Improving Cardiac Delivery of Antisense Oligonucleotides with Peptidomimetic Targeting Agents

by

Sarah An-ning Antilla

B.S., California Institute of Technology (2018)

Submitted to the Department of Materials Science and Engineering
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN MATERIALS SCIENCE AND ENGINEERING

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

SEPTEMBER 2023

Signature of Author……
Department of Materials Science and Engineering
August 10, 2023

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

Accepted by ……....
Robert J. Macfarlane
Chair
Departmental Committee on Graduate Studies

This doctoral thesis has been examined by a committee of the
Department of Materials Science and Engineering as follows:

Julia Ortony……..
Thesis Reader
Associate Professor of Chemistry and Biochemistry

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

Rafael Gomez-Bombarelli…….....
Thesis Committee Member
Associate Professor of Materials Science and Engineering

Improving Cardiac Delivery of Antisense Oligonucleotides with Peptidomimetic Targeting Agents

by

Sarah Antilla

Submitted to the Department of Materials Science and Engineering on
August 10, 2023
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Materials Science and Engineering

ABSTRACT

Cardiovascular disease (CVD) is one of the leading causes of death worldwide, both individually and as a comorbidity for other diseases such as diabetes and atherosclerosis. Unfortunately, CVD trends irreversibly toward heart failure. Current treatments only manage symptoms such as high blood pressure rather than addressing the root biological causes of the disease. Many micro-RNAs (miRNAs) are either over or under-expressed in CVD, making the regulation of these miRNAs a potential treatment strategy. Here, we investigate the delivery of antisense oligonucleotides (ASOs) to inhibit the expression of an overexpressed miRNA in CVD, miRNA-21.

One of the challenges in delivering ASOs and other gene therapies is achieving delivery to the desired tissue before the therapeutic is trafficked to the liver or kidneys. Our lab has a platform for discovering peptide-protein interactions, affinity selection-mass spectrometry (AS-MS), with which we can find short peptide or peptidomimetic targeting agents with nM binding affinity to target proteins. Here, I describe a platform for selecting and procuring cardiac-specific proteins or their extracellular domains (ectodomains), in some cases employing the automated fast-flow peptide synthesis (AFPS) our lab has developed, which can produce single domain proteins in hours in a single shot. We aim to discover and validate binders to these targets using AS-MS.

Because these targets were challenging to generate binders to, we began investigating the transferrin receptor (TfR1) and a peptide found in literature to bind to TfR1, T12. T12 binds to TfR1 with low tens of nM binding affinity, and a conjugate of anti-miRNA peptide nucleic acid (PNA) with T12 inhibits about 50% of miRNA-21 expression in mouse cardiac tissue at 30 mg/kg, while 30 mg/kg PNA alone does not show significant inhibition of miRNA-21 expression in the heart. To reduce the dose required for efficacy, we synthesized a linear dimer of T12, which exhibits tenfold stronger binding to TfR1. A PNA-T12 dimer conjugate exhibits just over 50% inhibition of miRNA-21 expression in cardiac tissue at only 5 mg/kg, out-performing the PNA-T12 monomer conjugate. We begin to investigate dimer architecture and its effects on the T12-TfR1 interaction. With these promising initial results, we hope to apply this simple peptide targeting platform to other cardiac-specific targets and their discovered binders.

Thesis Supervisor: Bradley L. Pentelute
Professor of Chemistry

Acknowledgements

There are so many people that grad school wouldn't have been possible without. First of all, I'd like to thank Prof. Brad Pentelute for being my advisor, especially given my unique story of joining the group. He truly saved me from not having a lab to call home, and I got so lucky with both Brad and the people in the lab, as I hadn't met anyone before joining. He's allowed me to work on a project that gained me experience in almost every major branch of science the lab does as well as a bonus dash of plumbing and engineering to take care of the synthesizers. It's really been a blessing to work in a lab where scientific curiosity is encouraged and with the instrumentation to try many things. I'd also like to thank my committee, Prof. Julia Ortony and Prof. Rafa Gomez-Bombarelli. They have offered great feedback and support over the years, always delivered with a smile. Many thanks to Dr. Andrei Loas as well; the lab truly would not function without him. He is always quick to reply to emails with feedback and suggestions as well as keeping track of orders, helping with safety and maintenance, and of course making sure the coffee machine is working. Finally, thank you to Emily Wensberg, Angelita Mireles, Dominique Altarejos, and Amanda Bendzel for all their help with the administrative side of lab and the department.

Of course, I could never have survived grad school without a wonderful group of people to collaborate, commiserate, and clown around with. Thank you so much to the staples of Clown College, Charlie Farquhar, Michael Lee, Dr. Joseph Brown, Dr. Ed Miller, Dr. Diomedes Dieppa-Matos, Dr. Wayne Vuong, Maysa Ilamanova, Amanda Cowfer, Dr. Azin Saebi, Roman Misteli, Gha Young Lee, Dr. Steve Byrne, and Dr. Corey Johnson. I'll miss all the laughs, memes, quotables, roasts, and of course stealing Ed's chair. I could always count on a(n un)necessary trip to 7/11 or having someone to clown around with, and that was so huge on the long days and late nights. Thank you as well to past lab members who played a big part in shaping lab culture and that I looked to as mentors when I was new, especially Dr. Carly Schissel and Dr. Anthony Quartararo. Mega thank you to the mentors who trained me in science, Prof. Nina Hartrampf, Dr. Joseph Brown, Dr. Genwei Zhang, and Dr. Aurelie Rondon. I learned so much from you, and I loved chatting and getting to know you as we worked! Also, huge thank you to Novo team: Dr. Nina Hartrampf, Amanda Cowfer, Stephanie Hanna, Mackenzie Poskus, and Dr. Alex Callahan (Phase I); Dr. Joseph Brown, Michael Lee, Dr. Tom Wood, Roman Misteli, Dr. Genwei Zhang, and Dr. Chengxi Li (Phase II). I'll miss scrambling to get data for the JSC, the traditional post-JSC lunches, and of course the legendary micromeetings. Thank you to Yehlin Cho for being my desk neighbor for the past few years. Thank you to everyone who was at any point on our lab's volleyball team with me (Carly, Ed, Azin, Anthony, Katsushi, Hiro, Dio, Cam, Aaron, Hannah, Beck, Coralie, Lee, Ethan, Corey, Steve, Giulio, Vlad, Peiyuan, Martin); it was so much fun to play and mostly win with y'all! Thank you to everyone who trained me on an instrument, taught me a technique, or ever interacted with me in the lab, past and present. You all really made grad school a great experience and I feel so lucky to know you all!

Thank you to all of the friends I made in my department who got me through classes and quals; it feels so long ago, but it wouldn't have been possible without you all. Kevin, Haihao, Michael, James, Daniel, Jatin, and Jonathan, thank you all for making the last academic classes I'll probably ever take bearable. Thank you to the high

school and college friends who have been a huge support since long before I came to Boston. Jack, Luke, Mahir, Nick, Miclos, Matt, Haris, Chris, it's wild that I've known you for over 10 years now, but I still love hanging out with y'all. The regular TableTop and Jackbox during Covid was the highlight of my weeks. Pedro, thank you for introducing me to Sam and your friend group; Tanner thank you for sending songs; Karim, Greg, Gene, Joe, and Schmidt thank you for keeping the snapchat streaks alive.

I definitely would have gone crazier than I am if it weren't for my sports. They both kept me active enough to eat all the desserts I wanted as well as boosting my confidence to be a leader, and I loved organizing events for both teams. I came to grad school knowing I wanted to play volleyball, so I joined the Women's Volleyball Club. Huge thank you to Dr. Tony Lee and Dr. Max Wolf for being such great coaches! Thank you to the players I've gotten to know over the years, Darcy Duke, Carly Schissel, Cathy Melnikow, Amy Leung, Laura Li (and Vika!), Amy Stoddard, Alina Stavroula, Lily Yan, Jackie Cho, and everyone else. I loved playing with you all and chatting on and off the court. The biggest and best surprise of grad school was joining sMITe. I had no intention of playing ultimate in grad school, but not only have I learned a new sport, I met an amazing community that's now in literally almost every facet of my life. Thank you to my amazing housemates past and present (who all play frisbee in some capacity), Dr. Theresa Cloutier, Dr. Amy Xiao, Stella Lau, Mary Zhong, and Grace Chen, for all the cute house activities, chats, photoshoots, paint nights, and of course traditional Punjabi Dhaba house dinners. Thank you to the people on sMITe and Grim past and present who have become amazing friends off the field, Josh Murray, Michael Kural, Michael Ma, Henrik Boecken, Richard Yip, Nic Arons, Kelly Chen, Munir Libos, Saroja Erabelli, Cynthia Ni, Hao Shen, Kavya Anbarasu, Jess Xu, Andrew Churchill, Nathan Fritz, Anastasia Dunca, Caitlin Lian, Luka Govedic, Ashley Chen, Megan Gillen, Federico Ramirez, Lili Wilson, Jenny Li, Axis Sivitz, Brooke Rethman, and so many others. Thank you to the coaches who taught me what I know about ultimate, Lisa Liu, Meredith Duffy, Theresa Cloutier, Dory Ziperstein, Maya Stokes, and Alisa Schor. And of course, biggest thank you to my absolute favorite person from Smim, Miles Cruice-Barnett. Thank you for being in my life and making me part of yours, for bringing me so much joy, laughter, comfort, support, calmness, and taking away my hiccups. Thank you for having the common sense I lack and for keeping me from losing my mind during this crazy time. And thank you to your family for being so welcoming to me!

Through sMITe I grew to love ultimate and joined some club teams in the summer and even got bold enough to captain a team. Thank you Rainbow for being such a great first mixed team and teaching me a lot when I was still new to the sport. Thank you Grims, Cass, JK, Kara, DJ, Tim, Shat, Jenny, and everyone who was on the team with me at any point. Huge thanks to Joe Rafalak, Robert Hsu, Souichi Terada, and Andrew Goh for creating the greatest car group ever, Silly Robo GoJo Wagon (and thank you Joe for driving me everywhere); I love all of our ongoing wagon adventures. Thank you to Candace, Ian, Haley, Noah, Axis, Alex, Andrew, Jazzy, Gabby, Sage, Gabe, Sean, Irene, Nisha, Caitlin, Umar, Jess, Drew, and all of my kiwis for bearing with me learning how to captain a team and being such amazing people to play with. And of course massive thank you to Josh, Hao, Nic, and Lucy for captaining with me and being so much fun to work with.

Last but not least, thank you to my family, who has supported me my whole life. Thank you to my parents, who drove me to and from school, helped me with homework, took care of me when I was sick, and supported me financially so I only had to worry about school. Thank you Kabe for making me actually try in school so I could beat you, and for paving the way so I could just follow you if I didn't feel like choosing a school, major, or class. Thank you Bug for being weird with me when no one else would. Thank you to my Uncle Richard for always making us all laugh and for being everyone's tech support. Thank you to my grandparents for always being supportive of me and helping us pay for school. Thank you Audrey for making me a duckling in Boston and feeding me now that I'm far from everyone else.

I hope I didn't miss anyone, but if you were at any time important in my life, thank you. Thank you everyone, I really couldn't have made it this far without you all!

Table of Contents

List of Figures

List of Tables

1 Background and Overview

1.1 Thesis Overview

This thesis focuses on the application of flow-based peptide and protein synthesis to improving the delivery of gene therapies, specifically antisense oligonucleotides, to cardiac tissue. Chapter 1 details several of the early phases of the project and the workflow that we developed for several steps in the discovery of cardiac targeting agents. It begins with a discussion of target selection and production, which posed a larger challenge than anticipated. Many targets did not satisfy our requirements, and several that did later proved cumbersome to purify or fold. Section 1.4 discusses the flow synthesizers in our lab and the optimizations we did on them, resulting in great success with producing several single-domain proteins, which are further detailed in Appendix I. Section 1.5 elaborates on the process of discovering peptide or peptidomimetic binders to target proteins as well as some of the challenges in this step. We have discovered putative binders to several of our targets, but many of these have failed at the validation stage. We are investigating alternative discovery methods and modifications to AS-MS to try to remedy this.

Because there have been unforeseen barriers to progressing to *in vitro* and *in vivo* tests with the highly cardiac specific proteins and their targeting peptides, we looked to another platform to develop as a proof of concept. Chapter 2 describes our investigation of the transferrin receptor (TfR1)-T12 interaction. Though TfR1 is not extremely cardiac specific, it is expressed in cardiac tissue. Because of this, we observe delivery of our ASO therapeutic to cardiac tissue by a readout of the level of target miRNA expression at doses competitive with therapeutics currently in the clinic. We

hope that the success of this platform can be translated to a protein-peptide or protein-peptidomimetic pair that is extremely cardiac specific, with potentially even greater effects in cardiac tissue in lower doses. The strategy discussed here, combined with others in the future such as co-delivery with a targeted cell-penetrating peptide, could have significant impact in improving the efficacy of cardiac therapeutics.

1.2 The significance and progression of CVD

Cardiovascular disease (CVD) is currently the leading cause of death, comprising about 30% of worldwide deaths as well as contributing to others as a comorbidity.[1,2] [1,2]Patients with diabetes are at a two- to four-fold higher risk of developing CVDs as a secondary condition due to a similarity in risk factors between CVD and diabetes, such as obesity. In fact, about 70% of diabetes patients die of CVD.[3,4] With the incidence of diabetes rising, from 30 million patients in 1985 to almost 400 million patients in 2014, and a projected 1 in 10 people suffering from the disease by 2035, the number of patients at high risk of CVDs is also ever-increasing.[3] Besides diabetes and obesity, other risk factors such as hypertension, atherosclerosis, renal dysfunction, inflammation, and old age can increase a person's chances of developing CVD. These risk factors can lead to structural changes in cardiac tissue, including volume overload, regurgitation, and cardiomyopathy. The structural changes will lead to symptoms including shortness of breath and fatigue. Eventually, this results in heart failure (HF) in which the heart decreases its blood flow even at rest. One of the reasons CVD is deadly is because it progresses irreversibly toward HF. [5,6] Over humans' lifetimes, they only replace <50% of their cardiomyocytes. Hence, ischemic events such as myocardial infarctions (heart attacks) can be catastrophic. Current treatments only manage

symptoms, such as blood pressure regulators (angiotensin-converting enzyme (ACE)-inhibitors and diuretics), rather than addressing the root cause of CVD.[7] As such, researchers are striving to develop novel treatments[8] as well as determine the associated biological causes of CVDs via genomic studies,[9] RNA-sequencing experiments,[10] and phage display studies.[11]

1.3 Cardiac-specific protein targets to improve CVD therapeutic delivery

1.3.1 Determination of cardiac-specific target proteins

Our approach to improving the targeted delivery of cardiac therapeutics is to leverage the binding of targeting peptides or peptidomimetics to cardiac-specific proteins. We formulated a set of requirements a protein must fulfill to be eligible for pursuit as a target. The primary requirements for a protein to be a viable target are tissue specificity – from tissue or transcriptomic detection from known datasets, the protein must be highly expressed in cardiac tissue only[12,13] – and a known extracellular domain, or ectodomain – the protein must be a membrane protein with a targetable domain outside of cells so that it is accessible by targeting agents.[12,14] To further narrow the list of proteins that satisfy these requirements, we prioritized proteins that have some known structural information such as a crystal structure,[15] existing antibodies against the ectodomain, high sequence homology between the mouse and human protein sequences for ease of translation between mouse *in vivo* studies and clinical applications with humans,[13,16] and supporting information in literature. We were especially interested in the biological function of the proteins and insight into the domains of the proteins implicated in their activity, as we aimed to select target domains of the proteins that would not interrupt the proteins' native function, and by extension

potentially disrupt large scale cardiac function. While the list of potential protein targets changed as information became available, we focused on several proteins including atrial natriuretic peptide (ANP)-converting enzyme (Corin), cadherin-2 (CDH2), calcium voltage-gated channel subunit alpha1 C ($Ca_v1.2$), and sodium voltage-gated channel subunit alpha 5 ($Na_v1.5$). We chemically synthesized proteins or ectodomains that were fewer than 200 residues with few to no post-translational modifications (PTMs)[13] using the automated fast-flow peptide synthesis method developed in our lab (Section 1.4 and Appendix I).[17,18] Proteins with more than 200 residues or several PTMs were purchased or expressed.

1.3.2 Synthesis and de novo folding of cardiac-specific protein ectodomains

Chemically synthesized proteins must be folded into their native conformations during binder discovery to ensure a realistic protein target is presented to the peptides. Our strategy for folding protein domains without known folding protocols was to dissolve them at high concentrations of a denaturing agent, typically 6 M guanidine, and slowly dilute it into a “folding” buffer. We developed a matrix of folding buffers that surveyed various salt concentrations and other additives to determine the optimal folding conditions for each individual protein (Table 1.1). To validate the folding of the proteins, we measured the change in protein mass from the formation of disulfide bonds. Because most of the chemically synthesized proteins were small, folding did not result in a significant shift in retention time by reverse-phase (RP) high performance liquid chromatography (HPLC).

Proteins with a single disulfide bond generally folded and appeared pure by RP-HPLC. However, proteins with multiple disulfide bonds adopted multiple conformations and required post-folding purification. We looked to ion exchange chromatography (IEC) as an orthogonal purification method to RP-HPLC.

Component	Conditions	Purpose
Guanidine	0.3, 0.6, 1.2 M	Denaturant; slows folding potential
Arginine	0, 250, 450 mM	Stabilizer; reduces precipitation
Reductant/Oxidant	200 – 2000 µM	Promotes disulfide shuffling
		Use Cysteine/Cystine (pH 7.8) or GSH/GSSG (pH 8.2)
Ratio	1:1, 1:0.2, 1:0.1	Establishes reduction / oxidation potential
Temperature	4°C	Controls folding kinetics; 4 °C gives less aggregation relative to room temperature
Salt (Ionic strength)	All buffers: 50 mM Tris, 19 mM NaCl, 0.8 mM KCl, 1 mM EDTA	

Table 1.1. **Conditions for *de novo* protein folding buffers.** For each chemically synthesized protein, we assessed 18 folding buffers at a 100 µg scale that surveyed the range of the above components and concentrations, 9 buffers per red/ox pair. In general, we found that cysteine/cystine was a more suitable red/ox pair for our proteins relative to glutathione (GSH/GSSG).

1.3.3 Atrial natriuretic peptide-converting enzyme (Corin)

Corin is highly enriched in cardiac muscle tissue by mRNA detection. [12] It cleaves NPPA and NPPB into active hormones, atrial natriuretic peptide (ANP) and ß-type natriuretic peptide (BNP), respectively. This implicates Corin in the pathway to regulate blood pressure and promote natriuresis, diuresis, and vasodilation.[19–23] It is also shed from the cell surface in healthy people, but in HF patients, there is reduced shedding of Corin, making it a potential biomarker for HF.[24] Because of its biological function, it is important to target the correct region of the protein to avoid harmful largescale effects. The protein domain structure is known, with an N-terminal cytoplasmic and transmembrane (TM) domain. It contains two frizzle-like domains (Fz1,2) as well as eight LDL receptor-like domains (LDLR) and a scavenger receptor-like domain (SR). At the C-terminus, Corin has a trypsin-like catalytic domain that is activated by cleavage at

R801. Post cleavage, the catalytic domain remains connected by a single disulfide.[25] Fz2, SR, and LDLR4 are the least important domains to Corin's function.[19] We expressed a soluble Corin reported in literature[19] without the catalytic domain using Expi293 cells in which the TM domain was replaced with an IgGk signal peptide, and added an N-terminal V5 tag and His tag. We captured the protein from the supernatant of the cells using a His-trap column and subsequently purified Corin by anion exchange (AEX). Fractions were analyzed by sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) for protein mass and western blot for protein identity, and the compiled fractions were analyzed by size-exclusion chromatography (SEC) to ensure a uniform product with approximately the expected size and mass based on an SEC standard (Figure 1.1).

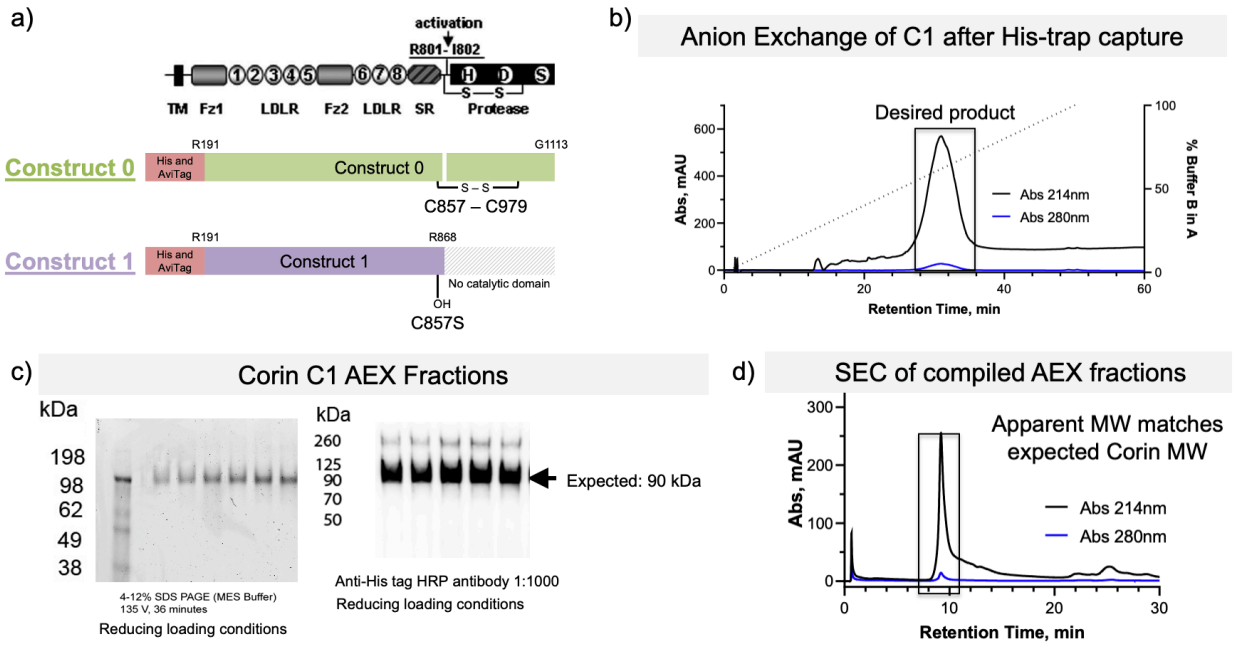


Figure 1.1. **Expression and purification of Corin from Expi293 cells.** A) The illustration of the domains of Corin (reproduced from Knappe et al.[19]) informed the constructs we expressed. B) Construct 1 (C1) was purified by AEX after His-trap capture from the cell supernatant. C) We verified the presence and mass of the Corin C1 by SDS-PAGE and western blot. D) Corin C1 is well-behaved by SEC after AEX purification.

1.3.4 Cadherin-2 (CDH2)

CDH2 shows high cardiac enrichment by mRNA detection and moderate cardiac enrichment by antibody detection.[12] It mediates homotypic cell-cell adhesion by homodimerizing with another CDH2 from another cell.[12,13,15] Predominantly ß-sheet structured in five domains, it contains no disulfides but seven glycosylation sites. To avoid the domains involved in the homodimerization, we purchased the two C-terminal domains 4 and 5.

1.3.5 Calcium voltage-gated channel subunit alpha1 C ($Ca_v1.2$)

$Ca_v1.2$ is highly expressed in cardiac tissue by mRNA detection and moderately expressed in cardiac tissue by antibody detection.[12,13] As its name implies, it is involved in the transport of calcium ions across the cell membrane, making it critical for cardiac muscle contractions. Thus, defects in its expression and function can result in arrhythmia.[10] $Ca_v1.2$ is a primarily transmembrane protein, but it has some small ectodomains that are suitable for chemical synthesis.[14,26] One such ectodomain is $Ca_v1.2_{1072\text{-}1164}$, which contains a single disulfide bond. After assessing folding conditions as described in Section 1.2.2, we determined that 600 mM guanidine, 250 mM L-arginine, 50 mM Tris, 19 mM NaCl, 0.8 mM KCl, 1 mM EDTA, and 1 mM cysteine/1 mM cystine at pH 7.8 was the optimal buffer. Because of the single disulfide, $Ca_v1.2_{1072\text{-}1164}$ did not require further purification post-folding (Figure 1.2). Another ectodomain we studied was the gamma domain, $Ca_v1.2\text{-}\gamma$, which has two disulfide bonds and ß-sheet structure. We determined the optimal folding conditions for this ectodomain to be 1200 mM guanidine, 50 mM Tris, 19 mM NaCl, 0.8 mM KCl, 1 mM EDTA, and 1 mM cysteine/1 mM cystine at pH 7.8. Because of the two disulfide bonds, we observed two

distinct folded conformations by RP-HPLC post folding. We separated the two conformations with AEX using a ProPac-SAX-10 column and 20mM Tris, 3 M urea, 20 mM morpholine pH 8.0 buffer with increasing NaCl gradient (Figure 1.3).

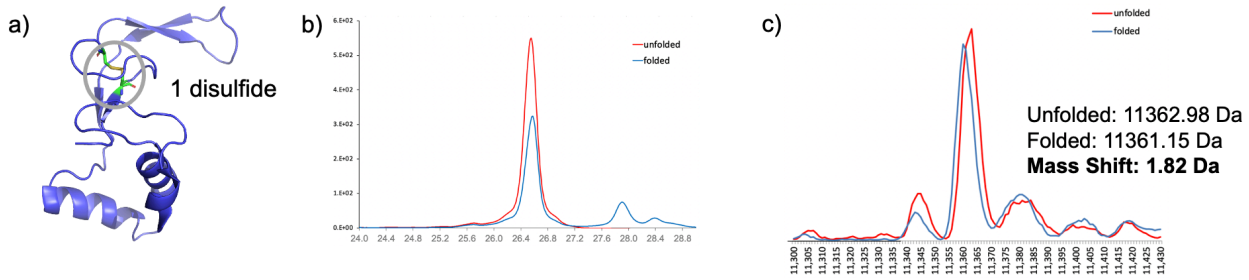


Figure 1.2. **Folding of $Ca_v1.2_{1072-1164}$.** A) $Ca_v1.2_{1072-1164}$ contains a single disulfide but is otherwise largely unstructured. B) RP-HPLC shows no change in retention time between folded and unfolded $Ca_v1.2_{1072-1164}$, likely due to similar interaction with the column before and after folding. C) The deconvoluted mass of $Ca_v1.2_{1072-1164}$ indicates close to complete folding, in which we would observe a mass shift of 2 Da.

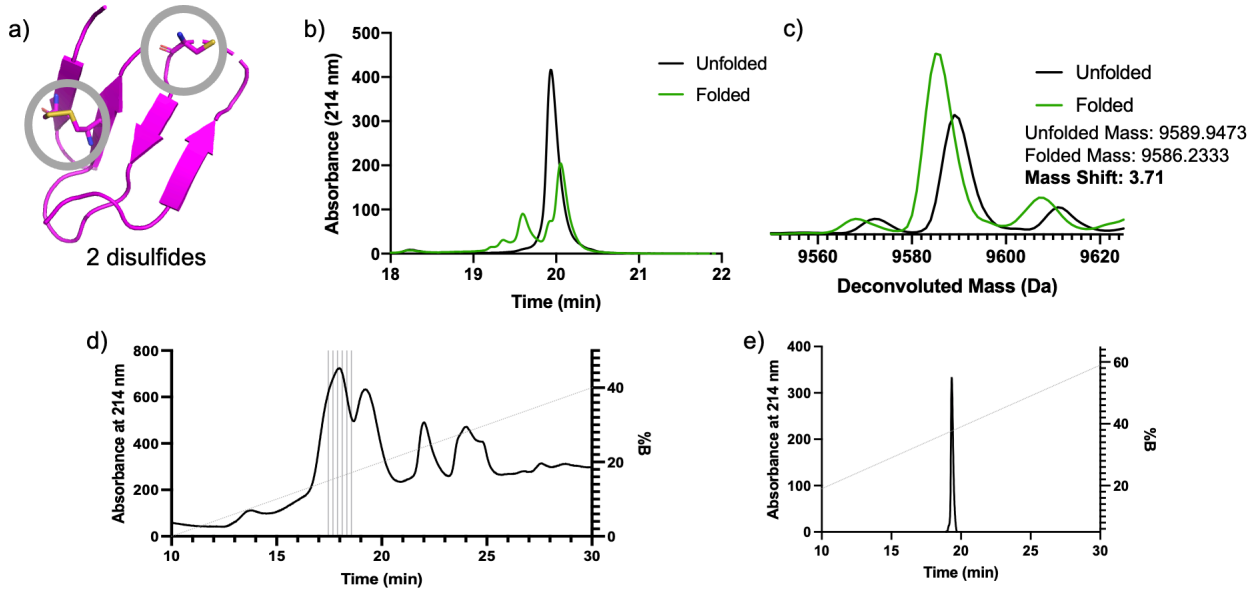


Figure 1.3. **Folding and purification of $Ca_v1.2$-γ.** A) $Ca_v1.2$-γ has a largely ß-sheet structure with two disulfides. B) RP-HPLC shows two main folded conformations that need to be separated. C) The deconvoluted mass of $Ca_v1.2$-γ shows a nearly 4 Da mass shift, indicating both disulfides formed during folding. D) Both folded conformations are visible in the AEX trace, with the first peak producing a pure folded conformation by RP-HPLC (E).

1.3.6 Sodium voltage-gated channel subunit alpha 5 ($Na_v1.5$)

$Na_v1.5$ is nearly exclusively expressed in cardiac tissue by mRNA detection.[12] It mediates the voltage-dependent sodium ion permeability of membranes. There are several sodium ion channel inhibitors that bind to other sodium ion channels, but not $Na_v1.5$.[13] Analysis of its structure shows an ectodomain, $Na_v1.5_{274\text{-}346}$, that contains two disulfide bonds.[14,27] From the folding matrix, most conditions resulted in nearly full oxidation of the ectodomain. Interestingly, unlike with $Ca_v1.2$-γ, only one folded conformation was observed despite there being two possible sets of disulfide bonds possible (Figure 1.4). The folded protein was purified by SEC to yield a pure, folded ectodomain.

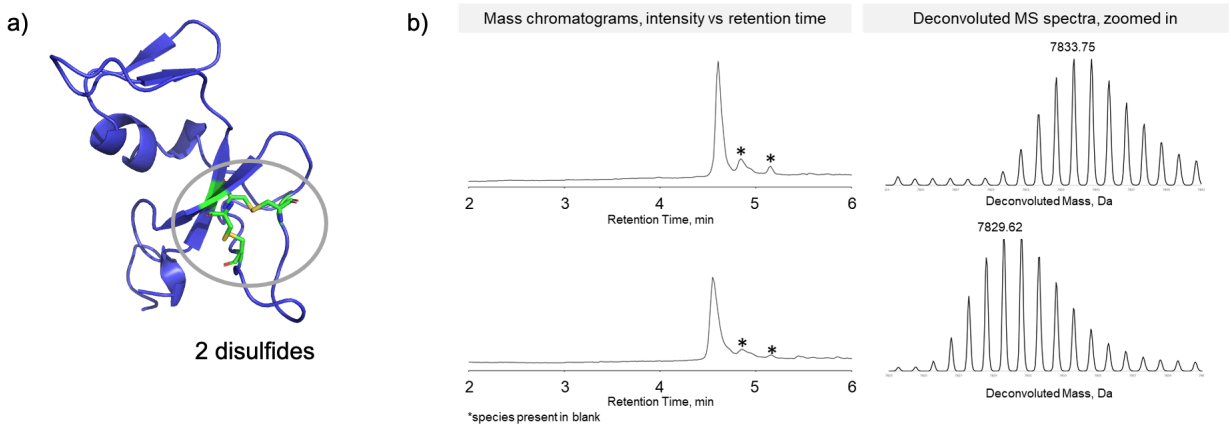


Figure 1.4. **Folding of $Na_v1.5$.** A) The crystal structure of $Na_v1.5$ is largely unstructured but contains two disulfides. B) Folding $Na_v1.5$ results in a single folded species with a nearly 4 Da mass shift, indicating both disulfides are formed.

1.4 Development of automated fast-flow peptide synthesizers (AFPS)

1.4.1 Solid phase peptide synthesis (SPPS)

While proteins and peptides are naturally expressed by cells and bacteria, chemically synthesized proteins offer several advantages. Synthetic proteins allow for more chemical freedom than recombinant proteins, for example the ability to incorporate

unnatural amino acids or access chemical PTMs.[28] Therefore, there has been longstanding interest in developing robust chemical synthesis of proteins and peptides. The first solution phase chemical synthesis of a peptide was performed by Fischer and Fourneau in 1901. With unprotected amino acids in solution, there was no way to prevent the polymerization of the same amino acid to itself, creating many side products in addition to the desired sequence-defined peptides.[29] In 1932, Bergmann and Zervas added reversible protecting groups to amino acids, allowing for sequence-defined peptide synthesis.[30] A major advancement came in 1963 when Merrifield introduced solid phase peptide synthesis (SPPS). In SPPS, the peptide grows on an insoluble resin, facilitating wash steps between amine couplings without sacrificing synthetic quality.[31] In SPPS, an N-terminally protected amino acid is activated in the presence of a base to form an activated ester, which is added to a growing peptide chain on a solid phase resin. The activated ester reacts with the exposed amine at the N-terminus of the peptide. Finally, the protecting group is removed from the N-terminus, resulting in an exposed amine group ready for the next coupling. Because of the exponential effect of the efficiency of each coupling, the theoretical yield of sequences drops to about 50% in about 40 residues even with 99% coupling efficiency, so SPPS was generally limited to less than 50 amino acids. To access full length single-domain proteins, Kent developed native chemical ligation (NCL) as a strategy to chemically synthesize pieces of a protein and ligate them to form a functional protein.[32]

1.4.2 Automated fast-flow peptide synthesis (AFPS)

Though traditional SPPS is still a functional method to synthesize peptides, introducing flow chemistry to peptide synthesis is a major advancement towards

accessing longer sequences in a single shot. Batch SPPS couplings generally take about an hour each and 10 min each for deprotection, making synthesis of longer sequences unrealistic from a time standpoint. Our lab developed automated fast-flow peptide synthesizers (AFPS). In AFPS, bottles of all reagents (amino acids, activators, base, and deprotection solution) are flowed by HPLC pumps through selector valves and to a mixing T, where for each coupling the amino acid, activator, and base meet at the same time. The activated amino acid is then flowed through a heating loop and then over heated reaction vessel containing the resin with the growing peptide. Excess reagents are flowed through an in-line ultraviolet-visible (UV-vis) detector before going to waste (Figure 1.5). The in-line UV detector allows us to monitor the sequence progression and aggregation in real time. For each coupling, two peaks appear – first a broad, saturated peak as the excess coupling agents flow through, followed by a smaller peak as the protecting group is removed from the peptide. The relative height and width of the deprotection peaks gives information on the synthetic quality. In addition to the real-time knowledge of synthetic progress, AFPS has improved heat and mass transfer and low residency times on resin relative to batch SPPS. Perhaps most importantly, the constant exposure to fresh reagents and high temperatures vastly speeds up the coupling reaction, allowing for as low as 2 minutes per coupling and 1 minute per deprotection.[17,33] With further optimizations, the AFPS machines in our lab are now able to synthesize single-domain proteins that are biologically functional over 200 amino acids long that are critical for research.[18,34] Many of these optimizations are described in detail in Appendix I.

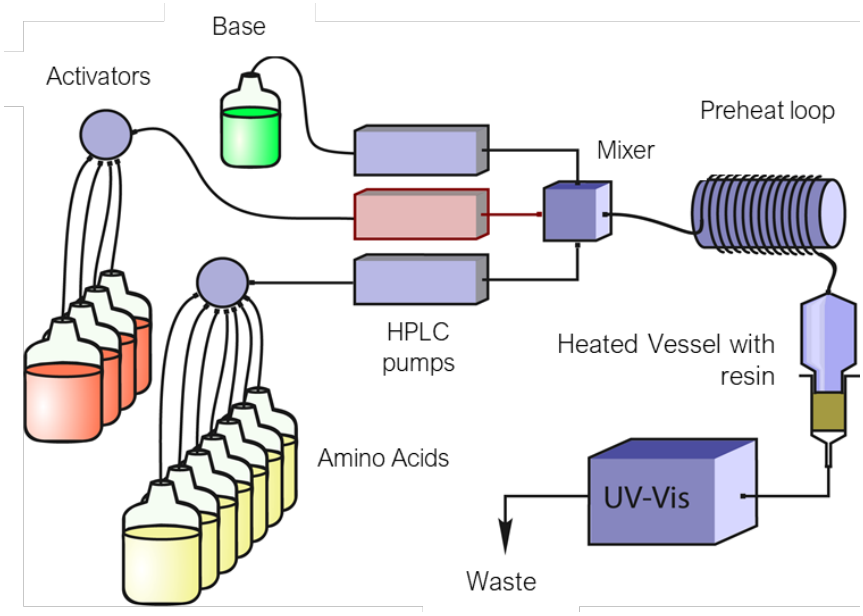


Figure 1.5. **Schematic of the AFPS machine.** There are several synthesizers in our lab, each with slight differences, but all following the same synthetic principles and general setup depicted here.

1.4.3 Renovation of the first generation AFPS

Over time, the AFPS machines get worn down from use. The first generation AFPS was in such a state, so we used the opportunity to fully renovate and redesign the layout of the synthesizer. Our goal was to make it safer and more user-friendly. Because of the orientation of the rack of amino acids, reaching several reagents was difficult to do safely, so one of the most noticeable changes we made was rotating the solution rack and redid the plumbing to remedy this. Additionally, to ensure that the amino acids and activators reach the mixing T at exactly the same time to avoid potentially capping the peptide, we carefully measured the length of the lines going to the pumps such that they were equal in length. Another safety hazard was the exposure of live wires and electrical components on the surface of the fume hood to solvents in the case of leaks. Several wires and fittings were damaged, so we replaced these and added an elevated protected track to prevent future chemical exposure. We also added

a box to protect a set of previously exposed electrical connections. Figure 1.6 displays the layout of the first-generation AFPS before and after renovations.

Another aspect to improving the machine was synchronizing the HPLC pumps. Pump synchronization may only have a small effect on a single coupling, but for longer sequences the effects add up and result in lower quality synthesis. The two main ways the pumps can become unsynchronized is either one tube being shorter than the other, which was remedied by fixing the lines, and cavitation, which can occur if the pump briefly creates a vacuum while refilling with solution. We used colored food dye and a 960 fps video camera to observe frame by frame the time at which color arrived at an endpoint (Figure 1.7). On another synthesizer, we found that too short of a refill time, 200 ms, induced cavitation which prevented the synthesis of barnase, a protein synthesized by a third AFPS machine. Increasing the refill time to 600 ms yielded comparable synthetic quality of barnase on the two machines. Additionally, even in a shorter test peptide GLP-1, synchronizing the pumps increased the purity of the produced peptide from 46% (98.8% efficiency per coupling) to 78% (99.6% efficiency per coupling). Following the renovations to the first generation AFPS, all machines in the lab were able to synthesize similar quality single-domain proteins.

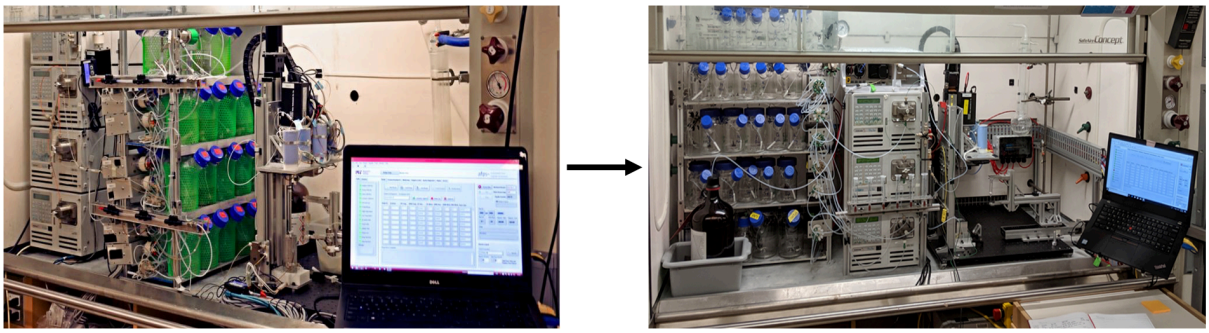


Figure 1.6. **First generation AFPS before and after renovation.** Several modifications were made to the first generation AFPS to improve its synthesis quality, safety, and user-friendliness.

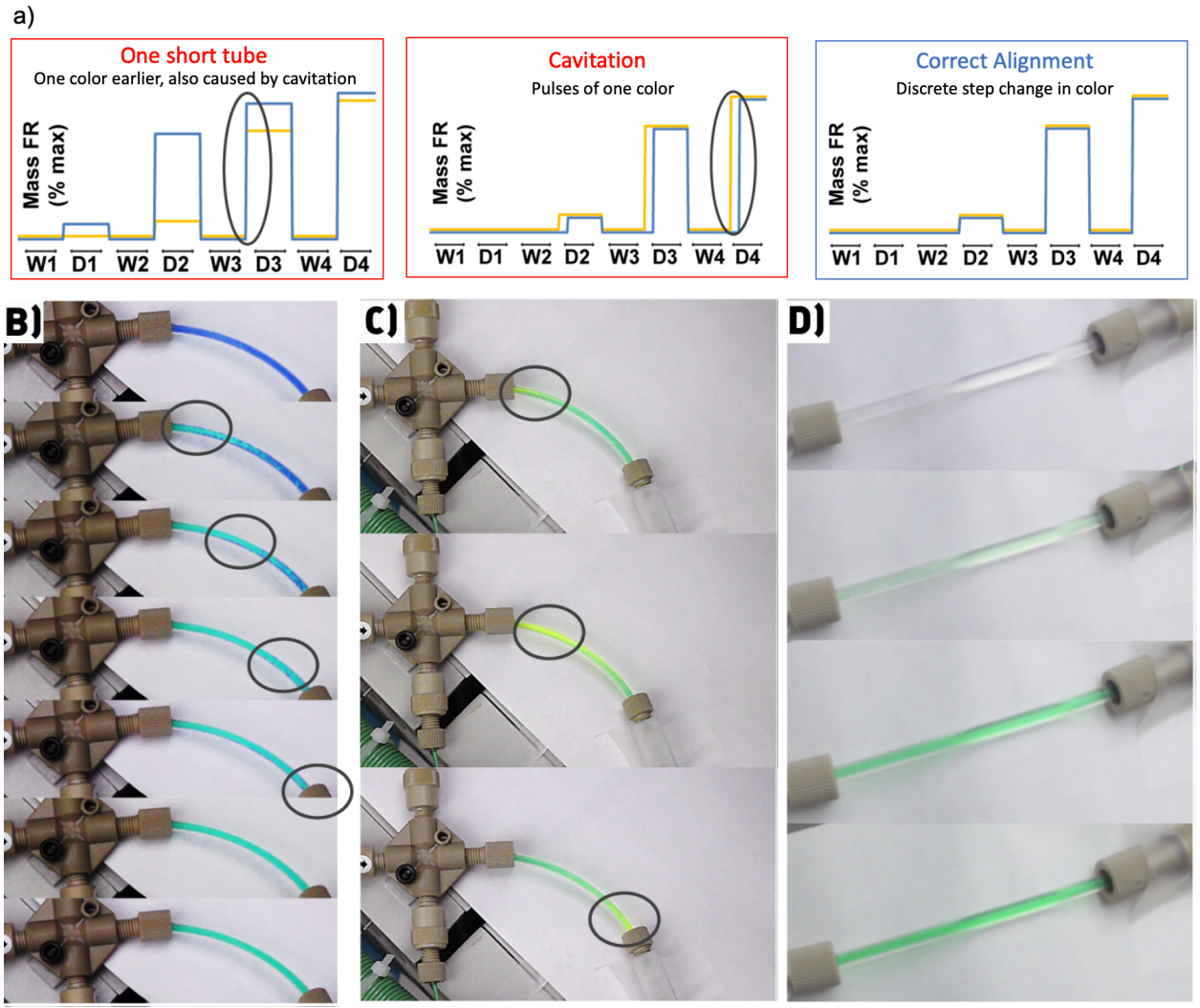


Figure 1.7. **Illustration of pump synchronization.** A) Diagrams show the two potential sources of a lack of sychronization and how they manifest visibly with food dye. Images of dye in the lines for a pump with one short tube (B), cavitation (C), and correct synchronization (D) are courtesy of Dr. Mark Simon.

1.5 Affinity selection mass spectrometry to discover peptidomimetic binders to cardiac-specific proteins

Affinity selection-mass spectrometry (AS-MS) balances chemical control and library diversity to discover putative binders to target proteins such as those described in Section 1.3. For the discovery of novel peptide binders, phage display[35] and mRNA display[36] are common because they routinely sample a sequence space of 10^8-10^{13} members.[37] The key drawback to these methods is that incorporation of D- and

noncanonical amino acids is difficult, as these are biological approaches to binder discovery.[38,39] Though chemical combinatorial techniques offer complete chemical control over the sequence and easy incorporation of any building blocks, they are historically limited to a library diversity of only 10^6.[40] Our lab has developed an AS-MS protocol that can identify sequences from libraries up to 10^8 members while maintaining total chemical control.[41]

AS-MS leverages the robust biotin-streptavidin interaction to facilitate the capture of putative binders to a target protein. The target protein is biotinylated either through a coupling on resin for synthetic proteins or using biotin-NHS ester to non-selectively label lysines or another free amine on a recombinant protein. The biotinylated protein is then immobilized on a streptavidin-coated magnetic bead. The immobilized protein is then incubated for an hour with a synthetic library of up to 10^8 members. Unbound members are washed away. Bound peptides are eluted with guanidine and sequenced using nano-liquid chromatography coupled with tandem mass spectrometry (nLC-MS/MS).[42] To ensure the selectivity of the discovered putative binders, we use anti-hemagglutinin monoclonal antibody (anti-HA mAb or 12ca5) as an off-target protein because it binds to peptides with a known binding motif (D**DYA) in a common library structure we use in selections.[41] Because we observe consistent results with 12ca5 in our AS-MS protocol, it acts as a positive control for the success of the selection overall. We also run multiple target proteins in parallel in addition to 12ca5 to act as off-target proteins for each other. Putative binders are validated using biolayer interferometry (BLI) to confirm their binding to the target protein and determine their binding affinities.

In AS-MS, we chemically synthesize libraries with the solid-phase one-bead-one-compound (OBOC) method.[43] For each coupling, the resin beads are split equally into as many aliquots as monomers for that library. After the amino acid is coupled, all the resin is pooled together for deprotection. Before the next coupling, the resin is equally aliquoted again. In this way, each bead will have a distinct sequence, hence the name OBOC. The most standard library we use in initial screens for binder discovery has twelve variable residues with a single C-terminal lysine to aid with solubility and sequencing (X12K), with the monomer set comprising the canonical L-amino acids except for cysteine and isoleucine. Cysteine is excluded to ensure that the peptides do not cyclize or multimerize, as this can create complications in the sequencing of the discovered binders. Isoleucine is excluded because it has the same mass as leucine, and the sequencing is done by measuring the mass of each peptide. Because the theoretical diversity of the X12K library is so much higher than the number of beads (10^{15} vs 10^{8}),[41] it has a very low redundancy, so each bead should have a distinct sequence. Some variations of the canonical-L X12K library are a slightly shorter and lower diversity canonical X9K library with nine variable residues and the same monomer set, and the canonical-D versions of both of these libraries. Macrocyclic-L libraries are often more likely to yield binding peptides because the steric constraint the peptides are already under reduces the entropic penalty of binding a protein.[35] We are in the process of creating a robust protocol to perform AS-MS with macrocyclic peptides, which is a greater challenge because they require extra steps to cyclize them after synthesis and to linearize them before sequencing. Additionally, we have a non-canonical X12K library that samples a much larger chemical space. These libraries were used to discover

serum-stable peptidomimetic binders to angiotensin converting enzyme 2 (ACE2), described in Appendix II.[44]

Some proteins have known binders or putative binders from a previous round of *de novo* binder discovery. In these cases, we can synthesize a lower diversity focused or specialized library. For example, if a protein has a known binding partner, and especially if a motif is known from that partner, a focused library can mimic the binding partner and include the binding motif to screen for an even stronger binder. As another example, a putative binder to CDH-2 was discovered, so we performed an alanine scan in which each residue in the sequence is mutated to alanine one by one. We check each alanine mutant with BLI to see which residues maintain or lose binding. If mutating a given residue to alanine removes all binding, we conclude that that residue is a "hot spot", or key to the binding interaction. Conversely, if mutating a given residue has no effect on the binding, we conclude that it is a "cold spot", or not important to the binding interaction. After determining the hot and cold spots for the CDH-2 binder, we synthesized two focused libraries. The hot spot library only had variable regions at the hot spot residues and aimed to make the "hot spots hotter". The cold spot library only had variable regions at the cold spot residues and aimed to make the cold spots warm.

Despite the advances in AS-MS, there are ongoing efforts to improve this discovery platform. Many parameters can affect the quality and number of recovered peptides from a given selection. We have varied the buffers used, incubation times, and number of washes (stringency of the selection) as physical variables. However, one of the largest losses of data comes from the sequencing and analysis step. Parameters of the sequencing, such as time spent sequencing and strength of the ionization of the

peptides can affect how many peptides are both discovered and sequenced with high fidelity. One solution is to look to machine learning to aid in handling massive datasets. There are efforts in the lab to develop a program PyBinder that will look at the raw MS spectra and detect peptides to later be sequenced. With this program, the nLC-MS/MS step can be optimized to spend a long time acquiring a high fidelity sequence only for real peptides.

2 A transferrin receptor 1-targeted PNA-peptide conjugate inhibits microRNA-21 expression in cardiac and other mouse tissues

The work presented in this chapter has been reproduced and adapted from the following publication:

Antilla, S.*; Zhang, G.*; Li, C.*; Loas, A.; Nielsen, T.E.; Pentelute, B.L., A transferrin receptor 1-targeted PNA-peptide conjugate inhibits microRNA-21 expression in cardiac and other mouse tissues. *Manuscript in progress.*

**These authors contributed equally to this work.*

2.1 Introduction

The implication of many microRNAs (miRNAs) in disease progression has been documented.[45,46] MiRNA genes are transcribed by RNA polymerase II in the nucleus, producing the precursor miRNAs, termed primary miRNAs, which can be further capped and polyadenylated into pre-miRNA transcripts.[47] Pre-miRNA typically exists in a hairpin-shaped structure and is processed by the Dicer enzyme into short single-stranded miRNA (mature miRNA) consisting of ~22 nucleotide bases after being exported into the cytosol via the nuclear pore.[48] Functional mature miRNA can incorporate into the RNA-induced silencing complex (RISC) in the cytoplasm and complementarily bind the 3'-untranslated region of mRNA targets, regulating protein expression through either mRNA degradation or translational suppression.[47,49] Over the past two decades, nearly 2,000 precursors and over 2,600 mature miRNAs have been identified and annotated in humans.[50] Many miRNAs were found to be potent regulators of intracellular signaling pathways and act as important effectors of different cellular functions, such as cell proliferation and differentiation,[51–53] survival,[54–56] extracellular matrix production,[57] and other regulatory events.[7,45,58,59] Individual miRNAs can affect the expression of several mRNAs and govern complex biological processes. Not surprisingly, dysregulation of miRNAs was found in some pathophysiological stress and disorder conditions.[7,55,60–62]

In the diseased state, relevant miRNAs can be either overexpressed or downregulated, therefore presenting opportunities for therapeutic intervention by using either oligonucleotide-based inhibitors or miRNA mimics.[43] To combat upregulated miRNAs in diseased tissue, targeting strategies with antisense oligonucleotides (ASOs)

have gained traction as potential therapies.[56,63,64] Antagomirs, as a novel class of chemically engineered oligonucleotides, first demonstrated efficacy toward silencing specific miRNAs *in vivo* in 2005[65] and have shown promise in developing treatments of overexpressed miRNAs.[66–68] Chemical modifications to the ASO backbone such as phosphorothioate (PS), 2'-*O*-methyl (OMe), 2'-*O*-methoxyethyl (MOE), and locked nucleic acids (LNAs) have greatly improved their metabolic stability, enhanced RNA target specificity, and alleviated off-target toxicity.[69,70] In other modifications of ASOs, the charged backbone is replaced with neutral components to reduce electrostatic repulsion, such as in phosphorodiamidate morpholino oligomers (PMO) and peptide nucleic acids (PNA). PMOs of at least 15 nucleotides are highly specific, easily soluble, and non-toxic.[69] Unlike the other ASOs, the backbones of PNAs consist of repeating neutral *N*-(2-aminoethyl)glycine units, rendering them chemically, thermally, and enzymatically stable. Additionally, PNAs exhibit higher binding affinity with their DNA or RNA targets due to their neutral backbone.[69,71]

Recent studies illustrate how delivery of ASOs can safely affect the progression of diseases in humans. Inhibiting miRNA-92a with ASOs led to improved vascularization in myocardial infarction models and more rapid wound healing.[72] In 2020, the first in-human study of MRG-110, an LNA-DNA mixed antagomir that targets miRNA-92a, showed time- and dose-dependent inhibition of miRNA-92a expression relative to the placebo group after a single dose over the course of twelve weeks.[73] At the highest dose of 1.5 mg/kg, >95% inhibition of miRNA-92a was observed after 24 h. No significant side effects were reported in the patients participating in this study.[73] In 2021, the first in-human study was reported of CDR132L, an LNA-based antagomir that

targets miRNA-132, which is upregulated in response to cardiac stress, eventually leading to heart failure (HF). Patients in this study received two doses or a placebo four weeks apart. At all doses, rapid and sustained inhibition of miRNA-132 was observed in plasma relative to the placebo group. Importantly, patients receiving at least 1 mg/kg of CDR132L showed plasma levels of miRNA-132 similar to that of a group of healthy volunteers. None of the patients in the study reported adverse side effects.[74,75] These two studies demonstrate that ASOs are tolerable and effective in humans, augmenting their promise as therapies.

In addition to miRNA-92a and miRNA-132, miRNA-21 is overexpressed in failing hearts, and its inhibition shows promise as a treatment for HF. MiRNA-21 is involved in the extracellular signal-regulated kinase—mitogen-activated protein kinase signaling pathway that modulates apoptotic cell death in cardiac fibroblasts. Increased expression of miRNA-21 reduces the percentage of apoptotic cells of failing hearts, and the greater survival of damaged cells leads to morphology changes exhibited in HF. Silencing miRNA-21 with ASOs in mouse models has shown reduced fibrosis and cardiac dysfunction.[7] Transverse aortic constriction (TAC) model mice treated with anti-miRNA-21 exhibited higher fractional shortening and decreased cardiac size relative to control mice, partially recovering to healthy measurements.[7]

Despite encouraging results demonstrating the efficacy of ASOs, delivering them to extrahepatic tissues is a challenge because they are often trafficked to the liver or kidney for excretion. Investigating various techniques to modulate the biodistribution of gene therapies is an area of active research. For example, studies have recently shown that *N*-acetylgalactosamine (GalNAc) conjugated to small interfering RNA (siRNA), a

double-stranded RNA that activates the RISC complex similarly to miRNAs, can target liver hepatocytes.[43,76] Targeting the central nervous system and lungs has been achieved by modifying siRNA with lipophilic moieties.[77] Another strategy leverages the specific binding interactions between antibodies and their target proteins to target non-liver tissue. Antibody-oligonucleotide conjugates (AOCs) were developed that help reduce gene splicing errors in target mRNAs implicated in myotonic dystrophy type 1 (DM1).[78] Though antibodies are highly specific, they are large (>150 kDa) and complex glycosylated molecules. It has been shown that using only the antigen-binding fragment (Fab) of an antibody to deliver PMOs to muscle tissue can maintain the selectivity of the targeting agent while reducing the molecular weight of the conjugate.[79] An alternative to antibody-based targeting strategies is employing adeno-associated virus vectors with low pre-existing immunogenicity to target skeletal and cardiac muscle tissue to treat diseases such as Duchenne muscular dystrophy (DMD).[80] These and other targeting strategies show promising results for improved delivery of gene therapies to desired tissue.

One protein of particular interest for targeted delivery of gene therapies is the transferrin receptor 1 (TfR1). TfR1, or CD71, is highly expressed in cardiac and skeletal muscle tissue and mediates iron transport into cells by binding its native partner transferrin (Tf), which has been used as a metallodrug carrier.[81,82] Several classes of targeting agents previously discussed are currently being studied in the context of targeting TfR1, including AOCs and Fabs, for the treatment of DMD and other muscular diseases.[78,79] Centyrins, antigen-specific proteins with specific sites for drug conjugation about 1/15 the size of an antibody, are being investigated as an siRNA delivery agent to

treat Pompe disease.[83] To avoid potentially disrupting iron cycling into cells, there are ongoing efforts to discover bicyclic peptides that target regions of TfR1 that do not bind to Tf.[84] All of these strategies show significantly improved delivery of the gene therapies to muscle tissue, but they rely on large molecular weight targeting agents that sometimes require high doses for efficacy. For example, one reported Fab complex is effective at 30 mg/kg PMO, corresponding to 68 mg/kg total dosage of compound.[79] Simplifying the structure of the targeting agent would allow for lower doses to achieve equivalent efficacy, further reducing treatment costs.

Here we report a structurally simple linear 12-mer peptide with nanomolar affinity to TfR1 that alters the biodistribution of anti-miRNA-21 PNA relative to PNA alone. Two peptide binders to TfR1, 12-mer and 7-mer long, respectively, were discovered in 2001 by phage display on whole cells transfected to express TfR1.[85] Competition assays with Tf indicated that neither binder competes with Tf binding to TfR1. Using phage titer, it was determined that the 12-mer peptide has 15 nM binding affinity to TfR1, while the 7-mer peptide's binding affinity was only in the micromolar range.[85] We leverage the strong interaction with TfR1 of this known 12-mer binder, termed T12, to modulate the expression of miRNA-21 *in vitro* and *in vivo* via the delivery of an anti-miRNA-21 PNA. By conjugating T12 to PNA, we observe knockdown of miRNA-21 expression in cardiac tissue *in vivo*, whereas no such activity was observed with unconjugated PNA.

2.2 Results and Discussion

2.2.1 Chemically synthesized PNA-T12 antagomir shows selective binding to its pre-miRNA target

Previously, we developed a single-shot strategy to synthesize peptide-conjugated PNAs on an automated platform,[86] which was controlled using a modular script in the 'Mechwolf' programming environment (Figure2.1a).[68] The PNA sequence was reverse-complemented to specifically target miRNA-21. Synthetic T12, amino acid sequence THRPPMWSPVWP,[85] was verified to exhibit nanomolar binding affinity to the recombinant mouse TfR1 ectodomain (residues C89-F763) by bio-layer interferometry (BLI, Figure 2.1b). We observed an apparent dissociation constant K_D = 26 nM, in line with the previously reported value of 15 nM determined by phage titer.[85] For the PNA-T12 conjugate, we included a simple Lys_3 linker between the two components, resulting in the final sequence of CATCAGTCTGATAAGCTA-KKK-THRPPMWSPVWP-$CONH_2$. After multiple rounds of synthesis and purification, >10 mg of PNA-T12 (>95% purity by LC-MS) were isolated. Following sample preparation, a gel shift assay was performed to demonstrate that chemically synthesized PNA-T12 could selectively bind to the target miRNA sequence. As shown in Figure 2.1c, the anti-miRNA-21 PNA-T12 construct bound and shifted pre-miRNA-21, but not pre-miRNA-34a used as a control, indicating miRNA target selectivity.

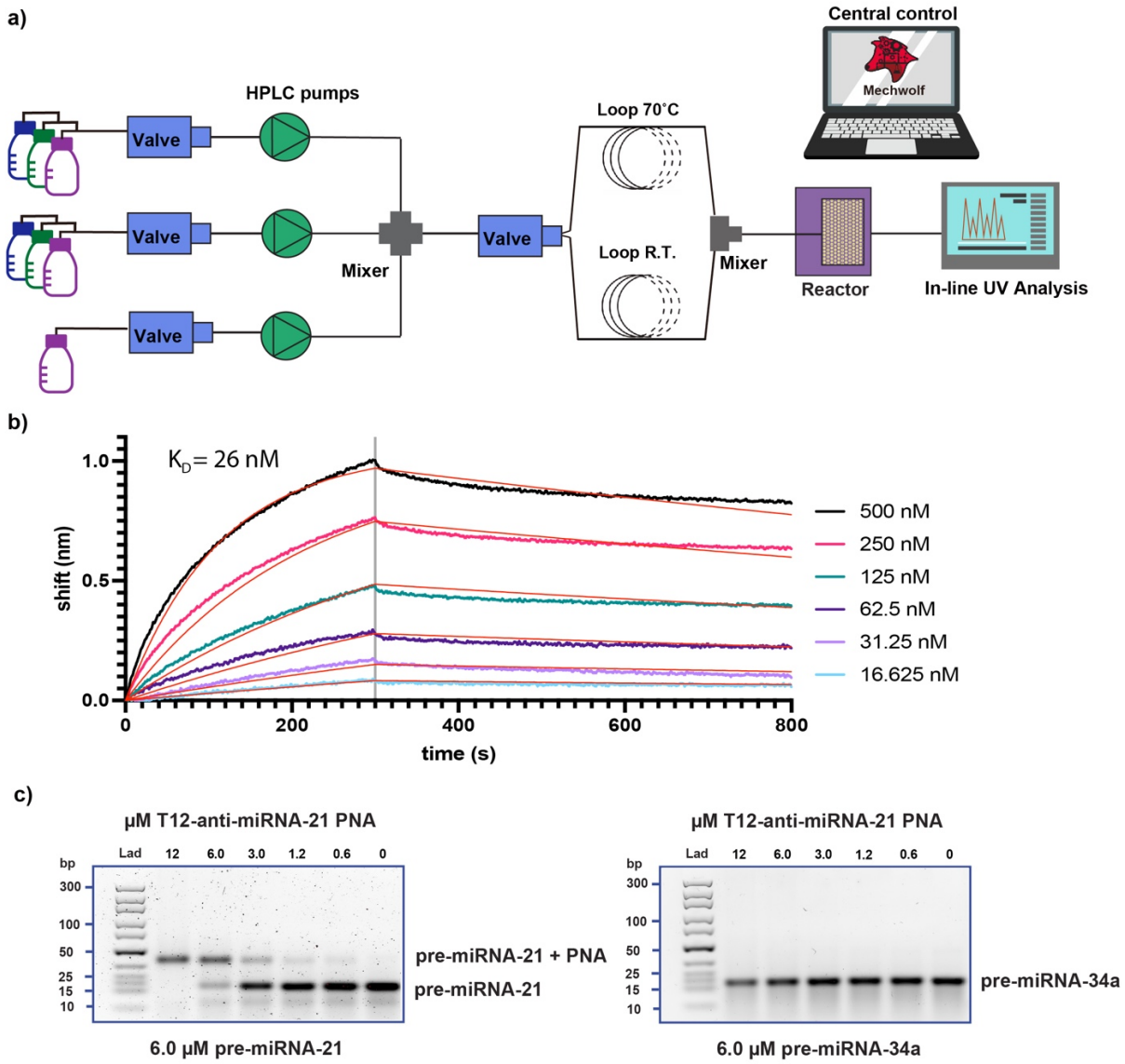


Figure 2.1. **Chemically synthesized PNA-T12 conjugate binds the targeted pre-miRNA.** (a) T12-conjugated PNA was prepared by automated single-shot flow synthesis. (b) Synthetic T12 exhibits concentration-dependent binding to recombinant TfR1 ectodomain with nanomolar affinity by BLI. Biotinylated T12 loaded on streptavidin (SA) tips at 1.5 μM was allowed to associate with mTfR1 at the above concentrations for 300 s, followed by a dissociation step for 500 s. (c) Gel shift assays show the selective binding between chemically synthesized anti-miRNA-21 PNA-T12 and pre-miRNA-21, with pre-miRNA-34a used as a control. For each gel, 6 μM miRNA was pre-mixed with 0-12 μM PNA-T12 at 37°C for 1 hour and loaded into a 5% agarose gel at 400 ng miRNA/well. Gels were run at 100 V for 70 min in 1X TBE buffer.

2.2.2 Anti-miRNA-21 PNA-T12 conjugate inhibits miRNA-21 expression in mouse cardiomyocytes

To assess the efficacy of our PNA-T12 conjugate for gene silencing, we performed a miRNA inhibition efficacy assay in a cardiac muscle cell line (HL-1). HL-1 cells were derived from the AT-1 mouse atrial cardiomyocyte tumor lineage and can be repeatedly passaged while maintaining a cardiac-specific phenotype.[87] Stable HL-1 cells were incubated with PNA-T12 or PNA alone at gradually increasing concentrations (for PNA alone 0.01, 0.1, 1, and 10 μM; for PNA-T12 0.1, 0.5, 1, and 5 μM). One day later, the attached cells were scraped off the plate and harvested for RNA extraction, followed by reverse transcription and quantitative real-time polymerase chain reaction (qRT-PCR). The mouse miRNA-16 was selected as an internal control for the qRT-PCR reactions based on a previous report,[88] and the comparative C-T method was used to calculate the cycle threshold.[89] A clear dose-dependent reduction inmiRNA-21 expression, up to >50% inhibition at 5μM, was observed in the cells incubated with PNA-T12 (Figure 2.2b). However, no dose- dependent inhibition was observed in the cells incubated with PNA alone (Figure 2.2a). These results indicate that conjugating the targeting ligand T12 to PNA improves uptake and intracellular delivery of PNA in cardiac cells and drives targeted miRNA silencing.

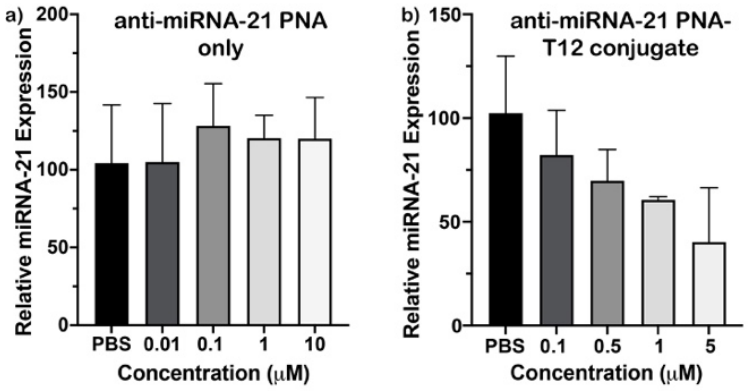


Figure 2.2. **T12-conjugated PNA antagomir displays miRNA-21 inhibition in mouse cardiomyocytes.** The bioactivity of the synthetic PNA-T12 conjugate was tested in cultured HL-1 mouse cardiomyocytes, and the targeted miRNA inhibition efficacy was measured via qRT-PCR. HL-1 cells were incubated at 37°C in complete Claycomb media for 24-36 h after being dosed with PNA or PBS. Anti-miRNA-21 PNA conjugated to T12 shows dose-dependent inhibition of miRNA-21 relative to the PBS control when conjugated to T12 (b), but not on its own (a). For each condition, N=3 technical replicates were used, with relative expression calculated using the comparative C_T method. Statistical significance is calculated using Student's t-test, ** $p < 0.01$; **** $p < 0.0001$; n.s. not significant.

2.2.3 The anti-miRNA-21 PNA-T12 conjugate affects miRNA-21 inhibition in the mouse heart

Minimizing off-target delivery could improve the therapeutic window of antagomir PNAs. We investigated the PNA targeting capabilities of the T12 ligand *in vivo* following the improved uptake in HL-1 mouse cardiomyocytes observed *in vitro*. Two experimental groups of wild-type mice (strain C57BL/6, n = 5) were dosed with 30 mg/kg or 15 mg/kg of PNA-T12 via retro-orbital injection. Two control groups (n = 5) were injected with equivalent volume of 1X phosphate-buffered saline (PBS) and 30 mg/kg PNA alone. All mice were sacrificed two weeks after compound administration. At the time of sacrifice, all mice appeared healthy, and we observed no significant adverse side effects in the physical appearance or behavior of the mice. To measure the efficacy of PNA-T12 toward miRNA-21 inhibition, five organs (heart, liver, kidney, lung, and

spleen, Figure 2.3a) were extracted from dissected mice and preserved in Trizol reagent for subsequent analysis. Total RNA was extracted from the preserved organ samples and qRT-PCR was performed on each sample using primers provided within the Taqman kit. We observed the naked PNA construct at 30 mg/kg caused significant miRNA-21 reduction in the kidney and liver, but not in the heart (Figure 2.3b-d), which reinforces the long-standing challenge of delivering ASO molecules outside of vascularized tissues.[90,91] In contrast, PNA-T12 inhibits approximately 50% of miRNA-21 expression in cardiac tissue at 30 mg/kg (Figure 2.3b). In lung tissue, including a targeting ligand reduces inhibition of miRNA-21 expression relative to PNA alone (Figure 2.3e), indicating that T12 is able to alter the biodistribution of PNA in mice. At 15 mg/kg, there is reduced gene silencing in cardiac liver, kidney, and spleen tissue (Figure 2.3b,c,d,f), consistent with dose dependence. While this outcome is promising in demonstrating that lower doses of these compounds can reduce off-target delivery, no significant efficacy was observed in the cardiac tissue below 30 mg/kg.

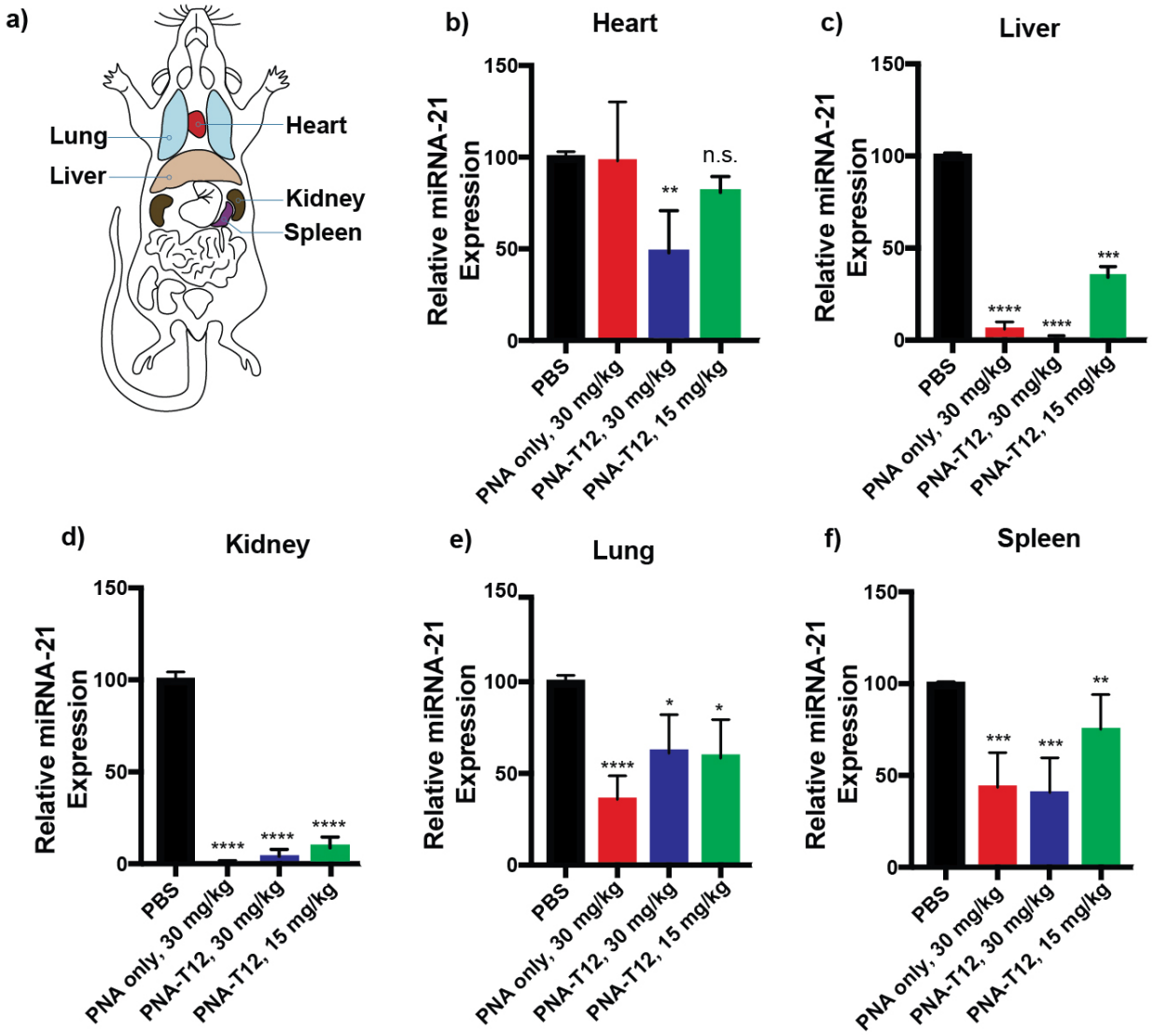


***Figure 2.3*. T12-conjugated anti-miRNA-21 PNA exhibits altered biodistribution relative to PNA alone.** (a) Schematic representation of the mouse organs dissected to assess miRNA-21 inhibition. Anti-miRNA-21 PNA alone drives substantial inhibition of miRNA-21 in the mouse liver (c) and kidney (d), but not in the heart (b). In contrast, T12-conjugated anti-miRNA-21 PNA renders reduced expression of miRNA-21 at equivalent doses in mouse cardiac tissue (b) along with a diminished inhibitory effect in lung tissue (e). The efficacy of the PNA-T12 conjugate in mouse cardiac tissue is reduced at doses below 30 mg/kg (b). Statistical significance is calculated using Student's t-test, ** $p < 0.01$; **** $p < 0.0001$; n.s. not significant.

2.2.4 Synthetic linear dimer of T12 exhibits tenfold stronger binding interaction with mTfR1

To reduce the dose of compound required to inhibit miRNA expression in cardiac tissue, we investigated the effect of dimerizing T12 on its interaction with TfR1. Using automated flow synthesis previously developed,[92] we synthesized a linear dimer of T12 with a Lys_3 linker between the two T12 monomers, with the sequence THRPPMWSPVWP-KKK-THRPPMWSPVWP- $CONH_2$. To compare the binding affinities of this dimer construct with the original monomer T12, we performed BLI to assess the interaction between the T12 dimer and mTfR1. We observed an apparent dissociation constant K_D = 2.5 nM (Figure 2.4a), which is tenfold stronger than the dissociation constant K_D = 26 nM we observed for the T12 monomer.

Following the confirmation of the T12 dimer having stronger binding to TfR1 than the T12 monomer, we synthesized a conjugate of the dimer with anti-miRNA-21 PNA using the same single-shot method[86] as the PNA-T12 monomer conjugate, with the final sequence CATCAGTCTGATAAGCTA-KKK- THRPPMWSPVWP-KKK-THRPPMWSPVWP-$CONH_2$. Following several rounds of synthesis and purification, we isolated >10 mg of PNA-T12 dimer.

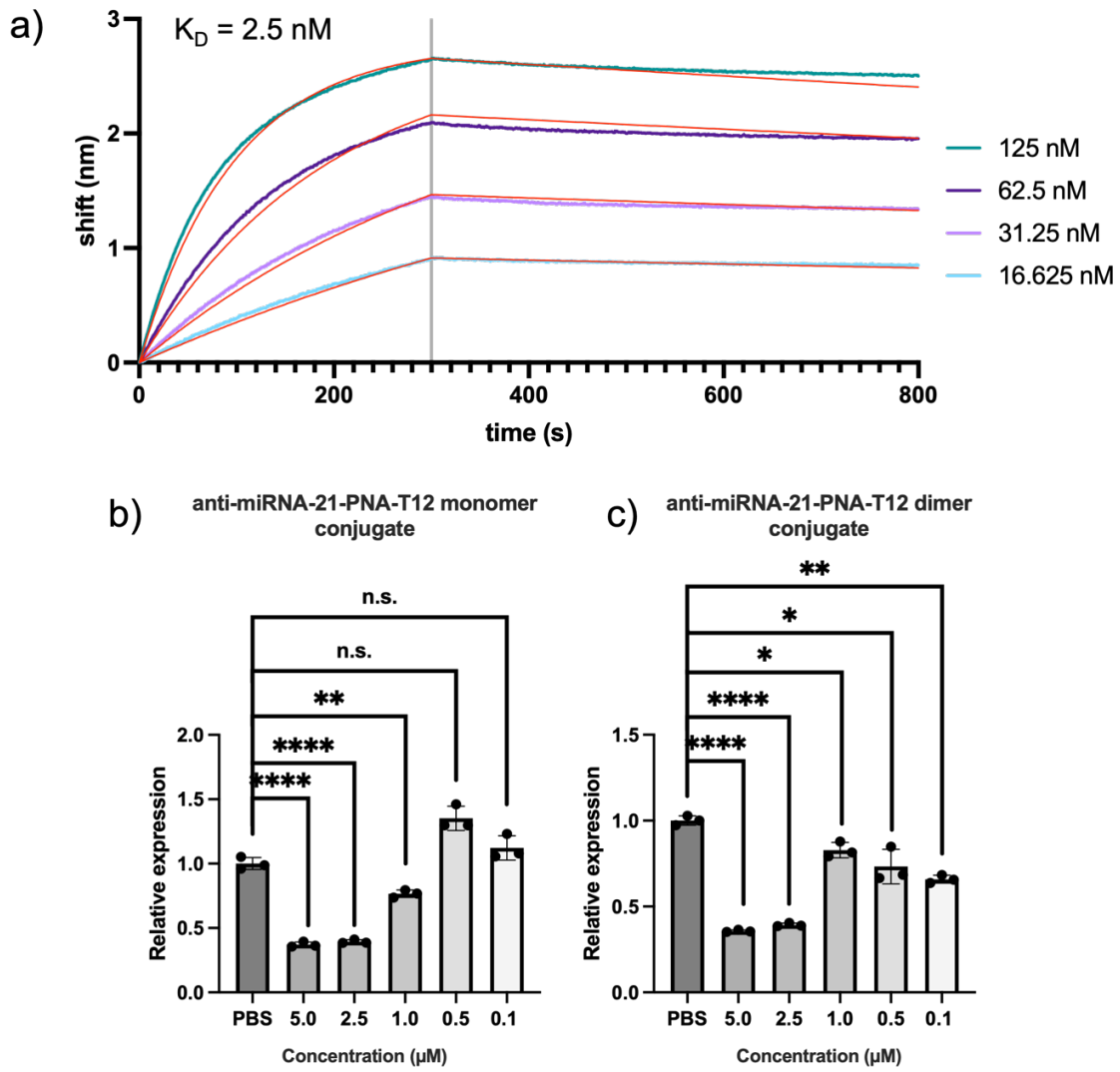


Figure 2.4. **Chemically synthesized T12 dimer exhibits stronger binding to mTfR1, and its PNA conjugate shows improved inhibition of miRNA-21 expression *in vitro* relative to PNA-T12 monomer conjugate.** (a) Synthetic T12 dimer exhibits concentration-dependent binding to recombinant mTfR1 ectodomain with nanomolar affinity by BLI, tenfold stronger than the analogous binding affinity demonstrated by synthetic T12 monomer. Biotinylated T12 dimer loaded on streptavidin (SA) tips at 1.5 μM was allowed to associate with TfR1 at the above concentrations for 300 s, followed by a dissociation step for 500 s. (b,c) The bioactivity of the synthetic PNA-T12 dimer conjugate was tested in cultured HL-1 mouse cardiomyocytes, and the targeted miRNA inhibition efficacy was measured via qRT-PCR. HL-1 cells were incubated at 37°C in complete Claycomb media for 24-36 h after being dosed with anti-miRNA-21 PNA-T12 monomer or dimer. At lower doses, T12 dimer conjugates inhibit miRNA expression (c) while T12 monomer conjugates do not (b) relative to PBS control. For each condition, N=3 technical replicates were used, with relative expression calculated using the comparative C_T method. Statistical significance is calculated using Student's t-test, * $p<0.1$; ** $p < 0.01$; **** $p < 0.0001$; n.s. not significant.

2.2.5 Anti-miRNA-21 PNA-T12 dimer conjugate inhibits miRNA-21 expression in mouse cardiomyocytes at sub-μM doses

To investigate the effect of the improved binding interaction with TfR1, we performed another miRNA inhibition efficacy assay. We cultured HL-1 cells as previously described then incubated stable HL-1 cells with increasing doses of either PNA-T12 monomer or PNA-T12 dimer (5.0, 2.5, 1.0, 0.5, and 0.1 μM). The following day, the cells were scraped from the plate and the RNA extracted. We performed reverse transcription and qRT-PCR as described above with miRNA-16 as an internal control. In this side-by-side comparison of PNA-T12 monomer vs PNA-T12 dimer, the two constructs have similar performance at higher concentrations (5.0 and 2.5 μM), inhibiting approximately 60-65% of miRNA-21 expression relative to the PBS control (Figure 2.4b,c). At lower doses (0.5 and 0.1 μM), the monomer construct does not effectively inhibit miRNA-21 expression, but the dimer construct shows statistically significant inhibition of miRNA-21 expression (Figure 2.4b,c). This result indicates that the T12 dimer construct is functional to deliver anti-miRNA-21 PNA to mouse cardiomyocytes, with similar or improved performance relative to the T12 monomer construct.

2.2.6 Anti-miRNA-21-T12 dimer construct inhibits miRNA-21 expression in the mouse heart at lower doses than its monomer analog

With promising results from our *in vitro* efficacy assay, we wanted to investigate whether the dimer construct would outperform the monomer construct at lower doses *in vivo*. Three experimental groups of wild-type mice (strain C57BL/6, n = 3) were dosed with 30 mg/kg, 15 mg/kg, or 5 mg/kg of PNA-T12 dimer via retro-orbital injection. Two

control groups (n = 3) were injected with equivalent volume of 1X PBS and 30 mg/kg PNA-T12 monomer (positive control from previous experiment). All mice were sacrificed one week after compound administration, at which time no adverse side effects in the appearance or behavior of the mice was observed. We harvested the same five organs as reported in Figure 2.3, the heart, liver, kidney, lungs, and spleen (Figure 2.3a, 5a) and preserved the organs in TriZol reagent for analysis. We extracted RNA and performed qRT-PCR as previously described. At all doses of compound, we observed >50% inhibition of miRNA-21 expression in cardiac tissue (Figure 2.5b). Even as low as 5 mg/kg PNA-T12 dimer, miRNA-21 expression was inhibited by >60%, a significant improvement over the performance of the T12 monomer construct, which showed little to no cardiac delivery at 15 mg/kg. While there is minimal trafficking of PNA to lung tissue for all doses (Figure 2.5e), we observe significant delivery of PNA to liver (Figure 2.5c), kidney (Figure 2.5d), and spleen (Figure 2.5f) tissue. However, there is a reduction in inhibition in kidney tissue at 5 mg/kg T12 dimer construct. This is a promising result that may indicate reduction in kidney delivery at further reduced doses of compound.

At 5 mg/kg we were able to enhance miRNA-21 inhibition in the cardiac tissue of mice using a directed targeting strategy without eliciting observable side effects. This indicates that at the doses studied, our PNA-T12 monomer and dimer conjugates are effective and well-tolerated in mice. Our peptide-conjugated antisense PNA construct opens a route to produce lead compounds to continue in-depth pre-clinical investigations.

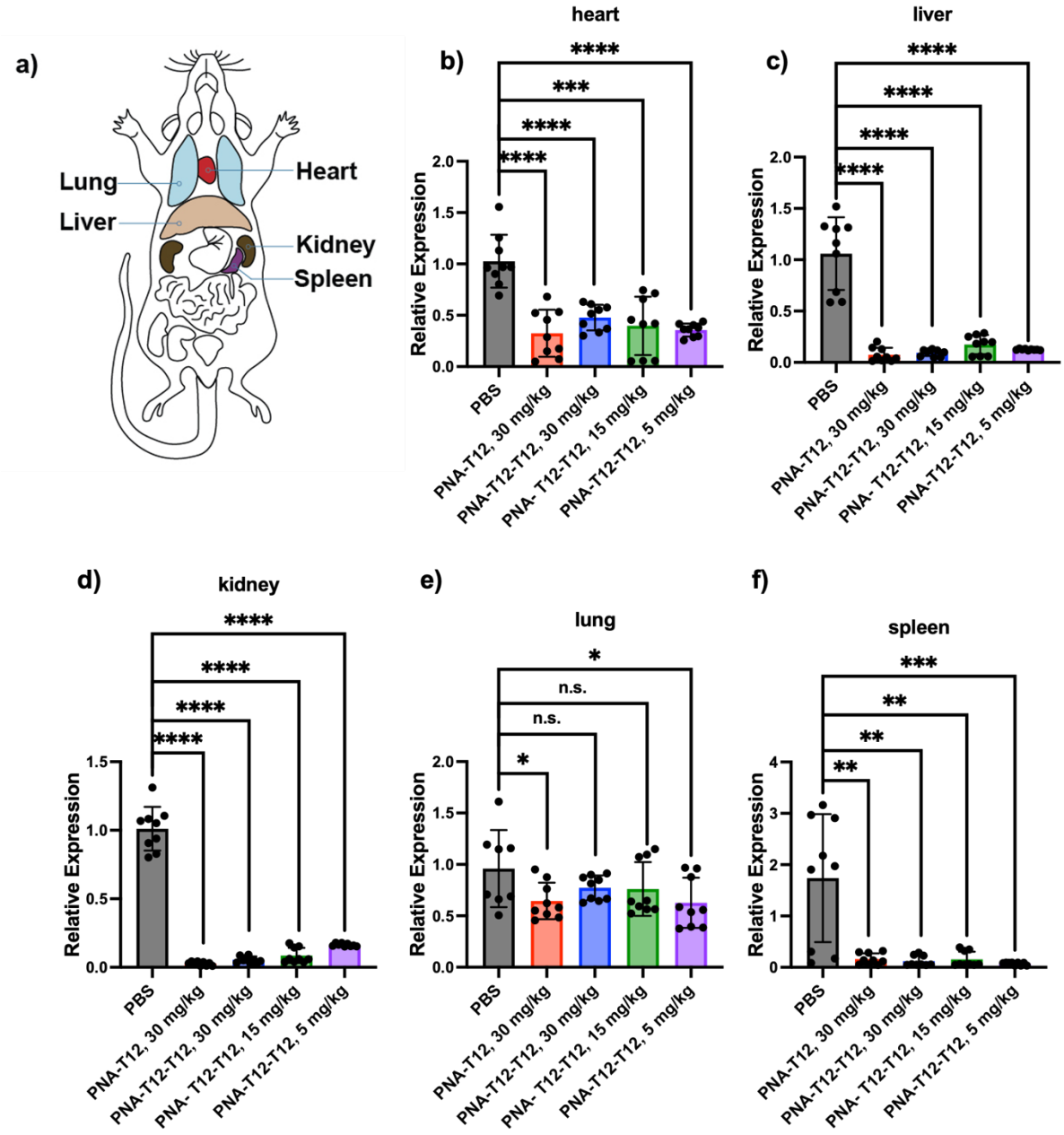


Figure 2.5. **T12 dimer-conjugated anti-miRNA-21 PNA inhibits miRNA-21 expression in cardiac tissue at lower doses than analogous PNA-T12 monomer conjugate.** (a) Schematic representation of the mouse organs dissected to assess miRNA-21 inhibition. While reduced efficacy of PNA-T12 monomer was observed below 30 mg/kg, the efficacy of PNA-T12 dimer conjugates remains stable down to 5 mg/kg in cardiac tissue (b). Significant levels of PNA delivery is observed in liver (c), kidney (d), and spleen (f) tissue, though a dose-dependent reduction of miRNA inhibition is seen in kidney tissue (d). Statistical significance is calculated using Student's t-test, ** $p < 0.01$; *** $p<0.001$; **** $p < 0.0001$; n.s. not significant.

2.2.7 A single residue mutation of T12 eliminates binding to TfR1

Following the strong performance of the linear T12 dimer, we are interested in optimizing the interaction between T12 constructs and TfR1. To do this, we want a negative control non-binding peptide related to T12. To achieve this and simultaneously gain insight into the residues critical to the binding interaction, we synthesized[50] four T12 mutants, mutating prolines and tryptophans that we suspected to be involved in the T12-TfR1 interaction: T12-P5A, T12-W11A, T12-P5A-W11A, and T12-W7A-P9A. We performed the same BLI experiment on each of these that we had done to validate T12 binding to TfR1 (Figure 2.6). T12-P5A bound to TfR1 with a K_D of 6.0 nM, indicating that this proline is not necessary for the interaction to occur (Figure 2.6a). In contrast, T12-W11A exhibited no binding to TfR1, strongly implicating the mutated tryptophan as a key part of the T12-TfR1 interaction (Figure 2.6b). Both double mutants also lost all binding (Figure 2.6c,d), and based on the single mutant results we hypothesize that the mutation of tryptophans plays a larger role in eliminating binding. Importantly, we determined that mutating a single tryptophan yields a non-binding T12 mutant that can be used as a negative control peptide.

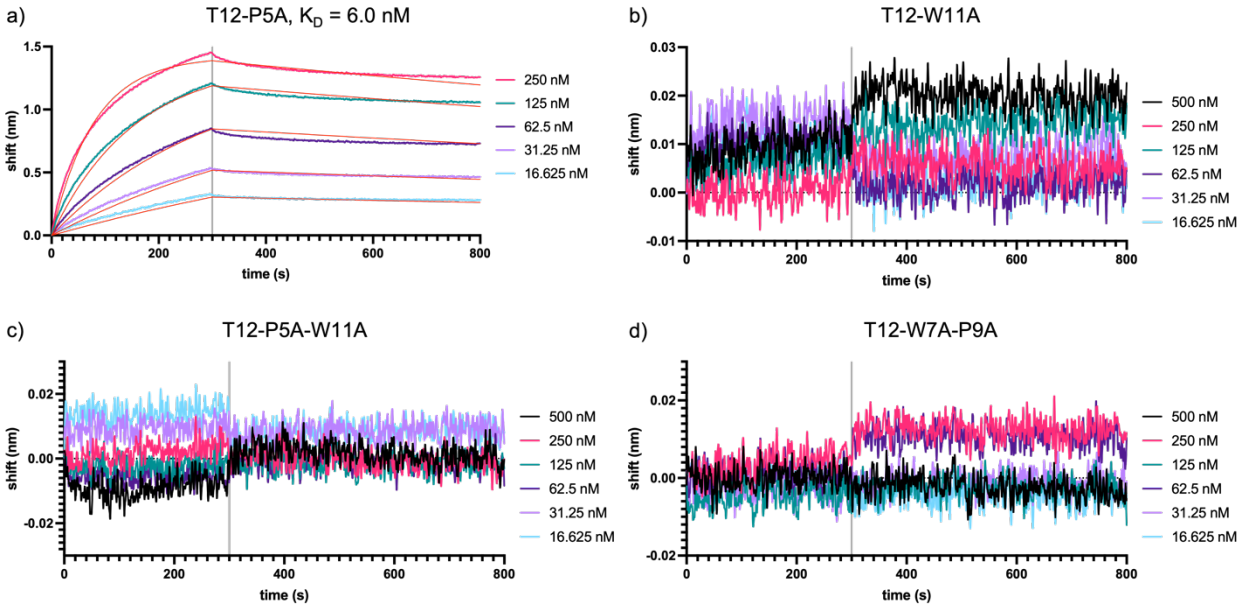


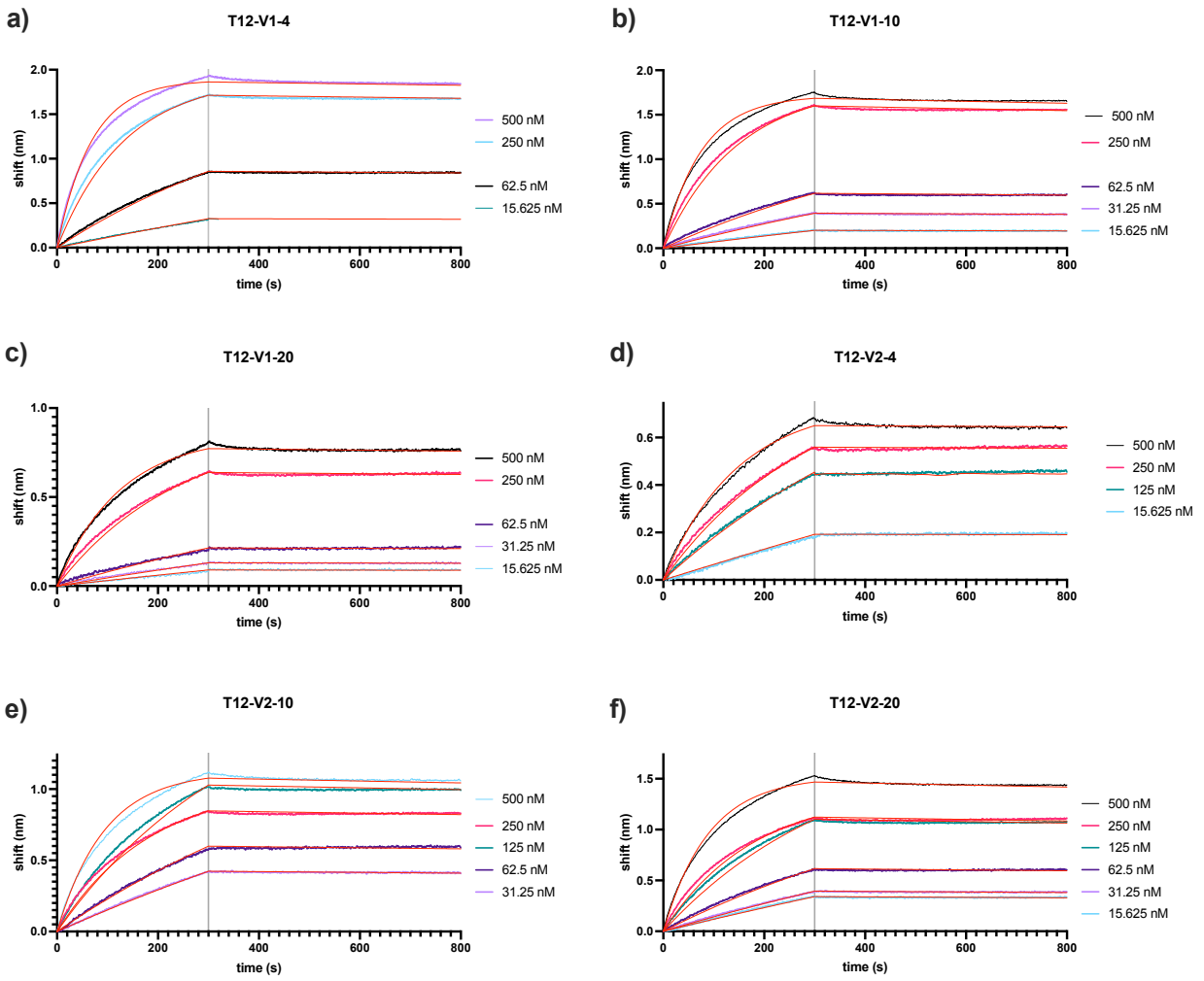
Figure 2.6. **Mutating tryptophans in T12 eliminates is binding to TfR1.** Biotinylated T12 mutants loaded on streptavidin (SA) tips at 1.5 μM were allowed to associate with mTfR1 at the above concentrations for 300 s, followed by a dissociation step for 500 s. (a) A single proline mutation (T12-P5A) maintains binding to TfR1. A single tryptophan mutation, T12-W11A (b) and two proline and tryptophan double mutants, T12-P5A-W11A (c) and T12-W7A-P9A (d), do not exhibit binding to TfR1, implicating the tryptophan at residue 11 and likely at residue 7 as key to the interaction between T12 and TfR1.

2.2.8 Branched T12 dimers exhibit similar binding affinities to TfR1 as the linear T12 dimer

To further understand the interaction between T12 and TfR1, we want to evaluate the effect of varying the architecture of the T12 dimers. There are two branched architectures we are interested in, where the ASO or biotin (for BLI) is between the two T12 monomers instead of at the N-terminus of the entire sequence: both C-termini of the T12 monomers are adjacent to the ASO or other cargo (referred to as “V1”) and where the ASO or cargo is between the N-and C-termini of the two monomers (referred to as “V2”) (Figure 2.7a). Additionally, rather than use a tri-lysine linker, we vary the length of the linker between the monomers using PEG_4, PEG_{10}, or

PEG_{20} linkers. In total, we synthesized[18] six branched T12 dimer constructs using automated flow synthesis. Each of these constructs was labeled with an azidolysine handle so that either PNA-DBCO or biotin-DBCO could be attached with a copper-free click reaction.

From BLI experiments, it is difficult to conclude a superior dimer architecture. We clicked biotin-DBCO to each of the six branched dimer constructs and performed BLI as with the other T12 compounds (Figure 2.7). All constructs exhibited low single digit nM binding affinities, meaning their performance was similar to that of the linear T12 dimer. Figure 2.7b summarizes the BLI performance of all T12 compounds tested. The linear dimer has the fastest association rate (k_{on}) and response by three to fourfold relative to all other compounds. The other branched dimers have a similar k_{on} as the T12 monomer. What gives the branched dimers a stronger overall binding affinity is that they have a slower dissociation rate (k_{off}) than the monomer by about tenfold. Overall, the K_D of the branched dimers is comparable to that of the linear dimer, so it is inconclusive from BLI alone which compound would perform better in a more biologically complex environment.



g)

Biotinylated peptide	K_D (nM)	k_{on}	k_{off}	Response at 500 nM (nm)
T12 monomer	26	2e4	3e-4	1.00
T12 linear dimer	2.5	9e4	6e-5	3.37
T12 V1-4	2.8	3e4	4e-5	1.94
T12 V1-10	3.8	3e4	1e-4	1.76
T12 V1-20	3.4	2e4	5e-5	0.83
T12 V2-4	1.1	1e4	2e-5	0.68
T12 V2-10	3.0	2e4	6e-5	1.12
T12 V2-20	3.1	2e4	7e-5	1.53

h)

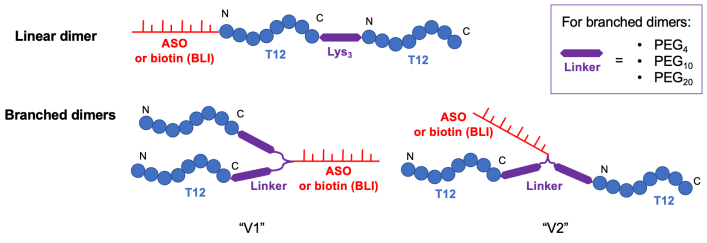


Figure 2.7. **Branched T12 dimers exhibit low single digit nM binding to TfR1.** Biotinylated T12 dimers loaded on streptavidin (SA) tips at 1.5 μM were allowed to associate with mTfR1 at the above concentrations for 300 s, followed by a dissociation step for 500 s. (a-f) Six branched T12 dimers bind to TfR1 in a concentration-dependent manner, with comparable binding affinities to the linear T12 dimer. (g) Comparing the K_D of the branched T12 dimers to the linear T12 dimer, all compounds perform similarly in BLI. Panel (h) depicts the branched structures of the dimers, with the final number in the compound name indicating the length of PEG linker in the dimer (e.g. T12-V1-4 has architecture V1 with PEG_4 linkers).

2.3 Conclusion

In this study we show that a TfR1-targeting ligand can modulate the biodistribution of ASOs *in vivo* in a beneficial manner. Addressing the root biological causes of several diseases that are driven by upregulation or downregulation of RNA expression could lead to reversal of the disease progression rather than managing symptoms. However, because of the high concentration of receptors that facilitate the uptake of ASOs in vascularized tissue such as liver, kidney, and spleen, targeted delivery of ASO-based therapeutics remains a challenge.[90] This proof-of-concept study illustrates that a simple receptor-targeting peptide has the capacity to modulate the activity of a conjugated ASO cargo toward miRNA expression in cardiac tissue.

This small linear peptide and its dimer perform on a comparable level with much larger, more complex targeting agents. The size and structural simplicity of T12 and the T12 dimer is reliably accessed with flow synthesis in hours,[86] offering a more rapid and straightforward production method relative to recombinant expression and other bioengineering procedures required for acquiring antibodies and other complex macromolecular targeting agents. Protein and antibody-based targeting agents typically require doses in the range of 1-30 mg/kg for efficacy.[78,79,83] The best reported efficacy for a Fab-PMO conjugate is achieved at 30 mg/kg of ASO equivalent, which corresponds to nearly 70 mg/kg of total compound,[79] a large amount of material required for human patients. Our PNA-T12 conjugate demonstrated about 50% efficacy in cardiac tissue at 30 mg/kg total compound, and our PNA-T12 dimer conjugate demonstrated about 60% efficacy at only 5 mg/kg total compound. Thus, T12 alters the biodistribution of ASOs at a competitive level with other larger targeting modalities.

Looking ahead, we seek to find the minimum dose required for the T12 dimer construct to remain efficacious and to investigate dimer architecture and other strategies to further increase the conjugate's efficacy. We hope that trough dose reduction we will achieve a reduction in trafficking to liver and kidney tissue. One other approach that can be used on its own or in conjunction with the dimer strategy is to co-deliver targeted cell-penetrating peptides (CPPs) that could aid in the entry of the ASO therapeutics into cells.[93]

Current advances in the field of ASO targeting are expanding the viability of RNA interference (RNAi) therapeutics. As the number of target miRNA sequences in the progression of various diseases grows, ASOs and other gene therapies are becoming a key area of interest.[7,55,60–62] Several ongoing efforts as well as this work demonstrate that active targeting can be used to beneficially direct where ASOs or other RNAi therapeutics are delivered in the body.[78–80,83,84] The development of rapid flow synthesis of peptides, ASOs, and their conjugates will allow for straightforward manufacturing of many candidate compounds on a short time scale not achievable by biological methods for larger, more complex conjugates.[68,86] These efforts have shown encouraging progress in achieving extrahepatic targeted delivery of gene therapies. As targeting strategies improve and off-target delivery is reduced, ASOs can become much more prevalent and widely available treatments for diseases.

We hope that this system can act as a proof of concept that can be applicable for more cardiac-specific proteins and their respective peptide or peptidomimetic binders. Moving forward, when the system is further understood, we would like to demonstrate efficacy with other backbone structures of ASOs such as LNAs and PMOs, and other

gene targets implicated in CVD or other diseases. Additionally, we believe that combining this and the co-delivery strategy can result in low doses competitive with those in current clinical development.

2.4 Materials and Methods

2.4.1 General information

Throughout this work, no unexpected or unusually high safety hazards were encountered.

Synthesis reagents

H-Rink Amide (0.50 mmol/g loading or 0.18 mmol/g) resin for peptide and PNA synthesis were purchased from PCAS Biomatrix. All the fluorenylmethyloxycarbonyl (Fmoc) protected amino acids were purchased from Novabiochem-line from Sigma Millipore and used as received. All the Fmoc-protected PNA monomers (moA, moC, moG, and moT) were purchased from PNA bio; O-(7-azabenzotriazol-1-yl)-*N*,*N*,*N*',*N*'-tetramethyluronium hexafluorophosphate (HATU, ≥97.0%) was purchased from P3 Biosystems; *N,N,N',N'*-tetramethyl-O-(1H-benzotriazol-1-yl)uronium hexafluoro-phosphate (HBTU, ≥97.0%) was purchased from P3 Biosystems; Diisopropylethylamine (DIEA; 99.5%, biotech grade, catalog number 387649), piperidine (≥99.0%), piperazine (≥99.0%), morpholine (≥99.5%), trifluoroacetic acid (HPLC grade, ≥99.0%), triisopropylsilane (≥98.0%), formic acid (FA, ≥95.0%), and 1,2-ethanedithiol (EDT, GC grade, ≥98.0%) were purchased from Sigma-Aldrich. Diisopropylethylamine (DIEA; 99.5%, biotech grade, catalog number 387649), piperidine (≥99.0%) and formic acid (FA, ≥95.0%) were purchased from Sigma-Aldrich. *N*, *N*-dimethylformamide (DMF,

Biosynthesis OmniSolv® grade) was purchased from EMD Millipore (DX1732-1); AldraAmine trapping agents (for 1000~4000 mL DMF, catalog number Z511706) were purchased from Sigma-Aldrich.

Cleavage reagents

Trifluoroacetic acid (TFA, HPLC grade, ≥99.0%), triisopropylsilane (TIPS, ≥98.0%), and 1,2-ethanedithiol (EDT, GC grade, ≥98.0%) were purchased from Sigma-Aldrich.

Analysis and purification reagents

Water for HPLC was purified to 18.2MΩ/cm resistivity on a Millipore Milli-Q system. Acetonitrile (HPLC-grade) was purchased from VWR International (Philadelphia, PA) and acetonitrile (LC-MS grade) was purchased from Sigma-Aldrich (St. Louis, MO).

Cell assay reagents

Claycomb medium, (+/-)-norepinephrine (+)-bitartrate salt, L-ascorbic acid, L-glutamine solution bioxtra, trypsin-EDTA solution, trypsin inhibitor, fibronectin from bovine plasma, and gelatin from bovine skin type B were all purchased from Sigma-Aldrich; fetal bovine serum was purchased from Thermo Fisher Scientific.

Animal experiments

Wild-type male mice (strain C57BL/6) were ordered from Charles River; Dulbecco's phosphate buffered saline was purchased from Sigma-Aldrich; TRIzol reagent, TaqMan

universal PCR master mix, and TaqMan MicroRNA assay kits were all purchased from Invitrogen; Anti-MiR21 LNA molecule was ordered from Integrated DNA Technologies.

Declaration of approval for animal experiments

Experiments with mice completed under this study were approved by the Committee on Animal Care (CAC) at MIT under protocol number 0821-058-24.

2.4.2 Procedure for automated PNA/PNA-peptide synthesis

Unmodified PNAs and peptide-conjugated PNAs were synthesized on an automated synthesizer named Tiny Tides[68] following our reported protocols.[86] We redescribe the procedure below for the reader:

Unmodified PNA synthesis

The automated synthesis of PNAs was performed using the self-designed oligonucleotide synthesizer (Tiny Tides). Rink amide resin (15 mg, 0.5 mmol/g loading) was loaded into the reactor. The reactor was connected to the reactor head and heated to 70 ˚C. DMF was delivered at 5 mL/min (2.5 mL/min per pump) for 20 seconds to remove air. The flow was stopped and the resin was allowed to swell at 70 ˚C for 5 minutes. The flow protocol was started with an initial DMF wash at 5 mL/min (2.5 mL/min per pump) 70 ˚C for 40 seconds, then coupling solution composed of one-part 0.2 M PNA monomer subunit in DMF, one-part 0.18 M HBTU in DMF, and one-part 10% DIEA (v/v) in DMF was delivered for 10 seconds (10 eq PNA monomer) at 70 ˚C. Next, DMF was delivered at 5 mL/min (2.5 mL/min per pump) under 70 ˚C for 20 seconds to ensure all the monomer solutions arrived at the reactor and clean the loop. The 6-

position valve then switched to the room-temperature loop. DMF was delivered at 5 mL/min (2.5 mL/min per pump) for 40 seconds at room-temperature. Cold DMF flow (rt) mixed with the hot reactor (70 ˚C) generated an in-situ 40 ˚C environment for deprotection. Deprotection was performed with one-part 40% piperidine, 2% formic acid (v/v) in DMF, and one-part DMF for 50 seconds in the room-temperature loop. After a 20-second room temperature DMF wash, the 6-position valve was switched to the 70 ˚C loop. DMF was delivered at 5 mL/min for 40 seconds to wash the resin and preheat the reactor. No capping or multiple couplings were needed, and each single coupling cycle took 3 minutes. Repeat the synthesis cycle to finish the long PNA chain assembly.

PNA-T12 monomer and dimer synthesis

For single-shot T12 conjugated PNA synthesis, T12 was pre-synthesized on the H-Rink Amide resin (0.49 mmol/g) with an automated peptide synthesizer developed in our lab.[50,52,53] The T12-bound resin was used directly for automated PNA-T12 synthesis on Tiny Tides. Both $(Lys)_3$ linker and PNA sequences were synthesized directly on Tiny Tides. Similarly, for the PNA-T12 dimer conjugate, T12-Lys_3-T12 was pre-synthesized on H-Rink Amide resin (0.49 mmol/g) with our automated peptide synthesizer.[50,52,53] This peptide-bound resin was transferred to Tiny Tides for the synthesis of PNA-Lys_3.

2.4.3 Cleavage protocols

After synthesis, the PNA bound resin was washed with dichloromethane (3 x 5 mL), dried in a vacuum chamber, and weighed. Then the resin was transferred into a 15 mL conical polypropylene tube. Approximately 1 mL of cleavage solution (94% TFA, 1% TIPS, 2.5% EDT, 2.5% water) was added to the tube and kept at room temperature

for 2 hours. After cleavage, the resin was removed by filtration, the filtrate was concentrated under a stream of nitrogen and the PNA product was precipitated in dry ice-cold diethyl ether (12 mL) by centrifugation and washed three times. The supernatant was discarded, and the residual was dissolved in 50% acetonitrile in water with 0.1% TFA. The PNA solution was filtered with a Nylon 0.22 μm syringe filter, frozen with liquid nitrogen, and lyophilized to dried powder. Finally, the crude PNA was weighed.[44]

2.4.4 Purification protocols

Method A:

HPLC purification was carried out on a reversed-phase preparative HPLC using an Agilent mass directed purification system (1260 Infinity LC and 6130 single quad MS) with UV detection at 260 nm. Column: Agilent Zorbax SB-C3 (9.4 x 250 mm, 5 μm). Flow rate 4.0 mL/min; Temperature: 60 ˚C. Solvent System: A linear gradient of acetonitrile with a 0.1% TFA additive (solvent B) in water with a 0.1% TFA additive (solvent A) was used. Gradient: 5 min hold 1% B, 1-31% B gradient from 5 to 105 min, 31-65% B gradient from 105 to 120 min, hold 65% B from 120 to 125 min. A final 5.5 min hold was performed with 1% B. The method in total lasted 125 min. Fractions were collected every minute.

