, ≥99.7%), piperazine (≥99.0%) and 1,2- ethanedithiol (EDT, GC grade, ≥98.0%) were purchased from Sigma-Aldrich. HRink Amide (0.49 mmol/g and 0.18 mmol/g loading) and HMPB ChemMatrix polyethylene glycol (0.45 mmol/g loading) resin were purchased from PCAS Biomatrix. Water was deionized using a Milli-Q Reference water purification system (Millipore). Nylon 0.22 µm syringe filters were TISCH brand SPEC17984. Coelenterazine substrate, NanoFuel solvent, and recombinant GLuc were purchased from NanoLight technologies. Diphenylenterazine was purchased from MedChemExpress. Guanidine HCl (Gdn HCl, molecular biology grade, ≥99%), tris(hydroxymethyl)aminomethane (TRIS, ≥99.9%), tris(hydroxymethyl)aminomethane hydrochloride (TRIS HCl, ≥99.0%), monopotassium phosphate (KPi, ≥99.0%), dipotassium phosphate (≥ 99.0%), sodium chloride (NaCl, BioXtra, ≥99.5%), glycerol (≥99.5%), sodium hydroxide (≥98%), and hydrochloric acid (36.5-38%) were purchased from Sigma Aldrich. A 0.5 M solution of tris(2-carboxyethyl)phosphine hydrochloride (TCEP, Bond-Breaker™, catalogue number 77720), was purchased from Thermo Fisher Scientific. 1,4-Dithio-DL-threitol (DTT, ≥99%) was purchased from Chem-Impex.

4.8.2 Automated fast-flow peptide synthesizer set-up

All peptides were synthesized on an automated-flow system, which was built in the Pentelute lab (“Amidator”) (19), which is similar to the published AFPS system. Capitalized letters refer to L-amino acids, uncapitalized letters refer to D-amino acids.

Unless otherwise noted, the following settings were used for peptide synthesis: flow-rate = 40 mL/min, temperature = 90 °C (loop) and 85–90 °C (reactor). The 50 mL/min pump head pumps 400 µL of liquid per pump stroke; the 5 mL/min pump head pumps 40 µL of liquid per pump stroke. The pump refill time is 600 ms. The standard synthetic cycle involves a first step of prewashing the resin at elevated temperatures for 60 s at 40 mL/min. During the coupling step, three HPLC pumps are used: a 50 mL/min pump head pumps the activating agent, a second 50 mL/min pump head pumps the amino acid and a 5 mL/min pump head pumps DIEA.

To start the machine a filtered syringe cartridge loaded with resin is mounted on an automated robot arm which moves the syringe into the reactor. Prior to reactor closing DMF is pumped into the syringe for 8 pumping strokes, to swell the resin. standard synthetic cycle involves a first step of prewashing the resin at elevated temperatures for 60 s at 40 mL/min. The first two pumps are activated for 8 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 8 pumping strokes, after which the DIEA pump is shut off and the other two pumps continue to wash the resin for another 40 pump strokes.

During the deprotection step, two HPLC pumps are used. Using a rotary valve, one HPLC pump selects deprotection stock solution and DMF. The pumps are activated for 13 pump strokes. Both solutions are mixed in a 1:1 ratio. Next, the rotary valves select DMF for both HPLC pumps, and the resin is washed for an additional 40 pump strokes. The coupling–deprotection cycle is repeated for all additional monomers. The final synthetic step is a postwash of the resin at elevated temperatures for 60 s at 40 mL/min. A full synthetic cycle is complete in 120 seconds.

4.8.3 Peptide cleavage protocol

Directly after synthesis, the resin was washed with dichloromethane (3 x 5 mL), dried in a vacuum chamber, and weighed. 50% of the resin was transferred into a 6 mL syringe with filter tip. For cleavage of peptides without cystine residues and peptides/protein sequences with cystine residues we used two different protocols.

A) Peptides[84]: Approximately 3 mL of cleavage solution (94% TFA, 1% TIPS, 2.5% EDT, 2.5% water) was added to the tube. If needed, more cleavage solution was added to ensure complete submersion. The capped syringe was kept at room temperature for 2 hours.

B) Proteins and Cys-containing peptides:[85] Approximately 5 mL of cleavage solution (82.5% TFA, 5% water, 5% phenol, 5% thioanisole, 2.5% EDT) was added to the tube. If needed, more cleavage solution was added to ensure complete submersion. The capped syringe was kept on a nutating mixer at room temperature for 4 h.

After incubation cleavage solution was filtered away from resin into a 50 mL conical and the resin was washed 3 times with neat TFA, access TFA was blown off using nitrogen and then ice cold diethyl ether (45 mL) was added to the cleavage mixture and the precipitate was collected

by centrifugation and triturated twice more with cold diethyl ether (45 mL). The supernatant was discarded at each step. Residual ether was allowed to evaporate for 20 minutes after which the peptide was dissolved in 50% acetonitrile in water with 0.1% TFA (long peptides were dissolved 70% acetonitrile in water with 0.1% TFA). The peptide solution was filtrated with a Nylon 0.22 µm syringe filter and frozen, lyophilized until dry, and weighed.

4.8.4 Liquid chromatography-mass spectrometry (LC-MS)

For mass analysis of peptides, a filtered peptide solution (10 µL of a 1mg/mL solution) was diluted in 50% acetonitrile in water with 0.1% TFA (90 µL) to a final concentration approximately 0.1 mg/mL. LC-MS chromatograms and associated high resolution mass spectra were acquired using an Agilent 6550 iFunnel Q-TOF LC-MS system (abbreviated as 6550). Solvent compositions used in the LC-MS are 0.1% formic acid in H2O (solvent A) and 0.1% formic acid in acetonitrile (solvent B).