Method B:

Reverse phase purification was carried out on a Biotage Selekt flash purification system with UV detection at 280 and 214 nm. Column: Biotage Sfär C18 (12g column

size, particle size 20 μM, pore size 300 Å, column volume (CV) 17 mL); Flow rate 12 mL/min; Temperature: r.t. Solvent System: A linear gradient of acetonitrile with a 0.1% TFA additive (solvent B) in water with a 0.1% TFA additive (solvent A) was used. Gradient: 1 CV hold 5% B, 5-15% B gradient over 1 CV, 15-45% B gradient over 10 CV, 45-95% B gradient over 0.5 CV, hold 95% B for 2 CV, return to 5% B over 0.1 CV, and hold 5% B for 1 CV. Fractions were collected by UV peak detection, with thresholds of 15 mAu at 280 nm and 50 mAu at 214 nm.[44]

Method C:

HPLC purification was carried out on a reversed-phase preparative HPLC using an Agilent mass directed purification system (1260 Infinity LC and 6130 single quad MS) with UV detection at 260 nm. Column: Agilent Zorbax SB-C3 (9.4 x 250 mm, 5 μm). Flow rate 4.0 mL/min; Temperature: 60 ˚C. Solvent System: A linear gradient of acetonitrile with a 0.1% TFA additive (solvent B) in water with a 0.1% TFA additive (solvent A) was used. Gradient: 5 min hold 1% B, 1-40% B gradient from 5 to 105 min, 40-65% B gradient from 105 to 120 min, hold 65% B from 120 to 125 min. A final 5.5 min hold was performed with 1% B. The method in total lasted 125 min. Fractions were collected every minute.

Method D:

HPLC purification was carried out on a reversed-phase preparative HPLC using an Agilent mass directed purification system (1260 Infinity LC and 6130 single quad MS) with UV detection at 260 nm. Column: Agilent Zorbax SB-C3 (9.4 x 250 mm, 5 μm).

Flow rate 4.0 mL/min; Temperature: 60 ˚C. Solvent System: A linear gradient of acetonitrile with a 0.1% TFA additive (solvent B) in water with a 0.1% TFA additive (solvent A) was used. Gradient: 5 min hold 5% B, 5-25% B gradient from 5 to 15 min, 25-32.5% B gradient from 15-40 min, 32.5-95% B gradient from 40-45 min, hold 95% B from 45-50 min. The method in total lasted 50 min. Fractions were collected every minute.

2.4.5 LC-MS analysis

Analysis was performed on an Agilent 1290 Infinity HPLC coupled to an Agilent 6550 Q-TOF with Dual Jet Stream ESI ionization and iFunnel. MS was run in positive ionization mode, extended dynamic range (2 GHz), and low mass range (m/z in range 100 to 1700). All peptide and PNA solutions were filtered and then diluted to approximately 0.1 mg/mL before sample loading. Buffer A: 0.1% formic acid in H_2O. Buffer B: 0.1% formic acid in acetonitrile.

The following LC-MS methods were used at different situation:

Method A:

Column: Phenomenex Aeris C4 (2) (3.6 um, 2.1 x 150 mm, 100 Å silica); Flow Rate: 0.2 mL/min; Gradient: 1% B 0-1.50 min, linearly ramp from 1% B to 61% B 1.50 to 6.50 min, hold 90% B from 6.51 to 8 min. Post time is 1% B for 5 min.

Method B:

Column: Phenomenex Aeris C4 (2) (3.6 um, 2.1 x 150 mm, 100 Å silica); Flow Rate: 0.2 mL/min; Gradient: 1% B 0-2 min, linearly ramp from 1% B to 61% B 1.50 to 10 min, hold 90% B from 10.1 to 12 min. Post time is 1% B for 5 min.

Method C:

Column: Agilent ZORBAX 300SB C3 (5 um, 2.1 x 150 mm); Flow Rate: 0.5 mL/min; Gradient: 1% B 0-2 min, linearly ramp from 1% B to 91% B 2 to 12 min. Post time is 1% B for 3 min.

Method D:

Column: Agilent ZORBAX 300SB C3 (5 um, 2.1 x 150 mm); Flow Rate: 0.5 mL/min; Gradient: 1% B 0-2 min, linearly ramp from 1% B to 61% B 2 to 7 min. Post time is 1% B for 1 min.

2.4.6 Biotinylated cardiac targeting ligand binding affinity study

Peptide binding validation was carried out using bio-layer interferometry (BLI) on a Gator Bio GatorPlus system. The plate agitation speed was kept at 1,000 rpm and the temperature was maintained constantly at 30 °C. Streptavidin (SA)-coated biosensor tips were used to load N-terminal biotinylated cardiac targeting ligand (CTL) T12 (sequence: THRPPMWSPVWP) or T12 dimer (sequence: THRPPMWSPVWPKKKTHRPPMWSPVWP) which was dissolved at 1.5 μM in a kinetic buffer (K.B.): 1 × PBS with 0.1% BSA and 0.02% tween20. After loading the peptide for 120 s, the biosensor tips were then moved into solutions containing various concentrations (15.625 nM, 31.25 nM, 62.5 nM, 125 nM, 250 nM, and 500 nM) of recombinant transferrin receptor (TfR1/CD71) protein (purchased from Sino Biological) in the K.B. to obtain the association curve. After a 300-s association step, the tips were moved back into the K.B. for 500 s to collect the dissociation curve. Peptide-only (1.5 μM) and protein-only (500 nM) conditions were used as references for background subtraction. The association and dissociation curves were analyzed in the Gator Bio

GatorPlus Data Analysis Software with defined parameters (linked global kinetic fitting algorithm, binding model 1:1) to calculate the apparent dissociation constant (K_D). The analysis results were shown in Figure 2.1, the K_D value is 26 nM (reported K_D = 15 nM by Scatchard plot from phage titer).[85]

2.4.7 Gel shift assay

Synthetic purified PNA-T12 was pre-mixed with 6 μM pre-miRNAA-21 or pre-miRNA-34a (ordered from IDT) at different concentrations (0, 0.6, 1.2, 3.0, 6.0 and 12 μM) for 1 hour at 37 °C. Subsequently, a 5% agarose gel was used for gel band separation and analysis, the gel running condition was 100 V for 70 min in 1×TBE buffer. The gel was imaged with SYBR gold nucleic acid stain purchased from Thermo Fisher.

2.4.8 Bioactivity study of anti-miRNA PNAs in HL-1 cells

A cardiac muscle cell line derived from AT-1 mouse atrial cardiomyocyte tumor lineage, named HL-1, was cultured at 37 °C and 5% CO_2 in complete Claycomb medium until 90% confluency. Purified PNA-T12 monomer and dimer constructs with serial concentrations were added into adherent HL-1 cells. After 24-36 hours inoculation, the total RNA was extracted from treated cells using TRIzol reagent following manufacture's instruction (also specified below). The TaqMan microRNA assay kit was used for quantitative real-time polymerase chain reaction (qRT-PCR). After data acquisition and cycle threshold calculation, the expression levels of relevant miRNAs were calculated using a comparative C_T method (in comparison with a PBS-

treated control after being internally normalized to a housing keeping microRNA, MiR-16).

Extract RNA

Add 500 µL TRIzol Reagent to cell pellets, then incubate for 5 min at room temperature to allow complete disruption of the cell membrane. Add 200 µL of chloroform per 1 mL of TRIzol reagent used for lysis, then securely cap the tube, and thoroughly mix by shaking. Incubate for 3 minutes. Centrifuge the sample for 15 minutes at 12,000 g at 4 ˚C. Transfer the upper colorless aqueous phase containing the RNA to a new tube carefully with a micropipette.

Isolate RNA

Add 500 µL of isopropanol to the aqueous phase, incubate for 10 min at 4 ˚C. Centrifuge the solution for 10 minutes at 12,000 g at 4 ˚C. Note: the total RNA precipitate forms a white gel-like pellet at the bottom of the tube. Discard the supernatant with a micropipette.

Wash RNA

Resuspend the pellet in 1 mL of 75% ethanol, and vortex the sample briefly. Then, centrifuge the sample for 5 minutes at 7500 g at 4 ˚C. Discard the supernatant with a micropipette. Air dry the RNA pellet for 10-30 minutes. Note: Do not let the pellet very dry to make sure the solubility of RNA.

Reverse transcription (RT)

TaqMan™ MicroRNA Reverse Transcription Kit was used for reverse transcription. The concentration of each RNA sample was measured using a BioTek Epoch microplate spectrophotometer. Prepare 50 ng/μL RNA solution according to the UV concentration. To a 96-well plate, add 0.1 μL dNTP substrate (100 mM wdTTP), 0.5 μL reverse transcriptase, 1 μL RT buffer (10X), 0.14 μL RNase inhibitor (20 U/μL), 5.25 μL nuclease-free water, 2 μL primer, and 1 μL RNA sample to each well. MiRNA-16 was used as an internal control. Perform the reverse transcription on a regular PCR cycler as follows: 16 °C for 30mins, 42 °C for 30mins, 85 °C for 5 min, and then chilled at 4 °C until further manipulation.

qRT-PCR

The qRT-PCR experiment was performed using the TaqMan microRNA assay kit and the PCR 2X super mix solution with three technical repeats on a Roche light cycler 480 instrument at MIT Proteomics Core Facility. The qRT-PCR running condition is: 95˚C – 10s, 95˚C – 15s, 60˚C – 60s, and repeat for 50 cycles. The fold change relative to control was calculated using the comparative C_T method, and equation was specified as follows:[89]

Fold change = $2^{-[\Delta CPsample - \Delta CPcontrol]}$

2.4.9 T12-conjugated PNA efficacy study in mice

Wild-type male mice were reared under a 12-hour light and 12-hour dark cycle and housed at the facility within MIT Division of Comparative Medicine (DCM). To study

the PNA-T12 *in vivo* efficacy, PNA-T12 was administered into wild-type mice via retro-orbital injection. Two experimental groups of mice (n=5) were treated with 30 mg/kg and 15 mg/kg PNA-T12, and two control groups of mice (n=5) were injected with equivalent volume of 1× PBS and 30 mg/kg PNA only. At day 14 post-injection, the mice were euthanized and sacrificed using approved protocols and in compliance with the MIT Committee on Animal Care (CAC) guidelines. To measure the efficacy of PNA-T12 injections, five mouse organs (heart, liver, kidney, liver, and spleen) from the mice were dissected, collected, and preserved in TRIzol Reagent for subsequent analysis.

In a separate experiment, PNA-T12 dimer was administered into wild-type mice reared as described above via retro-orbital injection. Three experimental groups of mice (n=3) were treated with 30, 15, or 5 mg/kg PNA-T12 dimer, and two control groups of mice (n=3) were injected with equivalent volume of 1× PBS and 30 mg/kg PNA-T12 monomer. 7 days post-injection, the mice were euthanized and sacrificed using approved protocols and in compliance with the MIT Committee on Animal Care (CAC) guidelines. The same five organs were collected and preserved in TRIzol Reagent as in the first experiment.

For both rounds of *in vivo* experiments, the harvested organs were homogenized in TRIzol using a Dounce homogenizer and incubated for 5 minutes to allow complete dissociation of the nucleoprotein complexes. RNA extraction, reverse transcription and qRT-PCR were performed similarly following the procedures outlined in Section 4.4.8.

2.4.10 Copper-free click reaction

For copper-free click reactions between azidolysine on a peptide and DBCO on an ASO or biotin, both components were dissolved at 5 mM. Generally, peptides were dissolved in MilliQ water and DBCO compounds were dissolved in dimethylsulfoxide (DMSO). At the scale of the reactions completed, this was less than 200 µL total volume. The two components were then combined 1:1 to a final concentration of 2.5 mM each reactant and allowed to react at room temperature for at least 2 hours with occasional vortexing. LC-MS was taken at the start time as well as after 2 hours (Section 2.4.5 Method D). Upon analysis of LC-MS traces, amounts of starting materials were adjusted until >90% conversion was achieved. Once the reaction was complete, it was diluted with 900 µL 50% acetonitrile in water and freeze dried overnight on the lyophilizer.

2.5 Acknowledgements

Financial support for this work was provided by Novo Nordisk A/S (to B.L.P.). We acknowledge the Preclinical Modeling, Imaging and Testing Facility at the MIT Koch Institute's Robert A. Swanson (1969) Biotechnology Center for technical support in the handling and care of animals (NCI Core Center Grant P30-CA14051). We thank the Division of Comparative Medicine at MIT and in particular Sarah Elmiligy and Virginia Spanoudaki for their assistance with animal studies and training; Dr. Joseph S. Brown and Michael A. Lee for insightful discussions and suggestions; and Andrew Wilson for his help in the design and production of the automated flow instrument.

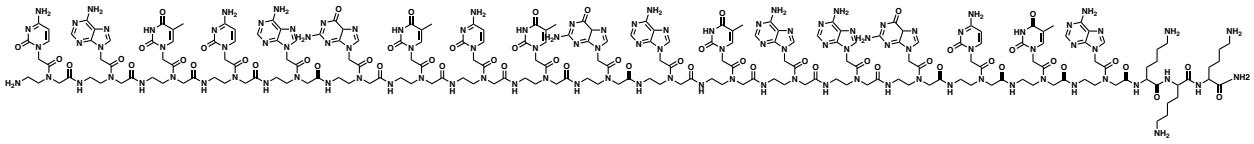
2.6 Author Contributions

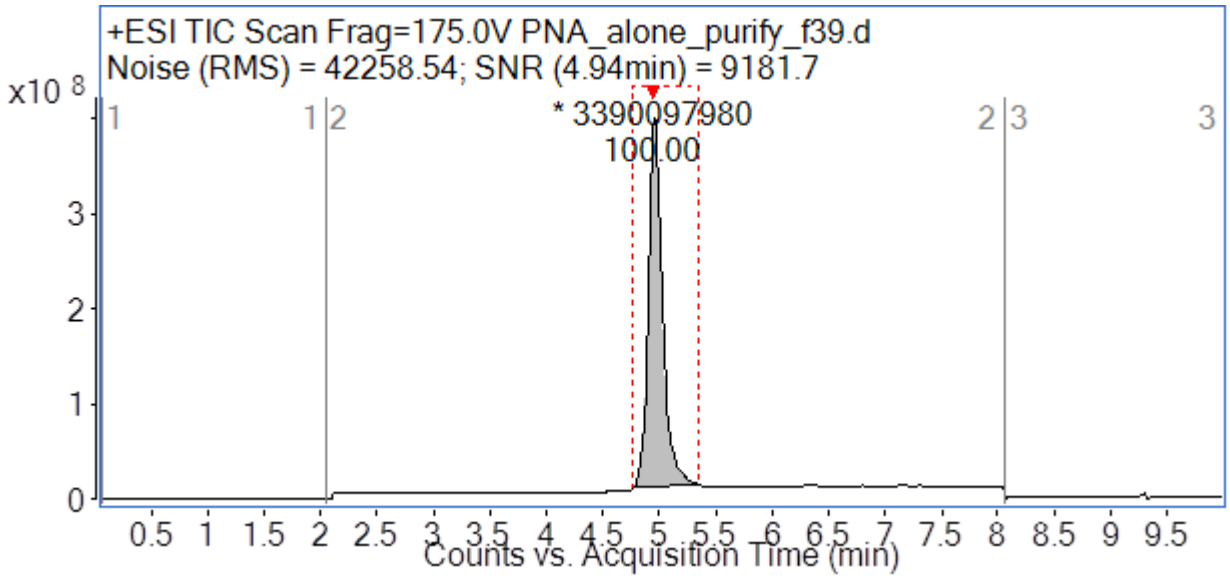
S.A., G.Z., C.L., B.L.P., A.L., and T.E.N. conceptualized the research; S.A., G.Z., and C.L. synthesized and purified relevant peptides and PNA conjugates, and performed BLI validations. S.A. developed the gel shift assay. S.A. and G.Z. performed tissue culture and *in vitro* and *in vivo* qRT-PCR assays. S.A., G.Z., and C.L. wrote the manuscript with input from all authors.

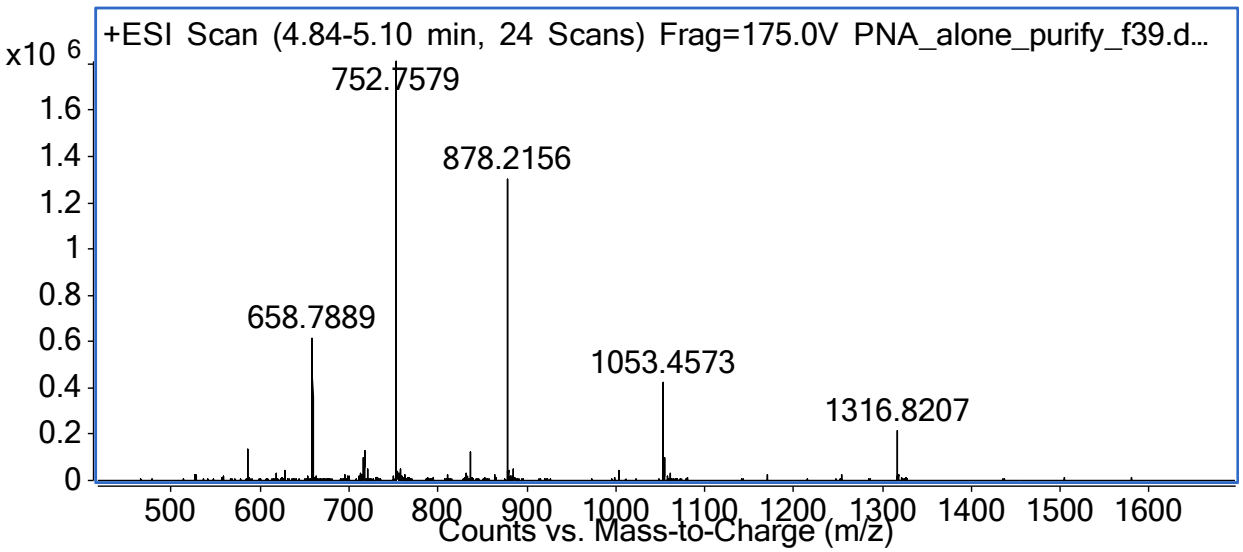
2.7 Appendix I: LC-MS Characterization

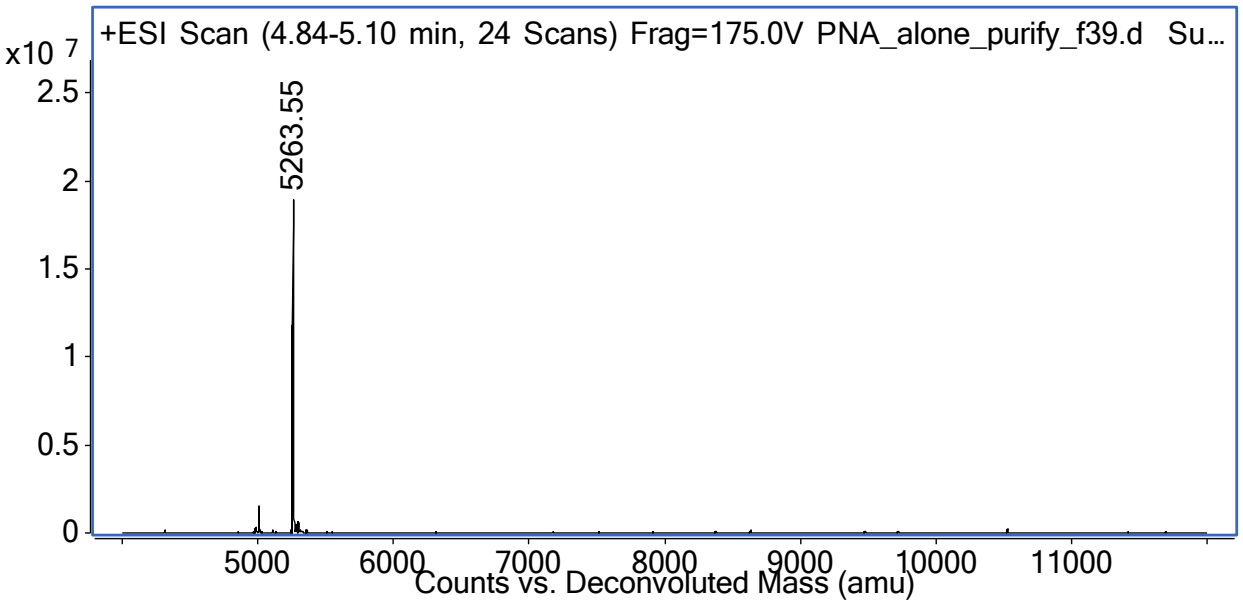
All the peptide and PNA samples were synthesized and purified according to the protocols in Section 2.4.2 and analyzed with LC-MS before use.

<u>Anti-MiRNA-21 PNA</u>
Sequence: CATCAGTCTGATAAGCTA-KKK-$CONH_2$.
Synthesis method: Automated flow synthesis.
Purification method: Section 2.4.4, method A
LC-MS method: Section 2.4.5, method A
Calculated: 5263.27 Da
Observed: 5263.55 Da

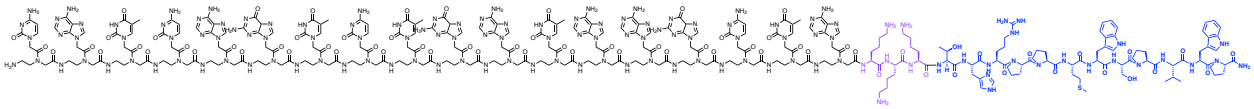


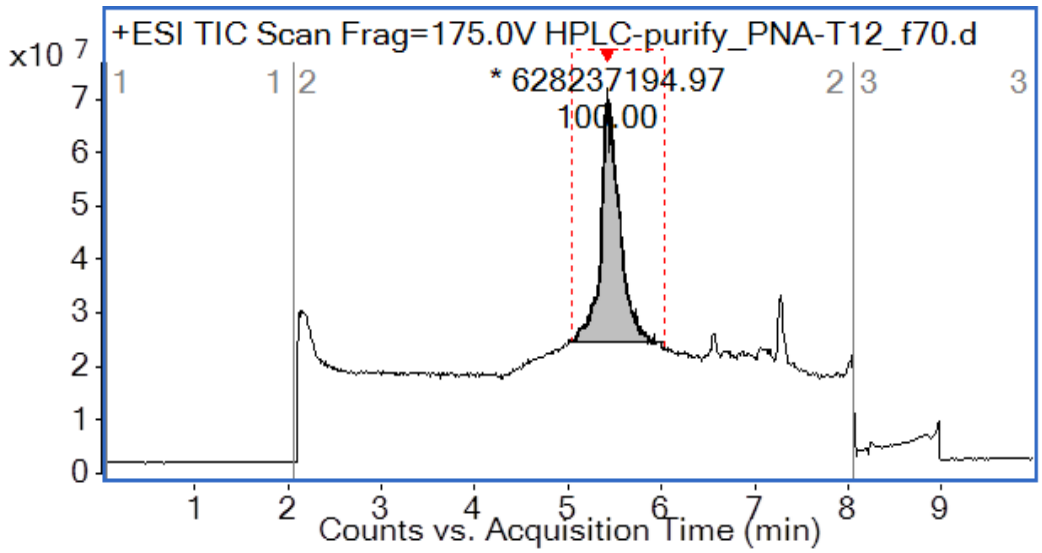


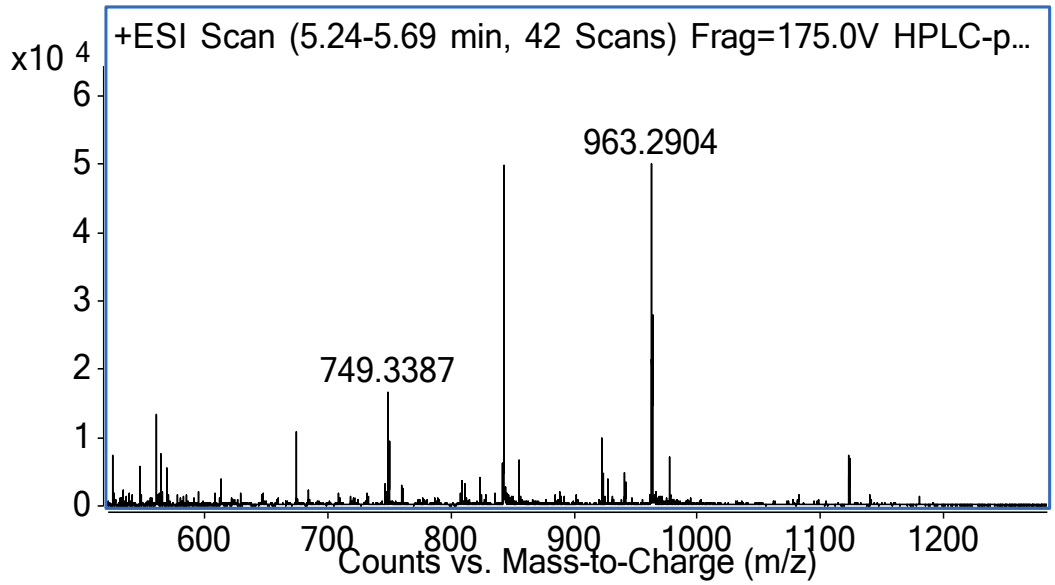

+ESI Scan (4.84-5.10 min, 24 Scans) Frag=175.0V PNA_alone_purify_f39.d...
x10 6
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
752.7579
878.2156
658.7889
1053.4573
1316.8207
500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600
Counts vs. Mass-to-Charge (m/z)

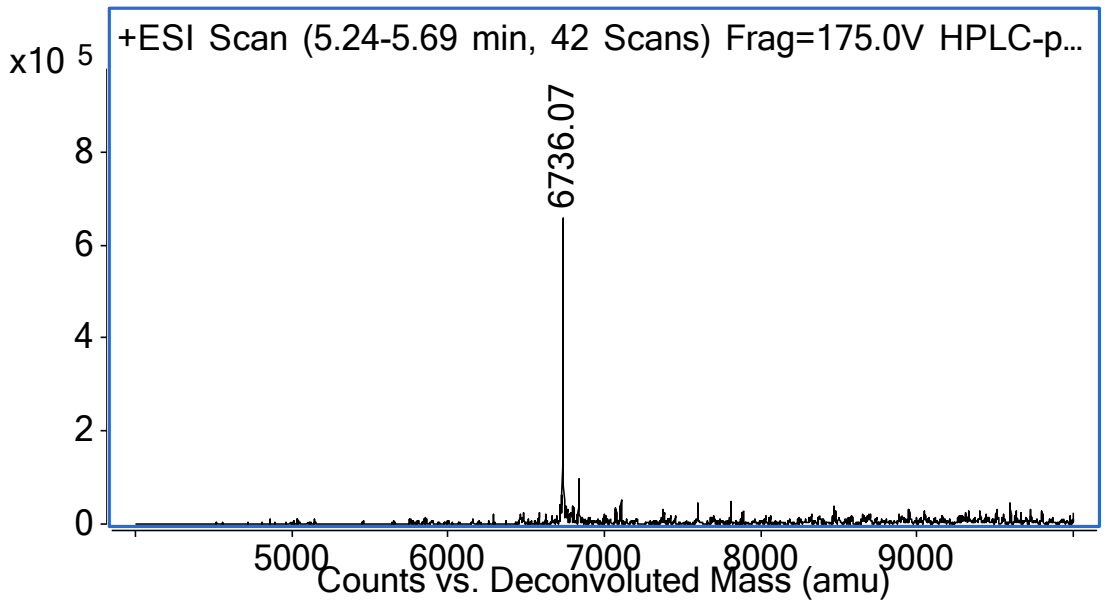

+ESI Scan (4.84-5.10 min, 24 Scans) Frag=175.0V PNA_alone_purify_f39.d Su...
x10 7
2.5
2
1.5
1
0.5
0
5263.55
5000 6000 7000 8000 9000 10000 11000
Counts vs. Deconvoluted Mass (amu)

Anti-MiRNA-21-T12
Sequence: CATCAGTCTGATAAGCTA-KKK-THRPPMWSPVWP-$CONH_2$.
Synthesis method: Automated flow synthesis.
Purification method: Section 2.4.4, method A
LC-MS method: Section 2.4.5, method B
Calculated: 6736.01 Da
Observed: 6736.07 Da

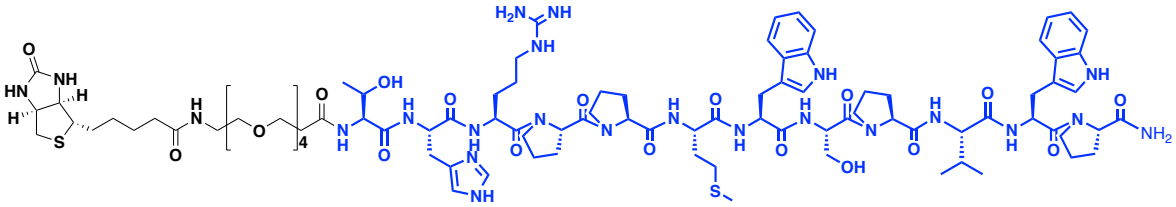


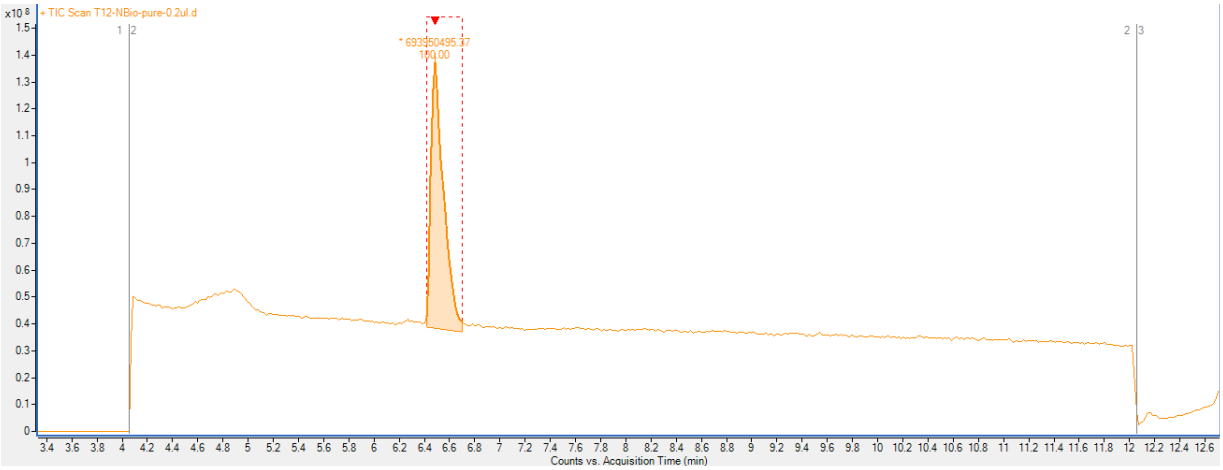


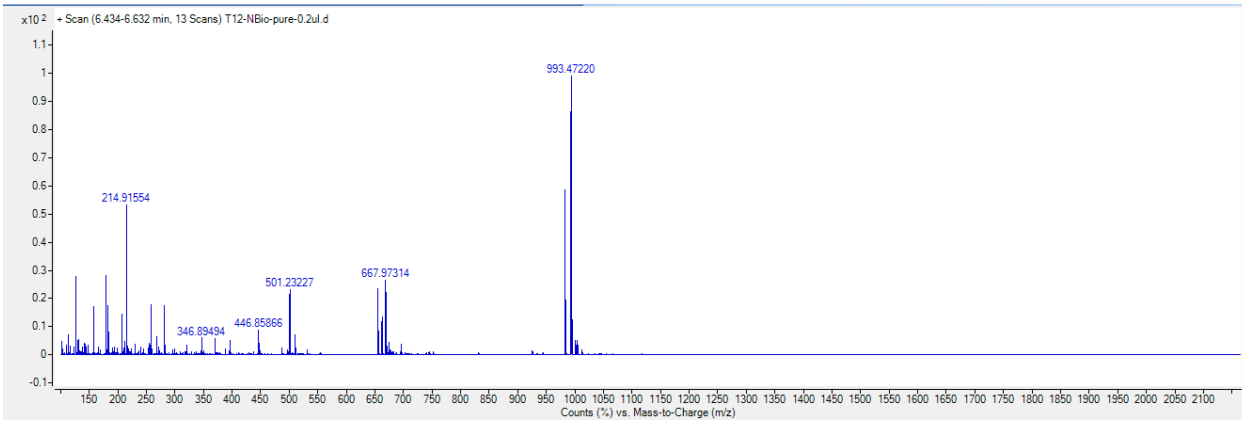




<u>biotin-T12</u>
Sequence: biotin-PEG$_4$-THRPPMWSPVWP-CONH$_2$.
Synthesis method: Automated flow synthesis.
Purification method: Section 2.4.4, method B
LC-MS method: Section 2.4.5, method C
Calculated: 1963.35 Da
Observed: 1962.95 Da

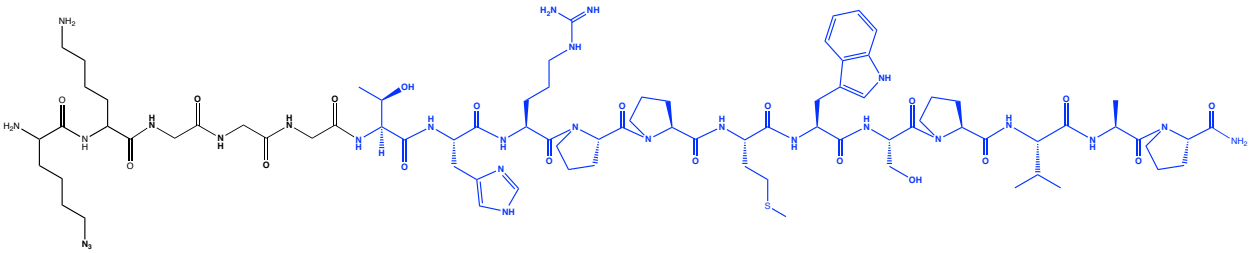


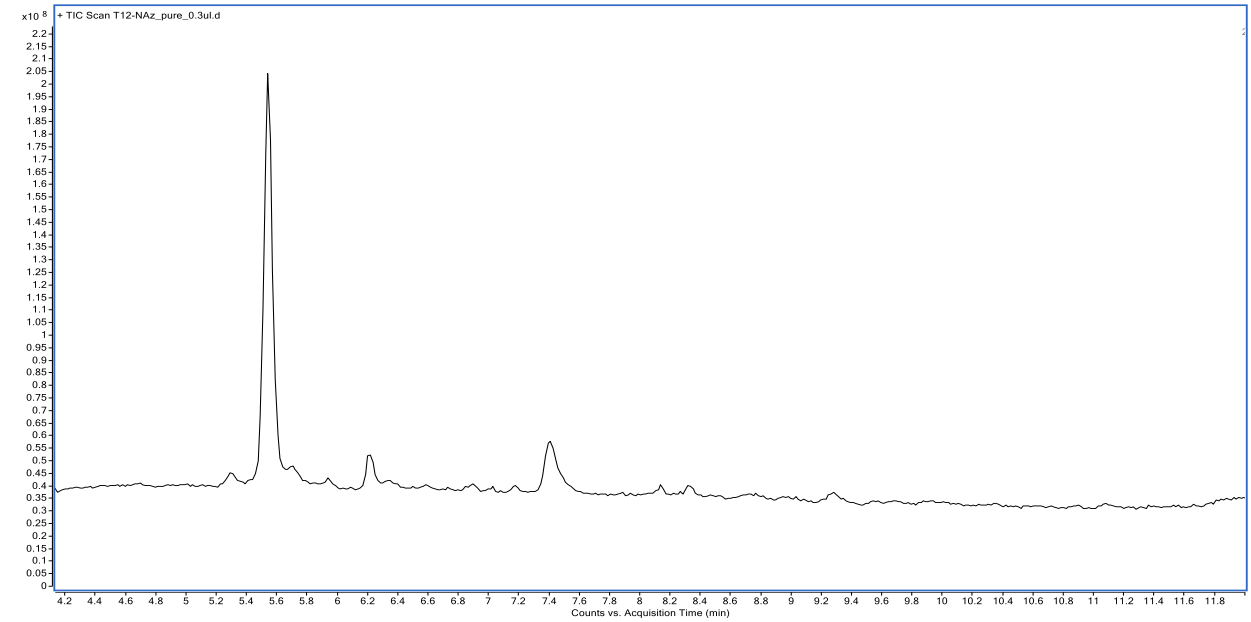


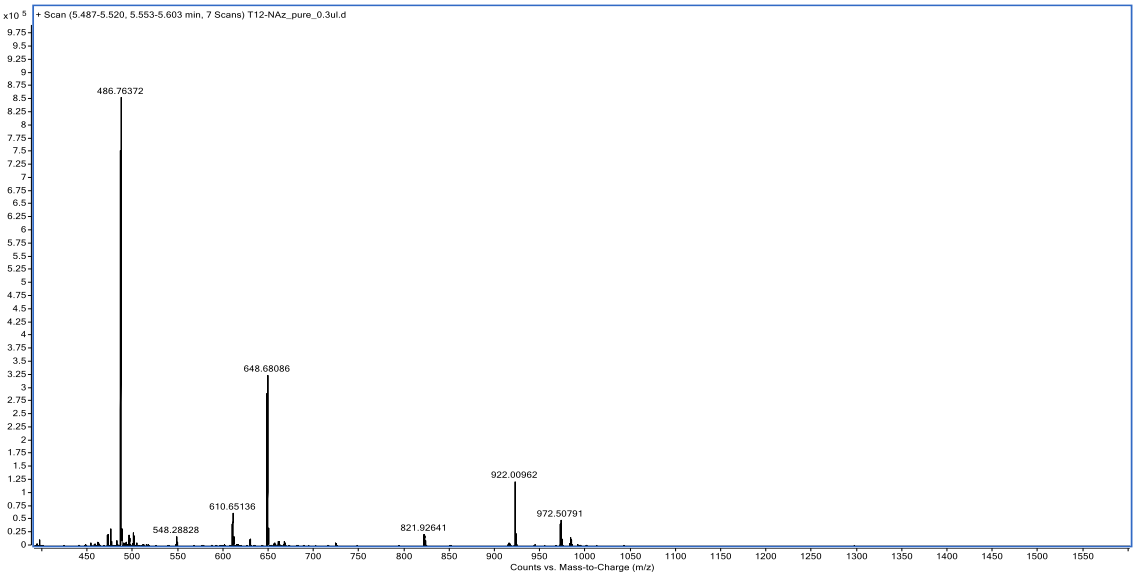


<u>T12-NAz</u>

Sequence: Lys(N_3)-KGGG-THRPPMWSPVWP-$CONH_2$.
Synthesis method: Automated flow synthesis.
Purification method: Section 2.4.4, method B
LC-MS method: Section 2.4.5, method C
Calculated: 1943.27 Da
Observed: 1943.11 Da

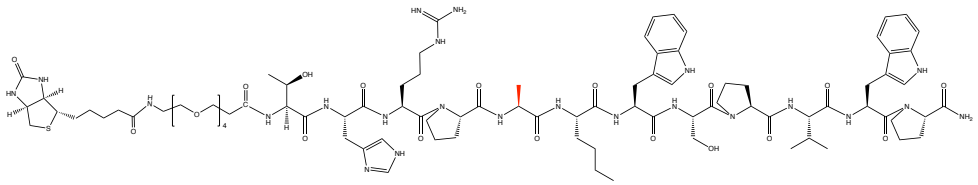


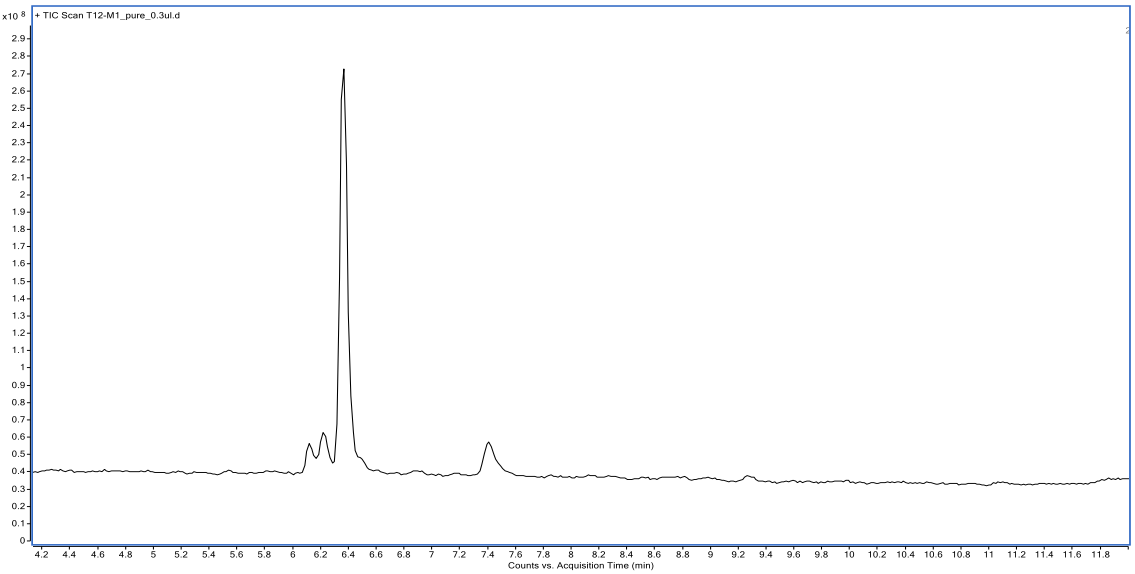


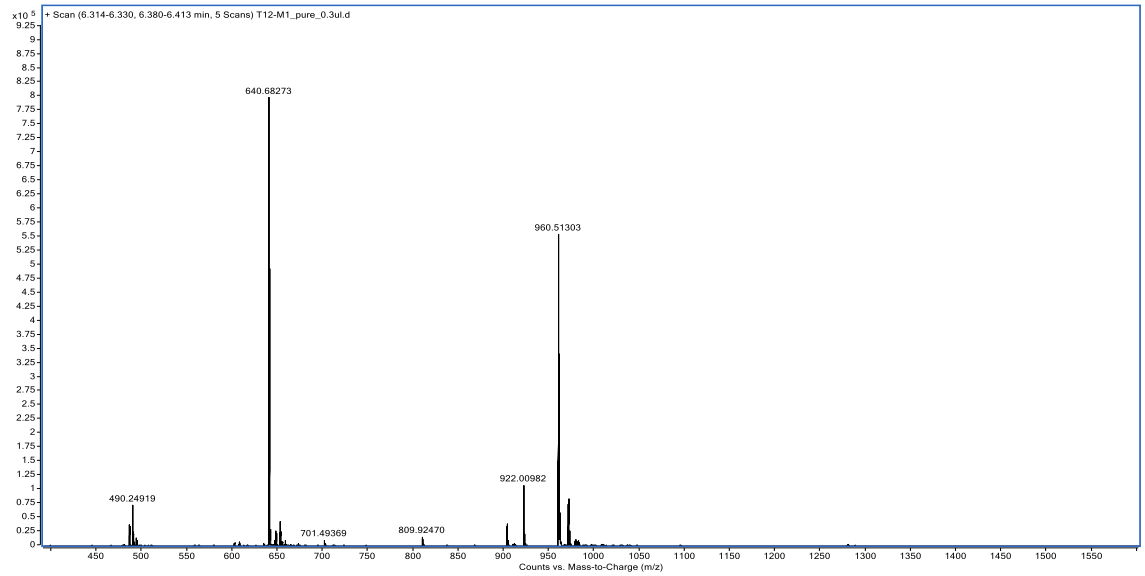


biotin-T12-P5A-M6Nle

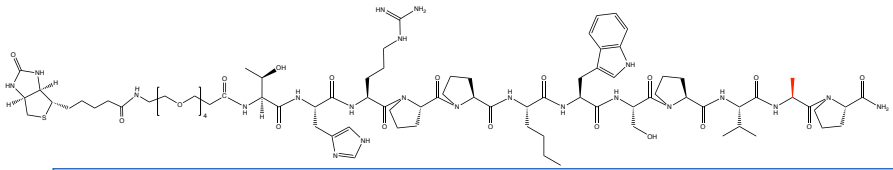
Sequence: biotin-PEG$_4$-THRPA(Nle)WSPVWP-CONH$_2$.
Synthesis method: Automated flow synthesis.
Purification method: Section 2.4.4, method B
LC-MS method: Section 2.4.5, method C
Calculated: 1919.29 Da
Observed: 1919.00 Da

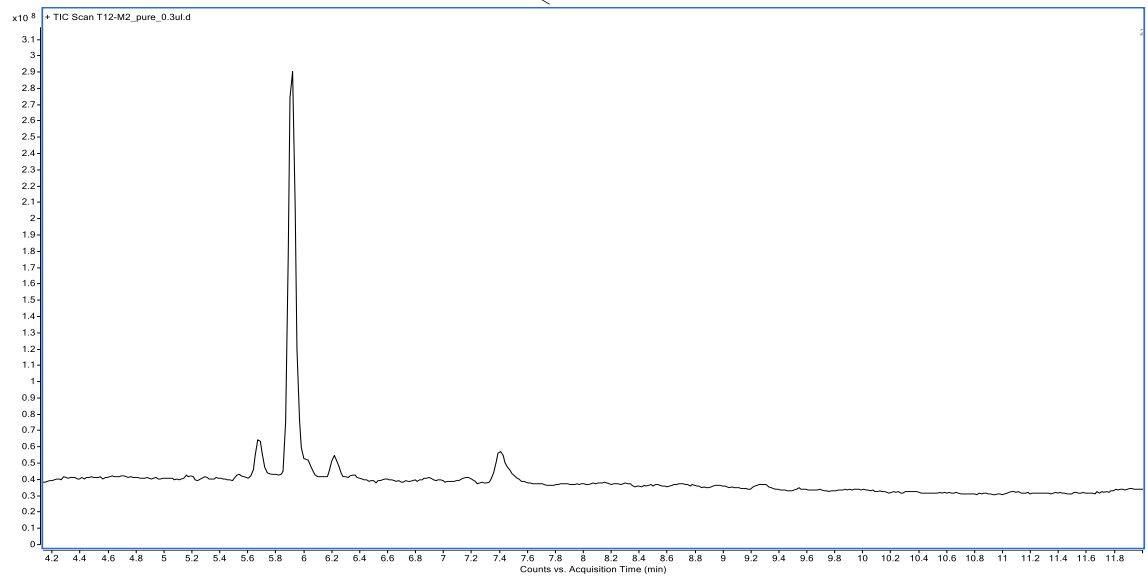


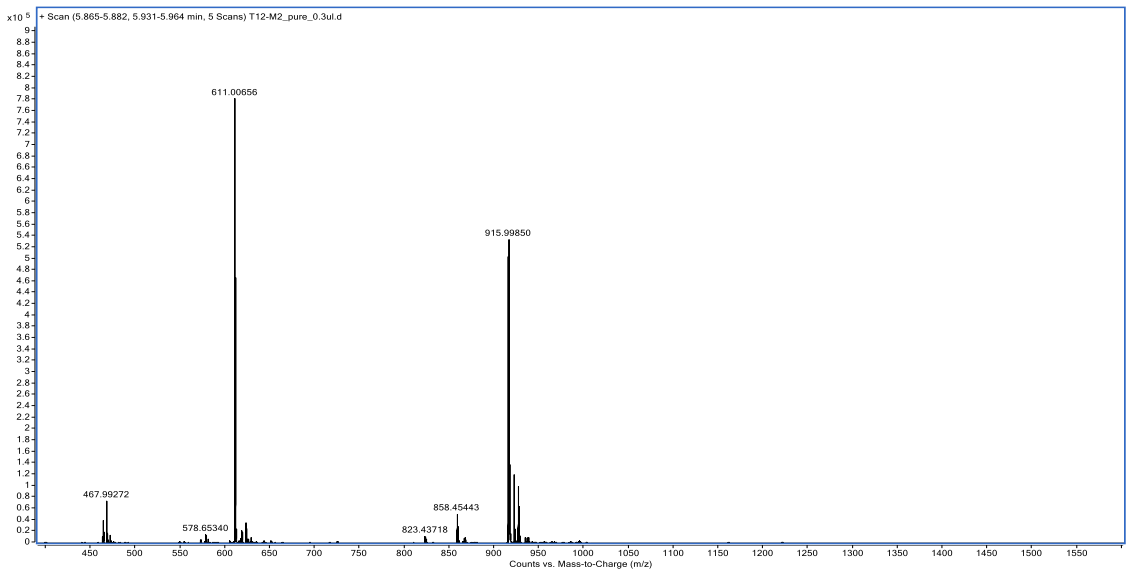




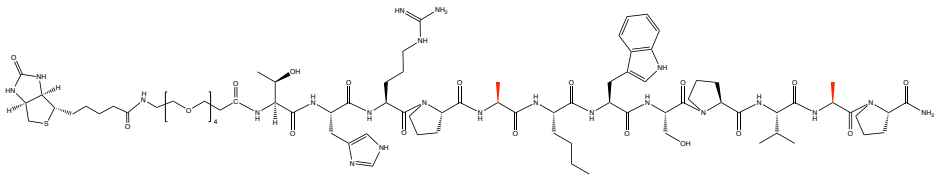
<u>biotin-T12-M6Nle-W11A</u>
Sequence: biotin-PEG$_4$-THRPP(Nle)WSPVAP-CONH$_2$.
Synthesis method: Automated flow synthesis.
Purification method: Section 2.4.4, method B
LC-MS method: Section 2.4.5, method C
Calculated: 1830.19 Da
Observed: 1829.97 Da

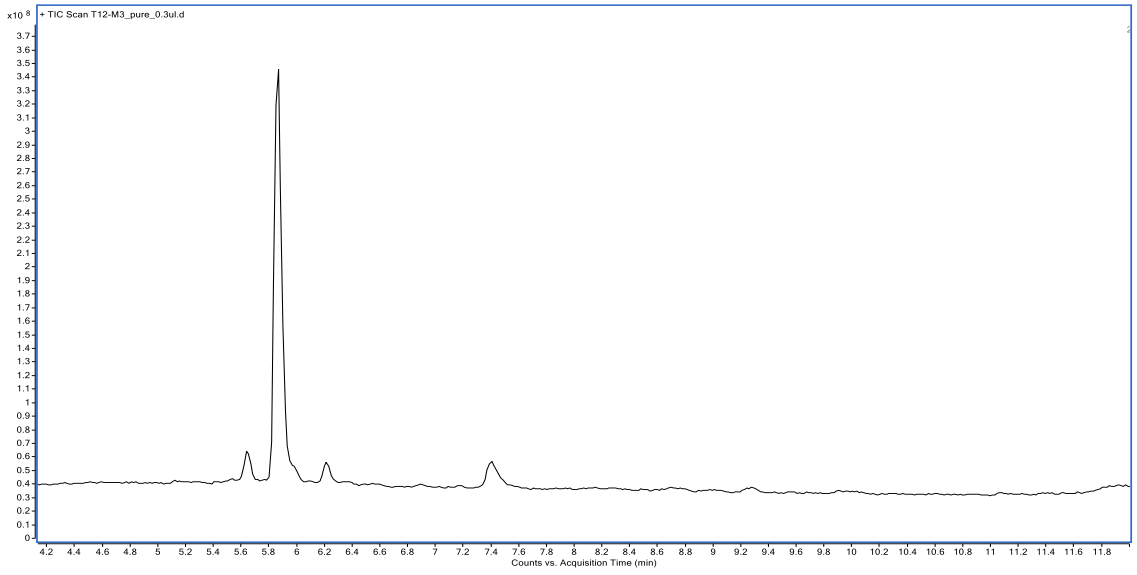


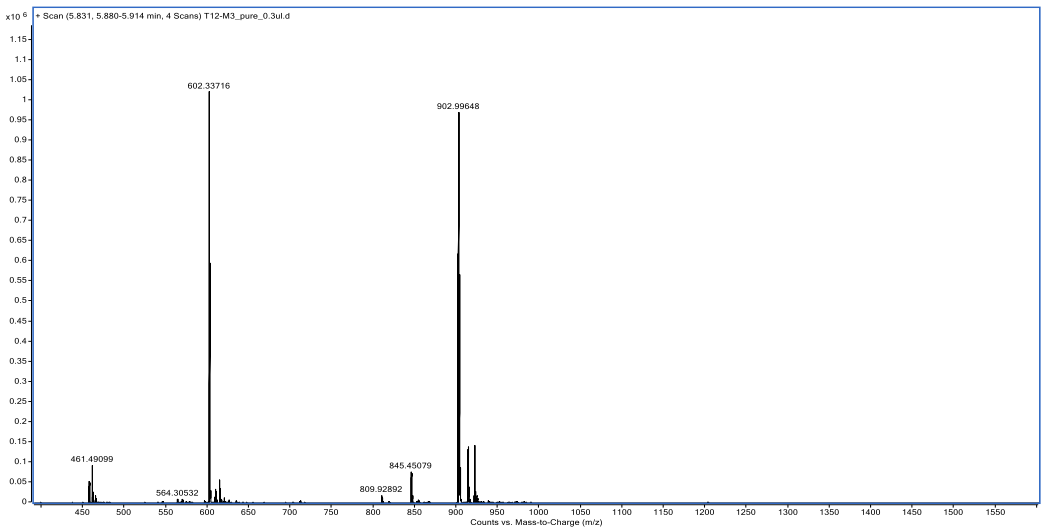




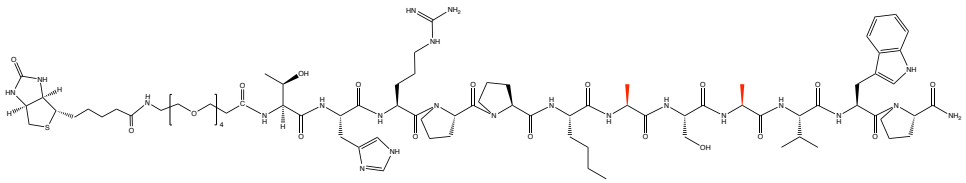
<u>biotin-T12-P5A-M6Nle-W11A</u>
Sequence: biotin-PEG$_4$-THRPA(Nle)WSPVAP-CONH$_2$.
Synthesis method: Automated flow synthesis.
Purification method: Section 2.4.4, method B
LC-MS method: Section 2.4.5, method C
Calculated: 1804.15 Da
Observed: 1804.05 Da

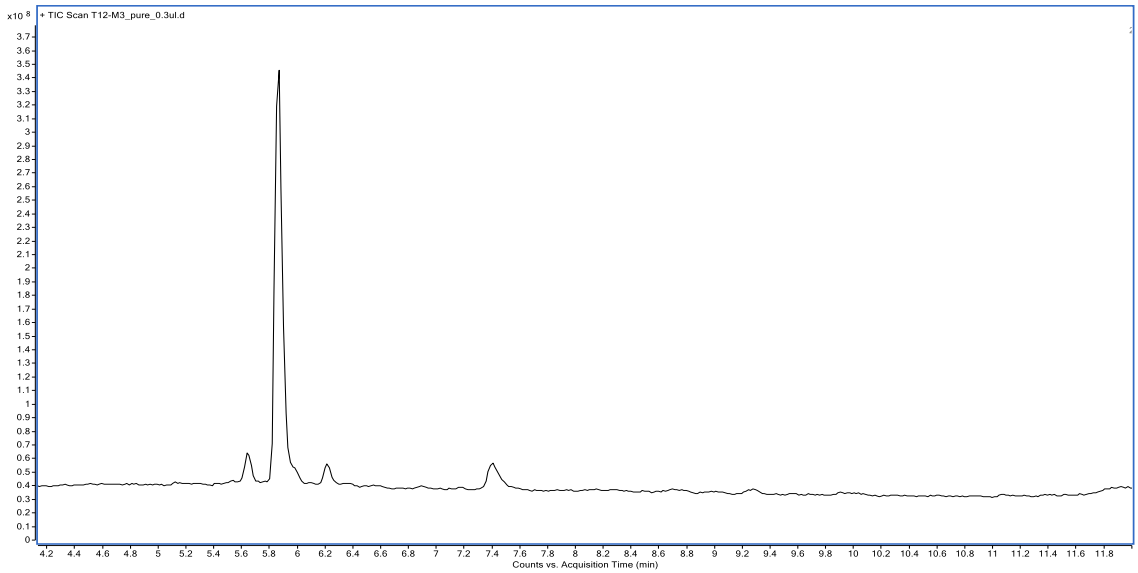


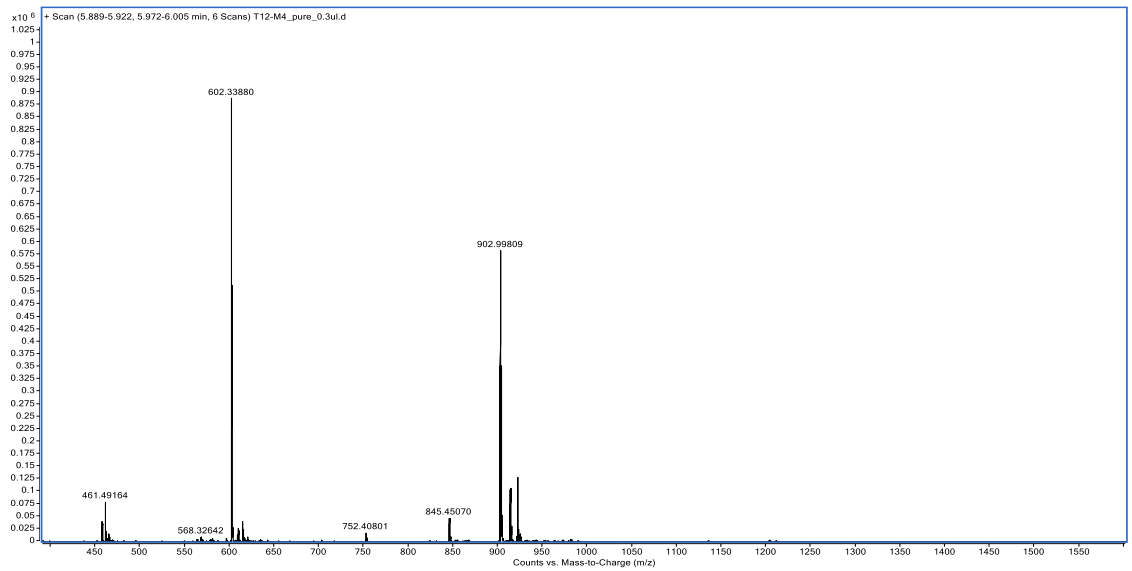




<u>biotin-T12-M6Nle-W7A-P9A</u>
Sequence: biotin-PEG_4-THRPP(Nle)ASAVWP-$CONH_2$.
Synthesis method: Automated flow synthesis.
Purification method: Section 2.4.4, method B
LC-MS method: Section 2.4.5, method C
Calculated: 1804.15 Da
Observed: 1804.04 Da

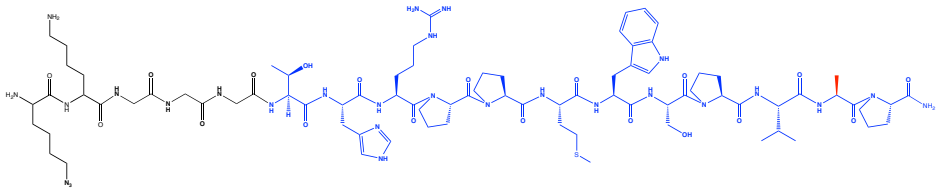


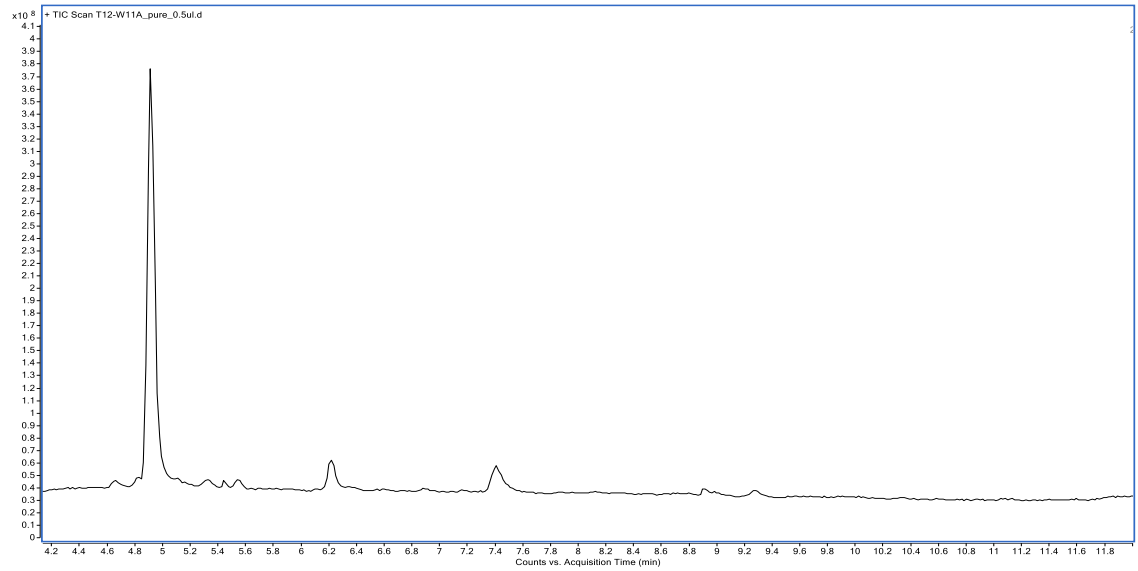


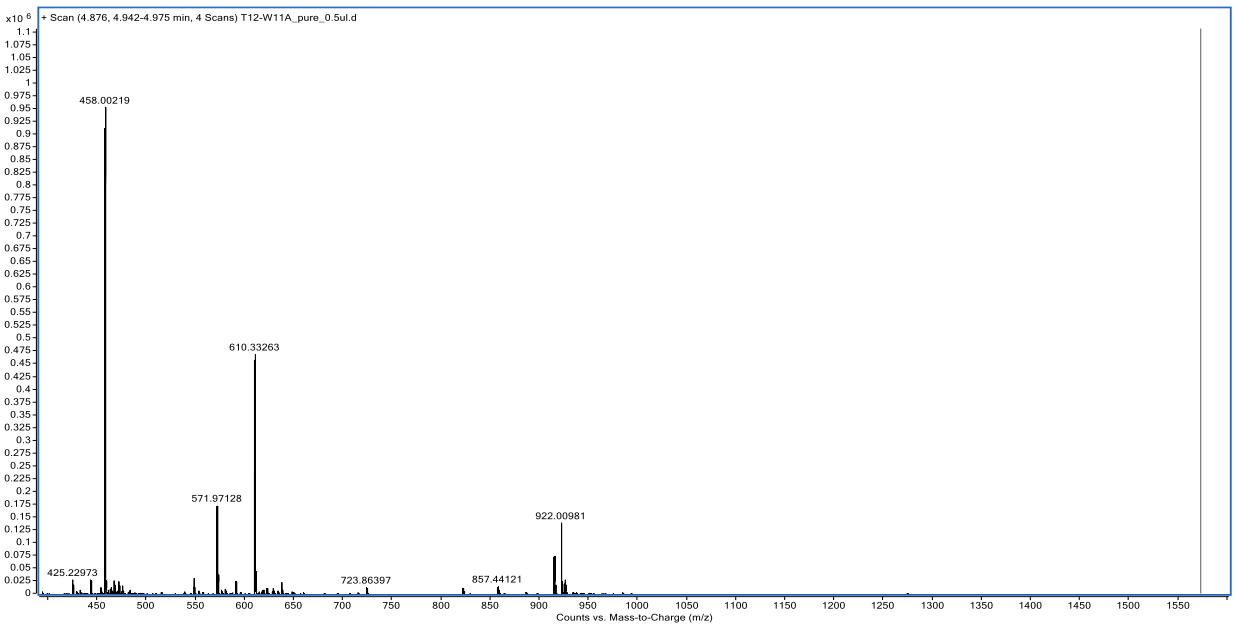


T12-W11A-NAz

Sequence: Lys(N_3)-KGGG-THRPPMWSPVAP-$CONH_2$.
Synthesis method: Automated flow synthesis.
Purification method: Section 2.4.4, method B
LC-MS method: Section 2.4.5, method C
Calculated: 1828.14 Da
Observed: 1828.07 Da







Anti-MiRNA-21-T12 dimer

Sequence: CATCAGTCTGATAAGCTA-KKK-THRPPMWSPVWP-KKK-THRPPMWSPVWP-$CONH_2$.

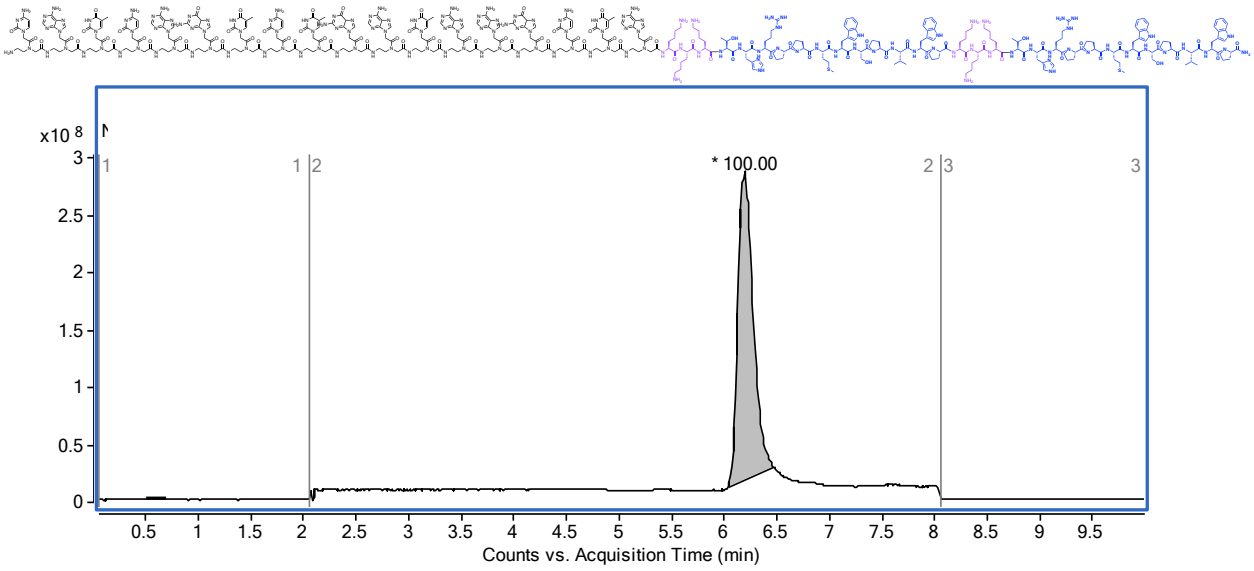
Synthesis method: Automated flow synthesis.

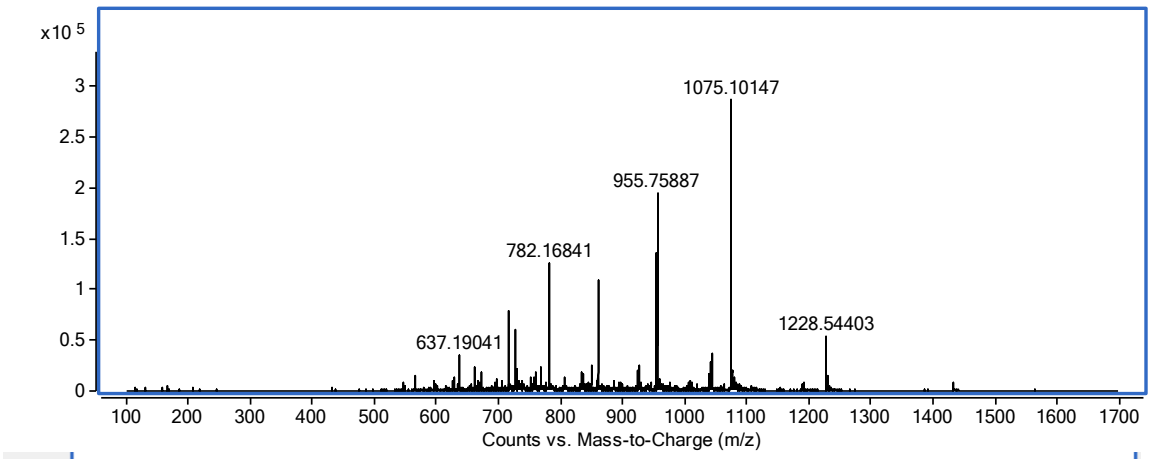
Purification method: Section 2.4.4, method C

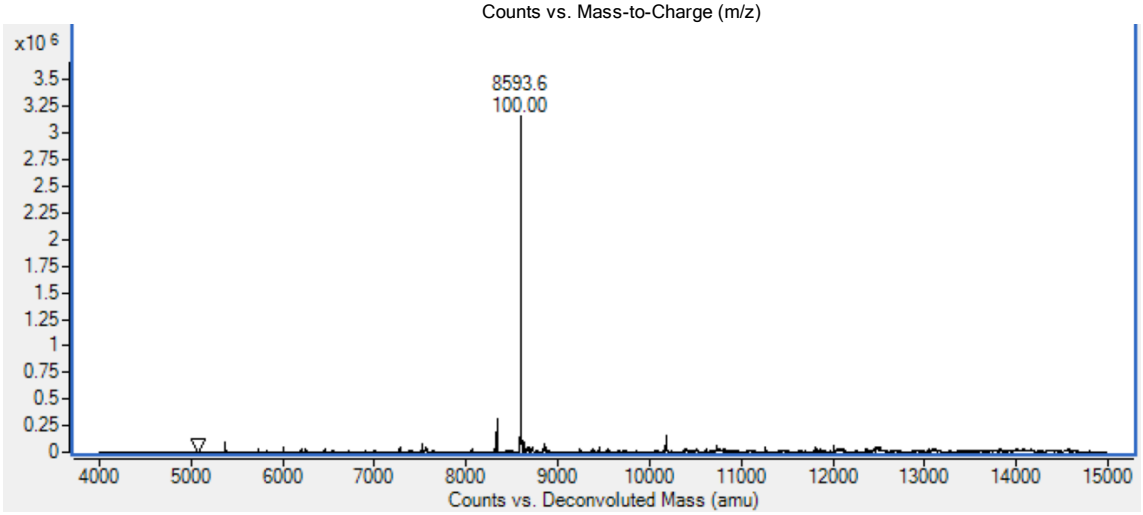
LC-MS method: Section 2.4.5, method C

Calculated: 8593.27 Da

Observed: 8593.60 Da

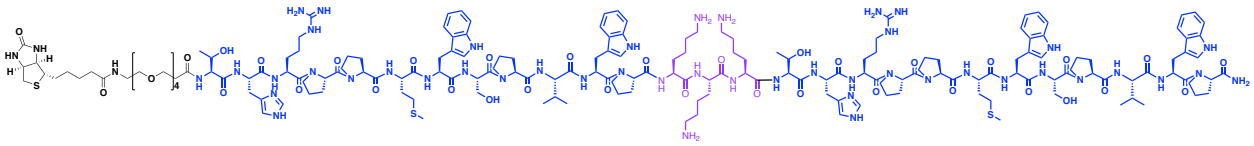


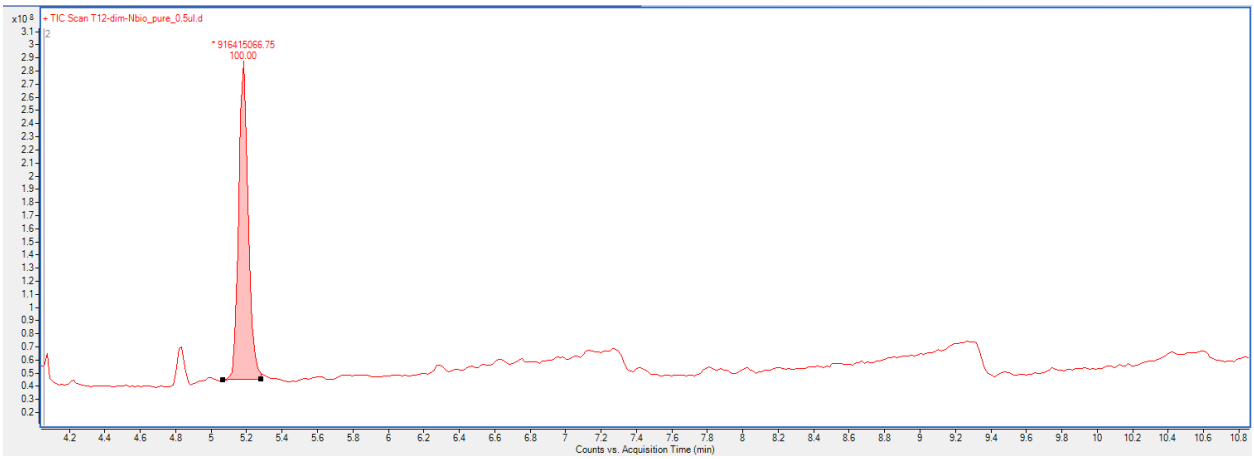

x10 5
3
2.5
2
1.5
1
0.5
0
1075.10147
955.75887
782.16841
637.19041
1228.54403
100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700
Counts vs. Mass-to-Charge (m/z)

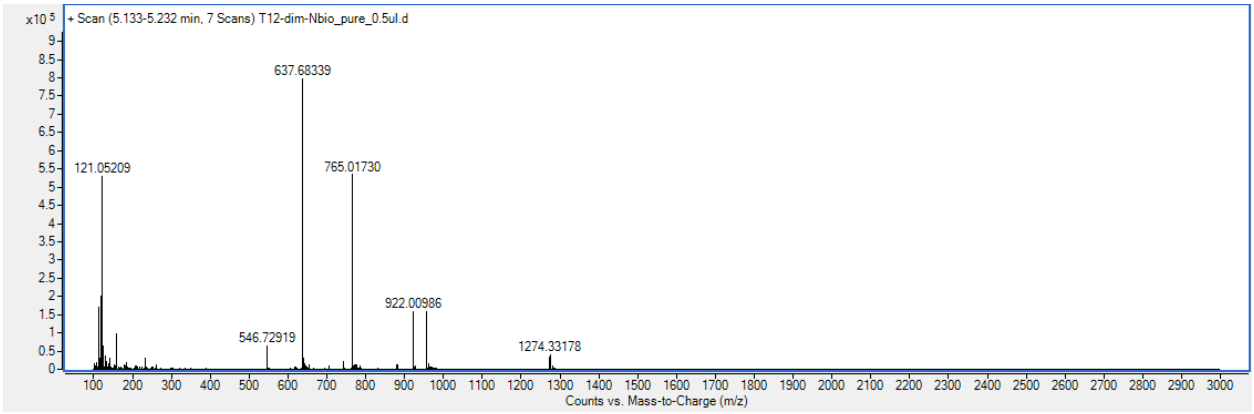

x10 6
3.5
3.25
3
2.75
2.5
2.25
2
1.75
1.5
1.25
1
0.75
0.5
0.25
0
8593.6
100.00
4000 5000 6000 7000 8000 9000 10000 11000 12000 13000 14000 15000
Counts vs. Deconvoluted Mass (amu)

biotin-T12 dimer

Sequence: biotin-PEG_4-THRPPMWSPVWP-KKK-THRPPMWSPVWP-$CONH_2$.
Synthesis method: Automated flow synthesis.
Purification method: Section S4, method D
LC-MS method: Section S5, method C
Calculated: 3820.61 Da
Observed: 3820.10 Da







T12-V1-4

Sequence: $(THRPP(Nle)WSPVWP)_2$-PEG_4-K-Lys(N_3)-$CONH_2$.

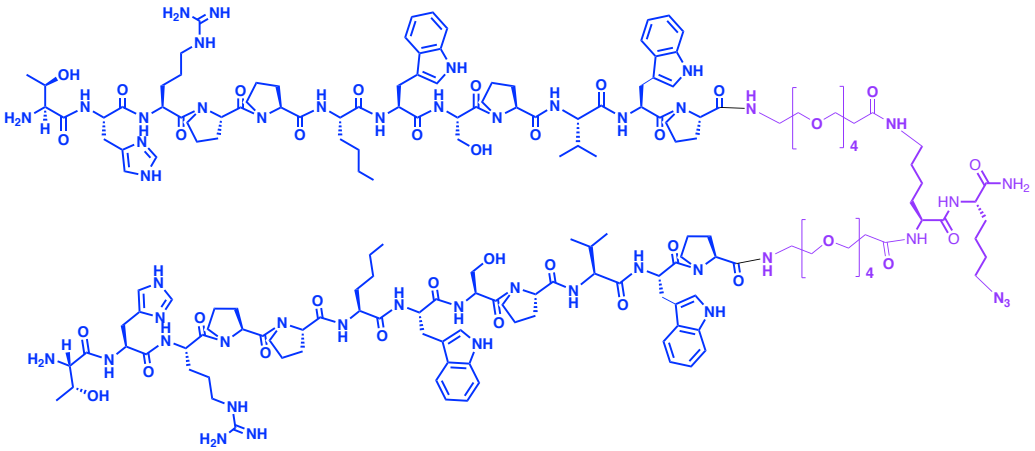
Synthesis method: Automated flow synthesis.

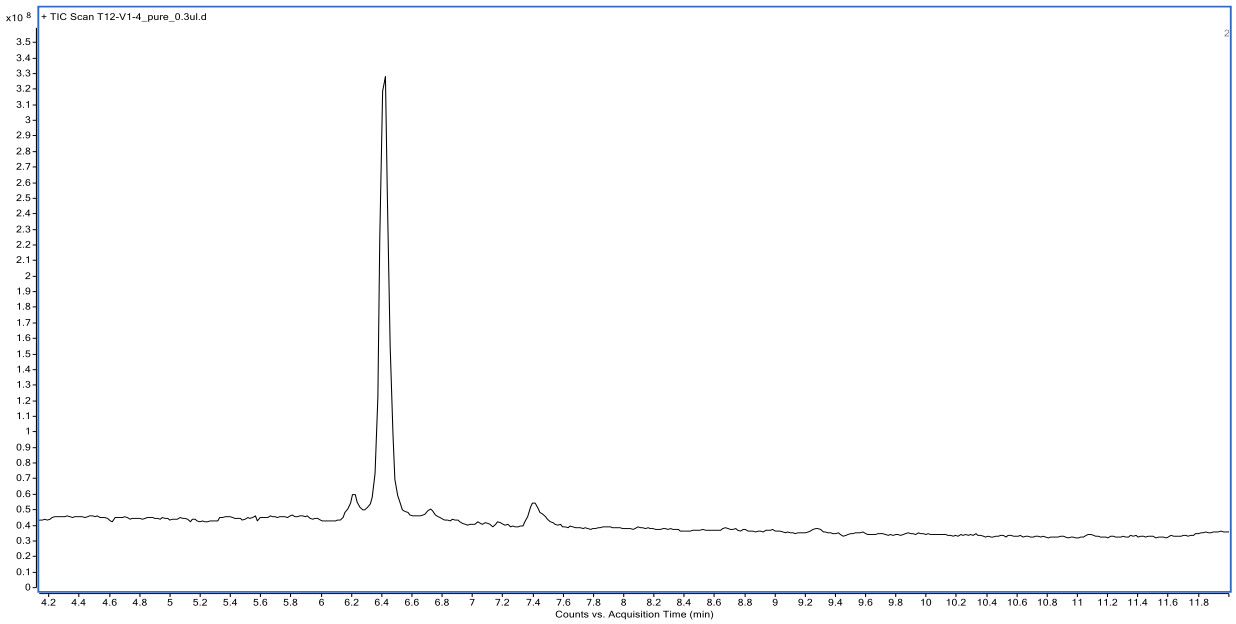
Purification method: Section S4, method D

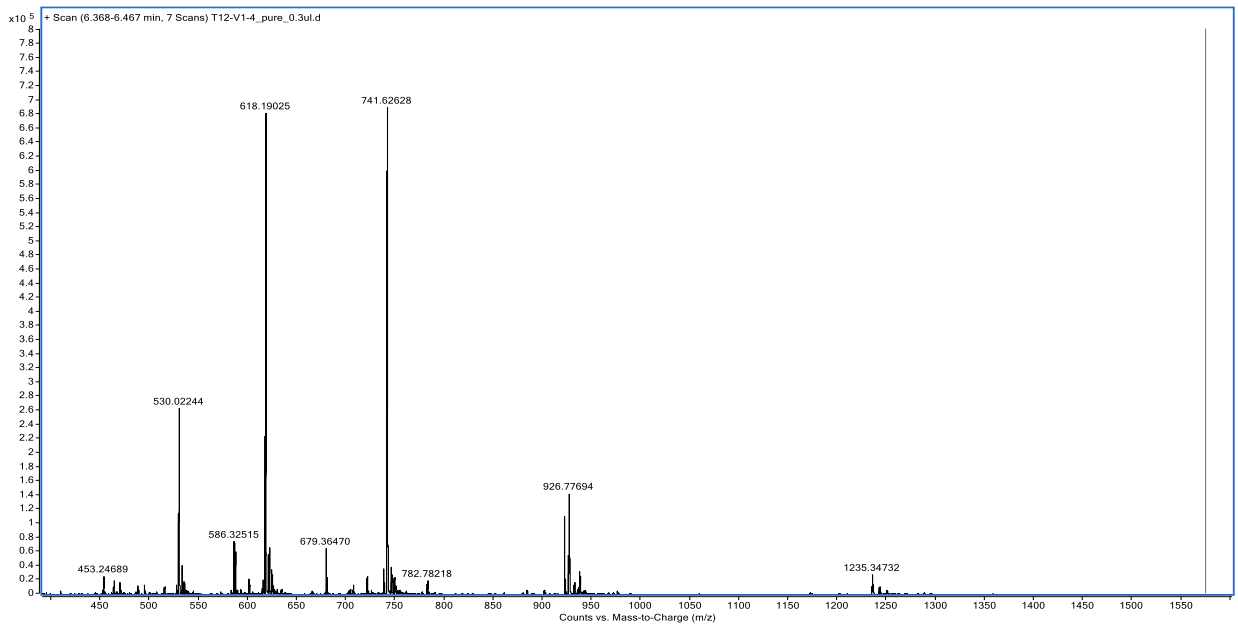
LC-MS method: Section S5, method C

Calculated: 3703.36 Da

Observed: 3703.03 Da

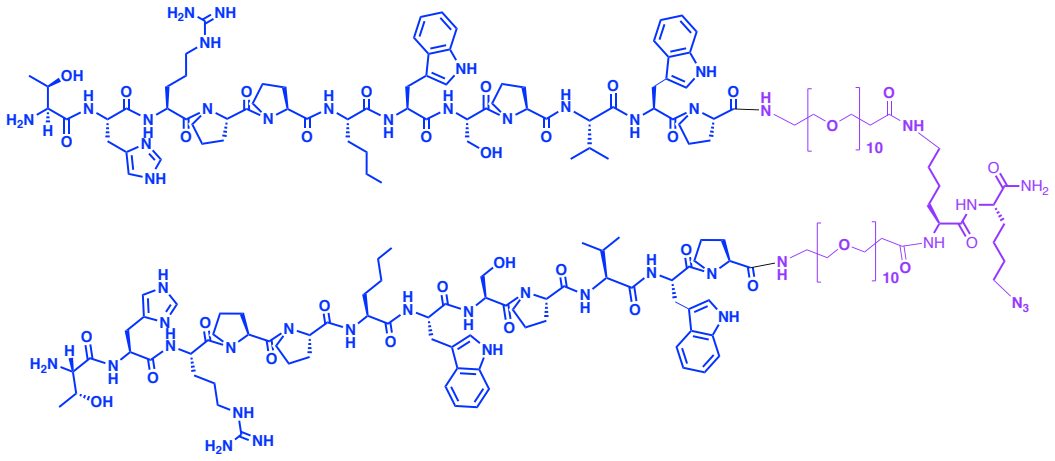


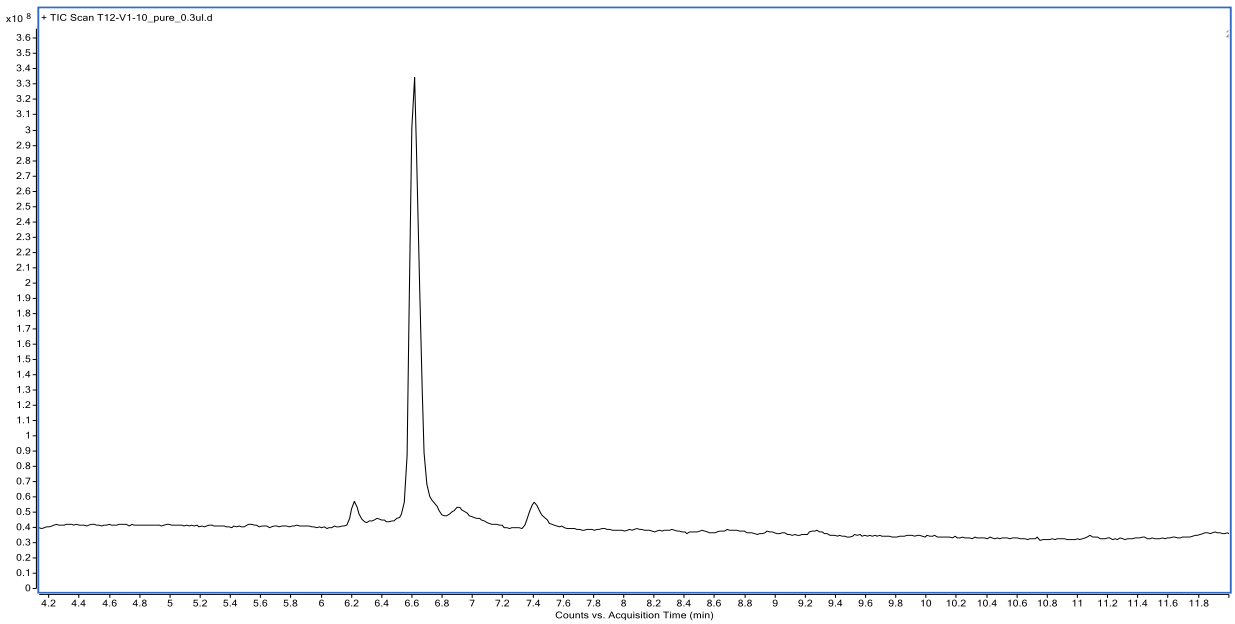


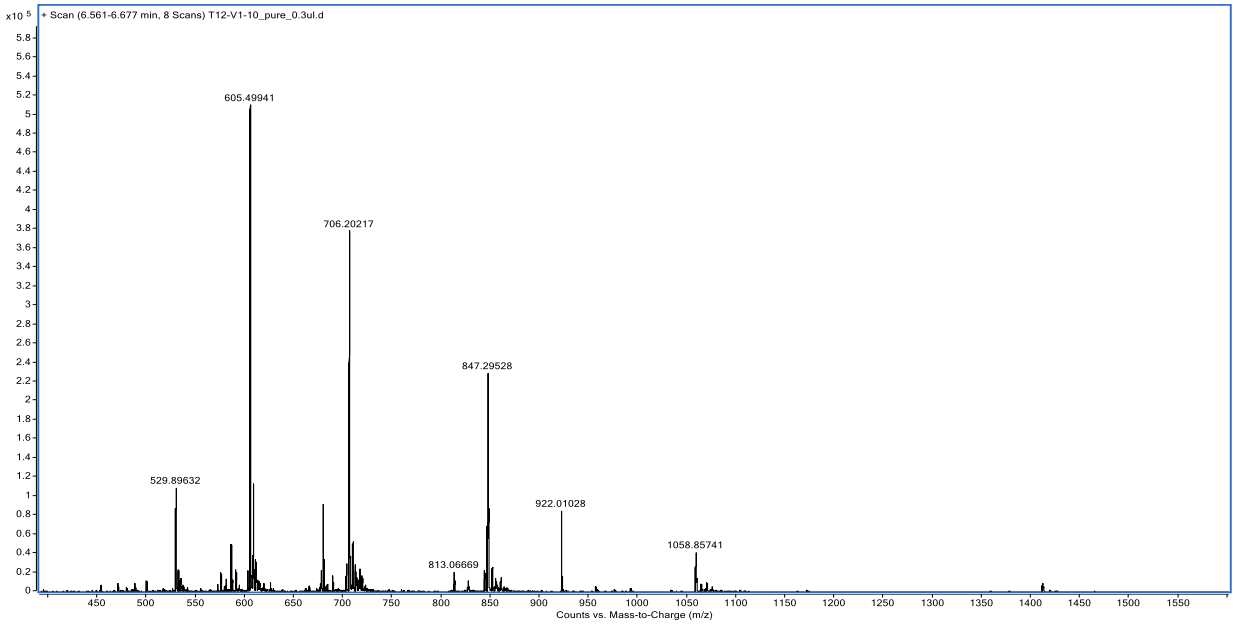


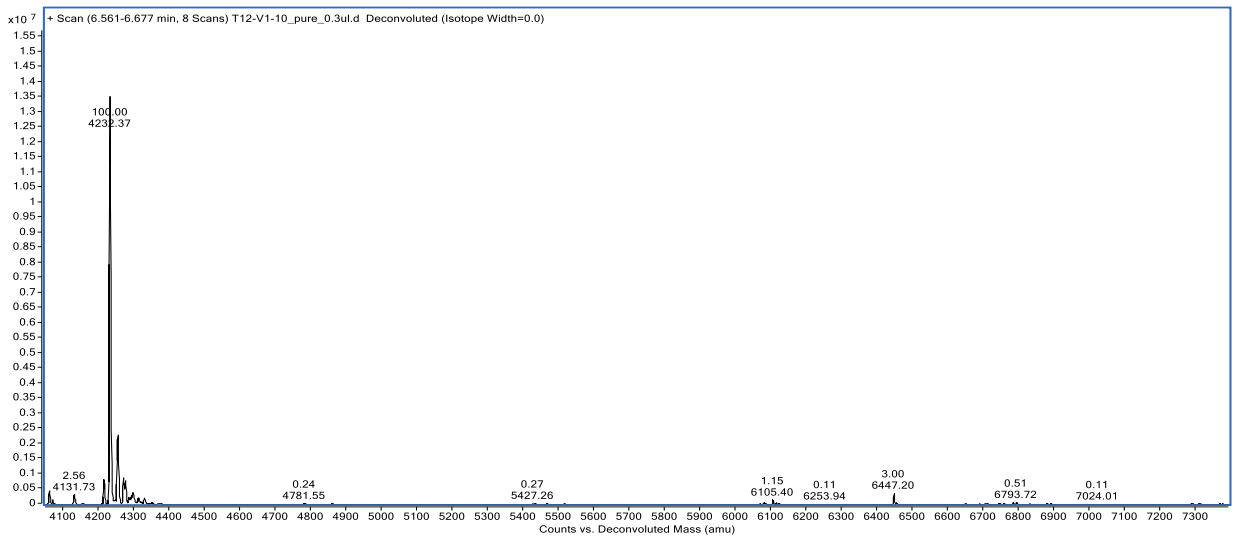
T12-V1-10

Sequence: (THRPP(Nle)WSPVWP)$_2$-PEG$_{10}$-K-Lys(N$_3$)-CONH$_2$.
Synthesis method: Automated flow synthesis.
Purification method: Section S4, method D
LC-MS method: Section S5, method C
Calculated: 4231.99 Da
Observed: 4232.37 Da



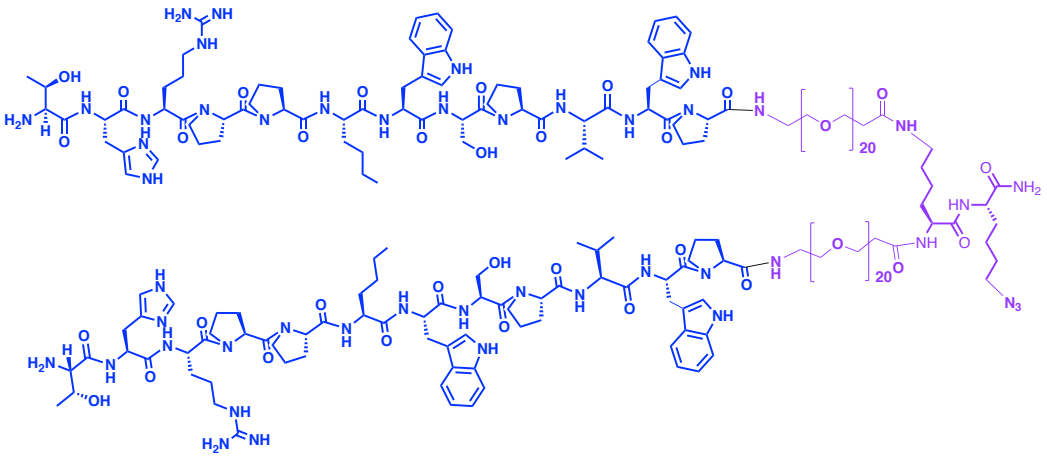

+ TIC Scan T12-V1-10_pure_0.3ul.d
x10 8
Counts vs. Acquisition Time (min)

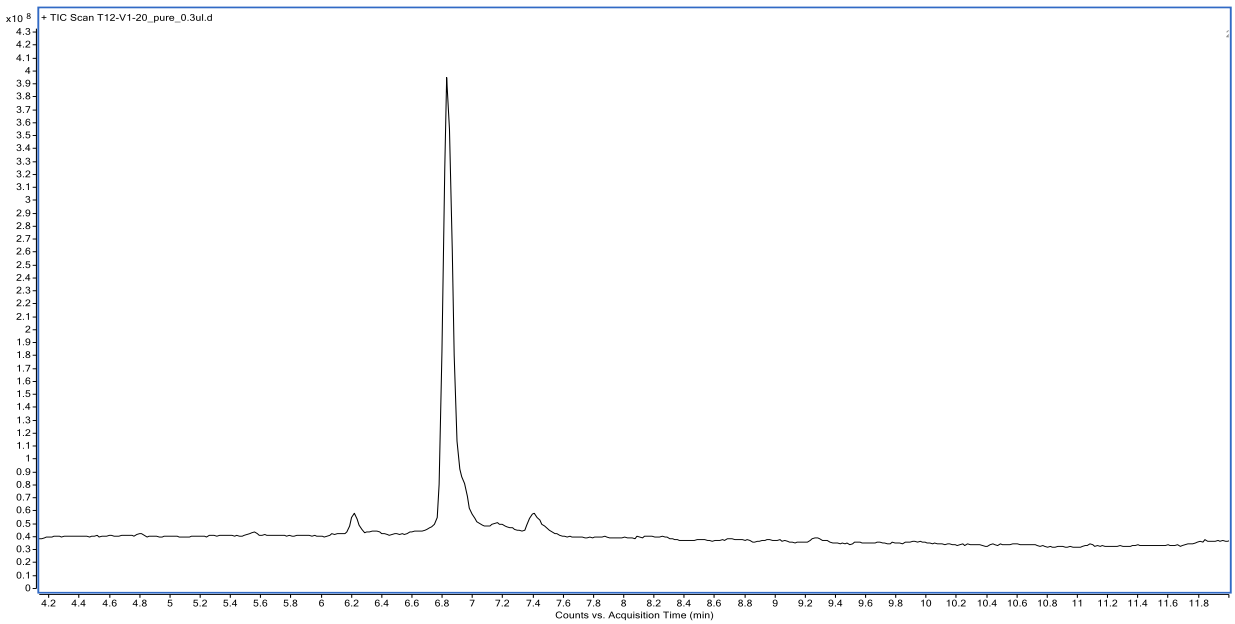

+ Scan (6.561-6.677 min, 8 Scans) T12-V1-10_pure_0.3ul.d
x10 5
605.49941
706.20217
847.29528
529.89632
922.01028
1058.85741
813.06669
Counts vs. Mass-to-Charge (m/z)

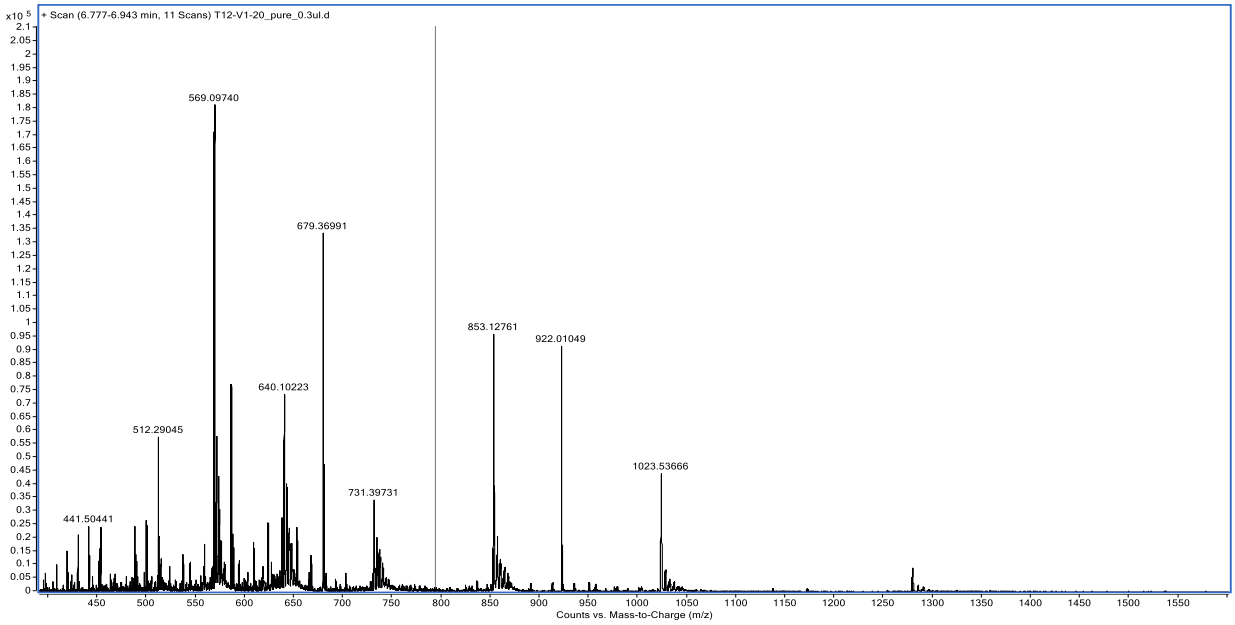


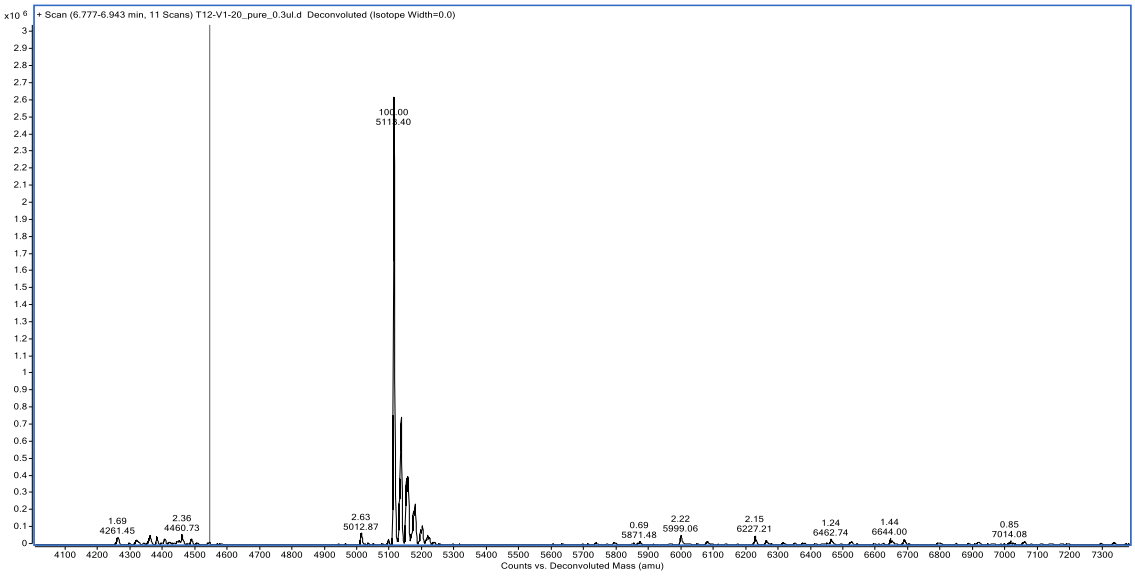
T12-V1-20

Sequence: $(THRPP(Nle)WSPVWP)_2$-PEG_{20}-K-Lys(N_3)-$CONH_2$.
Synthesis method: Automated flow synthesis.
Purification method: Section S4, method D
LC-MS method: Section S5, method C
Calculated: 5113.03 Da
Observed: 5113.40 Da



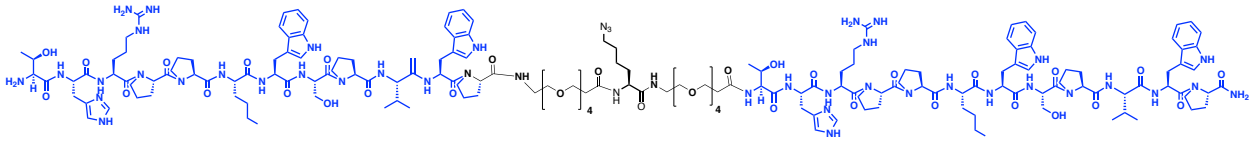

+ TIC Scan T12-V1-20_pure_0.3ul.d
x10 8
Counts vs. Acquisition Time (min)

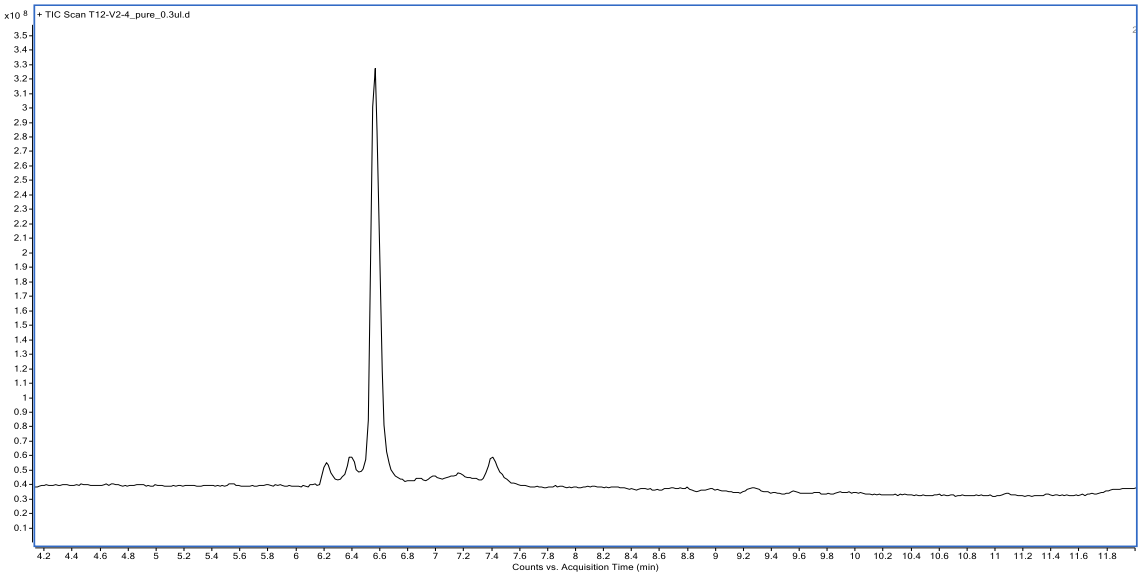

+ Scan (6.777-6.943 min, 11 Scans) T12-V1-20_pure_0.3ul.d
x10 5
569.09740
679.36991
853.12761
922.01049
640.10223
512.29045
1023.53666
731.39731
441.50441
Counts vs. Mass-to-Charge (m/z)

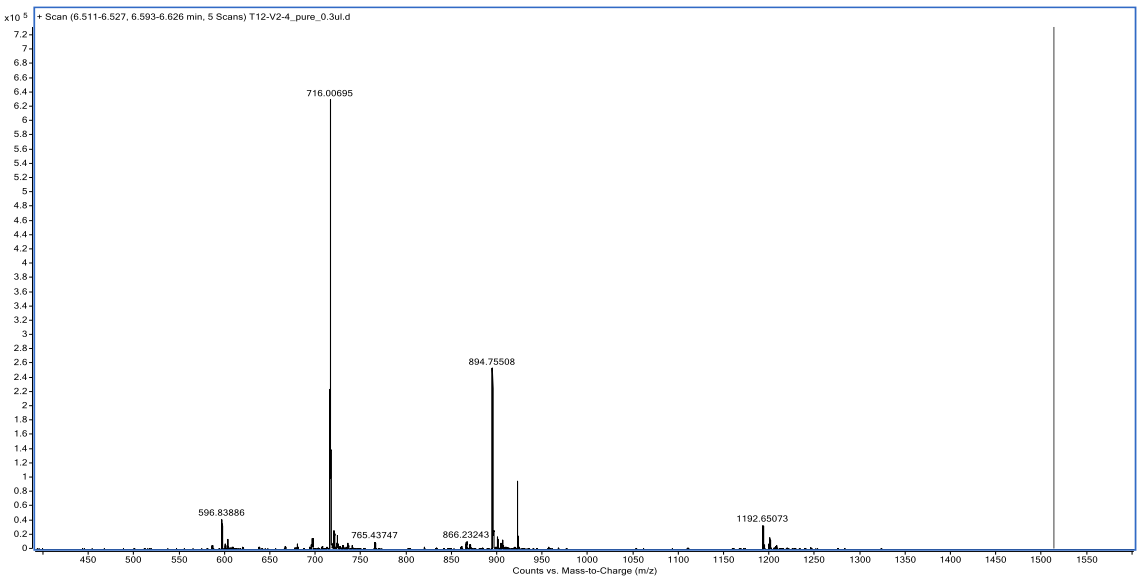


T12-V2-4

Sequence: THRPP(Nle)WSPVWP-PEG_4-Lys(N_3)-PEG_4-THRPP(Nle)WSPVWP-$CONH_2$.
Synthesis method: Automated flow synthesis.
Purification method: Section S4, method D
LC-MS method: Section S5, method C
Calculated: 3573.21 Da
Observed: 3574.93 Da







T12-V2-10

Sequence: THRPP(Nle)WSPVWP-PEG_{10}-Lys(N_3)-PEG_{10}-THRPP(Nle)WSPVWP-$CONH_2$.

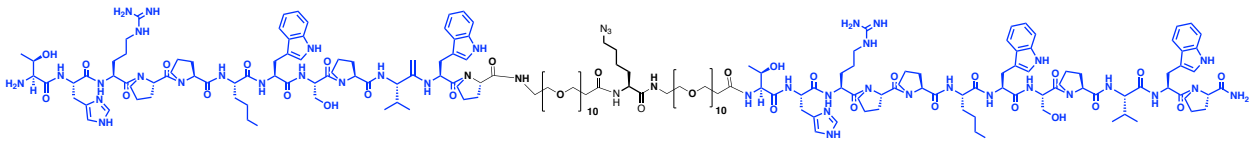
Synthesis method: Automated flow synthesis.