The following LC-MS methods were used:

H) 1-61% B over 15 min, Phenomenex Jupiter C4 column (6550)
   a. LC conditions: Phenomenex Jupiter C4 column: 1.0 × 150 mm, 5 µm, column temperature: 40 °C, gradient: 0-2 min 1% B, 2-12 min 1-61% B, 12-16 min 61-90% B; flow rate: 0.1 mL/min. A final 4-min hold was performed at a flow rate of 0.1 mL/min. The total method time was 20 min. MS is on from 4 to 12 min.
   b. MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 100–1700 m/z.

I) 1-61% B over 33 min, Phenomenex Jupiter C4 column (6550)
   a. LC conditions: Phenomenex Jupiter C4 column: 1.0 × 150 mm, 5 µm, column temperature: 40 °C, gradient: 0-2 min 1% B, 2-30 min 1-91% B, 30-34 min 61-90% B; flow rate: 0.1 mL/min. A final 4-min hold was performed at a flow rate of 0.1 mL/min. The total method time was 38 min. MS is on from 4 to 30 min.
   b. MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 100–1700 m/z.

J) 1-61% B over 15 min, Zorbax C3 column (6550)
   a. LC conditions: Zorbax 300SB-C3 column: 2.1 × 150 mm, 5 µm, column temperature: 40 °C, gradient: 0-2 min 1% B, 2-11 min 1-61% B, 11-12 min 61-95%

B, flow rate: 0.8 mL/min. A final 3-min hold was performed at a flow rate of 0.8 mL/min. The total method time was 15 min. MS is on from 4 to 12 min.

b. MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 100–1700 m/z.

K) 1-91% B over 20 min, Phenomenex Jupiter C4 column (6550)

a. LC conditions: Phenomenex Jupiter C4 column: 1.0 × 150 mm, 5 μm, column temperature: 40 °C, gradient: 0-2 min 1% B, 2-18 min 1-91% B, 18-21 min 91% B; flow rate: 0.1 mL/min. A final 4-min hold was performed at a flow rate of 0.1 mL/min. The total method time was 25 min. MS is on from 4 to 18 min.

b. MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 100–1700 m/z.

L) 1-61% B over 17 min, Luna C18 column (6550)

a. LC conditions: Phenomenex Luna C18 column: 0.5 × 150 mm, 5 μm, column temperature: 40 °C, gradient: 0-2 min 1% B, 2-14 min 1-61% B, 14-18 min 61-91% B; flow rate: 0.1 mL/min. A final 5-min hold was performed at a flow rate of 0.1 mL/min. The total method time was 23 min. MS is on from 4 to 14 min.

b. MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 100–1700 m/z.

M) 1-61% B over 9 min, Zorbax C3 column (6520)

a. LC conditions: Zorbax 300SB-C3 column: 2.1 × 150 mm, 5 μm, column temperature: 40 °C, gradient: 0-2 min 1% B, 2-11 min 1-61% B, 11-12 min 61-95% B, flow rate: 0.8 mL/min. A final 3-min hold was performed at a flow rate of 0.8 mL/min. The total method time was 15 min. MS is on from 4 to 12 min.

b. MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 300–3000 m/z.

N) 1-91% B over 9 min, Zorbax C3 column (6520)

a. LC conditions: Zorbax 300SB-C3 column: 2.1 × 150 mm, 5 μm, column temperature: 40 °C, gradient: 0-2 min 1% B, 2-11 min 1-91% B, 11-12 min 91-95% B; flow rate: 0.8 mL/min. A final 3-min hold was performed at a flow rate of 0.8 mL/min. The total method time was 15 min. MS is on from 4 to 12 min.

b. MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 300–3000 m/z.

Data were processed using Agilent MassHunter Workstation Qualitative Analysis Version B.06.00 with BioConfirm Software.

4.8.5 Analytical high-performance liquid chromatography (HPLC) protocol

For determination of purity by HPLC, the filtered peptide solution was diluted in 50% acetonitrile in water with 0.1% TFA (100 µL) to a final concentration of approximately 1.0 mg/mL.

For standard analysis of all peptide and protein sequence samples, analytical HPLC spectra were recorded on an analytical Agilent Zorbax 300SB-C3 column (2.1 mm × 150 mm, 5-µm particle size). A linear gradient of acetonitrile with a 0.08% TFA additive (solvent B) in water with a 0.1% TFA additive (solvent A) was used. After a 3-min hold, gradients of 1% B per minute ramped up over 60 min at a flow rate of 0.4 mL/min. Gradients either started at 1% B (annotated as "1–61% B over 60 min") or 5% B (annotated as "5–65% B over 60 min"). A final 3-min hold was performed. The total method time was 66 min. Crude HPLC purities were determined by manual integration of all signals in the area of 5−60 min.

For analysis of epimerization, analytical HPLC spectra were recorded on an analytical Agilent Zorbax 300SB-C18 column (2.1 mm × 150 mm, 5-µm particle size). A linear gradient of acetonitrile with a 0.08% TFA additive (solvent B) in water with a 0.1% TFA additive (solvent A) was used. After a 3-min hold, a gradient of 2% B per minute ramped up over 20 min at a flow rate of 0.4 mL/min (annotated as "1–41% B over 20 min"). A final 3-min hold was performed. The total method time was 30 min.

4.8.6 Mass-directed reversed-phase high performance liquid chromatography (RP-HPLC)

For RP-HPLC purification, the lyophilized peptide sample was dissolved in the gradient starting concentration (e.g., 5% acetonitrile in water with 0.1% TFA) or 6 M Guanidinium chloride containing 100 mM DTT. All samples filtrated with a Nylon 0.22 µm syringe filter prior to purification.

For all HPLC purifications, a gradient of acetonitrile with 0.1% TFA additive (solvent B) and water with a 0.1% TFA additive (solvent A) was used. All samples were purified on a semipreparative Agilent Zorbax 300SB-C3 column (9.4 mm × 250 mm, 5-µm particle size) at a flow rate of 4 mL/min unless otherwise noted.

Specific purification conditions such as column temperature and gradient are specified for each case. In order to optimize the purification for the longer peptide chains the procedure below was followed:

A 60-minute analytical run was performed prior to the purification. 1–2 mg of crude material from 900 µL of solvent was injected on a semipreparative Zorbax 300SB-C3 column at 60 °C (gradient: 5–65% B with 1% B/min). The method was then adjusted according to the elution profile.

The gradient was changed to 0.2% B around the B concentration at which the product eluted (e.g., if the compound eluted at 39% B we adjusted the method to: < 50 mg of crude on a semipreparative Zorbax 300SB-C3 column at 60 °C (gradient: 5-29% B with 1% B/min, 29-49% B with 0.2% B/min, 49-65% B with 1% B/min). These conditions were used unless otherwise noted.

4.8.7 Protein concentration determination protocol

The concentration of proteins in solution was determined by absorbance at 280 nm (A280). The detailed protocol is as followed:

- A280: The concentration of proteins was calculated using the Beer-Lambert law by measuring the absorbance of the protein sample at 280 nm. The absorbance was measured by averaging at least two independent readings of the same sample on a BioTek Synergy HT plate reader outfitted with a BioTek Take 3 micro-volume plate. Unless indicated otherwise, the molar extinction coefficient of the protein of interest was estimated based on the sequence of the protein via ExPASy Swiss Institute of Bioinformatics - Bioinformatics Resource Portal.

4.8.8 Determination of peptide and protein yield

Molecular weight of peptide sequences was determined via ChemDraw, accounting for the weight of a TFA counter-ion for each basic residue (K, R, H) in addition to the N-terminal amine.

For example, for a peptide with sequence “KALE” the molecular weight of the peptide as TFA salt is calculated as 916 g/mol (= 688 + 2 × 114).

The weight of lyophilized powders of the peptides was directly measured using analytical scales (XS205DU Analytical Balance, Mettler-Toledo) [note: use of deionizers such as SPI Westek Workstation Still Air Ionizer helps with measurements]. Following folding, protein concentration was measured based on the outlined procedures above.

Theoretical yield was determined based on weight of the resin, resin loading, and the molecular weight (with TFA) of each peptide.

For example, for the KALE sequence synthesized on 50 mg resin with 0.44 mmol/g loading, theoretical yield is:

theoretical yield = 0.44 mmol/g × 50 mg × 916 g mol = 20 mg

Yield of crude peptide was determined based on the ratio of weight of lyophilized crude peptide (as TFA salt) to theoretical yield multiplied by the purity determined by UV absorption at 280 nm (analytical HPLC).

In the example above, if 10 mg of crude KALE peptide is produced and the purity by analytical HPLC is 50%,

synthesis yield is: yield = 10 mg/20 mg × 0.50 × 100 = 25%

Yield past HPLC was calculated based on weight ratio of HPLC-purified peptides to theoretical yield (both as TFA salts).

In the example above, if 4.0 mg of KALE peptide is isolated from HPLC, the yield of the synthesis past HPLC is:

yield = 4 mg/20 mg × 100 = 20%

Folding and chromatography (size exclusion, ion exchange, etc.) yield was calculated based on concentration of protein in solution and the mass of lyophilized purified peptide (as TFA salt).

In the example above, if 1 mL of 1 mg/mL solution of KALE peptide is obtained from folding & size exclusion chromatography of 4 mg of HPLC-purified KALE, the purification yield is:

purification yield = (1 mg⁄mL × 1 mL)/4 mg × 100 = 25%

4.8.9 Extended folding and SEC purification of luciferases

The folding procedure of all luciferases was adapted from Yu and coworkers and 0.5 mg of HPLC-purified or crude luciferase was dissolved in 250 mM of a 6 M Gn·HCl solution in 55 mM NaH_2PO4 buffer pH 7. The mixture was then diluted 10-fold to 0.3 M Gn·HCl using a 50 mM KH_2PO_4, 50 mM NaCl, 1mM EDTA, 3 mM GSH, 2 mM GSSG and subjected to size exchange chromatography (GE Healthcare Superdex 75 increase 10/300 GL column), using a isocratic elution buffer of 200 mM NaCl, 1 mM EDTA, 20 mM Tris-HCl pH 7.5. The elution profile is depicted below.

Fractions containing the protein were isolated and concentrated using a 3K molecular weight cut off spin filter, flash frozen using liquid nitrogen, and stored at −80 °C. A total of 0.1 mg isolated luciferase was recovered corresponding to 20% isolated yield. The purity of the final product was assessed by LC-MS and HPLC.

4.8.10 Bioluminescent assay protocol

Bioluminescent assays were performed on one of two plate readers in the laboratory. The protocols are as follows:

1. 1450 microbeta liquid scintillation and luminescence counter: A 96-well black flat bottomed plate was filled with 180 uL of diluted luciferase in folding buffer (100 nM or 500nM) and then 50 uL of 3 uM CTZ in folding buffer was hand pipetted immediately prior to the start of data collection. Data was collected with a 10 second integration of the bioluminescence signal for at least 2 minutes.
2. Tecan Spark plate reader: A 96-well black flat bottomed plate was filled with 150 uL of diluted luciferase in folding buffer (100 nM or 500nM) and then 50 uL of 3 uM CTZ in folding buffer was autoinjected by the machine immediately prior to the start of data collection for each well and was mixed for 3 seconds via double orbital shaking. Data was then collected with a 10 second integration of the bioluminescence signal for at least 2 minutes via the looped kinetic analysis methods on the Tecan.

All data was then baseline corrected and processed using GraphPad Prism 10 software.

4.9 Acknowledgements

We thank Dr. Joesph Brown for the many helpful conversations in the early days of outlining this project. A massive thanks to David Sarabia-Castillo for coming in to offer additional scientific support as the project expanded and taking over the project as I leave MIT. Thank you to Gha Young-Lee for her help with luciferase modeling. Finally, a big thank you to Bradley Pentelute for allowing us to pursue this research.

Funding: A.E.C. gratefully acknowledges support from the National Science Foundation Graduate Research Fellowship under Grant No. 1122374; A.E.C. and M.A.L. are additionally supported by an MIT Dean of Science Fellowship.