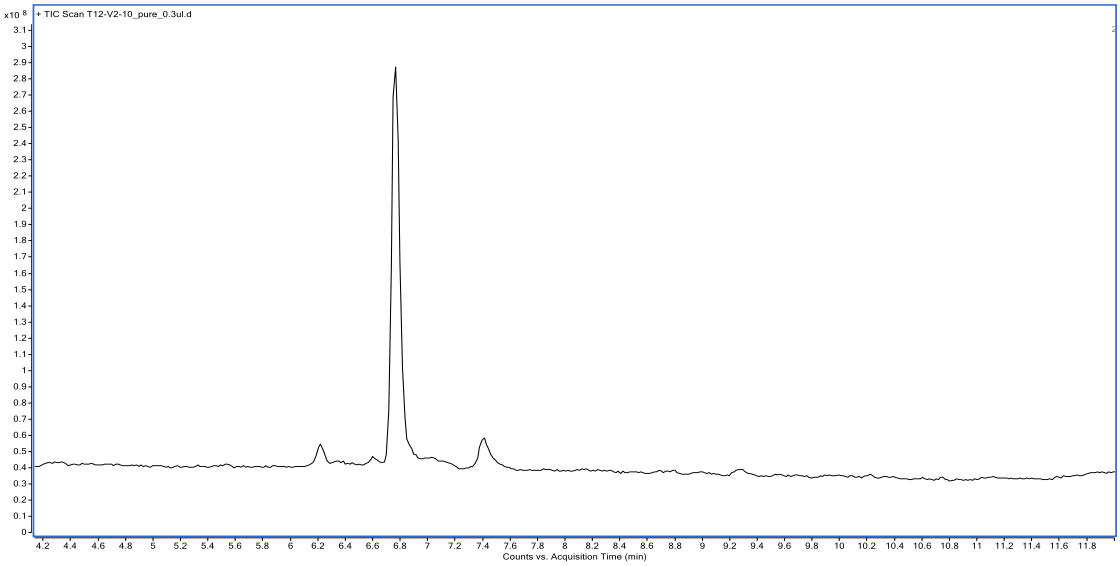
Purification method: Section S4, method D

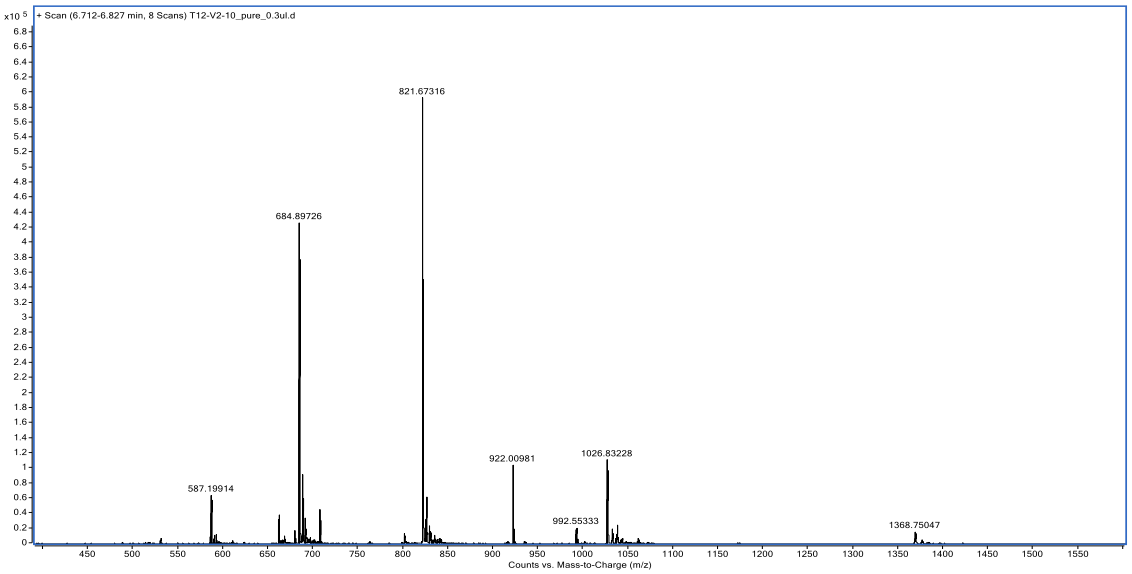
LC-MS method: Section S5, method C

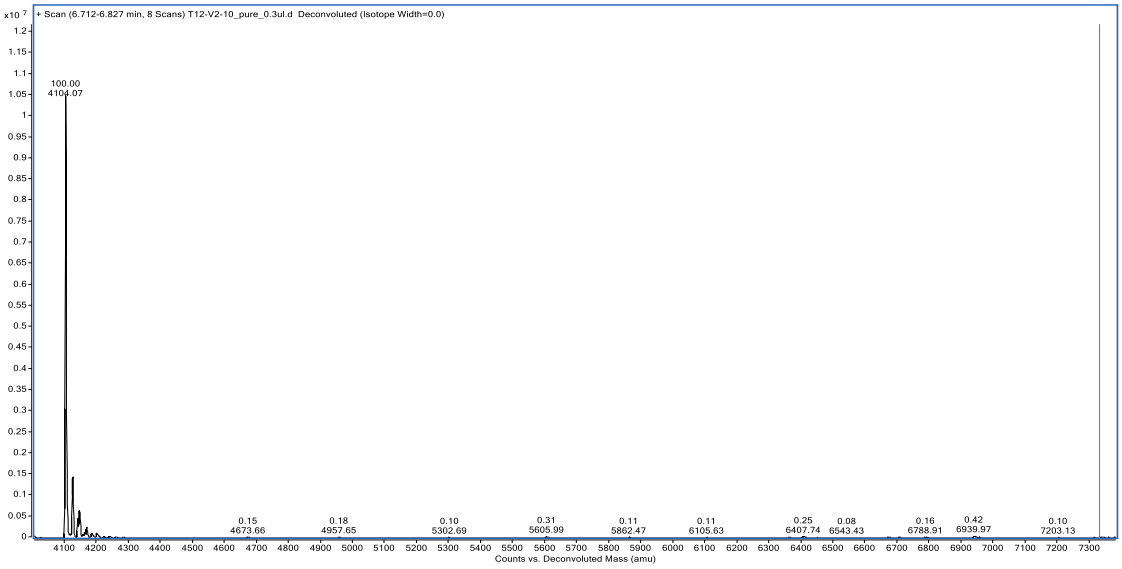
Calculated: 4101.84 Da

Observed: 4104.07 Da






+ Scan (6.712-6.827 min, 8 Scans) T12-V2-10_pure_0.3ul.d
x10 5
821.67316
684.89726
922.00981
1026.83228
587.19914
992.55333
1368.75047
Counts vs. Mass-to-Charge (m/z)


+ Scan (6.712-6.827 min, 8 Scans) T12-V2-10_pure_0.3ul.d Deconvoluted (Isotope Width=0.0)
x10 7
100.00
4104.07
0.15
4673.66
0.18
4957.65
0.10
5302.69
0.31
5605.99
0.11
5862.47
0.11
6105.63
0.25
6407.74
0.08
6543.43
0.16
6788.91
0.42
6939.97
0.10
7203.13
Counts vs. Deconvoluted Mass (amu)

T12-V2-20

Sequence: THRPP(Nle)WSPVWP-PEG_{20}-Lys(N_3)-PEG_{20}-THRPP(Nle)WSPVWP-$CONH_2$.

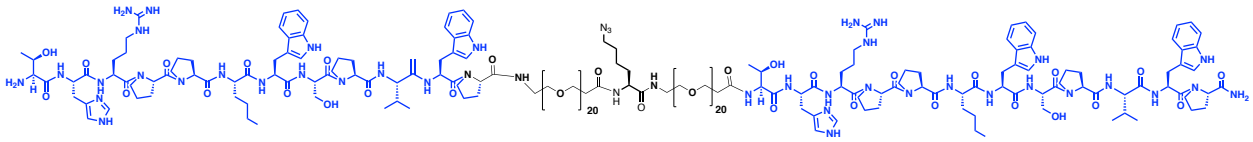
Synthesis method: Automated flow synthesis.

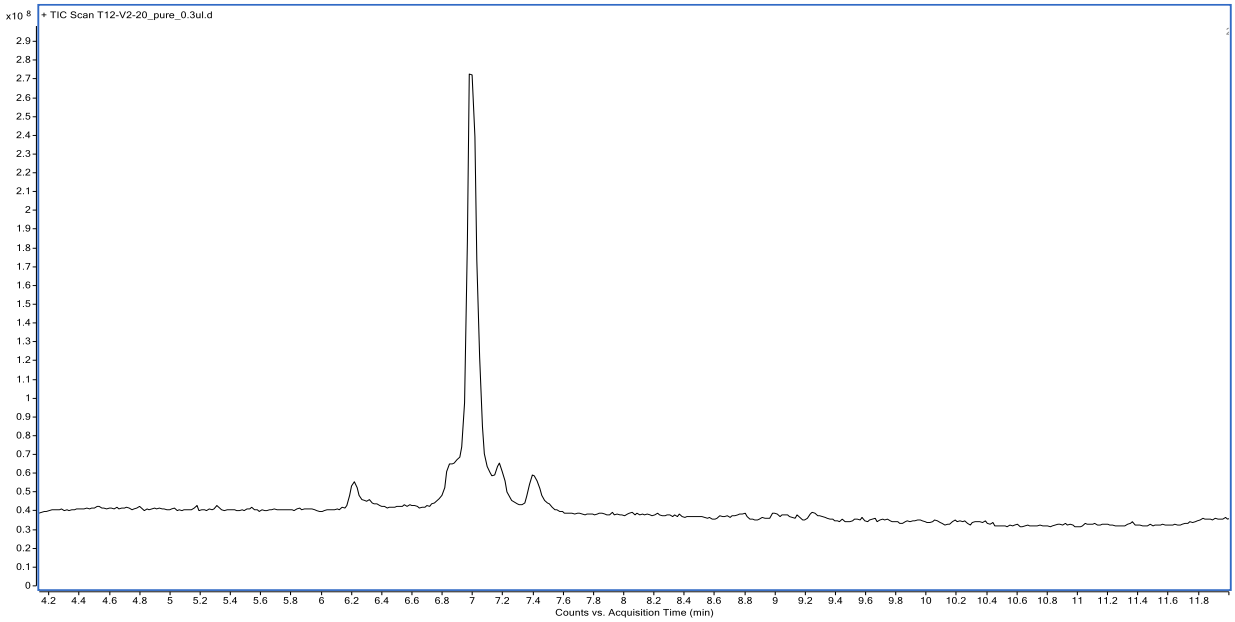
Purification method: Section S4, method D

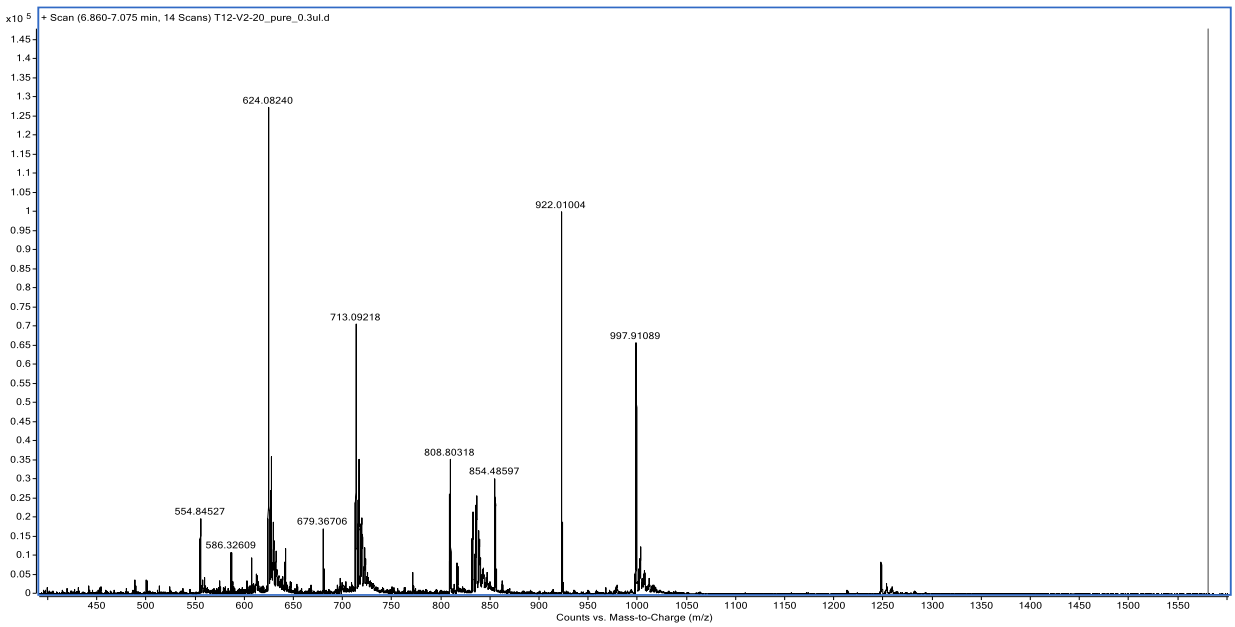
LC-MS method: Section S5, method C

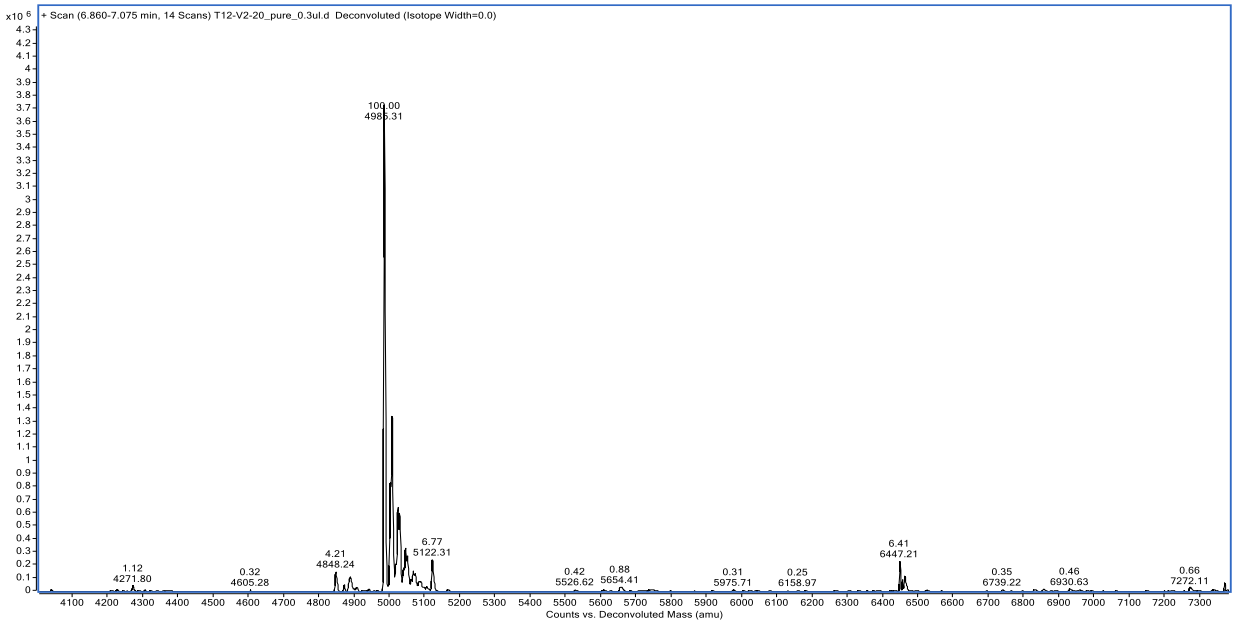
Calculated: 4982.88 Da

Observed: 4985.31 Da






x10 5
+ Scan (6.860-7.075 min, 14 Scans) T12-V2-20_pure_0.3ul.d
624.08240
922.01004
713.09218
997.91089
808.80318
854.48597
554.84527
679.36706
586.32609
Counts vs. Mass-to-Charge (m/z)


x10 6
+ Scan (6.860-7.075 min, 14 Scans) T12-V2-20_pure_0.3ul.d Deconvoluted (Isotope Width=0.0)
100.00
4985.31
6.77
5122.31
4.21
4848.24
1.12
4271.80
0.32
4605.28
0.42
5526.62
0.88
5654.41
0.31
5975.71
0.25
6158.97
6.41
6447.21
0.35
6739.22
0.46
6930.63
0.66
7272.11
Counts vs. Deconvoluted Mass (amu)

PMO-DBCO

Purchased anti-miRNA-21 PMO was dissolved in DMSO. 4 eq DBCO-acid was dissolved at 0.4 M in 0.38 M HBTU (3.8 eq) in amine-free DMF. 10 eq DIEA was added to the DBCO immediately before combining with PMO. The reaction was allowed to sit at room temperature overnight.

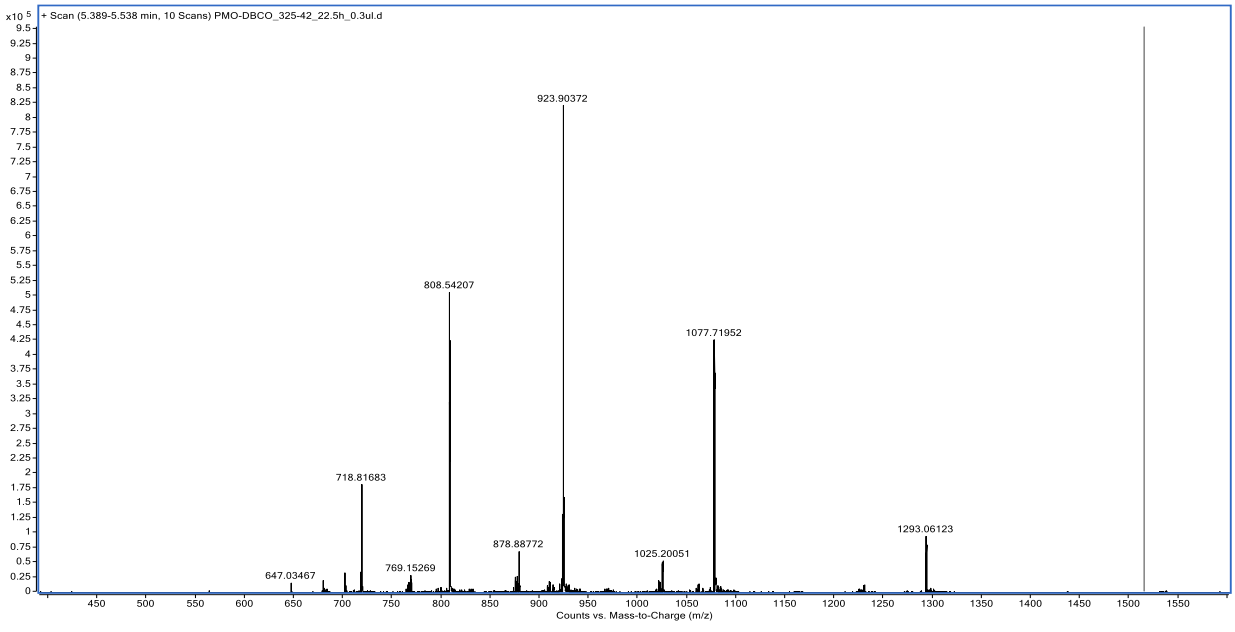
Purification method: Section S4, method D

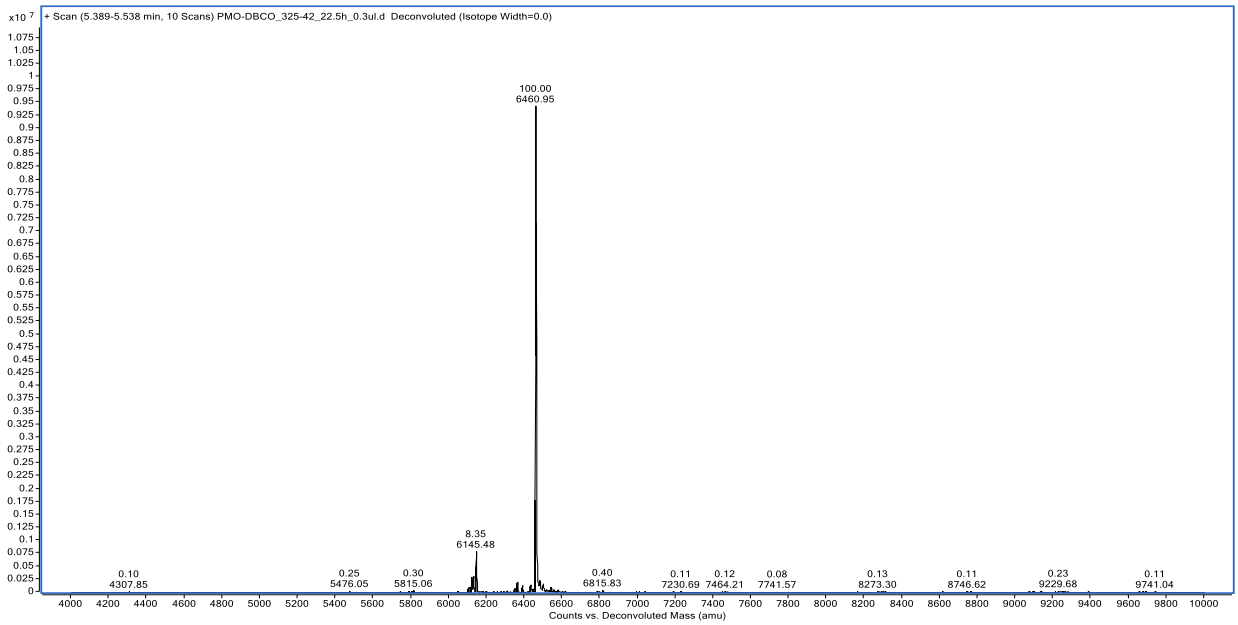
LC-MS method: Section S5, method C

Calculated: 6461.20 Da

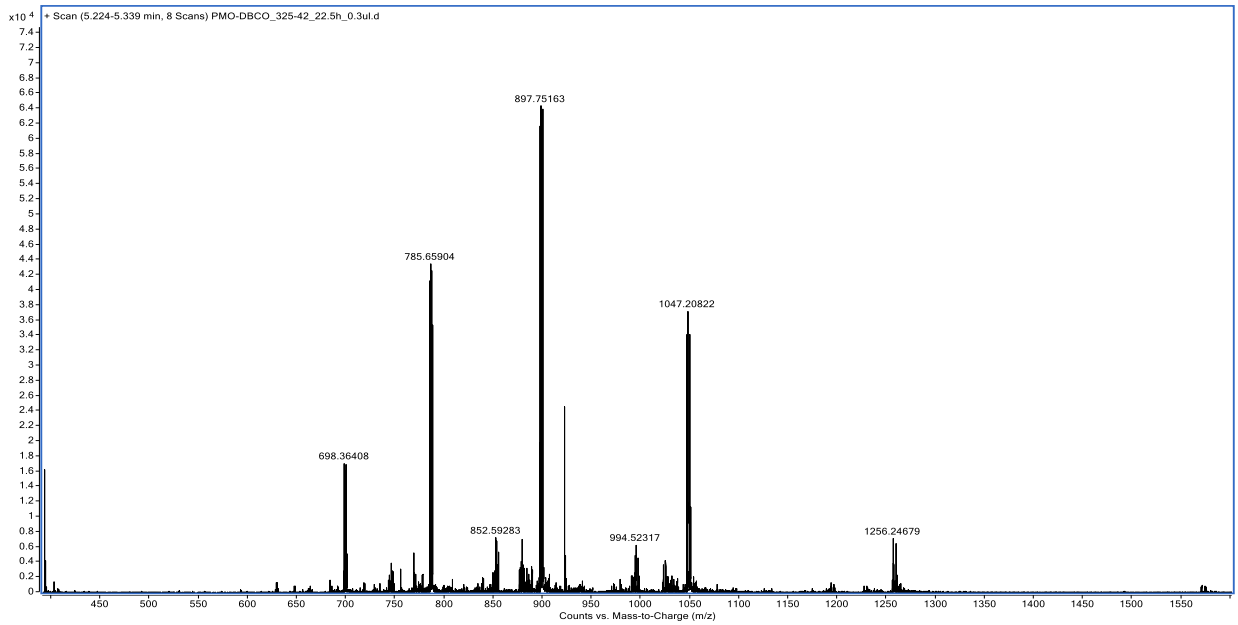
Observed: 6460.95 Da

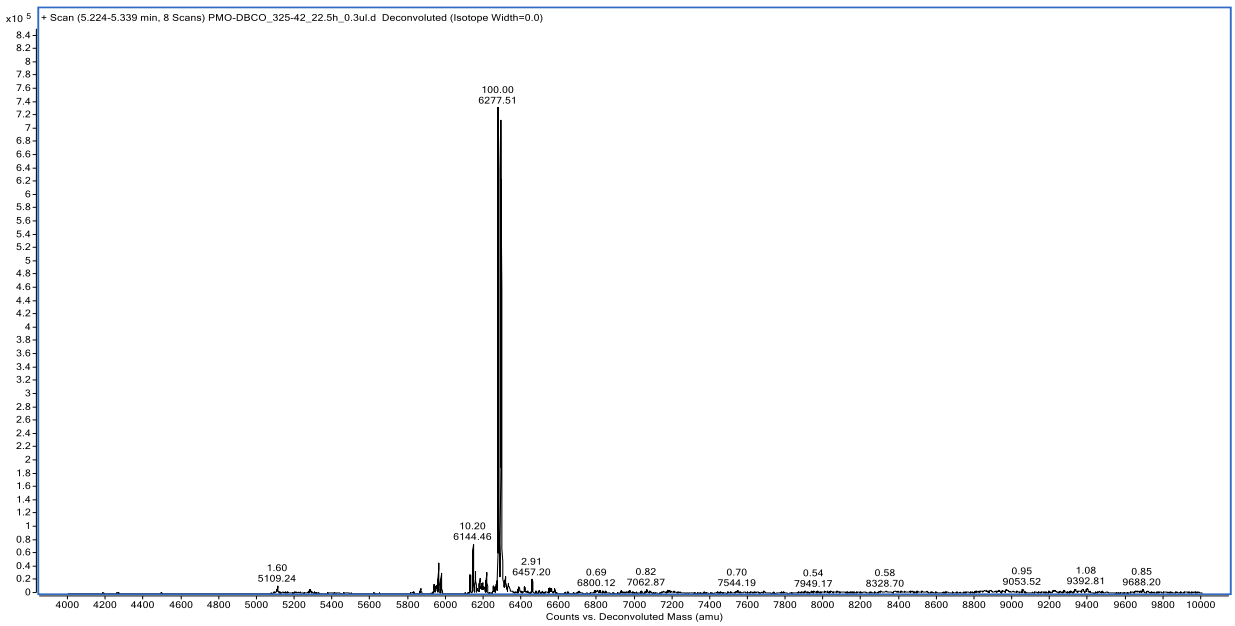
Right peak: PMO-DBCO





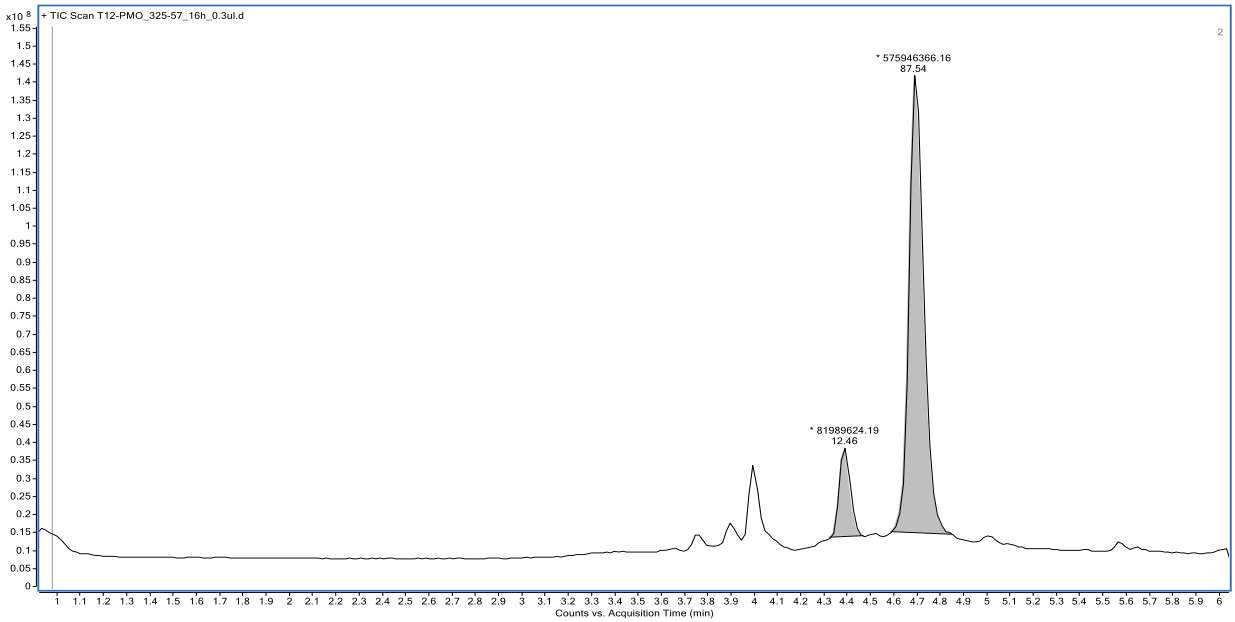
Left peak: unreacted PMO
Calculated: 6277 Da
Observed: 6277.51 Da



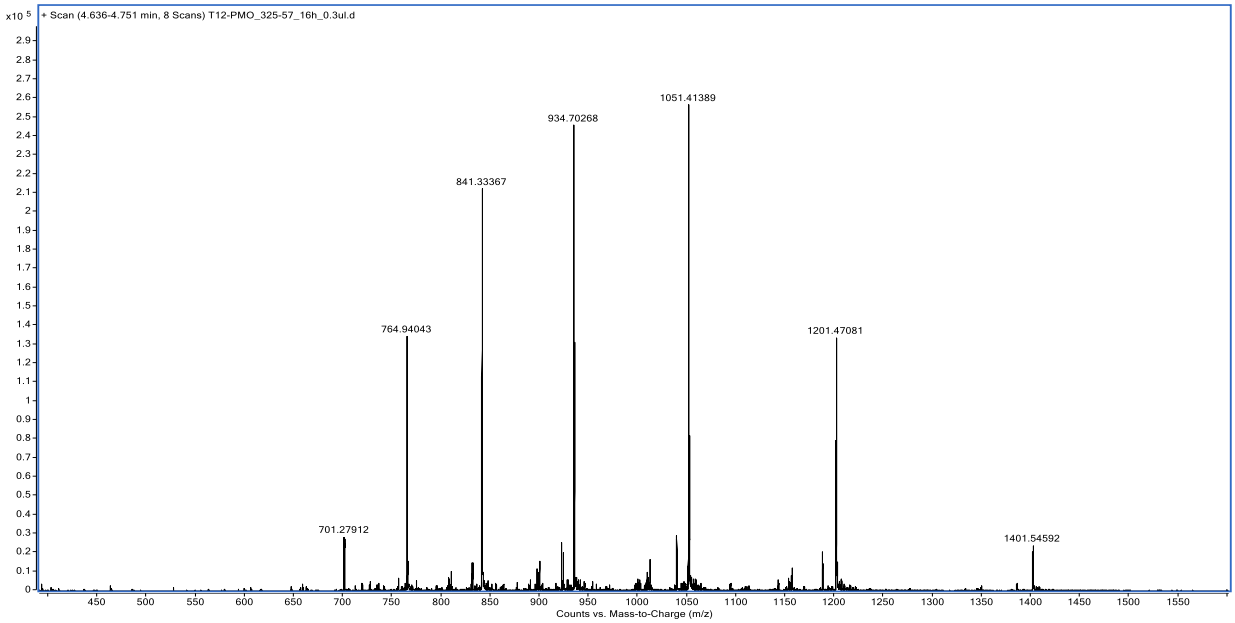


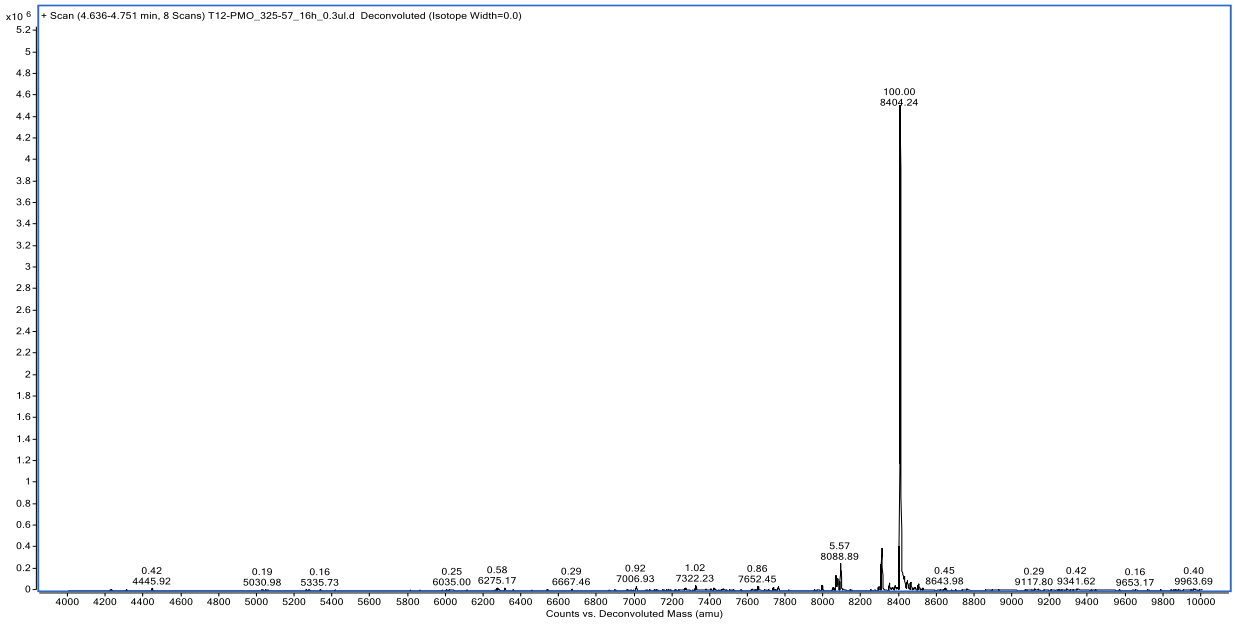
PMO-T12

PMO-DBCO was reacted to T12-NAz as described in Section 2.4.10
LC-MS method: Section S5, method C
Calculated: 8404.47 Da
Observed: 8404.24 Da

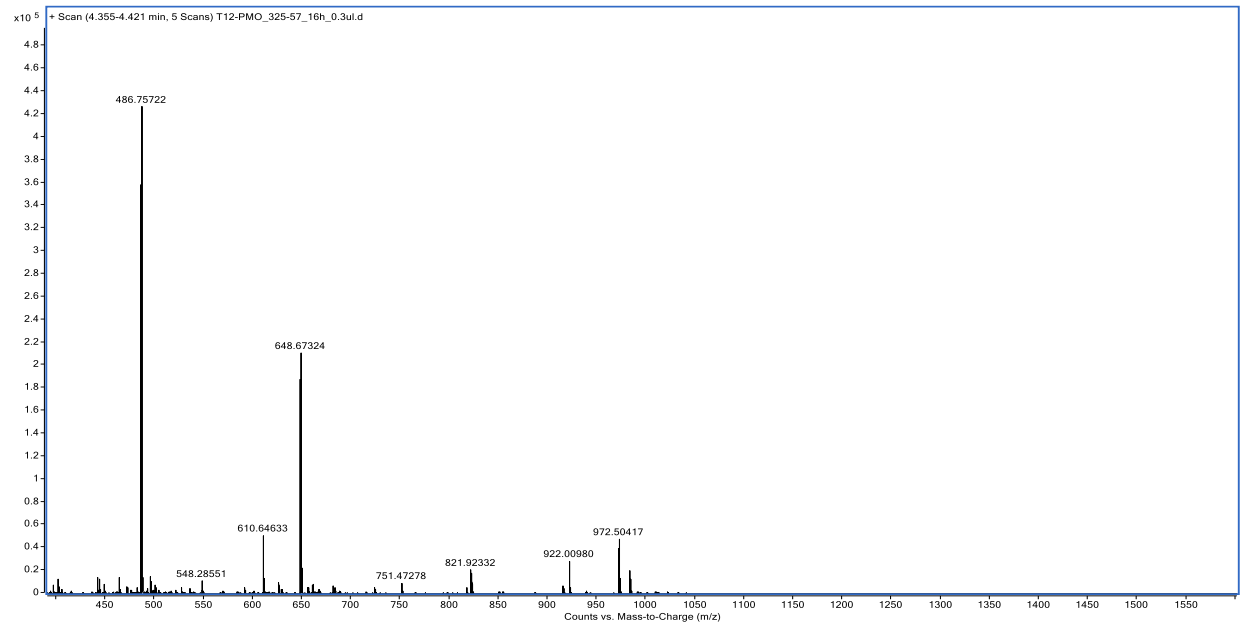


Right peak: PMO-T12





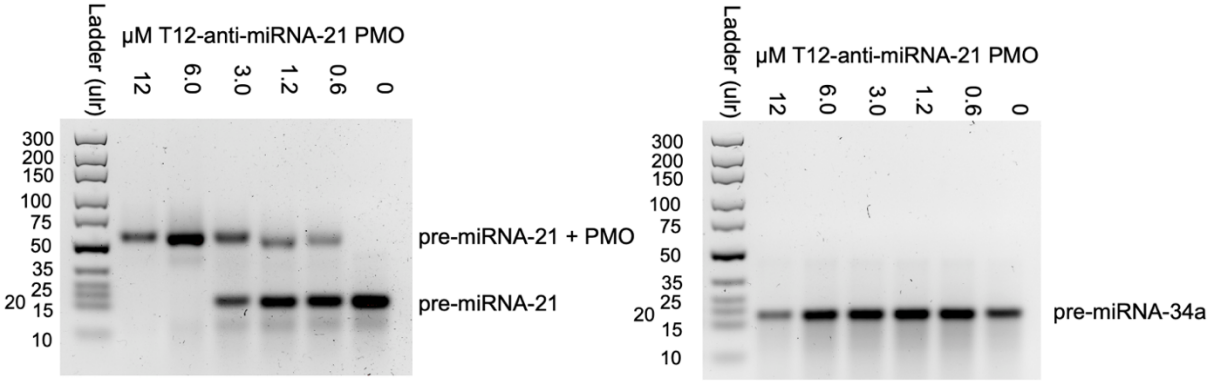
Left peak: excess T12-NAz
Calculated: 1943.27 Da
Observed: 1943.00 Da



2.8 Appendix II: Nucleic acid gels of alternative ASOs

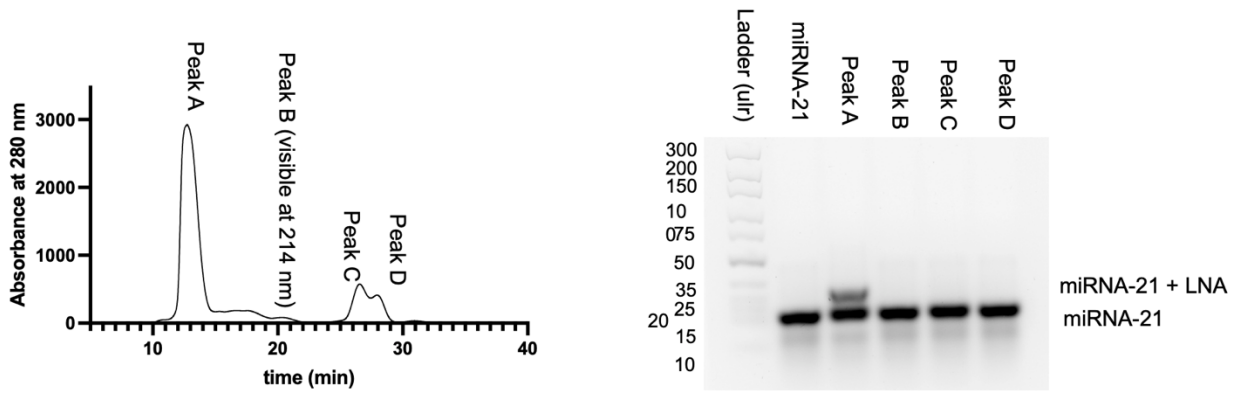
PMO-T12

Anti-miRNA-21 PMO was purchased from GeneTool with a C-terminal amine. PMO was coupled to DBCO-acid (see LC-MS in Appendix I) and purified as described in Section 2.4.4 Method C. PMO-DBCO was clicked to T12 with an N-terminal azidolysine handle. In the nucleic acid gel, 6 μM pre-miRNA-21 or pre-miRNA-34apurchased from IDT in nuclease-free water was pre-mixed with several concentrations of PMO-T12 (12, 6.0, 3.0, 1.2, 0.6, and 0 μM in nuclease-free water) for 1 hour at 37°C. Samples were loaded at ~400 ng pre-miRNA/well. Gels were 5% agarose with Cybr gold stain and were run in 1X TBE buffer at 100 V for 70 min. We observe concentration-dependent binding of PMO-T12 to pre-miRNA-21 but not pre-miRNA-34a, indicating the ASO binds specifically to its target pre-miRNA.

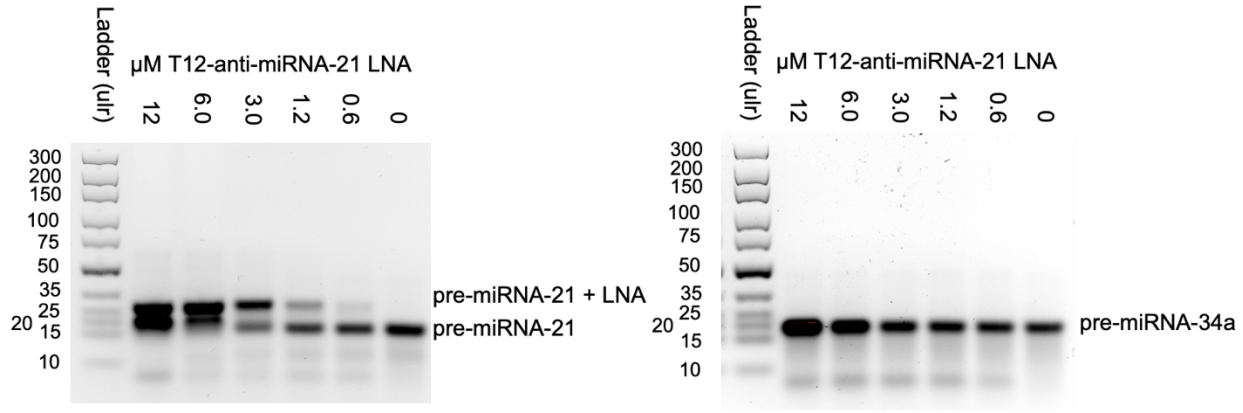


LNA-T12

Anti-miRNA-21 LNA was purchased from IDT with a C-terminal amine. LNA was coupled to DBCO-acid and purified by running through a NAP-10 column to remove excess DBCO. LNA-DBCO was clicked to T12 with an N-terminal azidolysine handle with excess 12 to ensure all LNA-DBCO was reacted. LNA-T12 was purified by SEC on a Superdex 75 Increase column at 0.80 mL/min 1X PBS for 40 min (2 CV). Fractions were collected and compiled into four peaks visible at 214 nm or 280 nm. Fractions were analyzed by nucleic acid gel under the same pre-mixing with pre-miRNA-21 and running conditions as described for PMO-T12. Peak A was the only fraction to indicate the presence of LNA.



Peak A of LNA-T12 was run in the same gel shift assay with pre-miRNA-21 and pre-miRNA-34a as with PMO-T12. Again, we observed concentration-dependent binding of LNA-T12 to pre-miRNA-21 but not pre-miRNA-34a.



3 References

(1) Mudd, J. O.; Kass, D. A. Tackling Heart Failure in the Twenty-First Century. *Nature* **2008**, *451* (7181), 919–928. https://doi.org/10.1038/nature06798.

(2) Simon Barquera; Andrea Pedroza-Tobias; Catalina Medina; Lucia Hernandez-Barrera; Kirsten Bibbins-Domingo; Rafael Lozano; Andrew E. Moran. Global Overview of the Epidemiology of Atherosclerotic Cardiovascular Disease. *Arch. Med. Res.* **2015**, *46* (5), 328–338.

(3) Leon, B. M.; Thomas M. Maddox. Diabetes and Cardiovascular Disease: Epidemiology, Biological Mechanisms, Treatment Recommendations and Future Research. *World J. Diabetes* **2015**, *6* (13), 1246. https://doi.org/10.4239/wjd.v6.i13.1246.

(4) Laakso, M. Cardiovascular Disease in Type 2 Diabetes: Challenge for Treatment and Prevention: REVIEW: CVD IN TYPE 2 DIABETES. *J. Intern. Med.* **2008**, *249* (3), 225–235. https://doi.org/10.1111/j.1365-2796.2001.00789.x.

(5) Steven R. Houser; Kenneth B. Margulies; Anne M. Murphy; Francis G. Spinale; Gary S. Francis; Sumanth D. Prabhu; Howard A. Rockman; David A. Kass; Jeffery D. Molkentin; Mark A. Sussman; Walter J. Koch. Animal Models of Heart Failure. *Circ. Res.* **2012**, *111* (1), 131–150.

(6) Riehle, C.; Bauersachs, J. Small Animal Models of Heart Failure. *Cardiovasc. Res.* **2019**, *115* (13), 1838–1849. https://doi.org/10.1093/cvr/cvz161.

(7) Thum, T.; Gross, C.; Fiedler, J.; Fischer, T.; Kissler, S.; Bussen, M.; Galuppo, P.; Just, S.; Rottbauer, W.; Frantz, S.; Castoldi, M.; Soutschek, J.; Koteliansky, V.; Rosenwald, A.; Basson, M. A.; Licht, J. D.; Pena, J. T. R.; Rouhanifard, S. H.; Muckenthaler, M. U.; Tuschl, T.; Martin, G. R.; Bauersachs, J.; Engelhardt, S. MicroRNA-21 Contributes to Myocardial Disease by Stimulating MAP Kinase Signalling in Fibroblasts. *Nature* **2008**, *456* (7224), 980–984. https://doi.org/10.1038/nature07511.

(8) Mc Namara, K.; Alzubaidi, H.; Jackson, J. K. Cardiovascular Disease as a Leading Cause of Death: How Are Pharmacists Getting Involved? *Integr. Pharm. Res. Pract.* **2019**, *Volume 8*, 1–11. https://doi.org/10.2147/IPRP.S133088.

(9) Lara-Pezzi, E.; Dopazo, A.; Manzanares, M. Understanding Cardiovascular Disease: A Journey through the Genome (and What We Found There). *Dis. Model. Mech.* **2012**, *5* (4), 434–443. https://doi.org/10.1242/dmm.009787.

(10) Lindskog, C.; Linné, J.; Fagerberg, L.; Hallström, B. M.; Sundberg, C. J.; Lindholm, M.; Huss, M.; Kampf, C.; Choi, H.; Liem, D. A.; Ping, P.; Väremo, L.; Mardinoglu, A.; Nielsen, J.; Larsson, E.; Pontén, F.; Uhlén, M. The Human Cardiac

and Skeletal Muscle Proteomes Defined by Transcriptomics and Antibody-Based Profiling. *BMC Genomics* **2015**, *16* (1), 475. https://doi.org/10.1186/s12864-015-1686-y.

(11) Zhang, L.; Hoffman, J. A.; Ruoslahti, E. Molecular Profiling of Heart Endothelial Cells. *Circulation* **2005**, *112* (11), 1601–1611. https://doi.org/10.1161/CIRCULATIONAHA.104.529537.

(12) *Protein Atlas*. Protein Atlas. https://www.proteinatlas.org/.

(13) *Uniprot*. Uniprot. https://www.uniprot.org/.

(14) *TMHMM*. TMHMM. https://services.healthtech.dtu.dk/services/TMHMM-2.0/.

(15) *PDB*. Protein Data Bank. https://www.rcsb.org/.

(16) *BLAST*. BLAST. https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome.

(17) 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.

(18) 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. **2020**.

(19) Knappe, S.; Wu, F.; Masikat, M. R.; Morser, J.; Wu, Q. Functional Analysis of the Transmembrane Domain and Activation Cleavage of Human Corin. *J. Biol. Chem.* **2003**, *278* (52), 52363–52370. https://doi.org/10.1074/jbc.M309991200.

(20) Yan, W.; Wu, F.; Morser, J.; Wu, Q. Corin, a Transmembrane Cardiac Serine Protease, Acts as a pro-Atrial Natriuretic Peptide-Converting Enzyme. *Proc. Natl. Acad. Sci.* **2000**, *97* (15), 8525–8529. https://doi.org/10.1073/pnas.150149097.

(21) Semenov, A. G.; Tamm, N. N.; Seferian, K. R.; Postnikov, A. B.; Karpova, N. S.; Serebryanaya, D. V.; Koshkina, E. V.; Krasnoselsky, M. I.; Katrukha, A. G. Processing of Pro–B-Type Natriuretic Peptide: Furin and Corin as Candidate Convertases2. *Clin. Chem.* **2010**, *56* (7), 1166–1176. https://doi.org/10.1373/clinchem.2010.143883.

(22) Peng, J.; Jiang, J.; Wang, W.; Qi, X.; Sun, X.-L.; Wu, Q. Glycosylation and Processing of Pro-B-Type Natriuretic Peptide in Cardiomyocytes. *Biochem. Biophys. Res. Commun.* **2011**, *411* (3), 593–598. https://doi.org/10.1016/j.bbrc.2011.06.192.

(23) Cui, Y.; Wang, W.; Dong, N.; Lou, J.; Srinivasan, D. K.; Cheng, W.; Huang, X.; Liu, M.; Fang, C.; Peng, J.; Chen, S.; Wu, S.; Liu, Z.; Dong, L.; Zhou, Y.; Wu, Q. Role of Corin in Trophoblast Invasion and Uterine Spiral Artery Remodelling in Pregnancy. *Nature* **2012**, *484* (7393), 246–250. https://doi.org/10.1038/nature10897.

(24) Dong, N.; Chen, S.; Yang, J.; He, L.; Liu, P.; Zheng, D.; Li, L.; Zhou, Y.; Ruan, C.; Plow, E.; Wu, Q. Plasma Soluble Corin in Patients With Heart Failure. *Circ. Heart Fail.* **2010**, *3* (2), 207–211. https://doi.org/10.1161/CIRCHEARTFAILURE.109.903849.

(25) Jiang, J.; Wu, S.; Wang, W.; Chen, S.; Peng, J.; Zhang, X.; Wu, Q. Ectodomain Shedding and Autocleavage of the Cardiac Membrane Protease Corin. *J. Biol. Chem.* **2011**, *286* (12), 10066–10072. https://doi.org/10.1074/jbc.M110.185082.

(26) Wu, J.; Yan, Z.; Li, Z.; Qian, X.; Lu, S.; Dong, M.; Zhou, Q.; Yan, N. Structure of the Voltage-Gated Calcium Channel Cav1.1 at 3.6 Å Resolution. *Nature* **2016**, *537* (7619), 191–196. https://doi.org/10.1038/nature19321.

(27) *AlphaFold*. AlphaFold Protein Structure Database. https://alphafold.ebi.ac.uk/.

(28) Okamoto, R.; Mandal, K.; Ling, M.; Luster, A. D.; Kajihara, Y.; Kent, S. B. H. Total Chemical Synthesis and Biological Activities of Glycosylated and Non-Glycosylated Forms of the Chemokines CCL1 and Ser-CCL1. *Angew. Chem.* **2014**, *126* (20), 5288–5293. https://doi.org/10.1002/ange.201310574.

(29) Emil Fischer; Ernest Fourneau. Ueber Einige Derivate Des Glykocolls. *Eur. J. Inorg. Chem. 34* (2), 2868–2877.

(30) 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.

(31) R. B. Merrifield. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc. 85* (14), 2149–2154.

(32) 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.

(33) 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.

(34) Azin Saebi; Joseph S. Brown; Victoria M. Marando; Nina Hartrampf; Nicole M. Chumbler; Stephanie Hanna; Mackenzie Poskus; Andrei Loas; Laura L. Kiessling; Deborah T. Hung; Bradley L. Pentelute. Rapid Single-Shot Synthesis of the 214

Amino Acid-Long N-Terminal Domain of Pyocin S2. *ACS Chem. Biol. 18* (3), 518–527.

(35) Clackson, T.; Wells, J. A. In Vitro Selection from Protein and Peptide Libraries. *Trends Biotechnol.* **1994**, *12* (5), 173–184. https://doi.org/10.1016/0167-7799(94)90079-5.

(36) Wilson, D. S.; Keefe, A. D.; Szostak, J. W. The Use of MRNA Display to Select High-Affinity Protein-Binding Peptides. *Proc. Natl. Acad. Sci.* **2001**, *98* (7), 3750–3755. https://doi.org/10.1073/pnas.061028198.

(37) Obexer, R.; Walport, L. J.; Suga, H. Exploring Sequence Space: Harnessing Chemical and Biological Diversity towards New Peptide Leads. *Curr. Opin. Chem. Biol.* **2017**, *38*, 52–61. https://doi.org/10.1016/j.cbpa.2017.02.020.

(38) Katoh, T.; Tajima, K.; Suga, H. Consecutive Elongation of D-Amino Acids in Translation. *Cell Chem. Biol.* **2017**, *24* (1), 46–54. https://doi.org/10.1016/j.chembiol.2016.11.012.

(39) Takayuki Katoh; Hiroaki Suga. Ribosomal Incorporation of Consecutive β-Amino Acids. *J. Am. Chem. Soc.* **2018**, *140* (38), 12159–12167.

(40) Gates, Z. P.; Vinogradov, A. A.; Quartararo, A. J.; Bandyopadhyay, A.; Choo, Z.-N.; Evans, E. D.; Halloran, K. H.; Mijalis, A. J.; Mong, S. K.; Simon, M. D.; Standley, E. A.; Styduhar, E. D.; Tasker, S. Z.; Touti, F.; Weber, J. M.; Wilson, J. L.; Jamison, T. F.; Pentelute, B. L. Xenoprotein Engineering via Synthetic Libraries. *Proc. Natl. Acad. Sci.* **2018**, *115* (23). https://doi.org/10.1073/pnas.1722633115.

(41) 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.

(42) Alexander A. Vinogradov; Zachary P. Gates; Chi Zhang; Anthony J. Quartararo; Kathryn H. Halloran; Bradley L. Pentelute. Library Design-Facilitated High-Throughput Sequencing of Synthetic Peptide Libraries. *ACS Comb. Sci. 19* (11), September 11, 2017.

(43) Lam, J. K. W.; Chow, M. Y. T.; Zhang, Y.; Leung, S. W. S. SiRNA Versus MiRNA as Therapeutics for Gene Silencing. *Mol. Ther. - Nucleic Acids* **2015**, *4*, e252. https://doi.org/10.1038/mtna.2015.23.

(44) Zhang, G.; Brown, J. S.; Quartararo, A. J.; Li, C.; Tan, X.; Hanna, S.; Antilla, S.; Cowfer, A. E.; Loas, A.; Pentelute, B. L. Rapid de Novo Discovery of Peptidomimetic Affinity Reagents for Human Angiotensin Converting Enzyme 2. *Commun. Chem.* **2022**, *5* (1), 8. https://doi.org/10.1038/s42004-022-00625-3.

(45) Liu, N.; Olson, E. N. MicroRNA Regulatory Networks in Cardiovascular Development. *Dev. Cell* **2010**, *18* (4), 510–525. https://doi.org/10.1016/j.devcel.2010.03.010.

(46) Jin, H. Y.; Xiao, C. MicroRNA Mechanisms of Action: What Have We Learned from Mice? *Front. Genet.* **2015**, *6*. https://doi.org/10.3389/fgene.2015.00328.

(47) Lucas, T.; Bonauer, A.; Dimmeler, S. RNA Therapeutics in Cardiovascular Disease. *Circ. Res.* **2018**, *123* (2), 205–220. https://doi.org/10.1161/CIRCRESAHA.117.311311.

(48) He, L.; Hannon, G. J. MicroRNAs: Small RNAs with a Big Role in Gene Regulation. *Nat. Rev. Genet.* **2004**, *5* (7), 522–531. https://doi.org/10.1038/nrg1379.

(49) Small, E. M.; Olson, E. N. Pervasive Roles of MicroRNAs in Cardiovascular Biology. *Nature* **2011**, *469* (7330), 336–342. https://doi.org/10.1038/nature09783.

(50) *miRBase: the microRNA database*. https://www.mirbase.org/ (accessed 2022-01-28).

(51) Xu, Y.; Hajdukiewicz, K.; Tiwari, A.; Przybys, J.; Parkitna, J. R.; Martin Novak; Ilya A. Vinnikov; Gunther Schutz; Witold Konopka. Micro RNAs Are Indispensable for the Proliferation and Differentiation of Adult Neural Progenitor Cells in Mice. *Biochem. Biophys. Res.* **2020**, *530* (1), 209–214. https://doi.org/10.1016/j.bbrc.2020.06.143.

(52) Zhang, Q.; Zhang, K.; Zhang, C.; Ge, H.; Yin, Y.; Feng, H.; Hu, R. MicroRNAs as Big Regulators of Neural Stem/Progenitor Cell Proliferation, Differentiation and Migration: A Potential Treatment for Stroke. *Curr. Pharm. Des.* **2017**, *23* (15), 2252–2257. https://doi.org/10.2174/1381612823666170228124657.

(53) Kim, D. Y.; Sung, J.-H. Regulatory Role of MicroRNAs in the Proliferation and Differentiation of Adipose-Derived Stem Cells. *Histol. Histopathol.* **2017**, *32* (1), 1–10. https://doi.org/10.14670/HH-11-798.

(54) Bezman, N. A.; Cedars, E.; Steiner, D. F.; Blelloch, R.; Hesslein, D. G. T.; Lanier, L. L. Distinct Requirements of MicroRNAs in NK Cell Activation, Survival, and Function. *J. Immunol.* **2010**, *185* (7), 3835–3846. https://doi.org/10.4049/jimmunol.1000980.

(55) Lin, J.; Zhao, J.; Lwin, T.; Luis, C.; Guan, F.; Dessureault, S.; Moscinski, L. C.; Dalton, W. S.; Sotomayor, E. M.; Cheng, J.; Tao, J. MiR-29 MicroRNAs Regulate IGF-1R Expression and Contribute Mantle Cell Lymphoma Growth and Survival. *Blood* **2009**, *114* (22), 1957. https://doi.org/10.1182/blood.V114.22.1957.1957.

(56) Montgomery, R. L.; Hullinger, T. G.; Semus, H. M.; Dickinson, B. A.; Seto, A. G.; Lynch, J. M.; Stack, C.; Latimer, P. A.; Olson, E. N.; van Rooij, E. Therapeutic Inhibition of MiR-208a Improves Cardiac Function and Survival During Heart Failure.

Circulation **2011**, *124* (14), 1537–1547. https://doi.org/10.1161/CIRCULATIONAHA.111.030932.

(57) Maegdefessel, L.; Azuma, J.; Toh, R.; Merk, D. R.; Deng, A.; Chin, J. T.; Raaz, U.; Schoelmerich, A. M.; Raiesdana, A.; Leeper, N. J.; McConnell, M. V.; Dalman, R. L.; Spin, J. M.; Tsao, P. S. Inhibition of MicroRNA-29b Reduces Murine Abdominal Aortic Aneurysm Development. *J. Clin. Invest.* **2012**, *122* (2), 497–506. https://doi.org/10.1172/JCI61598.

(58) Lanford, R. E.; Hildebrandt-Eriksen, E. S.; Petri, A.; Persson, R.; Lindow, M.; Munk, M. E.; Kauppinen, S.; Orum, H. Therapeutic Silencing of MicroRNA-122 in Primates with Chronic Hepatitis C Virus Infection. *Science* **2009**, *327* (5962), 198–201. https://doi.org/10.1126/science.1178178.

(59) Lin, Z.; Murtaza, I.; Wang, K.; Jiao, J.; Gao, J.; Li, P.-F. MiR-23a Functions Downstream of NFATc3 to Regulate Cardiac Hypertrophy. *Proc. Natl. Acad. Sci.* **2009**, *106* (29), 12103–12108. https://doi.org/10.1073/pnas.0811371106.

(60) Heymans, S.; Corsten, M. F.; Verhesen, W.; Carai, P.; van Leeuwen, R. E. W.; Custers, K.; Peters, T.; Hazebroek, M.; Stöger, L.; Wijnands, E.; Janssen, B. J.; Creemers, E. E.; Pinto, Y. M.; Grimm, D.; Schürmann, N.; Vigorito, E.; Thum, T.; Stassen, F.; Yin, X.; Mayr, M.; de Windt, L. J.; Lutgens, E.; Wouters, K.; de Winther, M. P. J.; Zacchigna, S.; Giacca, M.; van Bilsen, M.; Papageorgiou, A.-P.; Schroen, B. Macrophage MicroRNA-155 Promotes Cardiac Hypertrophy and Failure. *Circulation* **2013**, *128* (13), 1420–1432. https://doi.org/10.1161/CIRCULATIONAHA.112.001357.

(61) Chu, R.; Mo, G.; Duan, Z.; Huang, M.; Chang, J.; Li, X.; Liu, P. MiRNAs Affect the Development of Hepatocellular Carcinoma via Dysregulation of Their Biogenesis and Expression. **2014**, 12.

(62) Pereira, J. C.; Costa, J. O.; Costa, C. C. P.; Lima, N. S.; Barros, J. B. S.; Bento, D. C. P.; Reis, A. A. S.; Santos, R. S. Research Article New Insights of MiRNAs Dysregulation in the Molecular Pathological Basis of Neurodegenerative Sclerosis: A Systematic Review. *Genet. Mol. Res.* **2021**, *20* (3). https://doi.org/10.4238/gmr18843.

(63) van Rooij, E.; Sutherland, L. B.; Qi, X.; Richardson, J. A.; Hill, J.; Olson, E. N. Control of Stress-Dependent Cardiac Growth and Gene Expression by a MicroRNA. *Science* **2007**, *316* (5824), 575–579. https://doi.org/10.1126/science.1139089.

(64) Iannolo, G.; Sciuto, M. R.; Raffa, G. M.; Pilato, M.; Conaldi, P. G. MiR34 Inhibition Induces Human Heart Progenitor Proliferation. *Cell Death Dis.* **2018**, *9* (3), 368. https://doi.org/10.1038/s41419-018-0400-9.

(65) Krützfeldt, J.; Rajewsky, N.; Braich, R.; Rajeev, K. G.; Tuschl, T.; Manoharan, M.; Stoffel, M. Silencing of MicroRNAs in Vivo with 'Antagomirs.' *Nature* **2005**, *438* (7068), 685–689. https://doi.org/10.1038/nature04303.

(66) Elmén, J.; Lindow, M.; Silahtaroglu, A.; Bak, M.; Christensen, M.; Lind-Thomsen, A.; Hedtjärn, M.; Hansen, J. B.; Hansen, H. F.; Straarup, E. M.; McCullagh, K.; Kearney, P.; Kauppinen, S. Antagonism of MicroRNA-122 in Mice by Systemically Administered LNA-AntimiR Leads to up-Regulation of a Large Set of Predicted Target MRNAs in the Liver. *Nucleic Acids Res.* **2008**, *36* (4), 1153–1162. https://doi.org/10.1093/nar/gkm1113.

(67) Krützfeldt, J.; Kuwajima, S.; Braich, R.; Rajeev, K. G.; Pena, J.; Tuschl, T.; Manoharan, M.; Stoffel, M. Specificity, Duplex Degradation and Subcellular Localization of Antagomirs. *Nucleic Acids Res.* **2007**, *35* (9), 2885–2892. https://doi.org/10.1093/nar/gkm024.

(68) Li, C.; Callahan, A. J.; Simon, M. D.; Totaro, K. A.; Mijalis, A. J.; Phadke, K.-S.; Zhang, G.; Hartrampf, N.; Schissel, C. K.; Zhou, M.; Zong, H.; Hanson, G. J.; Loas, A.; Pohl, N. L. B.; Verhoeven, D. E.; Pentelute, B. L. Fully Automated Fast-Flow Synthesis of Antisense Phosphorodiamidate Morpholino Oligomers. *Nat. Commun.* **2021**, *12* (1), 4396. https://doi.org/10.1038/s41467-021-24598-4.

(69) Sardone, V.; Zhou, H.; Muntoni, F.; Ferlini, A.; Falzarano, M. Antisense Oligonucleotide-Based Therapy for Neuromuscular Disease. *Molecules* **2017**, *22* (4), 563. https://doi.org/10.3390/molecules22040563.

(70) Scoles, D. R.; Minikel, E. V.; Pulst, S. M. Antisense Oligonucleotides: A Primer. *Neurol. Genet.* **2019**, *5* (2), e323. https://doi.org/10.1212/NXG.0000000000000323.

(71) Demidov, V. V.; Potaman, V. N.; Frank-Kamenetskil, M. D.; Egholm, M.; Buchard, O.; Sonnichsen, S. H.; Nlelsen, P. E. Stability of Peptide Nucleic Acids in Human Serum and Cellular Extracts. *Biochem. Pharmacol.* **1994**, *48* (6), 1310–1313. https://doi.org/10.1016/0006-2952(94)90171-6.

(72) Corrie L Gallant-Behm; Joseph Piper; Brent A Dickinson; Christina M. Dalby; Linda A Pestando; Aimee L Jackson. A Synthetic MicroRNA-92a Inhibitor (MRG-110) Accelerates Angiogenesis and Wound Healing in Diabetic and Nondiabetic Wounds. *Wound Repair Regen.* **2018**, *26*, 311–323. https://doi.org/10.1111/wrr.12660.

(73) Abplanalp, W. T.; Fischer, A.; John, D.; Zeiher, A. M.; Gosgnach, W.; Darville, H.; Montgomery, R.; Pestano, L.; Allée, G.; Paty, I.; Fougerousse, F.; Dimmeler, S. Efficiency and Target Derepression of Anti-MiR-92a: Results of a First in Human Study. *Nucleic Acid Ther.* **2020**, *30* (6), 335–345. https://doi.org/10.1089/nat.2020.0871.

(74) Täubel, J.; Hauke, W.; Rump, S.; Viereck, J.; Batkai, S.; Poetzsch, J.; Rode, L.; Weigt, H.; Genschel, C.; Lorch, U.; Theek, C.; Levin, A. A.; Bauersachs, J.; Solomon, S. D.; Thum, T. Novel Antisense Therapy Targeting MicroRNA-132 in Patients with Heart Failure: Results of a First-in-Human Phase 1b Randomized, Double-Blind, Placebo-Controlled Study. *Eur. Heart J.* **2021**, *42* (2), 178–188. https://doi.org/10.1093/eurheartj/ehaa898.

(75) Baker, A. H.; Giacca, M. Antagonism of MiRNA in Heart Failure: First Evidence in Human. *Eur. Heart J.* **2021**, *42* (2), 189–191. https://doi.org/10.1093/eurheartj/ehaa967.

(76) Nair, J. K.; Willoughby, J. L. S.; Chan, A.; Charisse, K.; Alam, Md. R.; Wang, Q.; Hoekstra, M.; Kandasamy, P.; Kel'in, A. V.; Milstein, S.; Taneja, N.; O'Shea, J.; Shaikh, S.; Zhang, L.; Sluis, R. J. van der; Jung, M. E.; Akinc, A.; Hutabarat, R.; Kuchimanchi, S.; Fitzgerald, K.; Zimmerman, T.; Berkel, T. J. C. van; Maier, M. A.; Rajeev, K. G.; Manoharan, M. Multivalent N-Acetylgalactosamine-Conjugated SiRNA Localizes in Hepatocytes and Elicits Robust RNAi-Mediated Gene Silencing. *J. Am. Chem. Soc.* **2014**, *136* (49), 16958–16961. https://doi.org/10.1021/ja505986a.

(77) Brown, K. M.; Nair, J. K.; Janas, M. M.; Anglero-Rodriguez, Y. I.; Dang, L. T. H.; Peng, H.; Theile, C. S.; Castellanos-Rizaldos, E.; Brown, C.; Foster, D.; Kurz, J.; Allen, J.; Maganti, R.; Li, J.; Matsuda, S.; Stricos, M.; Chickering, T.; Jung, M.; Wassarman, K.; Rollins, J.; Woods, L.; Kelin, A.; Guenther, D. C.; Mobley, M. W.; Petrulis, J.; McDougall, R.; Racie, T.; Bombardier, J.; Cha, D.; Agarwal, S.; Johnson, L.; Jiang, Y.; Lentini, S.; Gilbert, J.; Nguyen, T.; Chigas, S.; LeBlanc, S.; Poreci, U.; Kasper, A.; Rogers, A. B.; Chong, S.; Davis, W.; Sutherland, J. E.; Castoreno, A.; Milstein, S.; Schlegel, M. K.; Zlatev, I.; Charisse, K.; Keating, M.; Manoharan, M.; Fitzgerald, K.; Wu, J.-T.; Maier, M. A.; Jadhav, V. Expanding RNAi Therapeutics to Extrahepatic Tissues with Lipophilic Conjugates. *Nat. Biotechnol.* **2022**. https://doi.org/10.1038/s41587-022-01334-x.

(78) Avidity Biosciences. Delivering on the RNA Revolution, 2022.

(79) Desjardins, C. A.; Yao, M.; Hall, J.; O'Donnell, E.; Venkatesan, R.; Spring, S.; Wen, A.; Hsia, N.; Shen, P.; Russo, R.; Lan, B.; Picariello, T.; Tang, K.; Weeden, T.; Zanotti, S.; Subramanian, R.; Ibraghimov-Beskrovnaya, O. Enhanced Exon Skipping and Prolonged Dystrophin Restoration Achieved by TfR1-Targeted Delivery of Antisense Oligonucleotide Using FORCE Conjugation in *Mdx* Mice. *Nucleic Acids Res.* **2022**, gkac641. https://doi.org/10.1093/nar/gkac641.

(80) Ingram, D.; Rodino-Klapac, L. SRP-9001: New Clinical Data and Integrated Analysis, 2022.

(81) Li, H.; Sun, H.; Qian, Z. M. The Role of the Transferrin–Transferrin-Receptor System in Drug Delivery and Targeting. *Trends Pharmacol. Sci.* **2002**, *23* (5), 206–209. https://doi.org/10.1016/S0165-6147(02)01989-2.

(82) Sugo, T.; Terada, M.; Oikawa, T.; Miyata, K.; Nishimura, S.; Kenjo, E.; Ogasawara-Shimizu, M.; Makita, Y.; Imaichi, S.; Murata, S.; Otake, K.; Kikushi, K.; Teratani, M.; Masuda, Y.; Kamei, T.; Takagahara, S.; Ikeda, S.; Ohtaki, T.; Matsumoto, H. Development of Antibody-SiRNA Conjugate Targeted to Cardiac and Skeletal Muscles. *J. Controlled Release* **2016**, *237*, 1–13. https://doi.org/10.1016/j.jconrel.2016.06.036.

(83) Aro Biotherapeutics. Centyrin-Targeted SiRNA Conjugates Demonstrate Potential New Therapeutic Approach for Reduction of Skeletal Muscle Glycogen in Pompe Disease, 2022.

(84) Ionis Pharmaceuticals; Bicycle Therapeutics. Bicycles, Bi-Cyclic Peptides, Novel Small Molecule Delivery Systems for RNA Therapeutics, 2021.

(85) Lee, J. H.; Engler, J. A.; Collawn, J. F.; Moore, B. A. Receptor Mediated Uptake of Peptides That Bind the Human Transferrin Receptor: Peptide Endocytosis by the Transferrin Receptor. *Eur. J. Biochem.* **2001**, *268* (7), 2004–2012. https://doi.org/10.1046/j.1432-1327.2001.02073.x.

(86) Li, C.; Callahan, A. J.; Phadke, K. S.; Bellaire, B.; Farquhar, C. E.; Zhang, G.; Schissel, C. K.; Mijalis, A. J.; Hartrampf, N.; Loas, A.; Verhoeven, D. E.; Pentelute, B. L. Automated Flow Synthesis of Peptide–PNA Conjugates. *ACS Cent. Sci.* **2022**, *8* (2), 205–213. https://doi.org/10.1021/acscentsci.1c01019.

(87) Claycomb, W. C.; Lanson, N. A.; Stallworth, B. S.; Egeland, D. B.; Delcarpio, J. B.; Bahinski, A.; Izzo, N. J. HL-1 Cells: A Cardiac Muscle Cell Line That Contracts and Retains Phenotypic Characteristics of the Adult Cardiomyocyte. *Proc. Natl. Acad. Sci.* **1998**, *95* (6), 2979–2984. https://doi.org/10.1073/pnas.95.6.2979.

(88) Schwarzenbach, H.; da Silva, A. M.; Calin, G.; Pantel, K. Data Normalization Strategies for MicroRNA Quantification. *Clin. Chem.* **2015**, *61* (11), 1333–1342. https://doi.org/10.1373/clinchem.2015.239459.

(89) Schmittgen, T. D.; Livak, K. J. Analyzing Real-Time PCR Data by the Comparative CT Method. *Nat. Protoc.* **2008**, *3* (6), 1101–1108. https://doi.org/10.1038/nprot.2008.73.

(90) Roberts, T. C.; Langer, R.; Wood, M. J. A. Advances in Oligonucleotide Drug Delivery. *Nat. Rev. Drug Discov.* **2020**, *19* (10), 673–694. https://doi.org/10.1038/s41573-020-0075-7.

(91) Crooke, S. T.; Wang, S.; Vickers, T. A.; Shen, W.; Liang, X. Cellular Uptake and Trafficking of Antisense Oligonucleotides. *Nat. Biotechnol.* **2017**, *35* (3), 230–237. https://doi.org/10.1038/nbt.3779.

(92) Nina Hartrampf; Azin Saebi; Mackenzie Poskus; Alex J. Callahan; Amanda E. Cowfer; Stephanie Hanna; Sarah Antilla; Carly K. Schissel; Anthony J. Quartararo; Xiyun Ye; Alex J. Mijalis; Mark D. Simon; Andrei Loas; Shunying Liu; Carsten Jessen; Thomas E. Nielsen; Bradley L. Pentelute. Synthesis of Proteins by Automated Flow Chemistry. *Science* **2020**, *368*, 980–987.

(93) Xie, J.; Bi, Y.; Zhang, H.; Dong, S.; Teng, L.; Lee, R. J.; Yang, Z. Cell-Penetrating Peptides in Diagnosis and Treatment of Human Diseases: From Preclinical Research to Clinical Application. *Front. Pharmacol.* **2020**, *11*, 697. https://doi.org/10.3389/fphar.2020.0069.

Appendix I Synthesis of proteins by automated flow chemistry

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.; 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, 980-987.

**These authors contributed equally to this work.*

The work presented in this appendix was a team effort and is intended to supplement and contextualize Chapter 1.3.

A1.1 Introduction

Mechanical pumps, valves, solid supports, and computers have transformed the way we perform chemical reactions. Continuous, multistep flow technology has enabled routine access to small molecules ranging from pharmaceutical ingredients to natural products and bulk commodities.[1] Advantages of flow synthesis over batch methods are in-line spectroscopic monitoring, efficient mixing, and precise control over the reaction parameters.[2] Translating these capabilities to the total chemical synthesis of peptides and proteins will provide rapid access to an expanded chemical space. Protein production is an essential part of research in academia and industry and can be accomplished by biological methods or chemical synthesis. Most proteins are obtained by biological expression, a process that largely limits their chemical composition to the canonical proteinogenic amino acids.[3] Advances in genetic code expansion have allowed for the incorporation of up to two unnatural amino acids in the structures of native proteins.[4] By contrast, chemical syn-thesis offers unmatched flexibility when incorporation of multiple unnatural amino acids, post-translational modifications, or artificial backbones is desired.[3] Synthetic proteins have become accessible with a combination of solid-phase and ligation methodologies. Yet, total chemical synthesis of proteins remains highly labor intensive. Solid-phase peptide synthesis (SPPS) is the foundation of chemical peptide and protein production.[5] Despite decades of optimization, peptides longer than 50 amino acids are difficult to synthesize with standard SPPS instrumentation, owing in large part to generation of by-products from deletion, truncation, and aggregation of the growing chains.[6,7] It was not until the development of native chemical ligation (NCL) that chemical synthesis of protein chains

became practical.[8] Despite the efforts dedicated to improving NCL techniques,[9] a major bottleneck resides in the absence of a routine, widely applicable protocol to access the requisite peptide fragments.[10,11] We set out to address this problem by developing a reliable method to synthesize long peptides and protein chains using flow chemistry. Flow-based SPPS is gaining momentum owing to its advantageous features— for example, control over physical parameters and greatly reduced formation of side products.[12-14] Studies carried out as early as 1970 found that automation and high fidelity of peptide synthesis could be achieved by containing the solid support in a reactor and operating it as a fixed bed.[15,16] Instead of complex systems for liquid handling to dispense reagents and wash the resin, high-performance liquid chromatography (HPLC) pumps were used to continuously deliver reagents, establishing the principles of peptide synthesis in flow. Inspired by this early work, over the past 5 years we developed rapid, automated fast-flow peptide synthesis (AFPS) instrumentation that incorporates amino acid residues in as little as 40 s at temperatures up to 90°C.[17-19] Even though prior work by us and others on flow-based SPPS considerably reduced the total synthesis time, the potential of flow chemistry to enable synthesis of peptide chains in the range of single-domain proteins has not been fully realized.[17-23] We set out to optimize our AFPS technology to meet this challenge.[19] We report here a routine protocol that allows for stepwise chemical total synthesis of peptide chains exceeding 50 amino acids in length, with a cycle time of ~2.5 min per amino acid (Fig. A1.1A). The optimized protocol was built on a collection of analytical data acquired with an AFPS system and delivers products with high fidelity and of high chiral purity. Using this protocol, single-domain protein chains ranging from barstar (90

amino acids) to sortase $_{A59–206}$ (sortase A*, 164 amino acids) were synthesized in 3.5 to 6.5 hours. To demonstrate production of functional proteins, these sequences were folded, and their biophysical properties and enzymatic activities were determined. The time scale of chemical protein synthesis is on par with that of recombinant expression and therefore offers a practical alternative to biochemical methods while opening up the chemical space beyond canonical amino acids.

A1.2 Rapid screening of reaction variables for AFPS protocol development

We chose to first optimize coupling efficiency and later investigate possible side reactions induced by the optimized coupling conditions. On a benchmark AFPS instrument previously developed in our laboratory,[17,19] reagents are mixed, heated, and delivered onto a pre-tempered solid support using three HPLC pumps. In-line ultraviolet-visible (UV-vis) detection of the reactor eluent is used to monitor removal of the N-terminal protecting group after each coupling cycle. Indirectly, this information reports on the efficiency of the preceding coupling step. We first optimized general parameters, including flow rate, reaction solvent, reagent concentration, temperature, and coupling agents (Fig. A1.1B and tables A1.1 to A1.7). Modifications to our original AFPS protocol included increasing reagent concentrations to 0.4 M,[24] the use of amine-free *N,N*-dimethylformamide (DMF), and an increase in temperature to 85°to 90°C for reagent activation and coupling. The performance of different activators for the coupling step was also investigated, identifying the azabenzotriazol-reagents PyAOP [(7-azabenzotriazol-1-yloxy)tripyrrolidino-phosphonium hexafluorophosphate] and HATU (hexafluorophosphate azabenzotriazole tetra-methyl uranium) as optimal. Automated

collection and analysis of data combined with the synthesis parameters allowed for optimization of residue-specific coupling conditions. By comparing data on amino acid deprotections, we were able to gain information on coupling efficiency for all canonical amino acids and generated a general amino acid–specific recipe (tables A1.8 and A1.9). Analytical comparison of the products obtained for glucagon-like peptide-1 (GLP-1) is illustrative of the improvement in crude peptide quality achieved with the optimized synthesis conditions (Fig. A1.1B).

We aimed to suppress aspartimide formation, a major side reaction in SPPS and AFPS. Because increased temperature leads to more aspartimide formation, various deprotection bases, additives, and aspartic acid protecting groups were screened to minimize this unwanted side reaction.[25,26] We found that milder deprotection bases [i.e., piperazine and 1-hydroxybenzotriazole (HOBt) with piperidine] and bulky aspartic acid protecting groups [i.e., O-3-methylpent-3-yl (OMpe)] decreased the level of aspartimide formation (fig. A1.8 and table A1.10). The most effective strategies, however, were the addition of formic acid as a piperidine additive and backbone protection with dimethoxybenzyl glycine. Formic acid (2% stock solution in 40:60 v/v piperidine: DMF) was therefore used as an additive for deprotection, and backbone protection was applied for collagen and fibroblast growth factor 1 (FGF1) syntheses.

We confirmed retention of chirality for amino acids at high risk of epimerization (i.e., cysteine and histidine) in a final optimization step (figs. A1.9 to A1.14).[27] The influence of temperature, time, and activating agent, as well as different side-chain protecting groups were screened (Fig. A1.1C and D).[17] For both amino acids, epimerization increases with activation time and temperature. The choice of protecting

group proved to be critical for histidine. Ultimately, activation of Fmoc-Cys(Trt)-OH and Fmoc-His(Boc)-OH with PyAOP with a shorter time at 60°C resulted in <2% D-epimer formation (Fmoc, fluorenylmethyloxycarbonyl; Trt, trityl; Boc, tert-butyloxycarbonyl). Next, we determined that the amount of epimerization under these optimized conditions does not increase over multiple coupling cycles (Fig.A1.1E). The amount of D-isomer did not change over 100 amino acid couplings executed after manual capping of the N-terminus, 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 [table A1.11 and Section A1.9.10].

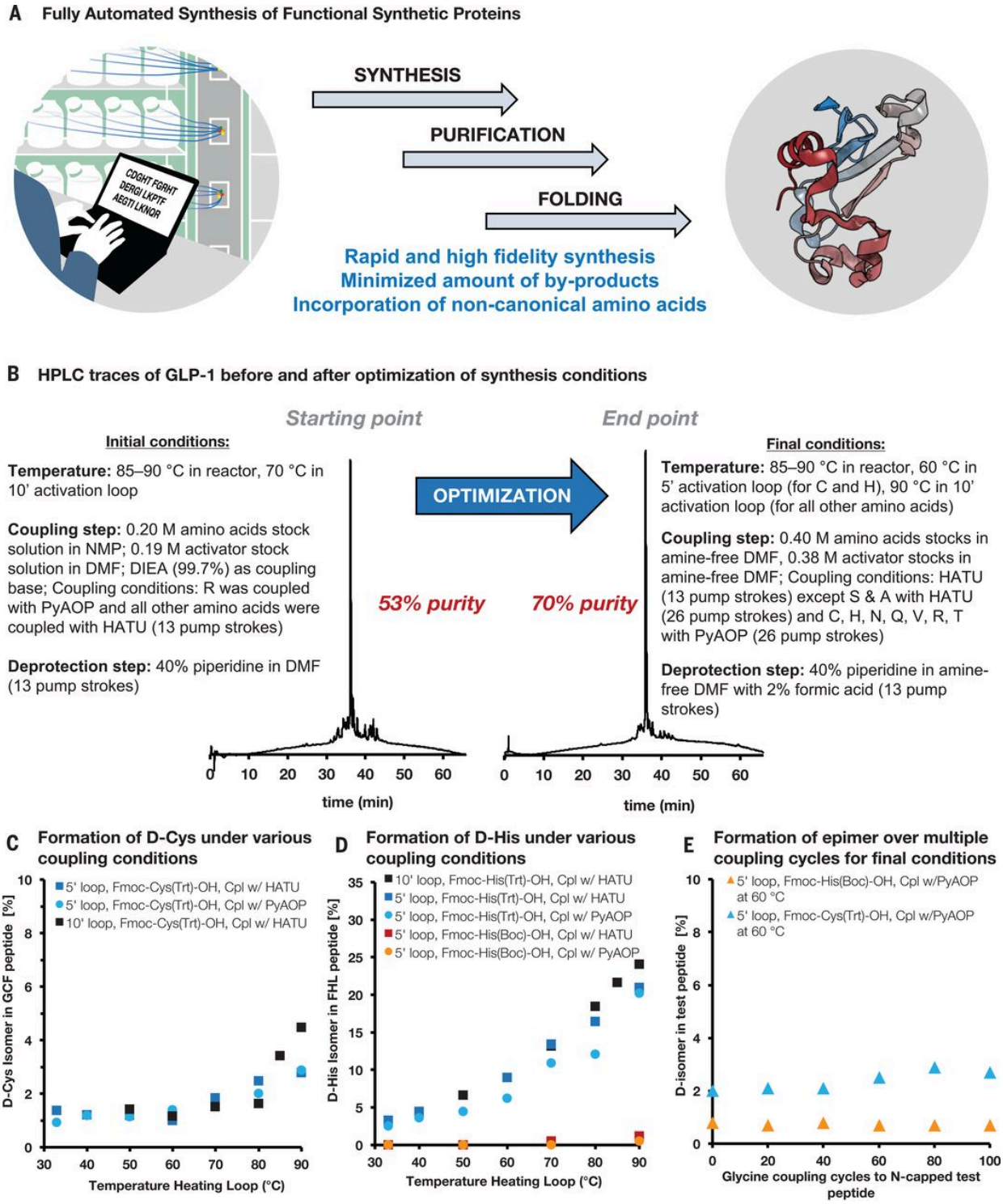


Figure A1.1. **Optimized conditions for automated fast-flow solid-phase peptide synthesis enable high-fidelity production of long amino acid sequences.** (A) Fully automated chemical flow synthesis yields peptide chains, which, after purification and folding, give functional proteins. Protein Data Bank (PDB) 1BRS (barnase)[57] was used. (B) Synthesis of GLP-1 using starting conditions and optimized conditions. The concentrations listed refer to stock solutions. NMP, N-methylpyrrolidone; DIEA,

diisopropylethylamine. (C) Quantification of cysteine epimerization as a function of activation temperature, heating time (5′ loop and 10′ loop), and activator in a GCF test peptide. Isomer was quantified from extracted ion chromatograms on LC-MS by comparison to reference peptides. (D) Quantification of histidine epimerization as a function of activation temperature, heating time (5′ loop and 10′ loop), and activator in a FHL test peptide. The d-isomer was quantified by analytical HPLC by comparison to reference peptides. (E) Quantification of epimerization over multiple coupling cycles. GCF and FHL were synthesized under optimized conditions, and the N terminus was manually capped with a Boc-protecting group. One-hundred glycine couplings were executed, and a sample was taken out for analysis every 20 amino acid couplings. Cpl w/, coupled with. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

A1.3 Optimized AFPS outperforms traditional synthesis methods

We investigated if our optimized AFPS conditions could facilitate the synthesis of longer sequences using proinsulin (86 amino acids) and human immunodeficiency virus-1 (HIV-1) protease (99 amino acids) as test sequences. The total synthesis of human proinsulin was previously reported using NCL of three peptide fragments individually prepared by SPPS.[28] HIV-1 protease was previously prepared using stepwise and chemical ligation routes under Boc-SPPS conditions.[29,30] 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%) of purified proinsulin and 5.3 mg (1%) of purified HIV-1 protease.

A comparison between AFPS and standard batch SPPS syntheses on commercially available synthesizers at room temperature, 70°C, and 90°C indicated substantially improved synthetic outcome for the optimized AFPS protocol (Fig. A1.2 and Section A1.10). On each instrument, machine-specific, optimized conditions were used to achieve the best synthesis outcome. For HIV-1 protease and proinsulin, 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). By contrast, synthesis on commercially available peptide synthesizers took approximately five times longer and resulted in a complex compound mixture. AFPS therefore offers a substantial improvement when directly compared with traditional SPPS methods, both with respect to time and performance.

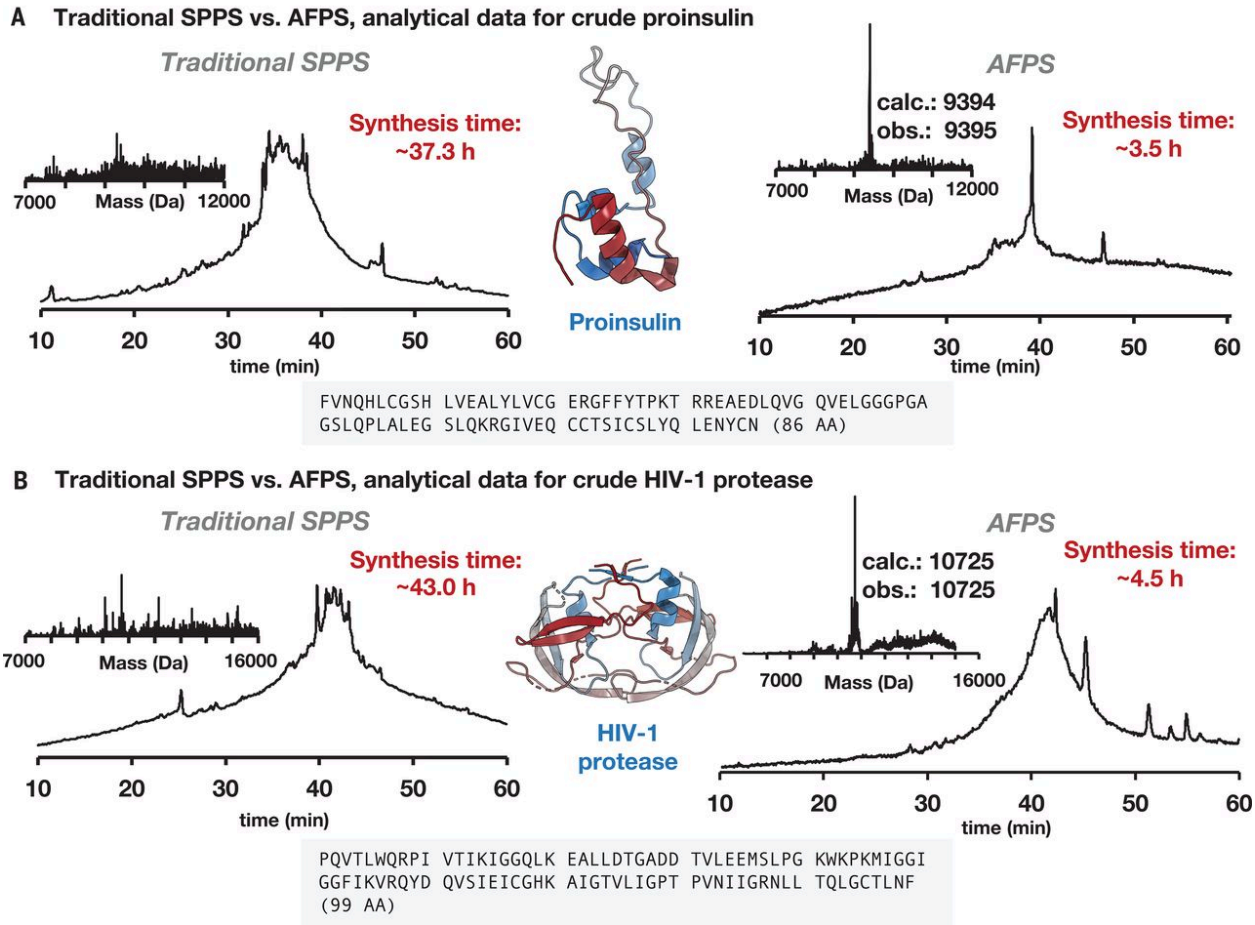


Figure A1.2. **Synthesis of proinsulin and HIV-1 protease demonstrates the advantage of AFPS over traditional SPPS methods.** (A and B) Analytical HPLC data of the crude proinsulin (A) and HIV-1 protease (B) are presented as the main chromatographic traces with absorbance detection at 214 nm (additional 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, not identical to the sequence synthesized) (30) were used. Crude material of higher quality was obtained in the case of HIV-1 protease by substitution of Cys and Met residues, as described in SM section 5. AA, amino acids.

A1.4 Optimized AFPS enables routine access to single-domain protein chains

To demonstrate general applicability of our AFPS protocol, the synthesis of additional protein chains ranging from ~70 to ~170 amino acids was performed (Fig. A1.3A and Section A1.11). These sequences were chosen to enable comparison with literature data. We chose not only historically relevant targets for drug discovery, such as HIV-1 protease and murine double minute 2 (MDM2),[31,32] but also proteins that serve as therapeutics themselves, such as FGF1 and proinsulin.[33,34] Barstar, barnase, lysozyme, MDM2, and sortase A* allowed for a direct comparison of recombinant and synthetic proteins. The ability of AFPS technology to rapidly and simultaneously incorporate noncanonical amino acids in greater number and of greater diversity than biological methods was 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.[35] Then, we produced synthetic HIV-1 protease in which two methionine and one cysteine residues were replaced as previously described with norleucine and aminobutyric acid, respectively (Fig. A1.3B), to avoid potential oxidation side products and increase synthetic efficiency.[30] 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 product. Isolated yields after HPLC

purification ranged from 2.2 to 19.0 mg (1 to 5%), a sufficient amount of material for folding and evaluation of tertiary structure and biological function (Fig. A1.3B and Secttion A1.11). In conclusion, optimized AFPS allows for the routine stepwise chemical synthesis of peptide chains of up to ~170 amino acids and therefore substantially decreases time and labor associated with the chemical production of single-domain proteins.

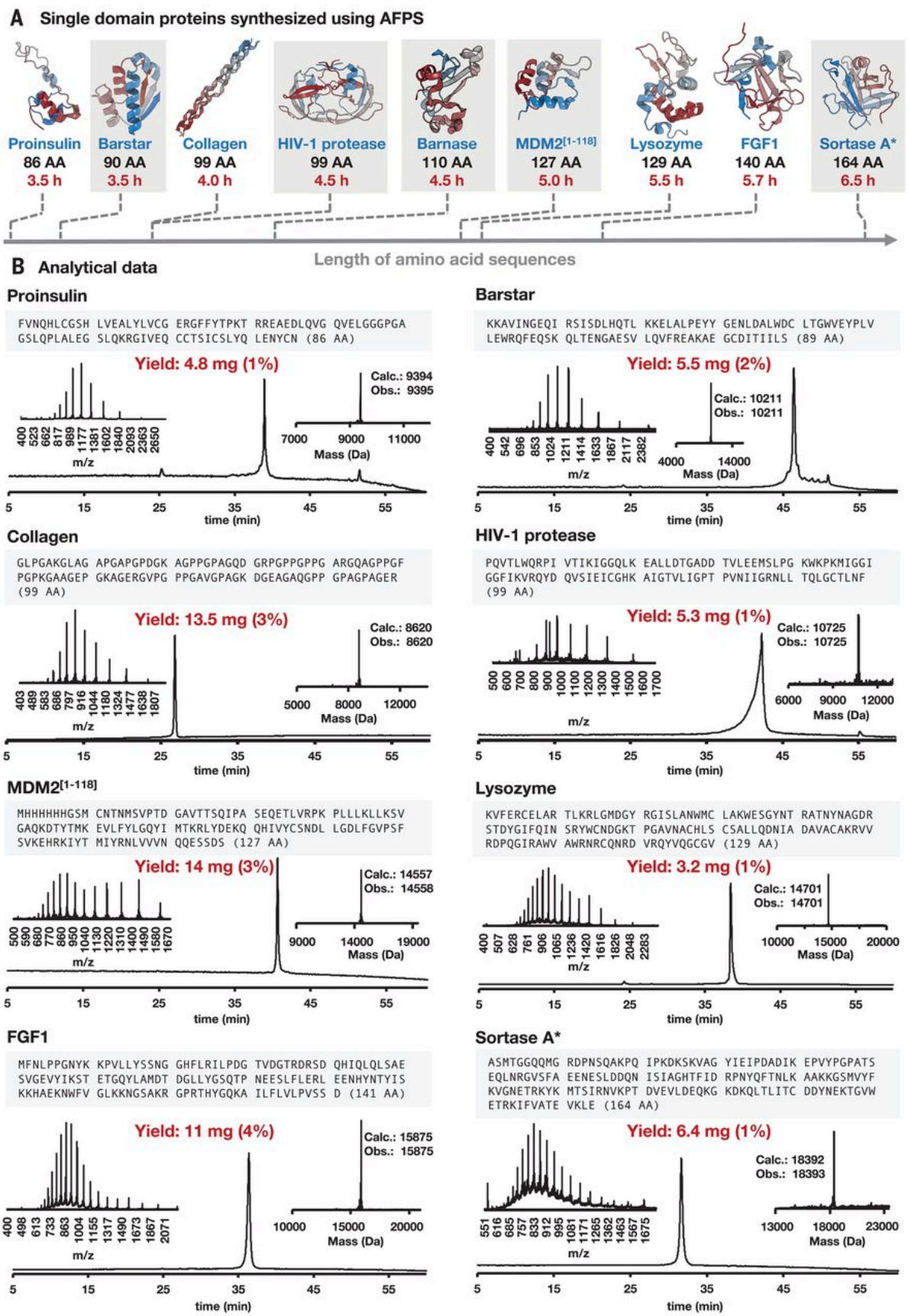

A Single domain proteins synthesized using AFPS
Proinsulin
86 AA
3.5 h
Barstar
90 AA
3.5 h
Collagen
99 AA
4.0 h
HIV-1 protease
99 AA
4.5 h
Barnase
110 AA
4.5 h
MDM2[1-118]
127 AA
5.0 h
Lysozyme
129 AA
5.5 h
FGF1
140 AA
5.7 h
Sortase A*
164 AA
6.5 h
Length of amino acid sequences
B Analytical data
Proinsulin
FVNQHLCGSH LVEALYLVCG ERGFFYTPKT RREAEDLQVG QVELGGGPGA
GSLQPLALEG SLQKRGIVEQ CCTSICSLYQ LENYCN (86 AA)
Yield: 4.8 mg (1%)
Calc.: 9394
Obs.: 9395
Barstar
KKAVINGEQI RSISDLHQTL KKELALPEYY GENLDALWDC LTGWVEYPLV
LEWRQFEQSK QLTENGAESV LQVFREAKAE GCDITIILS (89 AA)
Yield: 5.5 mg (2%)
Calc.: 10211
Obs.: 10211
Collagen
GLPGAKGLAG APGAPGPDGK AGPPGPAGQD GRPGPPGPPG ARGQAGPPGF
PGPKGAAGEP GKAGERGVPG PPGAVGPAGK DGEAGAQGPP GPAGPAGER
(99 AA)
Yield: 13.5 mg (3%)
Calc.: 8620
Obs.: 8620
HIV-1 protease
PQVTLWQRPI VTIKIGGQLK EALLDTGADD TVLEEMSLPG KWKPKMIGGI
GGFIKVRQYD QVSIEICGHK AIGTVLIGPT PVNIIGRNLL TQLGCTLNF
(99 AA)
Yield: 5.3 mg (1%)
Calc.: 10725
Obs.: 10725
MDM2[1-118]
MHHHHHHGSM CNTNMSVPTD GAVTTSQIPA SEQETLVRPK PLLLKLLKSV
GAQKDTYTMK EVLFYLGQYI MTKRLYDEKQ QHIVYCSNDL LGDLFGVPSF
SVKEHRKIYT MIYRNLVVVN QQESSDS (127 AA)
Yield: 14 mg (3%)
Calc.: 14557
Obs.: 14558
Lysozyme
KVFERCELAR TLKRLGMDGY RGISLANWMC LAKWESGYNT RATNYNAGDR
STDYGIFQIN SRYWCNDGKT PGAVNACHLS CSALLQDNIA DAVACAKRVV
RDPQGIRAWV AWRNRCQNRD VRQYVQGCGV (129 AA)
Yield: 3.2 mg (1%)
Calc.: 14701
Obs.: 14701
FGF1
MFNLPPGNYK KPVLLYSSNG GHFLRILPDG TVDGTRDRSD QHIQLQLSAE
SVGEVYIKST ETGQYLAMDT DGLLYGSQTP NEESLFLERL EENHYNTYIS
KKHAEKNWFV GLKKNGSAKR GPRTHYGQKA ILFLVLPVSS D (141 AA)
Yield: 11 mg (4%)
Calc.: 15875
Obs.: 15875
Sortase A*
ASMTGGQQMG RDPNSQAKPQ IPKDKSKVAG YIEIPDADIK EPVYPGPATS
EQLNRGVSFA EENESLDDQN ISIAGHTFID RPNYQFTNLK AAKKGSMVYF
KVGNETRKYK MTSIRNVKPT DVEVLDEQKG KDKQLTLITC DDYNEKTGVW
ETRKIFVATE VKLE (164 AA)
Yield: 6.4 mg (1%)
Calc.: 18392
Obs.: 18393
m/z
Mass (Da)
time (min)

Figure A1.3. **AFPS enables high-fidelity production of long amino acid sequences in hours.** (**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. The gradient for analytical HPLC was 5 to 65% B. A linear gradient of acetonitrile with 0.08% trifluoroacetic acid (TFA) added (solvent B) in water with 0.1% TFA added (solvent A) was used in all cases. 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; additional details on purification and analytical methods are in the Section A1.11.

A1.5 The structure and function of folded synthetic proteins are comparable to recombinant samples

Determining the purity of long synthetic peptides is challenging because of difficulties associated with identification and quantification of by-products by standard analytical techniques. In a physiological environment, the native folded structure of a globular protein, which gives rise to its distinctive biological activity, is determined by its amino acid sequence.[36] As a consequence, the tertiary structure of a protein can be used as a measure of the chemical integrity of the primary amino acid sequence.[37]

We folded and purified selected synthetic proteins by size exclusion chromatography and ion exchange chromatography and characterized their tertiary structure with biophysical and functional assays, alongside recombinant protein standards. Our goal was to demonstrate the fidelity of our AFPS protocol in delivering synthetic proteins of defined covalent structure and high chiral integrity. To this aim, we thoroughly characterized barnase and further investigated barstar, 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 various conditions.

Chemical denaturation is diagnostic for assessing structural integrity and stability of synthetic proteins. The globular protein barnase, a bacterial ribonuclease (RNase) isolated from Bacillus amyloliquefaciens, is a model system to investigate protein folding, denaturation, and binding to its inhibitor protein barstar (Fig. A1.4A).[38,39] The primary structures of synthetic and recombinant barnase were indistinguishable by LC-MS and HPLC methods (Fig. A1.4B). We used a chemical denaturation fluorometric assay as a readout for the integrity of the tertiary structure (Fig. A1.4C). In this assay, tryptophan fluorescence was used to monitor the folding equilibrium, as the concentration of urea was varied. Synthetic barnase exhibited a transition midpoint (the concentration at which half of the sample is unfolded, $[D]_{50\%}$) that compared well to both the authentic recombinant sample and literature value {[D]50%, synthetic = 4.68 ± 0.06 M; $[D]_{50\%,\ recombinant}$ = 4.63 ± 0.04 M (mean ± SE); $[D]_{50\%,\ literature}$ = 4.57 M}.[39] 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}].[39] If the synthetic protein were microheterogeneous (e.g., contained a distribution of isomers or deletion coproducts), then the apparent m value may be altered owing to the distribution of $[D]_{50\%}$ values represented within the mixture. Therefore, because the synthetic sample exhibited an m value within the error of the recombinant sample, we concluded that microheterogeneity was negligible.

Enzymatic assays show comparable activity of synthetic proteins obtained by AFPS and their recombinant equivalents. Enzymatic catalysis is sensitive to minor changes in the enzyme's tertiary structure, for which even single point mutations can have a major impact.[40,41] We evaluated the native activity of three synthetic variants of well-studied enzymes: barnase, HIV-1 protease, and sortase A*. Barnase catalyzes hydrolysis at diribonucleotide GpN sites. Its specific activity can be measured by monitoring hydrolysis of a DNA-RNA hybrid containing a Förster resonance energy transfer fluorophore pair.[42] The enzymatic efficiency of synthetic barnase was k_{cat}/K_M = (7.6 ± 0.2) × 106 M^{-1} s^{-1} (mean ± SE), which is comparable to that of recombinant barnase [k_{cat}/K_M = (9.0 ± 0.3) × 106 M^{-1} s^{-1} (mean ± SE)] determined using the same assay (Fig. A1.4D).

The primary structure of HIV-1 protease was confirmed by LC-MS and HPLC methods (Fig. A1.5B). HIV-1 protease hydrolyzes the peptides of HIV, and using a fluorogenic peptide allows for quantification of its proteolytic activity.[43] Synthetic HIV-1 protease displays a Michaelis constant of K_M = 20.9 ± 1.0 mM (mean ± SE) and a turnover number of k_{cat} = 29.6 ± 4.1 s^{-1} (mean ± SE), close to literature values published for a similar synthetic sample obtained by SPPS (Fig. A1.5C).[30] 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 (Fig. A1.5D).[29]

Sortase $A_{59–206}$ is a transpeptidase produced by Gram-positive bacteria that catalyzes a cell wall sorting reaction at a threonine-glycine bond in the LPXTG motif (Leu-Pro-X-Thr-Gly, where X is any amino acid).[44] We synthesized the 164–amino acid–long sortase A* variant (P94S/D160N/K196T; P, Pro; S, Ser; D, Asp; N, Asn; K,

Lys; T, Thr) to allow for direct comparison to a recombinant standard.[45,46] At a concentration of 0.01 mg/ml, synthetic sortase A* led to 47% product formation by LCMS within 24 hours (starting from 0.2 mg/ml GGGGGLY and AQALPETGEE as test substrates; G, Gly; L, Leu; Y, Tyr; A, Ala; Q, Gln; E, Glu) (Fig. A1.28). This conversion value is comparable to that determined for the recombinant protein (50% product formation within 24 hours). Enzymatic activity assays of synthetic proteins accessed by AFPS therefore confirmed both the high substrate specificity and comparable activity to recombinant enzymes and literature values.

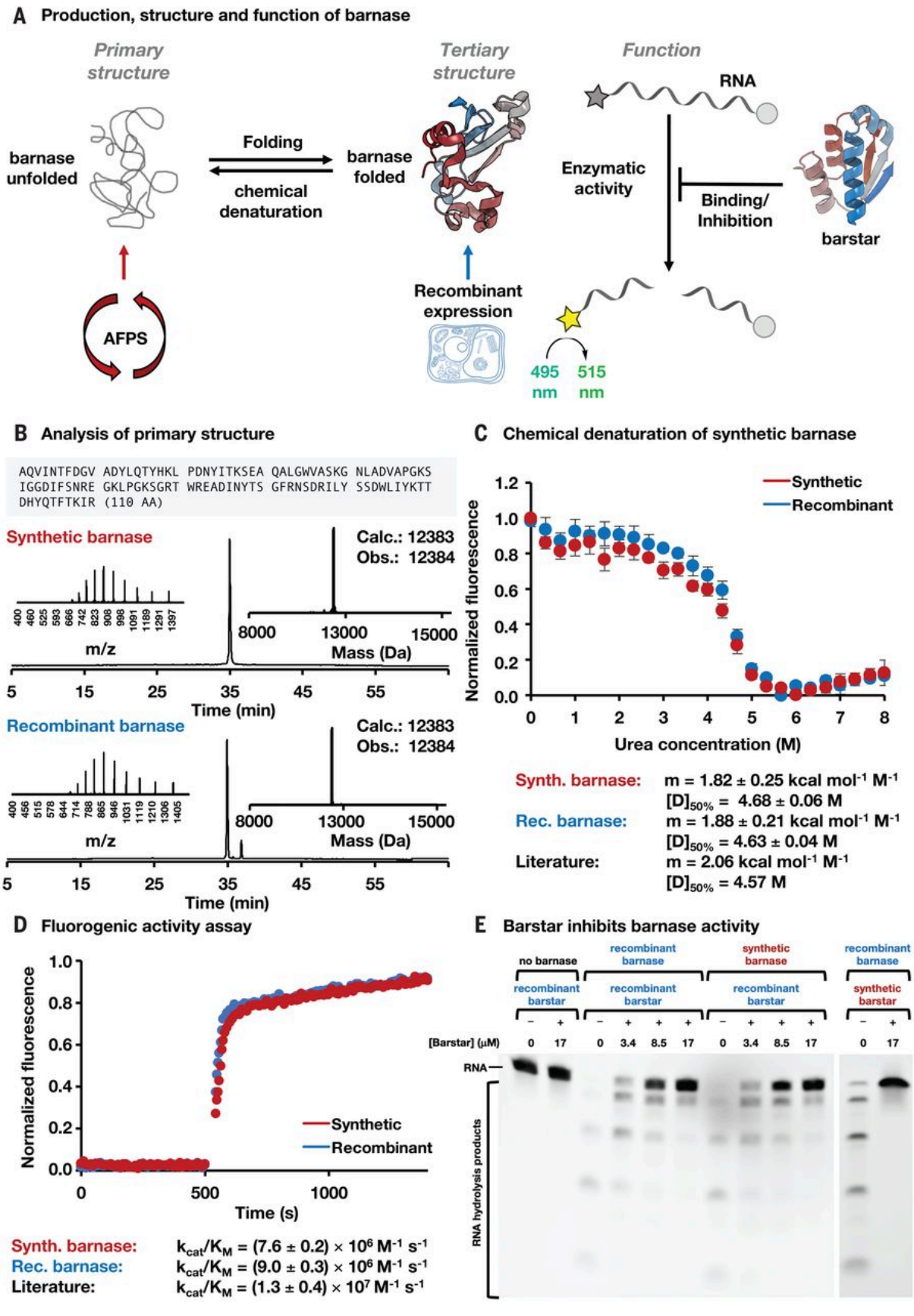

A Production, structure and function of barnase
Primary structure
Tertiary structure
Function
barnase unfolded
Folding
chemical denaturation
barnase folded
RNA
Enzymatic activity
Binding/ Inhibition
barstar
AFPS
Recombinant expression
495 nm
515 nm
B Analysis of primary structure
AQVINTFDGV ADYLQTYHKL PDNYITKSEA QALGWVASKG NLADVAPGKS
IGGDIFSNRE GKLPGKSGRT WREADINYTS GFRNSDRILY SSDWLIYKTT
DHYQTFTKIR (110 AA)
Synthetic barnase
Calc.: 12383
Obs.: 12384
m/z
Mass (Da)
Time (min)
Recombinant barnase
Calc.: 12383
Obs.: 12384
m/z
Mass (Da)
Time (min)
C Chemical denaturation of synthetic barnase
Synthetic
Recombinant
Normalized fluorescence
Urea concentration (M)
Synth. barnase: m = 1.82 ± 0.25 kcal mol-1 M-1
[D]50% = 4.68 ± 0.06 M
Rec. barnase: m = 1.88 ± 0.21 kcal mol-1 M-1
[D]50% = 4.63 ± 0.04 M
Literature: m = 2.06 kcal mol-1 M-1
[D]50% = 4.57 M
D Fluorogenic activity assay
Normalized fluorescence
Time (s)
Synthetic
Recombinant
Synth. barnase: kcat/KM = (7.6 ± 0.2) × 10^6 M-1 s-1
Rec. barnase: kcat/KM = (9.0 ± 0.3) × 10^6 M-1 s-1
Literature: kcat/KM = (1.3 ± 0.4) × 10^7 M-1 s-1
E Barstar inhibits barnase activity
no barnase
recombinant barnase
synthetic barnase
recombinant barstar
synthetic barstar
[Barstar] (μM)
RNA
RNA hydrolysis products

Figure A1.4. **Synthetic barnase and synthetic barstar fold into the native tertiary structure and display enzymatic activity comparable to recombinant samples.** (**A**) Conceptual overview of production and analysis methods. (**B**) Comparison of primary structures obtained from AFPS and recombinant expression. For both cases, analytical HPLC data of the purified barnase are presented as the main chromatographic trace with absorbance detection at 214 nm (additional details in the SM). ESI mass spectrum and deconvoluted mass spectrum of the purified peptide samples are displayed in the upper-left and the upper-right insets, respectively. Both spectra were obtained by summation over the entire LC peak in the chromatogram. (**C**) Structural evaluation of barnase in a chemical denaturation assay using urea as denaturant performed in triplicate; results are reported as mean ± SE. Error bars on the graph indicate SE. (**D**) Quantitative enzymatic activity assay performed in triplicate; error bars are not displayed for clarity. Details are outlined in the SM. k_{cat}/K_M values are reported as mean ± SE. (**E**) Barnase inhibition and binding assay using recombinant and synthetic barstar. 3.4 nM barnase was used in all conditions. Details are outlined in the SM. PDB 1BRS (barnase)[57]and 1AY7 (barstar)[59] were used.

Binding studies of synthetic MDM2 and barnase confirmed specific affinities for their respective substrates. Barnase binds selectively and with high affinity to its inhibitor barstar. In a gel-based assay, recombinant barstar inhibited RNase activity of synthetic and recombinant barnase in a concentration dependent manner (Fig. A1.4E).[47] In addition, synthetic barstar obtained with AFPS performed comparably to recombinant barstar. To quantify binding of a synthetic protein to a known ligand, we also characterized the N-terminal binding domain of MDM2$^{[1–118]}$.[32] The binding of MDM2 to p53 is a key interaction in multiple pathways up-regulated in cancer.[48,49] We folded milligram quantities of synthetic MDM2$^{[1–118]}$ and characterized its binding to immobilized p53$^{[14–29]}$ using biolayer interferometry (Figs. A1.29 and A1.30). Synthetic MDM2$^{[1–118]}$ displayed an affinity toward p53 [dissociation constant (K_d) = 6.25 mM] comparable to the literature value (K_d = 5.45 mM) obtained under the same folding conditions.