Author contributions: A.E.C. and B.L.P. conceptualized the research; A.E.C and D.S. synthesized, purified and analyzed protein samples; A.E.C. J.S.B, and M.A.L designed and synthesized peptide libraries and protein cleavage sights with aid from D. S. and G. Y. L. A.E.C.

and B.L.P. conceptualized folding and biological evaluation of the synthetic proteins; A.E.C and B.L.P wrote the manuscript with input of all co-authors

4.10 References

1. Brodl, E.; Winkler, A.; Macheroux, P. Molecular Mechanisms of Bacterial Bioluminescence. *Comput. Struct. Biotechnol. J.* **2018**, *16*, 551–564. https://doi.org/10.1016/j.csbj.2018.11.003.

2. Delroisse, J.; Duchatelet, L.; Flammang, P.; Mallefet, J. Leaving the Dark Side? Insights Into the Evolution of Luciferases. *Front. Mar. Sci.* **2021**, *8*, 673620. https://doi.org/10.3389/fmars.2021.673620.

3. Orlova, G.; Goddard, J. D.; Brovko, L. Yu. Theoretical Study of the Amazing Firefly Bioluminescence: The Formation and Structures of the Light Emitters. *J. Am. Chem. Soc.* **2003**, *125* (23), 6962–6971. https://doi.org/10.1021/ja021255a.

4. Mütze, J.; Iyer, V.; Macklin, J. J.; Colonell, J.; Karsh, B.; Petrášek, Z.; Schwille, P.; Looger, L. L.; Lavis, L. D.; Harris, T. D. Excitation Spectra and Brightness Optimization of Two-Photon Excited Probes. *Biophys. J.* **2012**, *102* (4), 934–944. https://doi.org/10.1016/j.bpj.2011.12.056.

5. Fleiss, A.; Sarkisyan, K. S. A Brief Review of Bioluminescent Systems (2019). *Curr. Genet.* **2019**, *65* (4), 877–882. https://doi.org/10.1007/s00294-019-00951-5.

6. Khakhar, A.; Starker, C. G.; Chamness, J. C.; Lee, N.; Stokke, S.; Wang, C.; Swanson, R.; Rizvi, F.; Imaizumi, T.; Voytas, D. F. Building Customizable Auto-Luminescent Luciferase-Based Reporters in Plants. *eLife* **2020**, *9*, e52786. https://doi.org/10.7554/eLife.52786.

7. Shimomura, O. Bioluminescence in the Sea: Photoprotein Systems. *Symp. Soc. Exp. Biol.* **1985**, *39*, 351–372.

8. DeLeo, D. M.; Bracken-Grissom, H. D. Lighting the Way: Forces Driving the Diversification of Bioluminescent Signalling in Sea Fireflies. *Mol. Ecol.* **2021**, *30* (8), 1747–1750. https://doi.org/10.1111/mec.15880.

9. Takakura, H. Molecular Design of D-Luciferin-Based Bioluminescence and 1,2-Dioxetane-Based Chemiluminescence Substrates for Altered Output Wavelength and Detecting Various Molecules. *Mol. Basel Switz.* **2021**, *26* (6), 1618. https://doi.org/10.3390/molecules26061618.

10. Feeney, K. A.; Putker, M.; Brancaccio, M.; O'Neill, J. S. In-Depth Characterization of Firefly Luciferase as a Reporter of Circadian Gene Expression in Mammalian Cells. *J. Biol. Rhythms* **2016**, *31* (6), 540–550. https://doi.org/10.1177/0748730416668898.

11. Zambito, G.; Hall, M. P.; Wood, M. G.; Gaspar, N.; Ridwan, Y.; Stellari, F. F.; Shi, C.; Kirkland, T. A.; Encell, L. P.; Löwik, C.; Mezzanotte, L. Red-Shifted Click Beetle

Luciferase Mutant Expands the Multicolor Bioluminescent Palette for Deep Tissue Imaging. *iScience* **2021**, *24* (1), 101986. https://doi.org/10.1016/j.isci.2020.101986.

12. Watkins, O. C.; Sharpe, M. L.; Perry, N. B.; Krause, K. L. New Zealand Glowworm (Arachnocampa Luminosa) Bioluminescence Is Produced by a Firefly-like Luciferase but an Entirely New Luciferin. *Sci. Rep.* **2018**, *8* (1), 3278. https://doi.org/10.1038/s41598-018-21298-w.

13. Srikantha, T.; Klapach, A.; Lorenz, W. W.; Tsai, L. K.; Laughlin, L. A.; Gorman, J. A.; Soll, D. R. The Sea Pansy Renilla Reniformis Luciferase Serves as a Sensitive Bioluminescent Reporter for Differential Gene Expression in Candida Albicans. *J. Bacteriol.* **1996**, *178* (1), 121–129. https://doi.org/10.1128/jb.178.1.121-129.1996.

14. Tessler, M.; Gaffney, J. P.; Crawford, J. M.; Trautman, E.; Gujarati, N. A.; Alatalo, P.; Pieribone, V. A.; Gruber, D. F. Luciferin Production and Luciferase Transcription in the Bioluminescent Copepod Metridia Lucens. *PeerJ* **2018**, *6*, e5506. https://doi.org/10.7717/peerj.5506.

15. Mitani, Y.; Yasuno, R.; Futahashi, R.; Oakley, T. H.; Ohmiya, Y. Luciferase Gene of a Caribbean Fireworm (Syllidae) from Puerto Rico. *Sci. Rep.* **2019**, *9* (1), 13015. https://doi.org/10.1038/s41598-019-49538-7.

16. Vysotski, E. S. Bioluminescent and Fluorescent Proteins: Molecular Mechanisms and Modern Applications. *Int. J. Mol. Sci.* **2022**, *24* (1), 281. https://doi.org/10.3390/ijms24010281.

17. Kubodera, T.; Koyama, Y.; Mori, K. Observations of Wild Hunting Behaviour and Bioluminescence of a Large Deep-Sea, Eight-Armed Squid, *Taningia Danae*. *Proc. R. Soc. B Biol. Sci.* **2007**, *274* (1613), 1029–1034. https://doi.org/10.1098/rspb.2006.0236.

18. Fallon, T. R.; Li, F.-S.; Vicent, M. A.; Weng, J.-K. Sulfoluciferin Is Biosynthesized by a Specialized Luciferin Sulfotransferase in Fireflies. *Biochemistry* **2016**, *55* (24), 3341–3344. https://doi.org/10.1021/acs.biochem.6b00402.

19. Wainwright, P. C.; Longo, S. J. Functional Innovations and the Conquest of the Oceans by Acanthomorph Fishes. *Curr. Biol.* **2017**, *27* (11), R550–R557. https://doi.org/10.1016/j.cub.2017.03.044.

20. Miyashiro, T.; Ruby, E. G. Shedding Light on Bioluminescence Regulation in Vibrio Fischeri. *Mol. Microbiol.* **2012**, *84* (5), 795–806. https://doi.org/10.1111/j.1365-2958.2012.08065.x.

21. Möller, A.; Jansson, J. K. Detection of Firefly Luciferase-Tagged Bacteria in Environmental Samples. In *Bioluminescence Methods and Protocols*; Humana Press: New Jersey, 1998; Vol. 102, pp 269–284. https://doi.org/10.1385/0-89603-520-4:269.

22. Donat, S.; Hasenberg, M.; Schäfer, T.; Ohlsen, K.; Gunzer, M.; Einsele, H.; Löffler, J.; Beilhack, A.; Krappmann, S. Surface Display of *Gaussia Princeps* Luciferase Allows

Sensitive Fungal Pathogen Detection during Cutaneous Aspergillosis. *Virulence* **2012**, *3* (1), 51–61. https://doi.org/10.4161/viru.3.1.18799.

23. Siebring-van Olst, E.; Van Beusechem, V. W. High-Throughput Firefly Luciferase Reporter Assays. In *Reporter Gene Assays*; Damoiseaux, R., Hasson, S., Eds.; Methods in Molecular Biology; Springer New York: New York, NY, 2018; Vol. 1755, pp 19–29. https://doi.org/10.1007/978-1-4939-7724-6_2.

24. Lang, Y.; Li, Z.; Li, H. Analysis of Protein-Protein Interactions by Split Luciferase Complementation Assay. *Curr. Protoc. Toxicol.* **2019**, *82* (1), e90. https://doi.org/10.1002/cptx.90.

25. Thorne, N.; Inglese, J.; Auld, D. S. Illuminating Insights into Firefly Luciferase and Other Bioluminescent Reporters Used in Chemical Biology. *Chem. Biol.* **2010**, *17* (6), 646–657. https://doi.org/10.1016/j.chembiol.2010.05.012.

26. Choy, G.; O'Connor, S.; Diehn, F. E.; Costouros, N.; Alexander, H. R.; Choyke, P.; Libutti, S. K. Comparison of Noninvasive Fluorescent and Bioluminescent Small Animal Optical Imaging. *BioTechniques* **2003**, *35* (5), 1022–1030. https://doi.org/10.2144/03355rr02.

27. Yao, Z.; Caldwell, D. R.; Love, A. C.; Kolbaba-Kartchner, B.; Mills, J. H.; Schnermann, M. J.; Prescher, J. A. Coumarin Luciferins and Mutant Luciferases for Robust Multi-Component Bioluminescence Imaging. *Chem. Sci.* **2021**, *12* (35), 11684–11691. https://doi.org/10.1039/d1sc03114g.

28. Neefjes, M.; Housmans, B. A. C.; Van Den Akker, G. G. H.; Van Rhijn, L. W.; Welting, T. J. M.; Van Der Kraan, P. M. Reporter Gene Comparison Demonstrates Interference of Complex Body Fluids with Secreted Luciferase Activity. *Sci. Rep.* **2021**, *11* (1), 1359. https://doi.org/10.1038/s41598-020-80451-6.

29. Baldwin, T. O. Firefly Luciferase: The Structure Is Known, but the Mystery Remains. *Structure* **1996**, *4* (3), 223–228. https://doi.org/10.1016/S0969-2126(96)00026-3.

30. Adams, S. T.; Miller, S. C. Enzymatic Promiscuity and the Evolution of Bioluminescence. *FEBS J.* **2020**, *287* (7), 1369–1380. https://doi.org/10.1111/febs.15176.

31. Shifera, A. S.; Hardin, J. A. Factors Modulating Expression of Renilla Luciferase from Control Plasmids Used in Luciferase Reporter Gene Assays. *Anal. Biochem.* **2010**, *396* (2), 167–172. https://doi.org/10.1016/j.ab.2009.09.043.

32. Nishihara, R.; Abe, M.; Nishiyama, S.; Citterio, D.; Suzuki, K.; Kim, S. B. Luciferase-Specific Coelenterazine Analogues for Optical Contamination-Free Bioassays. *Sci. Rep.* **2017**, *7* (1), 908. https://doi.org/10.1038/s41598-017-00955-6.

33. Guo, B.; Shi, X.; Ma, Z.; Ji, M.; Tang, C.; Wang, F. A Ratiometric Dual Luciferase Reporter for Quantitative Monitoring of Pre-mRNA Splicing Efficiency in Vivo. *J. Biol. Chem.* **2021**, *297* (2), 100933. https://doi.org/10.1016/j.jbc.2021.100933.

34. Kobayashi, H.; Picard, L.-P.; Schönegge, A.-M.; Bouvier, M. Bioluminescence Resonance Energy Transfer–Based Imaging of Protein–Protein Interactions in Living Cells. *Nat. Protoc.* **2019**, *14* (4), 1084–1107. https://doi.org/10.1038/s41596-019-0129-7.

35. Sadikot, R. T.; Blackwell, T. S. Bioluminescence Imaging. *Proc. Am. Thorac. Soc.* **2005**, *2* (6), 537–540, 511–512. https://doi.org/10.1513/pats.200507-067DS.