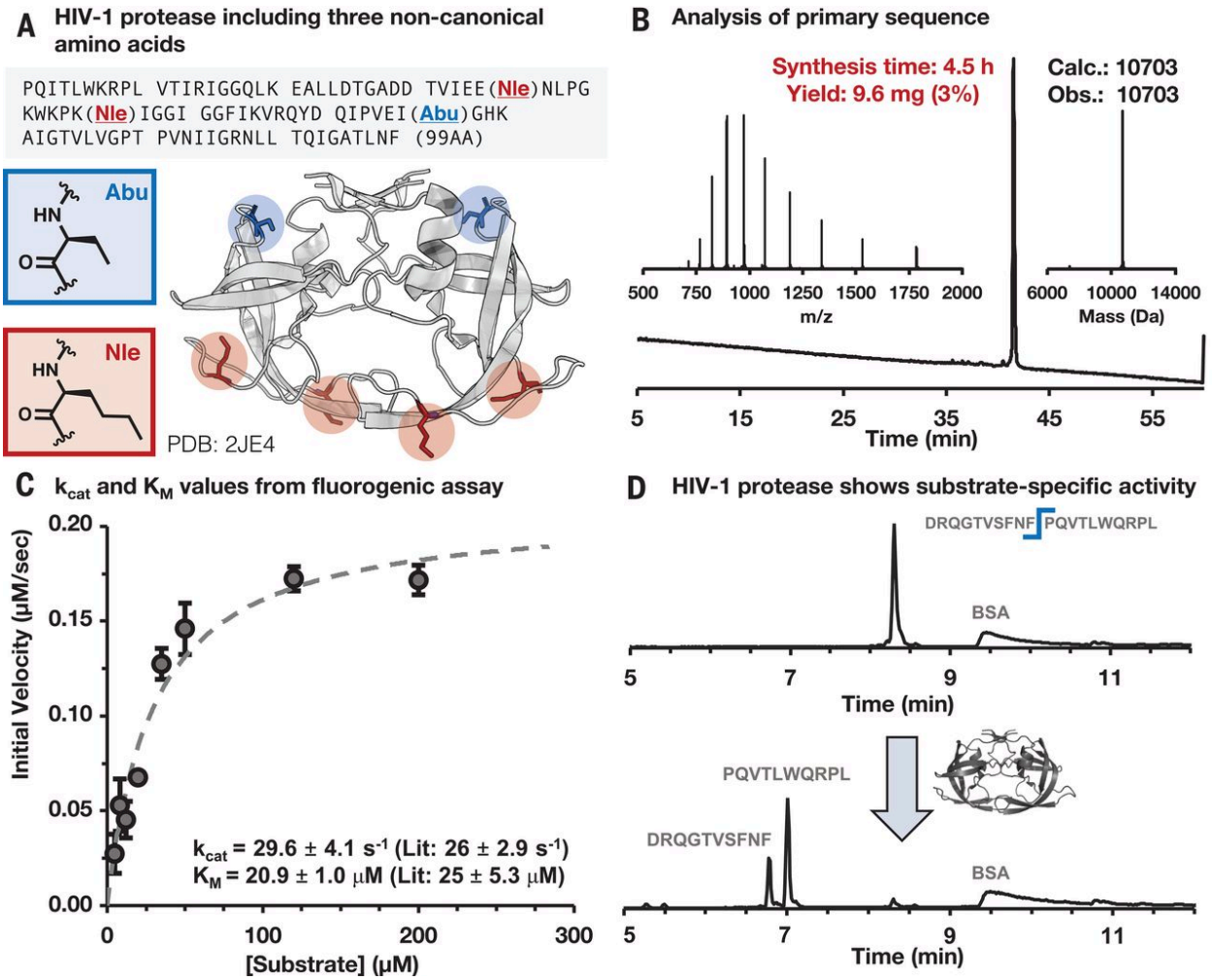


Figure A1.5. **Synthetic HIV-1 protease containing three noncanonical amino acids folds into the native dimer structure and displays enzymatic activity and substrate specificity comparable to literature samples.** (**A**) Crystal structure of HIV-1 protease dimer with highlighted noncanonical amino acids aminobutyric acid (Abu, blue) and norleucine (Nle, red). PDB 2JE4 (HIV-1 protease dimer with inhibitor)[30] was used. (**B**) Primary structure obtained from AFPS. Analytical HPLC data of the purified HIV-1 protease is presented as the main chromatographic trace with absorbance detection at 214 nm (additional details in the SM). ESI mass spectrum and deconvoluted mass spectrum of the purified sample are displayed in the upper-left and the upper-right insets, respectively. Both spectra were obtained by summation over the entire LC peak in the chromatogram. (**C**) Quantitative enzymatic activity assay performed in triplicate for the determination of k_{cat} and K_M values. Results are reported as mean ± SE. Error bars on the graph indicate SE. Lit., literature. (**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; details are outlined in the SM.

A1.6 Discussion

The optimized AFPS protocol demonstrates advantages of flow chemistry over common batch methods, yielding peptide chains more than three times longer than previously accessible by routine standard SPPS.[6] An improvement to existing flow protocols was achieved by rapid screening of variables in a reproducible reaction setup. Even though in this study AFPS yields superior results over traditional SPPS 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. A potentially limiting feature of our setup is synthesis scale. The capacity of the reactor used in our study allows up to 200 mg of resin with a loading of 0.49 mmol/g. Increased production output can be achieved by incorporating a larger reactor in the current system, but such a modification will likely require specific optimization, toward which we performed preliminary investigations.[19] Since we implemented AFPS, we have produced more than 5000 peptides and automatically collected in-line analysis data for all syntheses. Moving forward, this extensive, high-quality dataset could be leveraged to further improve peptide synthesis in flow using machine learning and other computational methods. Ultimately, we intend for this report to serve as a blueprint for the automated flow synthesis of other biopolymers and artificial sequence-defined polymers.[50]

A robust, widely available routine method for chemical production of proteins is poised to have a strong 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. 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 (those with a mass of up to ~30 kDa).[10,51] In this respect, we envisage adapting to our AFPS protocol the incorporation of peptide hydrazides for thioester-based ligation, an approach previously achieved with manual flow instrumentation.[52] Additional research avenues opened by our method include rapid access to mirror-image proteins, post-translationally modified proteins, and de novo–designed, abiotic proteins. Introduction of noncanonical amino acids as point mutations in native proteins will make accessible variants with considerably altered biological function, for example, catalytic activity.[53,54] 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.[55,56]

A1.7 References

1. J. Britton, C. L. Raston, Chem. Soc. Rev. 46, 1250–1271 (2017).
2. D. Webb, T. F. Jamison, Chem. Sci. 1, 675–680 (2010).
3. S. Bondalapati, M. Jbara, A. Brik, Nat. Chem. 8, 407–418 (2016).
4. J. W. Chin, Nature 550, 53–60 (2017).
5. R. B. Merrifield, J. Am. Chem. Soc. 85, 2149–2154 (1963).
6. S. B. H. Kent, Chem. Soc. Rev. 38, 338–351 (2009).
7. A. A. Zompra, A. S. Galanis, O. Werbitzky, F. Albericio, Future Med. Chem. 1, 361–377 (2009).
8. P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, Science, 266, 776–779 (1994).
9. V. Agouridas et al., Chem. Rev. 119, 7328–7443 (2019).
10. P. E. Dawson, S. B. H. Kent, Annu. Rev. Biochem. 69, 923–960, (2000).
11. A. El-Faham, F. Albericio, Chem. Rev. 111, 6557–6602 (2011).
12. N. Ahmed, Chem. Biol. Drug Des. 91, 647–650 (2018).
13. C. P. Gordon, Org. Biomol. Chem. 16, 180–196 (2018).

14. A. J. Mijalis, A. Steinauer, A. Schepartz, B. L. Pentelute, in Flow Chemistry in Organic Synthesis (Georg Thieme Verlag, Stuttgart, 2018), pp. 381–398.
15. E. Bayer, G. Jun, I. Halász, I. Sebestian, Tetrahedron Lett. 11,4503–4505 (1970).
16. T. J. Lukas, M. B. Prystowsky, B. W. Erickson, Proc. Natl. Acad. Sci. U.S.A. 78, 2791–2795 (1981).
17. A. J. Mijalis et al., Nat. Chem. Biol. 13, 464–466 (2017).
18. S. K. Mong, A. A. Vinogradov, M. D. Simon, B. L. Pentelute, ChemBioChem 15, 721–733 (2014).
19. M. D. Simon, “Fast flow biopolymer synthesis,” thesis, Massachusetts Institute of Technology, Cambridge, MA (2017).
20. S. Fuse, Y. Otake, H. Nakamura, Chem. Asian J. 13, 3818–3832 (2018).
21. S. Fuse, Y. Mifune, T. Takahashi, Angew. Chem. Int. Ed. 53, 851–855 (2014).
22. L. K. Spare, V. Laude, D. G. Harman, J. R. Aldrich-Wright, C. P. Gordon, React. Chem. Eng. 3, 875–882 (2018).
23. E. T. Sletten, M. Nuño, D. Guthrie, P. H. Seeberger, Chem. Commun. 55, 14598–14601 (2019).
24. M. Schnölzer, P. Alewood, A. Jones, D. Alewood, S. B. H. Kent, Int. J. Pept. Protein Res. 40, 180–193 (1992).
25. T. Michels, R. Dölling, U. Haberkorn, W. Mier, Org. Lett. 14, 5218–5221 (2012).
26. M. Mergler, F. Dick, B. Sax, P. Weiler, T. Vorherr, J. Pept. Sci. 9, 36–46 (2003).
27. S. A. Palasek, Z. J. Cox, J. M. Collins, J. Pept. Sci. 13, 143–148 (2007).
28. S. Luisier, M. Avital-Shmilovici, M. A. Weiss, S. B. H. Kent, Chem. Commun. 46, 8177–8179 (2010).
29. J. Schneider, S. B. H. Kent, Cell 54, 363–368 (1988).
30. E. C. B. Johnson et al., J. Am. Chem. Soc. 129, 11480–11490 (2007).
31. A. Brik, C.-H. Wong, Org. Biomol. Chem. 1, 5–14 (2003).
32. J. D. Oliner et al., Nature 362, 857–860 (1993).
33. D. Bresson et al., J. Clin. Invest. 116, 1371–1381 (2006).
34. E. Gasser, C. P. Moutos, M. Downes, R. M. Evans, Nat. Rev. Endocrinol. 13, 599–609 (2017).
35. L. Serrano, J. T. Kellis Jr., P. Cann, A. Matouschek, A. R. Fersht, J. Mol. Biol. 224, 783–804 (1992).
36. C. B. Anfinsen, Science 181, 223–230 (1973).
37. S. B. Kent, P. F. Alewood, Curr. Opin. Chem. Biol. 22, viii–xi (2014).
38. R. W. Hartley, Trends Biochem. Sci. 14, 450–454 (1989).
39. J. T. Kellis Jr., K. Nyberg, A. R. Fersht, Biochemistry 28, 4914–4922 (1989).
40. J. R. Knowles, Science 236, 1252–1258 (1987).
41. U. Arnold et al., J. Am. Chem. Soc. 125, 7500–7501 (2003).
42. B. R. Kelemen et al., Nucleic Acids Res. 27, 3696–3701 (1999).
43. M. V. Toth, G. R. Marshall, Int. J. Pept. Protein Res. 36, 544–550 (1990).
44. H. Ton-That, G. Liu, S. K. Mazmanian, K. F. Faull, O. Schneewind, Proc. Natl. Acad. Sci. U.S.A. 96, 12424–12429 (1999).
45. J. J. Ling, R. L. Policarpo, A. E. Rabideau, X. Liao, B. L. Pentelute, J. Am. Chem. Soc. 134, 10749–10752 (2012).
46. R. L. Policarpo et al., Angew. Chem. Int. Ed. 53, 9203–9208 (2014).
47. A. A. Vinogradov, E. D. Evans, B. L. Pentelute, Chem. Sci. 6, 2997–3002 (2015).

48. F. Touti, Z. P. Gates, A. Bandyopadhyay, G. Lautrette, B. L. Pentelute, Nat. Chem. Biol. 15, 410–418 (2019).
49. P. Chène, Mol. Cancer Res. 2, 20–28 (2004).
50. J.-F. Lutz, M. Ouchi, D. R. Liu, M. Sawamoto, Science 341, 1238149 (2013).
51. S. J. Wheelan, A. Marchler-Bauer, S. H. Bryant, Bioinformatics 16, 613–618 (2000).
52. M. D. Simon et al., ChemBioChem 15, 713–720 (2014).
53. J. C. Li, T. Liu, Y. Wang, A. P. Mehta, P. G. Schultz, J. Am. Chem. Soc. 140, 15997–16000 (2018).
54. H. Xiao et al., Proc. Natl. Acad. Sci. U.S.A. 112, 6961–6966 (2015).
55. R. J. Desnick, E. H. Schuchman, Nat. Rev. Genet. 3, 954–966 (2002).
56. N. L. Truex et al., Sci. Rep. 10, 723 (2020).
57. A. M. Buckle, G. Schreiber, A. R. Fersht, Biochemistry 33, 8878–8889 (1994).
58. Y. Yang et al., J. Biol. Chem. 285, 7847–7851 (2010).
59. J. Sevcík, L. Urbanikova, Z. Dauter, K. S. Wilson, Acta Crystallogr. D Biol. Crystallogr. 54, 954–963 (1998).
60. J. Bella, B. Brodsky, H. M. Berman, Structure 3, 893–906 (1995).
61. A. Czarna et al., Cell Cycle 8, 1176–1184 (2009).
62. T. Durek, V. Y. Torbeev, S. B. H. Kent, Proc. Natl. Acad. Sci. U.S.A. 104, 4846–4851 (2007).
63. X. Xia, L. M. Longo, M. Blaber, J. Pharm. Sci. 104, 566–576 (2015).
64. N. Suree et al., J. Biol. Chem. 284, 24465–24477 (2009).
65. A. J. Mijalis, amijalis/afps-integration: AFPS-analysis: Code used to extract and analyze data from AFPS synthesis files. Zenodo (2020).

A1.8 Materials and Methods

A1.8.1 Reagents and Solvents

All reagents were purchased and used as received. Fmoc-protected amino acids (FmocAla-OHxH2O, Fmoc-Arg(Pbf)-OH; Fmoc-Asn(Trt)-OH; Fmoc-Asp-(Ot-Bu)-OH; FmocCys(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-ProOH; Fmoc-Ser(But)-OH; Fmoc-Thr(t-Bu)-OH; Fmoc-Trp(Boc)-OH; Fmoc-Tyr(t-Bu)-OH; FmocVal-OH), N-α-Fmoc-L-α-aminobutyric acid (Fmoc-Abu-OH) and N-α-Fmoc-L-norleucine (Fmoc-Nle-OH) were purchased from the Novabiochem-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.

A1.8.2 Automated flow peptide synthesis 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 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 4 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.

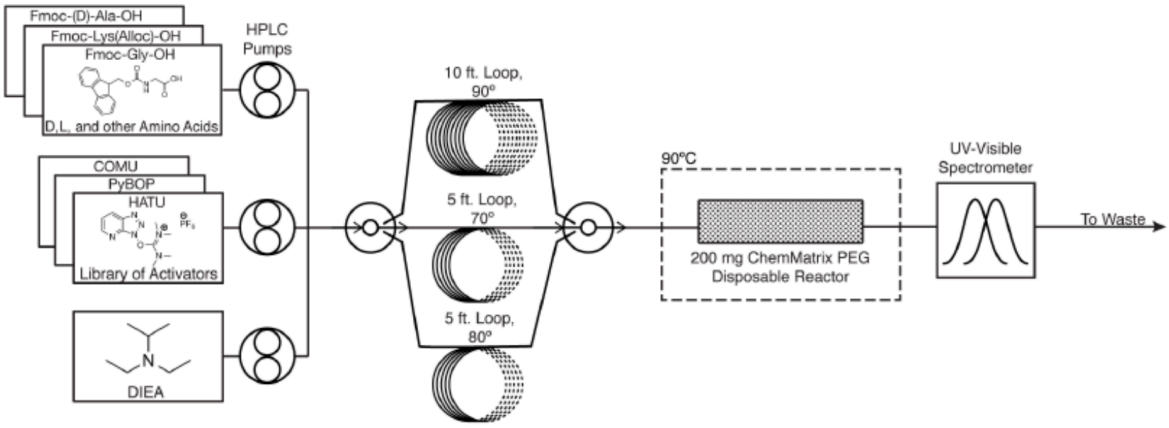


Figure A1.6. **Schematic AFPS setup used in this study.** Modified from Ref. 17.

A1.8.3 Cleavage Protocols

After synthesis, the peptidyl 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 and proteins we used two different protocols:

A) Peptides: 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 h.

B) Proteins and Cys-containing peptides (except for GCF peptide): Approximately 3 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.

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. Residual ether was allowed to evaporate and 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.

A1.8.4 Liquid chromatography-mass spectrometry (LC-MS)

For mass analysis, the 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 6520 Accurate-Mass Q-TOF LC-MS (abbreviated as 6520) or 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:

- 1-61% B over 9 min, Zorbax C3 column (6520)

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

- 1-91% B over 9 min, Zorbax C3 column (6520)

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

- 1-61% B over 15 min, Phenomenex Jupiter C4 column (6550)

  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. MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 100–1700 m/z.

- 1-61% B over 33 min, Phenomenex Jupiter C4 column (6550)

  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.

MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 100–1700 m/z.

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

  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.

  MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 100–1700 m/z.

- 1-91% B over 20 min, Phenomenex Jupiter C4 column (6550)

  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.

  MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 100–1700 m/z.

- 1-61% B over 17 min, Luna C18 column (6550)

  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.

  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.

A1.8.5 Analytical high performance liquid chromatography (HPLC)

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. Peptide samples containing cysteines were diluted in 6M Guanidinium chloride containing 100 mM DTT.[2] The samples were analyzed on Agilent Technologies 1200 Series, which was computer-controlled through Agilent ChemStation software.

For standard analysis of all peptide 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. Epimer ratios were determined as described in section A1.8.9.

A1.8.6 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 2 Except for GCF peptides for epimerization studies 7 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. 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).

A1.8.7 Determination of protein concentrations

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

1. A_{280}: 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.
2. 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.

A1.8.8 Determination of 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). 8 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 under “Determination of protein concentration”.

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:

$$\text{theoretical yield} = 0.44\ \text{mmol/g} \times 50\ \text{mg} \times 916\ \text{g/mol} = 20\ \text{mg}$$

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:

$$\text{yield} = 10\ \text{mg}/20\ \text{mg} \times 0.50 \times 100 = 25\%$$

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:

$$\text{yield} = 4\ \text{mg}/20\ \text{mg} \times 100 = 20\%$$

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:

$$\text{purification yield} = 1\ \text{mg}/\text{mL} \times 1\ \text{mL}/4\ \text{mg} \times 100 = 25\%$$

A1.9 Development of a general AFPS protocol

A1.9.1 Initial AFPS conditions

Parameter	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 and 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)

***Table A1.1*. Conditions that were used on the AFPS system "amidator" at the beginning of the optimization process.** NOTE: 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.

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)
Amyloid-β:	DAEFRHDSGY EVHHQKLVFF AEDVGSNKGA IIGLMVGGVV IA (42 AA)

A1.9.2 Temperature

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 A1.2. **Synthesis at 90 °C results in improved synthesis outcome for GLP-1 and GHRH.** Synthesis conditions: resin = 100 mg ChemMatrix RINK amide resin (0.5 mmol/g loading); temperature = as indicated in this table; flowrate = 40 mL/min; coupling step = 0.2 M amino acid stock solutions in NMP, 0.19 M activator in DMF, DIEA as coupling base; PyAOP for coupling of R, HATU for coupling of all other amino acids (13 pump strokes); deprotection = 40% piperidine as deprotection base (diluted to 20% during synthesis on the machine); washing steps = DMF (40 pump strokes).

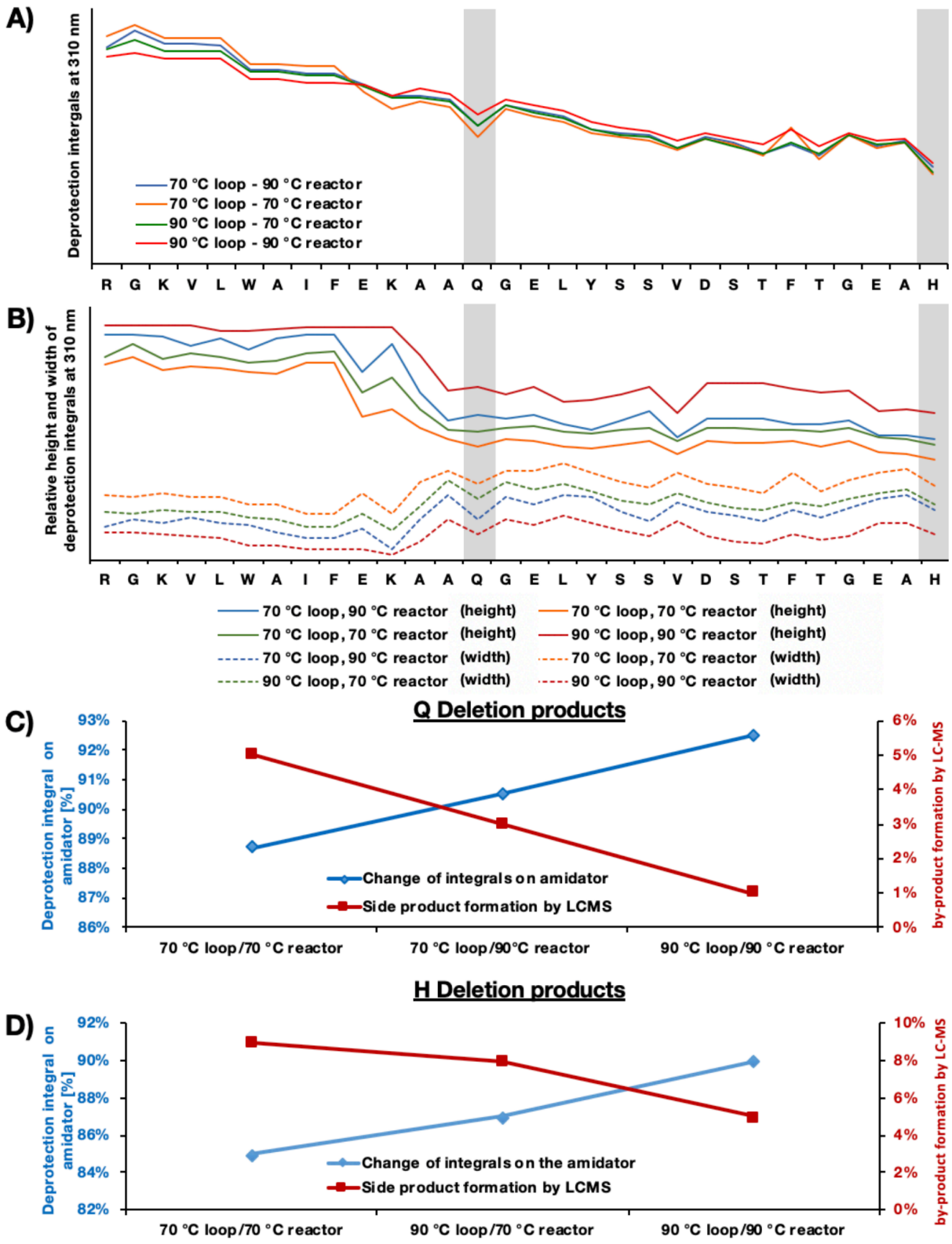


***Figure A1.7*. Change of UV deprotection integrals during the synthesis of GLP-1 at different temperatures**. A) Relative change of absolute deprotection signals. Higher temperatures lead to more consistent coupling efficiencies, which correlates to the

observation of increased formation of deletion by-products at lower temperature. Highlighted in grey are integrals for glutamine (Q) and histidine (H), two major deletion by-products by LC-MS; B) The shape of the peaks at 90 °C is overall much sharper than at 70 °C, which indicates increased mass transfer. Highlighted in grey are integrals for glutamine (Q) and histidine (H), two major deletion by-products by LC-MS; D) glutamine (Q) deprotection integral versus by-product detected by LC-MS; C) histidine (H) deprotection integral versus by-product detected by LC-MS.

Conclusion: Higher temperature leads to increased coupling efficiency and less deletion byproducts. 90 °C was therefore chosen as the optimal temperature for both activation loop and reactor.

A1.9.3 Solvent and reagent concentrations

Different solvent and reagents were evaluated for two temperature settings:

Entry	peptide	Solvent	Amino acid conc.	HPLC purity	Crude yield
1	GLP-1	DMF	0.2M	**54%**	18%
2	GLP-1	NMP	0.4M	**56%**	15%
3	GLP-1	NMP	0.2M	**46%**	17%
4	GLP-1	amine-free DMF	0.4M	**65%**	13%
5	GLP-1	amine-free DMF	0.2M	**61%**	11%

***Table A1.3*. High reagent concentration lead to increased synthesis outcome for the synthesis of GLP-1.** Synthesis conditions: resin = 100 mg ChemMatrix RINK amide resin (0.5 mmol/g loading); temperature = 85–90 °C in reactor, 70 °C in 10’ activation loop; flowrate = 40 mL/min; coupling step = concentrations and solvents as indicated in the table (“amine-free DMF” was pre-treated with AldraAmine trapping agents >24 h prior to usage), DIEA as coupling base; PyAOP for coupling of R, HATU for coupling of all other amino acids (13 pump strokes); deprotection = 40% piperidine as deprotection base (diluted to 20% during synthesis on the machine); washing steps = DMF (40 pump strokes).

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 A1.4. **High reagent concentration lead to increased synthesis outcome for the synthesis of GLP-1.** Synthesis conditions: resin = 100 mg ChemMatrix RINK amide resin (0.5 mmol/g loading); temperature = 85–90 °C in reactor, 90 °C in 10' activation loop; flowrate = 40 mL/min; coupling step = concentrations and solvents as indicated in the table, DIEA as coupling base; PyAOP for coupling of R, HATU for coupling of all other amino acids (13 pump strokes); deprotection = 40% piperidine as deprotection base (diluted to 20% during synthesis on the machine); washing steps = DMF (40 pump strokes). "Amine-free DMF" was pre-treated with AldraAmine trapping agents >24 h prior to usage.

Conclusion: A temperature of 90 °C in the heating loop, concentrations of 0.4 M (amino acids) and 0.38 M (activator) and amine-free DMF as a solvent significantly improve synthesis outcome for GLP-1. Temperature leads to the most significant improvement for GLP-1 synthesis, followed by the concentration of reagents.

A1.9.4 Coupling bases

Entry	peptide	Coupling base	HPLC purity	Crude yield
1	GLP-1	DIEA	**66%**	24%
2	GLP-1	NMM	**24%**	9%
3	GLP-1	Collidine + HOBt (1.0eq)	**5%**	2%
4	GLP-1	Anisole (7%, v/v)	**56%**	21%

Table A1.5. **DIEA is superior to other bases for the synthesis of GLP-1.** Synthesis conditions: resin = 100 mg ChemMatrix RINK amide resin (0.5 mmol/g loading); temperature = 85–90 °C in reactor, 90 °C in 10' activation loop; flowrate = 40 mL/min; coupling step = 0.4 M amino acid stock solutions in amine-free DMF, 0.38 M activator in amine-free DMF, coupling base as indicated in the table; PyAOP for coupling of R, HATU for coupling of all other amino acids (13 pump strokes); deprotection = 40% piperidine as deprotection base (diluted to 20% during synthesis on the machine);

washing steps = amine-free DMF (40 pump strokes). “Amine-free DMF” was pre-treated with AldraAmine trapping agents >24 h prior to usage.

Conclusion: DIEA is the most effective coupling base on AFPS.

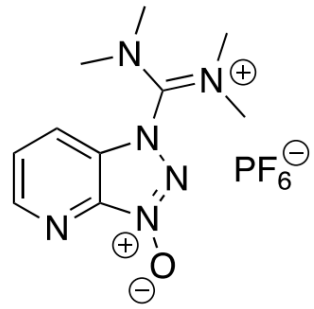
A1.9.5 Deprotection Solutions

Entry	peptide	base	solvent	Coupling agent	HPLC purity	Crude yield
1	GLP-1	Piperidine (40%, v/v)	DMF	HATU	**69%**	33%
2	GLP-1	Piperidine (40%, v/v)	DMF + TFE (1%)	HATU	**75%**	31%
3	GLP-1	piperazine (6%, w/v)	DMF	HATU	**41%**	14%
4	GLP-1	piperazine (6%, w/v)**	DMF	HATU	**59%**	22%
5	GLP-1	piperazine (8%, w/v)	DMF	HATU	**43%**	12%
6	GLP-1	DBU (2%)**	DMF	HATU	**78%**	33%
7	GLP-1	DBU (4%)	DMF	HATU	**66%**	31%
8	GLP-1	DBU (4%)**	DMF	HATU	**66%**	31%
9	GLP-1	DBU (8%)	DMF	HATU	**61%**	21%
10	GLP-1	DBU (2%) + piperazine (6%)	DMF	HATU	**75%**	34%
11	GLP-1	DBU (2%) + piperazine (6%)	DMF	PyAOP	**74%**	35%
12	GLP-1	DBU (2%) + piperazine (6%)	DMF + TFE (1%)	PyAOP	**76%**	41%
13	GLP-1	Piperidine (40%) + FA (2%)	DMF	HATU	**77%**	30%
14	GLP-1	Piperidine (40%) + FA (2%)	DMF	PyAOP	**78%**	36%
15	GLP-1	Piperidine (40%) + FA (2%)	DMF + TFE (1%)	PyAOP	**74%**	34%
16	GLP-1	Piperidine (40%) + 0.1M HOBt	DMF	HATU	**68%**	29%

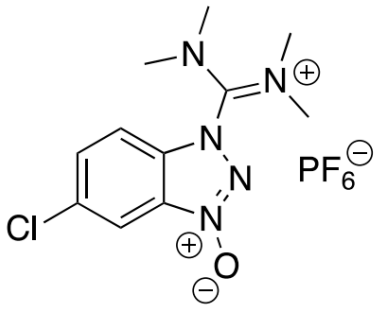
***Table A1.6*. Multiple deprotection solutions show improved synthesis outcome.** Synthesis conditions: resin = 100 mg ChemMatrix RINK amide resin (0.5 mmol/g loading); temperature = 85–90 °C in reactor, 90 °C in 10’ activation loop; flowrate = 40 mL/min; coupling step = 0.4 M amino acid stock solutions in amine-free DMF, 0.38 M activator in amine-free DMF, DIEA as coupling base; PyAOP for coupling of R, HATU for coupling of all other amino acids (13 pump strokes); deprotection = as indicated in the table (40 pump strokes); washing steps = amine-free DMF (40 pump strokes). “Amine-free DMF” was pre-treated with AldraAmine trapping agents >24 h prior to usage. ** = 80 pump strokes for deprotection.

Conclusion: Piperidine with formic acid additive improves synthesis outcome.

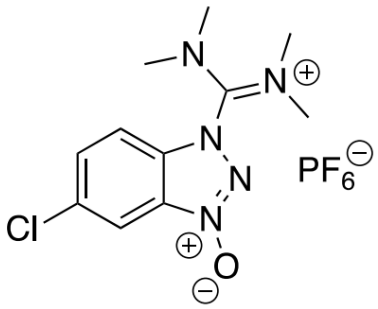
A1.9.6 Activators



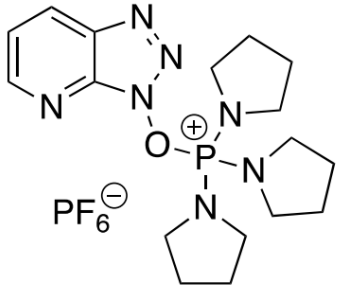
HATU



HCTU



HBTU



PyAOP

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 A1.7*. PyAOP was identified as the most effective activator for the synthesis of GLP-1.** Synthesis conditions: resin = 100 mg ChemMatrix RINK amide resin (0.5 mmol/g loading); temperature = 85–90 °C in reactor, 90 °C in 10' activation loop; flowrate = 40 mL/min; coupling step = 0.4 M amino acid stock solutions in amine-free DMF, 0.38 M activator in amine-free DMF, DIEA as coupling base; PyAOP for coupling of R, all other amino acids were coupled as described on this table (13 pump strokes); deprotection = 40% piperidine + 2% formic acid in DMF (40 pump strokes); washing steps = amine-free DMF (40 pump strokes). "Amine-free DMF" was pre-treated with AldraAmine trapping agents >24 h prior to usage.

Conclusion: PyAOP performed the best in our screening and was therefore used moving forward.

A1.9.7 Treatment of individual amino acids

For an amino acid-based optimization we used statistical data on amide bond formation that was indirectly obtained from the deprotection integrals on the AFPS system (real time UV data at 310 nm). Data from the synthesis of multiple peptides (excluding GLP-1 synthesis traces) was used to improve synthesis conditions as demonstrated in Table A1.8.

	A	D	E	F	G	H	I	K	L	M
initial coupling conditions	HATU, 13	HATU, 13	HATU, 13	HATU, 13	HATU, 13	HATU, 13	HATU, 13	HATU, 13	HATU, 13	HATU, 13
before individual amino acid optimization	97,9%	101,2%	99,9%	100,8%	103,2%	99,6%	99,6%	99,9%	99,7%	100,4%
changes to the protocol	PyAOP, 26	HATU, 13	HATU, 13	HATU, 13	HATU, 13	HATU, 13	HATU, 13	HATU, 13	HATU, 13	HATU, 13
after individual amino acid optimization	101,4%	98,1%	99,0%	100,6%	101,0%	98,7%	98,9%	98,4%	98,3%	99,8%

	N	P	Q	R	S	T	V	W	Y
initial coupling conditions	HATU, 13	HATU, 13	HATU, 13	PyAOP, 26	HATU, 13	HATU, 13	HATU, 13	HATU, 13	HATU, 13
before individual amino acid optimization	94,1%	99,8%	97,8%	96,2%	98,2%	98,3%	97,9%	98,7%	100,5%
changes to the protocol	PyAOP, 26	HATU, 13	PyAOP, 26	PyAOP, 26	HATU, 21	PyAOP, 26	PyAOP, 26	HATU, 13	HATU, 13
after individual amino acid optimization	97,4%	100,2%	100,5%	98,3%	101,3%	100,1%	99,7%	100,3%	99,4%

***Table A1.8*. UV integrals for deprotection signals before and after amino acid-based optimization of coupling protocol shows increased integral values for optimized residues.** NOTE Un-optimized amino acid coupling resulted in different integral values because coupling is residue and sequence-specific. This table is intended to display deprotection (and thereby coupling) trends rather than absolute values

Finally, GLP-1 synthesis under various coupling conditions was examined as demonstrated in Table A1.9. Since Valine and Histidine deletions were the most abundant by-products of GLP1 syntheses, these residues were coupled with various

conditions. After analyzing statistical peptide data for the various set of conditions (Table A1.8), GLP-1 was resynthesized with the optimal coupling conditions.

Entry	Peptide	Coupling conditions	Purity by HPLC	Crude Yield
1	GLP-1	H, V double coupling time	67%	30%
2	GLP-1	H, V double coupling time 1% TFE in Q,A,K solution	64%	27%
3	GLP-1	H, V double coupling time (deprotected by 2% DBU+6%Piperizine)	66%	30%
4	GLP-1	S,A = HATU (21 strokes) Q, N, V, T, R = PyAOP (21 strokes)	70%	24%

Table A1.9. **Set of developed coupling conditions from coupling statistics proves to increase synthesis outcome more than optimization based on by-product formation (by LCMS).** Synthesis conditions: resin = 100 mg ChemMatrix RINK amide resin (0.5 mmol/g loading); temperature = 85–90 °C in reactor, 90 °C in 10' activation loop; flowrate = 40 mL/min; coupling step = 0.4 M amino acid stock solutions in amine-free DMF, 0.38 M activator in amine-free DMF, DIEA as coupling base; PyAOP for coupling of R, all other amino acids were coupled as described on this table (13 pump strokes); deprotection = 40% piperidine + 2% formic acid in DMF (40 pump strokes); washing steps = amine-free. DMF (40 pump strokes). "Amine-free DMF" was pre-treated with AldraAmine trapping agents >24 h prior to usage.

A1.9.8 Aspartimide formation

Optimization effort was put into reducing aspartimide formation in sequences containing Asp. Peptide NN92 was used as a test sequence since it showed significant aspartimide formation in initial experiments. To determine the ratio of aspartimide to peptide, the integrals of the extracted ion count [EIC] of both compounds was integrated and compared as shown in Fig. A1.8.

Sequence: RVVVGEHNLS QNDGTEQYVN VQKIVSHPY (29 AA)

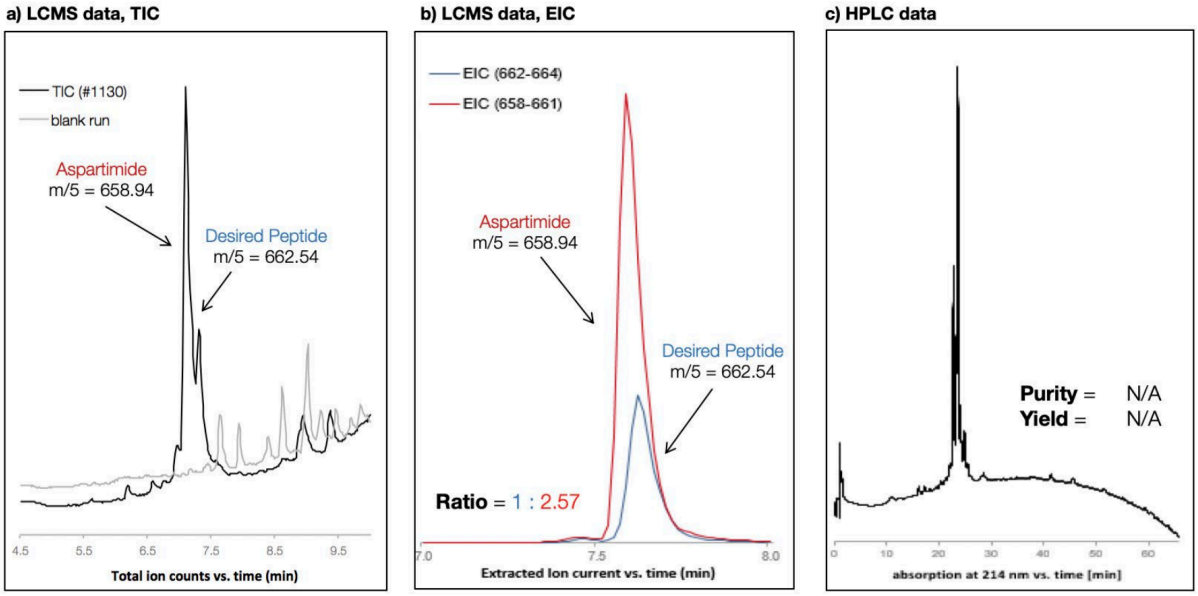
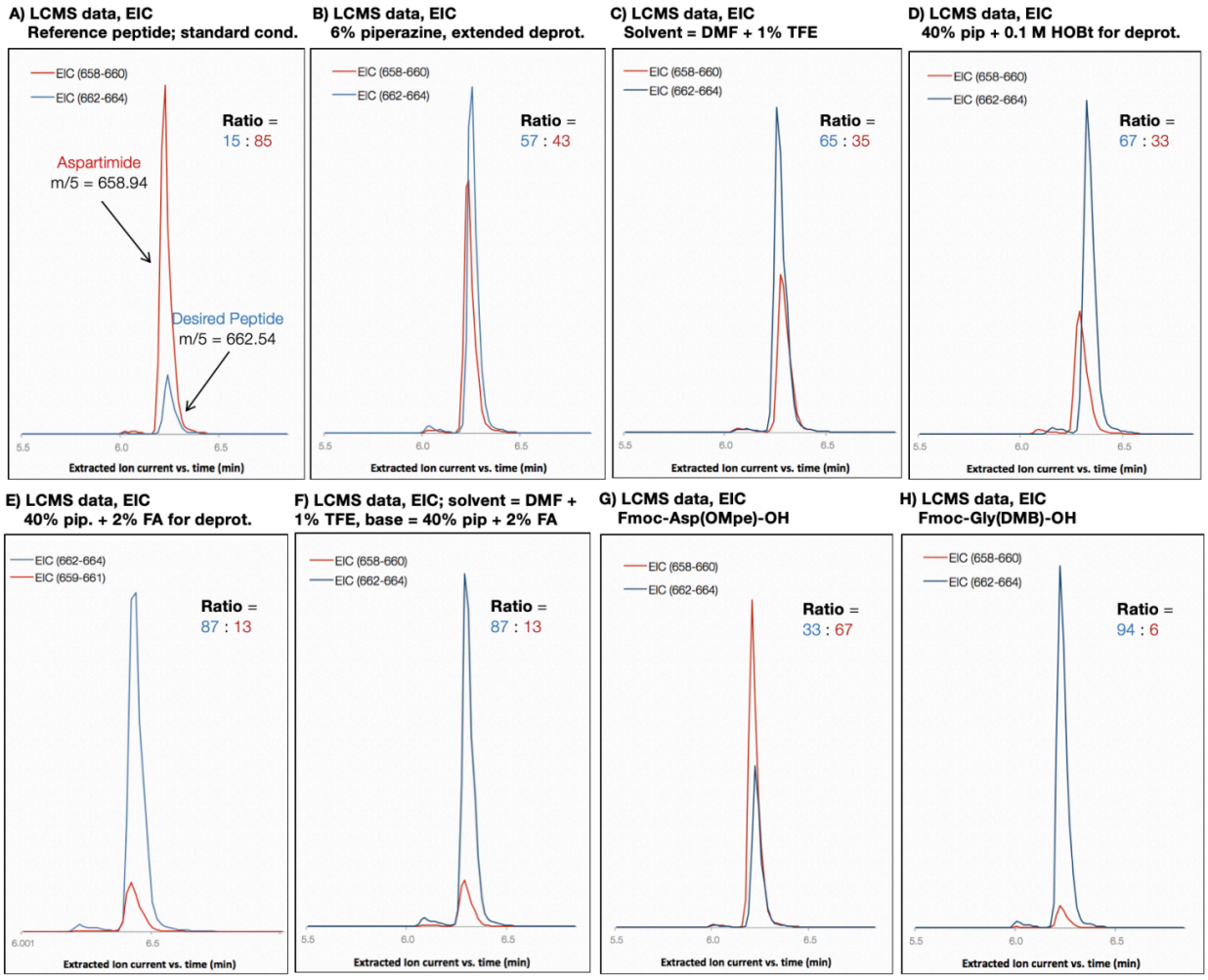


***Figure A1.8*. Synthesis data for aspartimide-forming peptide NN92**. A) LC-MS analysis of by-products; b) extracted ion counts of peptide m/z (662-664 Da) and aspartimide m/z (658-661); c) analytical HPLC trace.



Entry	base	solvent	Coupling agent	HPLC purity	Crude yield
1	Reference peptide; standard cond.	DMF	HATU [a)]	61%	17%
2	6% piperazine, extended deprot.	DMF	HATU [a)]	67%	31%
3	Solvent = DMF + 1% TFE	DMF	HATU [a)]	63%	27%
4	40% pip + 0.1 M HOBt for deprot.	DMF	HATU [a)]	63%	24%
5	40% pip + 2% formic acid for deprot.	DMF	HATU [a)]	65%	24%
6	solvent = DMF + 1% TFE, base = 40% pip+ 2% FA	DMF	HATU [a)]	74%	34%
7	Fmoc-Asp(OMpe)-OH	DMF	HATU [a)]	70%	15%
8	Fmoc-Gly(DMB)-OH	DMF	HATU [a)]	72%	20%

Table A1.10. **Formic acid additive and Fmoc-(DMB)Gly-OH were the most effective strategies to minimize aspartimide formation in the test peptide NN92.** Synthesis conditions: 100 mg ChemMatrix RINK amide resin (0.5 mmol/g loading), flowrate: 40 mL/min, temperature = 90 °C, 0.4 M amino acid stock solutions (diluted to 0.2 M solution during synthesis on the machine), 0.38 M activator (diluted to 0.19 M during synthesis on the machine), 40% piperidine as deprotection base (diluted to 20% during

synthesis on the machine) unless an alternative deprotection base was indicated in the figure.

Conclusion: In all cases the ratios of desired mass to aspartimide and HPLC purity (combined signal of product and aspartimide) improved; best results were achieved with formic acid additive and Fmoc-Gly(DMB)-OH.

A1.9.9 Epimerization experiments

Epimerization experiments were carried out for Fmoc-Cys(Trt)-OH, Fmoc-His(Trt)-OH and Fmoc-His(Boc)-OH. Initial experiments on epimerization and the correlation between flow rate and epimerization were already reported.[17] In additional experiments, we determined the correlation between epimerization and protecting group, temperature, heating time, reactor temperature, coupling strokes and additional coupling cycles. Four reference peptides (GCF, GcF, FHL and FhL) were prepared using standard batch solid-phase peptide synthesis. All peptides in this study were cleaved using a modified version of the protocol for short peptides (see section 2.3). Peptides were cleaved as described, the cleavage mixture was removed under a stream of nitrogen and the peptides were dissolved in 50% acetonitrile in water with 0.1% TFA. The peptide solution was filtrated by a Nylon syringe filter with a 0.22 µm pore size and directly submitted for analysis. For Fmoc-Cys(Trt)-OH epimerization was quantified by integrating EIC [m = 325–327] traces for both diastereomers. For Fmoc-His(Trt)-OH and Fmoc-His(Boc)-OH, epimerization was quantified by integrating analytical HPLC traces for both diastereomers (instrument and analysis method: analytical HPLC with Zorbax C18 column, 1–41 gradient over 20 minutes). Variability of approx. ±0.5% can be attributed to the inaccuracy of integration.

A1.9.9.1 Temperatures for activation in 10’ heating loop

First, we examined the dependence of epimerization on the temperature during activation in the 10’ heating loop. We synthesized the two test peptides GCF and FHL using our standard coupling protocol on 80 mg ChemMatrix Rink amide resin (0.49 mmol/g). All amino acids were coupled with different temperatures (90, 85, 80, 70, 60, and 50 °C) in the 10’ heating loop. The D-Cys values obtained from this analysis as a function of temperature as well as the respective EIC traces are plotted in Fig. A1.9 A and B. Approximately 4.5–5.0% D-isomer was observed under standard coupling conditions at 90 °C, which matches the previously reported value.(17) The D -His values obtained from this analysis as a function of temperature as well as the respective HPLC traces are plotted in Fig. A1.9 C and D.

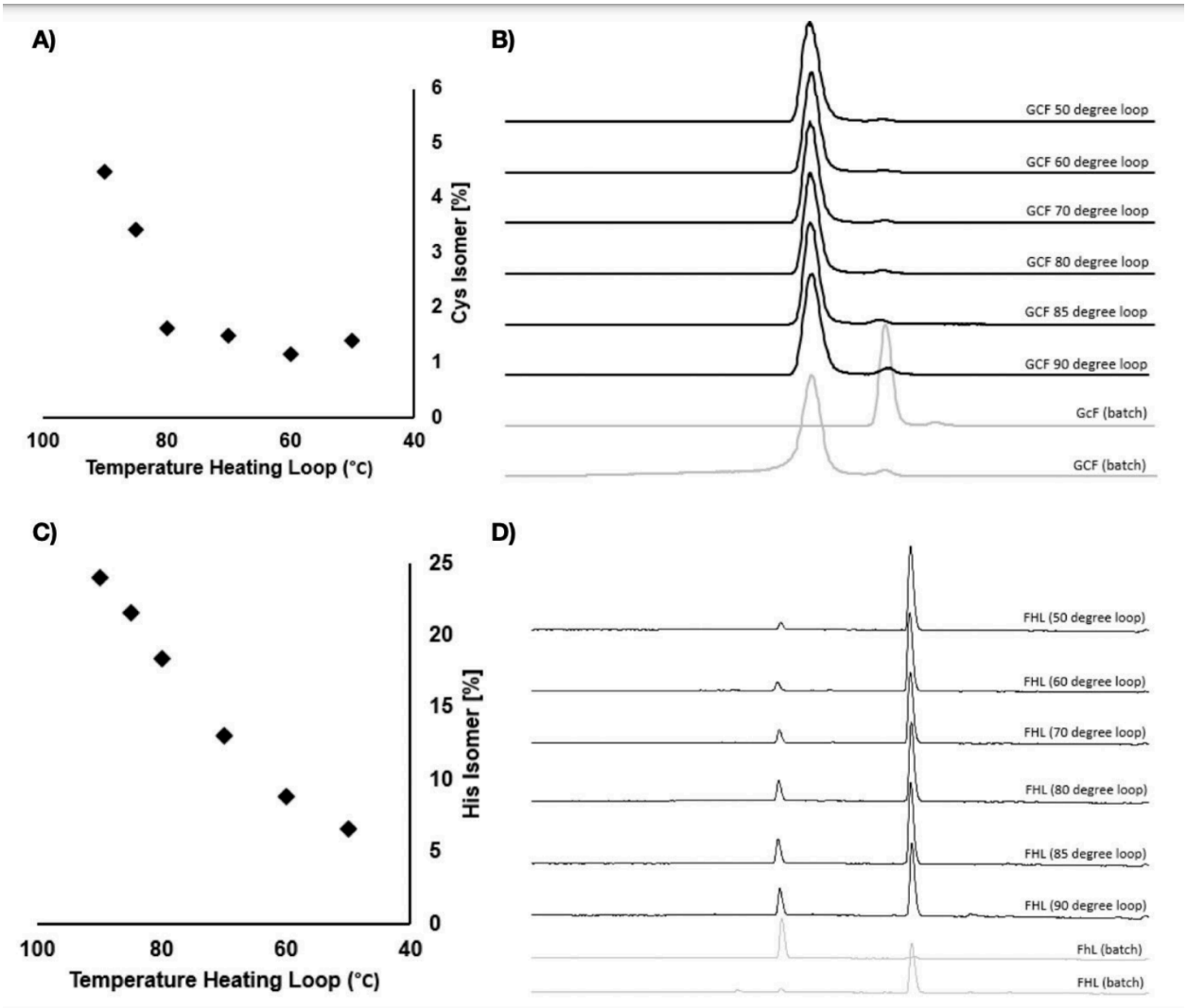


Figure A1.9. **Correlation between temperature and epimerization for cysteine (Fmoc-Cys(Trt)-OH) and histidine (Fmoc-His(Trt)-OH) coupling via a 10' activation loop at different temperatures.** A) Temperature in the heating loop vs. formation of D-Cys isomer in % resulting from Fmoc-Cys(Trt)-OH coupling; B) EIC [m = 325-327] plotted for different temperatures; analysis done on LC-MS 6550 with LunaC18 column, 1-61% B gradient over 18 minutes. Reference peptides (grey) were prepared using standard batch SPPS. C) Temperature in the heating loop vs. formation of D-His isomer in % resulting from Fmoc-His(Trt)-OH coupling; D) analytical HPLC traces plotted for syntheses at different temperatures; analysis was done on analytical HPLC with Zorbax C18 column, 1-41 gradient over 20 minutes. Reference peptides (grey) were prepared using standard batch SPPS.

A1.9.9.2 Temperatures for activation in 5’ heating loop

Next, we optimized conditions for His and Cys coupling by investigating a shorter heating loop, different protecting groups, different activators and different temperatures. GCF and FHL were synthesized on the amidator using our standard coupling protocol (10’ heating loop at 90 °C) on 80 mg ChemMatrix Rink amide resin (0.49 mmol/g), except for His and Cys which were diverted to a 5’ heating loop for activation (under the indicated conditions being studied).

Fmoc-Cys(Trt)-OH was coupled using either HATU or PyAOP as an activator with different temperatures (90, 80, 70, 60, 50, 40, 33 °C) in the 5’ heating loop. For Fmoc-Cys(Trt)-OH 20 epimerization was quantified by integrating EIC [m = 325–327] traces for both diastereomers (instrument and analysis method: LC-MS 6550 with LunaC18 column, 1–61 gradient over 18 minutes). The isomer values obtained from this analysis as a function of temperature as well as the respective EIC traces are plotted in Fig. A1.10. We found that both activators, HATU and PyAOP, led to an equal amount of epimerization, and epimerization was not further reduced below approx. 60 °C, where D-isomer remained approximately 1.5%. This is in accordance with experiments in the 10’ heating loop.

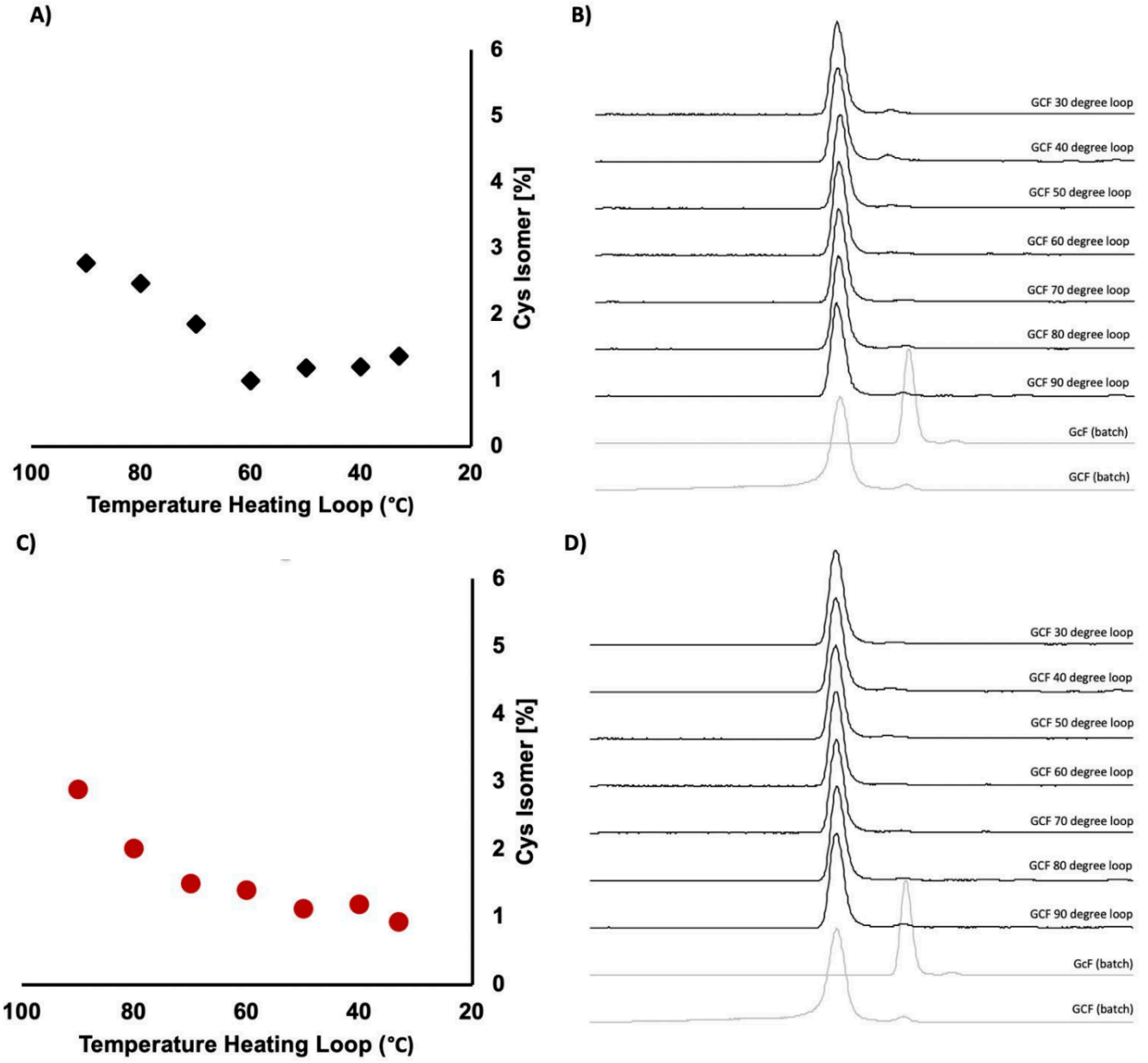


***Figure A1.10*. Correlation between temperature and epimerization for cysteine coupling via a 5' activation loop with different activators and at different temperatures**. A) Temperature in the heating loop vs. formation of Cys isomer in %, all Fmoc-Cys(Trt)-OH couplings with HATU; B) EIC [m = 325-327] plotted for different temperatures; analysis done on LCMS 6550-1 with LunaC18 column, 1-61 gradient over 18 minutes. Reference peptides (grey) were prepared using standard batch SPPS; C) Temperature in the heating loop vs. formation of Cys isomer in %, all FmocCys(Trt)-OH couplings with PyAOP; D) EIC [m = 325-327] plotted for different temperatures; analysis done on LCMS Agilent 6550 with LunaC18 column, 1-61% B gradient over 18 minutes. Reference peptides (grey) were prepared using standard batch SPPS.

Ideal coupling conditions for histidine-coupling on the AFPS systems were investigated next. In the first set of experiments, Fmoc-His(Trt)-OH was coupled using

HATU and PyAOP as an activator with different temperatures (90, 80, 70, 60, 50, 40, 30 °C) in the 5’ heating loop. The epimerization values obtained from this analysis as a function of temperature are plotted in Fig. A1.11.

As histidine-epimerization could not be suppressed further using Fmoc-His(Trt)-OH, we investigated Fmoc-His(Boc)-OH as an alternative building block. Fmoc-His(Boc)-OH was coupled using HATU and PyAOP as an activator with different temperatures (90, 70, 50, 30 °C) in the 5’ heating loop. The isomer values obtained from this analysis as a function of temperature are plotted in Fig. A1.11.

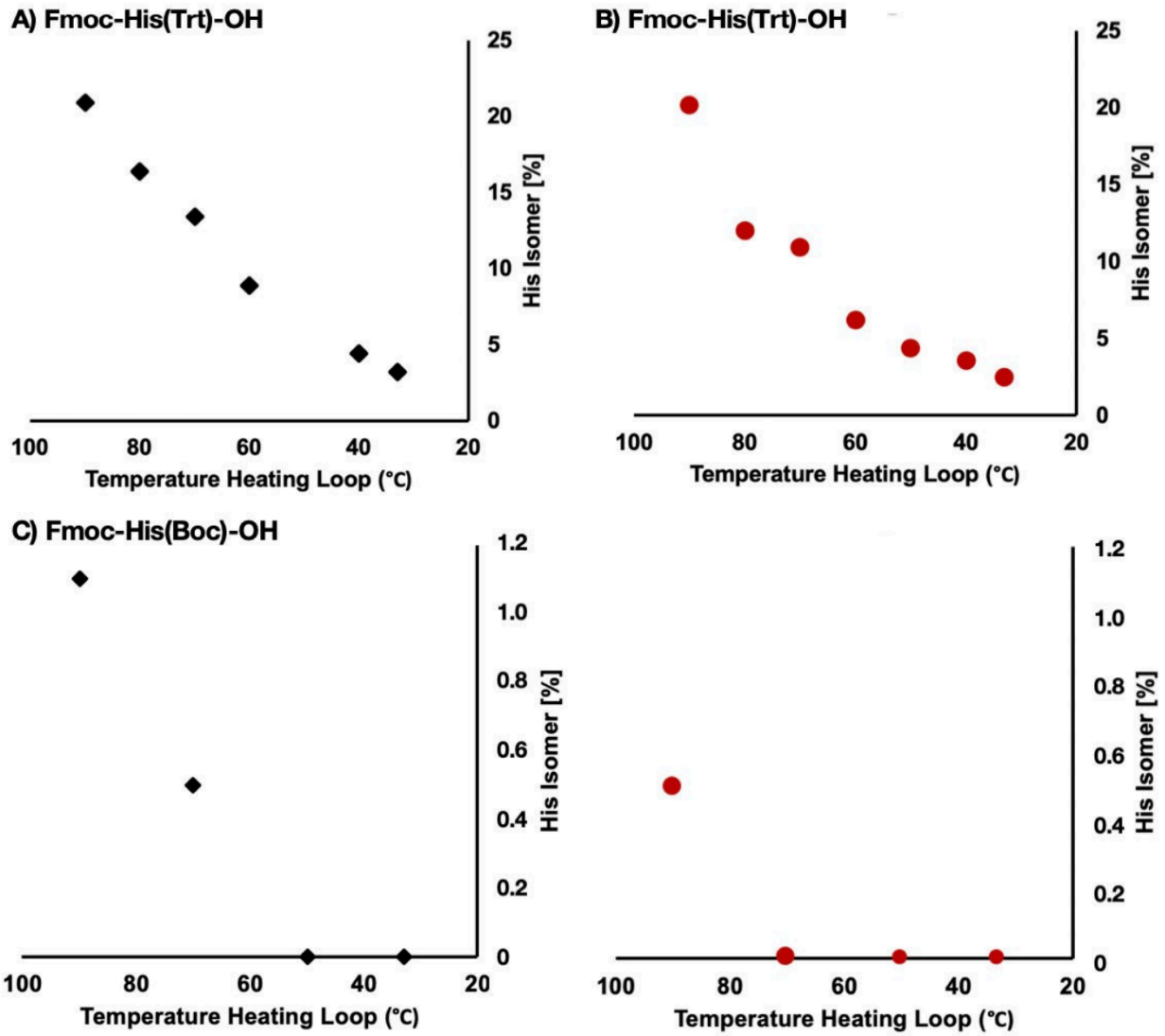


***Figure A1.11*. Correlation between temperature and epimerization coupling of histidine via a 5' activation loop at with different activators and at different temperatures**. A) Coupling of Fmoc-His(Trt)-OH; temperature in the heating loop vs. formation of His isomer in %, all Fmoc-His(Trt)-OH couplings with HATU; B) Coupling of FmocHis(Trt)-OH; temperature in the heating loop vs. formation of His isomer in %, all Fmoc-His(Trt)-OH couplings with PyAOP; C) Coupling of Fmoc-His(Boc)-OH; temperature in the heating loop vs. formation of His isomer in %, all 22 Fmoc-His(Boc)-OH couplings with HATU; D) Coupling of Fmoc-His(Boc)-OH; temperature in the heating loop vs. formation of His isomer in %, all Fmoc-His(Boc)-OH couplings with PyAOP.

A1.9.9.3 Extended coupling time

Then we determined if Fmoc-protected resin-bound Cys(Trt) is prone to epimerization under extended coupling times. GCF was synthesized on the amidator using our standard coupling protocol (heating loop to 90 °C in 10') on 80 mg ChemMatrix Rink amide resin (0.49 mmol/g) with various coupling strokes. The D-Cys values obtained from this analysis as a function of coupling cycles as well as the respective EIC traces are plotted in Fig. S7. Correlation between epimerization and coupling strokes for Fmoc-Cys(Trt)-OH. A) Fmoc-Cys(Trt)-OH coupling strokes vs. formation of Cys isomer in %, the study was performed in duplicates and both data sets are presented in the graph; B) EIC [m = 325-327] plotted for number of Fmoc-Cys(Trt)-OH coupling strokes; only first set of experiment is plotted; analysis done on an Agilent 6550 LC-MS with LunaC18 column, 1-61% B gradient over 18 minutes. Reference peptides (grey) were prepared using standard batch SPPS.

The experiment was carried out twice and experimental results suggest that the unusually high value for 73 coupling strokes leading to approximately 7% epimerization is likely an outlier, as all other values are within an error of +/- 1%. We therefore conclude that extended coupling times do not lead to an increased amount of epimerization.

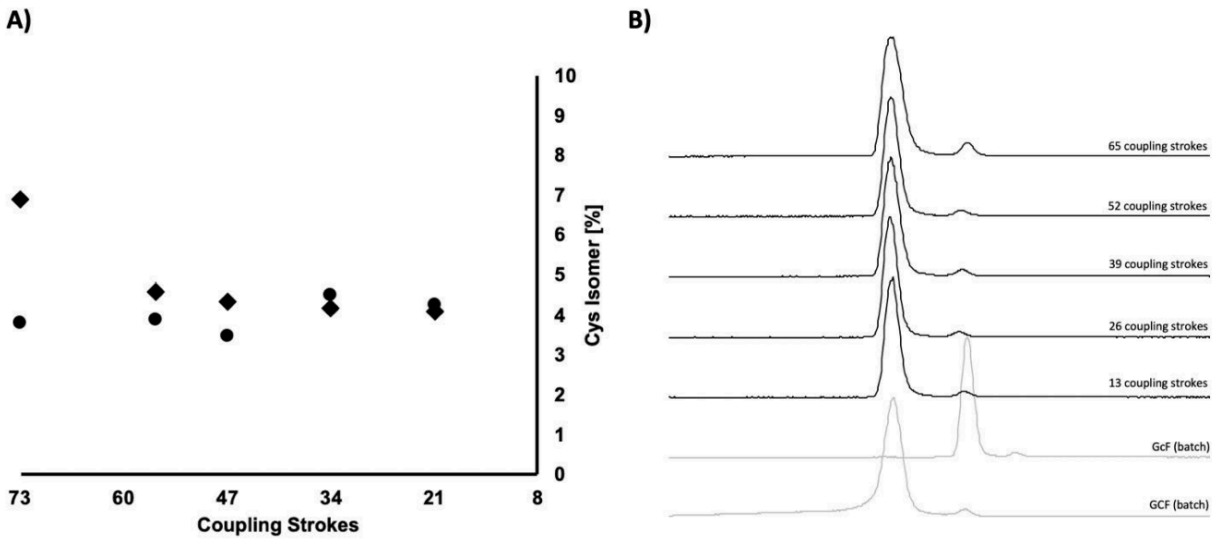


Figure A1.12. **Correlation between epimerization and coupling strokes for Fmoc-Cys(Trt)-OH.** A) Fmoc-Cys(Trt)-OH coupling strokes vs. formation of Cys isomer in %, the study was performed in duplicates and both data sets are presented in the graph; B) EIC [m = 325-327] plotted for number of Fmoc-Cys(Trt)-OH coupling strokes; only first set of experiment is plotted; analysis done on LCMS 6550-1 with LunaC18 column, 1-61 gradient over 18 minutes. Reference peptides (grey) were prepared using standard batch SPPS.

A1.9.9.4 Multiple coupling cycles

GCF and FHL were synthesized on the amidator using our standard coupling protocol (10' heating loop to 90 °C) on 200 mg ChemMatrix Rink amide resin (0.49 mmol/g). The N-terminus was capped with a Boc-protecting group. To simulate the synthesis of long sequences and the effect on cysteine- and histidine-isomerization the resin was exposed to 100 glycine couplings carried out under standard conditions. After every 20 amino acid couplings approximately 20 mg of resin was removed for cleavage and analysis. The D-Cys values obtained from this analysis as 23 a function of coupling cycles as well as the respective EIC traces are plotted in Fig. A1.138 A and B. The initial epimerization did not change over multiple coupling cycles. Variability of ±0.5% can be attributed to the inaccuracy of integration.

For Fmoc-His(Trt)-OH, epimerization was quantified by integrating analytical HPLC traces for both diastereomers. The D-His values obtained from this analysis as a function of coupling cycles as well as the respective HPLC traces are plotted in Fig. A1.13 C and D. The initial epimerization did not change over multiple coupling cycles.

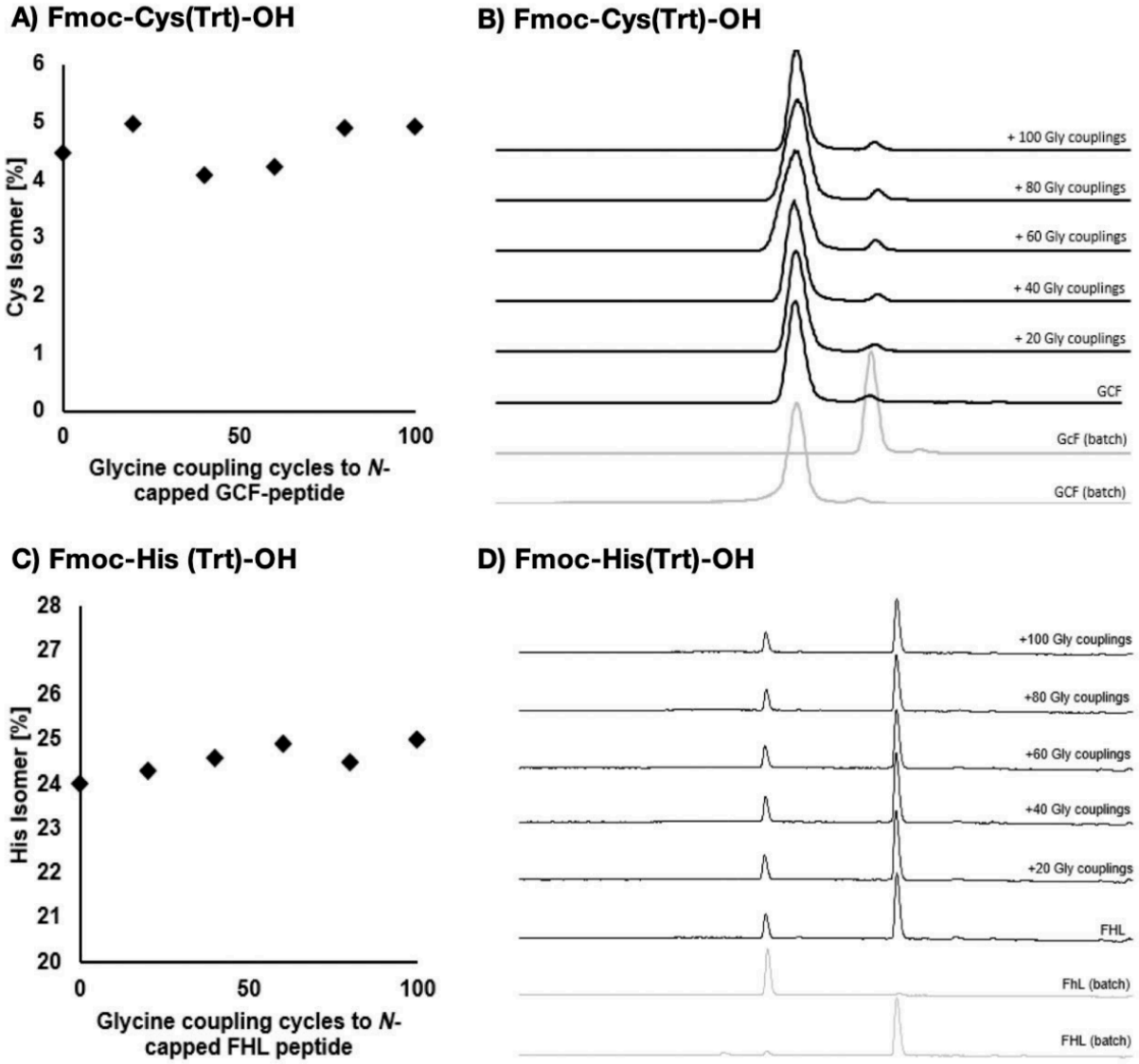


***Figure A1.13*. Correlation between epimerization and number of amino acid coupling cycles for Fmoc-Cys(Trt)-OH and Fmoc-His(Trt)-OH using coupling conditions resulting in highest epimer formation.** Fmoc-Cys(Trt)-OH and Fmoc-His(Trt)-OH were both coupled at 90 °C with HATU through the 10' heating loop. A) Glycine coupling cycles vs. formation of Cys isomer in %; B) EIC [m = 325-327] plotted

for number of Glycine coupling cycles; analysis done on LCMS 6550-1 with LunaC18 column, 1-61 gradient over 18 minutes. Reference peptides (grey) were prepared using standard batch SPPS. C) Glycine coupling cycles vs. formation of His isomer in %; D) analytical HPLC traces plotted for syntheses at different temperatures; analysis was done on analytical HPLC with Zorbax C18 column, 1-41 gradient over 20 minutes. Reference peptides (grey) were prepared using standard batch SPPS.

In the final coupling protocol, Fmoc-Cys(Trt)-OH and Fmoc-His(Boc)-OH were both coupled with PyAOP through the 5' heating loop at 60 °C to achieve a balance between epimerization and coupling efficiency. The iterative coupling experiment was repeated with the final set of synthesis conditions as demonstrated in Fig. A1.14.

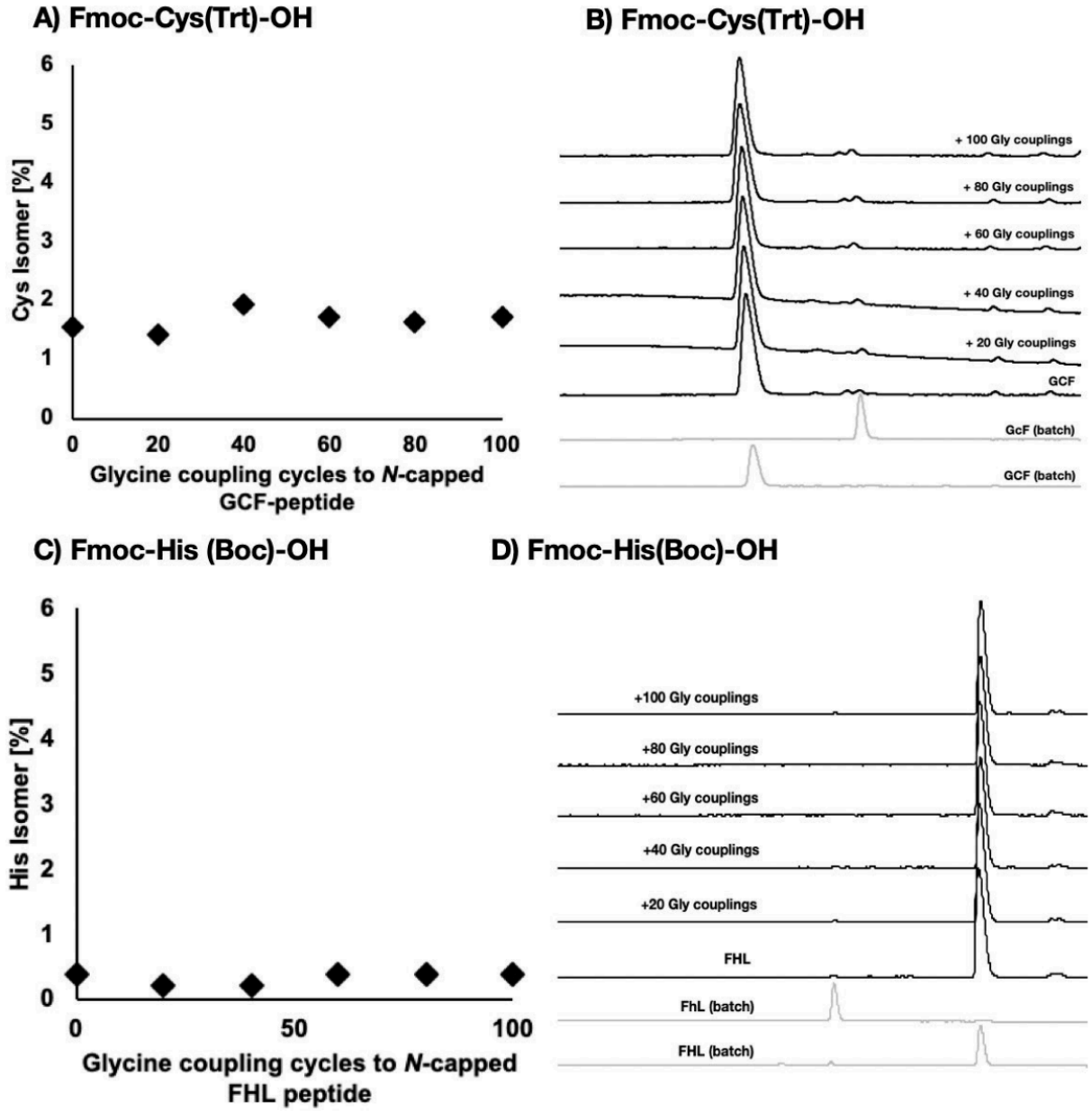


Figure A1.14. **Correlation between epimerization and number of amino acid coupling cycles for Fmoc-Cys(Trt)-OH and Fmoc-His(Boc)-OH using final coupling conditions**. Fmoc-Cys(Trt)-OH and Fmoc-His(Boc)-OH were both coupled with PyAOP at 60 °C through a 5' heating loop. A) Glycine coupling cycles vs. formation of Cys isomer in %; B) analysis was done on analytical HPLC with Zorbax C18 column, 1-41 gradient over 20 minutes. Reference peptides (grey) were prepared using standard batch SPPS. C) Glycine coupling cycles vs. formation of His isomer in %; D) analytical HPLC traces plotted for number of glycine coupling cycles; analysis was done on analytical HPLC with Zorbax C18 column, 1-41 gradient over 20 minutes. Reference peptides (grey) were prepared using standard batch SPPS.

A1.9.10 Final synthesis protocol

Moving forward all peptide sequences were synthesized using the general synthesis protocol developed in section A1.8

Parameter	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) and 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)

***Table A1.11*. Set of optimized synthesis conditions on the AFPS**. Pump strokes refer to volumes described in the general synthesis protocol (section A1.7.2).

A1.9.10.1 GLP-1

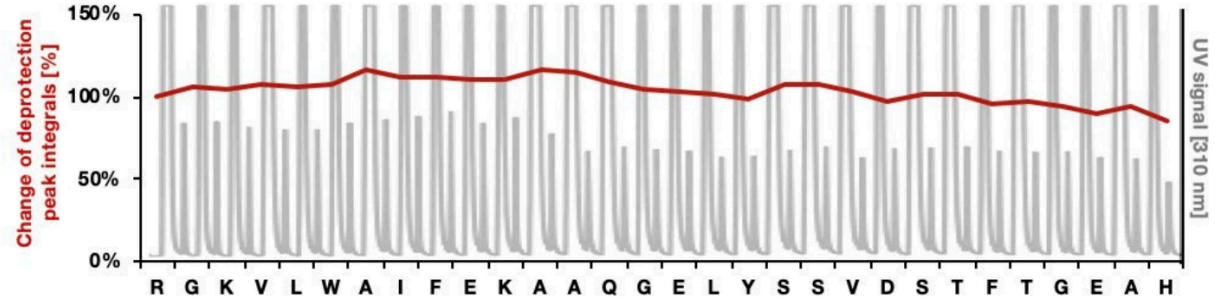
A) Synthesis Data for GLP-1

Sequence: HAEGTFTSDV SSYLEGQAAK EFIAWLVKGR (30 AA)

Resin: 100 mg of RINK amine ChemMatrix® (0.49 mmol/g), yielding the C-terminal amide after cleavage

Synthesis time: 1.1 h

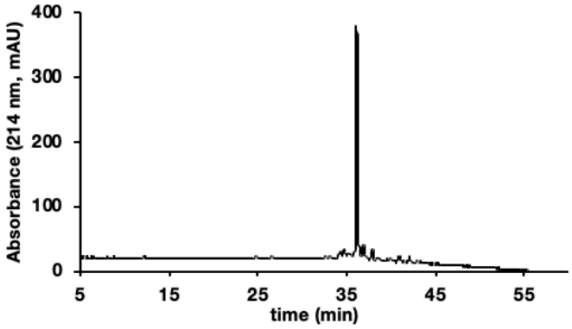
Synthesis UV trace from AFPS:



B) Cleavage and analytical Data for crude GLP-1

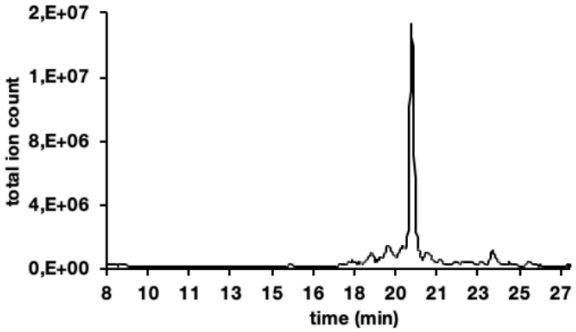
Cleavage: Cleavage protocol A for peptides

Analytical HPLC:



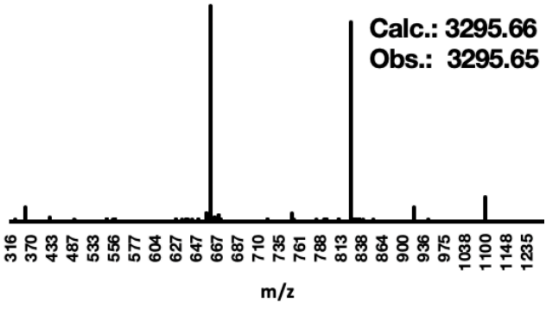
method: 5–65% B over 60 min

LC-MS:



method: 1-61% B over 34 min, Jupiter C4 column (6550)

MS data from LC-MS:



A1.9.10.2 NN92

A) Synthesis Data for NN92

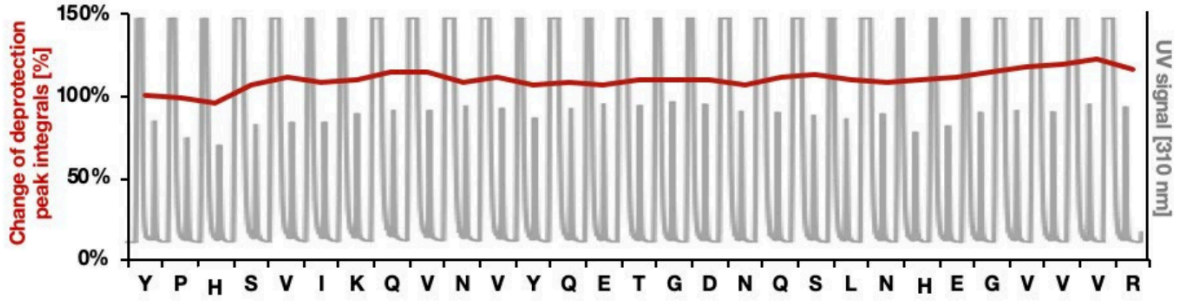
Sequence: RVVVGEHNLS QNDGTEQYVN VQKIVSHPY (29 AA)

Resin: 100 mg of RINK amine ChemMatrix® (0.49 mmol/g), yielding the C-terminal amide after cleavage

Synthesis time: 1.1 h

Synthesis UV trace from AFPS:

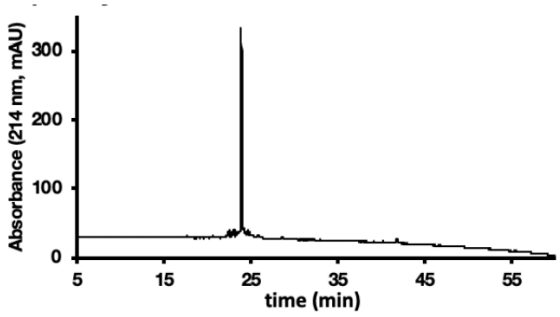
a) UV trace from AFPS



B) Cleavage and analytical Data for crude NN92

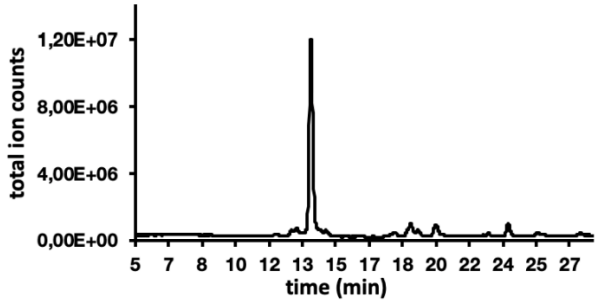
Cleavage: Cleavage protocol A for peptides

Analytical HPLC:



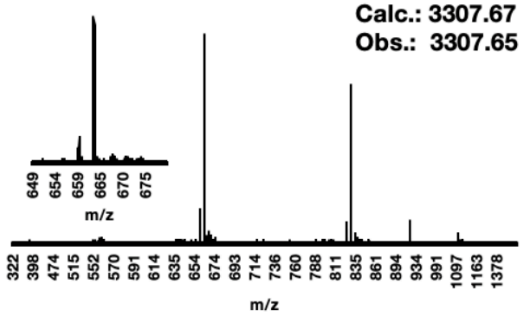
method: 5–65% B over 60 min

LC-MS:



method: 1-61% B over 34 min, Jupiter C4 column (6550)

MS data from LC-MS:



A1.9.10.3 Amyloid beta [1-42]

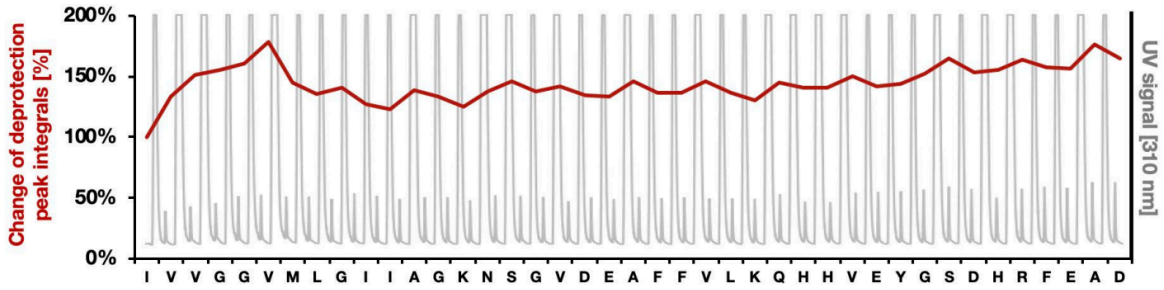
A) Synthesis Data for Abeta

Sequence: DAEFRHDSGY EVHHQKLVFF AEDVGSNKGA IIGLMVGGVV IA (42 AA)

Resin: 90 mg of HMPB ChemMatrix® (0.09 mmol/g) pre-coupled to alanine, yielding the C-terminal carboxylic acid after cleavage

Synthesis time: 1.7 h

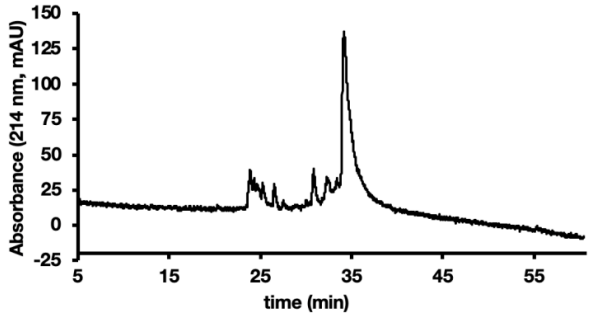
Synthesis UV trace from AFPS:



B) Cleavage and analytical Data for crude Abeta

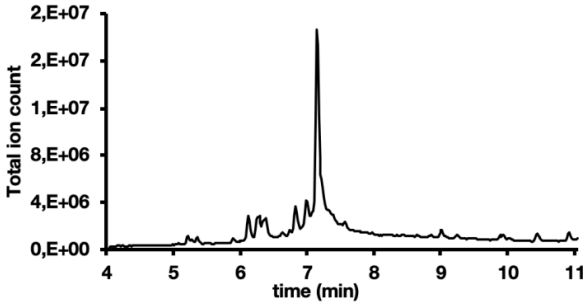
Cleavage: Cleavage protocol A for peptides

Analytical HPLC:



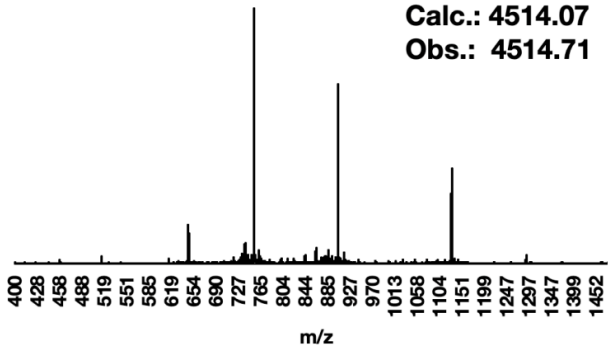
method: 5–65% B over 60 min

LC-MS:



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

MS data from LC-MS:



A1.10 Synthesis of longer peptides - comparison between AFPS and traditional synthesis methods

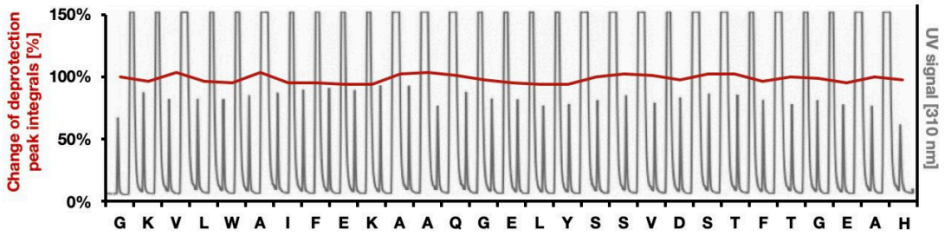
A1.10.1 Synthesis on AFPS

All proteins were synthesized using the general synthesis protocol developed in section A1.9 (unless otherwise noted).

A1.10.1.1 GLP-1

A) Synthesis Data for GLP-1	
Sequence:	HAEGTFTSDV SSYLEGQAAK EFIAWLVKGR (30 AA)
Resin:	140 mg of HMPB ChemMatrix® (0.44 mmol/g) pre-coupled to arginine, yielding the C-terminal carboxylic acid after cleavage.
Synthesis time:	1.1 h

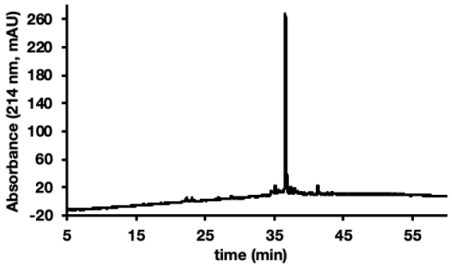
Synthesis UV trace from AFPS:



B) Cleavage and analytical Data for crude GLP-1

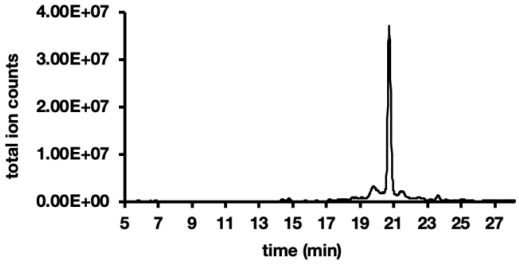
Cleavage: Cleavage protocol A for peptides

Analytical HPLC:



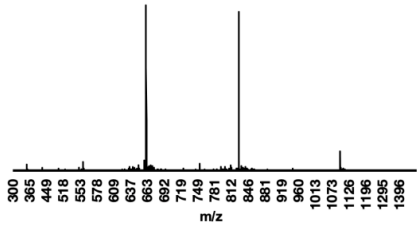
method: 5–65% B over 60 min

LC-MS:



method: 1-61%B over 34 min, Jupiter C4 column (6550)

MS data from LC-MS:

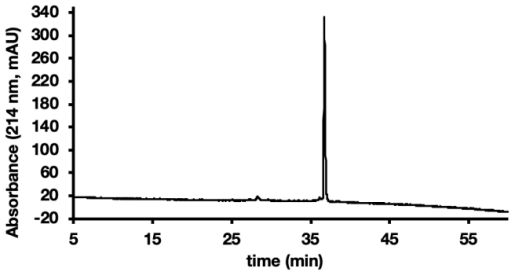


C) HPLC purification and analytical data for GLP-1

HPLC purification: Purified on semipreparative Zorbax 300SB-C3 column at 60 °C and 4 mL/min (gradient: 5–31% B at 1% B/min, 32–52% B at 0.2% B/min, 53–65% B at 1% B/min).

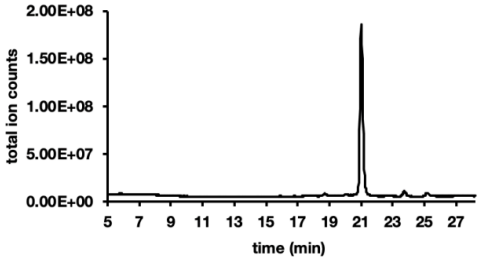
Yield past HPLC: 8 mg (8%)

Analytical HPLC:



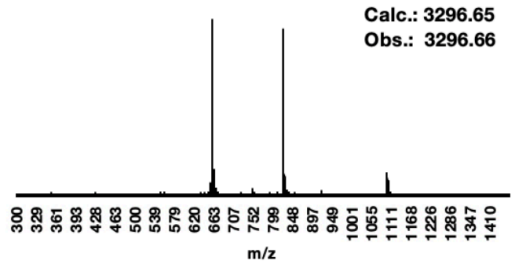
method: 5–65% B over 60 min

LC-MS:



method: 1-61%B over 34 min, Jupiter C4 column (6550)

MS data from LC-MS:



A1.10.1.2 Proinsulin

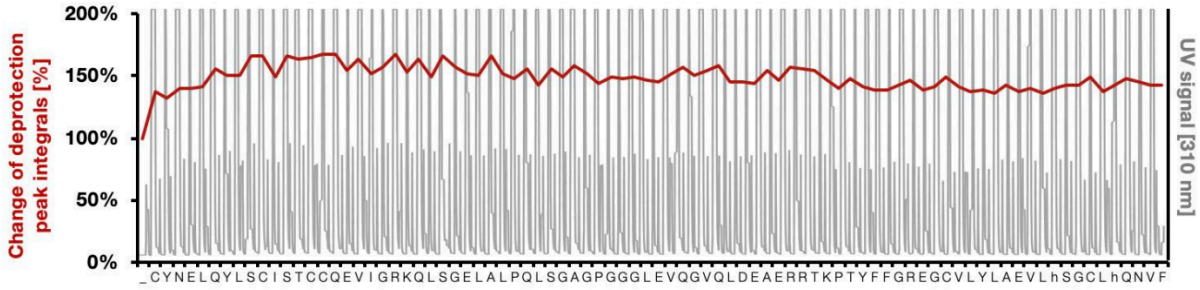
A) Synthesis Data for proinsulin

Sequence: FVNQHLCGSH LVEALYLVCG ERGFFYTPKT RREAEDLQVG QVELGGGPGA GSLQPLALEG SLQKRGIVEQ CCTSICSLYQ LENYCN (86 AA)

Resin: 80 mg of HMPB ChemMatrix® (0.44 mmol/g) pre-coupled to asparagine, yielding the C-terminal carboxylic acid after cleavage

Synthesis time: 3.5 h

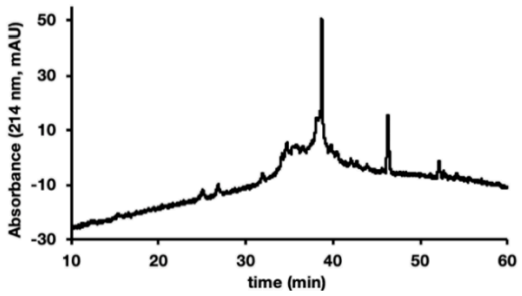
Synthesis UV trace from AFPS:



B) Cleavage and analytical Data for crude proinsulin

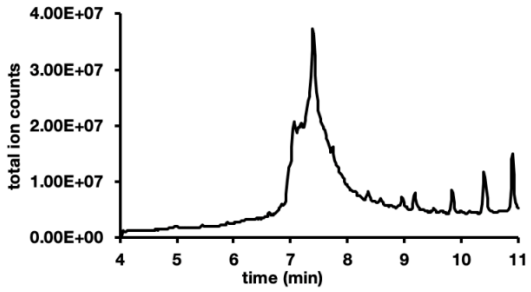
Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

Analytical HPLC:[*)]



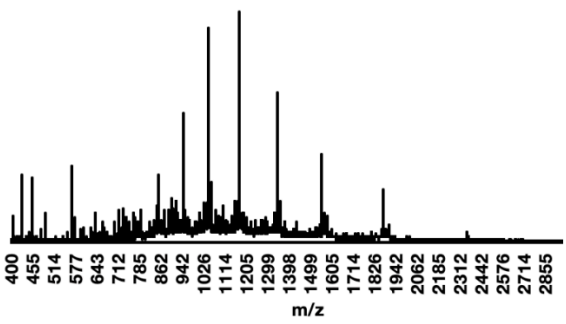
method: 5–65% B over 60 min

LC-MS:

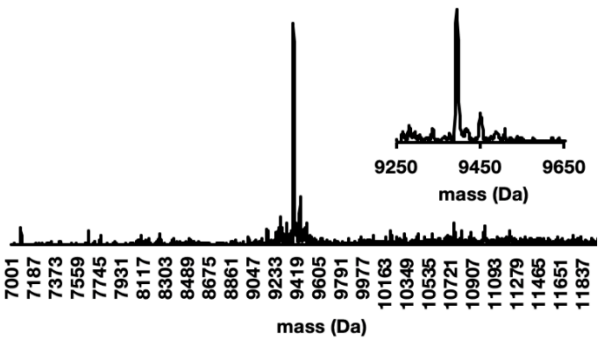


method: 1-91%B over 15 min, Zorbax C3 column (6520)

MS data from LC-MS:



Deconvolution of LC-MS data:

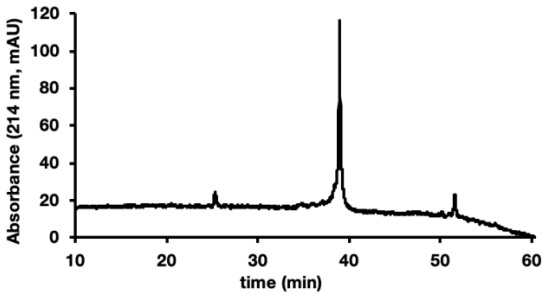


C) HPLC purification and analytical data for proinsulin

HPLC purification: Purified on semipreparative Zorbax 300SB-C3 column at 60 °C and 4 mL/min (gradient: 5–29% B at 1%/min, 29–49% B at 0.2% B/min, 49–65% B at 1% B/min).