36. Inouye, S.; Watanabe, K.; Nakamura, H.; Shimomura, O. Secretional Luciferase of the Luminous Shrimp *Oplophorus Gracilirostris* : cDNA Cloning of a Novel Imidazopyrazinone Luciferase. *FEBS Lett.* **2000**, *481* (1), 19–25. https://doi.org/10.1016/S0014-5793(00)01963-3.

37. Hall, M. P.; Unch, J.; Binkowski, B. F.; Valley, M. P.; Butler, B. L.; Wood, M. G.; Otto, P.; Zimmerman, K.; Vidugiris, G.; Machleidt, T.; Robers, M. B.; Benink, H. A.; Eggers, C. T.; Slater, M. R.; Meisenheimer, P. L.; Klaubert, D. H.; Fan, F.; Encell, L. P.; Wood, K. V. Engineered Luciferase Reporter from a Deep Sea Shrimp Utilizing a Novel Imidazopyrazinone Substrate. *ACS Chem. Biol.* **2012**, *7* (11), 1848–1857. https://doi.org/10.1021/cb3002478.

38. Takenaka, Y.; Masuda, H.; Yamaguchi, A.; Nishikawa, S.; Shigeri, Y.; Yoshida, Y.; Mizuno, H. Two Forms of Secreted and Thermostable Luciferases from the Marine Copepod Crustacean, Metridia Pacifica. *Gene* **2008**, *425* (1–2), 28–35. https://doi.org/10.1016/j.gene.2008.07.041.

39. Auld, D. S.; Narahari, J.; Ho, P.; Casalena, D.; Nguyen, V.; Cirbaite, E.; Hughes, D.; Daly, J.; Webb, B. Characterization and Use of TurboLuc Luciferase as a Reporter for High-Throughput Assays. *Biochemistry* **2018**, *57* (31), 4700–4706. https://doi.org/10.1021/acs.biochem.8b00290.

40. Tannous, B. A. Gaussia Luciferase Reporter Assay for Monitoring Biological Processes in Culture and in Vivo. *Nat. Protoc.* **2009**, *4* (4), 582–591. https://doi.org/10.1038/nprot.2009.28.

41. England, C. G.; Ehlerding, E. B.; Cai, W. NanoLuc: A Small Luciferase Is Brightening Up the Field of Bioluminescence. *Bioconjug. Chem.* **2016**, *27* (5), 1175–1187. https://doi.org/10.1021/acs.bioconjchem.6b00112.

42. Markova, S. V.; Larionova, M. D.; Vysotski, E. S. Shining Light on the Secreted Luciferases of Marine Copepods: Current Knowledge and Applications. *Photochem. Photobiol.* **2019**, *95* (3), 705–721. https://doi.org/10.1111/php.13077.

43. De Niz, M.; Stanway, R. R.; Wacker, R.; Keller, D.; Heussler, V. T. An Ultrasensitive NanoLuc-Based Luminescence System for Monitoring Plasmodium Berghei throughout Its Life Cycle. *Malar. J.* **2016**, *15* (1), 232. https://doi.org/10.1186/s12936-016-1291-9.

44. Kotlobay, A. A.; Kaskova, Z. M.; Yampolsky, I. V. Palette of Luciferases: Natural Biotools for New Applications in Biomedicine. *Acta Naturae* **2020**, *12* (2), 15–27. https://doi.org/10.32607/actanaturae.10967.

45. Hughes, K.; Maloy, S. R. *Brenner's Encyclopedia of Genetics*, 2. ed.; Academic Press, Elsevier Science: London [u.a.], 2013.

46. Hansen, J.; Floss, T.; Van Sloun, P.; Füchtbauer, E.-M.; Vauti, F.; Arnold, H.-H.; Schnütgen, F.; Wurst, W.; von Melchner, H.; Ruiz, P. A Large-Scale, Gene-Driven Mutagenesis Approach for the Functional Analysis of the Mouse Genome. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (17), 9918–9922. https://doi.org/10.1073/pnas.1633296100.

47. Kim, S. B. Labor-Effective Manipulation of Marine and Beetle Luciferases for Bioassays. *Protein Eng. Des. Sel.* **2012**, *25* (6), 261–269. https://doi.org/10.1093/protein/gzs016.

48. Kim, S. B.; Nishihara, R.; Citterio, D.; Suzuki, K. Fabrication of a New Lineage of Artificial Luciferases from Natural Luciferase Pools. *ACS Comb. Sci.* **2017**, *19* (9), 594–599. https://doi.org/10.1021/acscombsci.7b00081.

49. Kim, S. B.; Torimura, M.; Tao, H. Creation of Artificial Luciferases for Bioassays. *Bioconjug. Chem.* **2013**, *24* (12), 2067–2075. https://doi.org/10.1021/bc400411h.

50. Ohmuro-Matsuyama, Y.; Furuta, T.; Matsui, H.; Kanai, M.; Ueda, H. Miniaturization of Bright Light-Emitting Luciferase ALuc: picALuc. *ACS Chem. Biol.* **2022**, *17* (4), 864–872. https://doi.org/10.1021/acschembio.1c00897.

51. Takenaka, Y.; Yamaguchi, A.; Tsuruoka, N.; Torimura, M.; Gojobori, T.; Shigeri, Y. Evolution of Bioluminescence in Marine Planktonic Copepods. *Mol. Biol. Evol.* **2012**, *29* (6), 1669–1681. https://doi.org/10.1093/molbev/mss009.

52. Yu, T.; Laird, J. R.; Prescher, J. A.; Thorpe, C. Gaussia Princeps Luciferase: A Bioluminescent Substrate for Oxidative Protein Folding. *Protein Sci. Publ. Protein Soc.* **2018**, *27* (8), 1509–1517. https://doi.org/10.1002/pro.3433.

53. Mueller, L. K.; Baumruck, A. C.; Zhdanova, H.; Tietze, A. A. Challenges and Perspectives in Chemical Synthesis of Highly Hydrophobic Peptides. *Front. Bioeng. Biotechnol.* **2020**, *8*, 162. https://doi.org/10.3389/fbioe.2020.00162.

54. Lu, D.; Liu, Z. Dynamic Redox Environment-Intensified Disulfide Bond Shuffling for Protein Refolding in Vitro: Molecular Simulation and Experimental Validation. *J. Phys. Chem. B* **2008**, *112* (47), 15127–15133. https://doi.org/10.1021/jp804649g.

55. Boute, N.; Lowe, P.; Berger, S.; Malissard, M.; Robert, A.; Tesar, M. NanoLuc Luciferase – A Multifunctional Tool for High Throughput Antibody Screening. *Front. Pharmacol.* **2016**, *7*. https://doi.org/10.3389/fphar.2016.00027.

56. Dijkema, F. M.; Nordentoft, M. K.; Didriksen, A. K.; Corneliussen, A. S.; Willemoës, M.; Winther, J. R. Flash Properties of GAUSSIA LUCIFERASE Are the Result of Covalent Inhibition after a Limited Number of Cycles. *Protein Sci.* **2021**, *30* (3), 638–649. https://doi.org/10.1002/pro.4023.

57. Hunt, E. A.; Moutsiopoulou, A.; Ioannou, S.; Ahern, K.; Woodward, K.; Dikici, E.; Daunert, S.; Deo, S. K. Truncated Variants of Gaussia Luciferase with Tyrosine Linker for Site-Specific Bioconjugate Applications. *Sci. Rep.* **2016**, *6* (1), 26814. https://doi.org/10.1038/srep26814.

58. Yeh, A. H.-W.; Norn, C.; Kipnis, Y.; Tischer, D.; Pellock, S. J.; Evans, D.; Ma, P.; Lee, G. R.; Zhang, J. Z.; Anishchenko, I.; Coventry, B.; Cao, L.; Dauparas, J.; Halabiya, S.; DeWitt, M.; Carter, L.; Houk, K. N.; Baker, D. De Novo Design of Luciferases Using Deep Learning. *Nature* **2023**, *614* (7949), 774–780. https://doi.org/10.1038/s41586-023-05696-3.

59. Syed, A. J.; Anderson, J. C. Applications of Bioluminescence in Biotechnology and Beyond. *Chem. Soc. Rev.* **2021**, *50* (9), 5668–5705. https://doi.org/10.1039/D0CS01492C.

60. Jiang, T.; Du, L.; Li, M. Lighting up Bioluminescence with Coelenterazine: Strategies and Applications. *Photochem. Photobiol. Sci.* **2016**, *15* (4), 466–480. https://doi.org/10.1039/c5pp00456j.

61. Yeh, H.-W.; Karmach, O.; Ji, A.; Carter, D.; Martins-Green, M. M.; Ai, H. Red-Shifted Luciferase–Luciferin Pairs for Enhanced Bioluminescence Imaging. *Nat. Methods* **2017**, *14* (10), 971–974. https://doi.org/10.1038/nmeth.4400.

62. Anishchenko, I.; Pellock, S. J.; Chidyausiku, T. M.; Ramelot, T. A.; Ovchinnikov, S.; Hao, J.; Bafna, K.; Norn, C.; Kang, A.; Bera, A. K.; DiMaio, F.; Carter, L.; Chow, C. M.; Montelione, G. T.; Baker, D. De Novo Protein Design by Deep Network Hallucination. *Nature* **2021**, *600* (7889), 547–552. https://doi.org/10.1038/s41586-021-04184-w.

63. Aliashkevich, A.; Alvarez, L.; Cava, F. New Insights Into the Mechanisms and Biological Roles of D-Amino Acids in Complex Eco-Systems. *Front. Microbiol.* **2018**, *9*, 683. https://doi.org/10.3389/fmicb.2018.00683.

64. Milton, R. C. deL.; Milton, S. C. F.; Kent, S. B. H. Total Chemical Synthesis of a D-Enzyme: The Enantiomers of HIV-1 Protease Show Reciprocal Chiral Substrate Specificity. *Science* **1992**, *256* (5062), 1445–1448. https://doi.org/10.1126/science.1604320.

65. Feng, Z.; Xu, B. Inspiration from the Mirror: D-Amino Acid Containing Peptides in Biomedical Approaches. *Biomol. Concepts* **2016**, *7* (3), 179–187. https://doi.org/10.1515/bmc-2015-0035.

66. Zhou, X.; Zuo, C.; Li, W.; Shi, W.; Zhou, X.; Wang, H.; Chen, S.; Du, J.; Chen, G.; Zhai, W.; Zhao, W.; Wu, Y.; Qi, Y.; Liu, L.; Gao, Y. A Novel D -Peptide Identified by Mirror-Image Phage Display Blocks TIGIT/PVR for Cancer Immunotherapy. *Angew. Chem. Int. Ed.* **2020**, *59* (35), 15114–15118. https://doi.org/10.1002/anie.202002783.

67. Weidmann, J.; Schnölzer, M.; Dawson, P. E.; Hoheisel, J. D. Copying Life: Synthesis of an Enzymatically Active Mirror-Image DNA-Ligase Made of D-Amino Acids. *Cell Chem. Biol.* **2019**, *26* (5), 645-651.e3. https://doi.org/10.1016/j.chembiol.2019.02.008.

68. Callahan, A. J.; Gandhesiri, S.; Travaline, T. L.; Lozano Salazar, L.; Hanna, S.; Lee, Y.-C.; Li, K.; Tokareva, O. S.; Swiecicki, J.-M.; Loas, A.; Verdine, G. L.; McGee, J. H.; Pentelute, B. L. *Single-Shot Flow Synthesis of D-Proteins for Mirror-Image Phage Display*; preprint; Chemistry, 2023. https://doi.org/10.26434/chemrxiv-2023-x86xp.