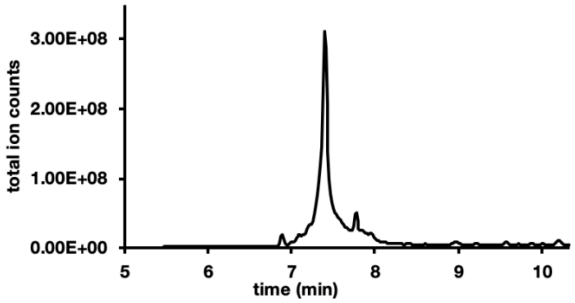
Yield past HPLC: 4.8 mg (1%)

Analytical HPLC:[*)]



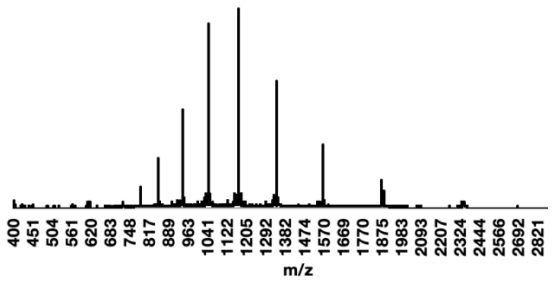
method: 5–65% B over 60 min

LC-MS:

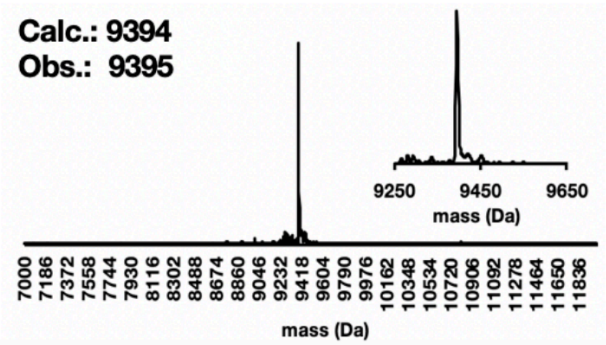


method: 1-91% B over 15 min, Zorbax C3 column (6520)

MS data from LC-MS:



Deconvolution of LC-MS data:



*) peptide is not very soluble, the absorbance of this sample is therefore low

A1.10.1.3 HIV protease (wild type sequence)

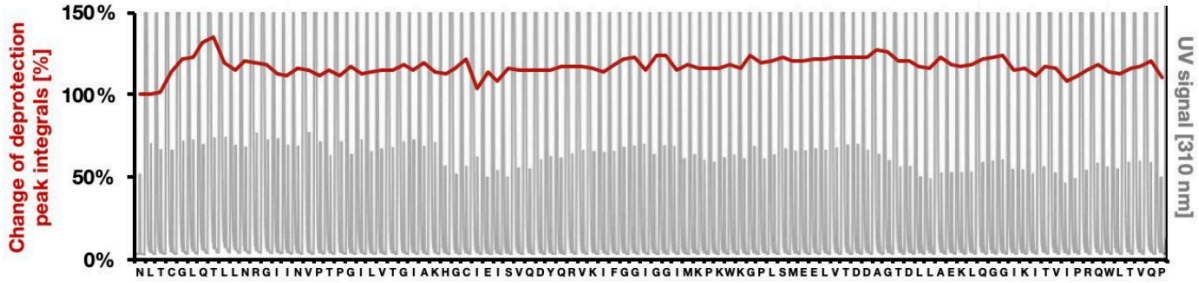
A) Synthesis Data

Sequence: PQVTLWQRPI VTIKIGGQLK EALLDTGADD TVLEEMSLPG KWKPKMIGGI GGFIKVRQYD QVSIEICGHK AIGTVLIGPT PVNIIGRNLL TQLGCTLNF
(99 AA)

Resin: 100 mg of HMPB ChemMatrix® (0.44 mmol/g) pre-coupled to phenylalanine, yielding the C-terminal carboxylic acid after cleavage

Synthesis time: 4.5 h

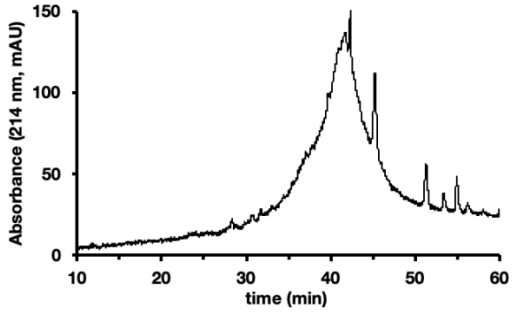
Synthesis UV trace from AFPS:



B) Cleavage and analytical Data for crude HIV-1 protease

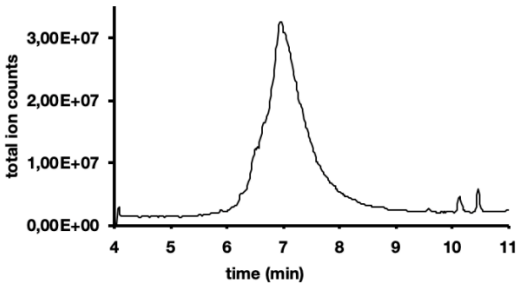
Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

Analytical HPLC:



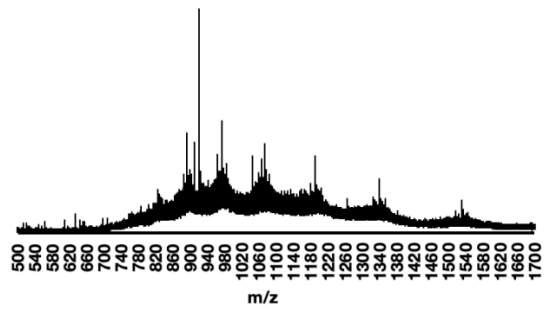
method: 5–65% B over 60 min

LC-MS:

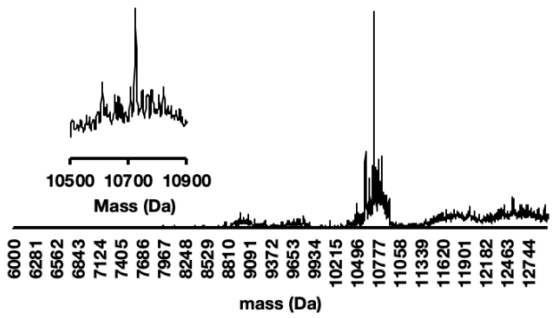


method: 1-91%B over 15 min, Zorbax C3 column (6550)

MS data from LC-MS:



Deconvolution of LC-MS data:

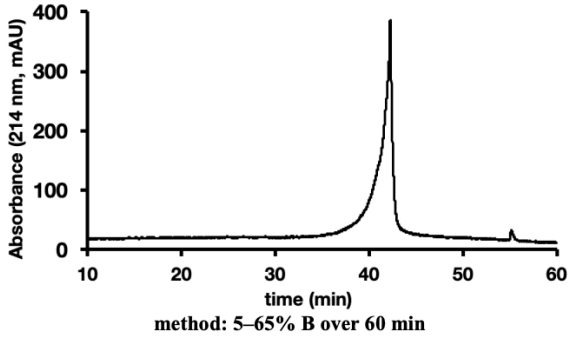


C) HPLC purification and analytical data for HIV-1 protease

HPLC purification: Purified 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)

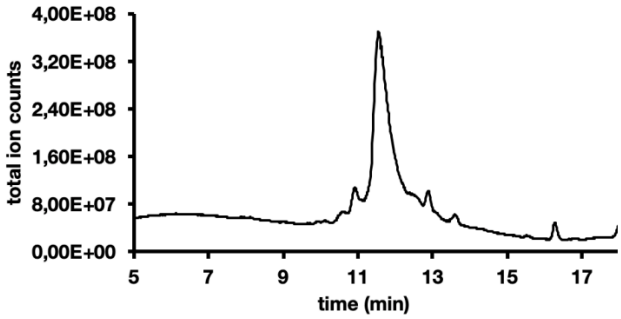
Yield past HPLC: 5.3 mg (1%)

Analytical HPLC:



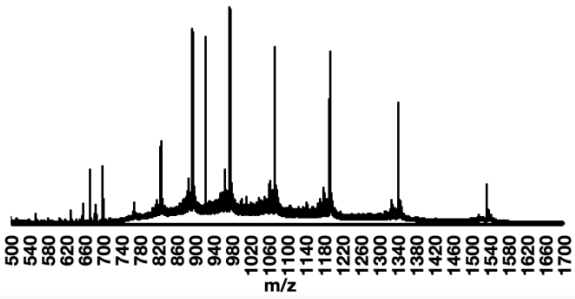
method: 5–65% B over 60 min

LC-MS:

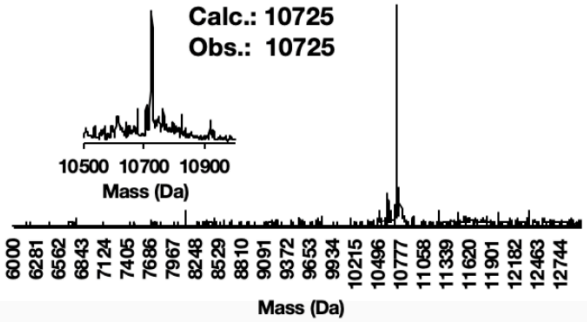


method: 1-91% B over 21 min, Jupiter C4 column (6550)

MS data from LC-MS:



Deconvolution of LC-MS data:



A1.10.2 Synthesis on commercially available synthesizers at r.t., 70 °C and 90 °C

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:

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

Total time per coupling and deprotection cycle: **3 h 37 min**

For 70 and 90 °C syntheses, the general protocol included the following steps:

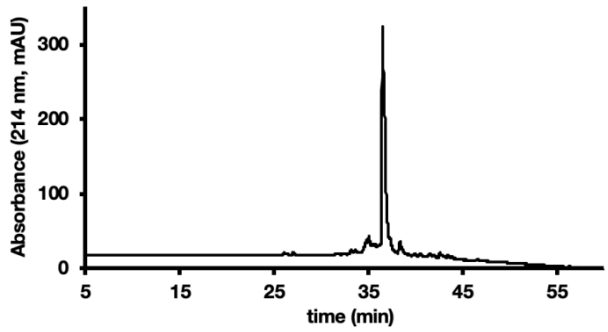
step	Volume (uL)	Mixing time (min)	Heated?	Repeat
Deprotection	3000	1	-	1
DMF wash	1000	0	-	1
Deprotection	3500	2	x	1
DMF wash	1000	0	-	4
DCM wash	3000	0	-	1
DMF wash	3000	0.3	-	3
Amino acid/Oxyma	3000	0	-	1
DIC	300	0	-	1
DCM wash	300	0	-	1
Mix	-	9	x	1
Drain/Dry	-	0	-	1
DCM wash	3000	0	-	1
DMF wash	1000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

Total time per coupling and deprotection cycle: 26 min

A1.10.2.1 GLP-1

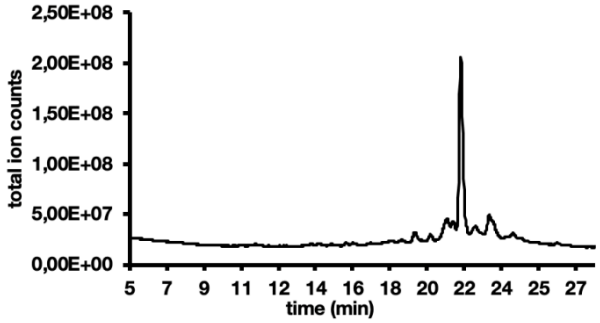
A) Synthesis Data for GLP-1	
Sequence:	HAEGTFTSDV SSYLEGQAAR EFIAWLVRGR GF(32 AA)
Resin:	50 mg of HMPB ChemMatrix® (0.44 mmol/g) pre-coupled to phenylalanine, yielding the C-terminal carboxylic acid after cleavage
Synthesis protocol:	**Commercially available peptide synthesizer at 70 °C**
B) Cleavage and analytical Data for crude GLP-1	
Cleavage:	Cleavage protocol A for peptides

Analytical HPLC:



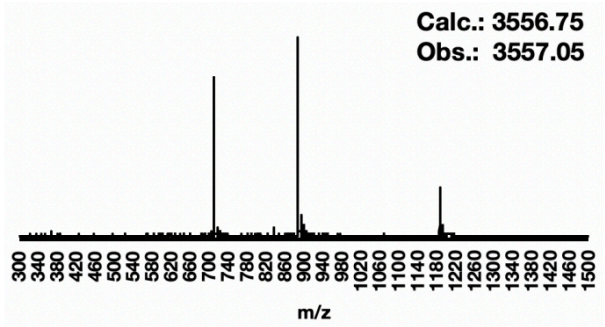
method: 5–65% B over 60 min

LC-MS:



method: 1-61%B over 34 min, Jupiter C4 column (6550)

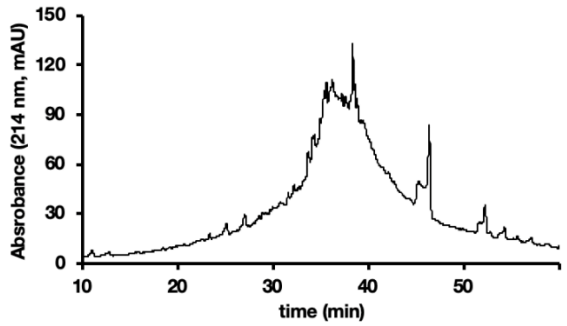
MS data from LC-MS:



A1.10.2.2 Proinsulin

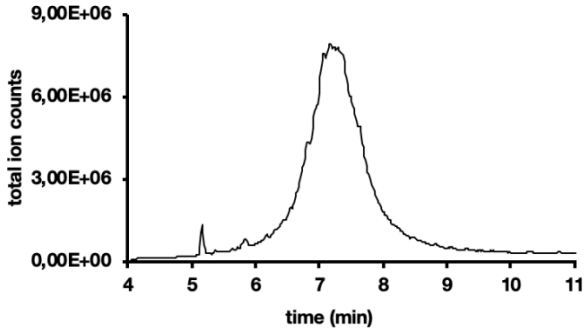
A) Synthesis Data for proinsulin	
Sequence:	FVNQHLCGSH LVEALYLVCG ERGFFYTPKT RREAEDLQVG QVELGGGPGA GSLQPLALEG SLQKRGIVEQ CCTSICSLYQ LENYCN (86 AA)
Resin:	50 mg of HMPB ChemMatrix® (0.44 mmol/g) pre-coupled to asparagine, yielding the C-terminal carboxylic acid after cleavage
Synthesis protocol:	**Commercially available peptide synthesizer at room temperature**
B) Cleavage and analytical Data for crude proinsulin	
Cleavage:	Cleavage protocol B for proteins and Cys-containing peptides

Analytical HPLC:[*)]



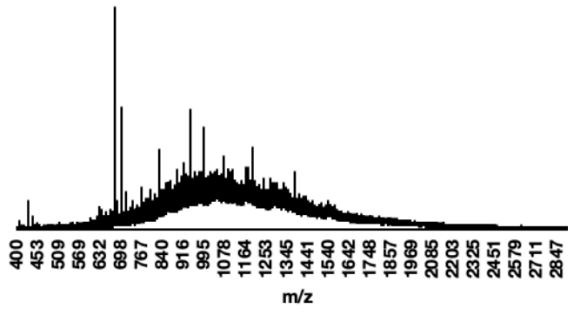
method: 5–65% B over 60 min

LC-MS:

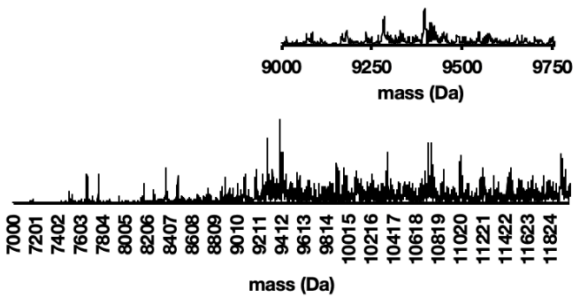


method: 1-91%B over 15 min, Zorbax C3 column (6520)

MS data from LC-MS:



Deconvolution of LC-MS data:



A) Synthesis Data for proinsulin

Sequence: FVNQHLCGSH LVEALYLVCG ERGFFYTPKT RREAEDLQVG QVELGGGPGA GSLQPLALEG SLQKRGIVEQ CCTSICSLYQ LENYCN (86 AA)

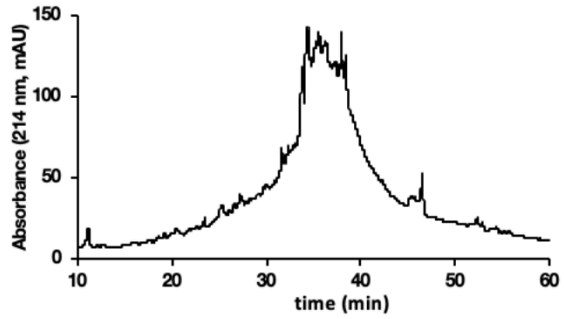
Resin: 50 mg of HMPB ChemMatrix® (0.44 mmol/g) pre-coupled to asparagine, yielding the C-terminal carboxylic acid after cleavage

Synthesis protocol: **Commercially available peptide synthesizer at 70 °C**

B) Cleavage and analytical Data for crude proinsulin

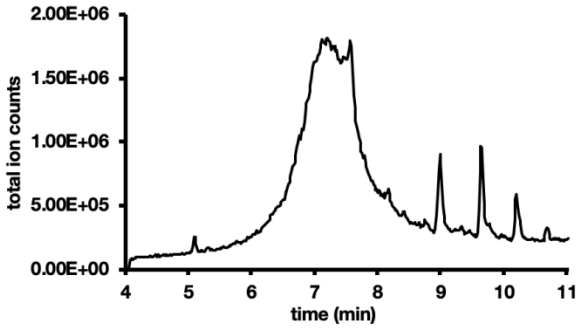
Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

Analytical HPLC:[*)]



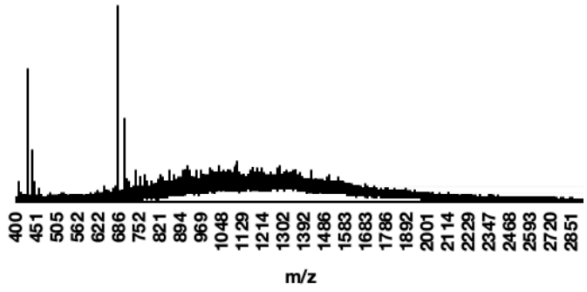
method: 5–65% B over 60 min

LC-MS:

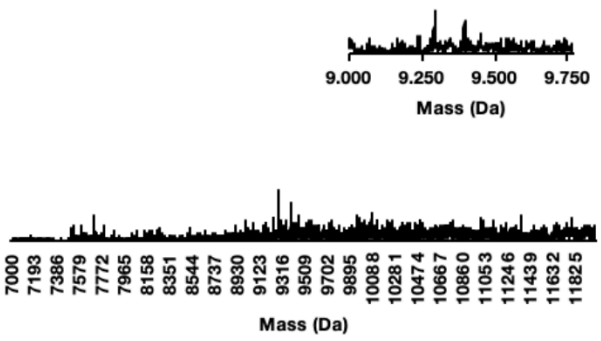


method: 1-91%B over 15 min, Zorbax C3 column (6520)

MS data from LC-MS:



Deconvolution of LC-MS data:



A) Synthesis Data for proinsulin

Sequence: FVNQHLCGSH LVEALYLVCG ERGFFYTPKT RREAEDLQVG QVELGGGPGA GSLQPLALEG SLQKRGIVEQ CCTSICSLYQ LENYCN (86 AA)

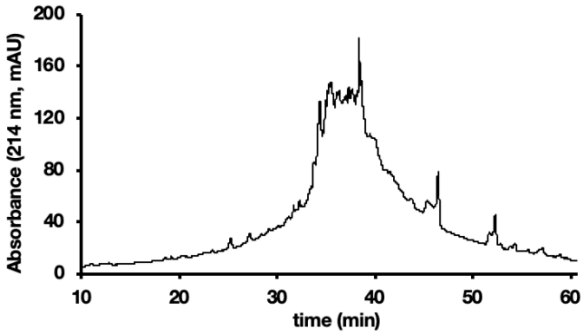
Resin: 50 mg of HMPB ChemMatrix® (0.44 mmol/g) pre-coupled to asparagine, yielding the C-terminal carboxylic acid after cleavage

Synthesis protocol: **Commercially available peptide synthesizer at 90 °C**

B) Cleavage and analytical Data for crude proinsulin

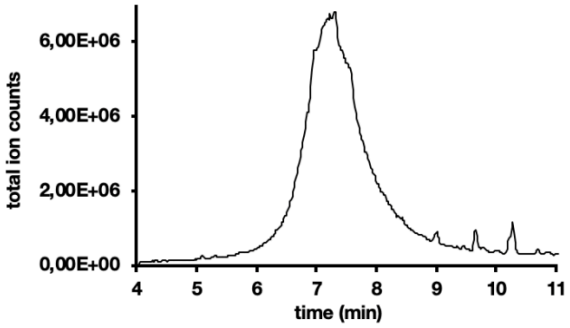
Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

Analytical HPLC:[*)]



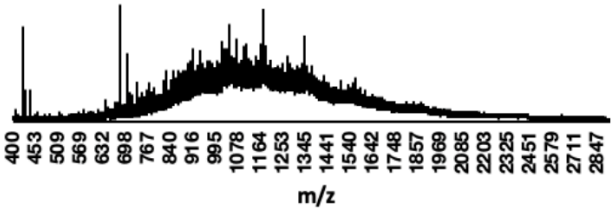
method: 5–65% B over 60 min

LC-MS:

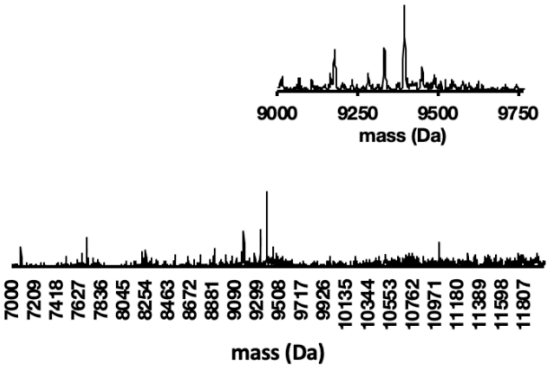


method: 1-91%B over 15 min, Zorbax C3 column (6520)

MS data from LC-MS:



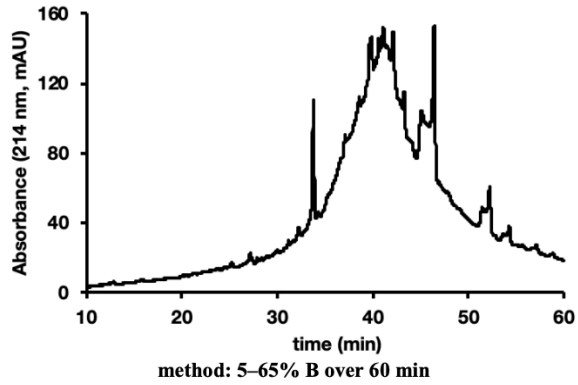
Deconvolution of LC-MS data:



A1.10.2.3 HIV protease (wild type sequence)

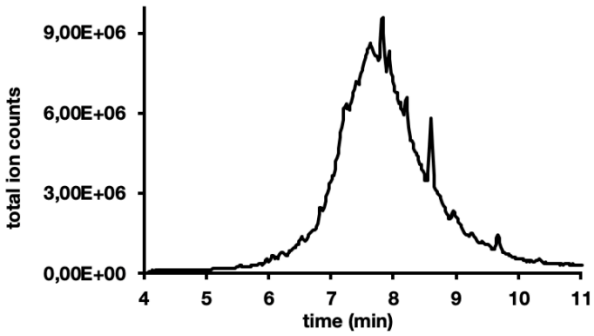
A) Synthesis Data	
Sequence:	PQVTLWQRPI VTIKIGGQLK EALLDTGADD TVLEEMSLPG KWKPKMIGGI GGFIKVRQYD QVSIEICGHK AIGTVLIGPT PVNIIGRNLL TQLGCTLNF (99 AA)
Resin:	50 mg of HMPB ChemMatrix® (0.44 mmol/g) pre-coupled to phenylalanine, yielding the C-terminal carboxylic acid after cleavage
Synthesis protocol:	**Commercially available peptide synthesizer at room temperature**
B) Cleavage and analytical Data for crude HIV-1 protease	
Cleavage:	Cleavage protocol B for proteins and Cys-containing peptides

Analytical HPLC:



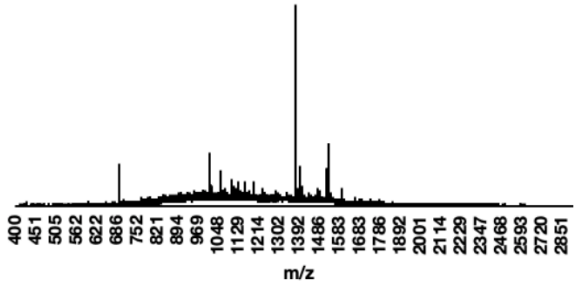
method: 5–65% B over 60 min

LC-MS:

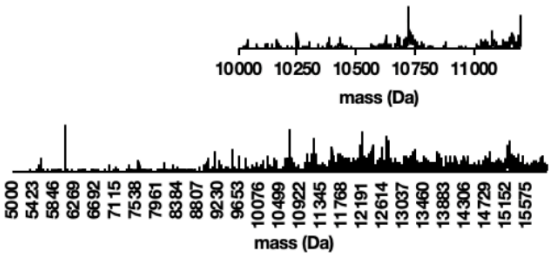


method: 1-91%B over 15 min, Zorbax C3 column (6520)

MS data from LC-MS:



Deconvolution of LC-MS data:



A) Synthesis Data for crude HIV-1 protease

Sequence: PQVTLWQRPI VTIKIGGQLK EALLDTGADD TVLEEMSLPG KWKPKMIGGI GGFIKVRQYD QVSIEICGHK AIGTVLIGPT PVNIIGRNLL TQLGCTLNF (99 AA)

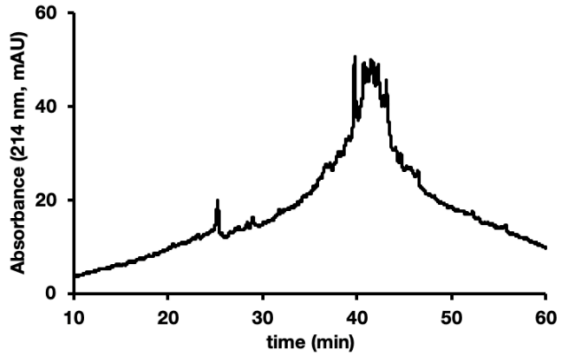
Resin: 50 mg of HMPB ChemMatrix® (0.44 mmol/g) pre-coupled to phenylalanine, yielding the C-terminal carboxylic acid after cleavage

Synthesis protocol: **Commercially available peptide synthesizer at 70 °C**

B) Cleavage and analytical Data for crude HIV-1 protease

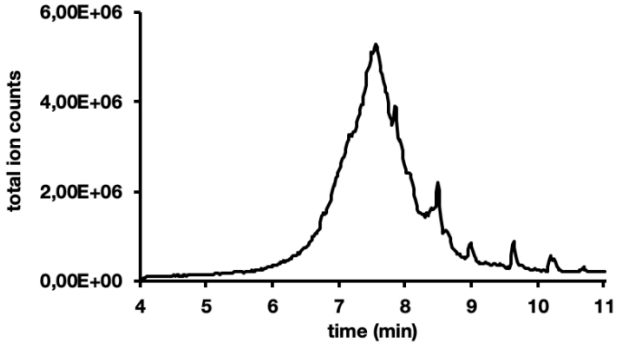
Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

Analytical HPLC:



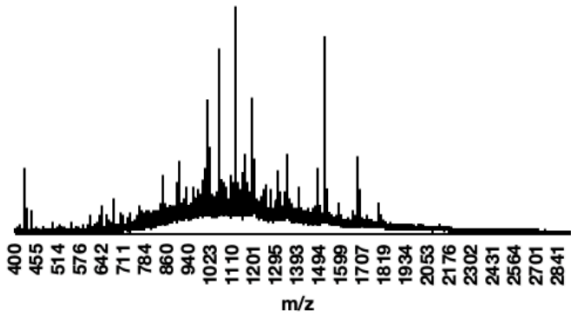
method: 5–65% B over 60 min

LC-MS:

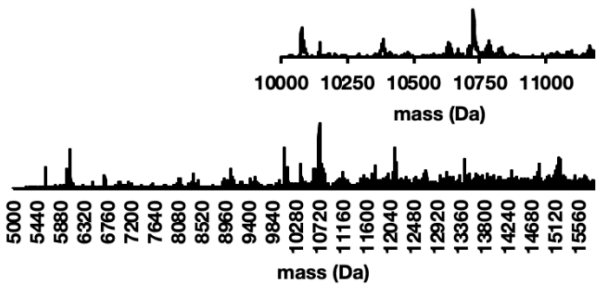


method: 1-91%B over 15 min, Zorbax C3 column (6520)

MS data from LC-MS:



Deconvolution of LC-MS data:



A1.11 AFPS Synthesis of peptides in the length range of single domain proteins

A1.11.1 MDM2[25-109]

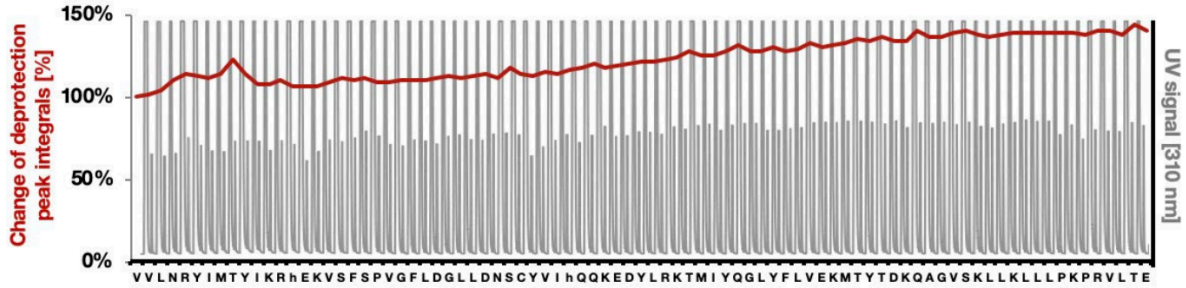
A) Synthesis Data for MDM2[25-109]

Sequence: ETLVRPKPLL LKLLKSVGAQ KDTYTMKEVL FYLGQYIMTK RLYDEKQQHI VYCSNDLLGD LFGVPSFSVK EHRKIYTMIY RNLVV (85 AA)

Resin: 150 mg of RINK amine ChemMatrix® LL (0.18 mmol/g), yielding the C-terminal amide after cleavage

Synthesis time: 3.5 h

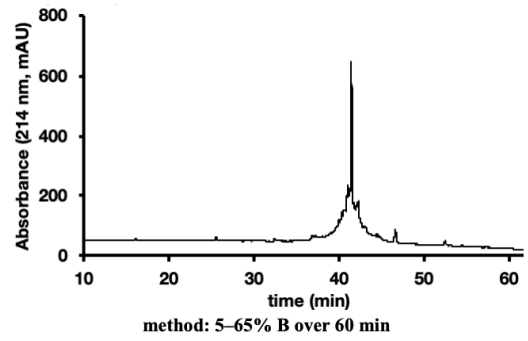
Synthesis UV trace from AFPS:



B) Cleavage and analytical Data for crude MDM2[25-109]

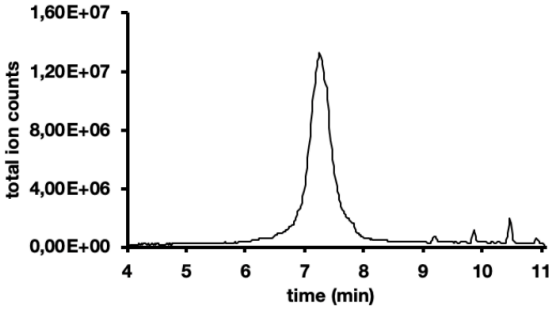
Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

Analytical HPLC:



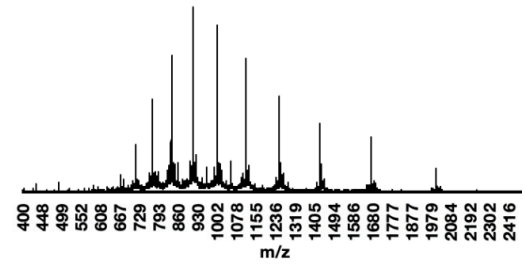
method: 5–65% B over 60 min

LC-MS:

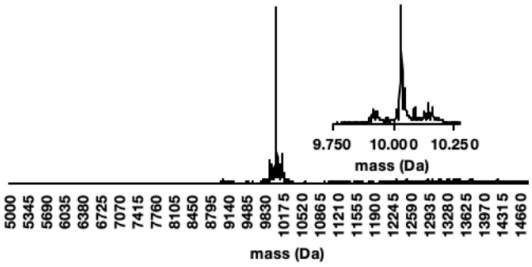


method: 1-91%B over 15 min, Zorbax C3 column (6520)

MS data from LC-MS:



Deconvolution of LC-MS data:

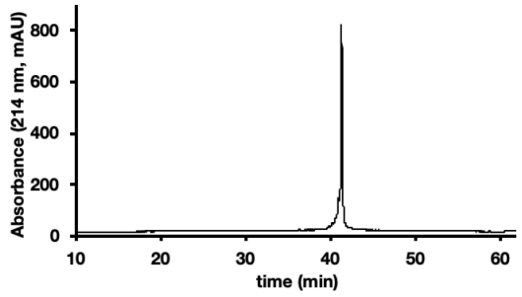


C) HPLC purification and analytical data for MDM2[25-109]

HPLC purification: Purified 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)

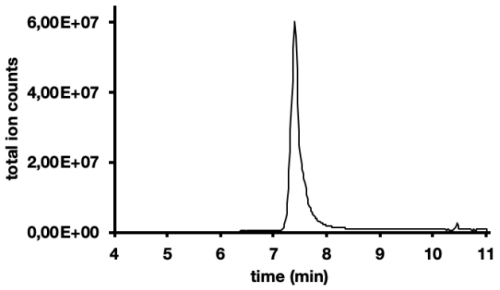
Yield past HPLC: 19.1 mg (6%)

Analytical HPLC:



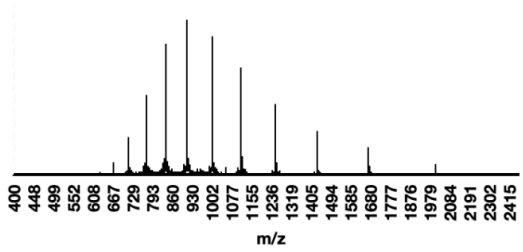
method: 5–65% B over 60 min

LC-MS:

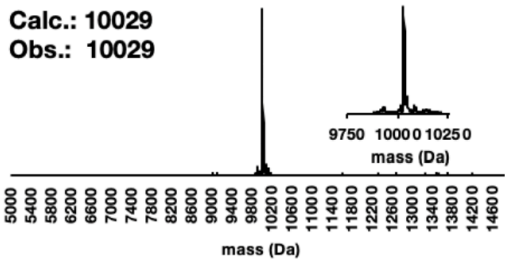


method: 1–91%B over 15 min, Zorbax C3 column (6520)

MS data from LC-MS:



Deconvolution of LC-MS data:



A1.11.2 Barstar

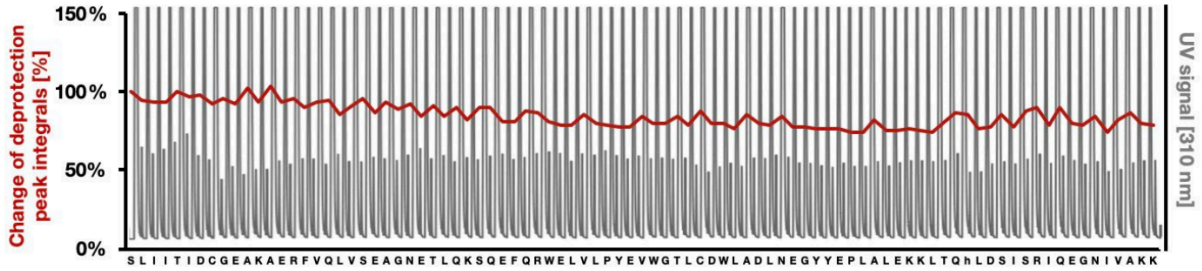
A) Synthesis Data of barstar

Sequence: KKAVINGEQI RSISDLHQTL KKELALPEYY GENLDALWDC LTGWVEYPLV LEWRQFEQSK QLTENGAESV LQVFREAKAE GCDITIILS (89 AA)

Resin: 150 mg of RINK amine ChemMatrix® LL (0.18 mmol/g), yielding the C-terminal amide after cleavage

Synthesis time: 3.5 h

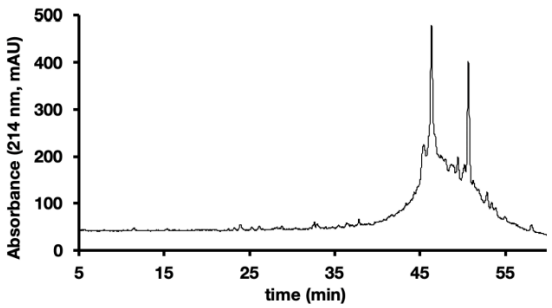
Synthesis UV trace from AFPS:



B) Cleavage and analytical Data for barstar

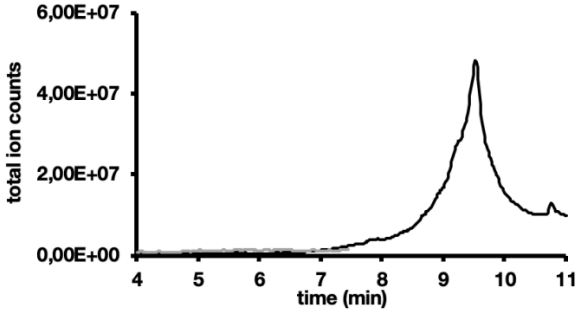
Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

Analytical HPLC:



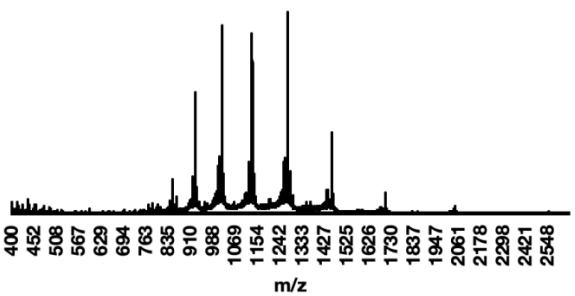
method: 5–65% B over 60 min

LC-MS:

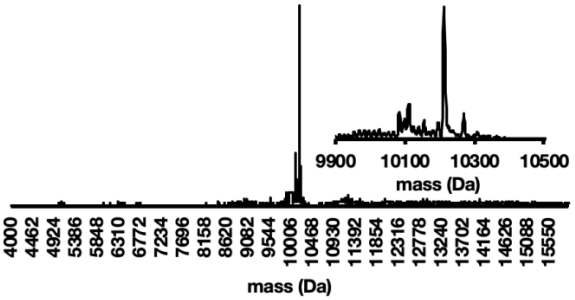


method: 1-91%B over 15 min, Zorbax C3 column (6520)

MS data from LC-MS:



Deconvolution of LC-MS data:

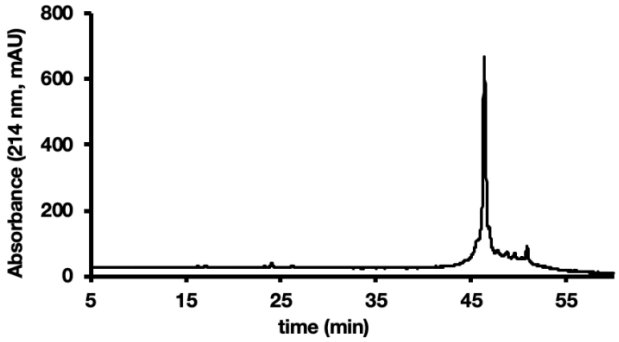


C) HPLC purification and analytical data for barstar

HPLC purification: Purified on a semipreparative Zorbax 300SB-C3 column at 60 °C (gradient: 5–43% B with 1.5% B/min, 43–63% B with 0.2% B/min, 63–73% B with 1% B/min)

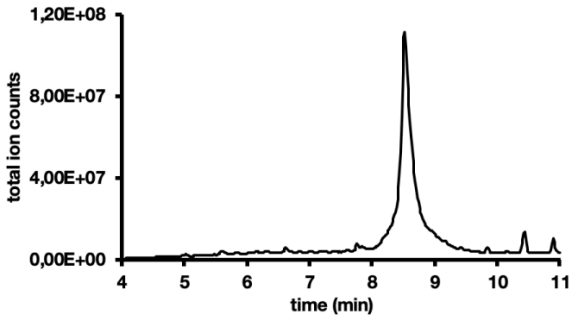
Yield: 5.5 mg (2%)

Analytical HPLC:



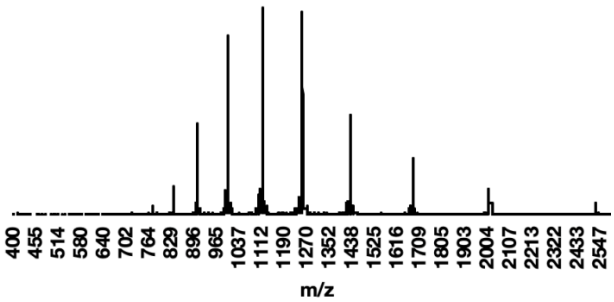
method: 5–65% B over 60 min

LC-MS:

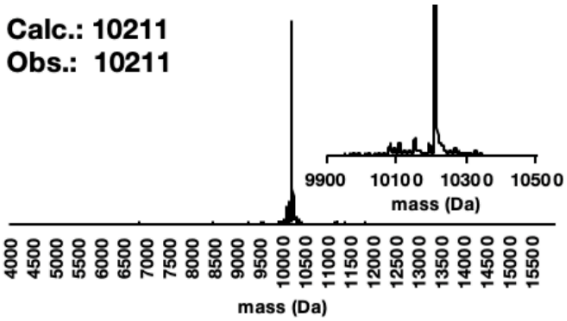


method: 1–91%B over 15 min, Zorbax C3 column (6520)

MS data from LC-MS:



Deconvolution of LC-MS data:



Comments: *) Desired deconvoluted mass is the primary peak in both crude and purified samples.

A1.11.3 Collagen

A) Synthesis Data

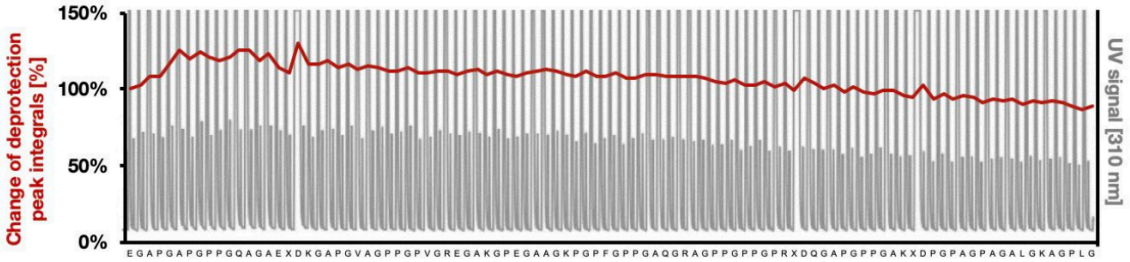
Sequence: GLPGAKGLAG APGAPGPDGK AGPPGPAGQD GRPGPPGPPG ARGQAGPPGF. PGPKGAAGEP GKAGERGVPG PPGAVGPAGK DGEAGAQGPP GPAGPAGER

Note: the three glycines (G) highlighted in blue were introduced as Fmoc-(Dmb)Gly-OH building blocks

Resin: 100 mg of HMPB ChemMatrix® (0.44 mmol/g) pre-coupled to arginine, yielding the C-terminal carboxylic acid after cleavage

Synthesis time: 4.0 h

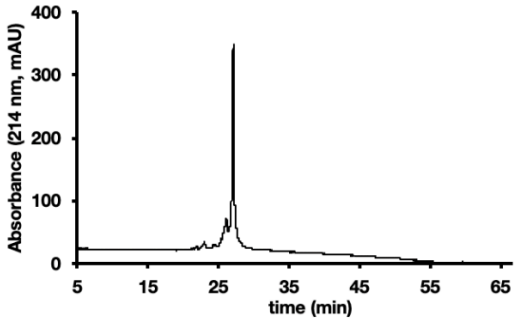
Synthesis UV trace from AFPS:



B) Cleavage and analytical Data for crude collagen

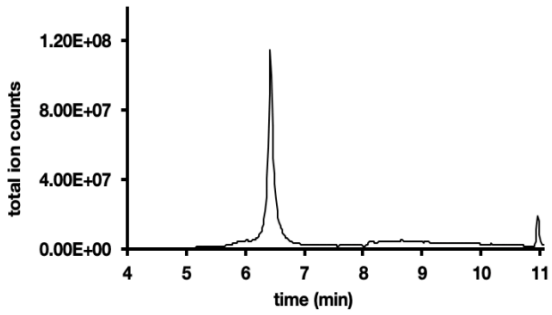
Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

Analytical HPLC:



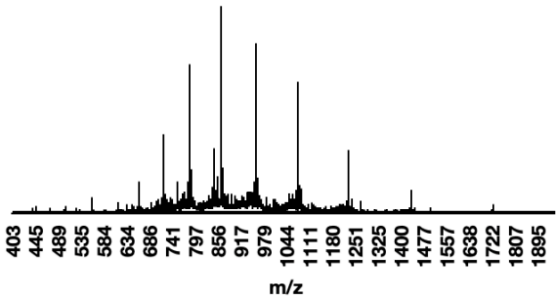
method: 1–61% B over 60 min

LC-MS:

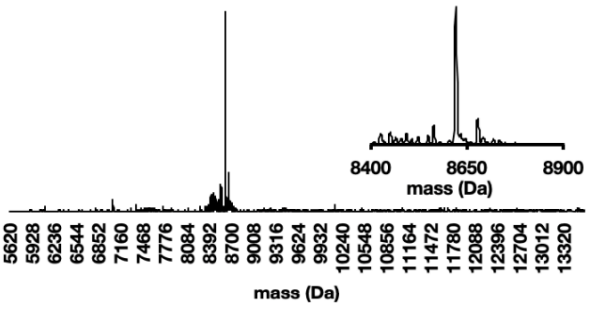


method: 1–91%B over 15 min, Zorbax C3 column (6520)

MS data from LC-MS:



Deconvolution of LC-MS data:

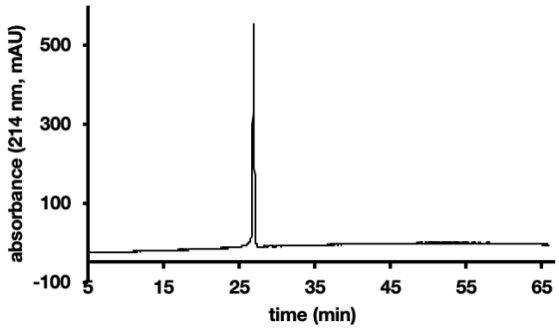


C) HPLC purification and analytical data for collagen

HPLC purification: Purified on a semipreparative Zorbax 300SB-C3 column at 60 °C (gradient: 1–31% B with 0.5% B/min)

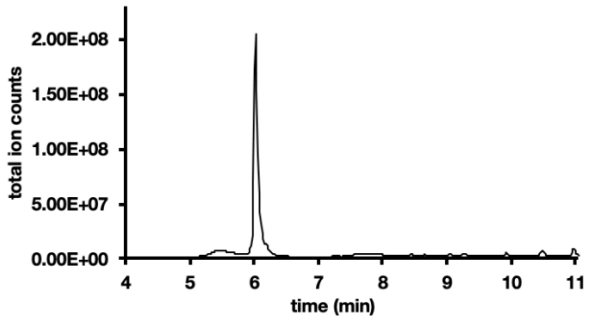
Yield past HPLC: 13.5 mg (3%)

Analytical HPLC:



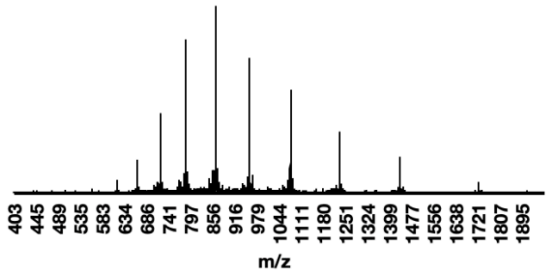
method: 1–61% B over 60 min

LC-MS:

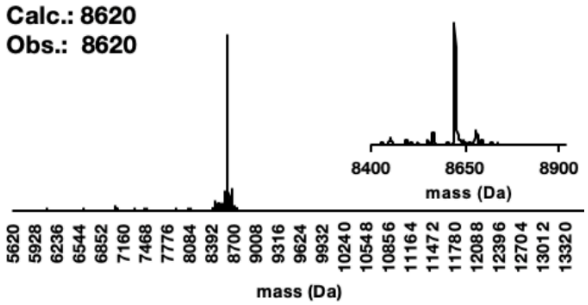


method: 1–91%B over 15 min, Zorbax C3 column (6520)

MS data from LC-MS:



Deconvolution of LC-MS data:



A1.11.4 HIV protease (Kent sequence)

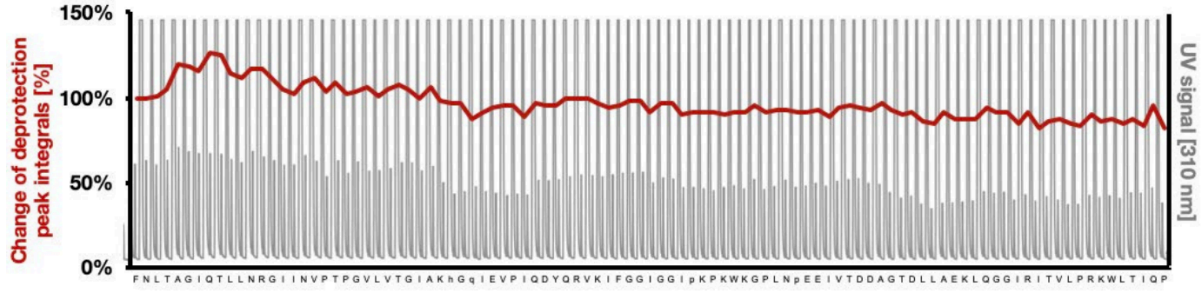
A) Synthesis Data

Sequence: PQITLWKRPL VTIRIGGQLK EALLDTGADD TVIEE(Nle)NLPG KWKPK(Nle)IGGI GGFIKVRQYD QIPVEI(Abu)GHK AIGTVLVGPT PVNIIGRNLL TQIGATLNF (99AA)
(Nle = norleucine, Abu = aminobutyric acid)

Resin: 130 mg of RINK amine ChemMatrix® LL (0.18 mmol/g), yielding the C-terminal amide after cleavage

Synthesis time: 4.5 h

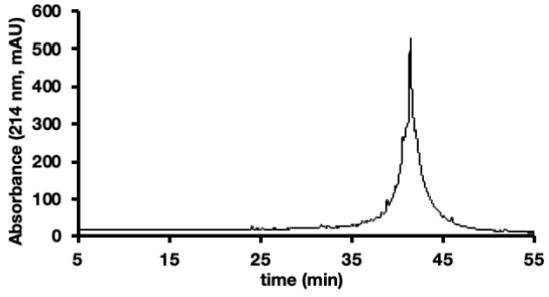
Synthesis UV trace from AFPS:



B) Cleavage and analytical Data for crude HIV-1 protease

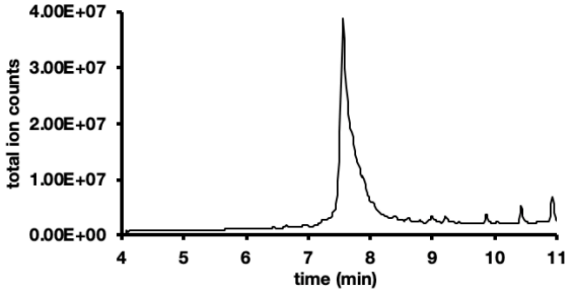
Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

Analytical HPLC:



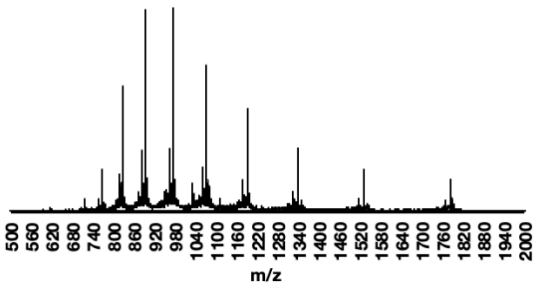
method: 5–65% B over 60 min

LC-MS:

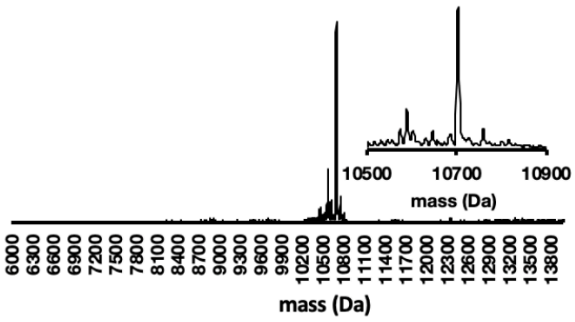


method: 1–91%B over 15 min, Zorbax C3 column (6520)

MS data from LC-MS:



Deconvolution of LC-MS data:

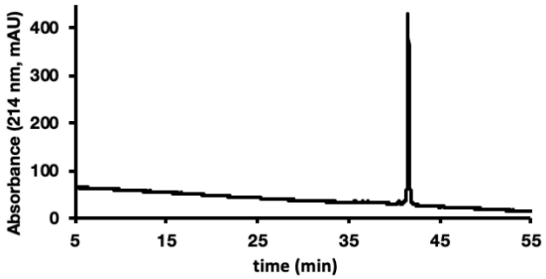


C) HPLC purification and analytical data for HIV-1 protease

HPLC purification: Purified 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)

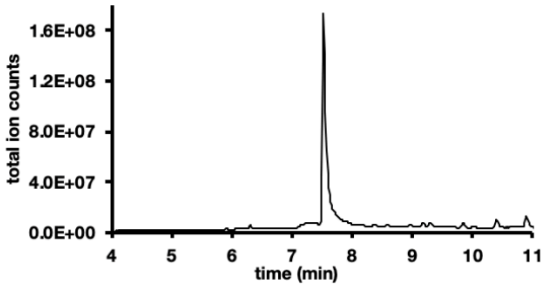
Yield past HPLC: 9.6 mg (3%)

Analytical HPLC:



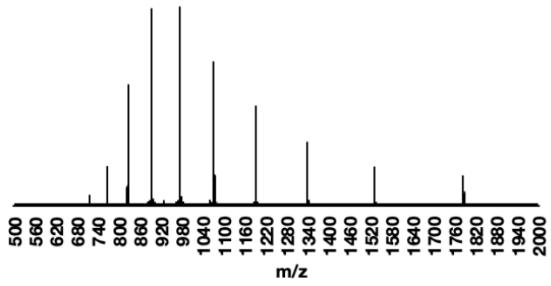
method: 5–65% B over 60 min

LC-MS:

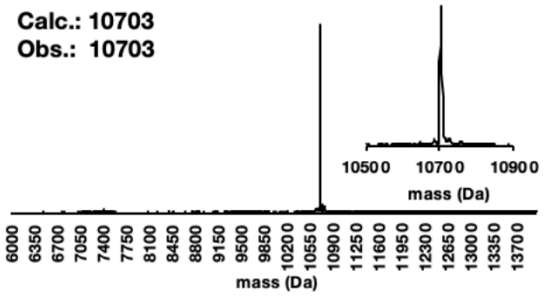


method: 1–91%B over 15 min, Zorbax C3 column (6520)

MS data from LC-MS:



Deconvolution of LC-MS data:



A1.11.5 Barnase

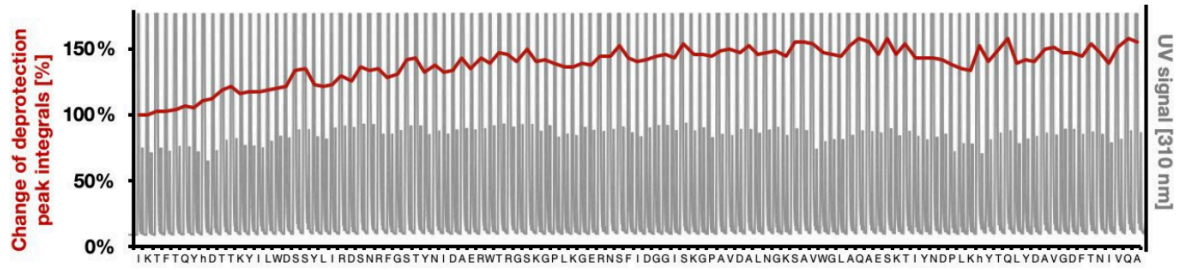
A) Synthesis Data

Sequence: AQVINTFDGV ADYLQTYHKL PDNYITKSEA QALGWVASKG NLADVAPGKS IGGDIFSNRE GKLPGKSGRT WREADINYTS GFRNSDRILY SSDWLIYKTT DHYQTFTKIR (110 AA)

Resin: 50 mg of HMPB ChemMatrix® (0.44 mmol/g) pre-coupled to arginine, yielding the C-terminal carboxylic acid after cleavage

Synthesis time: 4.5 h

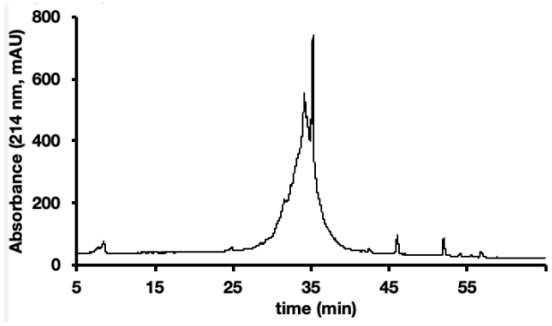
Synthesis UV trace from AFPS:



B) Cleavage and analytical Data for crude barnase

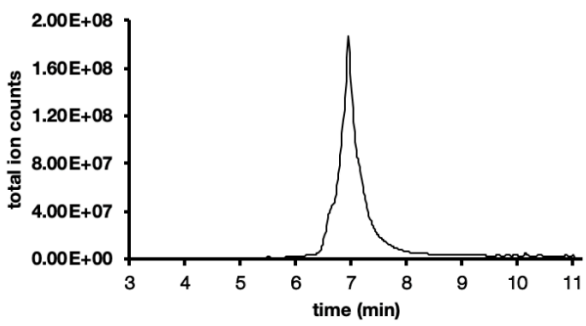
Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

Analytical HPLC:



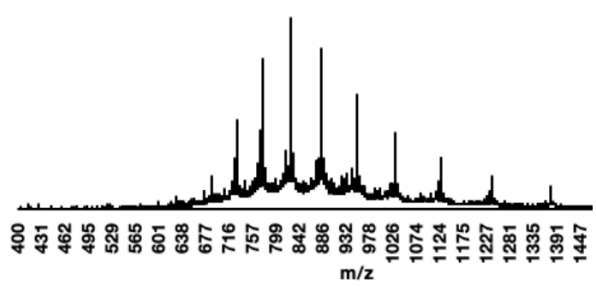
method: 1–61% B over 60 min

LC-MS:

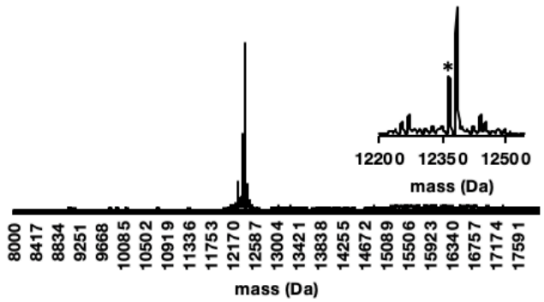


method: 1–91%B over 15 min, Zorbax C3 column (6520)

MS data from LC-MS:



Deconvolution of LC-MS data:

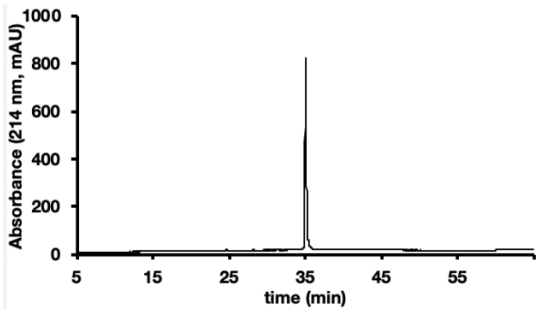


C) HPLC purification and analytical data for barnase

HPLC purification: Zorbax SB300 C3 column, 60 °C
Purified on a semipreparative Zorbax 300SB-C3 column at 60 °C (gradient: 5–25% B with 1% B/min, 25–45% B with 0.2% B/min, 45–65% B with 1% B/min)

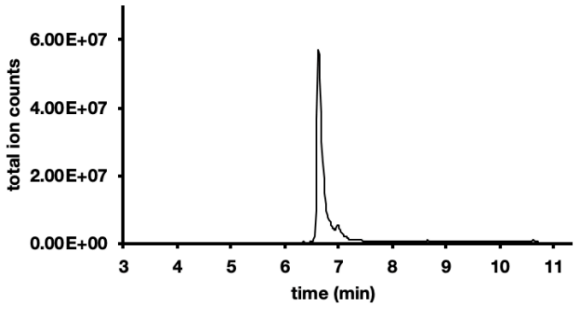
Yield past HPLC: 8.4 mg (2%)

Analytical HPLC:



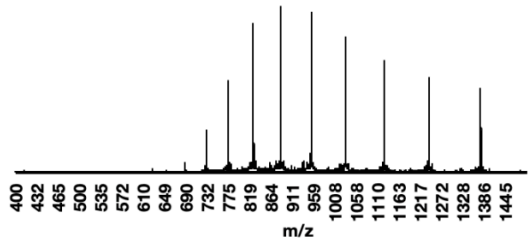
method: 1–61% B over 60 min

LC-MS:

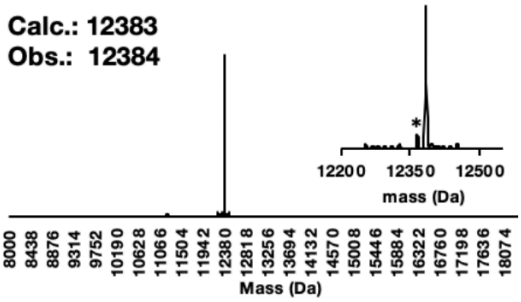


method: 1–91%B over 15 min, Zorbax C3 column (6520)

MS data from LC-MS:



Deconvolution of LC-MS data:



Comments: *) aspartimide by-product

A1.11.6 Barnase ($R110F_{Br}$)

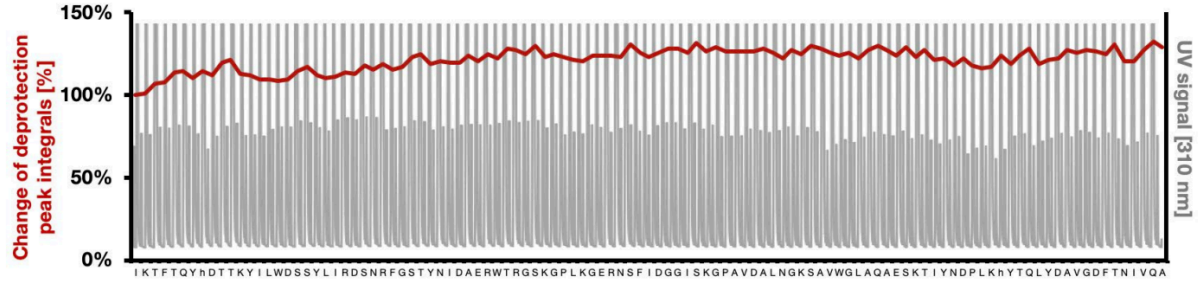
A) Synthesis Data

Sequence: AQVINTFDGV ADYLQTYHKL PDNYITKSEA QALGWVASKG NLADVAPGKS IGGDIFSNRE GKLPGKSGRT WREADINYTS GFRNSDRILY SSDWLIYKTT DHYQTFTKIF$_{Br}$ (110 AA)
(F_{Br} = *p*-bromo phenylalanine)

Resin: 80 mg of HMPB ChemMatrix® (0.44 mmol/g) pre-coupled to 4-bromophenylalanine, yielding the C-terminal carboxylic acid after cleavage

Synthesis time: 4.5 h

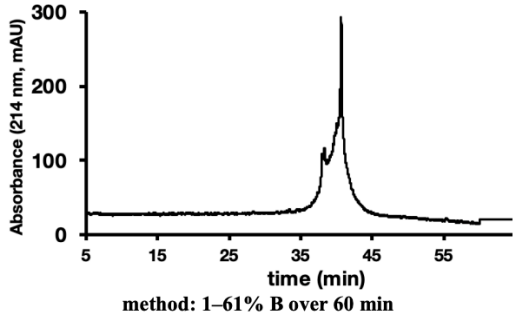
Synthesis UV trace from AFPS:



B) Cleavage and analytical Data for crude barnase($R110F_{Br}$)

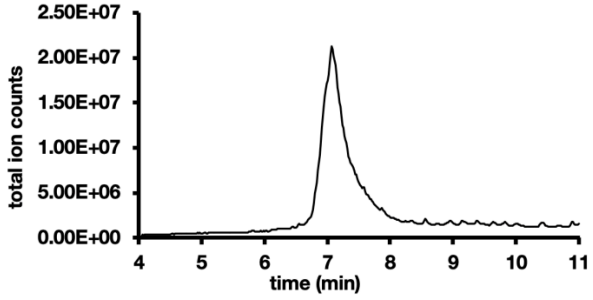
Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

Analytical HPLC:



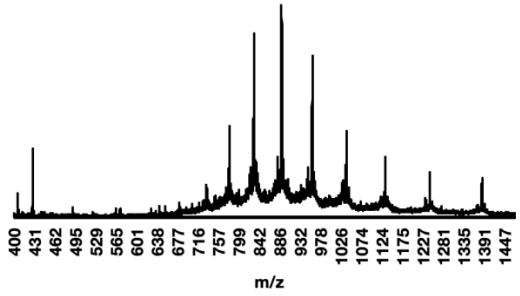
method: 1–61% B over 60 min

LC-MS:

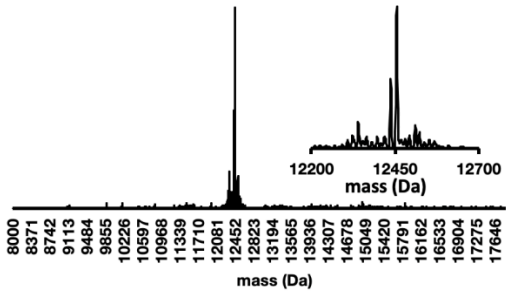


method: 1–91%B over 15 min, Zorbax C3 column (6520)

MS data from LC-MS:



Deconvolution of LC-MS data:

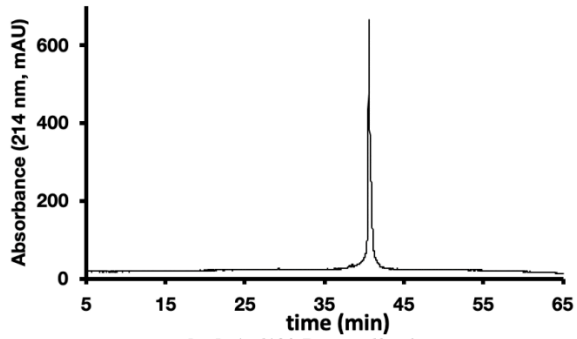


C) HPLC purification and analytical data for barnase(R110F$_{Br}$)

HPLC purification: Purified on a preparative Zorbax 300SB-C3 column at 25 °C at 20 mL/min. (gradient: 15–55% B with 0.33% B/min).

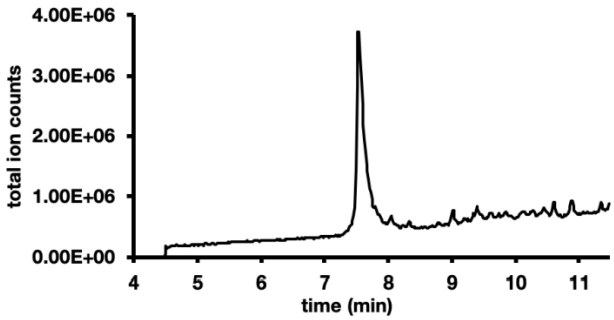
Yield past HPLC: 15 mg (3%)

Analytical HPLC:



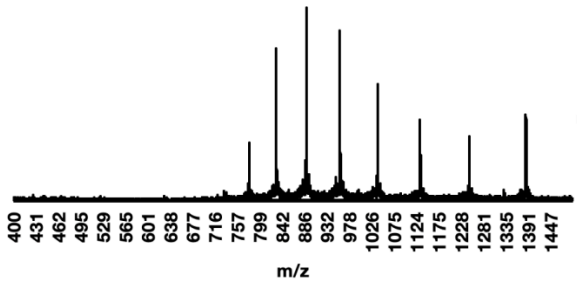
method: 1–61% B over 60 min

LC-MS:

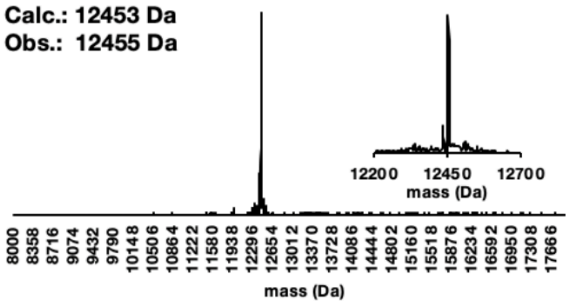


method: 1-91%B over 15 min, Zorbax C3 column (6520)

MS data from LC-MS:



Deconvolution of LC-MS data:



A1.11.7 MDM2[1-118]

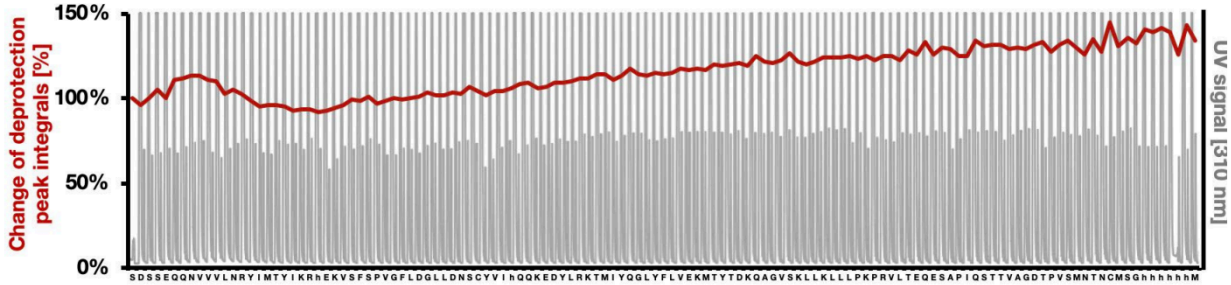
A) Synthesis Data

Sequence: MHHHHHHGSM CNTNMSVPTD GAVTTSQIPA SEQETLVRPK PLLLKLLKSV GAQKDTYTMK EVLFYLGQYI MTKRLYDEKQ QHIVYCSNDL LGDLFGVPSF SVKEHRKIYT MIYRNLVVVN QQESSDS (127 AA)

Resin: 150 mg of RINK amine ChemMatrix® LL (0.18 mmol/g), yielding the C-terminal amide after cleavage

Synthesis time: 5 h

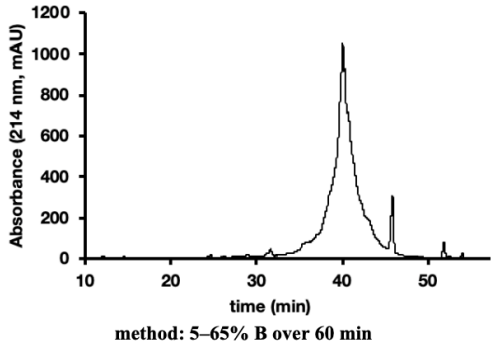
Synthesis UV trace from AFPS:



B) Cleavage and analytical Data for crude MDM2[1-118]

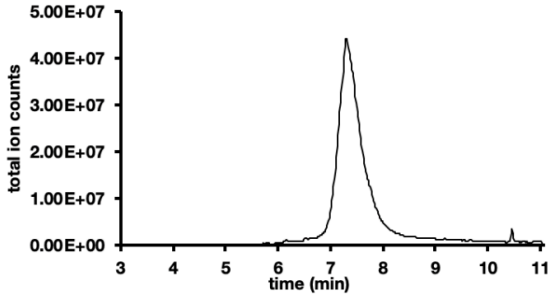
Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

Analytical HPLC:



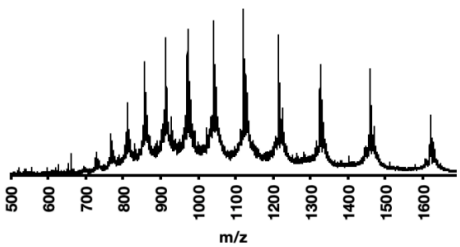
method: 5–65% B over 60 min

LC-MS:

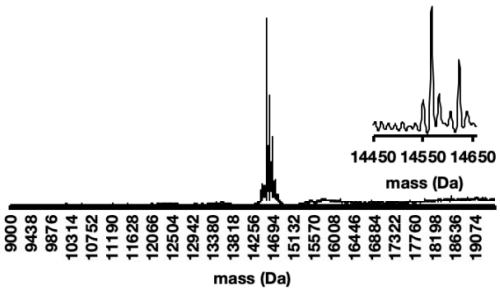


method: 1–91%B over 15 min, Zorbax C3 column (6520)

MS data from LC-MS:



Deconvolution of LC-MS data:

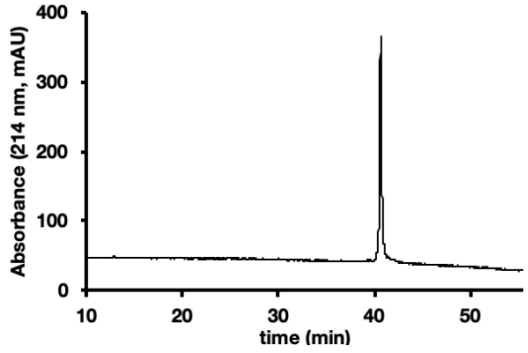


C) HPLC purification and analytical data for MDM2[1-118]

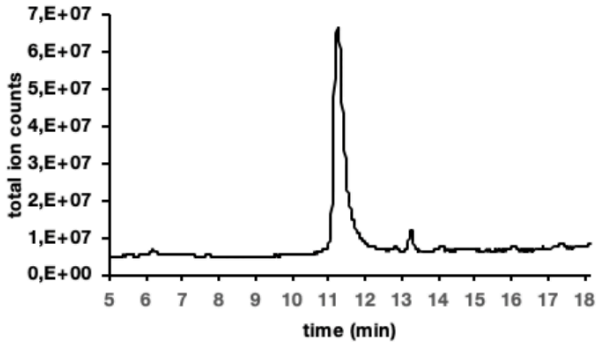
HPLC purification: Purified on a semipreparative Zorbax 300SB-C3 column at 60 °C (gradient: 5–29% B with 1% B/min, 29–49% B with 0.2% B/min, 49–65% B with 1% B/min)

Yield past HPLC: 14 mg (3%)

Analytical HPLC:

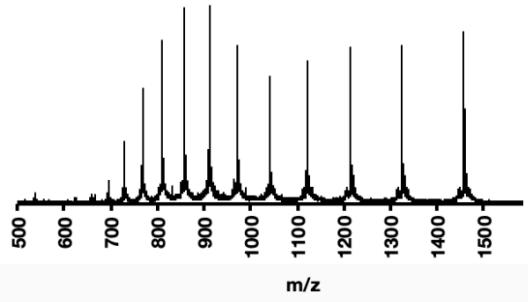


method: 5–65% B over 60 min

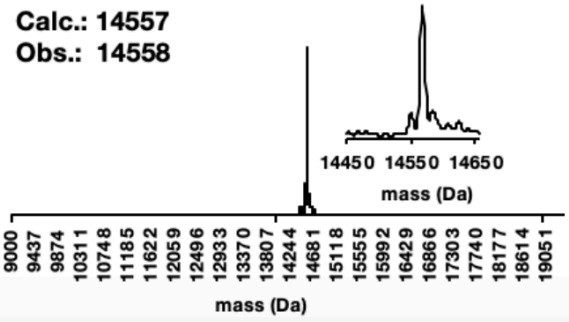


method: 1–91%B over 20 min, Zorbax C3 column (6550)

MS data from LC-MS:



Deconvolution of LC-MS data:



Comments: *) oxidation by-products detected

A1.11.8 Lysozyme

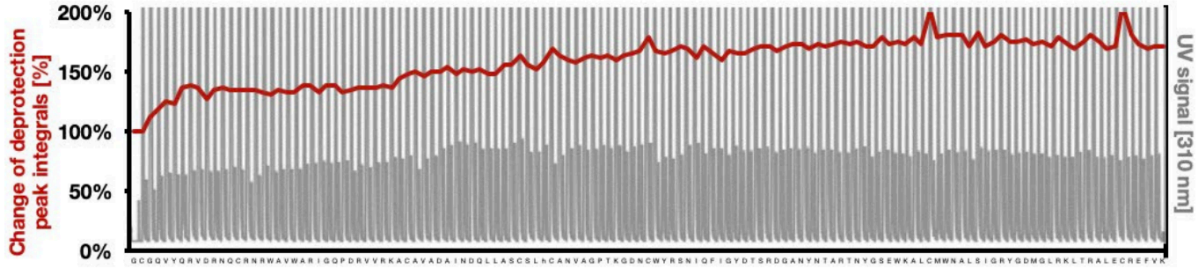
A) Synthesis Data for lysozyme

Sequence: KVFERCELAR TLKRLGMDGY RGISLANWMC LAKWESGYNT RATNYNAGDR STDYGIFQIN SRYWCNDGKT PGAVNACHLS CSALLQDNIA DAVACAKRVV RDPQGIRAWV AWRNRCQNRD VRQYVQGCGV (129 AA)

Resin: 80 mg of HMPB ChemMatrix® (0.44 mmol/g) pre-coupled to valine, yielding the C-terminal carboxylic acid after cleavage

Synthesis time: 5.5 h

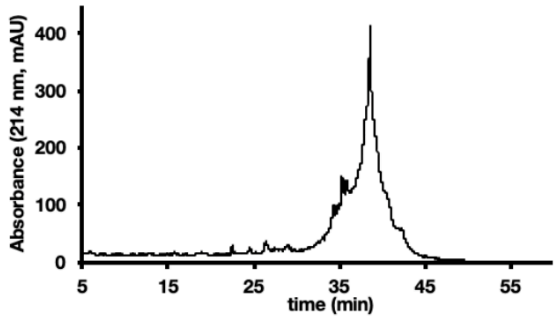
Synthesis UV trace from AFPS:



B) Cleavage and analytical Data for crude lysozyme

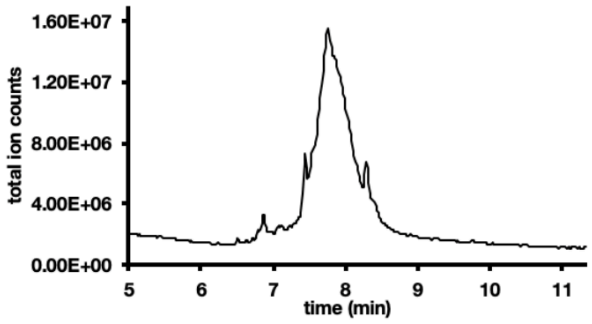
Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

Analytical HPLC:



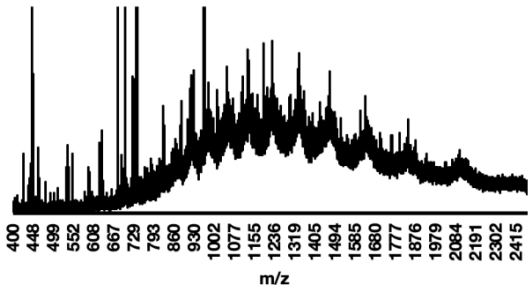
method: 5–65% B over 60 min

LC-MS:

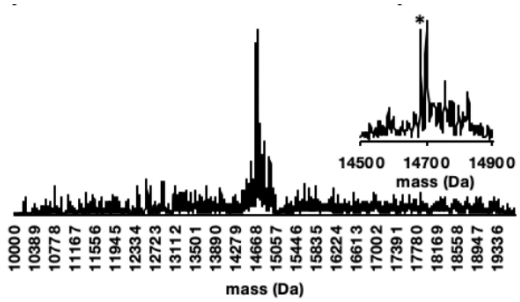


method: 1–91%B over 15 min, Zorbax C3 column (6520)

MS data from LC-MS:



Deconvolution of LC-MS data:

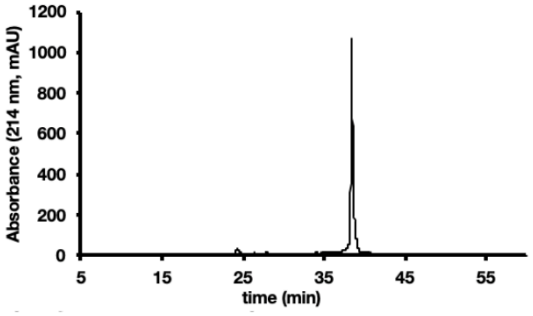


C) HPLC purification and analytical data for lysozyme

HPLC purification: Purified on Zorbax 300SB-C3 semipreparative column; Gradient: 5–28% B at 1% B/min; 29–49% B at 0.2% B/min; 50–65% B at 1% B/min.

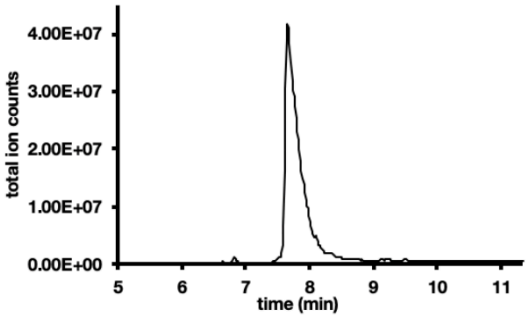
Yield past HPLC: 3.2 mg (1%)

Analytical HPLC:



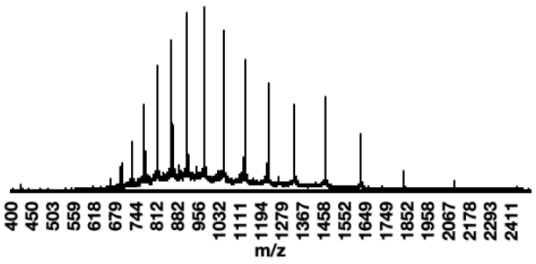
method: 5–65% B over 60 min

LC-MS:

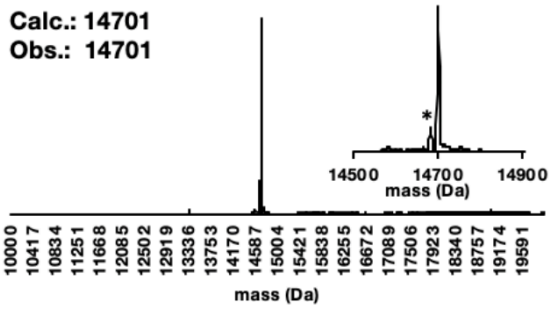


method: 1–91%B over 15 min, Zorbax C3 column (6520)

MS data from LC-MS:



Deconvolution of LC-MS data:



Comments: *) aspartimide by-product

A1.11.9 FGF1

A) Synthesis Data for FGF1

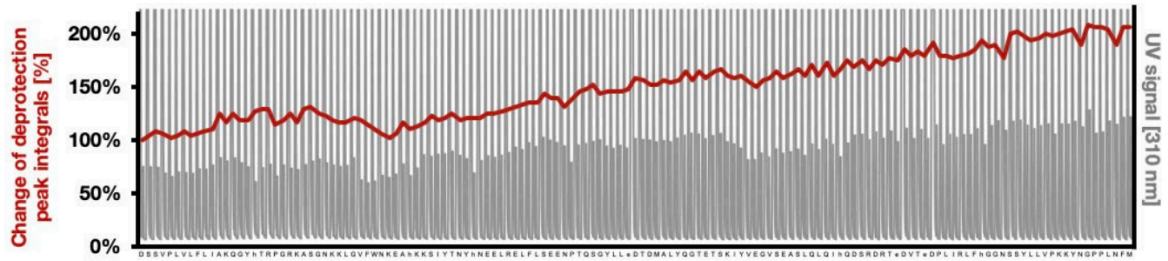
Sequence: MFNLPPGNYK KPVLLYSSNG GHFLRILPDG TVDGTRDRSD QHIQLQLSAE
SVGEVYIKST ETGQYLAMDT DGLLYGSQTP NEESLFLERL EENHYNTYIS
KKHAEKNWFV GLKKNGSAKR GPRTHYGQKA ILFLVLPVSS D (141 AA)

Note: the three glycines (G) highlighted in blue were introduced as Fmoc-(Dmb)Gly-OH building blocks

Resin: 100 mg of RINK amine ChemMatrix® LL (0.18 mmol/g), yielding the C-terminal amide after cleavage

Synthesis time: 5.7 h

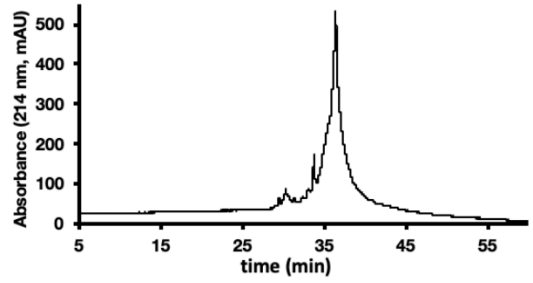
Synthesis UV trace from AFPS:



B) Cleavage and analytical Data for crude FGF1

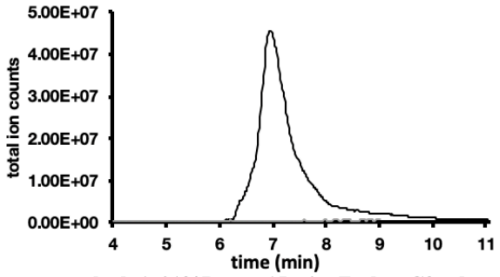
Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

Analytical HPLC:



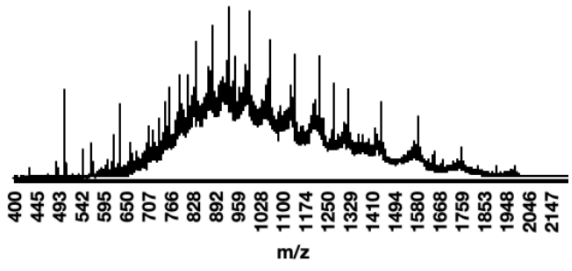
method: 5–65% B over 60 min

LC-MS:

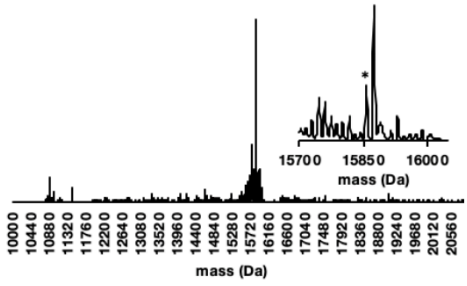


method: 1–91%B over 15 min, Zorbax C3 column (6520)

MS data from LC-MS:



Deconvolution of LC-MS data:

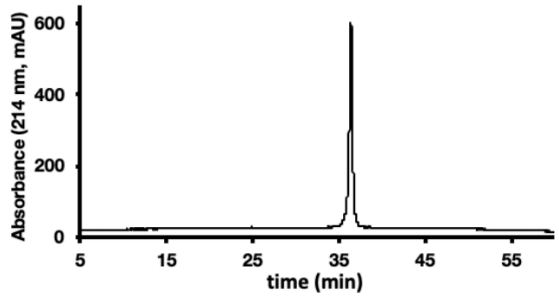


C) HPLC purification and analytical data for FGF1

HPLC purification: Purified 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)

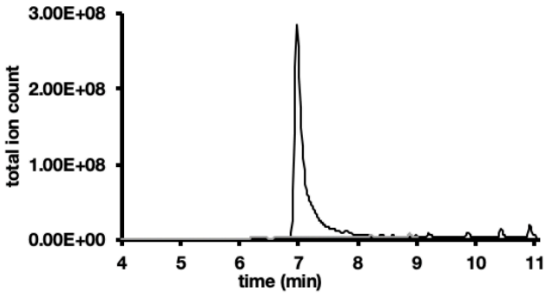
Yield past HPLC: 11 mg (4%)

Analytical HPLC:



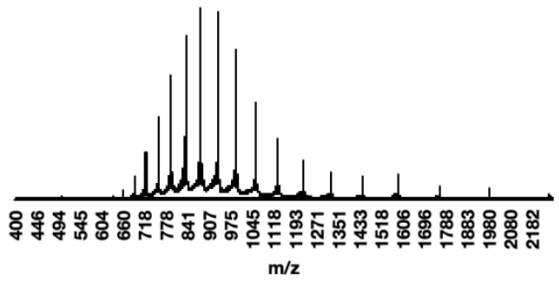
method: 5–65% B over 60 min

LC-MS:

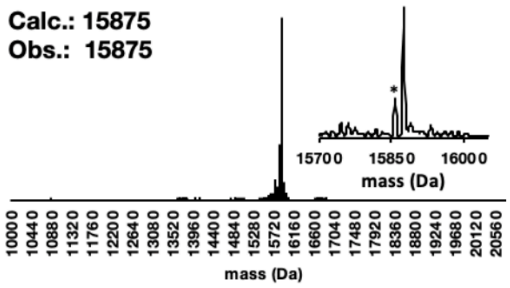


method: 1–91%B over 15 min, Zorbax C3 column (6520)

MS data from LC-MS:



Deconvolution of LC-MS data:



Comments: *)aspartimide by-product is present

A1.11.10 Sortase A_{59-206}; P94S/D160N/K196T

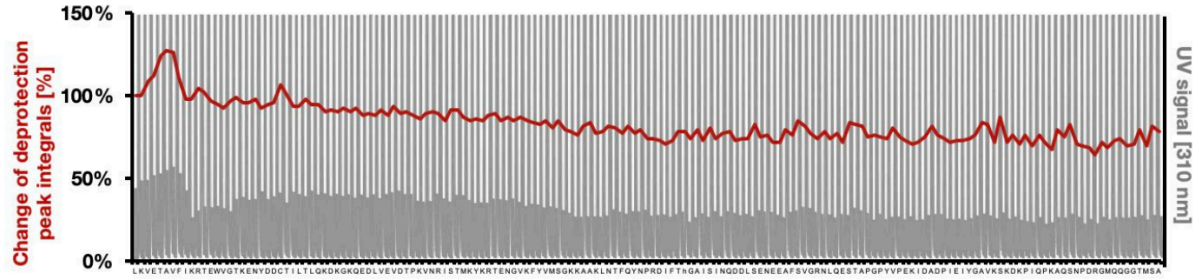
A) Synthesis Data

Sequence: ASMTGGQQMG RDPNSQAKPQ IPKDKSKVAG YIEIPDADIK EPVYPGPATS EQLNRGVSFA EENESLDDQN ISIAGHTFID RPNYQFTNLK AAKKGSMVYF KVGNETRKYK MTSIRNVKPT DVEVLDEQKG KDKQLTLITC DDYNEKTGVW ETRKIFVATE VKLE (164 AA)

Resin: 50 mg of HMPB ChemMatrix® (0.44 mmol/g) pre-coupled to glutamic acid, yielding the C-terminal carboxylic acid after cleavage

Synthesis time: 6.5 h

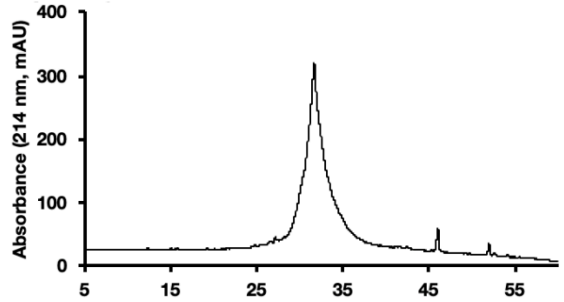
Synthesis UV trace from AFPS:



B) Cleavage and analytical Data for crude sortase A_{59-206};P94S/D160N/K196T

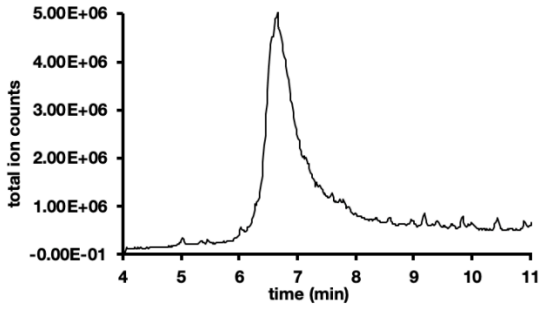
Cleavage: Cleavage protocol B for proteins and Cys-containing peptides

Analytical HPLC:



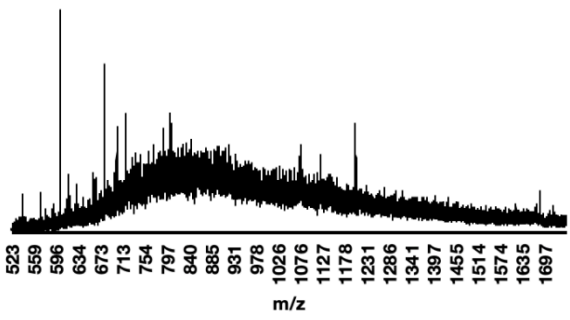
method: 5–65% B over 60 min

LC-MS:

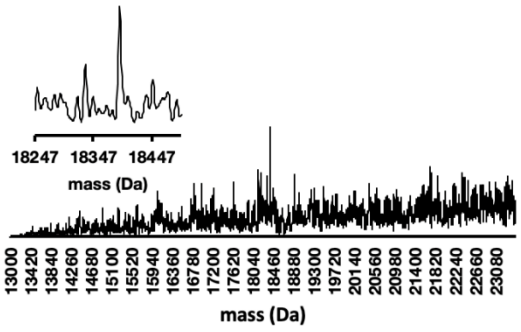


method: 1–91%B over 15 min, Zorbax C3 column (6520)

MS data from LC-MS:



Deconvolution of LC-MS data:

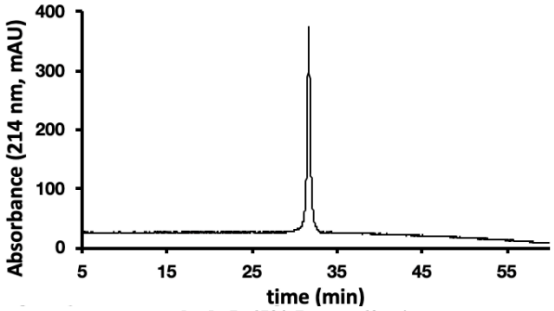


C) HPLC purification and analytical data for sortase A_{59-206};P94S/D160N/K196T

HPLC purification: Purified on a semipreparative Zorbax 300SB-C3 column at 60 °C (gradient: 5–24% B with 1% B/min, 24–44% B with 0.2% B/min, 44–65% B with 1% B/min)

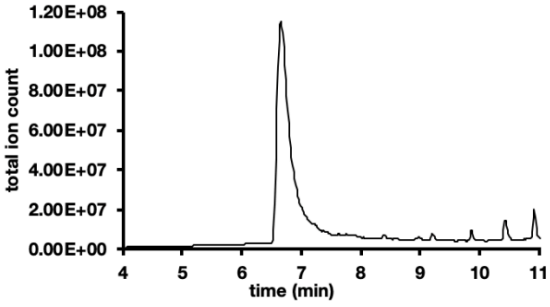
Yield past HPLC: 6.4 mg (1%)

Analytical HPLC:



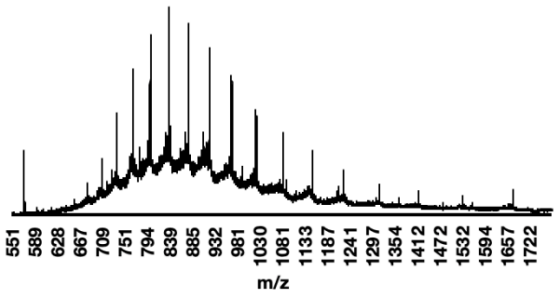
method: 5–65% B over 60 min

LC-MS:

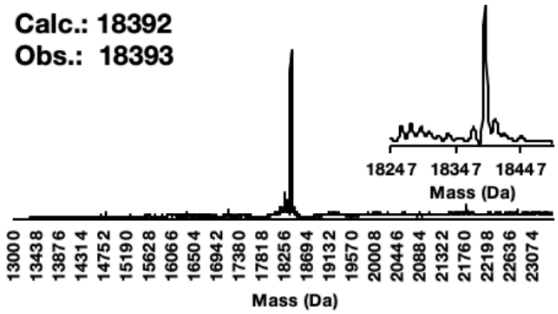


method: 1–91%B over 15 min, Zorbax C3 column (6520)

MS data from LC-MS:



Deconvolution of LC-MS data:



A1.12 Folding, Purification and Characterization of synthetic proteins

A1.12.1 Barnase and barstar

A1.12.1.1 Expression and SEC purification of recombinant barnase

Barnase was expressed based on previous report from Okorokov et al.(66) 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 (Fig. A1.15). Overall, 3 mg of recombinant barnase was isolated from 400 mL of the culture supernatant.

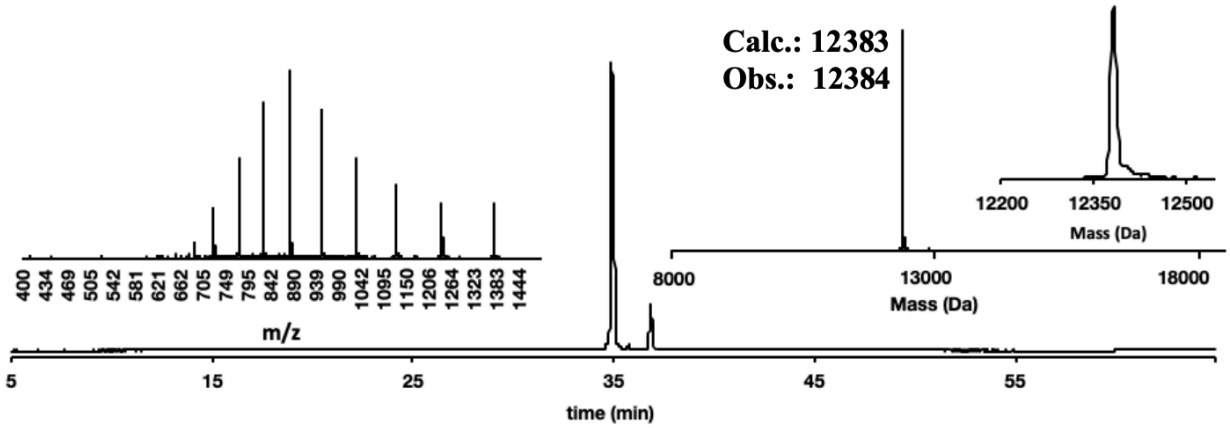


Figure A1.15. **Expression and purification yields recombinant barnase for comparison with synthetic enzyme**. Analytical HPLC traces of purified recombinant barnase (absorbance at 214 nm in mAU), LC-MS data (m/z in Da), and deconvoluted mass spectra (obtained from integration over all LC-MS signals). Calc.:12383 Da, Obs. 12384 Da.

A1.12.1.2 Folding and SEC purification of synthetic wild type 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 free[1] 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 in Fig. A1.16. 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.

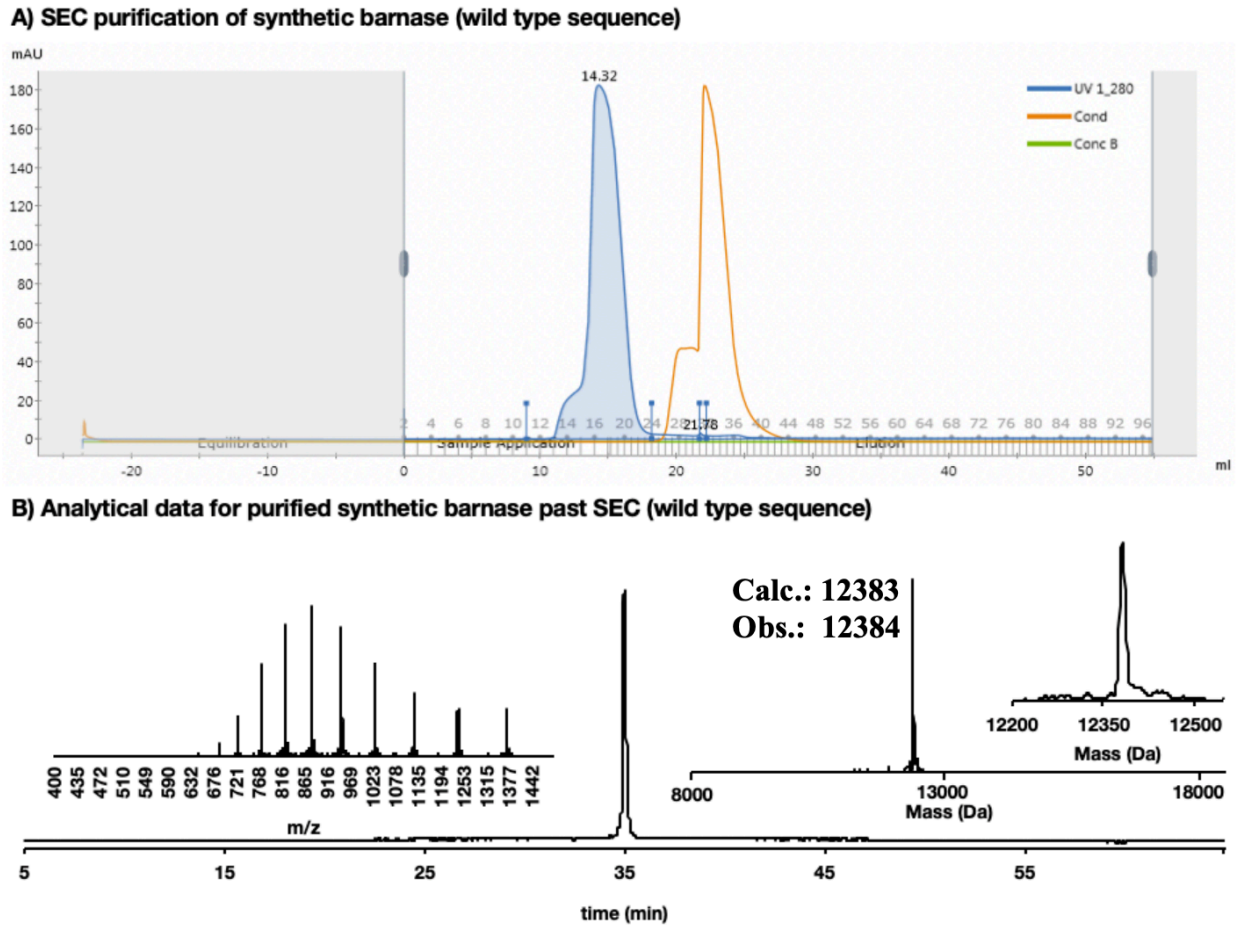


***Figure A1.16*. Folding and SEC purification of synthetic barnase yields folded protein indistinguishable from recombinant barnase.** A) HPLC-purified barnase was subjected to folding conditions followed by size exclusion chromatography. The elution profile from the size exclusion chromatography is displayed in this figure. B) analytical HPLC traces of purified synthetic barnase (absorbance at 214 nm in mAU), LC-MS data (m/z in Da), and deconvoluted mass spectra (obtained from integration over all LC-MS signals). Calc.:12383 Da, Obs.: 12384 Da.

A1.12.1.3 Comparison of timelines for production of synthetic wild type barnase versus expression of wild type barnase

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 77–82h.

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.

A1.12.1.4 Folding and SEC purification of synthetic barnase R110F$_{Br}$

The folding procedure of barnase R110FBr protein was adapted from Mong et al.(18) 5.9 mg of barnase R110FBr variant was dissolved in 150 µL of 6 M Gn·HCl solution in 50 mM NaH2PO4 RNase free4 buffer, pH = 7.8. The mixture was then 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 in Fig. A1.17. 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 1.3 mg of barnase R110FBr was isolated, corresponding to 22% yield. The purity of the final product was assessed by LC-MS and HPLC.

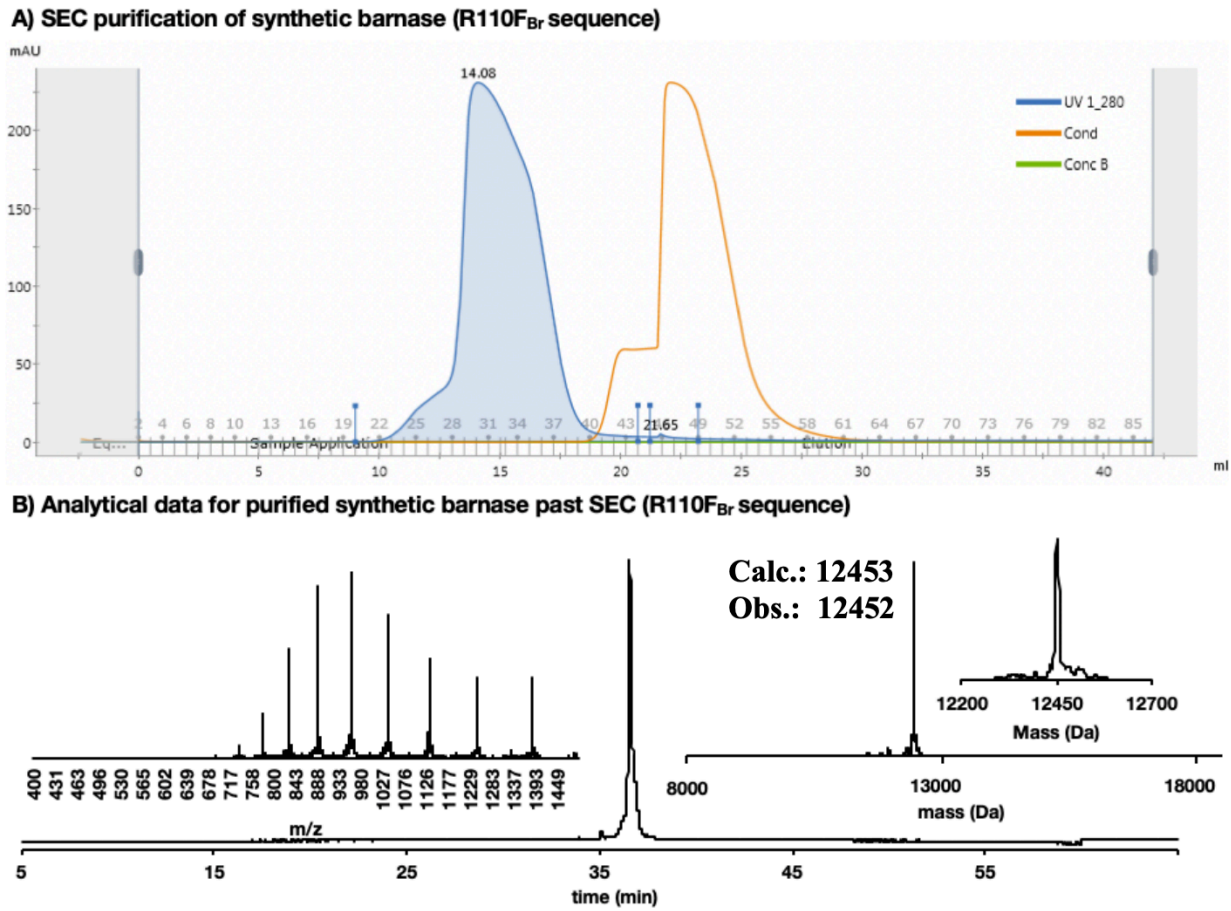


***Figure A1.17*. Folding and SEC purification of synthetic barnase R110F$_{Br}$ yields clean protein sample.** A) HPLCpurified barnase was subjected to folding conditions followed by size exclusion chromatography. The elution profile from the size exclusion chromatography is displayed in this figure. B) analytical HPLC traces of purified synthetic barnase R110FBr (absorbance at 214 nm in mAU), LC-MS data (m/z in Da), and deconvoluted mass spectra (obtained from integration over all LC-MS signals). Calc.:12453 Da, Obs.: 12452 Da.

A1.12.1.5 Folding and IEC purification of synthetic wild type Barstar

The folding procedure of wild type Barstar was adapted from Shastry et al.(67) 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 in Fig. A1.18.

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.

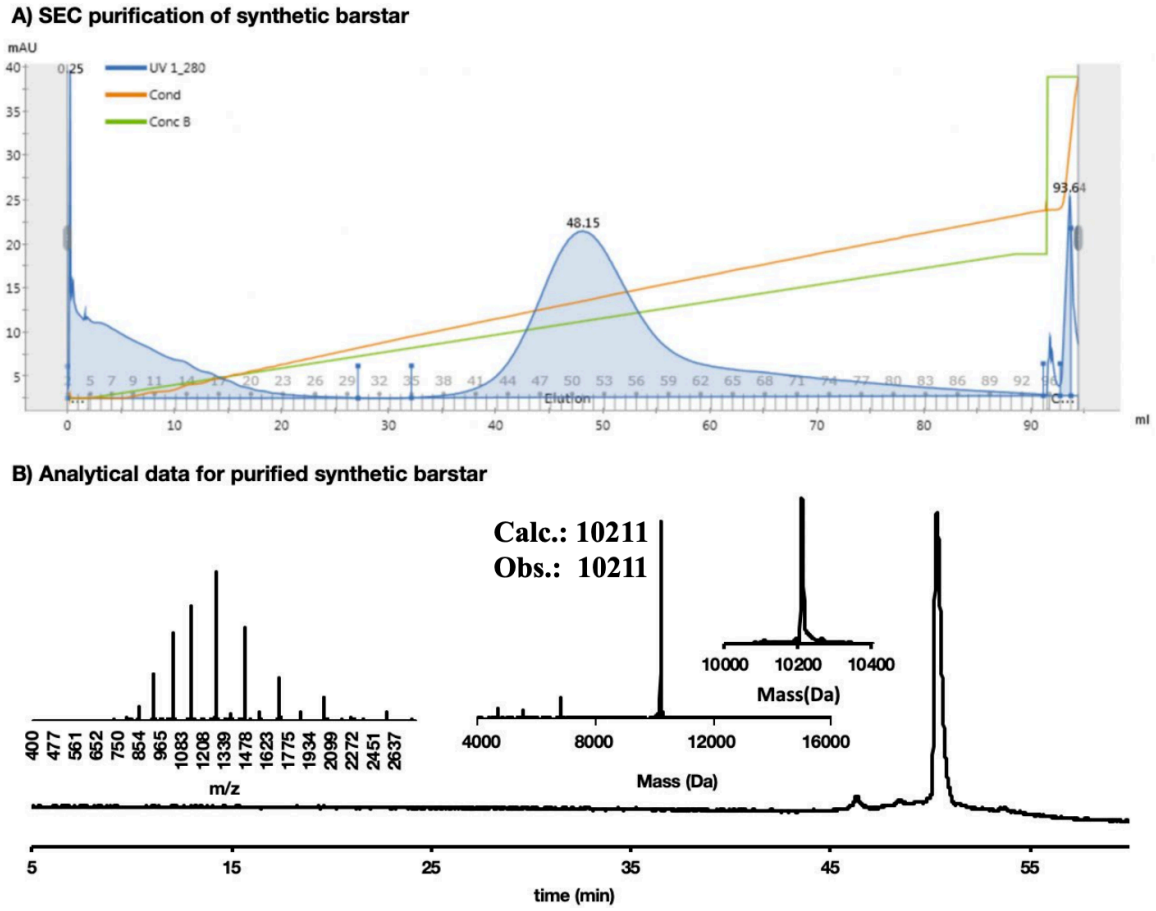


Figure A1.18. **Folding and IEC purification of synthetic barstar.** A) HPLC-purified barstar was subjected to folding conditions followed by anion exchange chromatography. The elution profile from the anion exchange chromatography is

displayed in this figure. B) analytical HPLC traces of purified synthetic barstar (absorbance at 214 nm in mAU), LC-MS data (m/z in Da), and deconvoluted mass spectra (obtained from integration over all LC-MS signals). Calc.:10211 Da, Obs.: 10211 Da.

A1.12.1.6 RNA hydrolysis assay

Hydrolysis of RNA by the synthesized 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. A1.19), or synthetic barstar (Fig. A1.20), 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. CAACAUCUUGCUAUACAAUGCCAAUCCAUGCUACACUACGUUACA

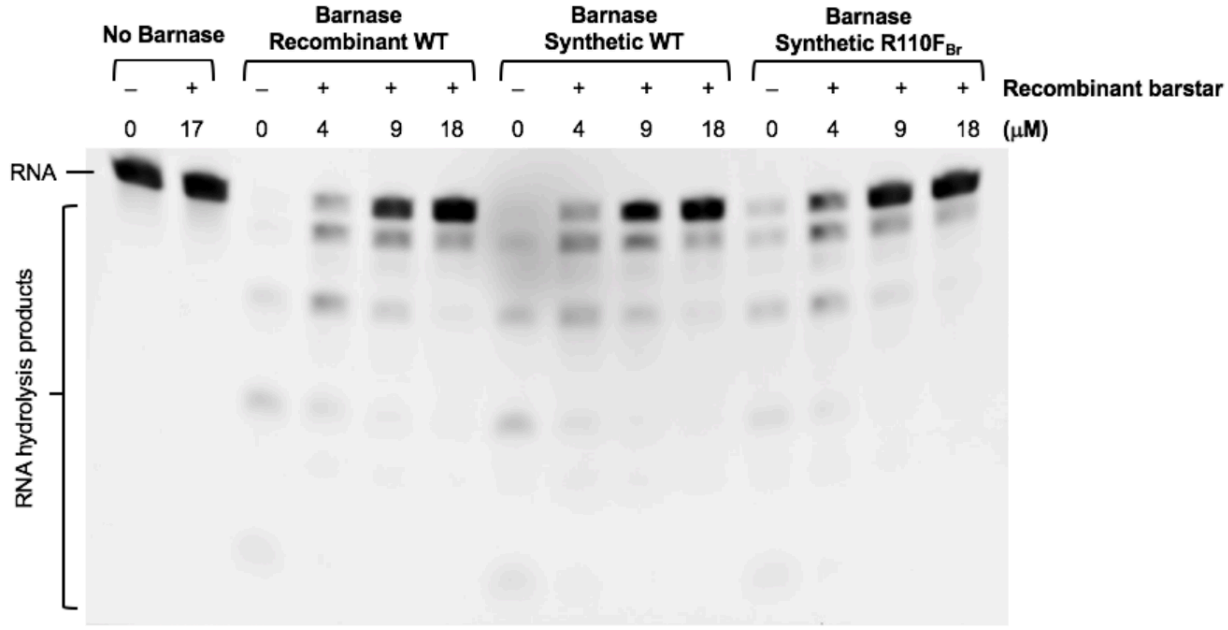


***Figure A1.19*. RNA hydrolysis assay shows comparable RNase activity for recombinant whild type barnase, synthetic wild type barnase, and synthetic R110F$_{Br}$ barnase**. The synthetic barnase proteins are able to digest native RNA within 20 minutes. The dose-dependent inhibition of barnase activity upon pre-incubation with recombinant barstar protein points to association of these proteins.

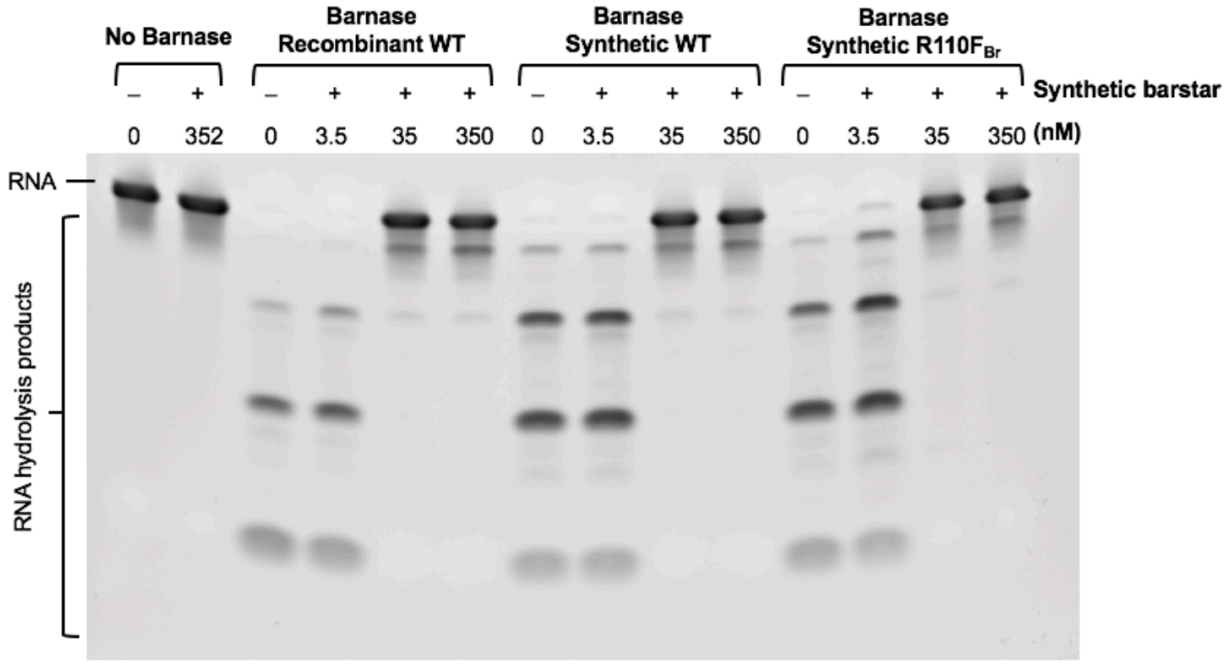


Figure A1.20. **Synthetic barstar inhibits the RNase activity of recombinant and synthetic barnase variants.** Preincubation of barnase variants with synthetic barstar protein leads to dose-dependent inhibition of RNase activity of barnases.

A1.12.1.7 Fluorogenic RNase activity assay

Hydrolysis of a fluorogenic substrate (6-FAM-dADrGDdADdAD-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, I_0 is fluorescence of the intact substrate, If is the fluorescence of the hydrolyzed substrate, [E] is the total enzyme concentration, k_{cat}, K_M are steady state enzyme kinetic parameters.[42,47] 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 (Fig. A1.21, Table A1.12). The fluorescence was monitored every 10 seconds, after 3 second shake, for at least 500 seconds to measure the starting fluorescence, I_0, 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 k_{cat} [E]/K_M 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}.[68]

Accordingly, k_{cat}/K_M (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 $(1.3 \pm 0.4) \times 10^{7}$.[66] Furthermore, the $k_{cat}/K_M \pm$ SE of the $R110F_{Br}$ synthetic variant barnase was calculated to be $(2.3 \pm 0.8) \times 10^{6}$ M^{-1} s^{-1} .