69. Braun, P. Interactome Mapping for Analysis of Complex Phenotypes: Insights from Benchmarking Binary Interaction Assays. *PROTEOMICS* **2012**, *12* (10), 1499–1518. https://doi.org/10.1002/pmic.201100598.

70. Rao, V. S.; Srinivas, K.; Sujini, G. N.; Kumar, G. N. S. Protein-Protein Interaction Detection: Methods and Analysis. *Int. J. Proteomics* **2014**, *2014*, 147648. https://doi.org/10.1155/2014/147648.

71. Michnick, S. W.; Ear, P. H.; Manderson, E. N.; Remy, I.; Stefan, E. Universal Strategies in Research and Drug Discovery Based on Protein-Fragment Complementation Assays. *Nat. Rev. Drug Discov.* **2007**, *6* (7), 569–582. https://doi.org/10.1038/nrd2311.

72. Sierecki, E. A Novel Split Reporter Uncages New Possibilities. *ACS Cent. Sci.* **2019**, *5* (11), 1744–1746. https://doi.org/10.1021/acscentsci.9b01051.

73. Magliery, T. J.; Wilson, C. G. M.; Pan, W.; Mishler, D.; Ghosh, I.; Hamilton, A. D.; Regan, L. Detecting Protein−Protein Interactions with a Green Fluorescent Protein Fragment Reassembly Trap: Scope and Mechanism. *J. Am. Chem. Soc.* **2005**, *127* (1), 146–157. https://doi.org/10.1021/ja046699g.

74. Rossi, F.; Charlton, C. A.; Blau, H. M. Monitoring Protein–Protein Interactions in Intact Eukaryotic Cells by β-Galactosidase Complementation. *Proc. Natl. Acad. Sci.* **1997**, *94* (16), 8405–8410. https://doi.org/10.1073/pnas.94.16.8405.

75. Romei, M. G.; Boxer, S. G. Split Green Fluorescent Proteins: Scope, Limitations, and Outlook. *Annu. Rev. Biophys.* **2019**, *48*, 19–44. https://doi.org/10.1146/annurev-biophys-051013-022846.

76. Moustaqil, M.; Bhumkar, A.; Gonzalez, L.; Raoul, L.; Hunter, D.; Carrive, P.; Sierecki, E.; Gambin, Y. A Split-Luciferase Reporter Recognizing GFP and mCherry Tags to Facilitate Studies of Protein–Protein Interactions. *Int. J. Mol. Sci.* **2017**, *18* (12), 2681. https://doi.org/10.3390/ijms18122681.

77. Verhoef, L. G. G. C.; Mattioli, M.; Ricci, F.; Li, Y.-C.; Wade, M. Multiplex Detection of Protein–Protein Interactions Using a next Generation Luciferase Reporter. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **2016**, *1863* (2), 284–292. https://doi.org/10.1016/j.bbamcr.2015.11.031.

78. Chen, H.; Zou, Y.; Shang, Y.; Lin, H.; Wang, Y.; Cai, R.; Tang, X.; Zhou, J.-M. Firefly Luciferase Complementation Imaging Assay for Protein-Protein Interactions in Plants. *Plant Physiol.* **2008**, *146* (2), 368–376. https://doi.org/10.1104/pp.107.111740.

79. Paulmurugan, R.; Gambhir, S. S. Combinatorial Library Screening for Developing an Improved Split-Firefly Luciferase Fragment-Assisted Complementation System for Studying Protein−Protein Interactions. *Anal. Chem.* **2007**, *79* (6), 2346–2353. https://doi.org/10.1021/ac062053q.

80. Krasitskaya, V. V.; Bashmakova, E. E.; Frank, L. A. Coelenterazine-Dependent Luciferases as a Powerful Analytical Tool for Research and Biomedical Applications. *Int. J. Mol. Sci.* **2020**, *21* (20), 7465. https://doi.org/10.3390/ijms21207465.

81. Dixon, A. S.; Schwinn, M. K.; Hall, M. P.; Zimmerman, K.; Otto, P.; Lubben, T. H.; Butler, B. L.; Binkowski, B. F.; Machleidt, T.; Kirkland, T. A.; Wood, M. G.; Eggers, C. T.; Encell, L. P.; Wood, K. V. NanoLuc Complementation Reporter Optimized for Accurate Measurement of Protein Interactions in Cells. *ACS Chem. Biol.* **2016**, *11* (2), 400–408. https://doi.org/10.1021/acschembio.5b00753.

82. Lam, K. S.; Lebl, M.; Krchňák, V. The "One-Bead-One-Compound" Combinatorial Library Method. *Chem. Rev.* **1997**, *97* (2), 411–448. https://doi.org/10.1021/cr9600114.

83. Quartararo, A. J.; Gates, Z. P.; Somsen, B. A.; Hartrampf, N.; Ye, X.; Shimada, A.; Kajihara, Y.; Ottmann, C.; Pentelute, B. L. Ultra-Large Chemical Libraries for the Discovery of High-Affinity Peptide Binders. *Nat. Commun.* **2020**, *11* (1), 3183. https://doi.org/10.1038/s41467-020-16920-3.

84. Ye, X.; Lee, Y.-C.; Gates, Z. P.; Ling, Y.; Mortensen, J. C.; Yang, F.-S.; Lin, Y.-S.; Pentelute, B. L. Binary Combinatorial Scanning Reveals Potent Poly-Alanine-Substituted Inhibitors of Protein-Protein Interactions. *Commun. Chem.* **2022**, *5* (1), 128. https://doi.org/10.1038/s42004-022-00737-w.

85. Thapa, P.; Cabalteja, C. C.; Philips, E. E.; Espiritu, M. J.; Peigneur, S.; Mille, B. G.; Tytgat, J.; Cummins, T. R.; Bingham, J.-P. T-Boc Synthesis of Huwentoxin-i through Native Chemical Ligation Incorporating a Trifluoromethanesulfonic Acid Cleavage Strategy: T-Boc Synthesis of Huwentoxin-I. *Biopolymers* **2016**, *106* (5), 737–745. https://doi.org/10.1002/bip.22887.