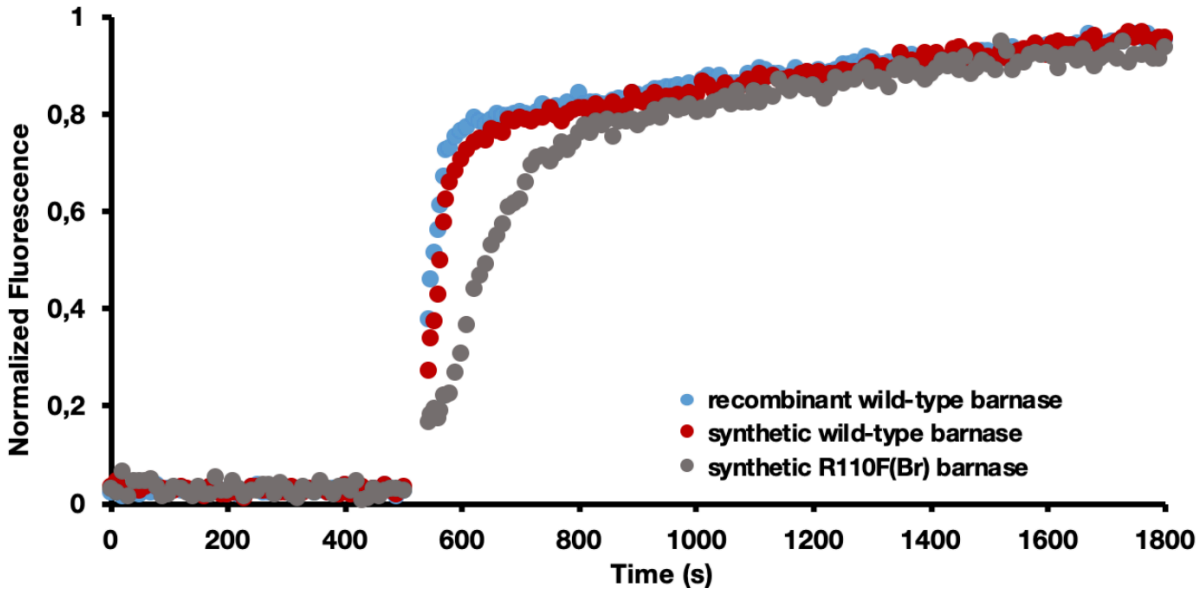


Figure A1.21. **Fluorogenic RNase activity determines the k_{cat}/K_M value of the barnase variants**. A sample of the data produced from cleavage of substrate 6-FAM-dADrGDdADdAD-6-TAMRA with synthetic and recombinant barnase. At 500 s, barnase was added to the substrate. The increase in fluorescence indicates cleavage of substrate by the barnase protein.

	recombinant wild-type barnase			synthetic wild-type barnase			synthetic R110F(Br) barnase		
	Normalized Fluoresence			Normalized Fluoresence			Normalized Fluoresence		
time	mean	SD	N	mean	SD	N	mean	SD	N
0	0.020037	0.017308	3	0.034435	0.007823	3	0.028378	0.023046	3
10	0.02132	0.015889	3	0.046331	0.016813	3	0.022022	0.007184	3
20	0.015086	0.015035	3	0.030857	0.019189	3	0.06372	0.012992	3
30	0.021942	0.017726	3	0.027624	0.02024	3	0.012052	0.008125	3
40	0.026442	0.018517	3	0.032546	0.004795	3	0.04555	0.002984	3
50	0.016073	0.00999	3	0.025117	0.008412	3	0.044298	0.018378	3
60	0.036291	0.016581	3	0.031619	0.006571	3	0.045795	0.004736	3
70	0.023665	0.007032	3	0.035765	0.003841	3	0.049083	0.009578	3
80	0.036835	0.003653	3	0.03402	0.005061	3	0.034325	0.001505	3
90	0.020951	0.004466	3	0.027982	0.006634	3	0.01559	0.013845	3
100	0.030705	0.009477	3	0.027342	0.01022	3	0.017765	0.005798	3
110	0.030156	0.009942	3	0.024142	0.012263	3	0.031525	0.002104	3
120	0.024709	0.008243	3	0.032683	0.003773	3	0.028929	0.012342	3
130	0.024256	0.005014	3	0.02834	0.015835	3	0.023533	0.005085	3
140	0.017455	0.017315	3	0.018469	0.004924	3	0.015224	0.014591	3
150	0.029419	0.004726	3	0.026854	0.008953	3	0.033465	0.004178	3
160	0.017336	0.006205	3	0.0151	0.0181	3	0.018083	0.003085	3
170	0.029225	0.008823	3	0.032867	0.0199	3	0.018952	0.010341	3
180	0.033916	0.011287	3	0.03369	0.0175	3	0.052823	0.011273	3

190	0.02599	0.007404	3	0.017313	0.017574	3	0.026656	0.011297	3
200	0.016351	0.002159	3	0.016372	0.015168	3	0.025757	0.012297	3
210	0.020221	0.015403	3	0.020087	0.006475	3	0.044108	0.009565	3
220	0.018462	0.016127	3	0.023417	0.020989	3	0.028391	0.012195	3
230	0.018707	0.00495	3	0.010505	0.003587	3	0.013807	0.005961	3
240	0.031657	0.01581	3	0.033561	0.006329	3	0.027224	0.017443	3
250	0.035476	0.01616	3	0.032288	0.016513	3	0.025987	0.012745	3
260	0.022506	0.009581	3	0.025767	0.011322	3	0.038104	0.007806	3
270	0.020404	0.013912	3	0.026544	0.020409	3	0.038312	0.010112	3
280	0.028206	0.011171	3	0.033368	0.010215	3	0.016367	0.014619	3
290	0.029076	0.013204	3	0.022812	0.004462	3	0.040011	0.012231	3
300	0.025502	0.021369	3	0.021907	0.018448	3	0.029034	0.006337	3
310	0.018464	0.016006	3	0.032555	0.01751	3	0.018069	0.004663	3
320	0.030029	0.026137	3	0.026226	0.008982	3	0.008987	0.008256	3
330	0.025448	0.004314	3	0.023299	0.015746	3	0.033125	0.006954	3
340	0.028496	0.007847	3	0.028521	0.011935	3	0.026245	0.012744	3
350	0.021273	0.00765	3	0.026386	0.020324	3	0.019992	0.006786	3
360	0.0253	0.010263	3	0.035107	0.021094	3	0.045855	0.006217	3
370	0.022551	0.011895	3	0.020279	0.021479	3	0.031424	0.006564	3
380	0.029261	0.004228	3	0.022874	0.023375	3	0.018642	0.004327	3
390	0.028017	0.002704	3	0.018335	0.0139	3	0.031547	0.011161	3
400	0.036284	0.009215	3	0.037668	0.014133	3	0.034496	0.005877	3
410	0.023297	0.012709	3	0.017666	0.017925	3	0.025033	0.017906	3
420	0.03321	0.014274	3	0.030426	0.003552	3	0.045924	0.013302	3
430	0.023832	0.009899	3	0.034993	0.022073	3	0.006082	0.005307	3
440	0.020559	0.005707	3	0.018712	0.007472	3	0.00967	0.009478	3
450	0.029683	0.009379	3	0.027875	0.013458	3	0.029794	0.002359	3
460	0.021359	0.024315	3	0.024734	0.022602	3	0.014599	0.009345	3
470	0.03843	0.022088	3	0.035663	0.018396	3	0.024636	0.020606	3
480	0.031492	0.015798	3	0.02752	0.015067	3	0.028337	0.003348	3
490	0.014854	0.00971	3	0.019215	0.015841	3	0.029691	0.005326	3
500	0.02735	0.013916	3	0.031689	0.025211	3	0.025918	0.005519	3
543	0.376124	0.036526	3	0.271574	0.032096	3	0.167876	0.059504	3
548,1	0.459421	0.047275	3	0.336779	0.039623	3	0.181806	0.053791	3
553,3	0.515691	0.046399	3	0.374202	0.04335	3	0.194694	0.068961	3
558,5	0.561773	0.04446	3	0.429502	0.037342	3	0.176133	0.064137	3
563,6	0.612995	0.036094	3	0.500059	0.060141	3	0.188731	0.083854	3
568,8	0.670432	0.053508	3	0.577149	0.056928	3	0.221164	0.088519	3
574	0.723481	0.068816	3	0.621854	0.052468	3	0.222829	0.071291	3
580	0.730962	0.074672	3	0.659043	0.058191	3	0.225588	0.061999	3
590	0.751674	0.056278	3	0.682233	0.044692	3	0.268627	0.047726	3
600	0.765715	0.063007	3	0.70632	0.066043	3	0.30802	0.032704	3
610	0.772352	0.054325	3	0.723838	0.063124	3	0.365948	0.020063	3

620	0.792288	0.042424	3	0.739504	0.061735	3	0.439824	0.014831	3
630	0.78369	0.058984	3	0.7489	0.066118	3	0.467312	0.047987	3
640	0.774608	0.043515	3	0.745366	0.064396	3	0.491913	0.039488	3
650	0.787995	0.070144	3	0.76974	0.060571	3	0.531004	0.054388	3
660	0.798371	0.060211	3	0.763413	0.051354	3	0.550379	0.038997	3
670	0.797081	0.072206	3	0.759797	0.047309	3	0.5723	0.035127	3
680	0.796562	0.072284	3	0.786097	0.052311	3	0.606657	0.057093	3
690	0.79919	0.067915	3	0.785046	0.065701	3	0.614952	0.05746	3
700	0.802912	0.059498	3	0.791514	0.073707	3	0.622057	0.075278	3
710	0.801541	0.055021	3	0.789526	0.054387	3	0.657701	0.059946	3
720	0.79664	0.070561	3	0.785719	0.067301	3	0.693549	0.061446	3
730	0.802072	0.052877	3	0.790921	0.046656	3	0.709792	0.044731	3
740	0.81815	0.048111	3	0.791962	0.064269	3	0.71556	0.043365	3
750	0.807192	0.056963	3	0.810828	0.060985	3	0.700964	0.043515	3
760	0.817287	0.052582	3	0.797686	0.059471	3	0.716954	0.031734	3
770	0.811981	0.062775	3	0.785714	0.061371	3	0.740284	0.027179	3
780	0.82347	0.048289	3	0.799313	0.077963	3	0.724472	0.026804	3
790	0.807626	0.042271	3	0.809195	0.053371	3	0.741614	0.020228	3
800	0.840929	0.051624	3	0.810611	0.078902	3	0.759457	0.028248	3
810	0.817837	0.054528	3	0.809661	0.061736	3	0.776538	0.007411	3
820	0.823256	0.055015	3	0.79439	0.061698	3	0.761966	0.010145	3
830	0.822055	0.071487	3	0.820107	0.06036	3	0.782278	0.018724	3
840	0.820815	0.052934	3	0.812052	0.062859	3	0.776735	0.002586	3
850	0.818606	0.052253	3	0.808896	0.058742	3	0.786466	0.013331	3
860	0.825116	0.061536	3	0.823394	0.067653	3	0.752564	0.006443	3
870	0.832907	0.063302	3	0.811726	0.059798	3	0.782646	0.012596	3
880	0.829368	0.062037	3	0.814136	0.053184	3	0.78504	0.02902	3
890	0.844227	0.053276	3	0.844583	0.045882	3	0.78659	0.009901	3
900	0.838947	0.055164	3	0.828261	0.037764	3	0.775951	0.029465	3
910	0.833839	0.040875	3	0.824427	0.044241	3	0.786218	0.025418	3
920	0.841965	0.043857	3	0.831476	0.05839	3	0.792628	0.017983	3
930	0.848654	0.056661	3	0.841001	0.045014	3	0.806894	0.012068	3
940	0.851123	0.064403	3	0.829143	0.057962	3	0.790717	0.012342	3
950	0.856127	0.059306	3	0.835491	0.046216	3	0.814523	0.022747	3
960	0.855458	0.055894	3	0.836658	0.044652	3	0.813464	0.002953	3
970	0.858759	0.043555	3	0.838659	0.048534	3	0.806437	0.007446	3
980	0.846229	0.039644	3	0.836323	0.052457	3	0.81123	0.010721	3
990	0.861021	0.051743	3	0.844542	0.047788	3	0.818862	0.017697	3
1000	0.860124	0.037773	3	0.838449	0.032969	3	0.802104	0.010954	3
1010	0.871405	0.047664	3	0.867588	0.030417	3	0.812119	0.010048	3
1020	0.879705	0.045614	3	0.857412	0.040826	3	0.807209	0.008592	3
1030	0.863406	0.032734	3	0.845084	0.027521	3	0.825026	0.015542	3
1040	0.876683	0.045794	3	0.843043	0.034408	3	0.825936	0.014285	3

1050	0.862598	0.048212	3	0.86368	0.025266	3	0.835548	0.010757	3
1060	0.860973	0.038292	3	0.849547	0.043789	3	0.819451	0.011808	3
1070	0.860911	0.051288	3	0.848751	0.041862	3	0.845981	0.009484	3
1080	0.863709	0.054628	3	0.859445	0.0489	3	0.821491	0.010403	3
1090	0.881335	0.042886	3	0.8713	0.033364	3	0.824353	0.007542	3
1100	0.859804	0.043836	3	0.855567	0.03273	3	0.845307	0.025729	3
1110	0.891695	0.037454	3	0.880997	0.052833	3	0.821864	0.024502	3
1120	0.890083	0.024555	3	0.872138	0.023785	3	0.842958	0.016154	3
1130	0.88306	0.056312	3	0.879636	0.029181	3	0.835432	0.005226	3
1140	0.882613	0.057051	3	0.870926	0.040086	3	0.86883	0.013364	3
1150	0.87425	0.05003	3	0.86957	0.038553	3	0.854641	0.007618	3
1160	0.87671	0.049984	3	0.87283	0.031383	3	0.847458	0.033206	3
1170	0.892104	0.052538	3	0.868942	0.038176	3	0.861389	0.022785	3
1180	0.87411	0.060786	3	0.880557	0.041124	3	0.858163	0.021748	3
1190	0.887225	0.051185	3	0.883772	0.00694	3	0.850603	0.010208	3
1200	0.879748	0.026432	3	0.88162	0.02692	3	0.859979	0.00223	3
1210	0.887951	0.036666	3	0.884597	0.02469	3	0.845525	0.009499	3
1220	0.881522	0.028912	3	0.874719	0.020365	3	0.830608	0.014017	3
1230	0.885875	0.045187	3	0.881455	0.037266	3	0.849051	0.016682	3
1240	0.888335	0.043274	3	0.883107	0.028542	3	0.874278	0.013654	3
1250	0.899468	0.018936	3	0.889966	0.014284	3	0.860377	0.00852	3
1260	0.902428	0.055806	3	0.884093	0.019233	3	0.868625	0.010584	3
1270	0.903998	0.037564	3	0.884141	0.010158	3	0.869546	0.009208	3
1280	0.898531	0.045361	3	0.883313	0.039547	3	0.885399	0.013973	3
1290	0.91715	0.035162	3	0.897579	0.037825	3	0.892564	0.014781	3
1300	0.912233	0.030149	3	0.905962	0.02742	3	0.865452	0.013866	3
1310	0.900096	0.042752	3	0.88913	0.042974	3	0.872594	0.009218	3
1320	0.901899	0.040921	3	0.897186	0.037983	3	0.86845	0.013984	3
1330	0.904747	0.030157	3	0.889655	0.026229	3	0.855536	0.009329	3
1340	0.905374	0.046754	3	0.892457	0.047966	3	0.889587	0.018802	3
1350	0.905217	0.042014	3	0.924449	0.051979	3	0.900231	0.010541	3
1360	0.903975	0.03494	3	0.898476	0.033105	3	0.876535	0.008822	3
1370	0.921486	0.039052	3	0.909217	0.030589	3	0.897478	0.012951	3
1380	0.917068	0.033473	3	0.908642	0.030608	3	0.881089	0.010173	3
1390	0.922752	0.029892	3	0.923588	0.011692	3	0.869858	0.002327	3
1400	0.917129	0.042799	3	0.905918	0.051671	3	0.882578	0.015365	3
1410	0.91162	0.031577	3	0.925903	0.018351	3	0.894472	0.018157	3
1420	0.913361	0.044968	3	0.913765	0.040699	3	0.908207	0.030894	3
1430	0.903894	0.04421	3	0.913699	0.036355	3	0.899196	0.01454	3
1440	0.92812	0.037626	3	0.934564	0.032934	3	0.909512	0.016236	3
1450	0.9104	0.035023	3	0.937841	0.046852	3	0.887865	0.011087	3
1460	0.91388	0.03966	3	0.922646	0.045705	3	0.916184	0.013709	3
1470	0.917405	0.038273	3	0.908025	0.046071	3	0.883533	0.004676	3

1480	0.926008	0.032702	3	0.926801	0.050733	3	0.895135	0.022068	3
1490	0.929518	0.052096	3	0.906598	0.049841	3	0.907352	0.015465	3
1500	0.927991	0.037908	3	0.91788	0.02429	3	0.900141	0.015125	3
1510	0.91493	0.03213	3	0.913883	0.041736	3	0.890612	0.009286	3
1520	0.927749	0.057806	3	0.929176	0.037133	3	0.949479	0.0082	3
1530	0.913	0.047798	3	0.913628	0.036506	3	0.928299	0.004685	3
1540	0.935032	0.034354	3	0.926339	0.038336	3	0.891125	0.013185	3
1550	0.928485	0.045773	3	0.927216	0.032867	3	0.903895	0.006207	3
1560	0.919521	0.049643	3	0.930735	0.039302	3	0.904905	0.017096	3
1570	0.931742	0.034419	3	0.931093	0.033827	3	0.92241	0.003846	3
1580	0.934999	0.039336	3	0.943812	0.043961	3	0.924075	0.016872	3
1590	0.935931	0.046387	3	0.930803	0.03432	3	0.920086	0.008909	3
1600	0.9401	0.030024	3	0.922722	0.029411	3	0.923504	0.024047	3
1610	0.93869	0.047519	3	0.943155	0.038442	3	0.91726	0.036838	3
1620	0.945422	0.034078	3	0.950145	0.02908	3	0.894994	0.028412	3
1630	0.931694	0.01677	3	0.940597	0.053523	3	0.921357	0.02071	3
1640	0.940147	0.025434	3	0.941288	0.039321	3	0.922655	0.015768	3
1650	0.946097	0.040666	3	0.937356	0.025513	3	0.907572	0.010849	3
1660	0.935475	0.031454	3	0.938526	0.042052	3	0.933259	0.02133	3
1670	0.963623	0.025302	3	0.950404	0.036139	3	0.913833	0.013341	3
1680	0.947625	0.042977	3	0.960243	0.033238	3	0.899055	0.022951	3
1690	0.921471	0.025743	3	0.936323	0.025505	3	0.913698	0.015741	3
1700	0.94573	0.03974	3	0.938312	0.028674	3	0.921568	0.005354	3
1710	0.948126	0.033203	3	0.941317	0.036537	3	0.926841	0.014384	3
1720	0.949213	0.029979	3	0.949526	0.044255	3	0.912354	0.011957	3
1730	0.948841	0.039448	3	0.955409	0.05122	3	0.94784	0.025639	3
1740	0.961175	0.019217	3	0.968252	0.047841	3	0.906015	0.019192	3
1750	0.963186	0.045429	3	0.95592	0.048029	3	0.922572	0.011168	3
1760	0.956187	0.026524	3	0.965879	0.013982	3	0.920943	0.020585	3
1770	0.963125	0.024549	3	0.952226	0.053125	3	0.924948	0.010928	3
1780	0.944291	0.035203	3	0.940995	0.040823	3	0.911972	0.006214	3
1790	0.95027	0.025093	3	0.954923	0.056191	3	0.91409	0.014675	3
1800	0.952615	0.036223	3	0.957022	0.030378	3	0.935238	0.009781	3
1810	0.95206	0.026894	3	0.959288	0.044304	3	0.946181	0.010277	3
1820	0.955887	0.027048	3	0.955439	0.020433	3	0.928095	0.018649	3
1830	0.961337	0.030839	3	0.950123	0.0436	3	0.932606	0.019155	3
1840	0.956388	0.035006	3	0.943586	0.0386	3	0.900708	0.017476	3
1850	0.948485	0.040128	3	0.958944	0.039501	3	0.9268	0.013184	3
1860	0.944026	0.032726	3	0.953577	0.024781	3	0.934471	0.007455	3
1870	0.969306	0.048112	3	0.975616	0.042234	3	0.94489	0.024365	3

Table A1.12. **Averaged data from fluorogenic RNase activity assay was used for the determination of k_{cat}/K_M values.**

A1.12.1.8 Chemical denaturation assay

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 $R110F_{Br}$) 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 (Fig. A1.22, Table A1.13). 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,[39,68] using the Prism 8 software from Graphpad. The values for these parameters were:

	m in [kcal mol^{-1} M^{-1}] ± SE	$[D]_{50\%}$ in [M] ± SE
Literature	2.06	4.57
recombinant wild-type barnase	1.88 ± 0.21	4.63 ± 0.04
synthetic wild-type barnase	1.82 ± 0.25	4.68 ± 0.06
synthetic R110F(Br) barnase	1.62 ± 0.21	4.46 ± 0.06

Table A1.13. **$[D]_{50\%}$ and m values for recombinant wild-type, synthetic wild-type and synthetic $R110F_{(Br)}$ barnase determined from a chemical denaturation assay shows comparable melting behavior for recombinant and synthetic samples.**

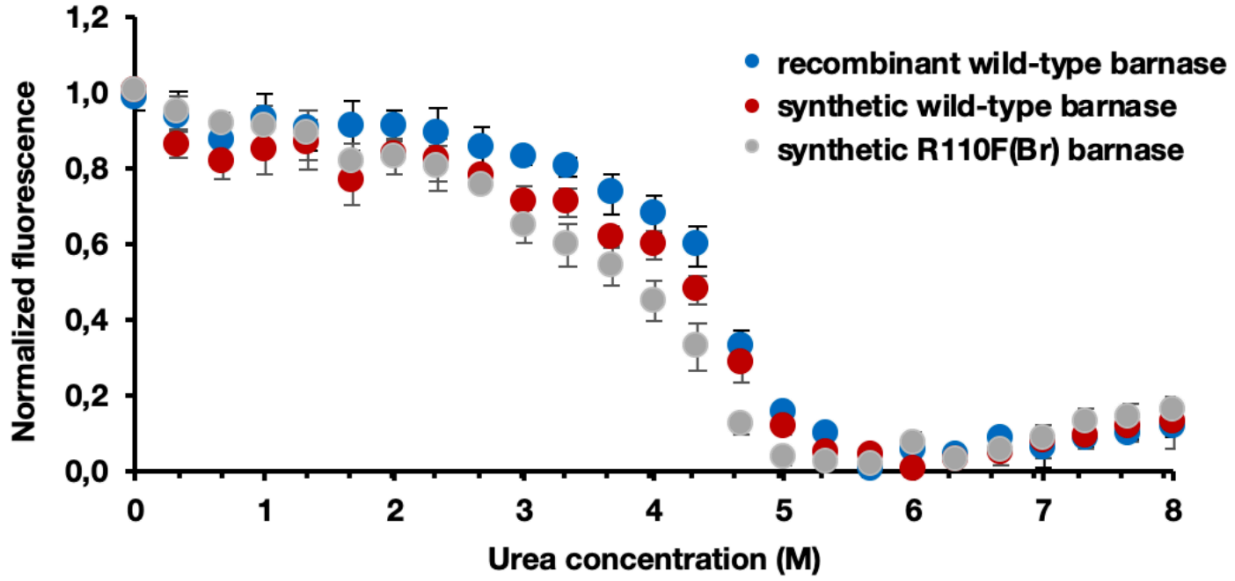


***Figure A1.22*. Urea denaturation of the barnase variants demonstrates the structural similarity of the synthetic and recombinant samples.** The fitted data was used to calculate the m-values as well as $[D]_{50\%}$ of recombinant wild type barnase, synthetic wild type barnase, and synthetic $R110F_{Br}$ of barnase. Error bars indicate SD.

	recombinant wild-type barnase			synthetic wild-type barnase			synthetic R110F(Br) barnase		
	Normalized Fluoresence			Normalized Fluoresence			Normalized Fluoresence		
Urea conc. [M]	Mean	SD	N	Mean	SD	N	Mean	SD	N
8.00	0.112	0.023	3	0.127	0.071	3	0.155	0.023	3
7.67	0.093	0.013	3	0.113	0.038	3	0.142	0.033	3
7.33	0.085	0.019	3	0.090	0.035	3	0.129	0.035	3
7.00	0.056	0.050	3	0.076	0.046	3	0.082	0.013	3
6.67	0.082	0.029	3	0.043	0.028	3	0.055	0.023	3
6.33	0.040	0.014	3	0.026	0.023	3	0.027	0.025	3
6.00	0.052	0.024	3	0.003	0.006	3	0.069	0.036	3
5.67	0.000	0.000	3	0.040	0.009	3	0.016	0.020	3
5.33	0.097	0.017	3	0.049	0.020	3	0.018	0.031	3
5.00	0.150	0.028	3	0.114	0.021	3	0.034	0.020	3
4.67	0.330	0.042	3	0.284	0.049	3	0.121	0.025	3
4.33	0.593	0.051	3	0.476	0.040	3	0.330	0.063	3
4.00	0.676	0.048	3	0.597	0.036	3	0.448	0.055	3
3.67	0.730	0.056	3	0.617	0.028	3	0.542	0.050	3
3.33	0.801	0.027	3	0.711	0.037	3	0.595	0.056	3
3.00	0.828	0.018	3	0.705	0.047	3	0.645	0.045	3
2.67	0.853	0.055	3	0.774	0.023	3	0.754	0.016	3
2.33	0.889	0.067	3	0.821	0.057	3	0.800	0.059	3
2.00	0.907	0.047	3	0.831	0.047	3	0.827	0.046	3
1.67	0.911	0.068	3	0.767	0.064	3	0.817	0.051	3
1.33	0.900	0.055	3	0.863	0.066	3	0.888	0.067	3
1.00	0.926	0.066	3	0.844	0.057	3	0.907	0.057	3
0.67	0.869	0.047	3	0.813	0.045	3	0.916	0.028	3
0.33	0.934	0.069	3	0.860	0.034	3	0.945	0.047	3
0.00	0.982	0.031	3	1.000	0.000	3	1.000	0.000	3

Table A1.14. **Normalized fluorescence and standard deviation from chemical denaturation assay with urea.**

A1.12.2 HIV-1 protease

A1.12.2.1 Folding of synthetic HIV-1 protease (Kent sequence)

The procedure for folding HIV Protease was taken from Johnson et al.(30) 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 denaturated HIV1 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.[70] using darunavir and was found to be 76%.

A1.12.2.2 Fluorogenic protease activity assay

Assay conditions were taken from Johnson et al with slight adaptions taken from the original report for the assay from Toth et al.[30,43] The fluorogenic substrate Abz-Thr-Ile-Nle-p-nitro-Phe-GlnArg-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.,[43] 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 (Fig. A1.23). Kinetic constants were derived from fitting the initial velocities to the Michaelis-Menten equation using Graphpad™ Prism's nonlinear least-squares regression (Fig. A1.24).

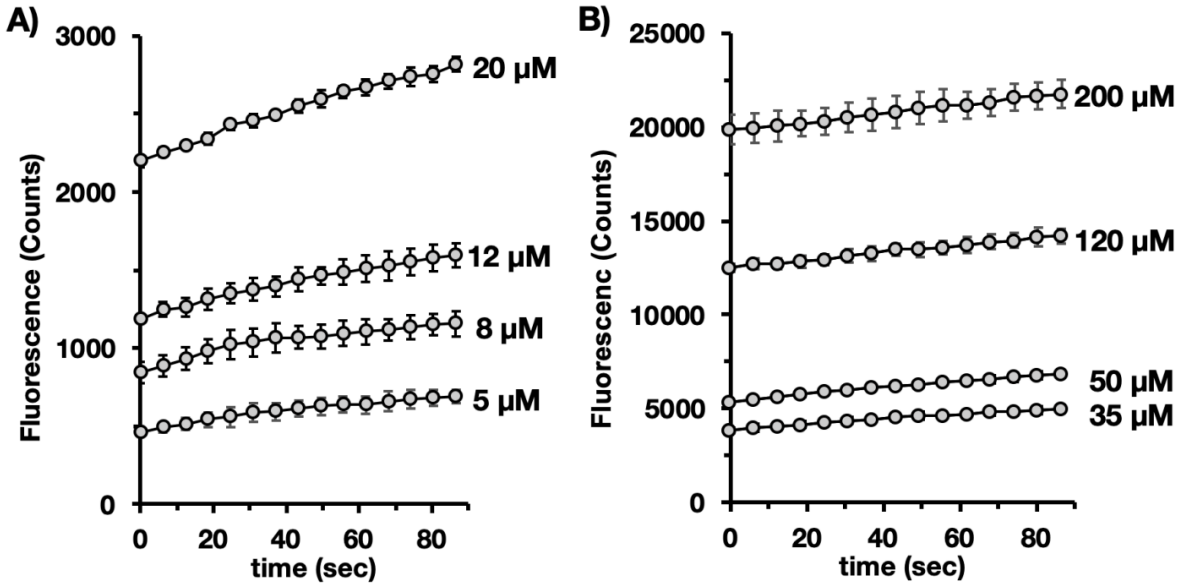


Figure A1.23. **Raw time course measurement data for fluorogenic HIV-1 protease activity assay.** Traces shown represent increases in fluorescence of each sample due to cleavage of the peptide substrate by the protease. All data represent averages over 3 replicates, with error bars plotted as standard deviations.

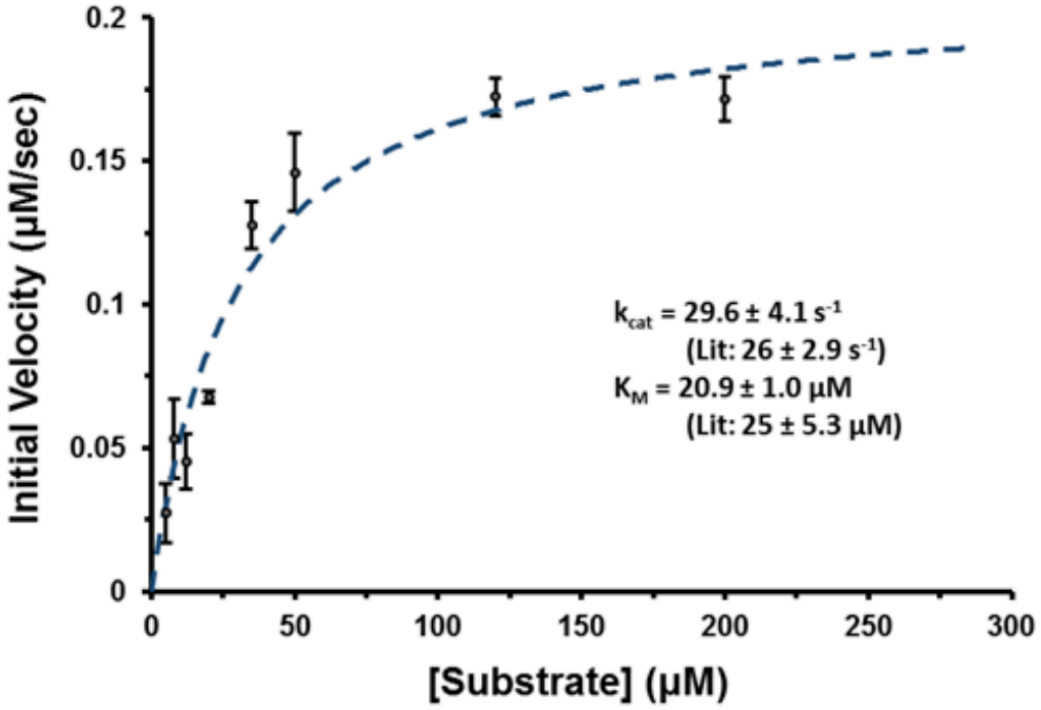


***Figure A1.24*. k_{cat} and K_M values of HIV-1 protease determined by fluorogenic assay.** Error bars plotted as standard error.

A1.12.2.3 Substrate specificity assay

The qualitative assay with model substrate p12nt was previously reported by Schneider et al.[29] 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. A1.26). The cleanest fraction affording 1 mg was used for all further assays, and analytical LC-MS data is shown in Fig. A1.25. 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.

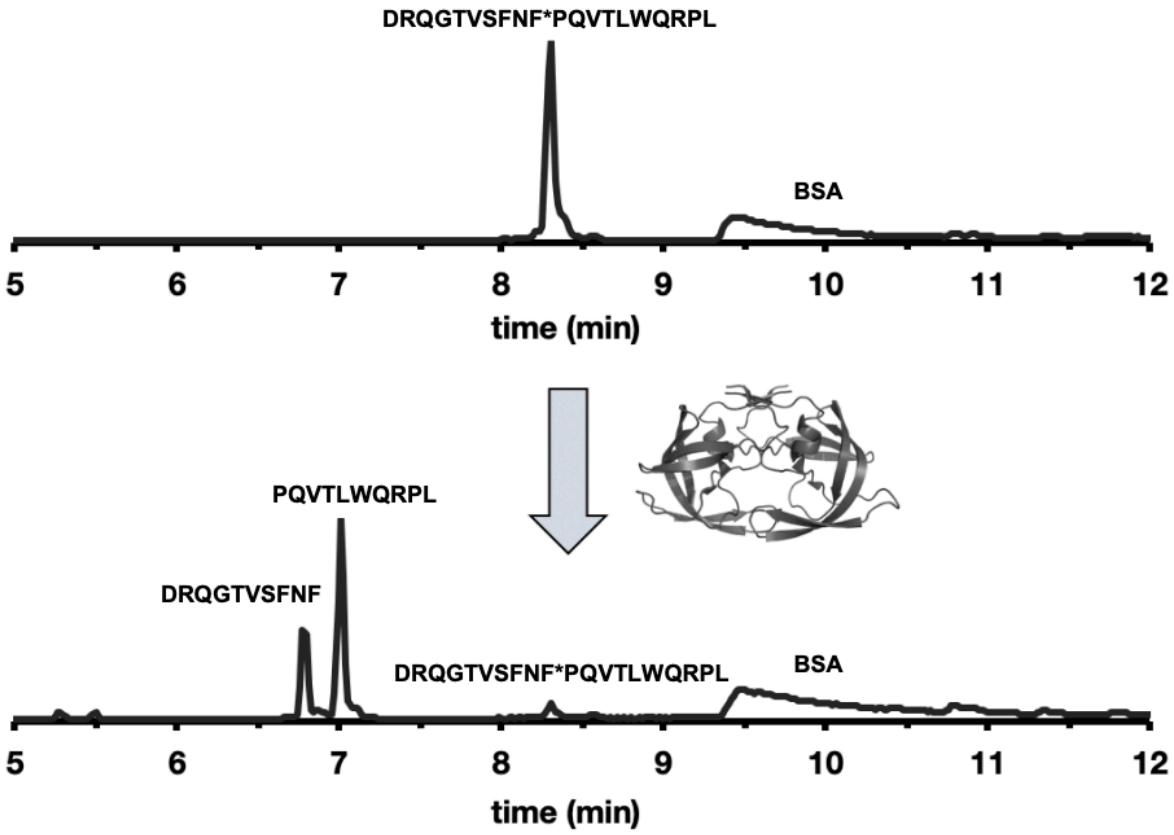


***Figure A1.25*. Synthetic HIV-1 protease shows substrate specific activity.** Incubation of the synthetic protease with a peptide containing a conserved HIV-1 protease consensus cleavage site leads to specific hydrolysis. No observation of off-target hydrolysis at separate sites, nor of native BSA in solution was observed. Traces shown represent separation of the reaction mixture onto LC-MS using standard method 1-61% B over 15 min, Zorbax C3 column (6550).

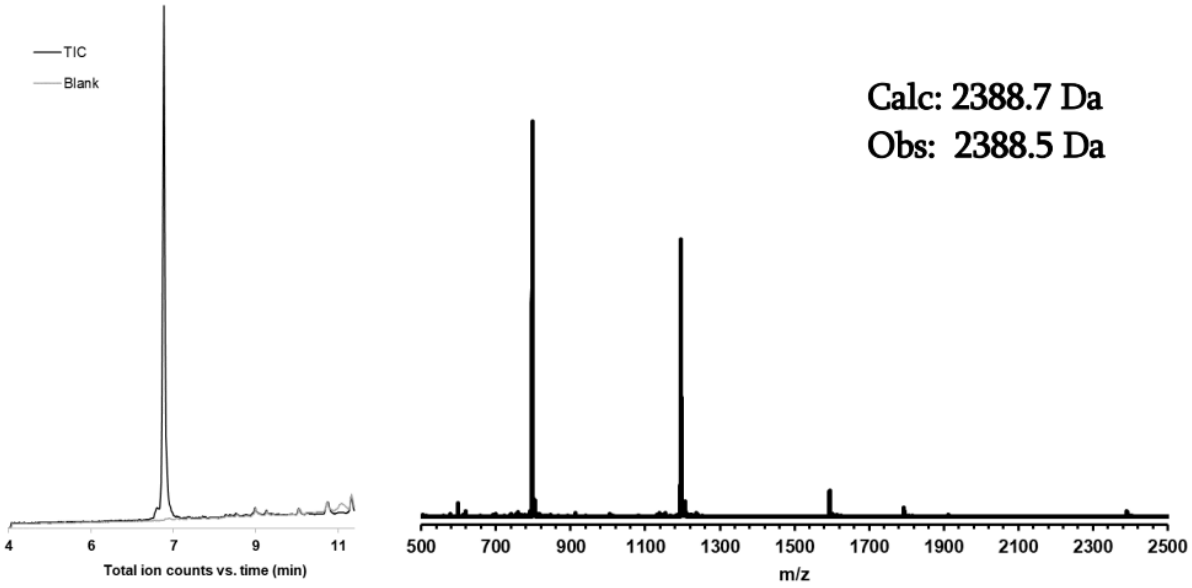


***Figure A1.26*. Characterization of purified qualitative substrate p12nt.** LC-MS trace recorded with standard method 1-91% B over 9 min, Zorbax C3 column (6520). Left figure shows the TIC trace, and the m/z extraction is shown for the major peak on the right. Calculated mass and observed mass are in good agreement.

A1.12.3 Sortase A$^{[59\text{-}206]}$; P94S/D160N/K196T

A1.12.3.1 Expression and purification of recombinant sortase A[59-206]; P94S/D160N/K196T

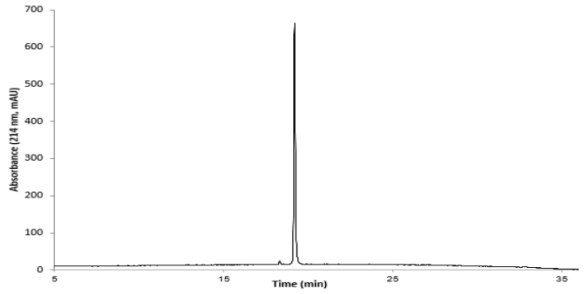
Recombinant sortase A[59-206] was recombinantly expressed following a protocol described in the literature.[44]

A) Recombinant Sortase A[59-206];P94S/D160N/K196T

Sequence: ASMTGGQQMG RDPNSQAKPQ IPKDKSKVAG YIEIPDADIK EPVYPGPATS EQLNRGVSFA EENESLDDQN ISIAGHTFID RPNYQFTNLK AAKKGSMVYF KVGNETRKYK MTSIRNVKPT DVEVLDEQKG KDKQLTLITC DDYNEKTGVW ETRKIFVATE VKLEHHHHHH (170 AA)

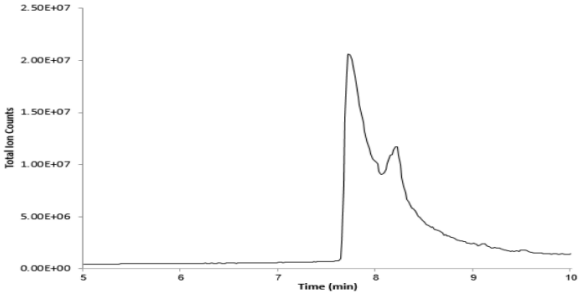
B) Analytical data for recombinant Sortase A[59-206];P94S/D160N/K196T

Analytical HPLC:



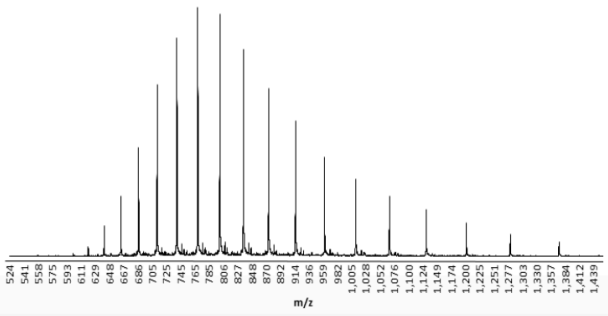
method: 5–65% B over 30 min

LC-MS:

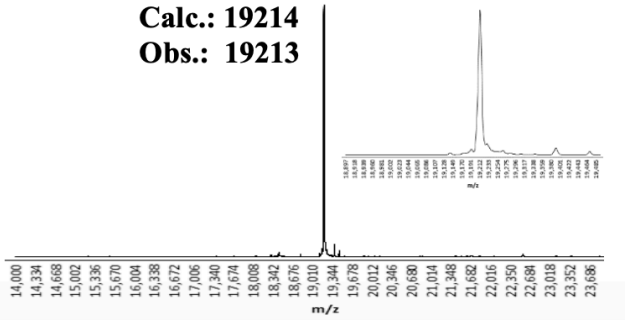


method: 1-61%B over 15 min, Zorbax C3 column (6520)

MS Data from LC-MS:



Deconvolution of LC-MS data:



A1.12.3.2 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 A_{280} 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 (Superdex^{TM} 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 in Fig. A1.27A. 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 (Fig. A1.27B).

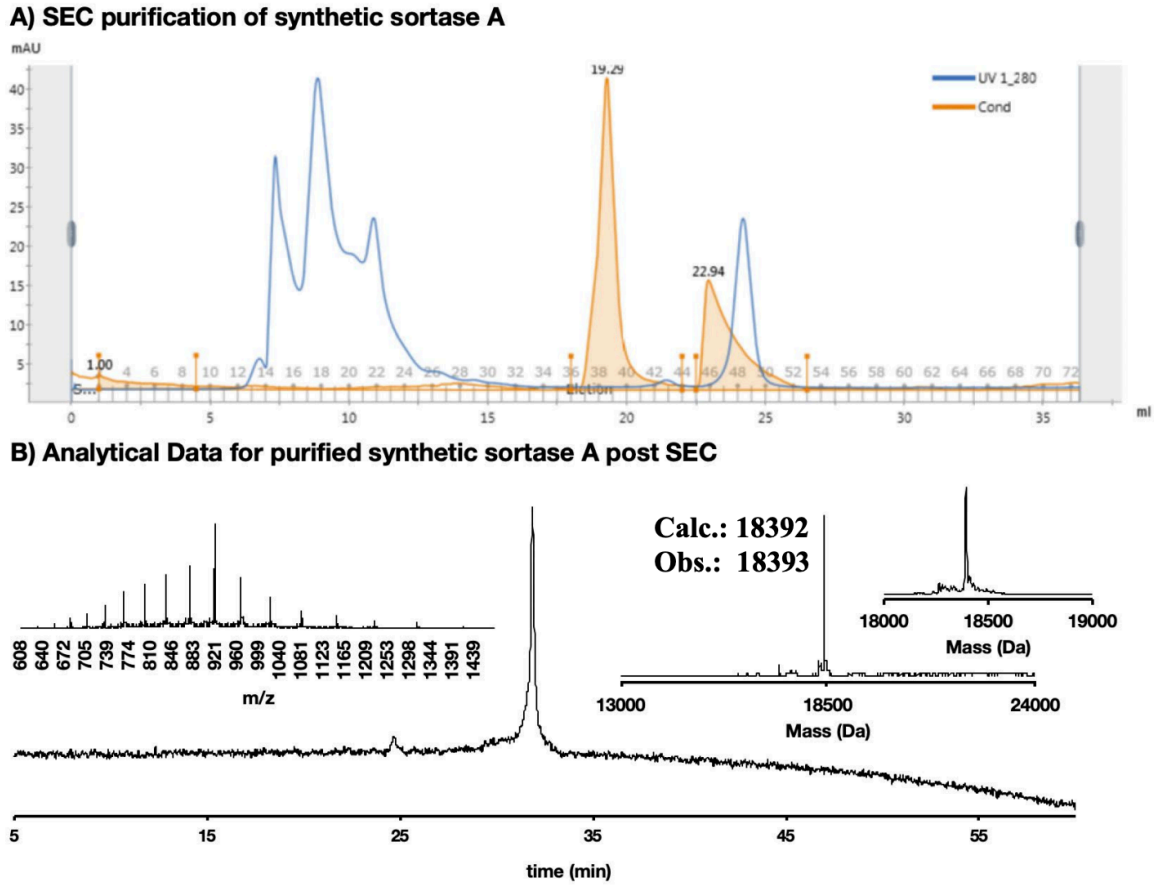
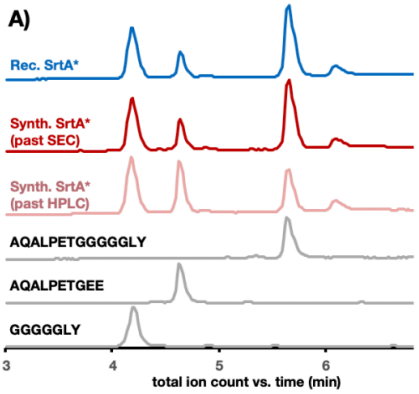


***Figure A1.27*. Folding and SEC purification of synthetic sortase A yields clean protein sample.** A) HPLC-purified sortase was subjected to folding conditions followed by size exclusion chromatography. The elution profile from the size exclusion chromatography is displayed in this figure. B) Analytical HPLC trace of purified synthetic

sortase A post size exclusion chromatography (absorbance at 214 nm), LC-MS data (m/z in Da), and deconvoluted mass spectra (obtained from integration over all LC-MS signals).

A1.12.3.3 Semiquantitative activity assay

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 (Fig. A1.28).



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 A1.28*. Synthetic sortase A shows enzymatic activity in semiquantitative assay.** 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.

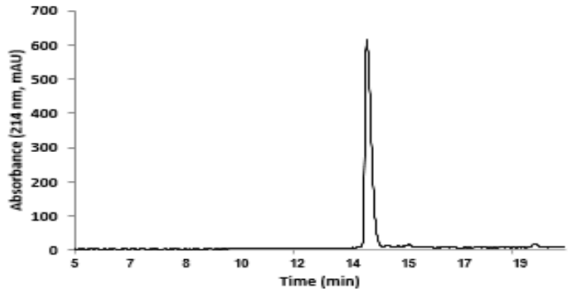
A) Synthesis data for substrate 1 (Sortase A semiquantitative assay)

Sequence:	GGGGGLY (7 AA)
Resin:	200 mg of RINK amine ChemMatrix® (0.49 mmol/g), yielding the C-terminal amide after cleavage

B) Cleavage and analytical data for crude substrate 1 (Sortase A semiquantitative assay)

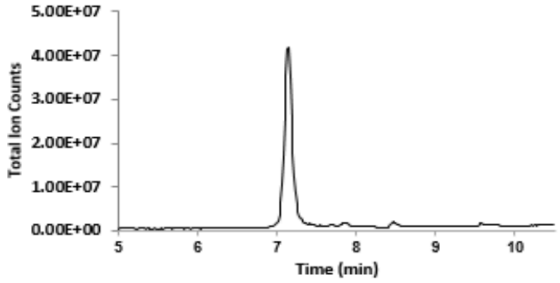
Cleavage:	Cleavage protocol A for peptides, no ether wash, evaporated under N_2

Analytical HPLC:



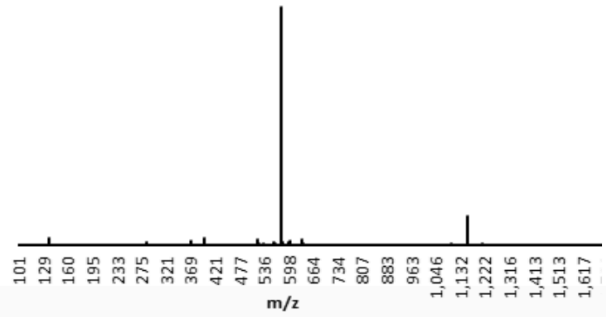
method: 1–31% B over 20 min

LC-MS:



method: 1-61%B over 15 min, Zorbax C18 column (6520)

MS Data from LC-MS:



calc.: 578.27
obs.: 578.22

A) Synthesis data for substrate 2 (Sortase A semiquantitative assay)

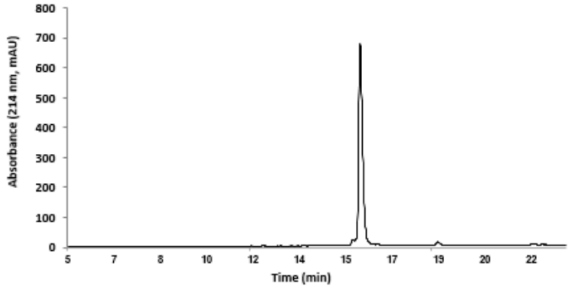
Sequence: AQALPETGEE (10 AA)

Resin: 150 mg of RINK amine ChemMatrix® (0.49 mmol/g), yielding the C-terminal amide after cleavage

B) Cleavage and analytical data for crude substrate 2 (Sortase A semiquantitative assay)

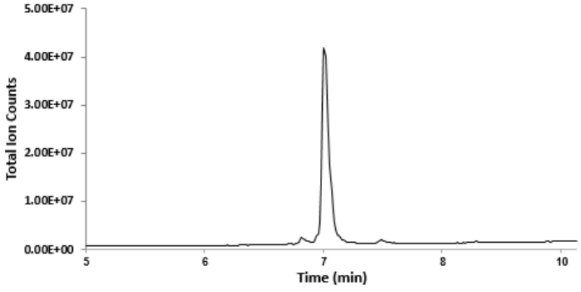
Cleavage: Cleavage protocol A for peptides

Analytical HPLC:



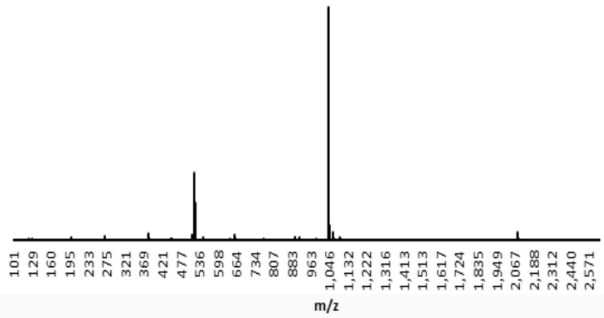
method: 1–31% B over 20 min

LC-MS:



method: 1-61%B over 15 min, Zorbax C18 column (6520)

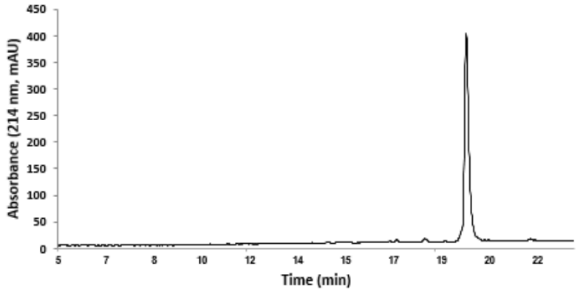
MS Data from LC-MS:



calc.: 1042.48
obs.: 1042.53

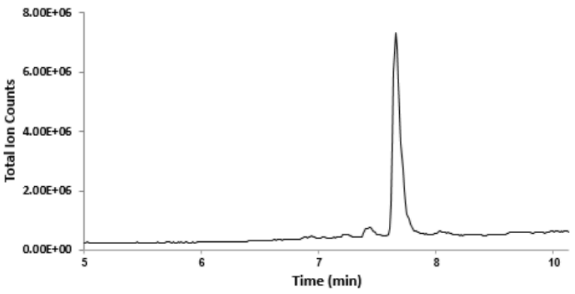
A) Synthesis Data for ligation standard (SrtA semiquantitative assay)	
Sequence:	AQALPETGGG GGLY (14 AA)
Resin:	150 mg of RINK amine ChemMatrix® (0.49 mmol/g), yielding the C-terminal amide after cleavage
B) Cleavage and analytical data for crude ligation standard (SrtA semiquantitative assay)	
Cleavage:	Cleavage protocol A for peptides

Analytical HPLC:



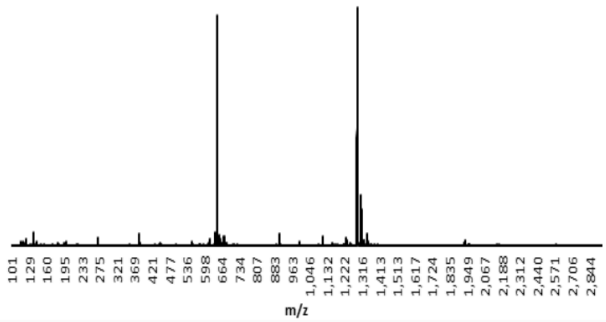
method: 1–31% B over 20 min

LC-MS:



method: 1-61%B over 15 min, Zorbax C18 column (6520)

MS Data from LC-MS:



calc.: 1288.63
obs.: 1288.86

A1.12.4 MDM2[1-118]

A1.12.4.1 Folding of synthetic MDM2[1-118]

Synthetic MDM2[1-118] was refolded according to the procedure from Zhan et al.[70] 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.

A1.12.4.2 Preparation of biotinylated p53(15-29)

15-29P53-like peptide (SQETFSDLWKLLPEN) was synthesized by Fmoc-based SPPS, with HRink 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 described in section A1.8.3 (method A).

A1.12.4.3 Assay of MDM2 binding activity, by biolayer interferometry

Synthetic MDM2 in refolding buffer (see section A1.12.4.1) 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 (Fig. A1.29, indicated concentrations). BLI assay was run in triplicates. Equilibrium response (nm) was plotted against MDM2 concentration to determine Kd of synthetic and recombinant MDM2 (Fig. A1.30).

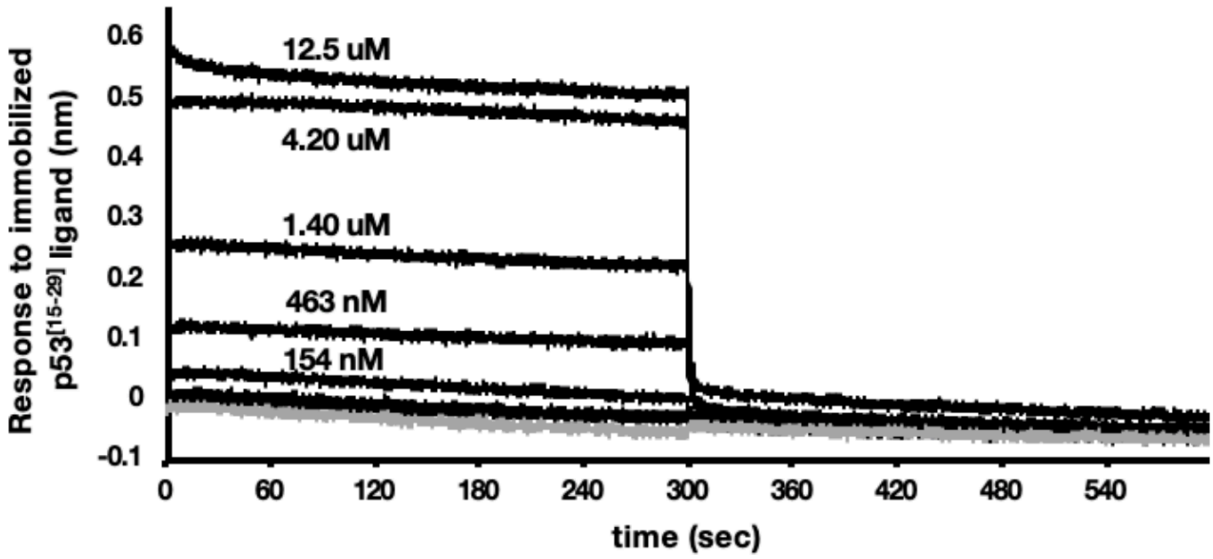


***Figure A1.29*. Sensogram of MDM2[1-118] .** Experiment was conducted in triplicates, for clarity only a single data set is displayed here.

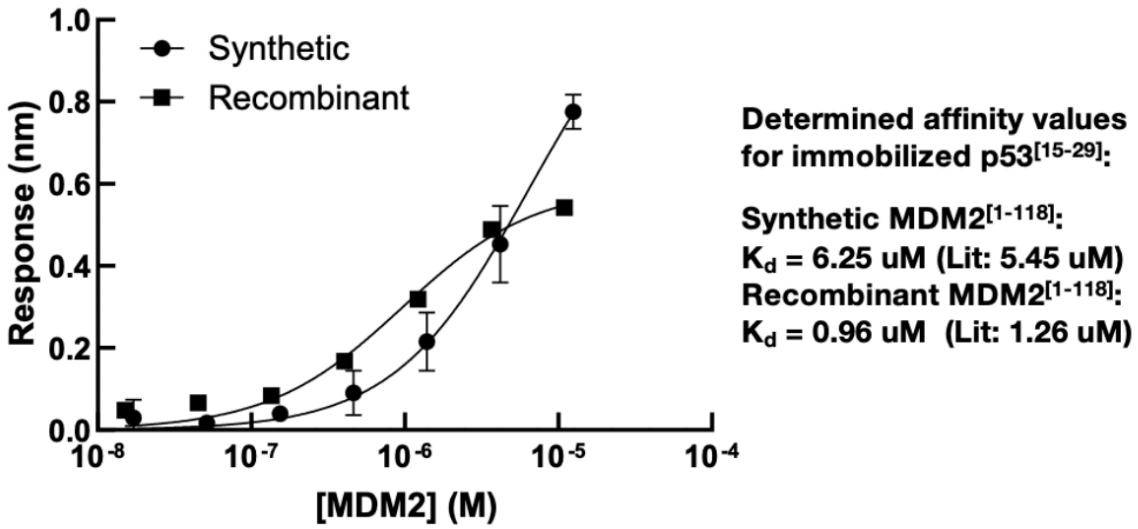


***Figure A1.30*. Determination of K_d from concentration-dependent equilibrium binding responses**. All experiments were carried out in triplicates and K_d values were determined from equilibrium binding responses of immobilized p53 to MDM2 at various concentrations.[71]

A1.13 Acknowledgements

We thank T. F. Jamison, H. U. Stilz, L. F. Iversen, K. Little, and 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 P. E. Dawson, R. T. Raines, and S. B. H. Kent, for helpful discussions and for suggesting additional experiments. We also thank E. D. Evans, F. W. W. Hartrampf, and R. L. Holden for careful proofreading of the manuscript. We are grateful to N. L. 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., and 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

A1.14 Author contributions

N.H., T.E.N., and B.L.P. conceptualized the research; N.H., M.P., A.J.M., and S.L. optimized synthesis conditions; A.J.M. provided the software for UV data analysis, and M.D.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 coauthors.

A1.15 Supplementary references for A1.8-A1.12

1. J. Britton, C. L. Raston, Multi-step continuous-flow synthesis. Chem. Soc. Rev. 46, 1250–1271 (2017). doi:10.1039/C6CS00830E Medline

2. D. Webb, T. F. Jamison, Continuous flow multi-step organic synthesis. Chem. Sci. 1, 675–680 (2010). doi:10.1039/c0sc00381f

3. S. Bondalapati, M. Jbara, A. Brik, Expanding the chemical toolbox for the synthesis of large and uniquely modified proteins. Nat. Chem. 8, 407–418 (2016). doi:10.1038/nchem.2476 Medline

4. J. W. Chin, Expanding and reprogramming the genetic code. Nature 550, 53–60 (2017). doi:10.1038/nature24031 Medline

5. R. B. Merrifield, Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. J. Am. Chem. Soc. 85, 2149–2154 (1963). doi:10.1021/ja00897a025

6. S. B. H. Kent, Total chemical synthesis of proteins. Chem. Soc. Rev. 38, 338–351 (2009). doi:10.1039/B700141J Medline

7. A. A. Zompra, A. S. Galanis, O. Werbitzky, F. Albericio, Manufacturing peptides as active pharmaceutical ingredients. Future Med. Chem. 1, 361–377 (2009). doi:10.4155/fmc.09.23 Medline

8. P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, Synthesis of proteins by native chemical ligation. Science 266, 776–779 (1994). doi:10.1126/science.7973629 Medline

9. V. Agouridas, O. El Mahdi, V. Diemer, M. Cargoët, J. M. Monbaliu, O. Melnyk, Native chemical ligation and extended methods: Mechanisms, catalysis, scope, and limitations. Chem. Rev. 119, 7328–7443 (2019). doi:10.1021/acs.chemrev.8b00712 Medline

10. P. E. Dawson, S. B. H. Kent, Synthesis of native proteins by chemical ligation. Annu. Rev. Biochem. 69, 923–960 (2000). doi:10.1146/annurev.biochem.69.1.923 Medline

11. A. El-Faham, F. Albericio, Peptide coupling reagents, more than a letter soup. Chem. Rev. 111, 6557–6602 (2011). Doi:10.1021/cr100048w Medline

12. N. Ahmed, Peptide bond formations through flow chemistry. Chem. Biol. Drug Des. 91, 647–650 (2018). doi:10.1111/cbdd.13115 Medline

13. C. P. Gordon, The renascence of continuous-flow peptide synthesis – an abridged account of solid and solution-based approaches. Org. Biomol. Chem. 16, 180–196 (2018). doi:10.1039/C7OB02759A Medline

14. A. J. Mijalis, A. Steinauer, A. Schepartz, B. L. Pentelute, “Pushing the limits of solid-phase peptide synthesis with continuous flow” in Flow Chemistry in Organic Synthesis (Georg Thieme Verlag, Stuttgart, 2018), pp. 381–398.

15. E. Bayer, G. Jun, I. Halász, I. Sebestian, A new support for polypeptide synthesis in columns. Tetrahedron Lett. 11, 4503–4505 (1970). doi:10.1016/S0040-4039(01)83961-7 Medline

16. T. J. Lukas, M. B. Prystowsky, B. W. Erickson, Solid-phase peptide synthesis under continuous-flow conditions. Proc. Natl. Acad. Sci. U.S.A. 78, 2791–2795 (1981). doi:10.1073/pnas.78.5.2791 Medline 91

17. A. J. Mijalis, D. A. Thomas III, M. D. Simon, A. Adamo, R. Beaumont, K. F. Jensen, B. L. Pentelute, A fully automated flow-based approach for accelerated peptide synthesis. Nat. Chem. Biol. 13, 464–466 (2017). doi:10.1038/nchembio.2318 Medline

18. S. K. Mong, A. A. Vinogradov, M. D. Simon, B. L. Pentelute, Rapid total synthesis of DARPin pE59 and barnase. ChemBioChem 15, 721–733 (2014). doi:10.1002/cbic.201300797 Medline

19. M. D. Simon, “Fast flow biopolymer synthesis,” thesis, Massachusetts Institute of Technology, Cambridge, MA (2017).

20. S. Fuse, Y. Otake, H. Nakamura, Peptide synthesis utilizing micro-flow technology. Chem. Asian J. 13, 3818–3832 (2018). doi:10.1002/asia.201801488 Medline

21. S. Fuse, Y. Mifune, T. Takahashi, Efficient amide bond formation through a rapid and strong activation of carboxylic acids in a microflow reactor. Angew. Chem. Int. Ed. 53, 851–855 (2014). doi:10.1002/anie.201307987 Medline

22. L. K. Spare, V. Laude, D. G. Harman, J. R. Aldrich-Wright, C. P. Gordon, An optimised approach for continuous-flow solid-phase peptide synthesis utilising a rudimentary flow reactor. React. Chem. Eng. 3, 875–882 (2018). doi:10.1039/C8RE00190A

23. E. T. Sletten, M. Nuño, D. Guthrie, P. H. Seeberger, Real-time monitoring of solid-phase peptide synthesis using a variable bed flow reactor. Chem. Commun. 55, 14598–14601 (2019). doi:10.1039/C9CC08421E Medline

24. M. Schnölzer, P. Alewood, A. Jones, D. Alewood, S. B. H. Kent, In situ neutralization in Boc-chemistry solid phase peptide synthesis. Int. J. Pept. Protein Res. 40, 180–193 (1992). doi:10.1111/j.1399-3011.1992.tb00291.x Medline

25. T. Michels, R. Dölling, U. Haberkorn, W. Mier, Acid-mediated prevention of aspartimide formation in solid phase peptide synthesis. Org. Lett. 14, 5218–5221 (2012). doi:10.1021/ol3007925 Medline

26. M. Mergler, F. Dick, B. Sax, P. Weiler, T. Vorherr, The aspartimide problem in Fmoc-based SPPS. Part I. J. Pept. Sci. 9, 36–46 (2003). doi:10.1002/psc.430 Medline

27. S. A. Palasek, Z. J. Cox, J. M. Collins, Limiting racemization and aspartimide formation in microwave-enhanced Fmoc solid phase peptide synthesis. J. Pept. Sci. 13, 143–148 (2007). doi:10.1002/psc.804 Medline

28. S. Luisier, M. Avital-Shmilovici, M. A. Weiss, S. B. H. Kent, Total chemical synthesis of human proinsulin. Chem. Commun. 46, 8177–8179 (2010). doi:10.1039/c0cc03141k Medline

29. J. Schneider, S. B. H. Kent, Enzymatic activity of a synthetic 99 residue protein corresponding to the putative HIV-1 protease. Cell 54, 363–368 (1988). doi:10.1016/0092-8674(88)90199-7 Medline

30. E. C. B. Johnson, E. Malito, Y. Shen, D. Rich, W.-J. Tang, S. B. H. Kent, Modular total chemical synthesis of a human immunodeficiency virus type 1 protease. J. Am. Chem. Soc. 129, 11480–11490 (2007). doi:10.1021/ja072870n Medline 92

31. A. Brik, C.-H. Wong, HIV-1 protease: Mechanism and drug discovery. Org. Biomol. Chem. 1, 5–14 (2003). doi:10.1039/b208248a Medline

32. J. D. Oliner, J. A. Pietenpol, S. Thiagalingam, J. Gyuris, K. W. Kinzler, B. Vogelstein, Oncoprotein MDM2 conceals the activation domain of tumour suppressor p53. Nature 362, 857–860 (1993). doi:10.1038/362857a0 Medline

33. D. Bresson, L. Togher, E. Rodrigo, Y. Chen, J. A. Bluestone, K. C. Herold, M. von Herrath, Anti-CD3 and nasal proinsulin combination therapy enhances remission from recentonset autoimmune diabetes by inducing Tregs. J. Clin. Invest. 116, 1371–1381 (2006). doi:10.1172/JCI27191 Medline

34. E. Gasser, C. P. Moutos, M. Downes, R. M. Evans, FGF1 – a new weapon to control type 2 diabetes mellitus. Nat. Rev. Endocrinol. 13, 599–609 (2017). doi:10.1038/nrendo.2017.78 Medline

35. L. Serrano, J. T. Kellis Jr., P. Cann, A. Matouschek, A. R. Fersht, The folding of an enzyme: II. Substructure of barnase and the contribution of different interactions to protein stability. J. Mol. Biol. 224, 783–804 (1992). doi:10.1016/0022-2836(92)90562-X Medline

36. C. B. Anfinsen, Principles that govern the folding of protein chains. Science 181, 223–230 (1973). doi:10.1126/science.181.4096.223 Medline

37. S. B. Kent, P. F. Alewood, Editorial overview: Synthetic biomolecules. Curr. Opin. Chem. Biol. 22, viii–xi (2014). doi:10.1016/j.cbpa.2014.09.037 Medline

38. R. W. Hartley, Barnase and barstar: Two small proteins to fold and fit together. Trends Biochem. Sci. 14, 450–454 (1989). doi:10.1016/0968-0004(89)90104-7 Medline

39. J. T. Kellis Jr., K. Nyberg, A. R. Fersht, Energetics of complementary side-chain packing in a protein hydrophobic core. Biochemistry 28, 4914–4922 (1989). doi:10.1021/bi00437a058 Medline

40. J. R. Knowles, Tinkering with enzymes: What are we learning? Science 236, 1252–1258 (1987). doi:10.1126/science.3296192 Medline

41. U. Arnold, M. P. Hinderaker, J. Köditz, R. Golbik, R. Ulbrich-Hofmann, R. T. Raines, Protein prosthesis: A nonnatural residue accelerates folding and increases stability. J. Am. Chem. Soc. 125, 7500–7501 (2003). doi:10.1021/ja0351239 Medline

42. B. R. Kelemen, T. A. Klink, M. A. Behlke, S. R. Eubanks, P. A. Leland, R. T. Raines, Hypersensitive substrate for ribonucleases. Nucleic Acids Res. 27, 3696–3701 (1999). doi:10.1093/nar/27.18.3696 Medline

43. M. V. Toth, G. R. Marshall, A simple, continuous fluorometric assay for HIV protease. Int. J. Pept. Protein Res. 36, 544–550 (1990). doi:10.1111/j.1399-3011.1990.tb00994.x Medline

44. H. Ton-That, G. Liu, S. K. Mazmanian, K. F. Faull, O. Schneewind, Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of Staphylococcus aureus at the LPXTG motif. Proc. Natl. Acad. Sci. U.S.A. 96, 12424– 12429 (1999). doi:10.1073/pnas.96.22.12424 Medline 93

45. J. J. Ling, R. L. Policarpo, A. E. Rabideau, X. Liao, B. L. Pentelute, Protein thioester synthesis enabled by sortase. J. Am. Chem. Soc. 134, 10749–10752 (2012). doi:10.1021/ja302354v Medline

46. R. L. Policarpo, H. Kang, X. Liao, A. E. Rabideau, M. D. Simon, B. L. Pentelute, Flowbased enzymatic ligation by sortase A. Angew. Chem. Int. Ed. 53, 9203–9208 (2014). doi:10.1002/anie.201403582 Medline

47. A. A. Vinogradov, E. D. Evans, B. L. Pentelute, Total synthesis and biochemical characterization of mirror image barnase. Chem. Sci. 6, 2997–3002 (2015). doi:10.1039/C4SC03877K Medline

48. F. Touti, Z. P. Gates, A. Bandyopadhyay, G. Lautrette, B. L. Pentelute, In-solution enrichment identifies peptide inhibitors of protein-protein interactions. Nat. Chem. Biol. 15, 410–418 (2019). doi:10.1038/s41589-019-0245-2 Medline

49. P. Chène, Inhibition of the p53-MDM2 interaction: Targeting a protein-protein interface. Mol. Cancer Res. 2, 20–28 (2004). Medline

50. J.-F. Lutz, M. Ouchi, D. R. Liu, M. Sawamoto, Sequence-controlled polymers. Science 341, 1238149 (2013). doi:10.1126/science.1238149 Medline

51. S. J. Wheelan, A. Marchler-Bauer, S. H. Bryant, Domain size distributions can predict domain boundaries. Bioinformatics 16, 613–618 (2000). doi:10.1093/bioinformatics/16.7.613 Medline

52. M. D. Simon, P. L. Heider, A. Adamo, A. A. Vinogradov, S. K. Mong, X. Li, T. Berger, R. L. Policarpo, C. Zhang, Y. Zou, X. Liao, A. M. Spokoyny, K. F. Jensen, B. L. Pentelute, Rapid flow-based peptide synthesis. ChemBioChem 15, 713–720 (2014). doi:10.1002/cbic.201300796 Medline

53. J. C. Li, T. Liu, Y. Wang, A. P. Mehta, P. G. Schultz, Enhancing protein stability with genetically encoded noncanonical amino acids. J. Am. Chem. Soc. 140, 15997–16000 (2018). doi:10.1021/jacs.8b07157 Medline

54. H. Xiao, F. Nasertorabi, S. H. Choi, G. W. Han, S. A. Reed, R. C. Stevens, P. G. Schultz, Exploring the potential impact of an expanded genetic code on protein function. Proc. Natl. Acad. Sci. U.S.A. 112, 6961–6966 (2015). doi:10.1073/pnas.1507741112 Medline

55. R. J. Desnick, E. H. Schuchman, Enzyme replacement and enhancement therapies: Lessons from lysosomal disorders. Nat. Rev. Genet. 3, 954–966 (2002). doi:10.1038/nrg963 Medline

56. N. L. Truex, R. L. Holden, B.-Y. Wang, P.-G. Chen, S. Hanna, Z. Hu, K. Shetty, O. Olive, D. Neuberg, N. Hacohen, D. B. Keskin, P. A. Ott, C. J. Wu, B. L. Pentelute, Automated flow synthesis of tumor neoantigen peptides for personalized immunotherapy. Sci. Rep. 10, 723 (2020). doi:10.1038/s41598-019-56943-5 Medline

57. A. M. Buckle, G. Schreiber, A. R. Fersht, Protein-protein recognition: Crystal structural analysis of a barnase-barstar complex at 2.0-Å resolution. Biochemistry 33, 8878–8889 (1994). doi:10.1021/bi00196a004 Medline

58. Y. Yang, Q. X. Hua, J. Liu, E. H. Shimizu, M. H. Choquette, R. B. Mackin, M. A. Weiss, Solution structure of proinsulin: Connecting domain flexibility and prohormone 94 processing. J. Biol. Chem. 285, 7847–7851 (2010). doi:10.1074/jbc.C109.084921 Medline

59. J. Sevcík, L. Urbanikova, Z. Dauter, K. S. Wilson, Recognition of RNase Sa by the inhibitor barstar: Structure of the complex at 1.7 Å resolution. Acta Crystallogr. D Biol. Crystallogr. 54, 954–963 (1998). doi:10.1107/S0907444998004429 Medline

60. J. Bella, B. Brodsky, H. M. Berman, Hydration structure of a collagen peptide. Structure 3, 893–906 (1995). doi:10.1016/S0969-2126(01)00224-6 Medline

61. A. Czarna, G. M. Popowicz, A. Pecak, S. Wolf, G. Dubin, T. A. Holak, High affinity interaction of the p53 peptide-analogue with human Mdm2 and Mdmx. Cell Cycle 8, 1176–1184 (2009). doi:10.4161/cc.8.8.8185 Medline

62. T. Durek, V. Y. Torbeev, S. B. H. Kent, Convergent chemical synthesis and high-resolution x-ray structure of human lysozyme. Proc. Natl. Acad. Sci. U.S.A. 104, 4846–4851 (2007). doi:10.1073/pnas.0610630104 Medline

63. X. Xia, L. M. Longo, M. Blaber, Mutation choice to eliminate buried free cysteines in protein therapeutics. J. Pharm. Sci. 104, 566–576 (2015). doi:10.1002/jps.24188 Medline

64. N. Suree, C. K. Liew, V. A. Villareal, W. Thieu, E. A. Fadeev, J. J. Clemens, M. E. Jung, R. T. Clubb, The structure of the Staphylococcus aureus sortase-substrate complex reveals how the universally conserved LPXTG sorting signal is recognized. J. Biol. Chem. 284, 24465–24477 (2009). doi:10.1074/jbc.M109.022624 Medline

65. A. J. Mijalis, amijalis/afps-integration: AFPS-analysis: Code used to extract and analyze data from AFPS synthesis files. Zenodo (2020); doi:10.5281/zenodo.3727779.

66. A. L. Okorokov, R. W. Hartley, K. I. Panov, An improved system for ribonuclease Ba expression. Protein Expr. Purif. 5, 547–552 (1994). doi:10.1006/prep.1994.1075 Medline

67. M. C. R. Shastry, J. B. Udgaonkar, The folding mechanism of barstar: Evidence for multiple pathways and multiple intermediates. J. Mol. Biol. 247, 1013–1027 (1995). doi:10.1006/jmbi.1994.0196 Medline

68. R. Loewenthal, J. Sancho, A. R. Fersht, Fluorescence spectrum of barnase: Contributions of three tryptophan residues and a histidine-related pH dependence. Biochemistry 30, 6775– 6779 (1991). doi:10.1021/bi00241a021 Medline

69. I. W. Windsor, R. T. Raines, Fluorogenic assay for inhibitors of HIV-1 protease with subpicomolar affinity. Cell 54, 363–368 (1988).

70. C. Zhan, K. Varney, W. Yuan, L. Zhao, W. Lu, Interrogation of MDM2 phosphorylation in p53 activation using native chemical ligation: The functional role of Ser17 phosphorylation in MDM2 reexamined. J. Am. Chem. Soc. 134, 6855–6864 (2012). doi:10.1021/ja301255n Medline

71. E. M. Nicholson, J. M. Scholtz, Conformational stability of the Escherichia coli HPr protein: Test of the linear extrapolation method and a thermodynamic characterization of cold denaturation. Biochemistry 35, 11369–11378 (1996). doi:10.1021/bi960863y Medline

Appendix II Rapid de novo discovery of peptidomimetic affinity reagents for human angiotensin converting enzyme 2

The work presented in this appendix has been reproduced from the following publication with permission through a license signed by MIT with the Nature:

Zhang, G.*; Brown, J.S.*; Quartararo, A.J.; Li, C.; Tan, X.; Hanna, **S.; Antilla**, S.; Cowfer, A.E.; Loas, A.; Pentelute, B.L., Rapid de novo discovery of peptidomimetic affinity reagents for human angiotensin converting enzyme 2. *Communications Chemistry,* **2022**, 5, 8.

**These authors contributed equally to this work.*

The work presented in this appendix was a team effort and is intended to supplement and contextualize Chapter 1.4.

A2.1 Introduction

The discovery of high affinity reagents is a critical initial step in drug discovery, diagnostic development, and proteome profiling.[1-3] High affinity ligands are under constant development for the modulation of activity or function of target proteins, including challenging protein–protein interactions (PPIs) and intracellular targets.[4,5] Diagnostics inform clinical decision making and require sensitive and selective detection within highly complex media often containing a broad range of other protein concentrations (e.g., plasma).[6] Lastly, affinity reagents have been crucial for understanding the proteome, facilitating the compilation of a knowledge base for protein expression profiles and localization across normal and disease tissue (e.g., Human Protein Atlas).[2,7,8] Yet, the vast majority of known proteins have no corresponding affinity reagent.[3,9] Overall, each of these fields relies on the rapid discovery and development of selective, high affinity binders to specific protein targets.

Current methods to discover high affinity reagents against protein targets vary in production speed and chemical diversity, ranging from antibody production to panning fully synthetic libraries. While they are the ‘gold standard,’ antibodies require a long production timeline, demonstrate low tissue penetration, and can exhibit batch variability.[10,11] Thus, several non-antibody systems including DNA-encoded libraries,[12,13] aptamers,[14,15] and peptide discovery platforms,[16] have been developed to discover specific, high affinity binders. Most ‘hit’ discovery techniques rapidly isolate and enrich binders from libraries based on their high affinity to the target protein, though high-throughput screening of individual compounds has also been used.[17,18] The design and diversity of these libraries greatly affect the rate of discovery against a novel target.

Larger molecular scaffolds can bind broader portions of protein surfaces, enabling the efficient targeting of PPIs.[3,5,15,19,20] The chemical and structural diversity of the curated library can improve discovery success.[17,18,21,22] Thus, an ideal discovery platform should combine high chemical diversity and rapid responsiveness to new clinically relevant targets.

Much research has been devoted to the discovery and engineering of peptidomimetic binders because of their broad access to diverse chemical spaces, amenability to rapid synthesis or modification, and availability of multiple rapid discovery platforms.[21,23] The use of noncanonical amino acids, macrocyclization, and chemical stapling, in particular, have been proven to be useful in promoting cell uptake, proteolytic stability, and improved pharmacokinetics.[16,21,24-27] Discovery via affinity selection using genetically-encoded techniques including phage display[28,29] and mRNA display[30] samples vast libraries up to 10^{13} members, being amenable for *de novo* discovery of high affinity reagents. However, these techniques are not well suited to the incorporation of highly noncanonical library members, even in cell-free systems.[31-34] Thus, following the initial identification of high affinity peptides, further development is required via iterative synthetic cycles of derivatization and screening.

High affinity peptidomimetic binders can also be identified by affinity selection-mass spectrometry (AS-MS), enabling the straightforward use of entirely noncanonical synthetic libraries without expending rapid discovery.[35-37] AS-MS generally functions through the enrichment and identification of peptidomimetic binders to the target protein through a single enrichment step since it cannot be genetically amplified. Thus, discovery efforts with AS-MS have generally been limited to small combinatorial libraries

(10^3–10^6 members), which were biased toward the target protein in a 'focused' or structure-based manner.[38,39] AS-MS of these focused libraries remains a reliable way to rapidly identify key binding 'hot-spot' residues and combinatorically introduce noncanonical amino acids.[38,40,41] Recent advancements made by our group in the MS/MS sequencing of complex peptidomimetic mixtures[42] and optimized AS-MS selection conditions[43] have enabled *de novo* discovery of high affinity binders from fully randomized peptidomimetic libraries up to 10^8 members.[43] Thus, AS-MS can enable rapid discovery across highly diverse libraries.[44,45] With these methods, we set out to perform *de novo* discovery with synthetic highly noncanonical peptidomimetic libraries against recently identified clinically relevant targets.

Angiotensin converting enzyme 2 (ACE2) has been identified as an important plasma biomarker for cardiovascular disease-induced events and death in a global, population-based study.[46,47] Also, ACE2 is ubiquitously known as the receptor utilized for cell entry by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) beta-coronavirus.[48,49] Thus, high affinity reagents for the specific serum detection of ACE2 are increasingly important. Here we demonstrate rapid discovery of high affinity peptidomimetic binders to ACE2 through a single-pass AS-MS experiment utilizing fully randomized canonical and noncanonical libraries. By comparing selection results from the noncanonical library over a 'standard' canonical library quantitatively and qualitatively, we highlight that the noncanonical binders exhibit improved proteolytic serum stability. In further tests, our noncanonical peptidomimetic ACE2 binder, ABP N1, demonstrated ACE2 binding specificity in a serum pulldown experiment and as low as

picomolar detection in an enzyme-linked immunosorbent assay (ELISA), highlighting the development as promising diagnostic tools.

A2.2 Results

A2.2.1 AS-MS identifies low-nanomolar affinity canonical and noncanonical peptide binders to ACE2

We recently optimized in-solution affinity selection combined with nano-liquid chromatography-tandem mass spectrometry (nLC-MS/MS) sequencing to enable the identification of high affinity binders from fully randomized synthetic libraries with a diversity of up to 10^8 members, a 100-fold increase over the standard practice.[43] Noncanonical amino acids can be extensively used in the preparation of the synthetic libraries utilized in AS-MS, provided there is no isobaric monomer mass overlap and sufficient tandem sequencing fidelity. Thus, we sought to compare the results of our selections against human ACE2 protein using a standard, canonical-L library, and a noncanonical-L library, each containing 200 million members (Fig. A2.1a).

The two libraries share the same design, 12 variable positions followed by an amidated C-terminal lysine to facilitate sequence identification and filtering.[42] The monomer set utilized in the canonical library (Library 1) is fully proteogenic, except Cys because it could form intra- or intermolecular disulfides and Ile because it is isobaric in mass with Leu, to give 18 amino acids total (Fig. A2.1b). For side-by-side comparison and exploration of noncanonical chemical space, another library (Library 2), composing 17 non-proteogenic monomers out of 21 amino acids total (81% noncanonical, significantly more than the available genetically encoded techniques) was synthesized (Fig. A2.1 b, c).[31-34] Library 2 was designed to sample an entirely new chemical space,

while still capturing similar chemical properties of the natural amino acid set. While nearly all the canonical monomers were replaced with non-proteogenic analogs, the number of positively and negatively charged residues at physiological pH was kept constant at 2 each, respectively. Nonstandard backbones including β-amino acids and achiral linkers were included to diversify structural availability and improve proteolytic stability. The final monomers were selected based on considerations of balanced chemical diversity, mass uniqueness, library solubility, stereochemical purity, and compatibility with Fmoc solid-phase peptide synthesis (SPPS).

Following the workflow depicted in Fig. A2.1a, we performed the affinity selection using both the canonical and noncanonical library against human ACE2 in parallel with 12ca5 as an unrelated control protein, and the enriched peptides were eluted and analyzed by nLC-MS/MS. The sequences were decoded using the PEAKS software[50] and filtered[42] to isolate peptidomimetic sequences that matched the original library design. The peptidomimetics were sorted to reveal those that were unique to ACE2 in comparison to the off-target control protein and are reported in Supplementary Tables A2.2 and A2.3. From both the canonical and noncanonical selections, we observed a preferred N-terminal motif, indicating a potential new class of ACE2 binders (see Figures A2.6 and A2.7). For the canonical L-peptides, leucine (L) and valine (V) were preferred at the N-terminus position followed by glutamine (Q) and asparagine (N) with some cationic or ionizable residues nearby (H, K, and R, see Fig. A2.6). Similarly, the noncanonical L-peptidomimetics discovered preferred cyclopropylalanine (Cpa, C) and isoleucine (Ile, I) at the N-terminal position followed by 3-(4-thiazolyl)-alanine (Tha,

T, see Fig. A2.7). The observation of a preferred motif added confidence in the discovery of a class of peptidomimetic binders to ACE2.

An additional refinement step was taken to analyze the extracted ion chromatogram (EIC) of the observed ion from each MS-identified peptide. This EIC-based refinement provides an additional level of confirmation that each peptide was uniquely selected against ACE2 or if it binds nonspecifically to the selection matrix (e.g., magnetic beads or streptavidin). A comparable number of sequences was identified as EIC-selective between the two libraries (48 noncanonical sequences and 60 canonical sequences), indicating that sequencing and identification of noncanonical binders can robustly be achieved with a similar throughput to canonical peptides on our AS-MS platform using Orbitrap nLC-MS/MS. However, the noncanonical selection showed a higher primary MS baseline that obscured more peptidomimetics from being discovered (Table A2.3; Fig. A2.22). This ambiguity could be explained by a higher level of baseline binding from the noncanonical peptidomimetics library or poorer ionization of the eluted library members. Lastly, both selections showed a low rate of nonspecific binder recovery (Fig. A2.23).

With the discovered sequences in hand, nine canonical and five noncanonical binders were chosen for synthesis and validation efforts based on their high average local confidence (ALC) scores. Each binder was re-synthesized and purified individually and their binding affinities (dissociation constant, K_D) to ACE2 were measured using bio-layer interferometry (BLI). As a result, low-nanomolar ACE2-binding affinities were determined for the noncanonical binders ABP N1 (K_D = 19 nM, Fig. A2.2a), ABP N4 (K_D = 123 nM, Fig. A2.2b), and ABP N6 (K_D = 33 nM, Fig. A2.2c), and for the canonical

binders ABP C3 (K_D = 35 nM, Fig. A2.2d), ABP C7 (K_D = 26 nM, Fig. A2.2e), and ABP C8 (K_D = 36 nM, Fig. A2.2f). All other peptides either presented lower affinity or no binding to ACE2 (Fig. A2.2g), indicating that approximately half of the ACE2-selective sequenced binders following AS-MS were non-binders by BLI.

To determine whether the observed binding is sequence-specific, two scrambled variants of ABP N1 (Scrm N1.1 and Scrm N1.2) were synthesized and their binding to ACE2 was tested by BLI. Under these conditions, no binding to ACE2 was observed (Fig. A2.2g; Fig. A2.26). We also observed minimal binding to an unrelated protein (12ca5) for ABP N1, ABP N4, and ABP N6 (Fig. A2.27). However, the binding of 500 nM of ACE2 to ABP N1 results in ~1.3 nm BLI response signal, compared to 0.25 nm signal from 500 nM of the off-target protein (12ca5), translating to ~ 5-fold higher signal and indicating selectivity toward ACE2.

Additionally, we investigated whether known binders to ACE2 inhibited the binding of the ABPs, including the native ACE2 substrate angiotensin 2 (AngII),[51] an ACE2 inhibitor MLN-4760,[52,53] and the receptor binding domain (RBD) of SARS-CoV-2.[48,49] First, we performed competitive binding experiments on all binders discovered (ABP N1, N4, N6, C1, C2, C3, C4, C7, and C8) versus AngII and MLN-4760 (Figs. A2.32-40). We observed that the native ACE2 substrate AngII did not inhibit the binding of any of our ABPs, indicating that the binding sites of ABPs and AngII likely do not overlap. However, MLN-4760 was able to partially block the binding of ABP N1 and C8 at the 10-fold excess of MLN-4760 mixed with ACE2, indicating that the binding sites likely overlap with the active site of the ACE2 enzyme. Secondly, we performed a competitive BLI experiment to determine if any of the discovered binders could disrupt

the interaction between ACE2 and SARS-CoV-2 spike protein receptor binding domain (RBD). We did not observe any competition by our peptides on the binding of ACE2 to RBD (Figs. A2.28-31) at 5- or 50-fold excess, indicating these ABPs do not bind at the RBD site of interaction.

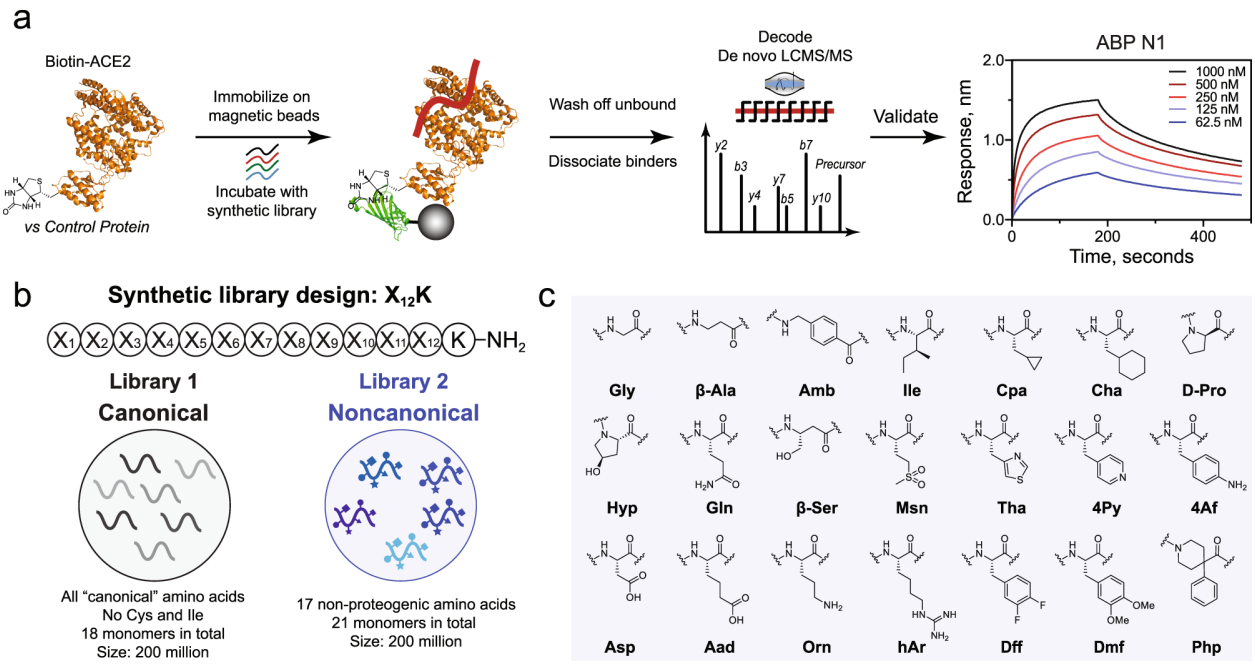
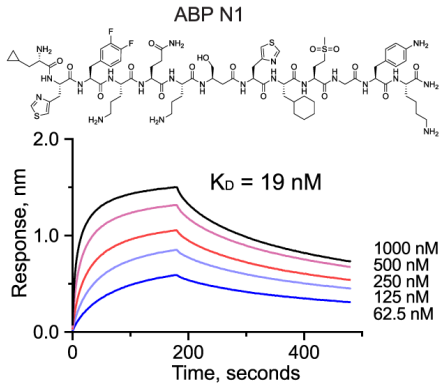
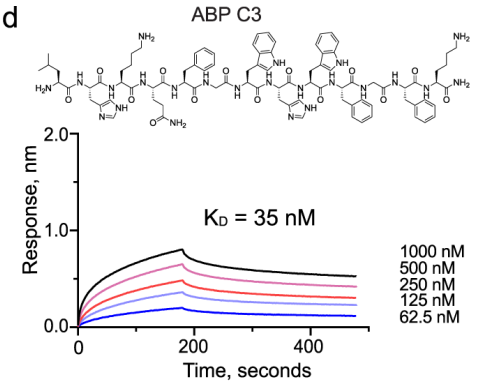


Figure A2.1. **Magnetic bead-based affinity selection-mass spectrometry (AS-MS) enables rapid discovery of both canonical and noncanonical binders in a single experiment.** A) Schematic representation of the AS-MS workflow used in this study. In brief, the biotinylated protein (ACE2 represented in orange from PDB: 6M17 or control) was immobilized onto streptavidin (SA)-coated magnetic beads and then sampled in synthetic peptide libraries to enrich peptide binders. Subsequently, unbound peptides were washed away, and bound material was eluted and then sequenced by nLC-MS/MS. Individual hits were synthesized and validated at the final step. B) The design for synthetic canonical and noncanonical libraries. C) The monomer set used for the synthesis of the noncanonical library (Library 2).

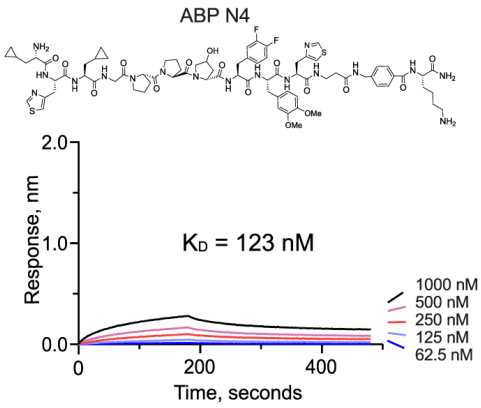
a



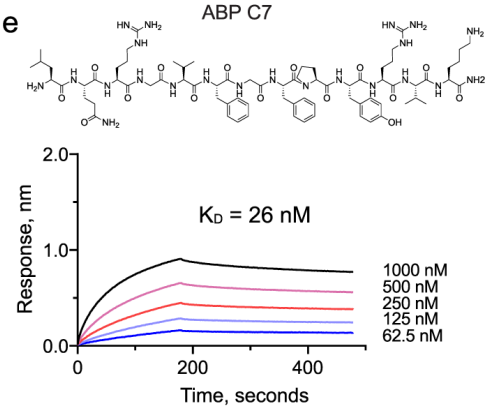
d



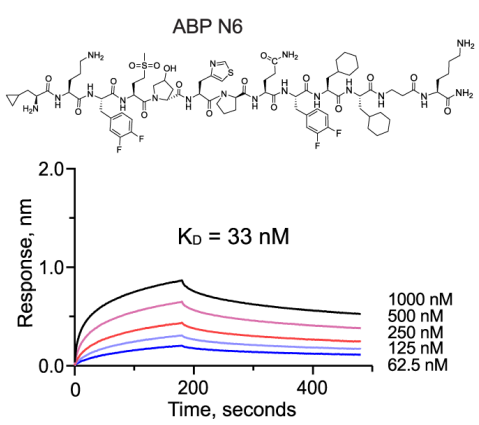
b



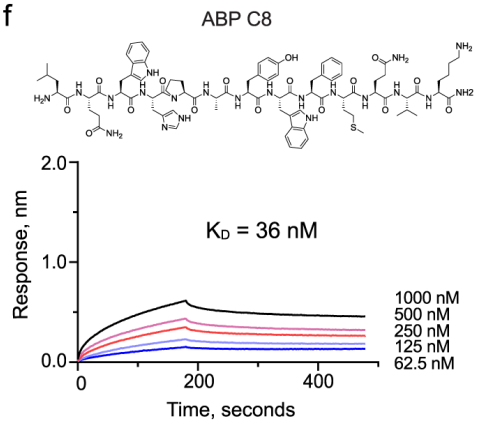
e



d



f



g

ID	Sequence, $X_{12}K$													ALC	Error, ppm	$K_{D,obs}$, nM
ABP C1	Leu	Phe	Arg	Gly	Tyr	Phe	Ala	Ala	Met	Asn	Asn	Leu	Lys	99	-1.0	66 ± 1.3
ABP C2	Leu	His	Phe	Ala	Lys	Trp	Asn	His	Val	Trp	Ser	Trp	Lys	94	-1.7	57 ± 0.8
ABP C3	Leu	His	Lys	Gln	Phe	Gly	Trp	His	Trp	Phe	Gly	Phe	Lys	94	-0.8	35 ± 0.7
ABP C4	Leu	Lys	Val	Lys	Tyr	Met	Trp	Asp	Tyr	Leu	Phe	Gly	Lys	98	-2.2	58 ± 1.0
ABP C5	Leu	Gln	Ala	Lys	Pro	Val	Pro	Gln	Phe	Trp	Pro	Phe	Lys	93	-1.8	*NB*
ABP C6	Leu	Gln	Lys	Phe	Leu	Gly	Glu	Arg	Phe	Pro	Gly	Trp	Lys	96	2.0	*NB*
ABP C7	Leu	Gln	Arg	Gly	Val	Phe	Gly	Phe	Pro	Tyr	Arg	Val	Lys	95	-1.4	**26 ± 0.4**
ABP C8	Leu	Gln	Trp	His	Pro	Ala	Tyr	Trp	Phe	Met	Gln	Val	Lys	99	-1.3	36 ± 0.8
ABP C9	Leu	Gln	Trp	Met	Asn	Lys	Tyr	Thr	Asn	Tyr	Gly	Leu	Lys	98	-2.8	316 ± 11
ABP N1	Cpa	Tha	Dff	Orn	Gln	Orn	βSer	Tha	Cha	Msn	Gly	4Af	Lys	99	-0.3	**19 ± 0.3**
Scrm N1.1	Tha	Cpa	βSer	Msn	Cha	Tha	Orn	Dff	Orn	Gln	Gly	4Af	Lys	*N/A*	*N/A*	*NB*
Scrm N1.2	Msn	Orn	Cha	Dff	Dff	βSer	Gly	Tha	Tha	Cpa	Gln	Orn	Lys	*N/A*	*N/A*	*NB*
ABP N4	Cpa	Tha	Cpa	Gly	DPro	DPro	Hyp	Dff	Dmf	Tha	β-Ala	Amb	Lys	99	-2.6	123 ± 2.0
ABP N5	Cpa	Tha	Dff	Gly	Gln	DPro	Msn	Cha	Gln	Tha	β-Ala	Tha	Lys	99	-2.1	*NB*
ABP N6	Cpa	Orn	Dff	Msn	Hyp	Tha	DPro	Gln	Dff	Cha	Cha	β-Ala	Lys	99	-0.8	33 ± 0.6
ABP N8	Cpa	Tha	hAr	DPro	Gly	β-Ala	Php	Asp	Cha	Msn	Dff	Msn	Lys	98	1.1	*NB*

Figure A2.2. **Nanomolar affinity binders were identified from both canonical and noncanonical libraries.** A–C) ACE2-binding traces, measured by bio-layer interferometry, for noncanonical binders ABP N1, ABP N4, and ABP N6, respectively. D-F) ACE2-binding traces, measured by bio-layer interferometry, for canonical binders ABP C3, ABP C7, and ABP C8, respectively. G) A summary of all individually synthesized peptides. Column headers: 'ID', the peptide identifiers; 'Sequence, $X_{12}K$', the peptide sequences with a lysine at the C-terminus; 'ALC', the exported average local confidence score from sequence decoding; 'Error, ppm', the mass error (in ppm) between the precursor and assigned sequence; '$K_{D,\ obs}$, nM', the apparent dissociation constant, in nM, measured by bio-layer interferometry. Cyan highlights the canonical (ABP C7) and noncanonical (ABP N1) peptides with the highest binding affinity.

A2.2.2 The noncanonical binder ABP N1 demonstrates enhanced serum stability

A significant limitation to the development of canonical L-peptides is their susceptibility to enzymatic degradation within physiological environments. Thus, the use of noncanonical amino acids, macrocyclization, and chemical stapling have each grown as approaches to improve their stability and pharmacokinetics.[16,21,24-27] To determine whether our noncanonical binders present improved stability over the canonical binders, we performed a serum stability assay on the two most potent variants: noncanonical ABP N1 and canonical ABP C7. After incubating both peptides in 5% normal human serum at 37 °C over time, serum proteins were precipitated by addition of trichloroacetic acid, and the binders retained in the supernatant were subjected to LC-MS analysis. As shown from the LC-MS traces, we observed significant degradation of ABP C7 over 12 hours when compared with the no-serum control (Fig. A2.3a), with an approximate half-life of 2 h (Fig. A2.3c). However, little-to-no degradation of ABP N1 was observed (Fig. A2.3b,c), even after 12 h of incubation. ABP N1 presented enhanced stability in human serum, and the side-by-side comparison with a canonical binder from a similar library selection demonstrated the immediate benefit of employing noncanonical

libraries. Thus, we could rapidly proceed to determine the binding capability of ABP N1 in biological milieu.

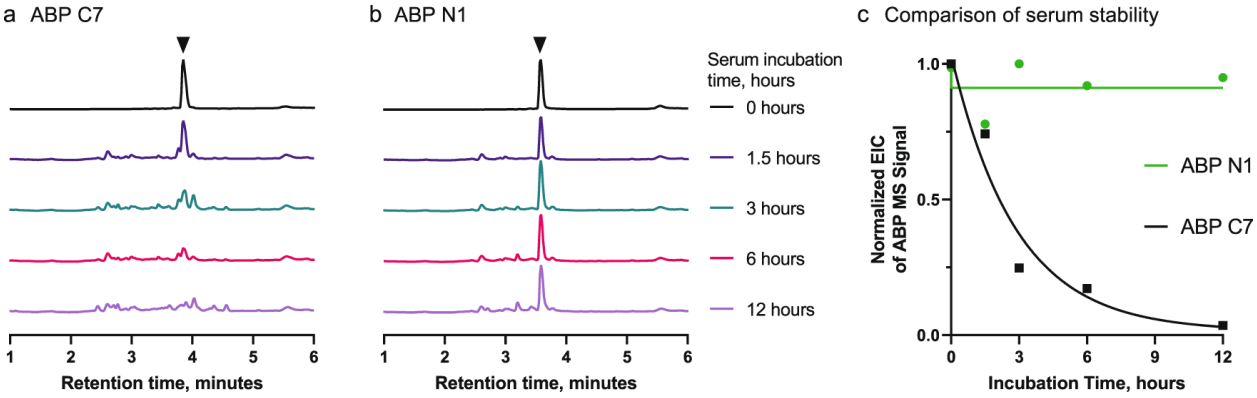


***Figure A2.3*. ABP N1 demonstrates enhanced serum stability relative to the canonical binder ABP C7.** LC-MS chromatograms (total ion current) of ABP N1 (a) and ABP C7 (b) incubated at 37 °C in 1 × PBS ($t = 0$ h) or 5% normal human serum. Spectra were normalized to the intensity of the peptide at $t = 0$ h to observe degradation of the original amount of peptide. c Comparison of the serum stability of ABP N1 and ABP C7 by the integral of the extracted ion count (EIC) of the monoisotopic $[M + 2H]^{2+}$ ion of the starting peptide. ABP C7 is quickly degraded and the starting peptide mass disappears over 12 h, with a calculated half-life of approximately 2 h (single phase decay model).

A2.2.3 ABP N1 pulls down ACE2 from human serum selectively

To further investigate the ACE2-binding capability of the identified high affinity reagents within a biological matrix and demonstrate binding selectivity, we performed a human serum pull-down experiment using the most potent noncanonical binder ABP N1. As depicted in Fig. A2.4a, the biotinylated ABP N1 was immobilized onto streptavidin-coated magnetic beads followed by incubation with a mixture of normal human serum and ACE2 protein. After removing the supernatant and washing off the unbound fraction, the bound material was eluted with a high concentration of urea and analyzed using the sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE). We observed a selective enrichment of ACE2 protein from the human serum

complex, and no other proteins were pulled down by ABP N1 (Fig. AA2.4b). The ability of selective binding and isolation of ACE2 from a complex biological matrix indicates a promising diagnostic application of ABP N1.

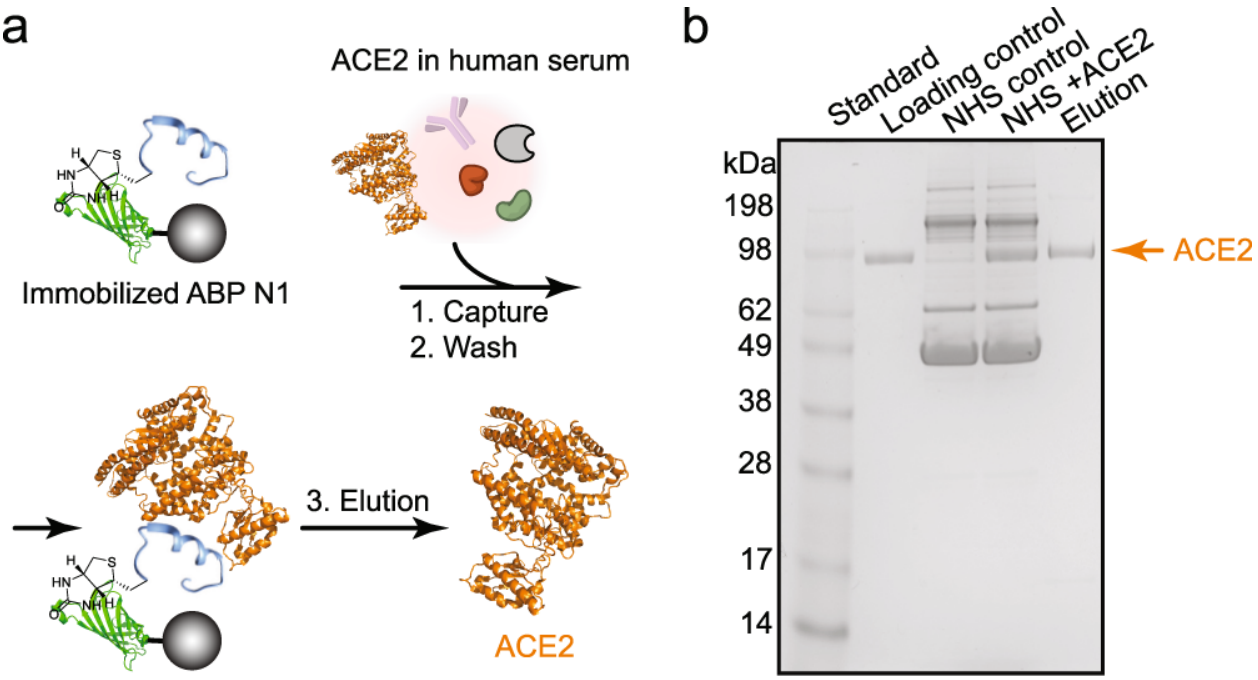


***Figure A2.4*. ABP N1 pulls down ACE2 from human serum selectively.** A) Schematic representation of the serum pull-down experiment. ABP N1 was immobilized onto streptavidin-coated magnetic beads and then incubated with ACE2 (represented in orange from PDB: 6M17) in human serum. Bound ACE2 was eluted for subsequent analysis. B) The SDS-PAGE image with samples showing from left to right lanes: (1) molecular weight standard; (2) purified ACE2 protein (1.5 μg) as a loading control; (3) normal human serum as a control; (4) normal human serum mixed with ACE2 (1.5 μg); (5) elution of the bound fraction from the magnetic beads. NHS, Normal Human Serum (5%).

A2.2.4 Picomolar ACE2 can be detected with ABP N1 via ELISA

Human ACE2 was recently identified as the top biomarker for cardiovascular disease and an elevated level of plasma ACE2 significantly associates with death, heart failure, stroke, and myocardial infarction.[47] To demonstrate the utility of our noncanonical ACE2 binders as detection probes of the plasma ACE2 level, we

developed an ELISA-based detection assay (Fig. A2.5a). After immobilization on an ELISA plate, the picomolar concentration of ACE2 was detected by biotinylated ABP N1 in a dose-dependent manner (Fig. A2.5b). As a negative control, no binding was observed from biotinylated ABP N8 at the tested ACE2 concentrations (10 pM–100 nM in Fig. A2.5b), which is consistent with the non-binding observation as demonstrated in Fig. A2.2g. Interestingly, when a similar ELISA experiment was performed with the plate being coated with a mixture of normal human serum and exogenous ACE2, the dose-dependent ACE2 detection using ABP N1 was retained (Fig. A2.5c). However, we observed a decreased detection sensitivity presumably due to the decreased amount of ACE2 being immobilized on the plate with many other serum proteins present. As a summary, with an ELISA-based approach, we demonstrated that ABP N1 could selectively detect soluble ACE2 from the human serum at concentrations ranging from picomolar to low nanomolar levels.

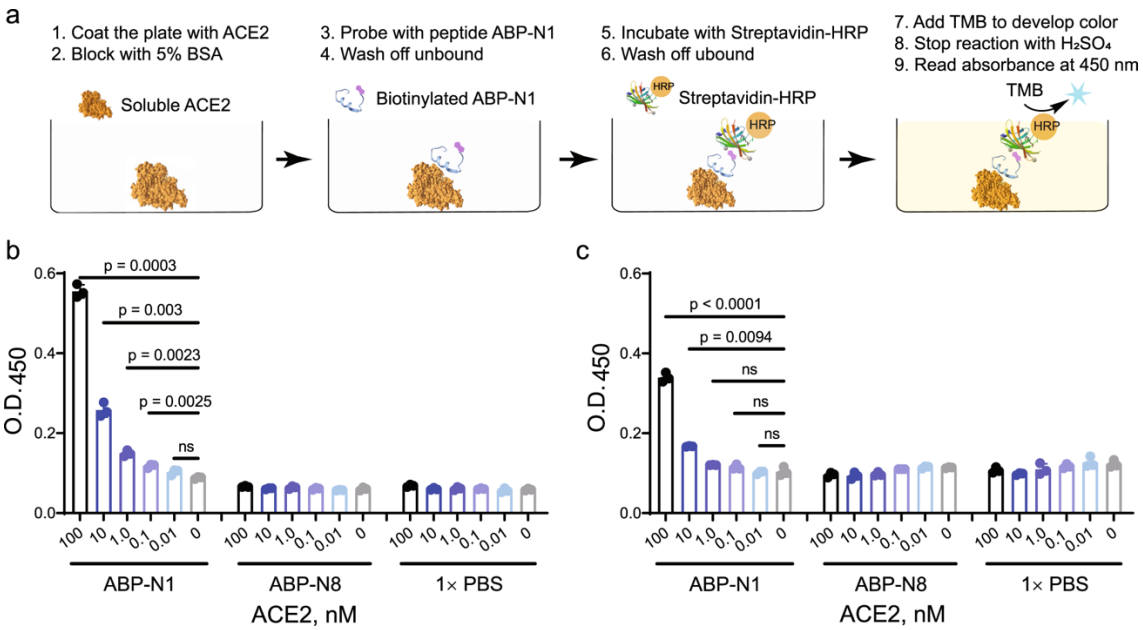


Figure A2.5. **Picomolar ACE2 concentration was detected by ABP N1 by ELISA.** A) Schematic representation of the ELISA workflow. B) ACE2 in 1 × PBS at different concentrations (100 nM–10 pM) was immobilized on an ELISA plate. C) ACE2 in human

serum at different concentrations (100 nM–10 pM) was immobilized on an ELISA plate. The plate was incubated with ABP N1 or control peptide, followed by streptavidin-HRP and TMB substrate. Absorbance was measured at 450 nm. The experiment was performed in technical triplicates ($n = 3$). Each data point, the mean signal as a bar, and error bars from experimental standard deviation, and statistical significance calculated with the unpaired Student's *t*-test are shown.

A2.3 Discussion

In this study, we report rapid discovery of high affinity peptidomimetic binders to ACE2 via AS-MS. After screening two ultra-large synthetic libraries, a series of ACE2-binding peptides with low-nanomolar affinities (K_D from 19 to 123 nM) were identified. We showed that the most potent noncanonical peptide binder, ABP N1, demonstrated enhanced serum stability in comparison with the most potent canonical binder, ABP C7. Furthermore, ABP N1 demonstrated a high selectivity to ACE2 over human serum proteins, indicated by a serum pull-down experiment. Lastly, in an ELISA-based format, picomolar to low nanomolar ACE2 concentrations in human serum could be detected using ABP N1. The *de novo* discovered high affinity reagents offer specific serum detection of ACE2 and represent a promising diagnostic tool of related diseases.

We chemically synthesized both canonical and noncanonical peptide libraries (each containing 200 million peptides) and subjected them to a single-pass affinity selection experiment as a side-by-side comparison. Of note, a comparable number of putative binders was observed from both canonical and noncanonical library selections, with a similar portion of binders measured by BLI to show no apparent binding to ACE2. Therefore, even though the binders were identified as ACE2-selective with high sequencing confidence, individual synthesis and validation of the putative hits are necessary to confirm ACE2-specific interaction. We observed an N-terminal motif from

both the canonical and noncanonical selections, indicating a potentially preferred binder class to ACE2 (Fig. A2.6 and A2.7). However, after individual validation, binders containing such a motif do not warrant ACE2-binding capability (Fig. A2.2g), indicating that the other amino acids within the sequence also play an important role in binding. Notably, all of the discovered binders are linear, whereas previous phage display efforts to discover linear ACE2 binding peptides were unsuccessful and only discovered macrocyclic binders.[54]

Even though the canonical or noncanonical binders discovered here are linear, they do not have a clear similarity to other known binders of ACE2, including the substrate angiotensin 2 (AngII, Sequence: DRVYIHPF) or the ACE2 inhibitor MLN-4760. The design of both libraries is similar in size to AngII. Yet, the lack of peptidomimetics discovered that are similar to AngII could be due to the libraries' carboxamide C-terminus interfering with the substrate binding site or because of ACE2 enzymatic activity cleaving peptides closely bound to the AngII active site. Cleaved peptides would be filtered out during analysis.[42] Moreover, ACE2 receptor blockers (ARBs) also do not appear similar to the discovered binders here, though peptidomimetic ARBs have been a source of inspiration for their small-molecule design.[53] With the use of the ARB MLN-4760, competitive BLI binding experiments were able to identify that ABP N1 and C8 are partially inhibited by MLN-4760, suggesting N1 and C8 bind near the active site and could be further developed.

Natively, ACE2 is a counter-regulator of the renin-angiotensin hormonal cascade. However, elevated levels of ACE2 in plasma have been found to be the highest-ranked predictor of death, heart failure, stroke, and myocardial infarction compared to several

established risk factors.[47,55-59] While this relationship is poorly understood, ACE2 plasma concentrations are a measurable indicator of renin-angiotensin system dysregulation.[47] As the target of our work here, we sought to design serum-stable, high affinity peptidomimetic ACE2 binders as the response to such a demand.

This work highlights that optimized AS-MS methods[43] can be used for the discovery of high affinity peptidomimetics with enhanced stability from a single rapid experiment. It is challenging to use existing approaches including phage-display, one-bead-one-compound (OBOC), or mRNA-display, to achieve selections at high chemical diversity and speed simultaneously. Multiple putative ACE2 binders were discovered from a single affinity selection experiment and several of the top hits were validated to bind to ACE2 with low-nanomolar affinities. Moreover, the immediate discovery of serum-stable noncanonical peptidomimetics like ABP N1 demonstrates the potential toward acceleration of the development timeline that our AS-MS platform can provide. We envision that the reported ACE2 binders in this study may be further developed as reliable and sensitive diagnostics to detect plasma ACE2 concentrations, as well as peptide conjugates for tissue-targeting or directed delivery of therapeutics.

A2.4 Materials and Methods

A2.4.1 Materials

H-Rink Amide-ChemMatrix resin was obtained from PCAS BioMatrix. Tentagel M NH_2 resin was obtained from Rapp Polymere. 4-[(*R*,*S*)-α-[1-(9*H*-Fluoren-9-yl)-methoxyformamido]-2,4-dimethoxybenzyl]-phenoxyacetic acid (Fmoc-Rink amide linker) was obtained from Chem-Impex International. Biotin-PEG4-propionic acid and Biotin-

PEG4-NHS were both purchased from ChemPep Inc. *N*,*N*-Diisopropylethylamine (DIEA) was obtained from a Seca Solvent Purification System by Pure Process Technology. Peptide synthesis-grade *N*,*N*-dimethylformamide (DMF), dichloromethane (DCM), diethyl ether, and HPLC-grade acetonitrile were obtained from VWR International. 1-[Bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-b]pyridinium-3-oxid-hexafluorophosphate (HATU) was purchased from P3 BioSystems. Trifluoroacetic acid (TFA; for HPLC, ≥ 99%), piperidine (ReagentPlus; 99%), triisopropylsilane (98%), 1,2-ethanedithiol (≥ 98%), phenylsilane (97%) and tetrakis(triphenylphosphine)palladium(0) (99%), were purchased from MilliporeSigma. Biotechnology grade bovine serum albumin (BSA) was obtained from VWR International. Ultrapure water was obtained by filtering deionized water with a Milli-Q water purification system (Millipore). Canonical amino acid monomers Fmoc-Ala-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Asp(*t*Bu)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Glu(*t*Bu)-OH, Fmoc-Gly-OH, Fmoc-His(Trt)-OH, Fmoc-Leu-OH, Fmoc-Lys(Boc)-OH, Fmoc-Met-OH, Fmoc-Phe-OH, Fmoc-Pro-OH, Fmoc-Ser(*t*Bu)-OH, Fmoc-Thr(*t*Bu)-OH, Fmoc-Trp(Boc)-OH, Fmoc-Tyr(*t*Bu)-OH, and Fmoc-Val-OH were purchased from Advanced ChemTech (Louisville, KY). Noncanonical monomers Fmoc-3,4-difluoro-L-phenylalanine, Fmoc-β-cyclopropyl-L-alanine, Fmoc-D-4-thiazolyl-alanine, Fmoc-(4-aminomethyl) benzoic acid, Fmoc-β-cyclohexyl-L-alanine, Fmoc-L-α-aminoadipic acid δ-*tert*-butyl ester, Fmoc-D-proline, Fmoc-L-methionine sulfone, Fmoc-β-alanine-OH, Fmoc-4-phenylpiperidine-4-carboxylic acid, Fmoc-L-ornithine(Boc), Fmoc-4-(Boc-amino)-L-phenylalanine, Fmoc-3-(4′-pyridyl)-L-alanine, Fmoc-L-homoarginine(Pbf)-OH,

Fmoc-hydroxyproline(*t*Bu)-OH, Fmoc-3,4-dimethoxy-L-phenylalanine, and Fmoc- β-Serine(*t*Bu)-OH were purchased from Chem-Impex International.

A2.4.2 Automated fast-flow synthesis of canonical peptides

Canonical peptide binders were synthesized on an automated fast-flow synthesizer[60,61] on a scale of 0.075 mmol using H-Rink Amide-ChemMatrix resin (loading capacity 0.49 mmol/g). The reactor temperature was set at 90 °C and the monomer activation was achieved by HATU. Amide bond formation was effected in 8 s, and removal of the Fmoc groups was performed in 8 s with 40% (v/v) piperidine and 2% (v/v) formic acid in DMF. Overall cycle times were about 90 s. After completion of synthesis, resins were retrieved from the synthesizer and washed with DCM (5 × 10 mL), and dried in a desiccator under reduced pressure.

A2.4.3 Manual solid-phase synthesis of noncanonical peptides

Noncanonical amino-acid-containing peptides were synthesized manually in batch by standard solid-phase protocols. The synthesis was carried out at a 0.05 mmol scale on H-Rink Amide-ChemMatrix resin within Torviq syringes (size 6 mL). In brief, resins were swelled with DMF for 10–15 min before the peptide sequence assembly. Coupling steps were performed at room temperature for 30 min with Fmoc-protected amino acids (0.5 mmol, 10 equivalents) dissolved in 1.25 mL of HATU (0.38 M solution in DMF, 9.5 equivalents) and 250 µL of DIEA for activation (3 equivalents). The resin was stirred multiple times during coupling and then washed (5 × 10 mL) with DMF before deprotection with 20% (v/v) piperidine in DMF (2 × 10 mL with 5 min each time). As a

final step, the resin was washed again (5 × 10 mL) with DMF to remove piperidine residuals and finish the cycle. After completing all monomer couplings, the resin was washed sequentially with DMF (5 × 10 mL) and DCM (5 × 10 mL) and then dried under reduced pressure before peptide cleavage.

A2.4.4 Peptide cleavage

Peptide cleavages from the resin and all side-chain deprotections were performed simultaneously with 2.5% (v/v) 1,2-ethanedithiol (EDT), 2.5% (v/v) water, and 1% (v/v) triisopropylsilane in neat TFA for 2.0 h at room temperature. Five mL of cleavage cocktail was used for 0.1 mmol peptides. The resulting cleaved solution was washed with dry ice-cold diethyl ether and followed by centrifugation at 4000 rpm for 3 min to precipitate the crude peptides. The obtained solids were briefly dried and dissolved in water/acetonitrile (50:50, v/v) before freezing and lyophilization.

A2.4.5 Peptide purification

Peptides with crude purity >85% were purified using a Biotage Selekt© flash purification system and peptides with crude purity <85% were purified using reverse-phase high-performance liquid chromatography (HPLC). 1) *Reverse-Phase HPLC*. Crude peptides with purity below 85% were dissolved in water with 0.1% TFA and then purified by semipreparative reverse-phase HPLC using an Agilent 1260 HPLC system (Agilent Zorbax SB-C3 column: 9.4 × 250 mm, 5 µm; or Agilent Zorbax SB-C18 column: 9.4 × 250 mm, 5 µm). The gradient used was 10 to 61% acetonitrile over 60 min with a 4.0 mL/min flow rate. HPLC fractions were analyzed by LC-MS and pure fractions were

combined and lyophilized. 2) *Flash chromatography purification*. Synthetic peptides with the estimated LC-MS crude purity above 85% were purified with a Biotage Selekt© instrument using Biotage Sfär C18 D column (Duo 100 Å 30 µm, 12 g) under a 5–60% acetonitrile gradient over 12 column volumes, and the flow rate was set to 12 mL/min. Fractions were collected based on absorbance at 214 nm and then subjected to LC-MS analysis before combining the pure fractions and lyophilization.

A2.4.6 Analytical high-performance liquid chromatography (HPLC)

Analytical HPLC was performed on an Agilent 1200 series system with UV detection at 280 nm. The column used was Phenomenex Kinetex (100 × 2.1 mm, 2.6 µm, 100 Å silica) with a flow rate of 0.375 mL/min. Solvents used are acetonitrile with 0.08% TFA additive (solvent B) and water with 0.1% TFA additive (solvent A). A linear gradient: 2 min hold 2% B, 2–32% B gradient from 2 to 17 min, 32–65% B gradient from 17 to 17.5 min, hold 65% B from 17.5 to 19 min. A final 6 min hold was performed with 2% B. The total method time was 25 min. HPLC purities were determined by manual integration of all signals in the area of 2–13 min.

A2.4.7 LC-MS analysis

Crude synthetic peptides or peptide fractions from the purification steps were analyzed by LC-MS (Agilent 6545 or 6550 ESI Q-TOF) using an Agilent Zorbax 300SB-C3 or Phenomenex Luna C18 column. Total ion chromatograms and integrated MS over the main peak are provided in the supplementary information.

Condition 1

Analysis was performed on an Agilent 1290 Infinity HPLC coupled to an Agilent 6545 ESI Q-TOF mass spectrometer. MS was run in positive ionization mode, extended dynamic range (2 GHz), and standard mass range (m/z in the range of 300–3000 a.m.u.). The solvent mixtures used for LC-MS chromatography were: *A* = water + 0.1% formic acid (LC-MS grade), B = acetonitrile + 0.1% formic acid (LC-MS grade). The following conditions were used for peptide analysis. Column: Zorbax 300-SB C3 (5 µm, 150 × 2.1 mm, 300 Å silica); Flow Rate: 0.8 mL/min; Gradient: 1% B 0–2 min, linearly ramp from 1% B to 61% B 2 to 11 min, 61% B to 95% B 11 to 12 min. Post time is 1% B for 3 min. MS data were acquired from 4 to 11 min.

Condition 2

Analysis was performed on an Agilent 1290 Infinity HPLC coupled to an Agilent 6550 Q-TOF with Dual Jet Stream ESI ionization and iFunnel. MS was run in positive ionization mode, extended dynamic range (2 GHz), and low mass range (m/z in the range of 100–1700 a.m.u.). The solvent mixtures used were as above. Column: Phenomenex Luna C18 (3 µm, 150 × 1 mm, 100 Å silica); Flow Rate: 0.5 mL/min; Gradient: 1% B 0–2 min, linearly ramp from 1% B to 61% B 2 to 14 min, 61% B. Post time is 1% B for 3 min. MS data were acquired from 4 to 14 min.

A2.4.8 Split-and-pool synthesis of peptide libraries

Split-and-pool synthesis was carried out on a 30 µm TentaGel resin (0.26 mmol/g) for a 10^8-member library. Each coupling step was carried out as follows: solutions of Fmoc-protected amino acids, HATU (0.38 M in DMF; 0.9 eq. relative to an amino acid), and DIEA (1.1 eq. for histidine; 3 eq. for all other amino acids) were each

added to individual portions of resin. Couplings were allowed to proceed for 20 min, and resin portions were recombined and washed with DMF for deprotection. Fmoc removal step was carried out by treatment of the resin with 20% piperidine in DMF (1 × flow wash; 2 × 5 min batch treatments). The resin was washed again with DMF before splitting. For the canonical library preparation, splitting was performed by suspending the resin in DMF and evenly dividing among 18 plastic fritted syringes on a manifold (18 canonical amino acids except for isoleucine and cysteine). For the noncanonical library preparation, 21 plastic fritted syringes were used (4 canonical amino acids and 17 non-proteogenic amino acids). The cycle was then continued with the next coupling until the completion of all random positions.

A2.4.9 Affinity selection from both canonical and noncanonical libraries

HEK293 cell-expressed biotinylated ACE2 was obtained from AcroBiosystems (Catalog: AC2-H82E6). ACE2 protein was immobilized onto streptavidin-coated magnetic beads (Thermo Fisher, Catalog number 65001) in the presence of 10% FBS in 1 × PBS on a rotating mixer for 30 min at 4 °C. Subsequently, functionalized beads were washed with 1 × PBS (3 × 1 mL) and then incubated with peptide library (typically, the concentration was at 10–20 pM/member and the reaction volume was 1.0 mL 1 × PBS containing 10% FBS) on a rotating mixer for 1.0 h at 4 °C. The beads were then washed with PBS (3 × 1 mL) on a magnetic separation rack. Bound peptides were eluted with 200 µL of 6 M guanidine hydrochloride, 200 mM phosphate, pH 6.8. Elution volume was concentrated using C18 ZipTip® pipette tips. After lyophilization, powders were resuspended in water (0.1% formic acid) and submitted for nLC-MS/MS analysis.

A2.4.10 Nano LC-MS/MS (nLC-MS/MS) sequencing

Peptide sequence analysis was performed on an EASY-nLC 1200 (Thermo Fisher Scientific) nano-liquid chromatography handling system with an Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Thermo Fisher Scientific) as previously described[43]. Samples prepared from the library selection steps were run on a PepMap RSLC C18 column (2 µm particle size, 15 cm × 50 µm ID; Thermo Fisher Scientific, P/N ES801). A nanoViper Trap Column (C18, 3 µm particle size, 100 Å pore size, 20 mm × 75 µm ID; Thermo Fisher Scientific, P/N 164946) was used for desalting. The standard nano-LC method was run at 40 °C and a flow rate of 300 nL/min with the following gradient: 1% solvent B in solvent A ramping linearly to 61% B in A over 60 or 90 min, where solvent A = water (0.1% FA), and solvent B = 80% acetonitrile, 20% water (0.1% FA). Positive ion spray voltage was set to 2200 V. Orbitrap detection was used for primary MS, with the following parameters: resolution = 120,000; quadrupole isolation; scan range = 200–1400 m/z; RF lens = 30%; AGC target = 1×10^6; maximum injection time = 100 ms; 1 microscan. Secondary MS spectra acquisition was done in a data-dependent manner: dynamic exclusion was employed such that a precursor was excluded for 30 s if it was detected four or more times within 30 s (mass tolerance: 10.00 ppm); monoisotopic precursor selection used to select for peptides; intensity threshold was set to 5×10^4; charge states 2–10 were selected; and precursor selection range was set to 200–1400 m/z. The top 15 most intense precursors that met the preceding criteria were subjected to subsequent fragmentation. Three fragmentation modes collision-induced dissociation (CID), higher-energy collisional dissociation (HCD), and electron-

transfer/higher-energy collisional dissociation (EThcD) were used for the acquisition of secondary MS spectra. Only precursors with charge states 3 and above were subjected to all three fragmentation modes; precursors with charge states of 2 were subjected to CID and HCD only. For all three modes, the detection was performed in the Orbitrap (resolution = 30,000; quadrupole isolation; isolation window = 1.3 m/z; AGC target = 2×10^4; maximum injection time = 100 ms; 1 microscan). For CID and HCD, a collision energy of 30 and 25% was used, respectively. For EthcD, a supplemental activation collision energy of 25% was used.

A2.4.11 Bio-layer interferometry

Peptide binding validation was carried out using the ForteBio Octet RED96 system. The chamber temperature was kept constant at 30 °C with a plate agitation speed at 1000 rpm. Briefly, streptavidin (SA)-coated biosensor tips were dipped into 200 µL of 1.0–2.0 µM biotinylated peptide solution (in a kinetic buffer (K.B.): 1 × PBS with 0.1% BSA and 0.02% Tween-20) for peptides immobilization. Once loaded with peptides, the tips were then moved into solutions containing various concentrations (1000, 500, 250, 125, and 62.5 nM) of recombinant ACE2 protein (purchased from Sino Biological, Catalog number 10108-H08H) in the K.B. to obtain the association curve. After the 180 s association step, the tips were moved back into the K.B. to obtain the dissociation curve. Buffer-only and protein-only conditions (concentration at 1000 nM) were used as references for background subtraction. The association and dissociation curves were fitted with the ForteBio Biosystems Data Analysis Software under 5

experimental conditions (n = 5, global fitting algorithm, binding model 1:1) to calculate the apparent dissociation constant (K_D).

A2.4.12 Serum stability study

Following our previous protocol,[62] the normal human serum (NHS) (Sigma Catalog number H4522) was thawed, centrifuged at 14,000 rpm for 10 min and the supernatant was pre-warmed in a 37 °C water bath. 500 µL of tested peptides were placed in a 1.7 mL Eppendorf tube. Five hundred µL of either 10% NHS (diluted in 1 × PBS) or 1 × PBS for the control, was added into the tube (5% NHS after dilution), and immediately the tubes were vortexed. Assay tubes were placed at 37 °C water bath for 3.0 h, and 200 µL were transferred to a fresh microfuge tube with 40 µL of 15% trichloroacetic acid (TCA). Tubes were then placed on ice for 15 min and centrifuged at 14,000 rpm for 10 min. Supernatant from each tube was collected and subjected to LC-MS analysis.

A2.4.13 ACE2 pulldown from human serum

Following a previous protocol,[44] streptavidin-coated magnetic beads (100 µL, at 10 mg/mL) were pre-washed with blocking buffer (1 × PBS with 0.05% Tween-20, pH 7.4) and captured on a magnetic separation rack. The beads were then re-suspended in 1.0 mL of the same blocking buffer. Biotinylated ABP N1 peptide (25 µL, 0.1 mM) was then added to the beads. After 30 min of incubation at 4 °C, the supernatant was removed and the beads were washed with 1 × PBS (3 × 1.0 mL) before incubation with the protein mixture. Soluble ACE2 (5 µL, 0.62 mg/mL) was mixed with 50 µL normal

human serum in 1 × PBS (pH 7.4) to a final volume of 1.0 mL. The magnetic beads displaying immobilized ABP N1 were then added into the protein complex. After 1.0 h incubation at 4 °C with gentle rotation, the supernatant was removed and the beads were washed with 1 × PBS (3 × 1.0 mL) and captured on a magnetic separation rack. The captured protein was eluted with 20 µL of 6 M urea solution, followed by SDS-PAGE gel analysis. A precast polyacrylamide gel, Bolt™ Bis-Tris Plus gel (Invitrogen, Catalog number NW04120BOX), was used for an optimal separation under denaturing conditions with a running voltage of 165 V for 36 min. The SeeBlue™ Plus2 pre-stained protein standard (Thermo Fisher, Catalog number LC5925) was used as the molecular weight reference, and Bolt™ LDS Sample Buffer (Thermo Fisher, Catalog number B0007) was used to prepare and load the protein samples.

A2.4.14 Enzyme-linked immunosorbent assay (ELISA)

A serial dilution of soluble ACE2 was coated onto an ELISA plate (96-well format) overnight at 4 °C. The next day, the plate was blocked with 5% BSA in 1 × PBS supplement with 0.05% Tween-20 (PBST, pH 7.4) for 2–3 h at room temperature. After a brief wash, 100 µL solution of biotinylated peptide ABP N1 (100 nM in PBST) was added to the wells and incubated for 1.0 h at room temperature. Control groups are biotinylated peptide ABP N8 and PBST alone. The supernatant was then removed, and the wells were washed with PBST (3× 200 µL). Streptavidin-HRP conjugate in PBST (0.25 µg/mL, 100 µL) was added to each well and incubated for 30 min at room temperature. The plate was PBST-washed again (3 × 200 µL) before color development with 100 µL 1-Step™ Ultra TMB-ELISA Substrate Solution (Thermo Fisher, Catalog

number 34028). After 10 min, the reaction was quenched with 2.0 M sulfuric acid (100 µL). Finally, the absorbance at 450 nm was measured on a microplate reader (BioTek) for all treated wells. The same procedure was repeated when the ELISA plate was coated with a mixture of ACE2 and normal human serum. Three replicates were used throughout.

A2.3 Acknowledgements

Financial support for this work was provided by a COVID-19 Fast Grant award sponsored by Emergent Ventures at the Mercatus Center, George Mason University (to B.L.P.). A.E.C. gratefully acknowledges support from the National Science Foundation Graduate Research Fellowship under Grant No. 1122374, and additional support from an MIT Dean of Science Fellowship. J.S.B. acknowledges support from the Pharmaceutical Research and Manufacturers of America for the Postdoctoral Fellowship in Drug Discovery. We thank the Biophysical Instrumentation Facility at MIT for providing access to the Octet Red96 Bio-Layer Interferometry System (NIH S10OD016326).

A2.4 Author contributions

G.Z., J.S.B., A.J.Q., A.L., and B.L.P. designed research; G.Z., J.S.B., A.J.Q., C.L., X.T., S.H., S.A., and A.E.C. performed research with support from B.L.P.; G.Z., J.S.B., A.J.Q., and B.L.P. analyzed data; G.Z., J.S.B., and C.L. generated figures; G.Z. and J.S.B. wrote the manuscript with input from all authors.

A2.5 Appendix I: Library selection using AS-MS

A2.5.1 Summary of the discovered hits

Library	ACE2 outcome
Noncanonical L-library (**Library 2**) 2 x 10^8, $X_{12}K$	**From BLI:** 3 of 5 binders synthetized were true binders **From AS-MS:** 48 EIC selective binders, 6 obscured by MS baseline, 2 nonspecific binders
Canonical L-library (**Library 1**) 2 x 10^8, $X_{12}K$	**From BLI:** 7 of 9 binders synthetized were true binders **From AS-MS:** 60 EIC selective binders, 2 nonspecific binders

Table A2.1. **The summary of identified canonical and noncanonical library selection hits.**

Name	Sequence	ALC	m/z	z	RT, min	Calc Mass, Da	Error, ppm	EIC Specific?
ABP C1	L F R G Y F A A M N N L K	99	515.2775	3	55.7	1542.8130	-1.6	Yes
ABP C8	L Q W H P A Y W F M Q V K	99	578.2968	3	68.9	1731.8708	-1.3	Yes
	L V K G D Q Y G W V W G K	99	512.2767	3	49.7	1533.8093	-0.7	Yes
	L V K S L N S Y F F V F K	99	530.9725	3	67.4	1589.8970	-0.9	Yes
	V P K N F S W N L W R P K	99	418.4868	4	49.6	1669.9204	-1.3	Yes
	W T Y D L F S M N F G R K	99	555.2733	3	68.2	1662.7976	0.2	Yes
	Y N K G P F K N G H L F K	99	387.9659	4	24.6	1547.8362	-1.2	Yes
ABP C9	L Q W M N K Y T N Y G L K	98	553.2873	3	50.4	1656.8447	-2.8	Yes
	L W T V H W F Q T Y T P K	98	569.2995	3	63.0	1704.8777	-0.6	Yes
	N P W W P L P Y H G M Y K	98	563.2769	3	64.1	1686.8130	-2.4	Yes
	Y N V G L Y T S M E K Y K	98	532.2693	3	46.5	1593.7861	-0.1	Yes
ABP C4	L K V K Y M W D Y L F G K	98	563.9764	3	67.6	1688.9111	-2.2	Yes
	L N T T R S V L W Y W P K	96	554.9756	3	70.3	1661.9043	0.4	Yes
ABP C6	L Q K F L G E R F P G W K	96	535.6412	3	72.5	1603.8987	2.0	Yes
	L H F S R Q W Q W N V R K	95	446.7453	4	33.4	1782.9543	-1.3	Yes
ABP C7	L Q R G V F G F P Y R V K	95	392.2314	4	42.8	1564.8989	-1.4	Yes
	L T F A N H T T Q V R P K	95	378.7159	4	25.0	1510.8369	-1.2	Yes
	T H V V A H A R D A Y R K	95	381.4610	4	27.3	1521.8276	-8.5	Yes
	V Q R Q P P N Y W N G F K	95	544.9509	3	40.8	1631.8320	-0.8	Yes
	V T Y Y G S G P Q W F W K	95	809.3956	2	68.0	1616.7776	-0.6	Yes
ABP C2	L H F A K W N H V W S W K	94	435.2328	4	49.0	1736.9053	-1.7	Yes
	N K F K N W N T W V G G K	94	526.6152	3	56.6	1576.8262	-1.5	Yes
ABP C3	L H K Q F G W H W F G F K	94	429.9771	4	49.8	1715.8838	-2.7	Yes
	L P F H S A Y W Y H M E K	93	427.7078	4	45.5	1706.8027	-0.5	Yes
ABP C5	L Q A K P V P Q F W P F K	93	528.9722	3	59.7	1583.8977	-1.8	Yes
	L Q R A S L Y F P W K V K	93	409.4927	4	48.4	1633.9456	-2.3	Yes
	L A F H N K P E W Y W P K	92	429.4764	4	48.3	1713.8779	-0.9	Yes
	L R F Q Q V S F Y P W R K	92	439.2460	4	51.8	1752.9575	-1.5	Yes
	L E L S T N W V W N P Y K	91	550.2910	3	47.8	1647.8408	6.3	Yes
	T R A L S N F D F F R P K	91	400.2198	4	46.8	1596.8525	-1.5	Yes
	V Q W N M L G F Y N P W K	91	561.2819	3	62.9	1680.8235	0.3	Yes

K A L T V R L M F Y V W K	90	418.2440	4	49.9	1668.9539	-4.2	Yes
L Q F W N G P V H E E H K	90	540.6081	3	56.6	1618.8005	1.2	Yes
L Q Y K P A Q M L N W P K	90	539.2975	3	51.6	1614.8704	0.3	Yes
L V K W T Y T L N M D Q K	90	546.9651	3	64.4	1637.8599	8.4	Yes
L V Y A W G S R D Y P W K	90	820.4226	2	77.1	1638.8306	0.1	Yes
Y A Y Q D D L W G V F P K	90	800.8931	2	76.4	1599.7722	-0.4	Yes
D N R G P D V S Y Y K W K	89	407.4561	4	37.7	1625.7949	0.1	Yes
L Q R T M L L W T R P F K	89	422.9993	4	49.0	1687.9707	-1.6	Yes
V Q W H S Y P N Y L Q Y K	89	575.6227	3	51.1	1723.8472	-0.4	Yes
Q L F S W W T T Q Y A Y K	88	860.9265	2	77.9	1719.8408	-1.4	Yes
V K K Y L L L N A N Y P K	88	521.6507	3	38.0	1561.9343	-2.6	Yes
K A N N R F K N F Y E R K	86	571.9835	3	61.1	1712.9224	3.7	Yes
T T R H T A R V H W V F K	86	410.2341	4	40.9	1636.9062	0.6	Yes
L A F Q H G S T A P S W K	84	476.9171	3	39.2	1427.7310	-0.9	Yes
L H F M N W A M N R G G K	84	520.9266	3	42.1	1559.7603	-1.4	Yes
W F R L N W A T N F S R K	84	431.9831	4	56.5	1723.9060	-1.7	Yes
W H L S T A P S F N P W K	84	523.9368	3	57.1	1568.7888	-0.2	Yes
L A N F G E Q N E W F W K	84	556.6021	3	71.3	1666.7893	-2.9	Yes
K W A Q D Y Y Q N M A W K	83	582.9432	3	65.3	1745.7983	5.4	Yes
V S N R P P Y F W T R F K	83	424.9823	4	43.9	1695.8997	0.3	Yes
W N V D P Y Y F M Q A W K	83	582.9409	3	65.5	1745.8025	-1.0	Yes
G R F R W H A N D H H T K	82	554.2863	3	46.5	1659.8242	7.8	Yes
L H Q Y N H K R P W G W K	82	437.9839	4	60.1	1747.9172	-6.1	Yes
L Q F E K K M Q Q H H D K	82	571.2938	3	55.7	1710.8623	-1.6	Yes
L V E F E Y R W R M F V K	82	451.2451	4	55.9	1800.9497	1.0	Yes
L D F K Q M S G W L H S K	81	394.7071	4	46.2	1574.8027	-2.3	Yes
W H R A V V S D L L F E K	81	533.6312	3	53.3	1597.8728	-0.7	Yes
W L N N Y P W W W S T P K	80	592.9593	3	62.9	1775.8572	-0.6	Yes
K N E L V G Y G W V W G K	74	512.2767	3	49.8	1533.8093	-0.7	Yes
W N V G V Y H K W F R V K	96	430.2407	4	43.2	1716.9365	-1.7	No
F W W S N P Y L R Q G D K	87	565.9498	3	65.1	1694.8318	-2.4	No

Table A2.2. **All AS-MS discovered canonical L-peptides from Library 1.**

Figure A2.6. **Graphical representation of sequence alignment from AS-MS with Library 1.** Weblogo plot of the binders discovered via AS-MS against ACE2 from the canonical L-library (2 x 108 members, X12K) demonstrating the presence of a modest N-terminal motif of '(L/V)(Q/N)'.

Name	Peptide sequence, 3 letter code													ALC (%)	m/z	z	RT, min	Calc Mass, Da	Error, ppm	EIC Selectivity
	Cpa	Tha	4Py	Aad	Cha	Cpa	Gly	bAla	hArg	Ile	Cpa	bAla	Lys	87	520.6335	3	40.9	1558.8801	-0.8	Yes
	Cpa	Tha	4Py	Cpa	Ile	Orn	Msn	D-Pro	Cpa	bAla	Ile	Tha	Lys	99	536.2723	3	39.7	1605.7976	-1.5	Yes
	Cpa	Gln	Dff	Orn	bAla	Aad	Cha	D-Pro	Gln	Dff	Gly	Cpa	Lys	96	542.6180	3	58.0	1624.8384	-3.7	Yes
	Aad	Tha	Ile	Amb	Amb	Cpa	Tha	Gly	4Af	Cha	bSer	bAla	Lys	98	544.5994	3	62.2	1630.7786	-1.4	Yes
ABP N4	Cpa	Tha	Cpa	Gly	D-Pro	D-Pro	Hyp	Dff	Dmf	Tha	bAla	Amb	Lys	99	545.5731	3	60.1	1633.7017	-2.6	Yes
	Ile	Tha	Ile	bAla	Dmf	Gly	Tha	bSer	hArg	Ile	Dff	bAla	Lys	84	551.9450	3	54.2	1652.8130	0.1	Yes
	Cpa	Gln	Gly	4Py	Cpa	Tha	D-Pro	Cha	Asp	hArg	Cpa	Cha	Lys	95	552.3120	3	59.3	1653.9172	-1.9	Yes
	Cha	Tha	Gly	Php	D-Pro	bSer	Ile	bAla	hArg	Gly	4Af	Php	Lys	84	552.6462	3	73.0	1654.9170	-0.1	Yes
	Cpa	Gln	Dff	bSer	Gly	Tha	Orn	Ile	Dmf	Gln	Gly	4Af	Lys	99	554.6065	3	38.6	1660.7993	-1.0	Yes
	Cpa	Cpa	Tha	4Py	Asp	D-Pro	Msn	Gly	Hyp	Ile	Dff	Php	Lys	87	555.5832	3	53.7	1663.7334	-3.4	Yes
	Cpa	Tha	Dff	Aad	Gly	Tha	bAla	Gly	Amb	Dff	hArg	4Py	Lys	85	559.5663	3	51.7	1675.6797	-1.6	Yes
	Cpa	Tha	Dff	Hyp	Cpa	Orn	Dff	Dff	bAla	Cpa	bAla	Amb	Lys	97	562.2561	3	65.3	1683.7488	-1.4	Yes
	Hyp	4Af	Msn	Ile	D-Pro	Amb	Amb	Amb	Amb	Php	bSer	bAla	Lys	97	562.6097	3	59.7	1684.8110	-2.3	Yes
ABP N5	Cpa	Tha	Dff	Gly	Gln	D-Pro	Msn	Cha	Gln	Tha	bAla	Tha	Lys	99	567.2306	3	63.7	1698.6736	-2.1	Yes
	Php	Orn	Gly	Cpa	Tha	D-Pro	Ile	Orn	Tha	Asp	Php	Tha	Lys	80	568.6047	3	56.5	1702.7930	-0.5	Yes
	Cpa	Tha	Dff	Aad	Cpa	Hyp	Gly	bAla	Hyp	Dff	hArg	Tha	Lys	90	570.5817	3	52.1	1708.7258	-1.6	Yes
	Cpa	Gln	4Py	Msn	Cha	bSer	Php	bSer	Ile	bSer	4Af	D-Pro	Lys	84	571.3110	3	44.4	1710.9165	-3.2	Yes
	Cpa	Tha	Dmf	bSer	bSer	Ile	Ile	bSer	Asp	Cha	Orn	Dff	Lys	99	571.6335	3	70.5	1711.8816	-1.7	Yes
	Ile	Dff	D-Pro	Asp	Aad	Amb	Dff	Gly	Gly	Amb	hArg	Php	Lys	98	573.2747	3	64.3	1716.8074	-3.0	Yes
ABP N1	Cpa	Tha	Dff	Orn	Gln	Orn	bSer	Tha	Cha	Msn	Gly	4Af	Lys	99	435.9548	4	28.7	1739.7905	-0.3	Yes
	Cpa	Gln	Dff	bAla	Gly	Ile	4Py	Dmf	Dmf	Cpa	Dff	Cpa	Lys	82	581.6208	3	58.3	1741.8486	-4.5	Yes
	Cha	Dff	bAla	Ile	Aad	Asp	4Py	Gly	Php	Cha	Ile	Msn	Lys	92	582.6428	3	80.2	1744.9070	-0.2	Yes
	Asp	D-Pro	Msn	4Af	Amb	Amb	Amb	Dff	Ile	Gly	Gln	Php	Lys	98	584.2676	3	66.0	1749.7825	-0.9	Yes
	Php	4Py	Amb	Msn	bSer	Cha	Aad	Asp	Aad	Msn	Gly	Amb	Lys	83	584.5899	3	58.3	1750.7581	-5.8	Yes
	Php	Gly	D-Pro	4Af	Tha	Gln	Dff	Cha	Gln	Ile	Gly	Php	Lys	80	584.9670	3	76.6	1751.8818	-1.6	Yes
	Cpa	Tha	Ile	Amb	Tha	Ile	Gly	Dff	bAla	Dmf	Gln	Dff	Lys	98	585.2582	3	83.6	1752.7563	-2.1	Yes
	Cpa	Tha	4Py	Orn	hArg	Asp	Php	Gly	Asp	Amb	Cha	Cha	Lys	84	586.3202	3	47.9	1755.9280	6.1	Yes
	Cpa	Tha	Dff	Gln	Cpa	Orn	Ile	bSer	Amb	4Af	Orn	Php	Lys	86	440.2345	4	46.3	1756.9084	0.3	Yes
	Ile	Tha	Ile	Ile	D-Pro	Amb	D-Pro	Dff	bSer	Dmf	Orn	Php	Lys	90	586.9777	3	76.8	1757.9177	-3.7	Yes
ABP N6	Cpa	Orn	Dff	Msn	Hyp	Tha	D-Pro	Gln	Dff	Cha	Cha	bAla	Lys	99	590.6218	3	73.1	1768.8450	-0.8	Yes
ABP N8	Cpa	Tha	hArg	Dpro	Gly	bAla	Php	Asp	Cha	Msn	Dff	Msn	Lys	98	590.9379	3	60.7	1769.7900	1.1	Yes
	Cpa	Aad	Dff	Dff	Amb	Gly	Dff	Hyp	Gln	hArg	Gly	Dff	Lys	95	597.5940	3	65.2	1789.7646	-2.5	Yes
	Asp	Msn	Gly	Msn	Amb	Dmf	bSer	D-Pro	4Af	Amb	Php	Amb	Lys	94	599.9252	3	55.5	1796.7578	-2.2	Yes
	4Af	Tha	Cpa	Amb	Dff	Amb	Cpa	D-Pro	Dmf	Amb	4Py	D-Pro	Lys	82	605.9469	3	62.4	1814.8240	-2.8	Yes
	Tha	Dff	hArg	Amb	D-Pro	Tha	Asp	Msn	Php	bAla	Php	Gly	Lys	99	606.5896	3	57.1	1816.7485	-0.9	Yes
	Amb	4Py	Asp	Tha	Aad	Amb	Dff	Msn	D-Pro	bAla	Php	4Py	Lys	86	456.1885	4	44.6	1820.7288	-2.0	Yes
	Ile	Tha	Dff	Gly	bAla	Amb	4Af	Tha	Dff	bSer	Dmf	Msn	Lys	99	609.9074	3	63.2	1826.7026	-1.2	Yes
	Php	Aad	Aad	Tha	4Af	Ile	Gly	Gln	Php	Msn	Asp	Amb	Lys	92	916.4117	2	68.2	1830.8108	-1.0	Yes
	Ile	Tha	hArg	bAla	Amb	Cha	hArg	Gly	Dff	Dmf	Amb	Tha	Lys	99	461.9810	4	58.5	1843.8977	-1.6	Yes
	Cpa	Tha	Tha	Msn	Ile	Gly	bAla	Dmf	Gln	hArg	Dmf	Php	Lys	81	623.6255	3	54.6	1867.8569	-1.2	Yes
	Cpa	Tha	Tha	Msn	Ile	Gln	Dmf	bAla	Gly	hArg	Dmf	Php	Lys	81	623.6255	3	54.7	1867.8569	-1.2	Yes
	Cha	Tha	Amb	D-Pro	Hyp	Tha	Tha	Php	Asp	4Py	Dff	Php	Lys	97	630.9304	3	74.6	1889.7688	0.2	Yes
	Dmf	Tha	hArg	Php	Aad	hArg	bAla	Ile	4Af	4Af	Msn	Gly	Lys	99	381.9974	5	28.7	1904.9541	-1.9	Yes
	Cha	Tha	Dff	bAla	Msn	Hyp	Tha	4Af	Dff	Msn	bSer	hArg	Lys	97	479.9479	4	51.6	1915.7649	-1.2	Yes
	Tha	Tha	hArg	bAla	Dmf	Php	Hyp	Php	Dff	Ile	Gly	Php	Lys	99	643.9741	3	72.3	1928.9067	-3.2	Yes
	Cha	Tha	Amb	hArg	Dff	Gly	Tha	hArg	Php	Amb	Amb	hArg	Lys	99	389.5978	5	50.4	1942.9563	-1.9	Yes
	Cpa	Tha	Dff	Msn	Tha	Orn	Dff	Gln	hArg	Amb	Amb	Php	Lys	99	490.7114	4	58.2	1958.8191	-1.3	Yes
	Php	bAla	Dff	Amb	Msn	Tha	Aad	Amb	Dmf	Asp	Dmf	Amb	Lys	83	659.2672	3	69.7	1974.7805	-0.4	Yes
	4Py	Tha	Cpa	Gly	D-Pro	bAla	Php	Orn	Cpa	Amb	D-Pro	Tha	Lys	88	516.2507	3	52.4	1545.7371	-4.4	Obscured in baseline
	Php	Gly	bAla	Hyp	Dff	bAla	Amb	Amb	4Py	Orn	Amb	Php	Lys	88	419.9605	4	47.6	1675.8149	-1.1	Obscured in baseline
	Cpa	4Py	hArg	4Af	Aad	4Py	Cpa	Cpa	4Af	4Af	Asp	Amb	Lys	86	439.4809	4	48.7	1753.9048	-5.9	Obscured in baseline
	Ile	4Af	Aad	Aad	Amb	Dff	Amb	Amb	Php	Gly	bSer	Amb	Lys	99	589.9507	3	63.6	1766.8306	-0.1	Obscured in baseline
	Tha	Tha	Cpa	Cha	bAla	4Py	Cha	Dmf	Gly	Dff	Php	D-Pro	Lys	97	607.9648	3	88.3	1820.8740	-0.7	Obscured in baseline
	Aad	Tha	Dff	bAla	Dmf	Cpa	Asp	Php	D-Pro	4Af	Ile	Tha	Lys	89	615.2768	3	70.5	1842.8069	0.9	Obscured in baseline
	Cpa	Gln	Dff	Orn	Ile	Amb	Gly	Amb	Gly	Cha	Ile	hArg	Lys	96	403.7390	4	56.6	1610.9258	0.7	No
	Cpa	Tha	4Py	Hyp	Orn	4Af	Dmf	Dmf	Cha	Cpa	bAla	Ile	Lys	80	604.3242	3	47.2	1809.9634	-7.0	No

Table A2.3. **All AS-MS discovered noncanonical-L peptides from Library 2.**

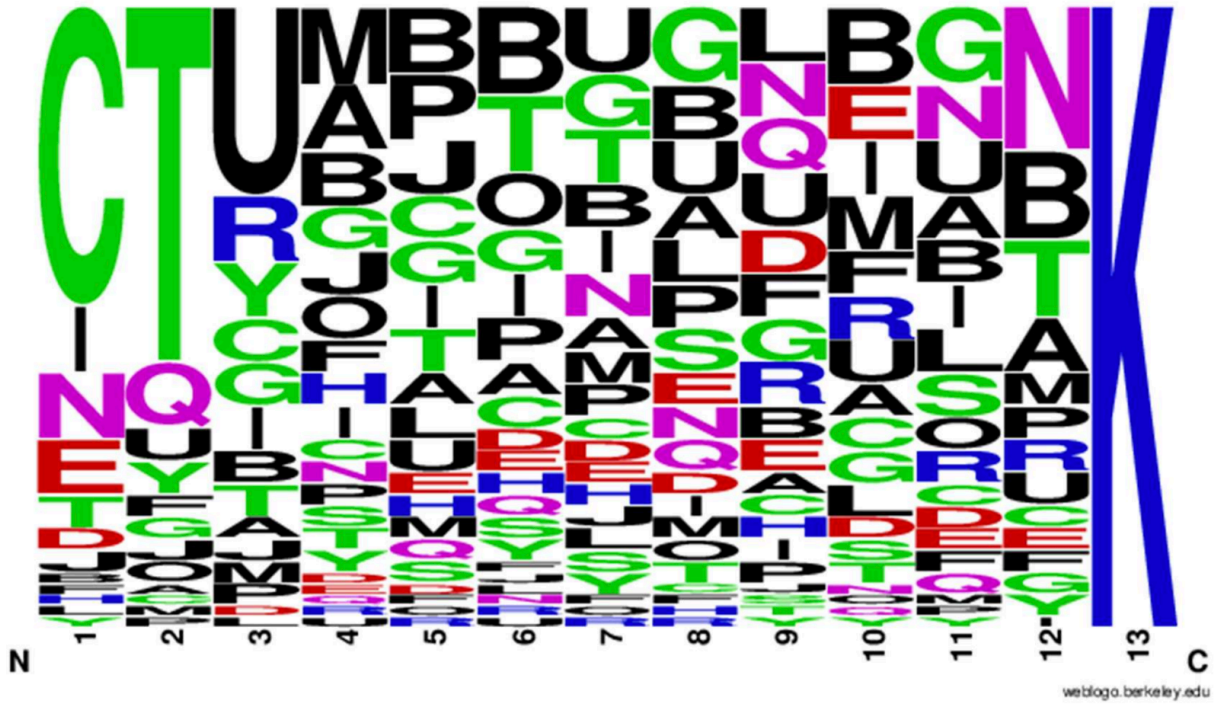


***Figure A2.7*. Graphical representation of sequence alignment from AS-MS with Library 2.** Weblogo plot of the binders discovered via AS-MS against ACE2 from the noncanonical library (2 x 108 members, X12K) demonstrating the presence of a modest N-terminal motif of '(C/I)(T/Q)U'. The amino acid single letter code correspondence: Gly-G; bAla-A; Amb-B; Ile-I; CpaC; Cha-E; D-Pro-P; Hyp-H; Gln-Q; bSer-S; Msn-M; Tha-T; 4Py-Y; 4AF-F; Asp-D; Aad-J; Orn-O; hAr-R; Dff-U; Dmf-L; Php-N; Lys-K.

A2.5.2 Extracted ion counts of representative peptides

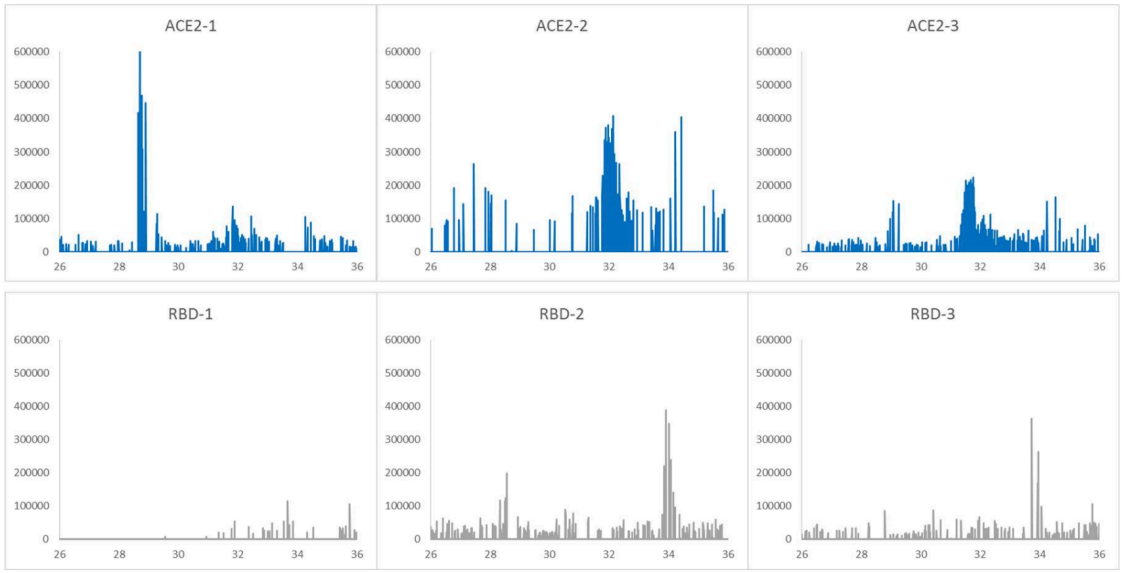


***Figure A2.8*. Extracted ion count (EIC) of ABP N1**. ABP N1 was sequenced with an ALC of 99 from the observed 435.955 m/z, z = 4 ion, and retention time of 29 minutes for a calculated parent mass of 1739.791, meaning the observed error was -0.3 ppm. An EIC was performed from 435.95 - 435.96 m/z and demonstrates that ABP N1 showed ACE2 specificity in the library selection.

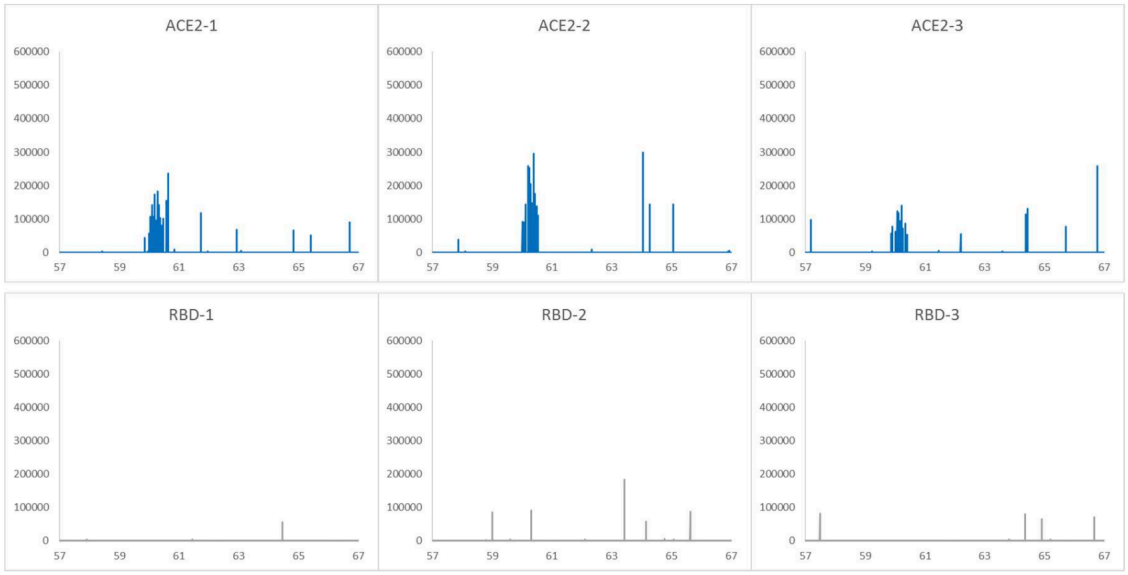


***Figure A2.9.* EIC of ABP N4.** ABP N4 was sequenced with an ALC of 99 from the observed 545.574 m/z, z = 3 ion, and retention time of 60 minutes for a calculated parent mass of 1633.7017, meaning the observed error was -1.9 ppm. An EIC was performed from 545.57 - 545.58 m/z and demonstrates that ABP N4 showed ACE2 specificity in the library selection.

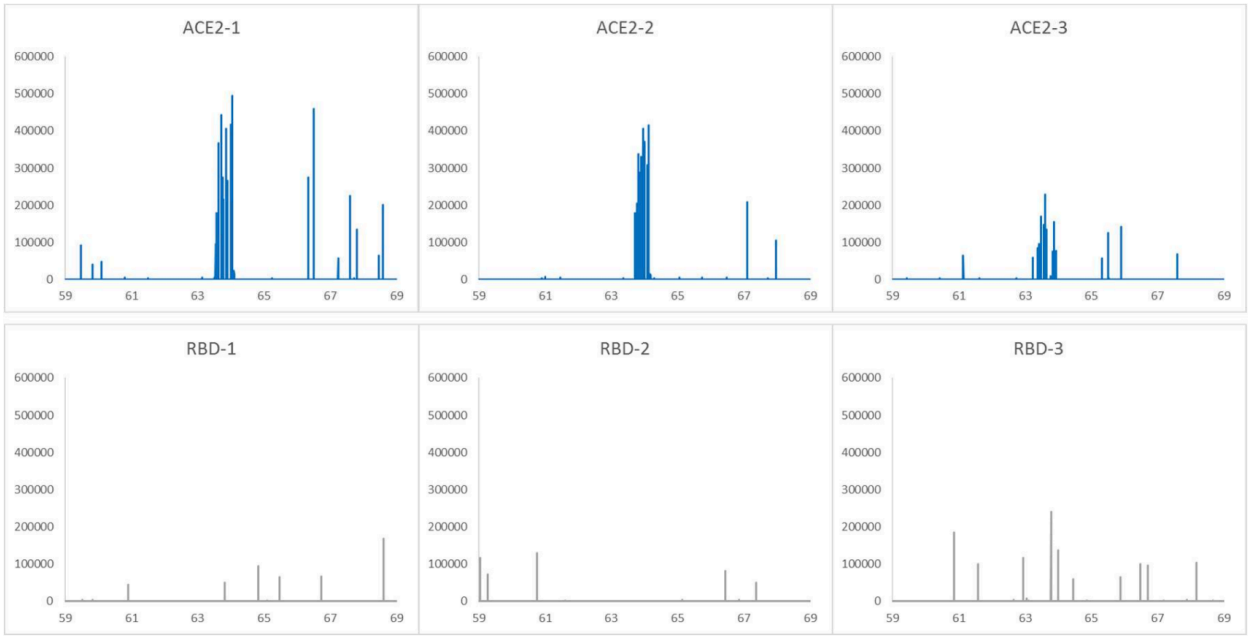


***Figure A2.10*. EIC of ABP N5.** ABP N5 was sequenced with an ALC of 99 from the observed 567.231 m/z, z = 3 ion, and retention time of 64 minutes for a calculated parent mass of 1698.6740, meaning the observed error was -2.1 ppm. An EIC was performed from 567.23 – 567.24 m/z and demonstrates that ABP N5 showed ACE2 specificity in the library selection.

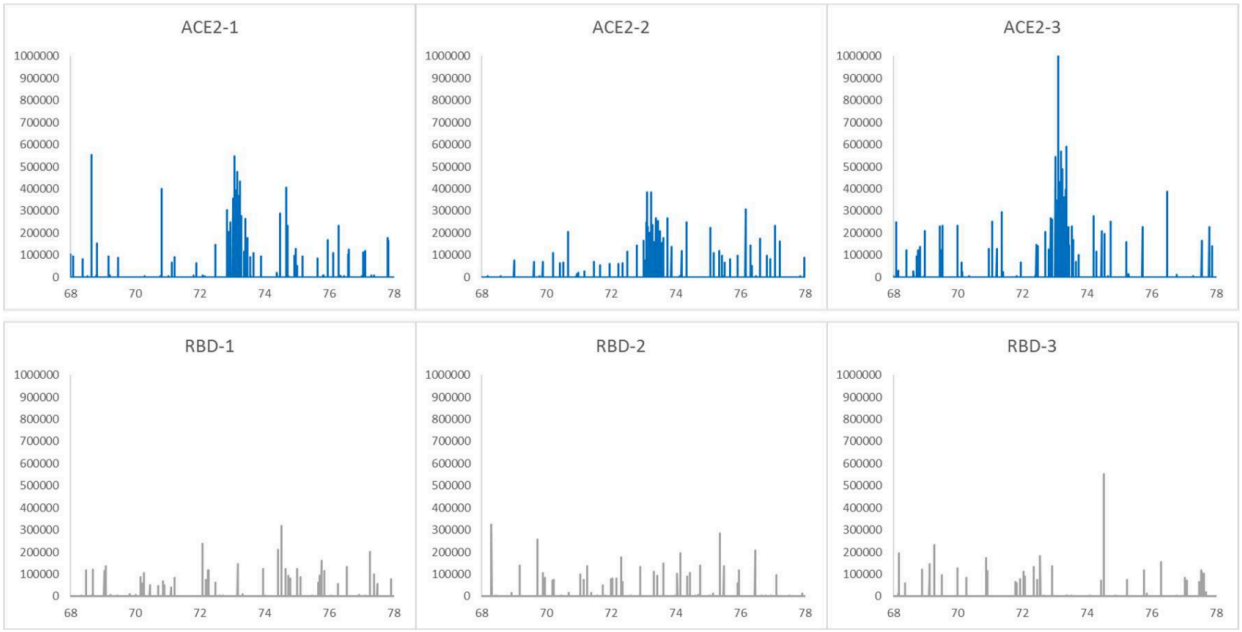


***Figure A2.11*. EIC of ABP N6.** ABP N6 was sequenced with an ALC of 99 from the observed 590.622 m/z, z = 3 ion, and retention time of 73 minutes for a calculated parent mass of 1768.845, meaning the observed error was -0.8 ppm. An EIC was performed from 590.62 - 590.63 m/z and demonstrates that ABP N6 showed ACE2 specificity in the library selection.

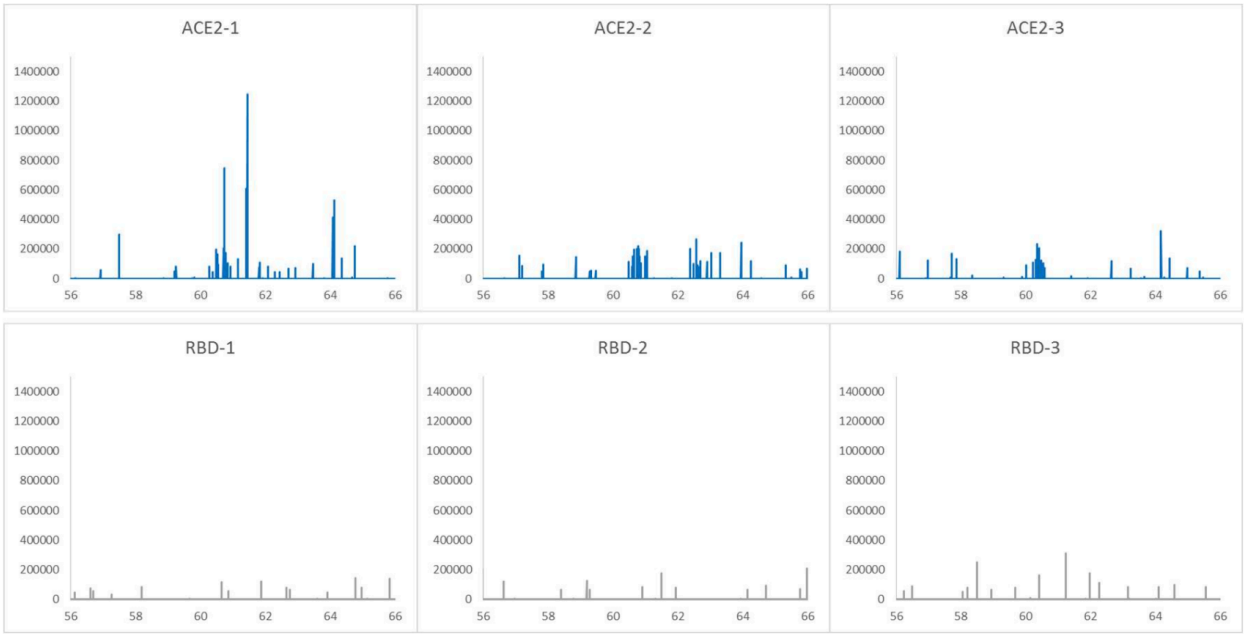


***Figure A2.12*. EIC of ABP N8.** ABP N8 was sequenced with an ALC of 98 from the observed 590.9379 m/z, z = 3 ion, and retention time of 61 minutes for a calculated parent mass of 1769.7900, meaning the observed error was 1.1 ppm. An EIC was performed from 590.93 - 590.94 m/z and demonstrates that ABP N8 showed ACE2 specificity in the library selection.

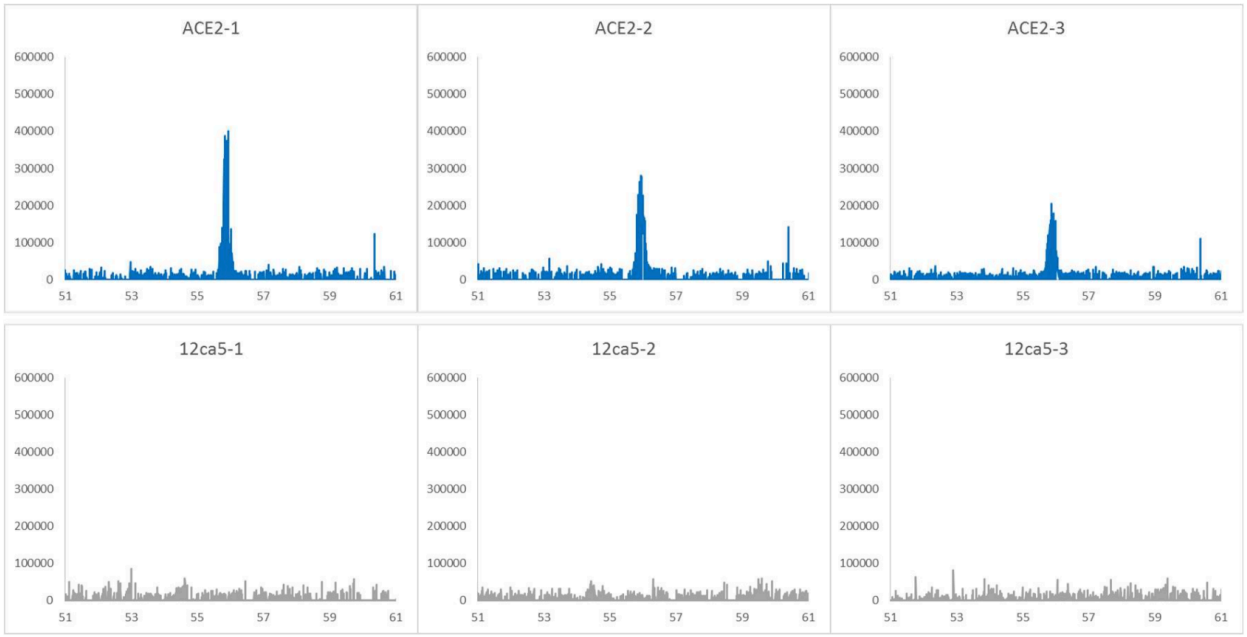


***Figure A2.13.* EIC of ABP C1.** ABP C1 was sequenced with an ALC of 99 from the observed 515.2775 m/z, z = 3 ion, and retention time of 56 minutes for a calculated parent mass of 1542.8130, meaning the observed error was -1.6 ppm. An EIC was performed from 515.27 - 515.28 m/z and demonstrates that ABP C1 showed ACE2 specificity in the library selection.

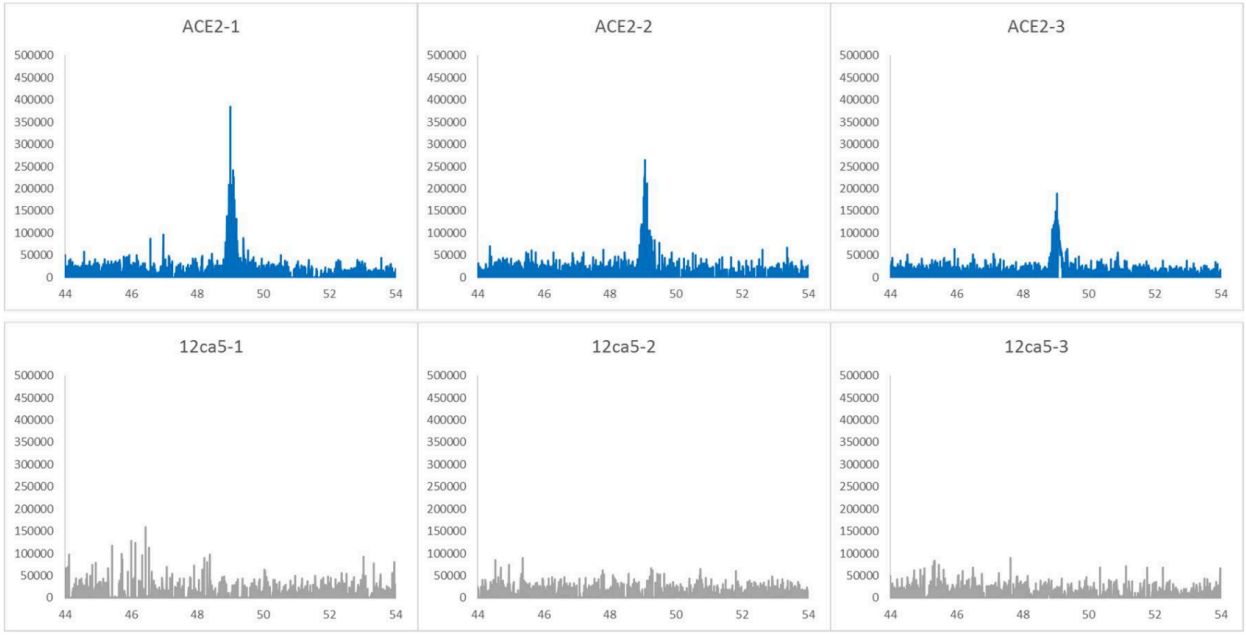


***Figure A2.14*. EIC of ABP C2.** ABP C2 was sequenced with an ALC of 94 from the observed 435.2328 m/z, z = 4 ion, and retention time of 49 minutes for a calculated parent mass of 1736.9053, meaning the observed error was -1.7 ppm. An EIC was performed from 435.23 - 435.24 m/z and demonstrates that ABP C2 showed ACE2 specificity in the library selection.

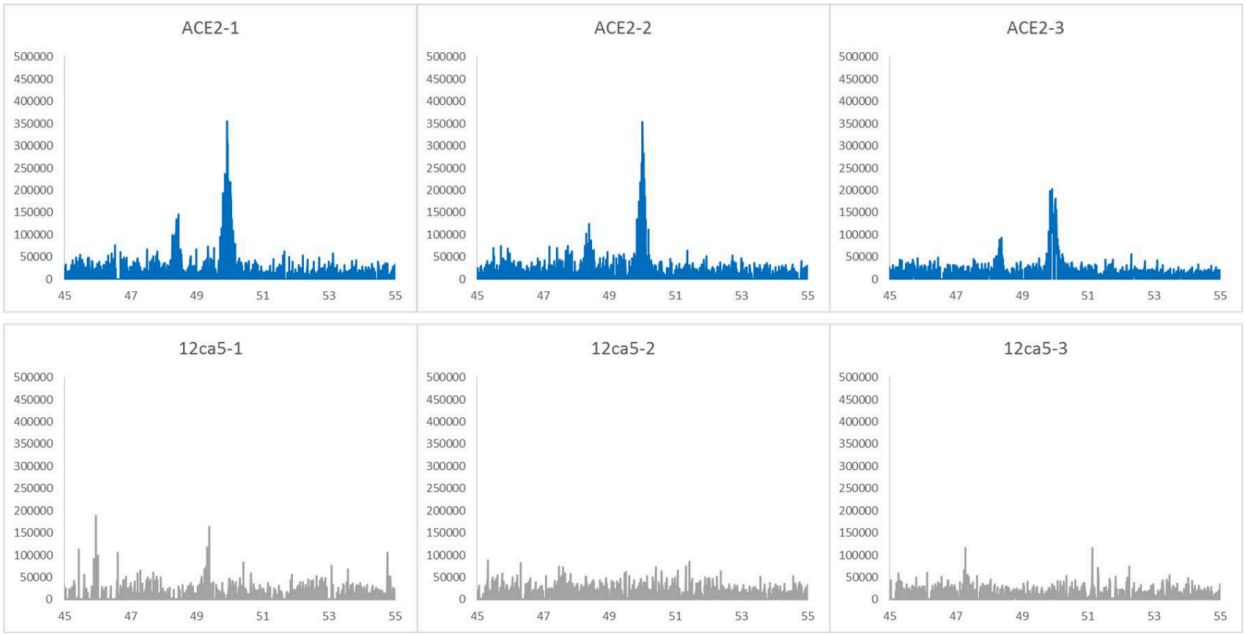


***Figure A2.15.* EIC of ABP C3.** ABP C3 was sequenced with an ALC of 94 from the observed 429.9771 m/z, z = 4 ion, and retention time of 50 minutes for a calculated parent mass of 1715.8838, meaning the observed error was -2.7 ppm. An EIC was performed from 429.97 - 429.98 m/z and demonstrates that ABP C3 showed ACE2 specificity in the library selection.

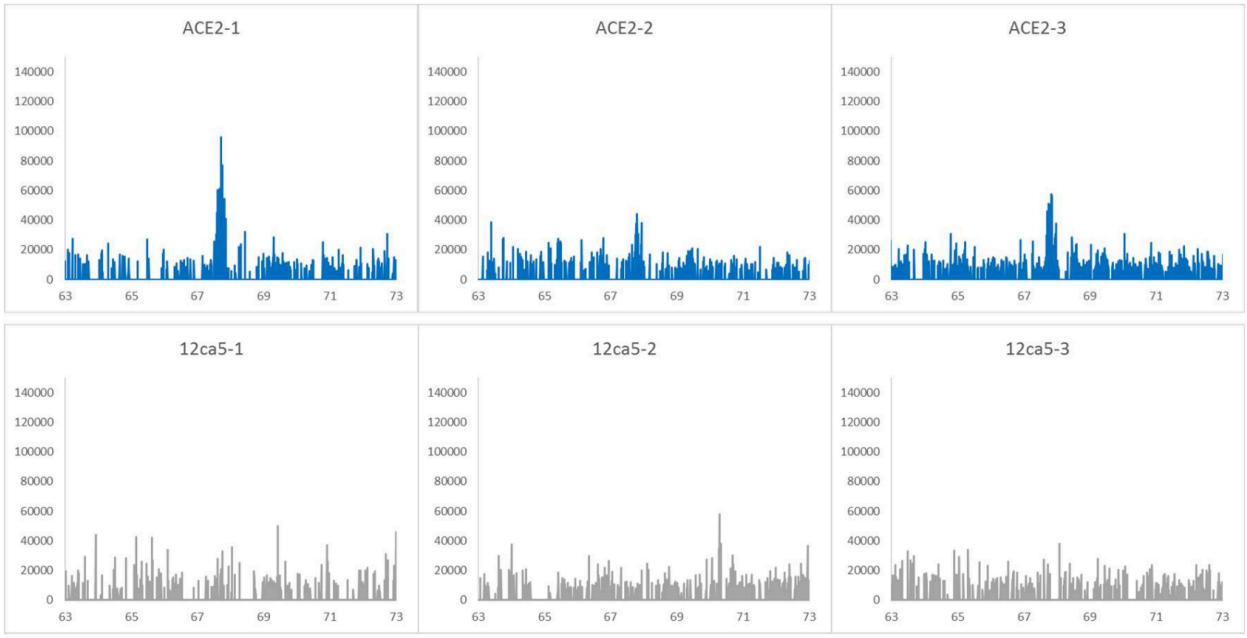


***Figure A2.16*. EIC of ABP C4.** ABP C4 was sequenced with an ALC of 94 from the observed 563.9764 m/z, z = 3 ion, and retention time of 68 minutes for a calculated parent mass of 1688.9111, meaning the observed error was -2.2 ppm. An EIC was performed from 563.97 - 563.98 m/z and demonstrates that ABP C4 showed ACE2 specificity in the library selection.

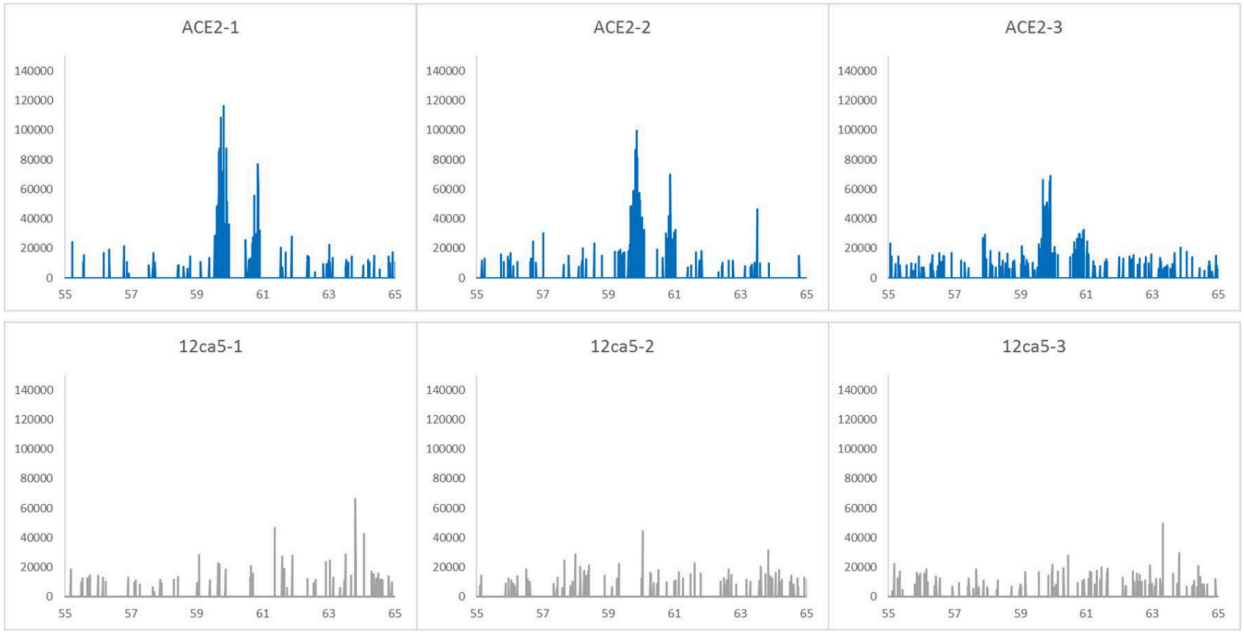


***Figure A2.17*. EIC of ABP C5.** ABP C5 was sequenced with an ALC of 93 from the observed 528.9722 m/z, z = 3 ion, and retention time of 60 minutes for a calculated parent mass of 1583.8977, meaning the observed error was -1.8 ppm. An EIC was performed from 528.97 - 528.98 m/z and demonstrates that ABP C5 showed ACE2 specificity in the library selection.

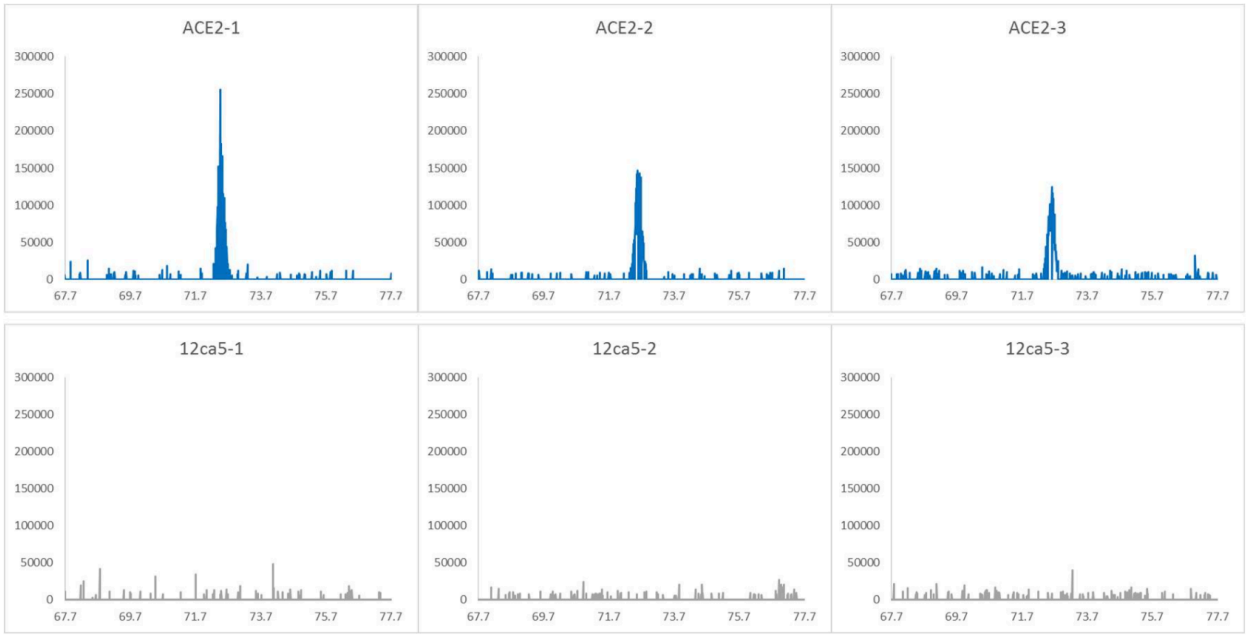


***Figure A2.18*. EIC of ABP C6.** ABP C6 was sequenced with an ALC of 96 from the observed 535.6412 m/z, z = 3 ion, and retention time of 73 minutes for a calculated parent mass of 1603.8987, meaning the observed error was 2.0 ppm. An EIC was performed from 535.64 - 535.65 m/z and demonstrates that ABP C6 showed ACE2 specificity in the library selection.

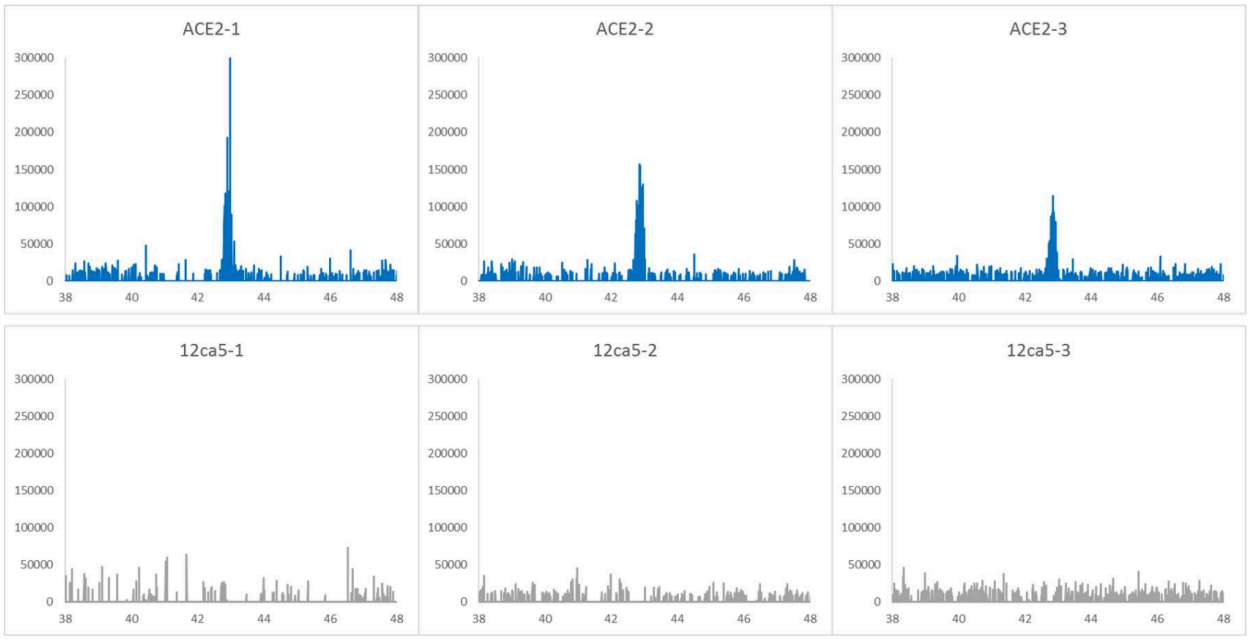


***Figure A2.19.* EIC of ABP C7.** ABP C7 was sequenced with an ALC of 95 from the observed 392.2314 m/z, z = 4 ion, and retention time of 43 minutes for a calculated parent mass of 1564.8989, meaning the observed error was -1.4 ppm. An EIC was performed from 392.23 - 392.24 m/z and demonstrates that ABP C7 showed ACE2 specificity in the library selection.

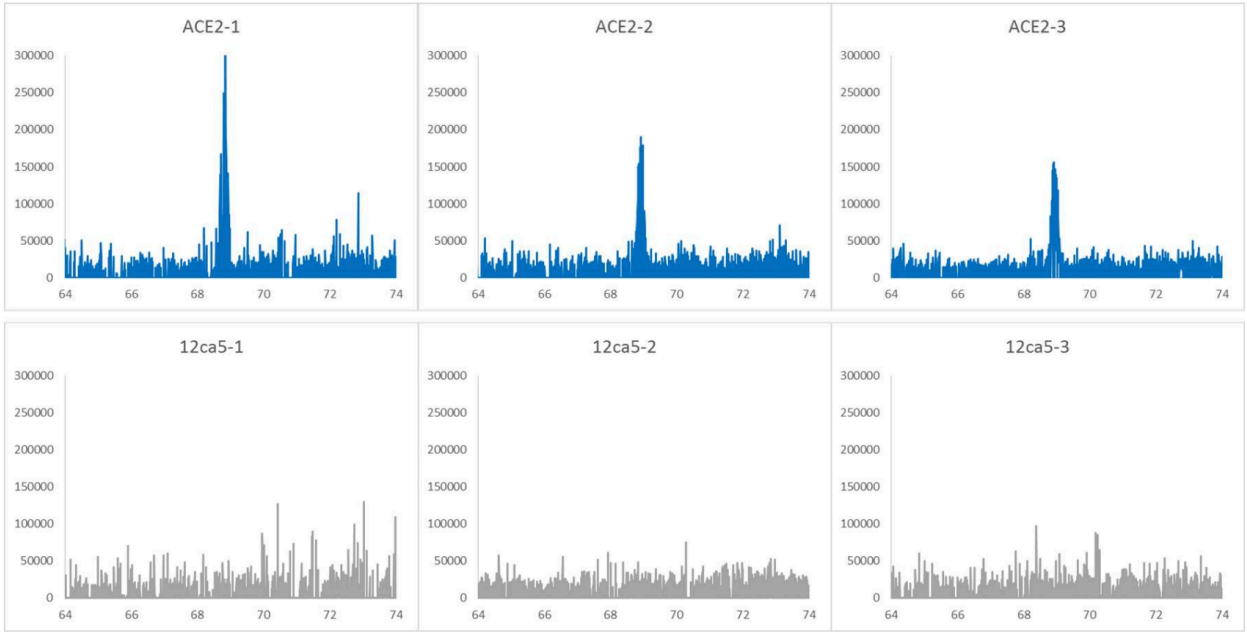


***Figure A2.20*. EIC of ABP C8.** ABP C8 was sequenced with an ALC of 99 from the observed 578.2968 m/z, z = 3 ion, and retention time of 69 minutes for a calculated parent mass of 1731.8708, meaning the observed error was -1.3 ppm. An EIC was performed from 578.29 - 578.30 m/z and demonstrates that ABP C8 showed ACE2 specificity in the library selection.

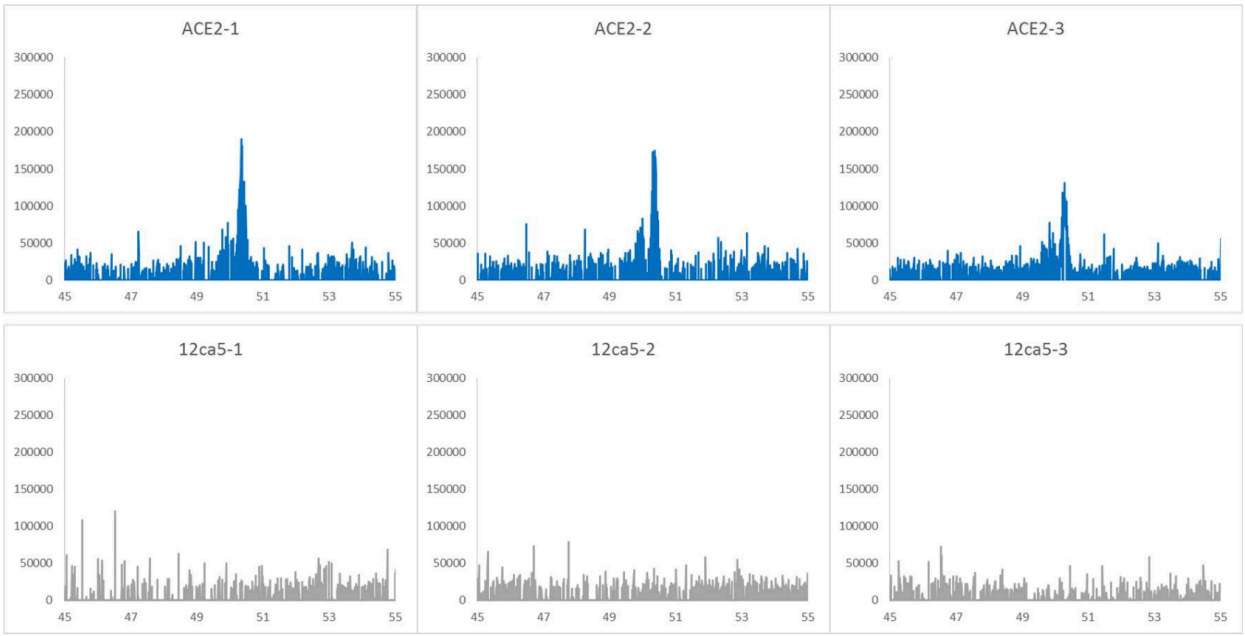


***Figure A2.21*. EIC of ABP C9.** ABP C9 was sequenced with an ALC of 98 from the observed 553.2873 m/z, z = 3 ion, and retention time of 50 minutes for a calculated parent mass of 1656.8447, meaning the observed error was -1.3 ppm. An EIC was performed from 553.28 - 553.29 m/z and demonstrates that ABP C9 showed ACE2 specificity in the library selection.

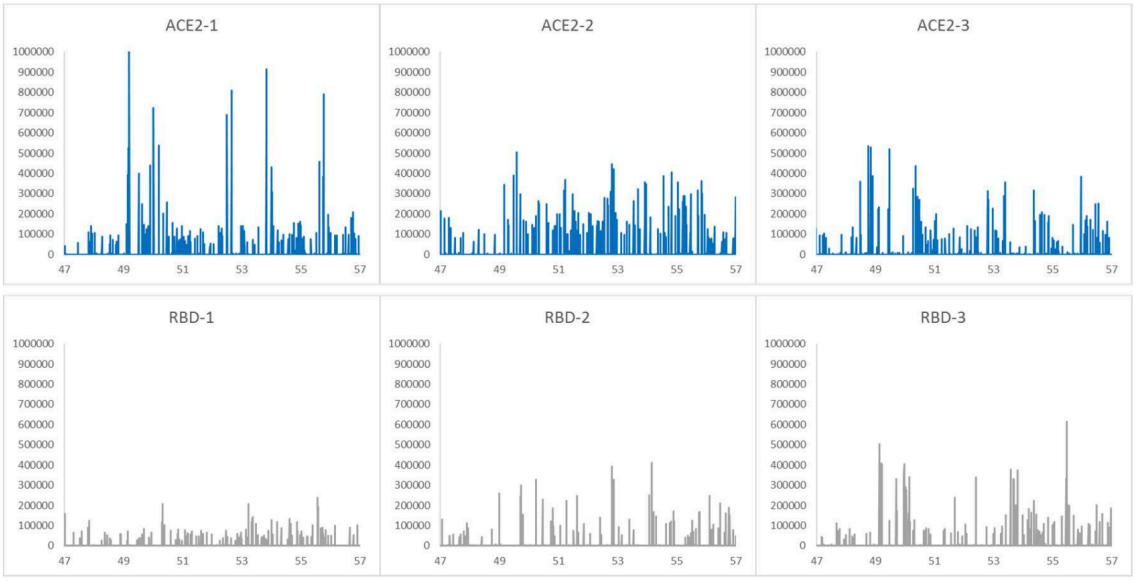


***Figure A2.22*. EIC of '4Py, Tha, Cpa, Gly, D-Pro, bAla, Php, Orn, Cpa, Amb, D-Pro, Tha, Lys'** that was sequenced, identified by PEAKS, and filtered through the standard library selection protocol. However, the EIC of this noncanonical peptide does not reveal convincing or distinct peaks that match its observed retention time of 52 minutes for an observed 516.2507 m/z and z = 3 ion. Thus, this noncanonical peptide serves as a representative example considered to be 'obscured' in the baseline of the mass chromatogram.

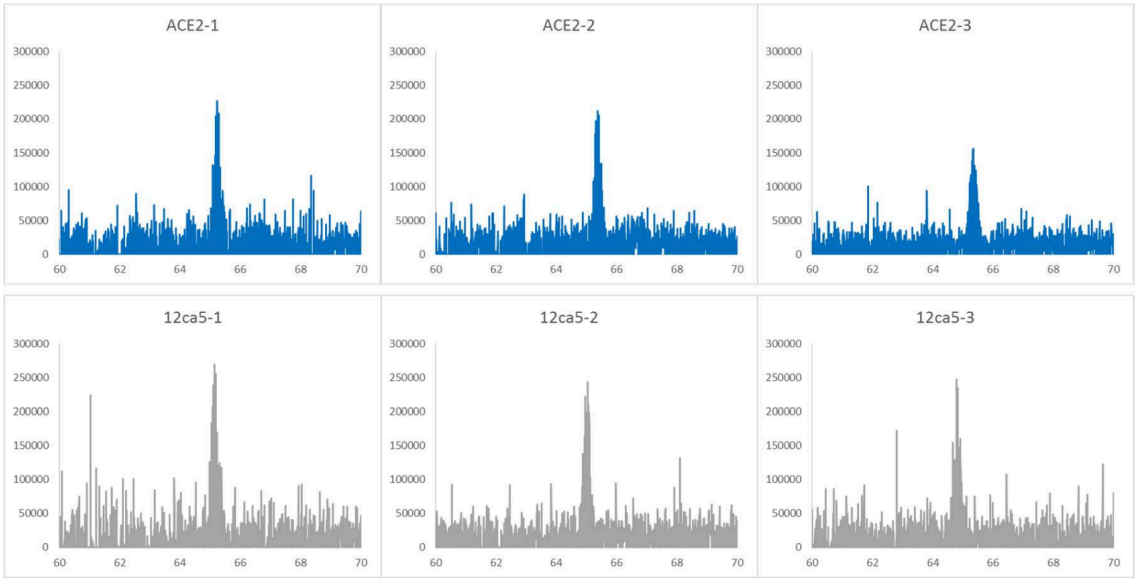


***Figure A2.23.* EIC of FWWSNPYLRQGDK** that was sequenced, assigned in PEAKS, and filtered through the standard library selection protocol. This peptide was sequenced with an ALC of 87, observed as 565.9498 m/z, z = 3 at a retention time of 65 for a calculated mass of 1694.8318 Da, meaning the observed error was -2.4 ppm. However, the EIC of this peptide revealed that it nonspecifically binds to the anti-hemagglutinin monoclonal antibody clone 12ca5 in similar abundance as it does to ACE2. Therefore, this peptide is considered to be 'nonspecific'.

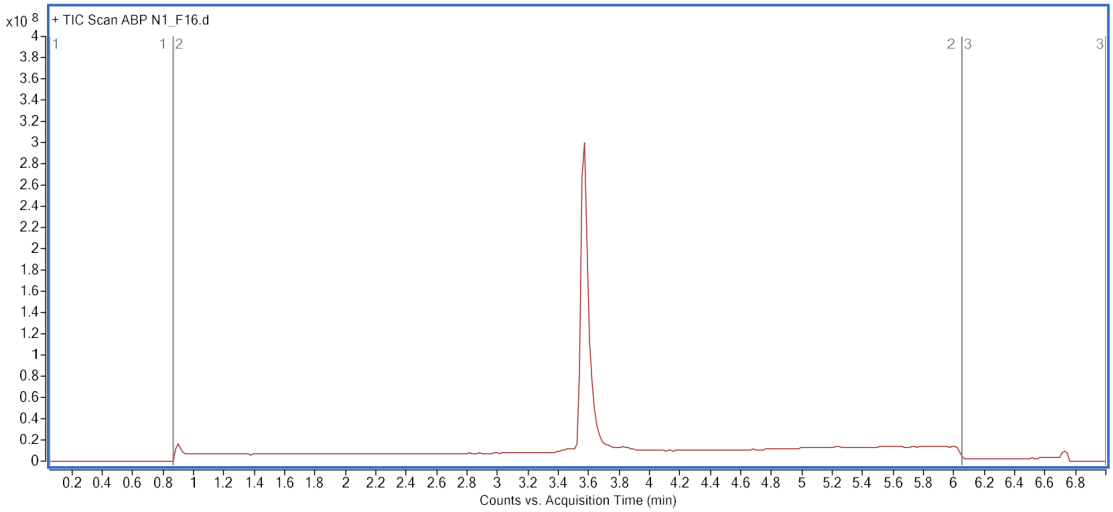
A2.6 Appendix II: Raw LC-MS data

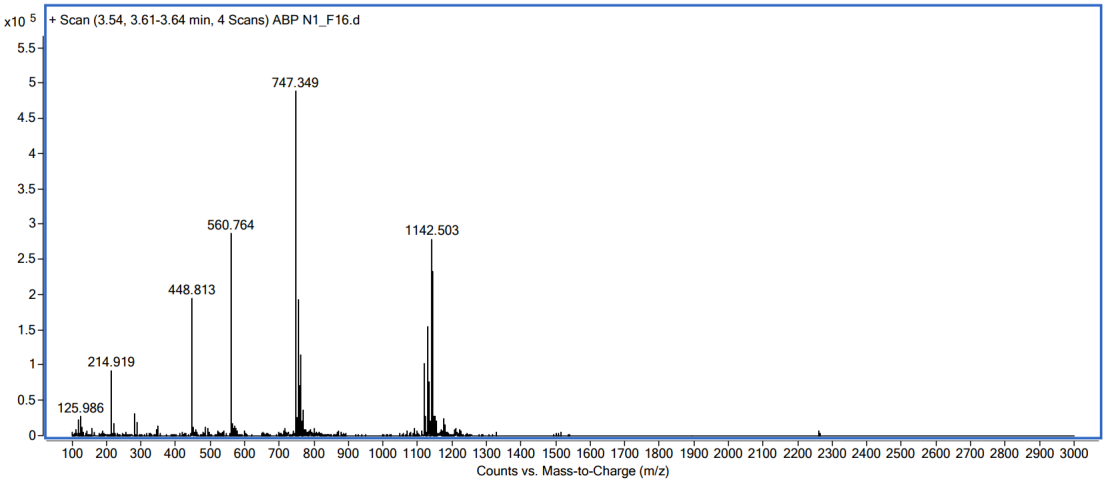
Note: After synthesis and purification, peptide purity was assessed under either condition 1 or condition 2 specified in the Methods section. The total ion chromatograph (TIC) was analyzed using software Agilent MassHunter Qualitative Analysis v10.0 and the MS spectrum was extracted over the main product peaks.

A2.6.1 LC-MS traces of purified noncanonical peptides

A2.6.1.1 Biotin-ABP-N1

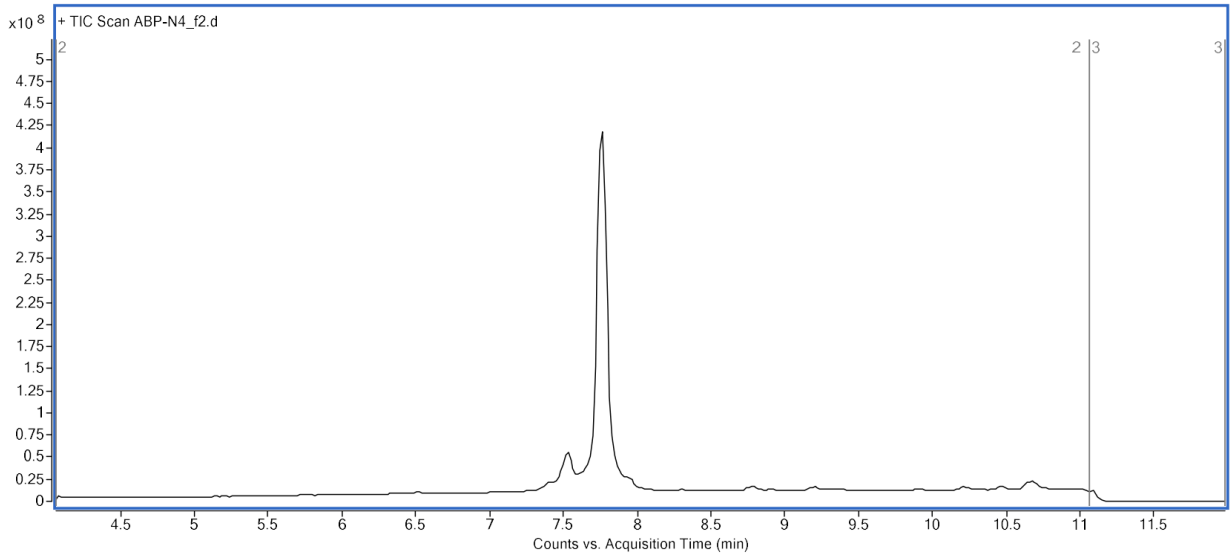
Observed: 2237.95 Da; Calculated: 2238.02 Da

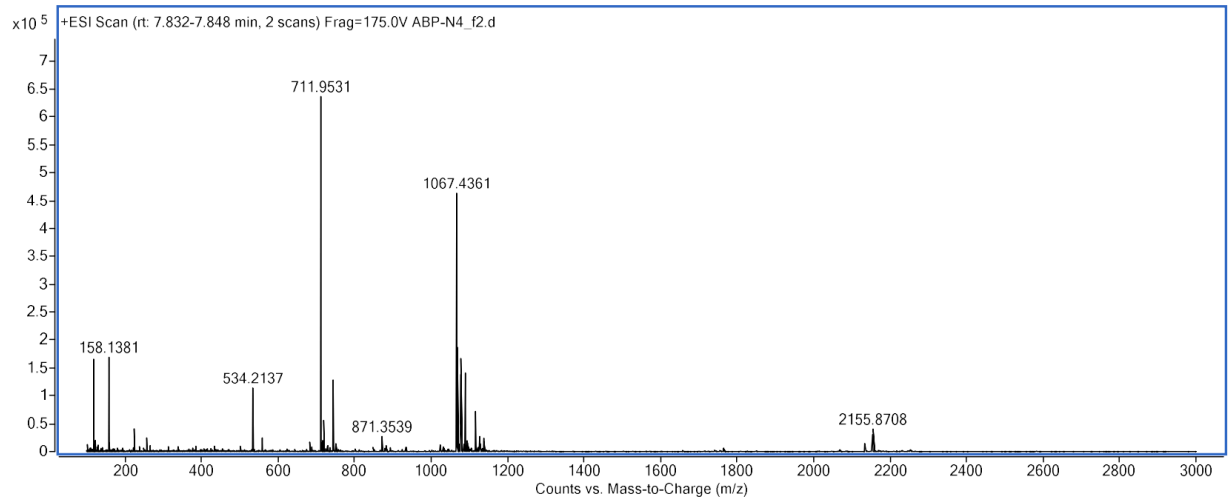




A2.6.1.2 Biotin-ABP-N4

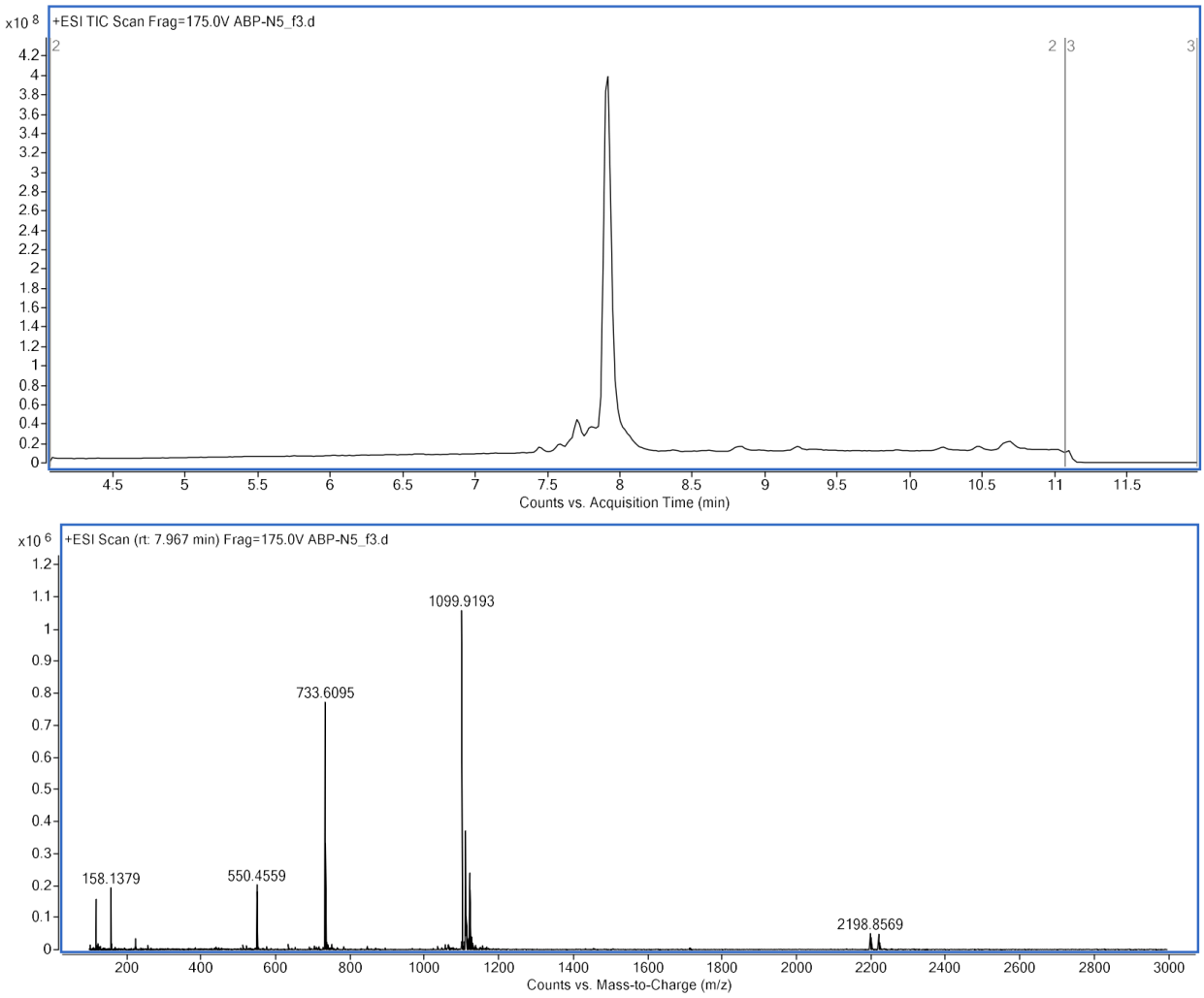
Observed: 2131.88 Da; Calculated: 2131.93 Da





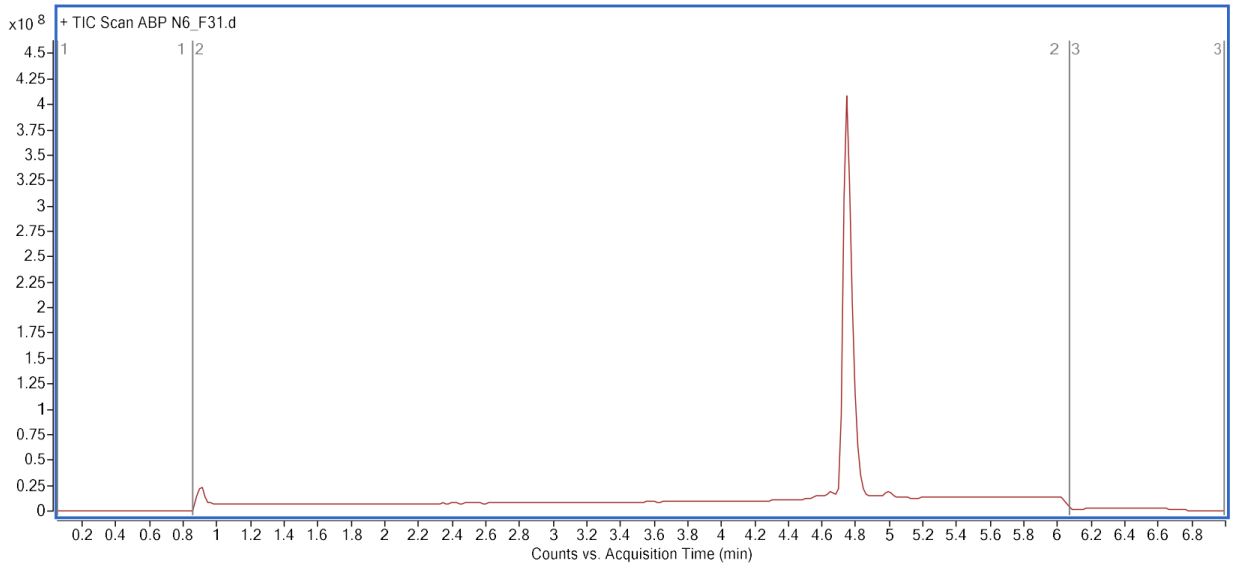
A2.6.1.3 Biotin-ABP-N5

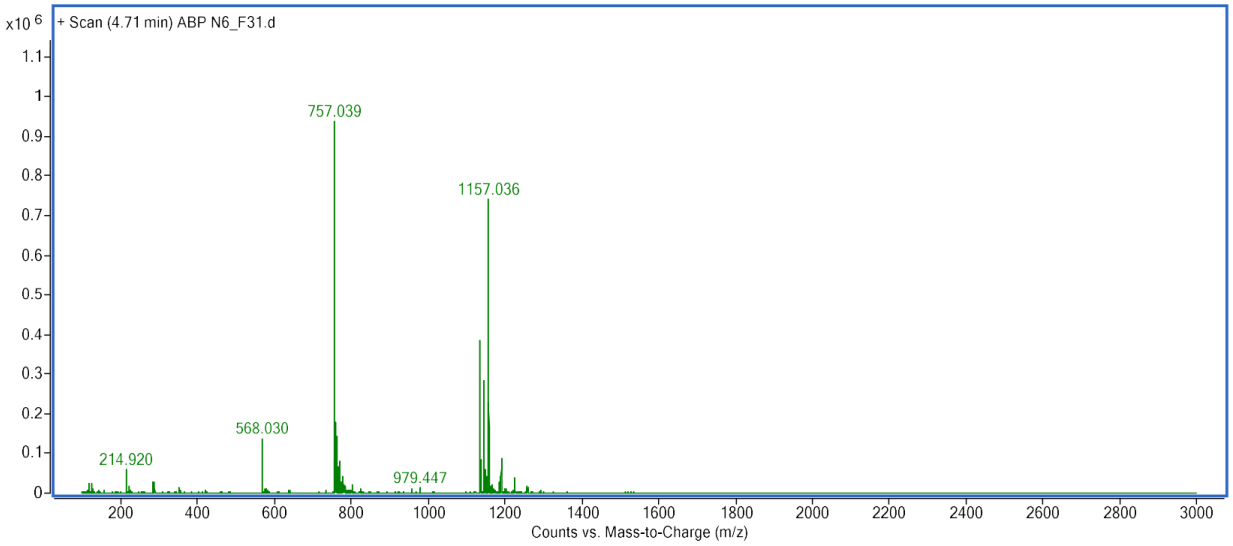
Observed: 2196.86 Da; Calculated: 2196.90 Da



A2.6.1.4 Biotin-ABP-N6

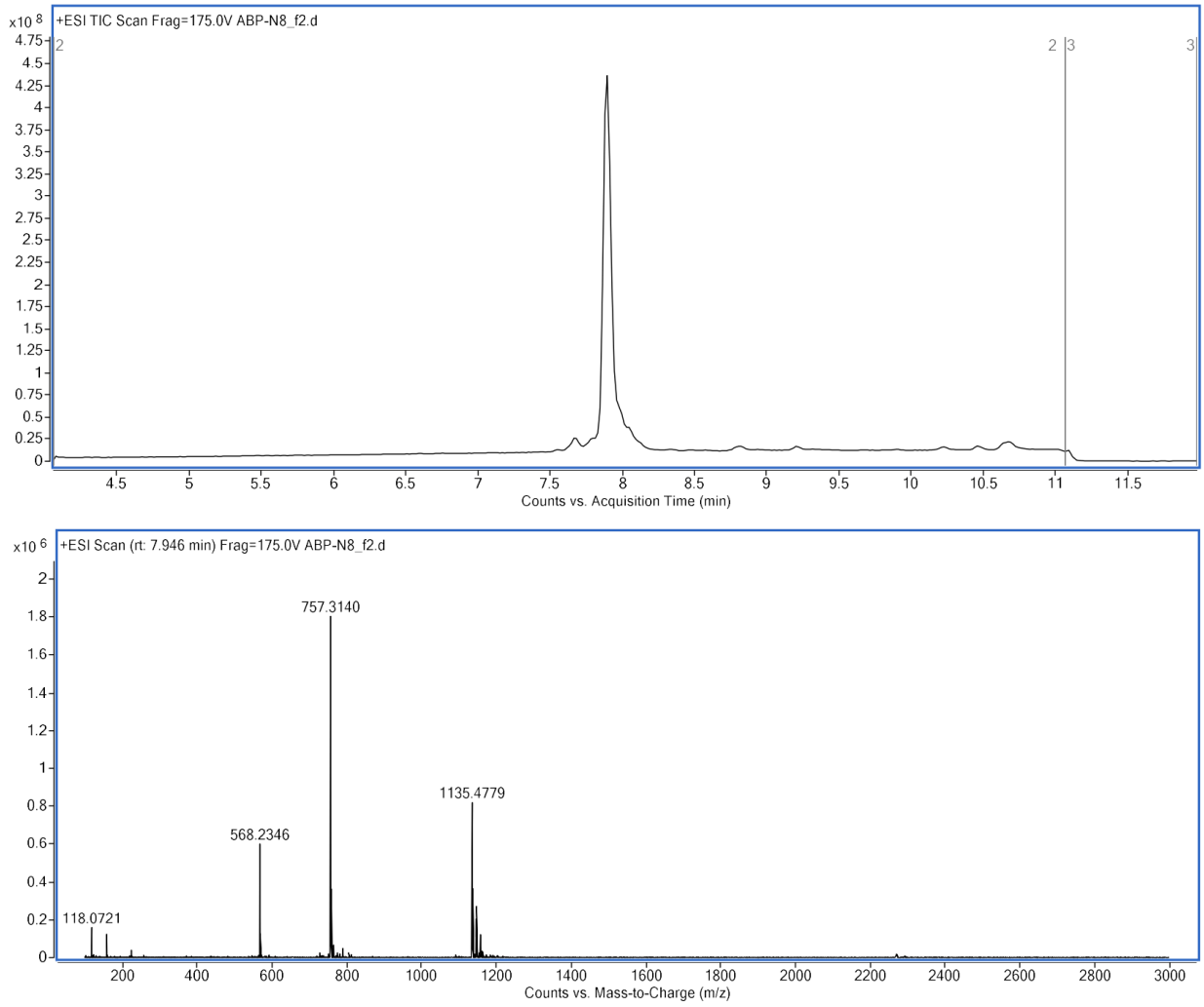
Observed: 2267.03 Da; Calculated: 2267.07 Da





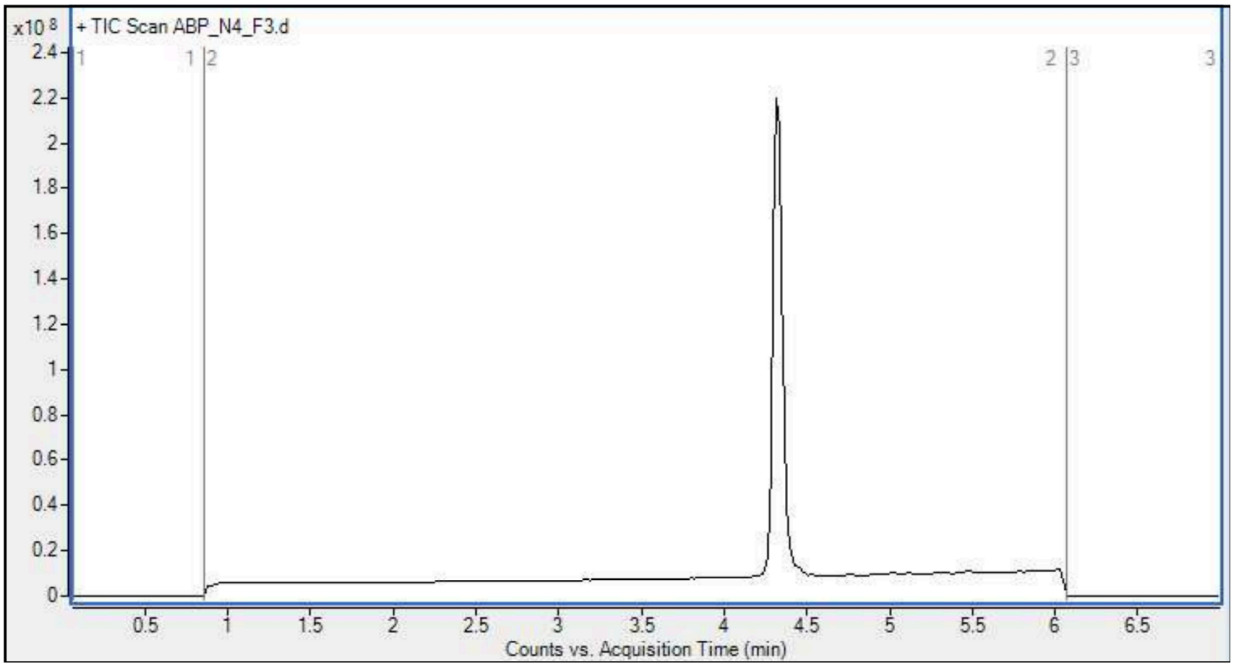
A2.6.1.5 Biotin-ABP-N8

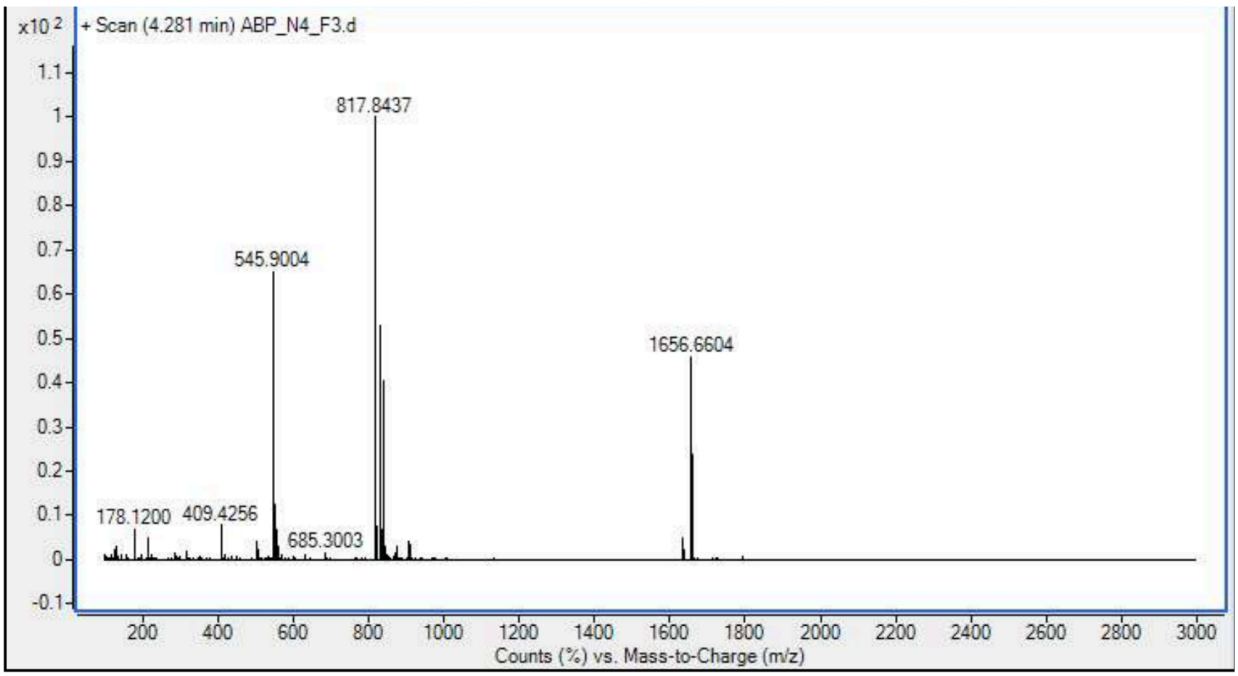
Observed: 2267.97 Da; Calculated: 2268.02 Da



A2.6.1.6 ABP-N4

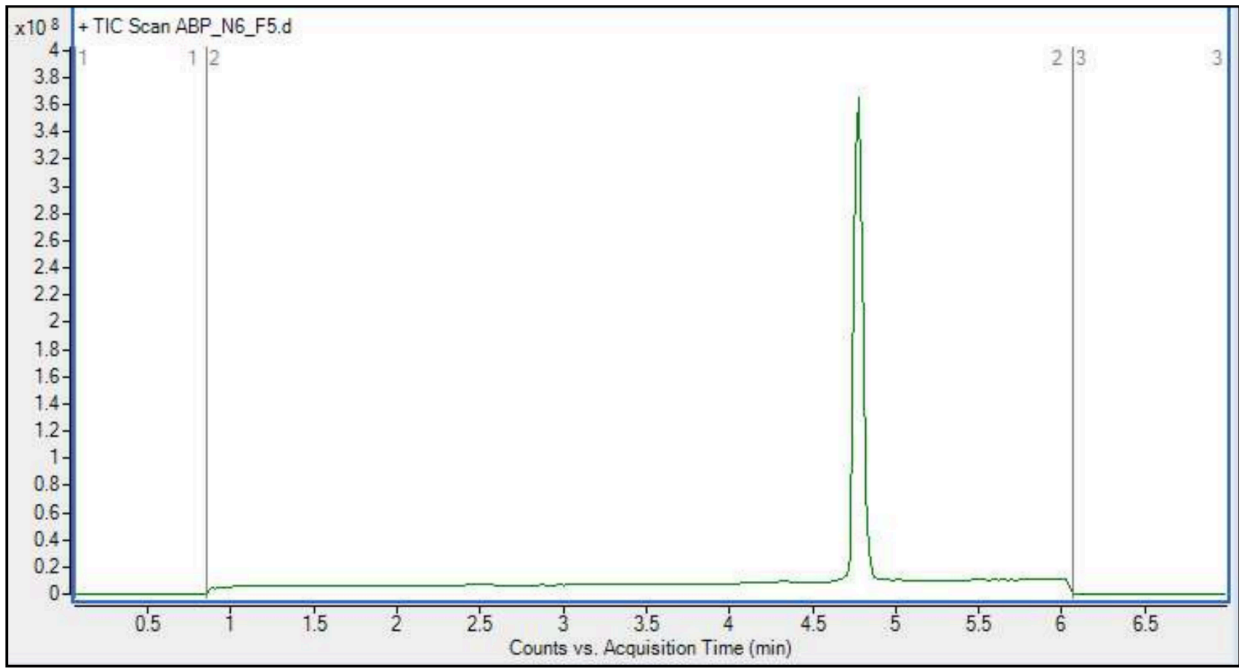
Observed: 1633.68 Da; Calculated: 1633.70 Da

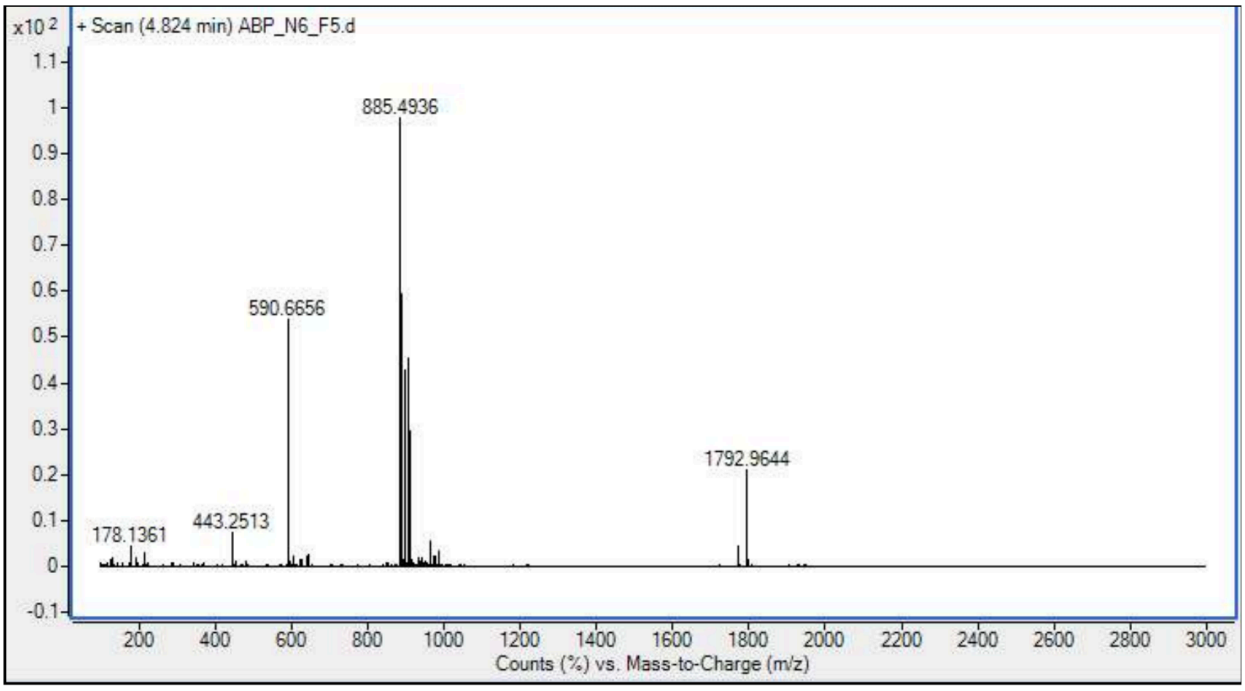




A2.6.1.7 ABP-N6

Observed: 1768.97 Da; Calculated: 1768.80 Da

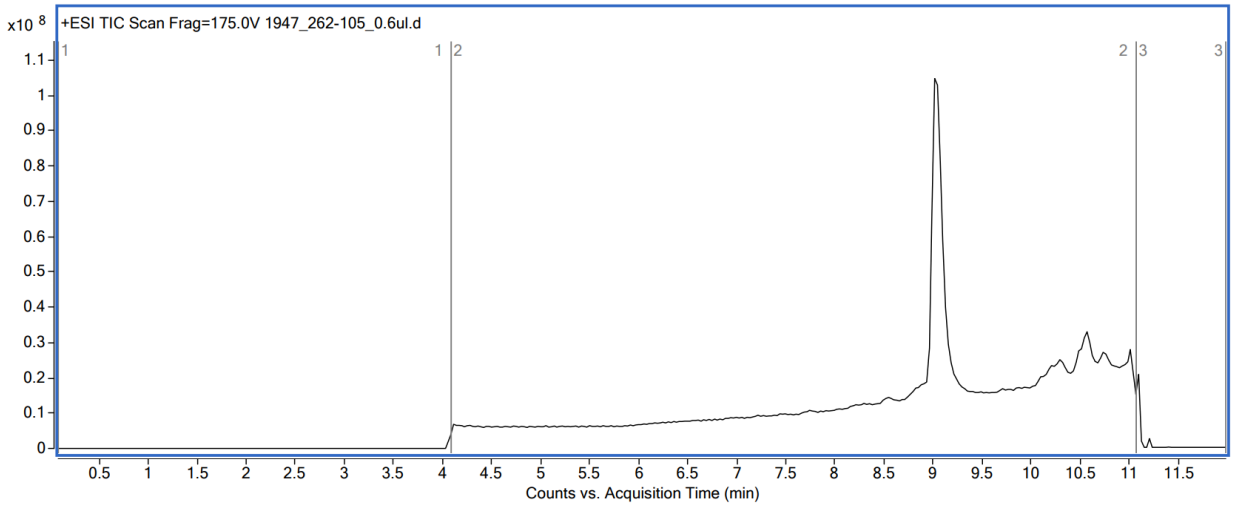


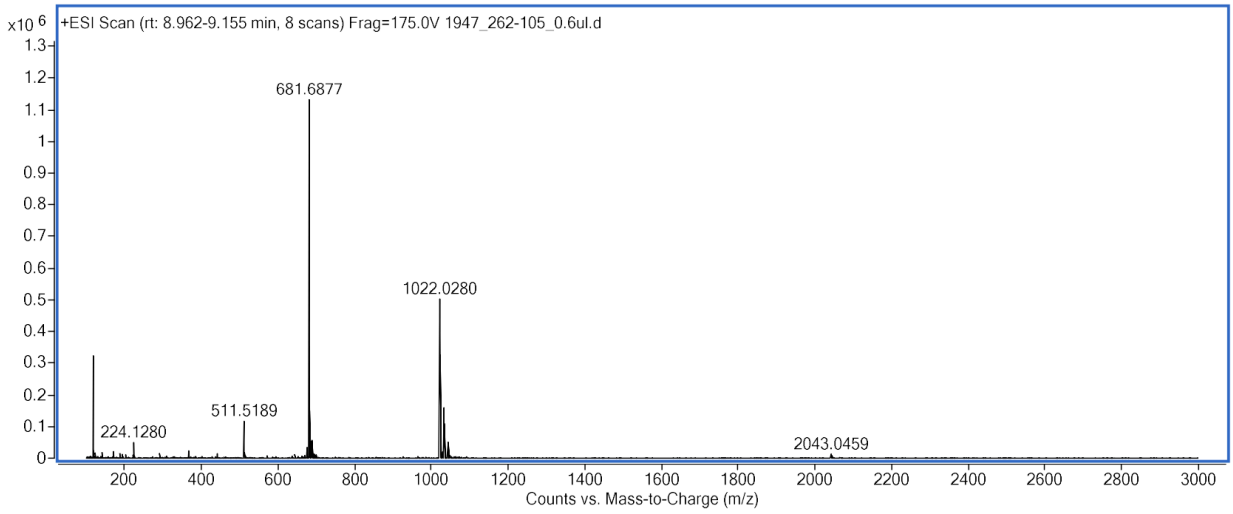


A2.6.2 LC-MS traces of purified canonical peptides

A2.6.2.1 Biotin-ABP-C1

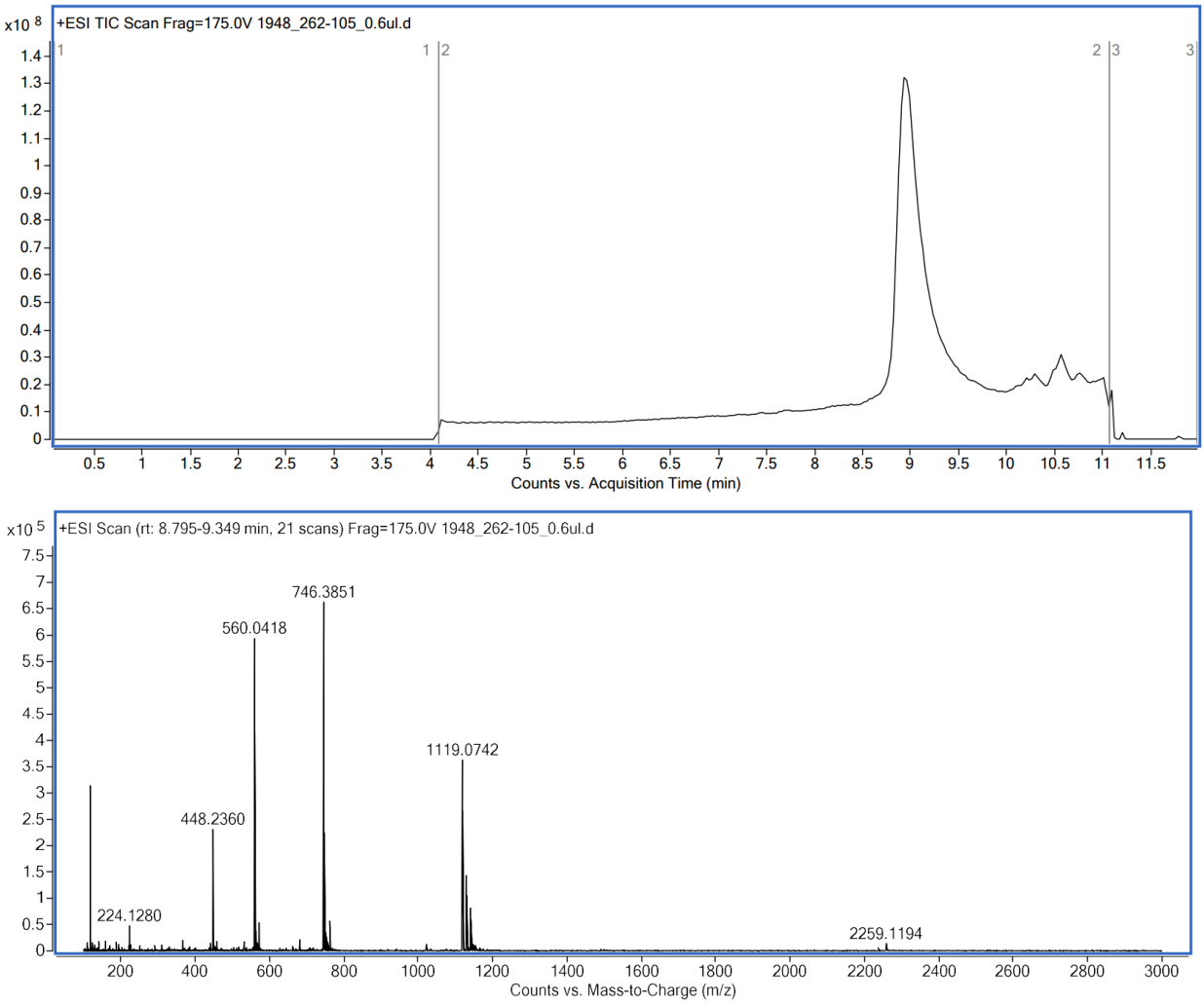
Observed: 2041.06 Da; Calculated: 2041.11 Da





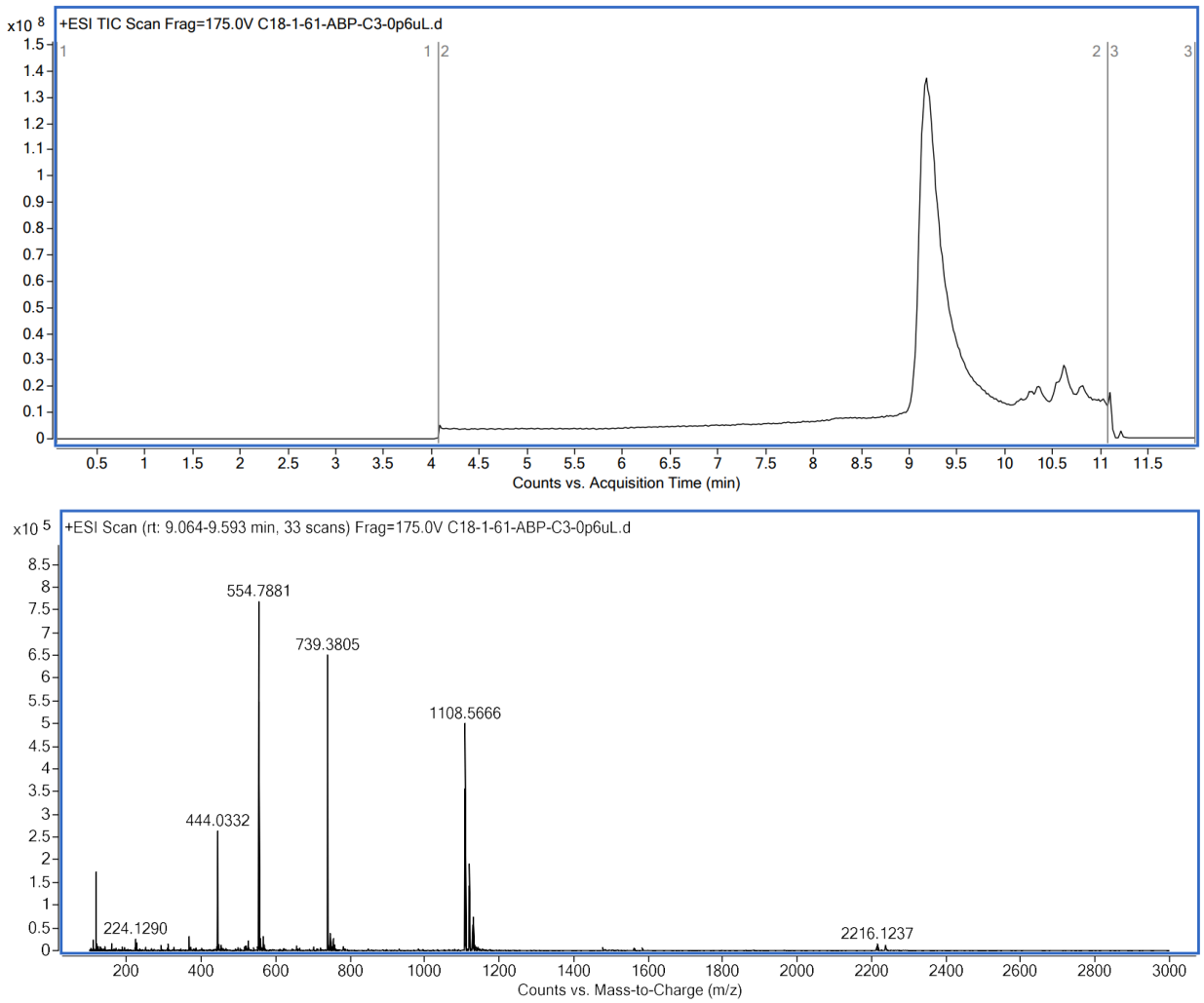
A2.6.2.2 Biotin-ABP-C2

Observed: 2235.15 Da; Calculated: 2235.19 Da



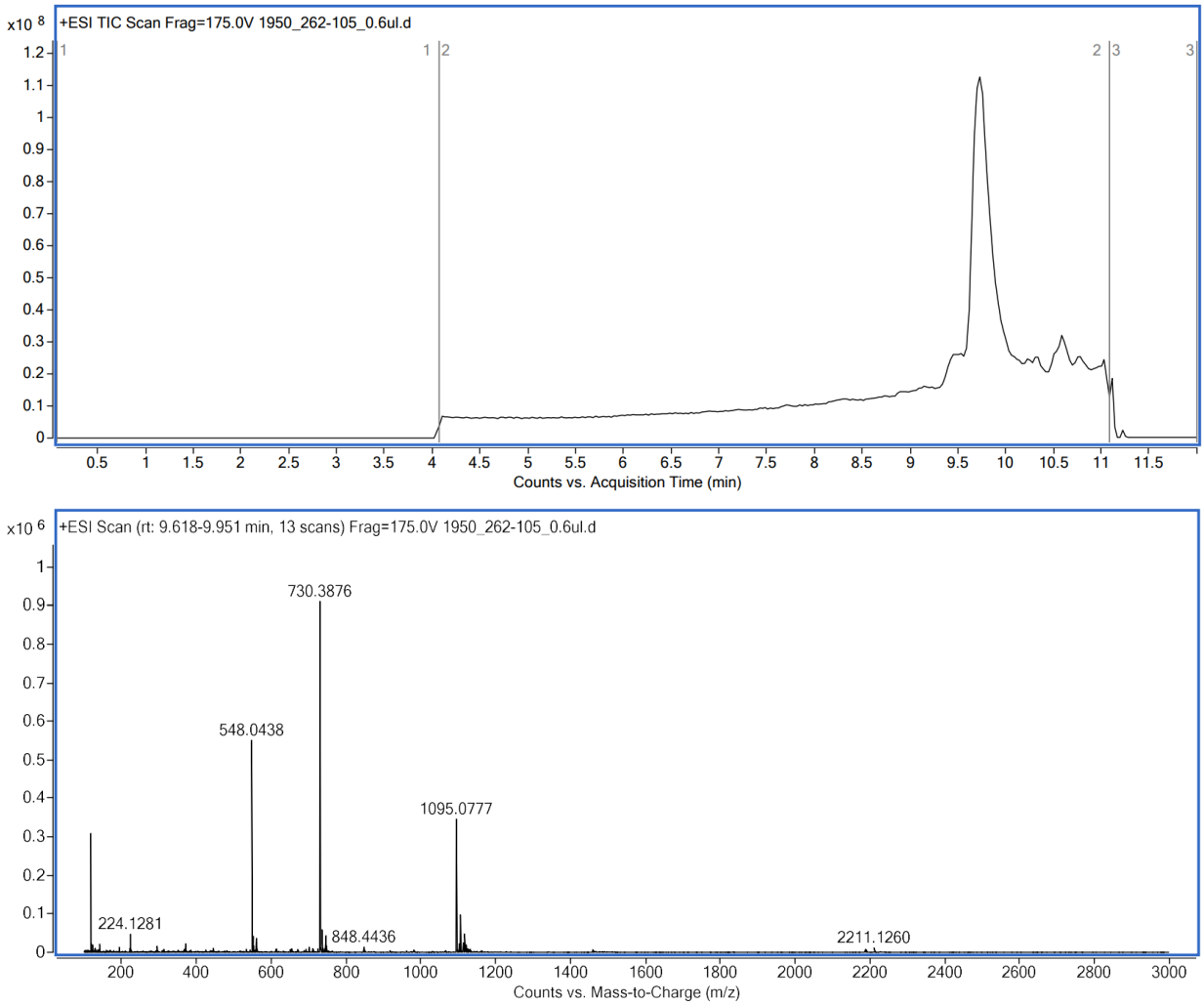
A2.6.2.3 Biotin-ABP-C3

Observed: 2214.14 Da; Calculated: 2214.15 Da



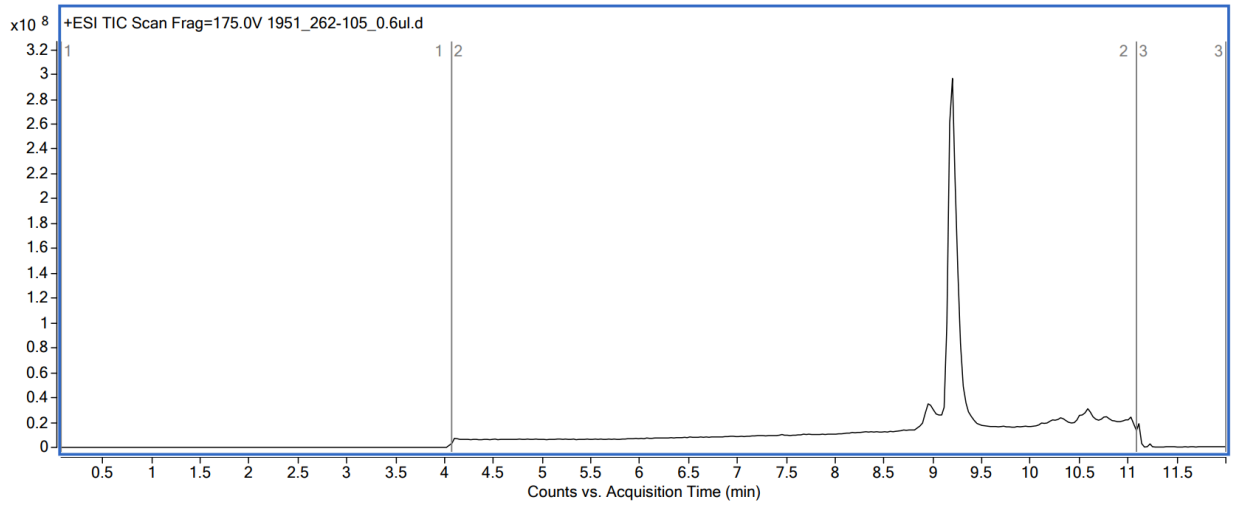
A2.6.2.4 Biotin-ABP-C4

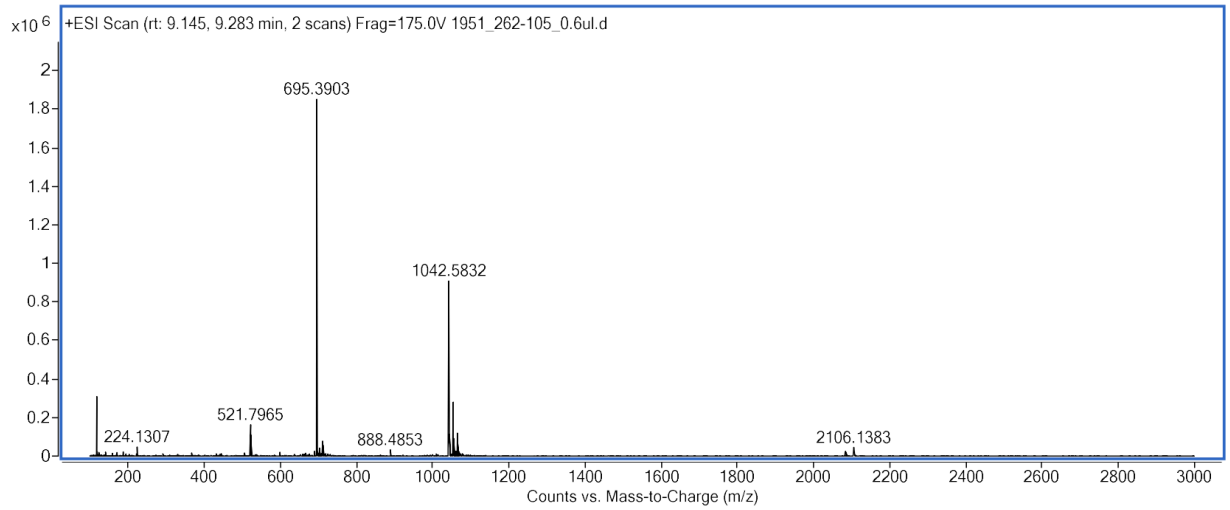
Observed: 2187.15 Da; Calculated: 2187.17 Da



A2.6.2.5 Biotin-ABP-C5

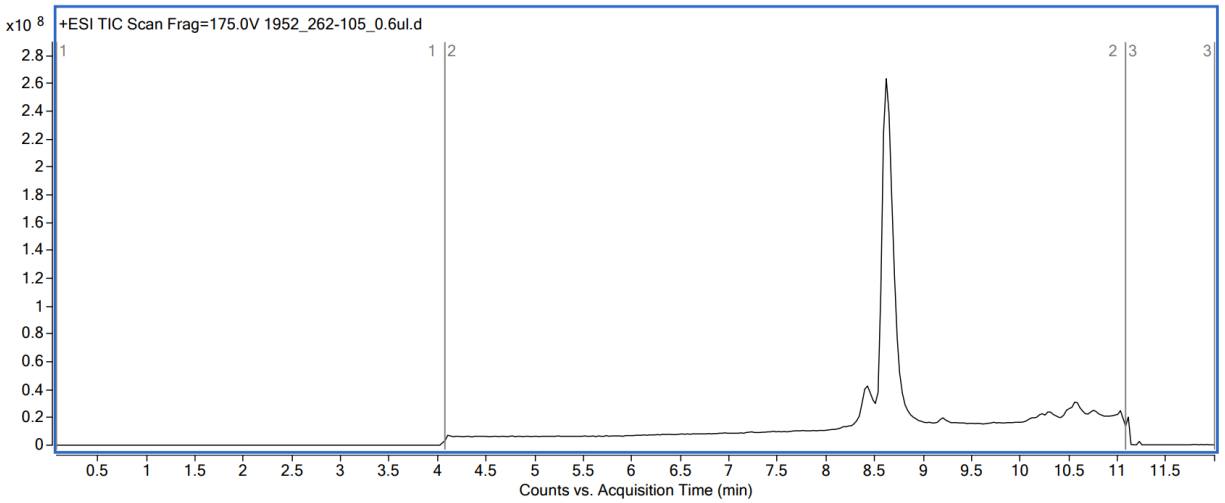
Observed: 2082.16 Da; Calculated: 2082.18 Da

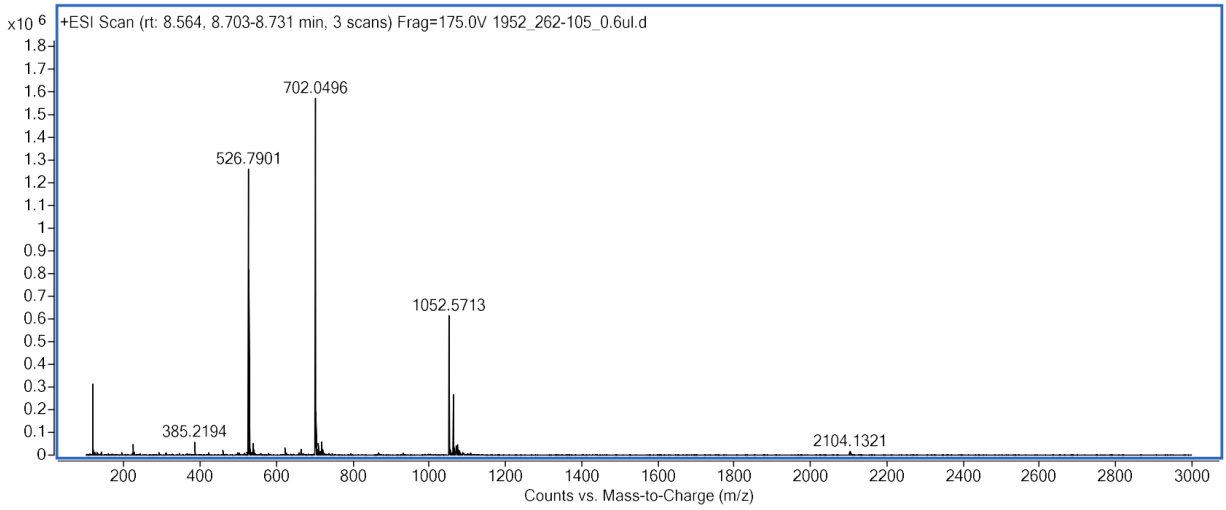




A2.6.2.6 Biotin-ABP-C6

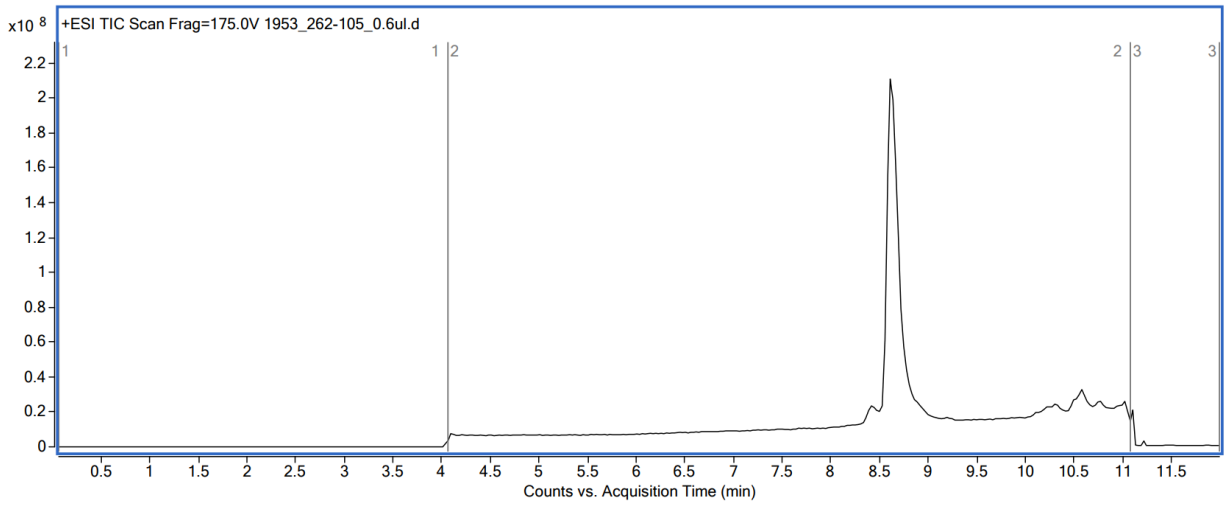
Observed: 2102.14 Da; Calculated: 2102.16 Da

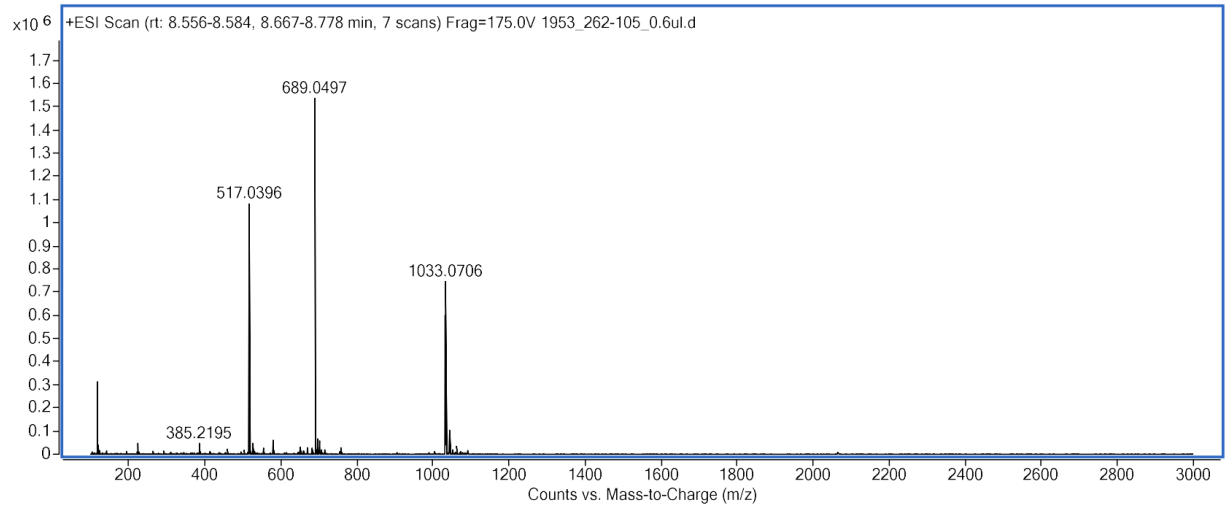




A2.6.2.78 Biotin-ABP-C7

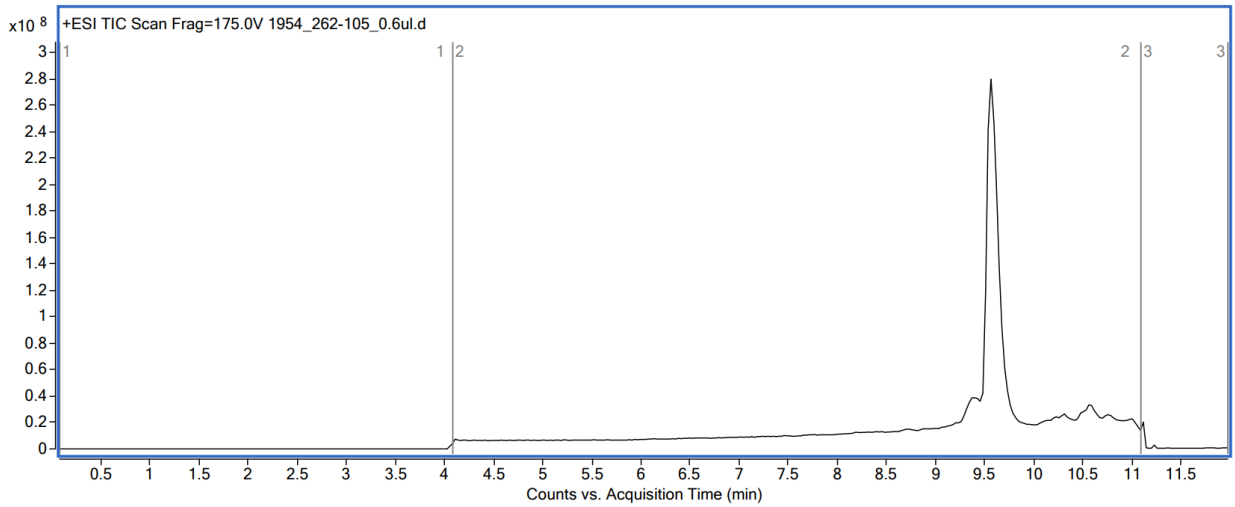
Observed: 2063.14 Da; Calculated: 2063.16 Da

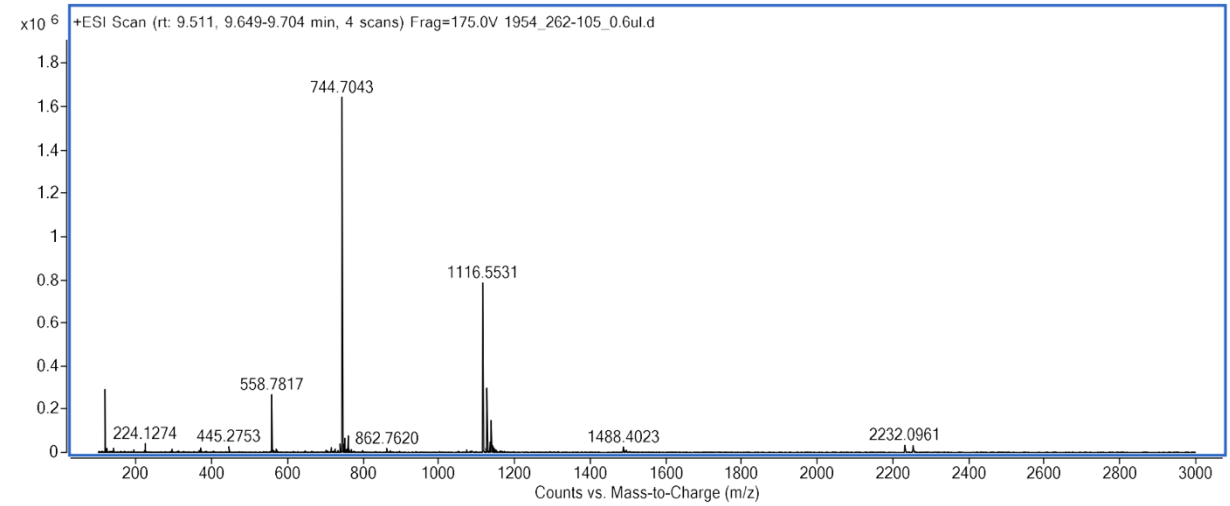




A2.6.2.8 Biotin-ABP-C8

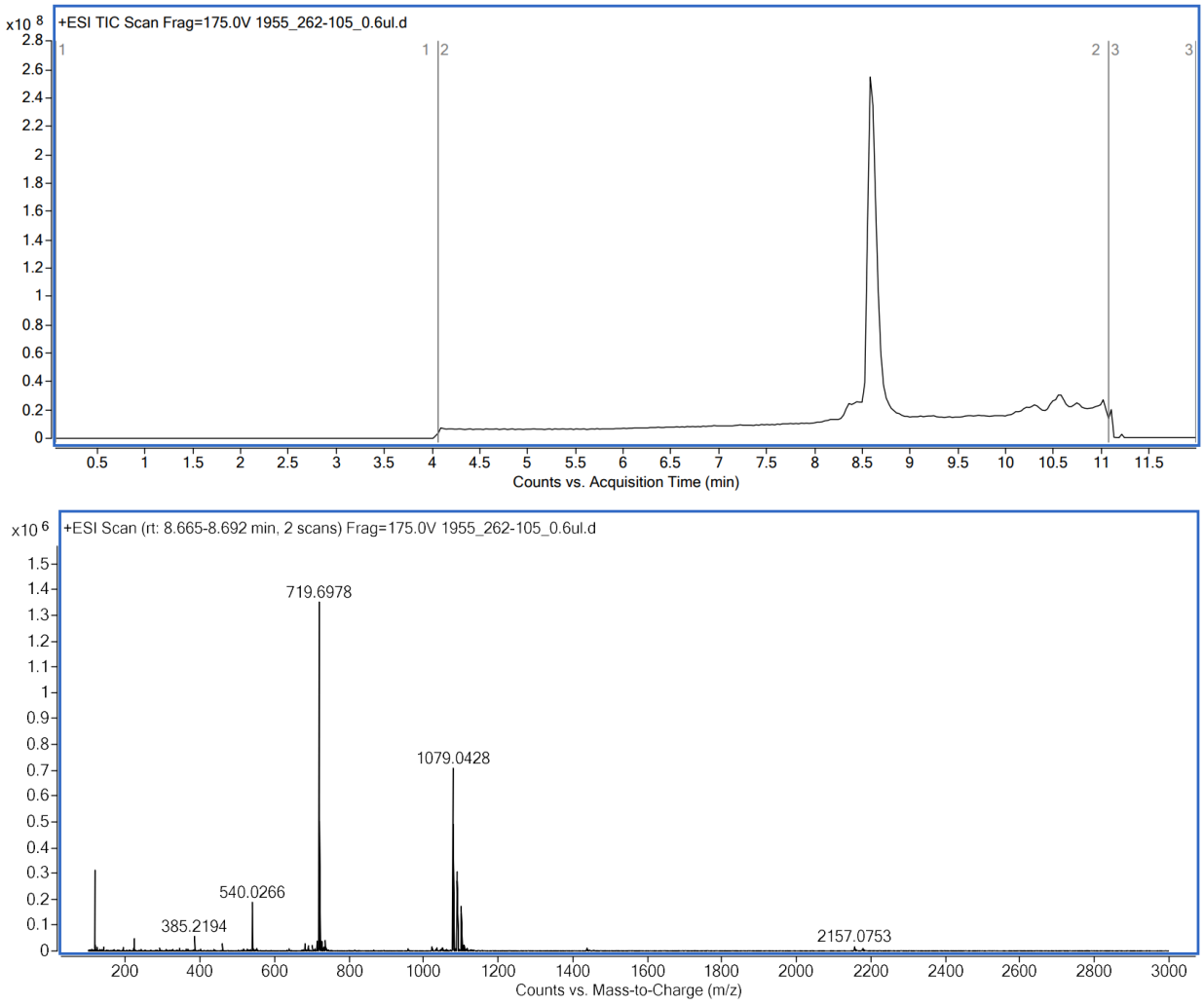
Observed: 2230.11 Da; Calculated: 2230.13 Da





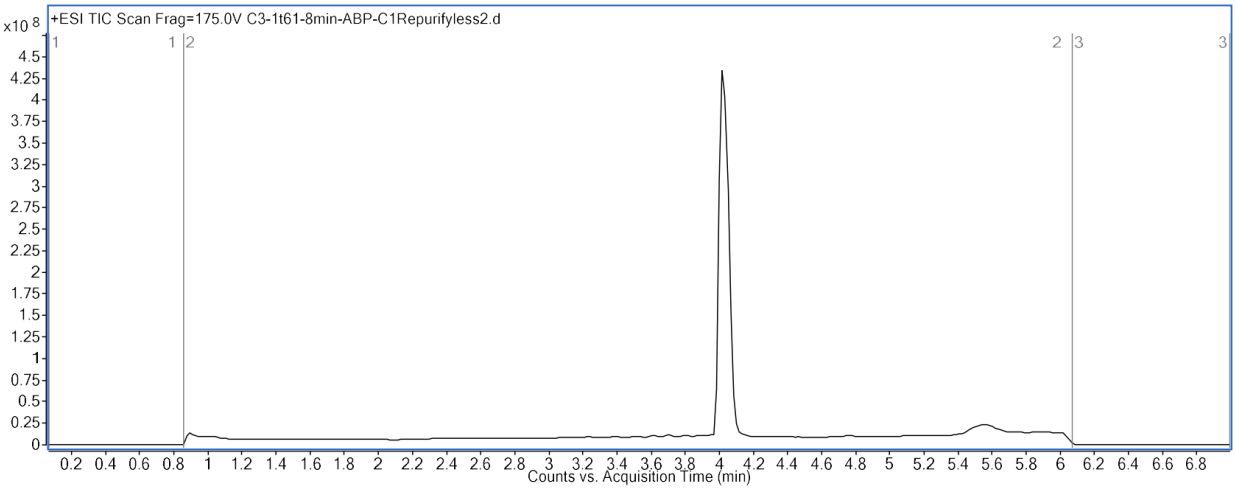
A2.6.2.9 Biotin-ABP-C9

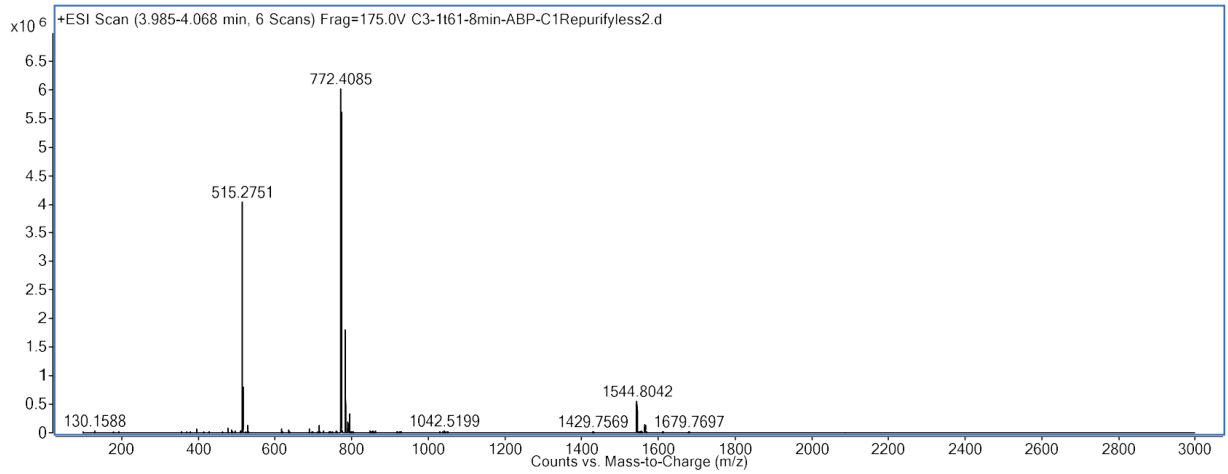
Observed: 2155.09 Da; Calculated: 2155.12 Da



A2.6.2.10 ABP-C1

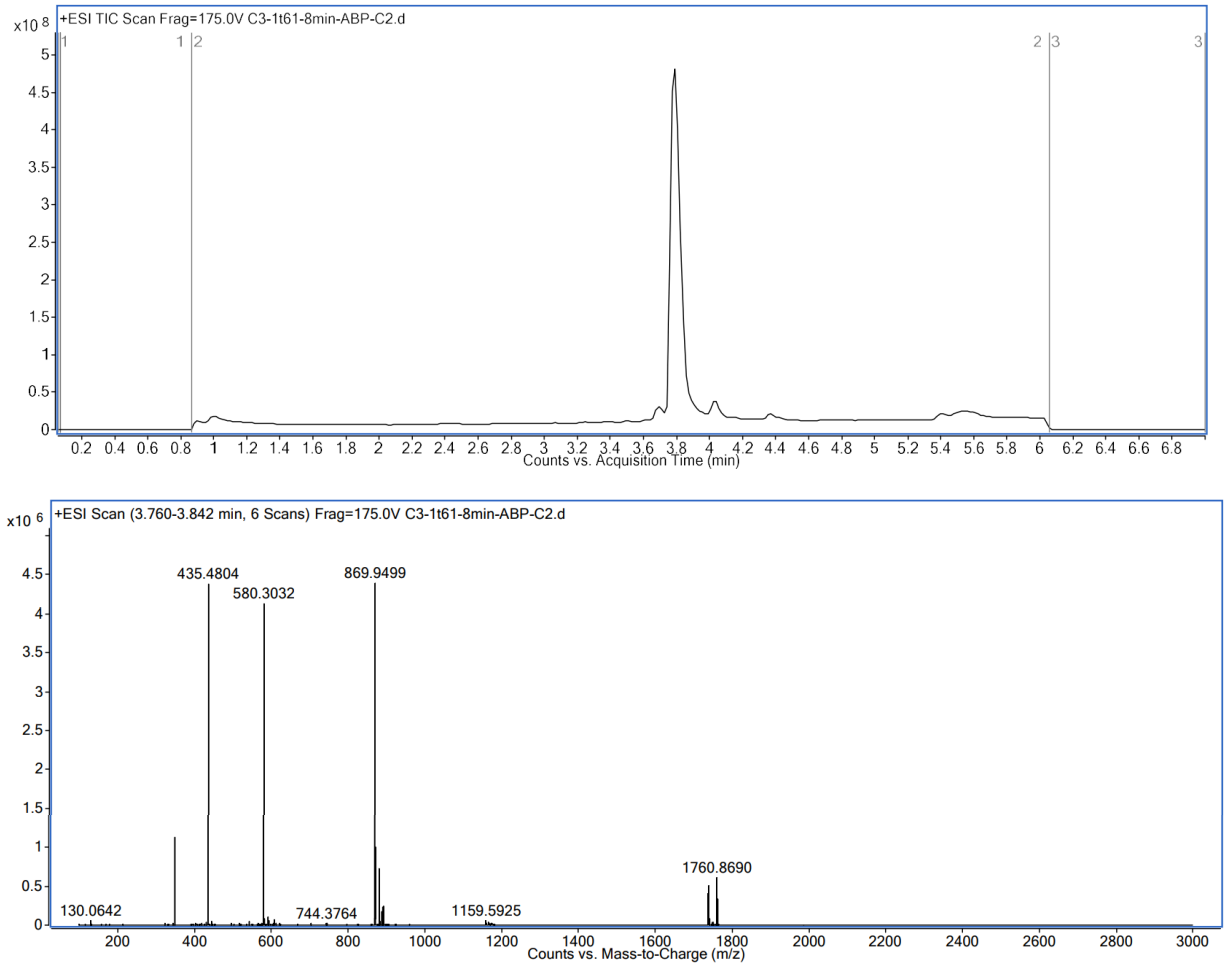
Observed: 1543.80 Calculated [M+H]: 1543.82





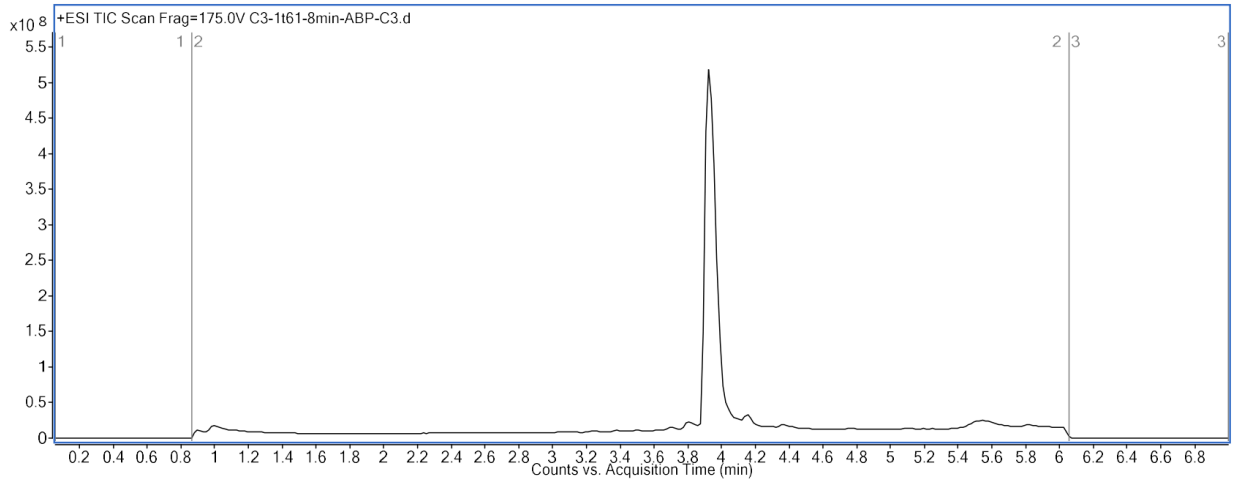
A2.6.2.11 ABP-C2

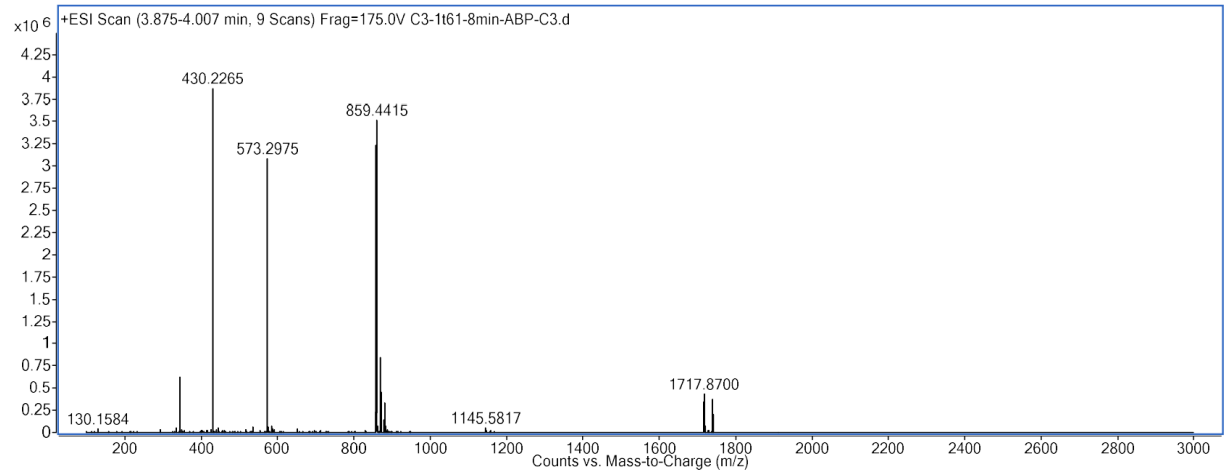
Observed: 1737.89 Calculated [M+H]: 1737.91



A2.6.2.12 ABP-C3

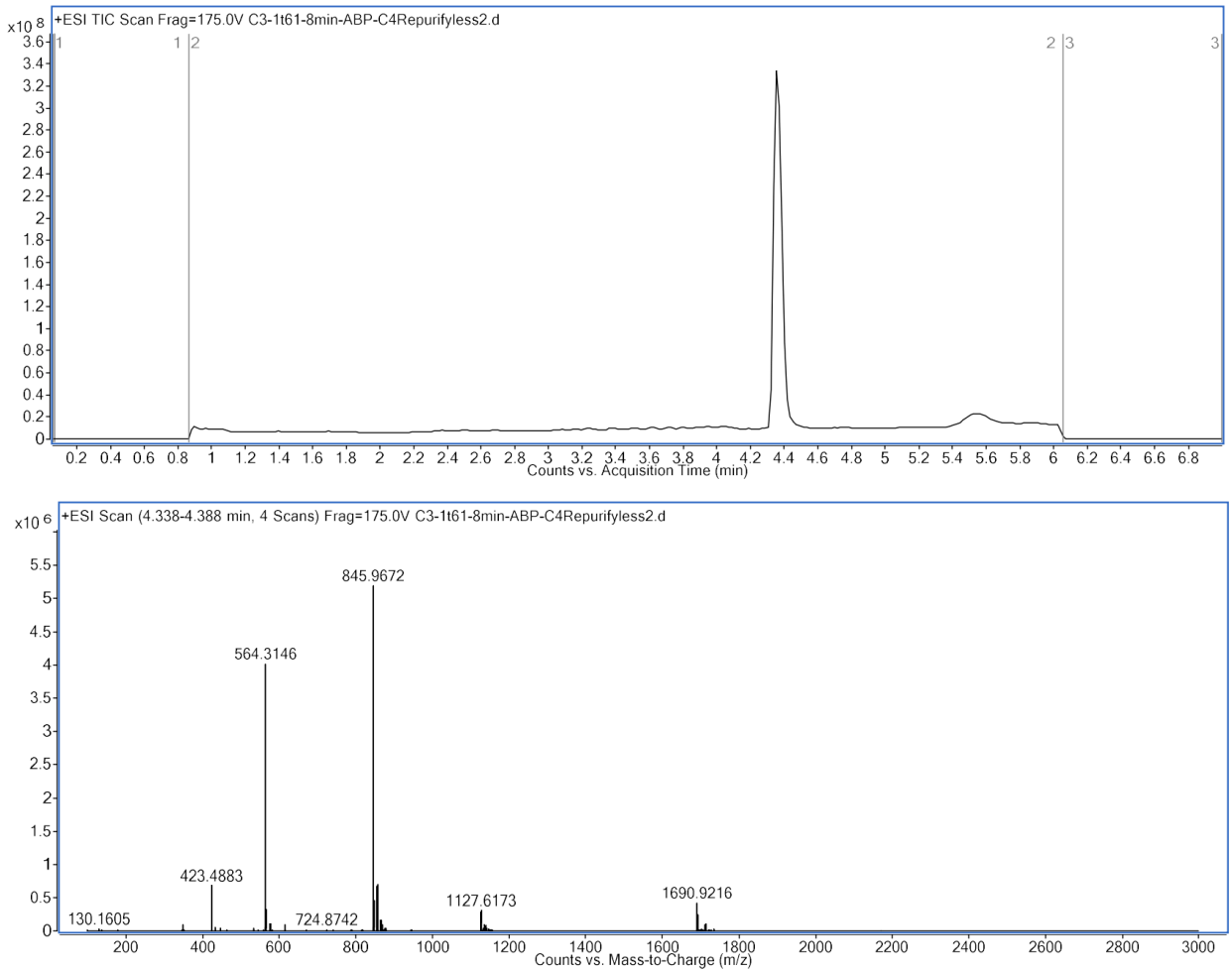
Observed: 1716.87 Calculated [M+H]: 1716.89





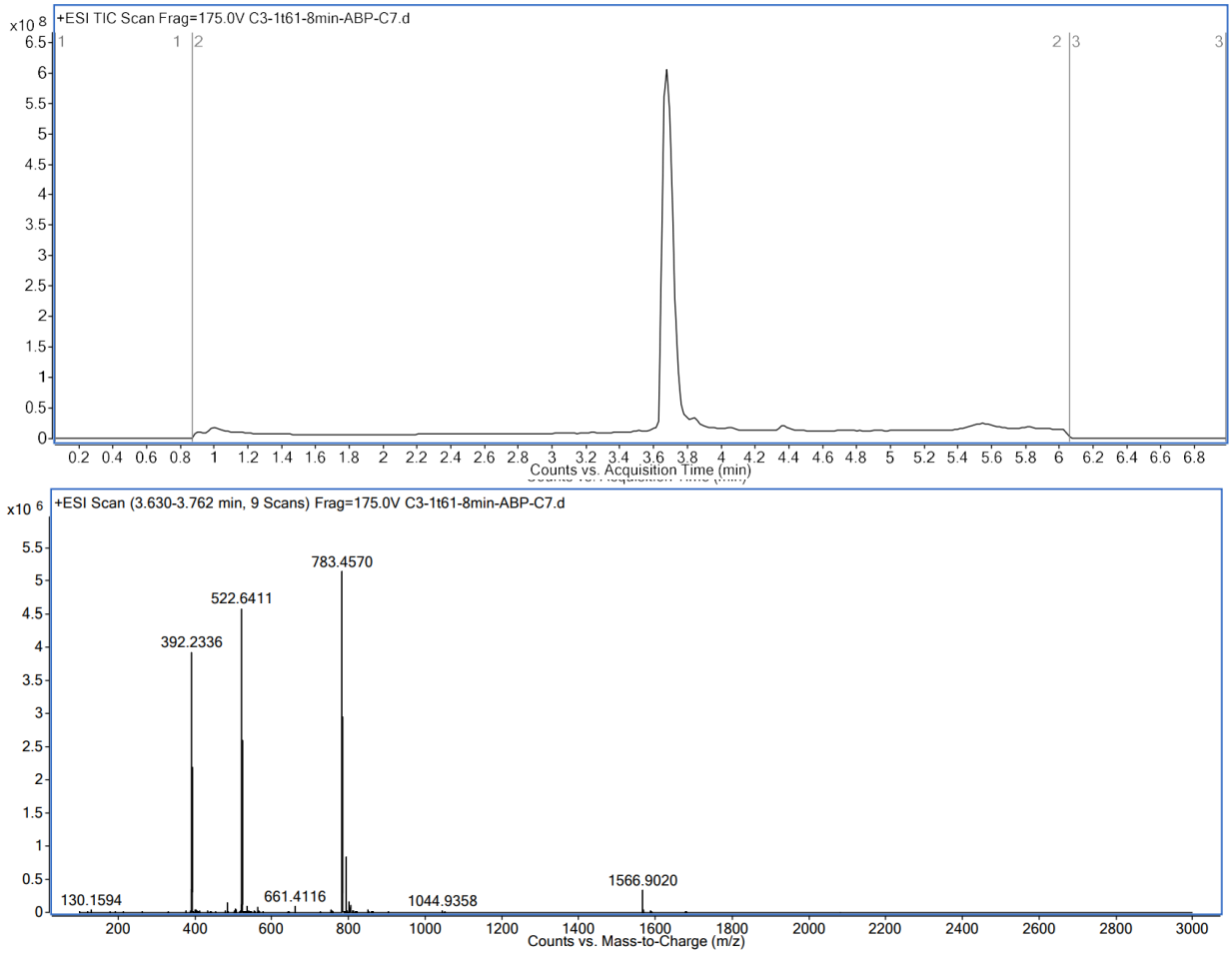
A2.6.2.13 ABP-C4

Observed: 1689.92 Calculated [M+H]: 1689.92



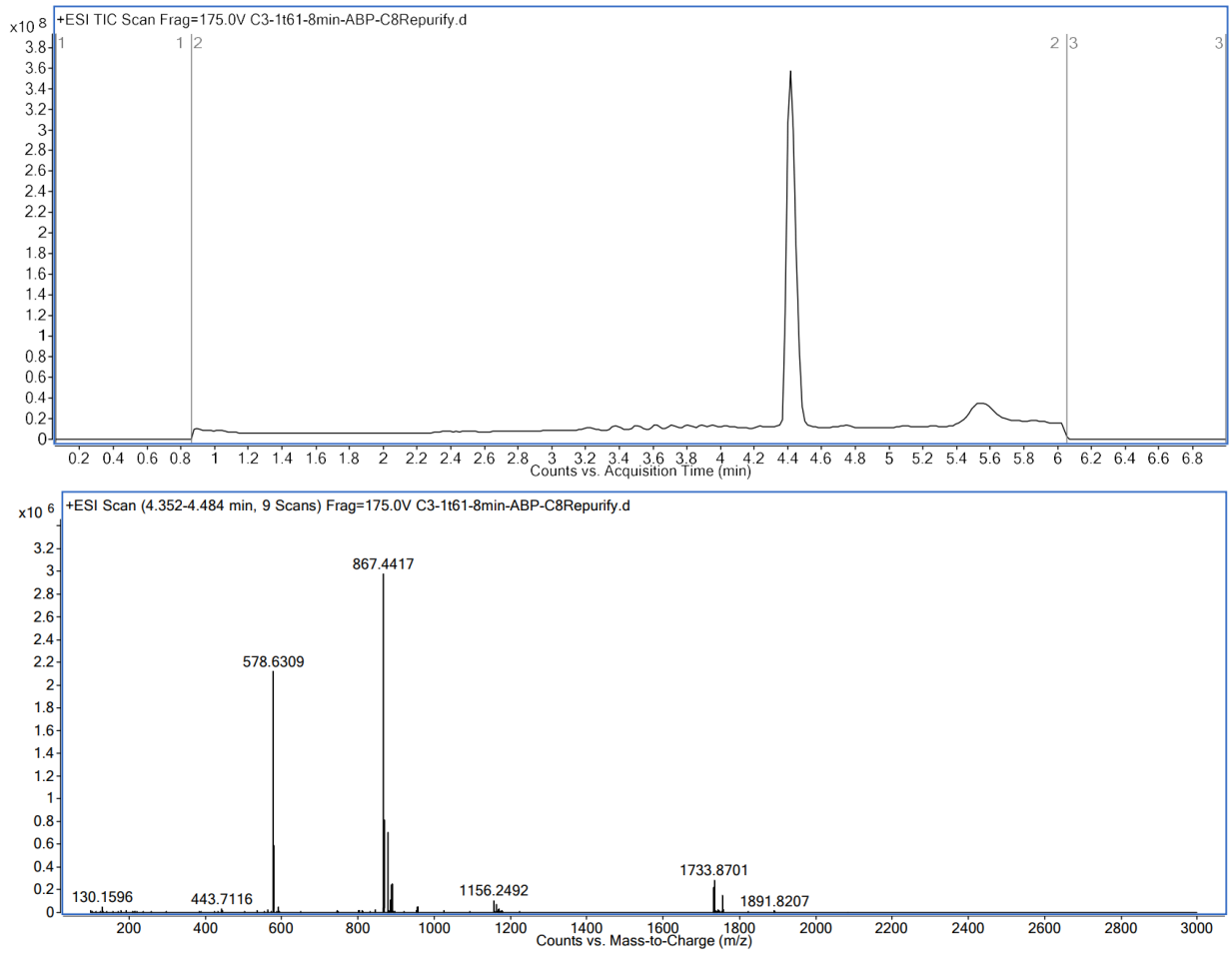
A2.6.2.14 ABP-C7

Observed: 1565.90 Calculated [M+H]: 1565.91



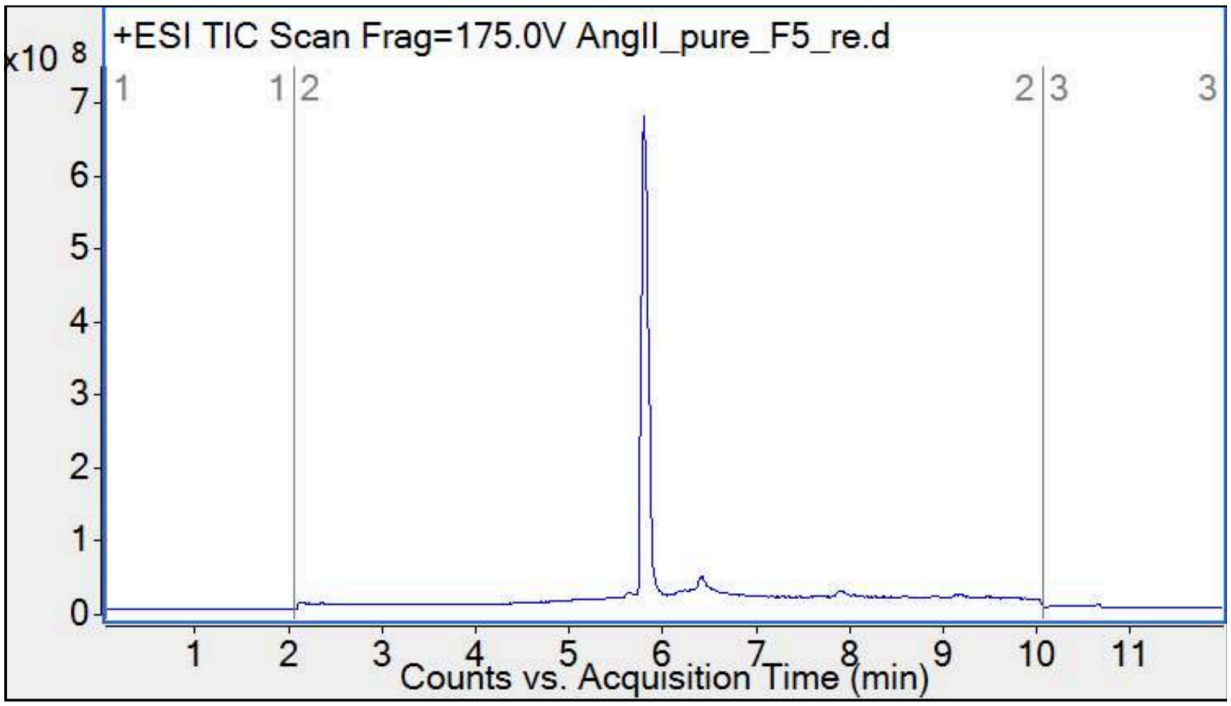
A2.6.2.15 ABP-C8

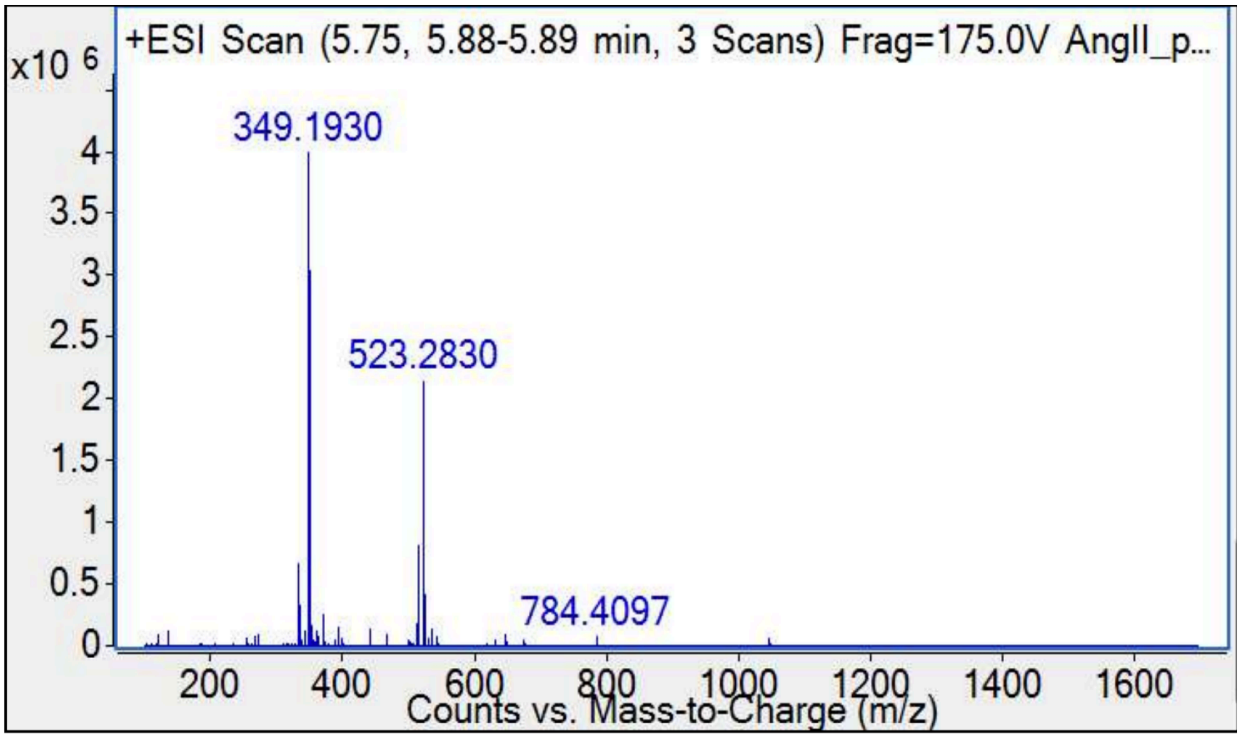
Observed: 1732.87 Calculated [M+H]: 1732.88



A2.6.2.16 AngII

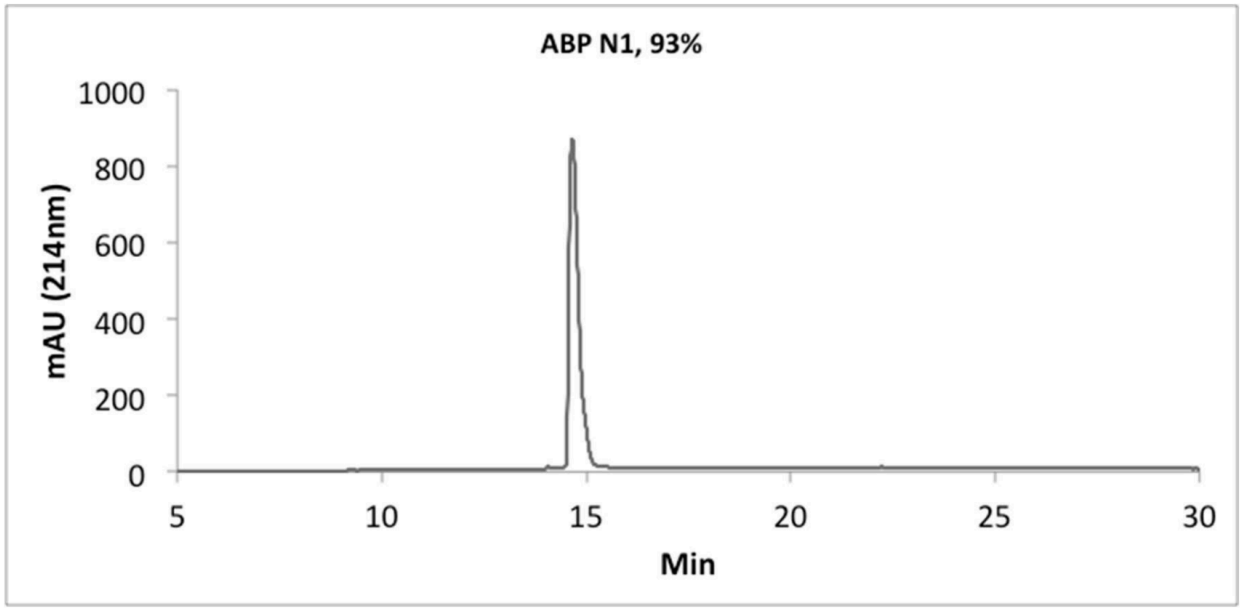
Observed: 1044.57 Da; Calculated: 1044.54 Da

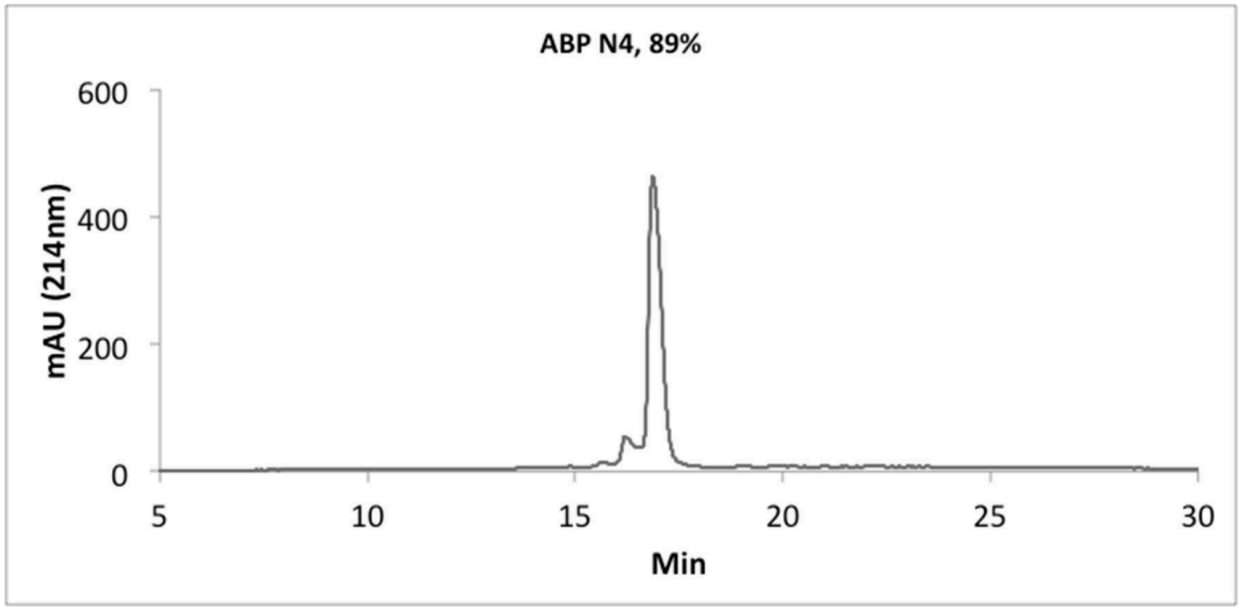


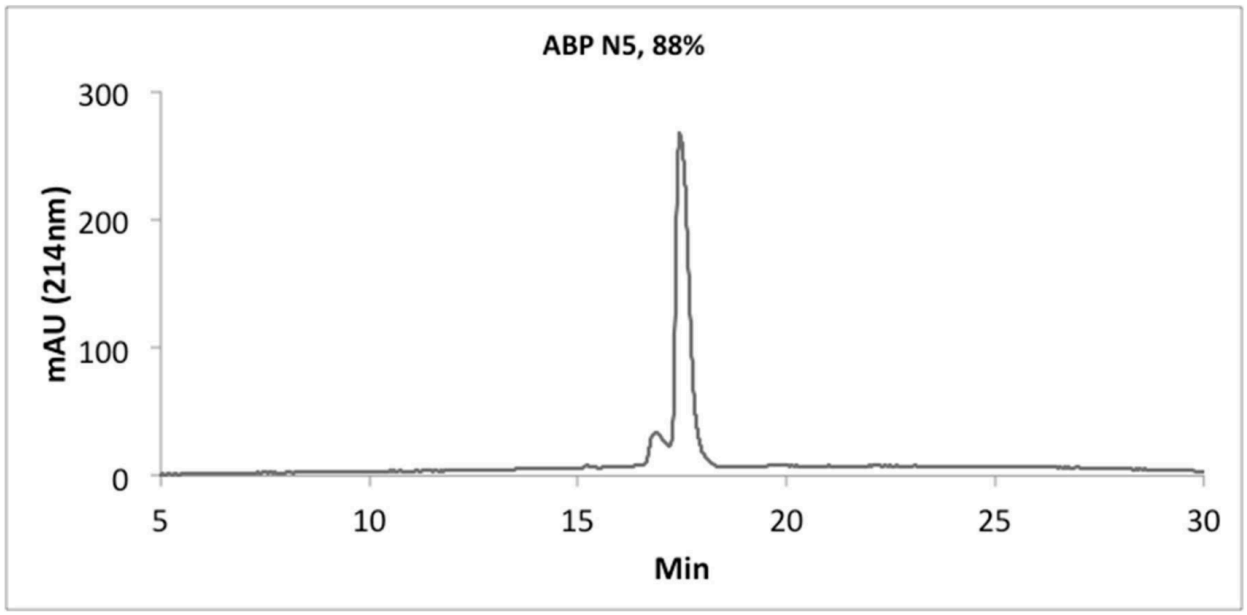


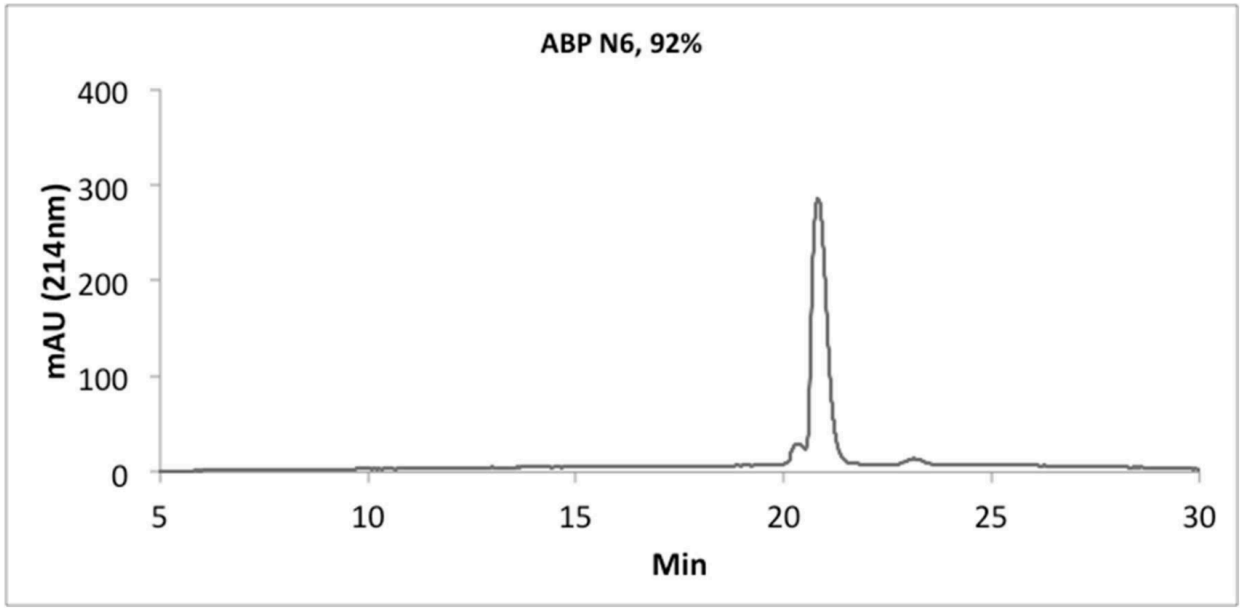
A2.7 Analytical HPLC data

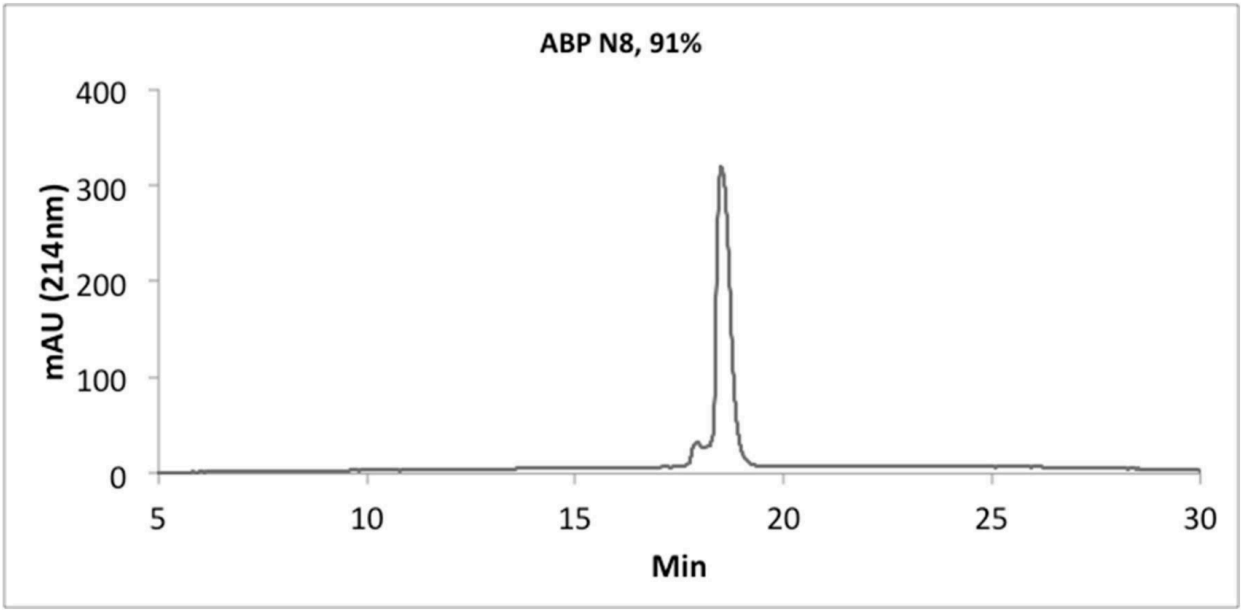
A2.7.1 HPLC trace of noncanonical peptides (biotinylated)



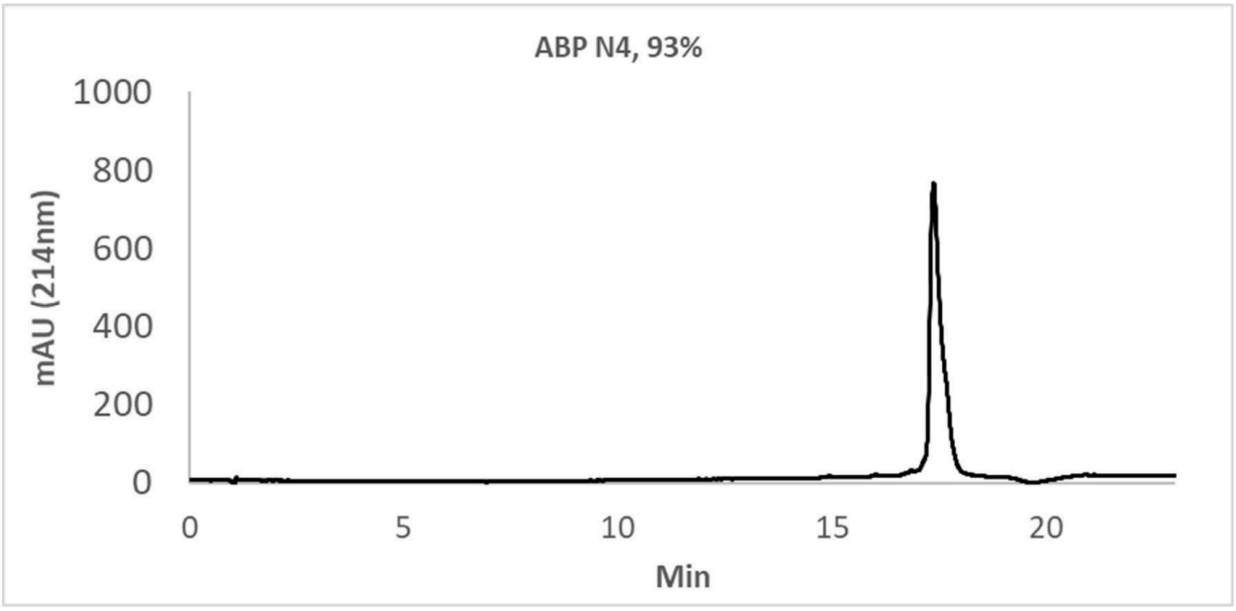


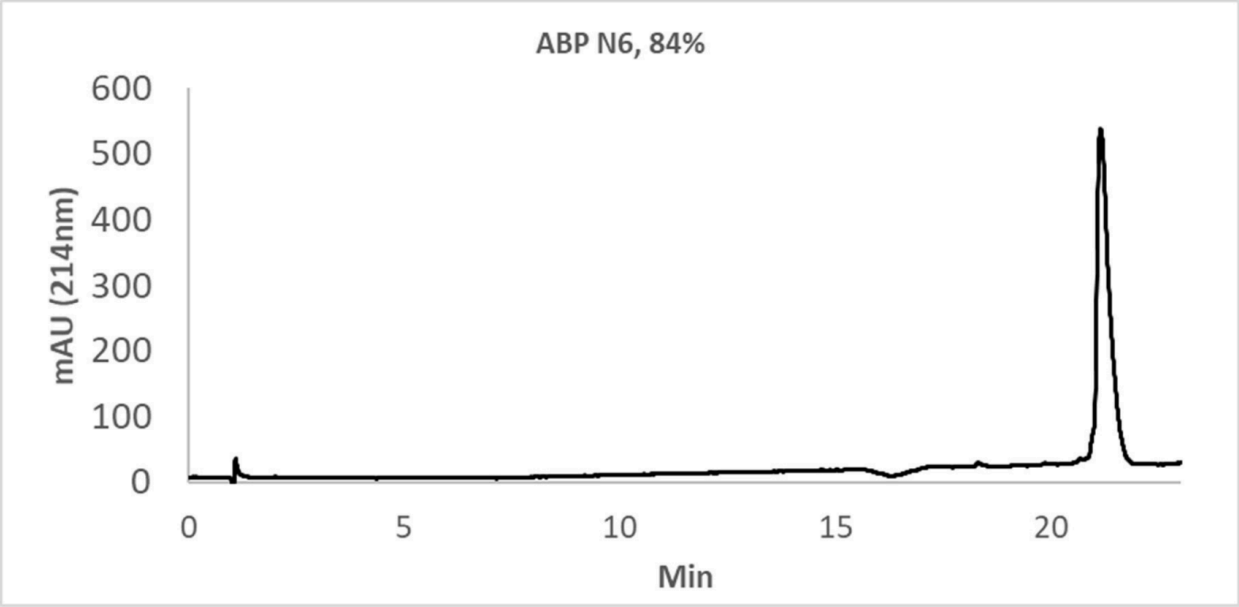

ABP N5, 88%
mAU (214nm)
300
200
100
0
5
10
15
20
25
30
Min


ABP N6, 92%
mAU (214nm)
400
300
200
100
0
5
10
15
20
25
30
Min

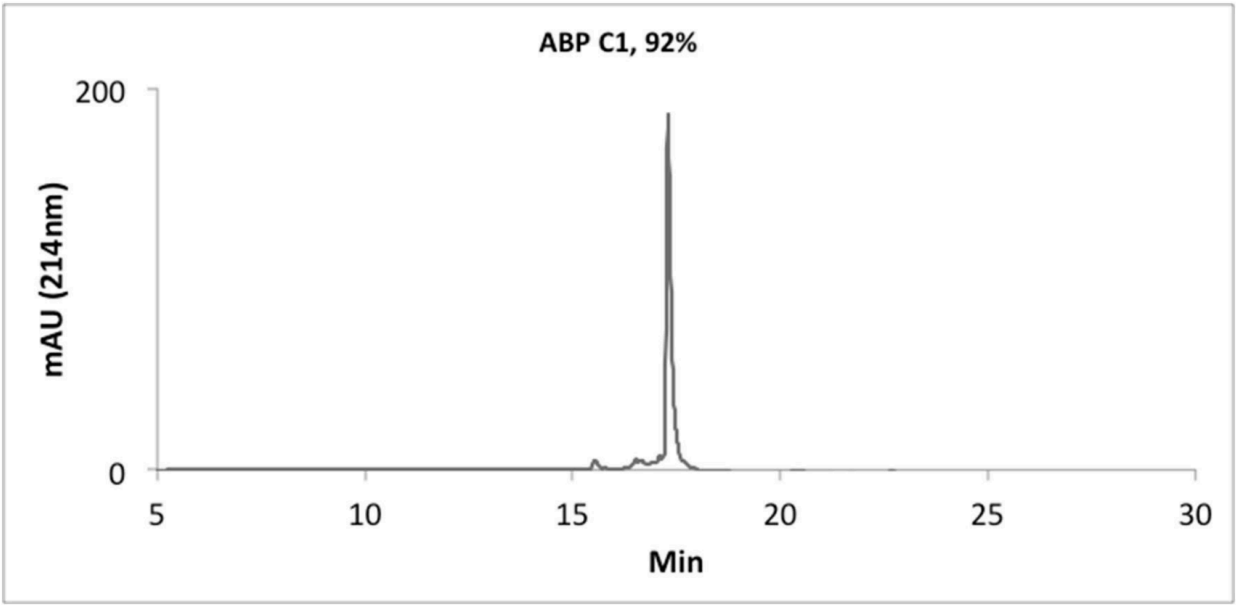


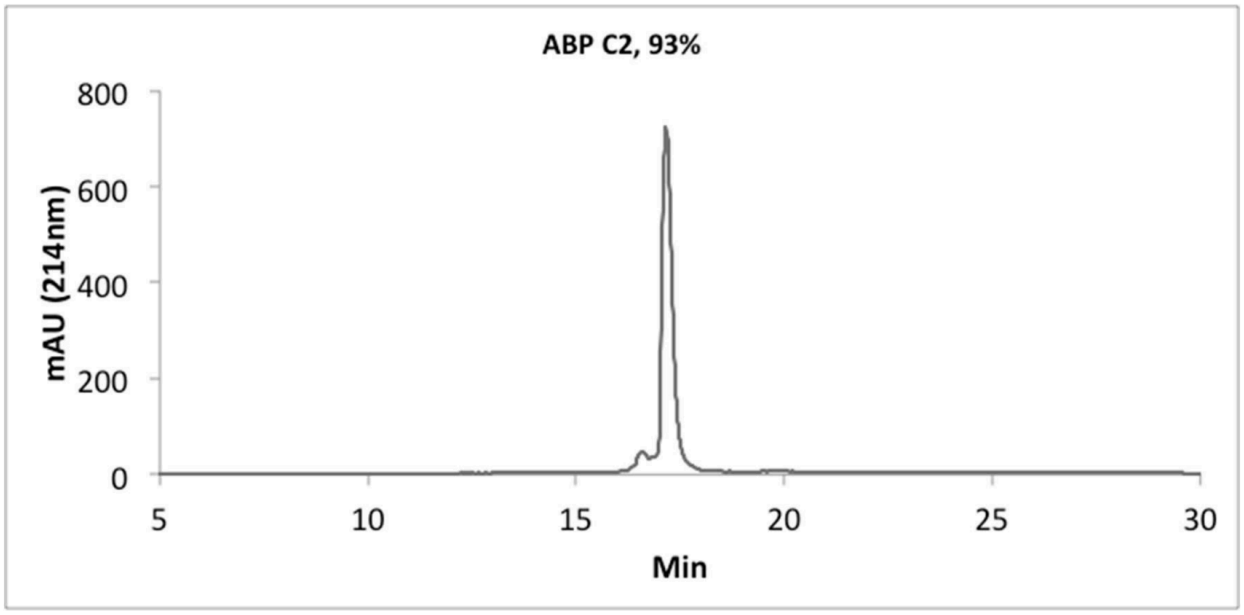
A2.7.2 HPLC trace of noncanonical peptides (nonbiotinylated)

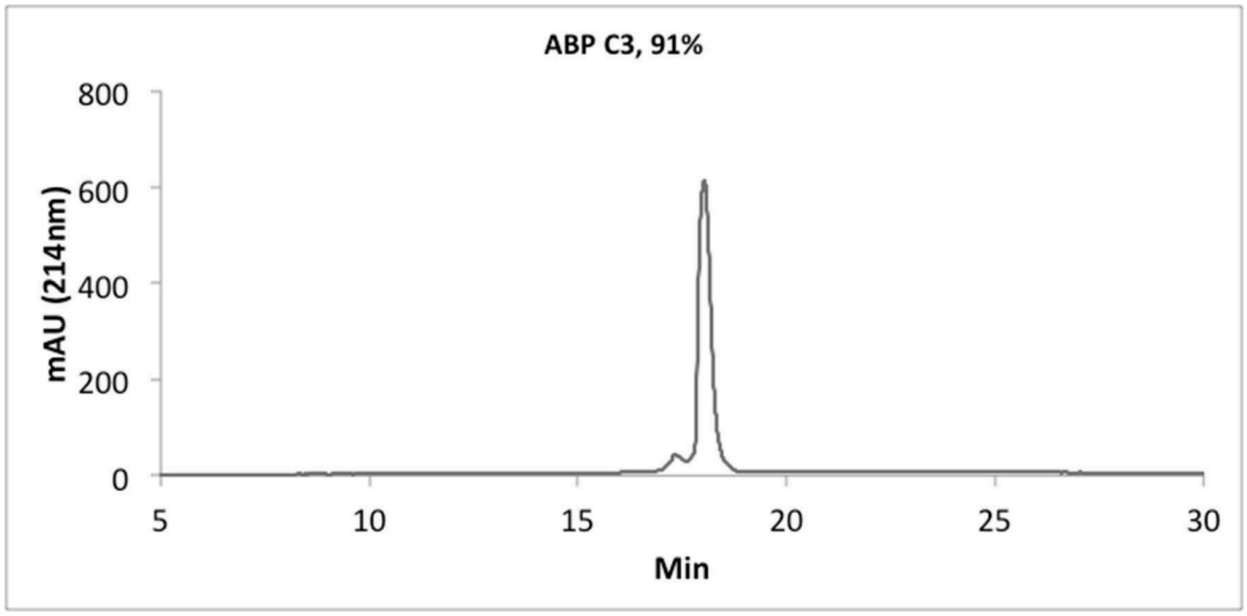


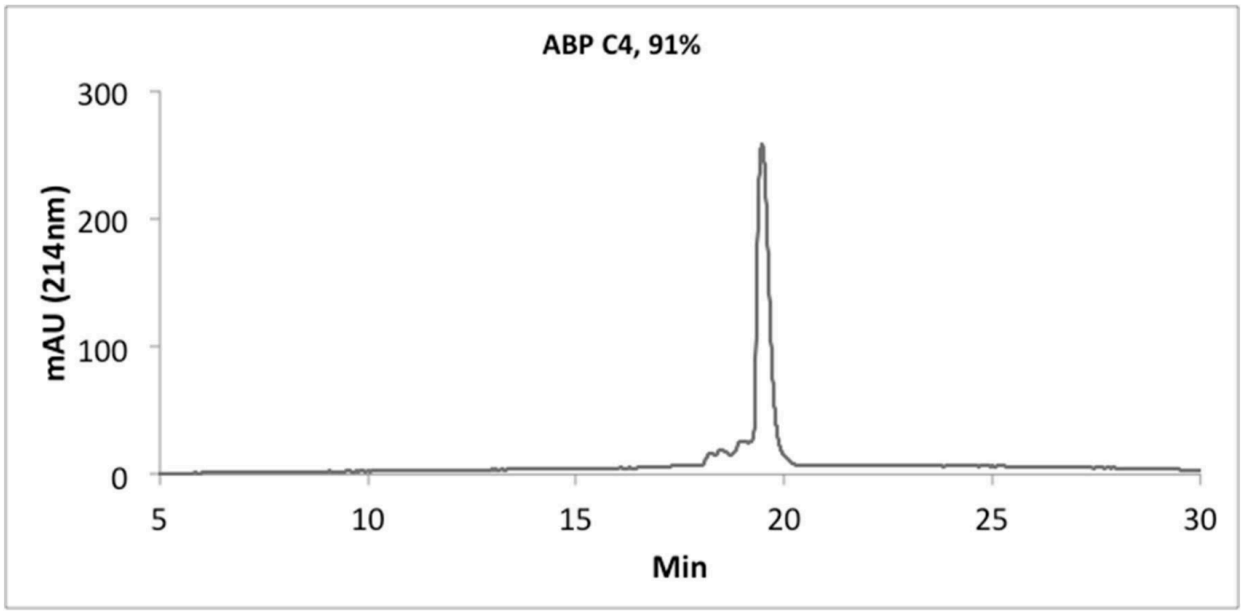


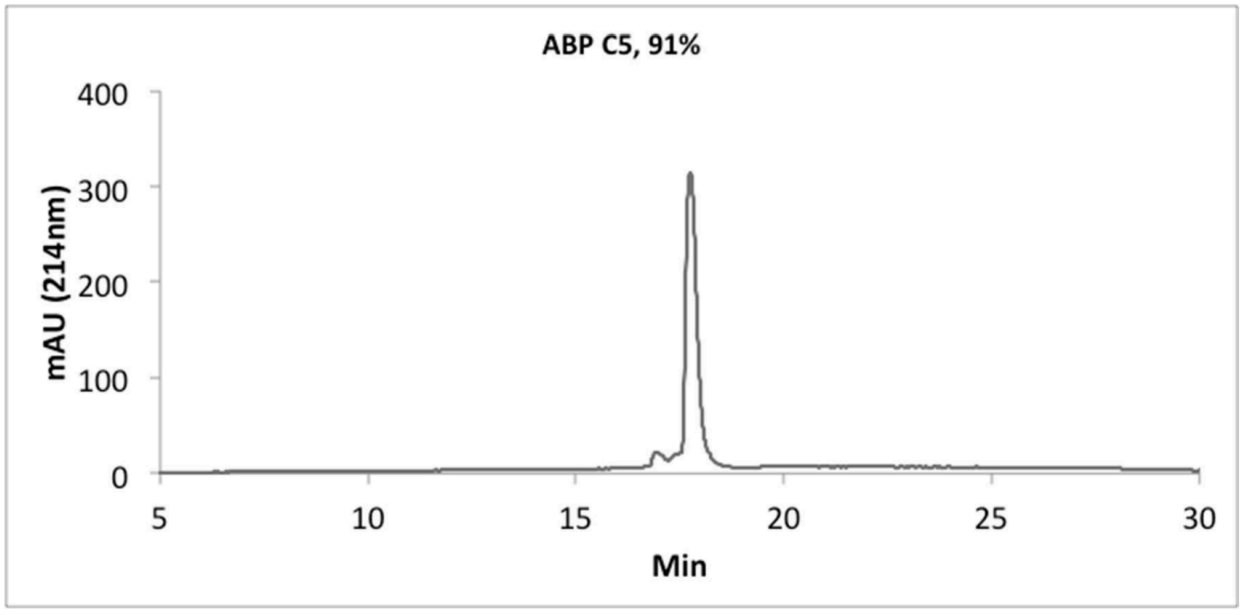
A2.7.3 HPLC trace of canonical peptides (biotinylated)

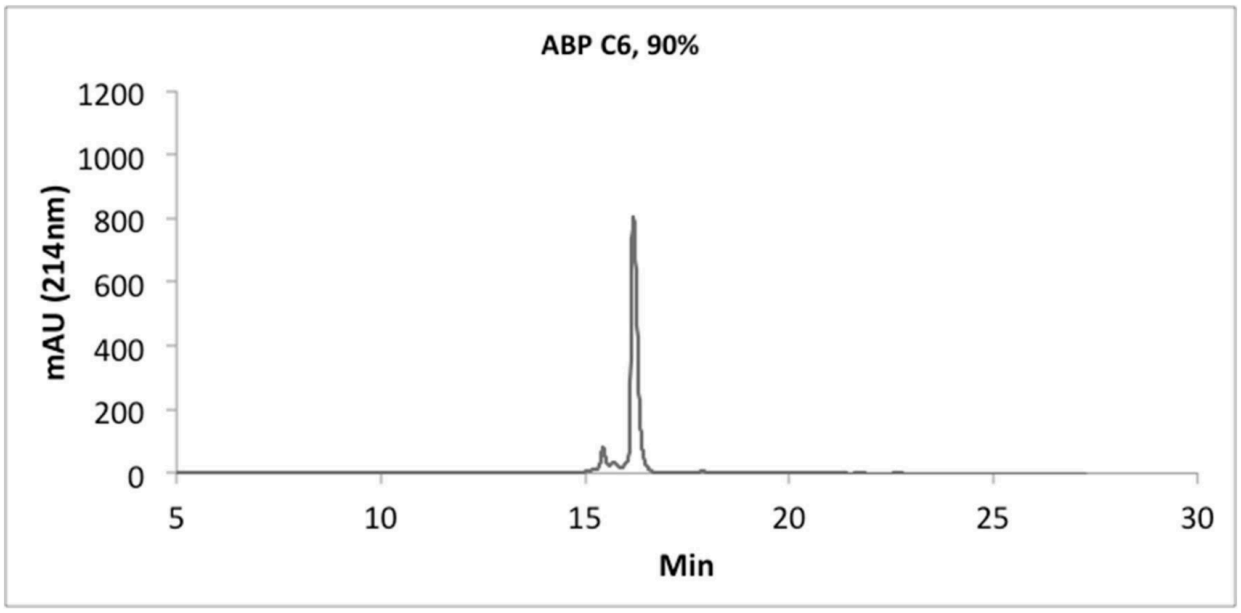


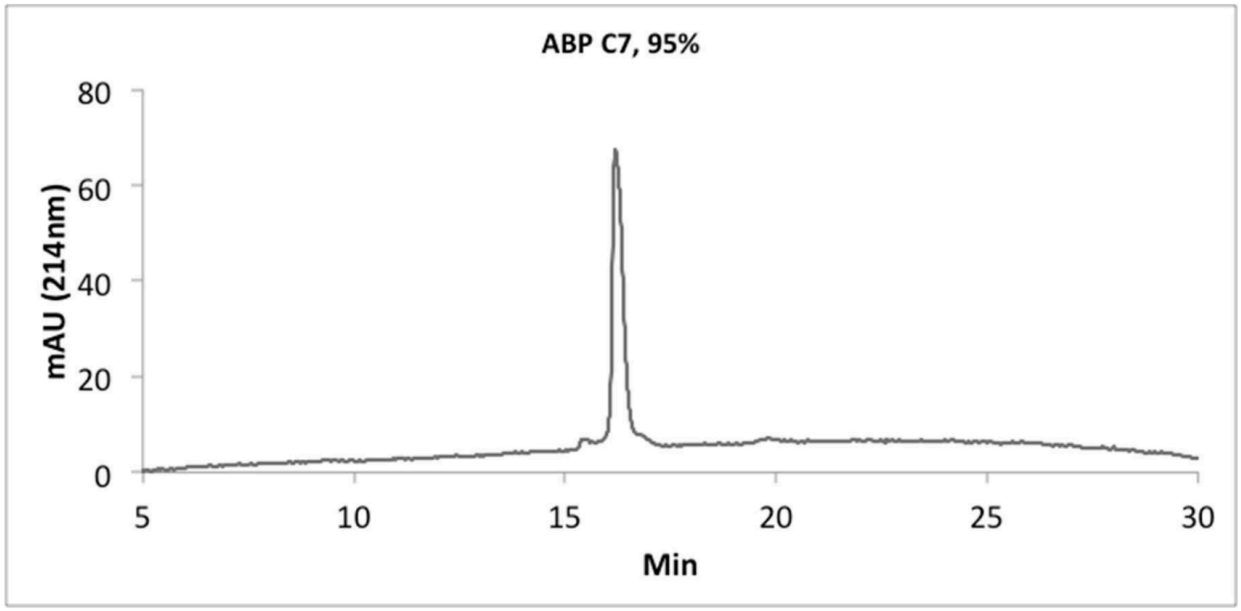

ABP C2, 93%
800
600
400
200
0
mAU (214nm)
5
10
15
20
25
30
Min

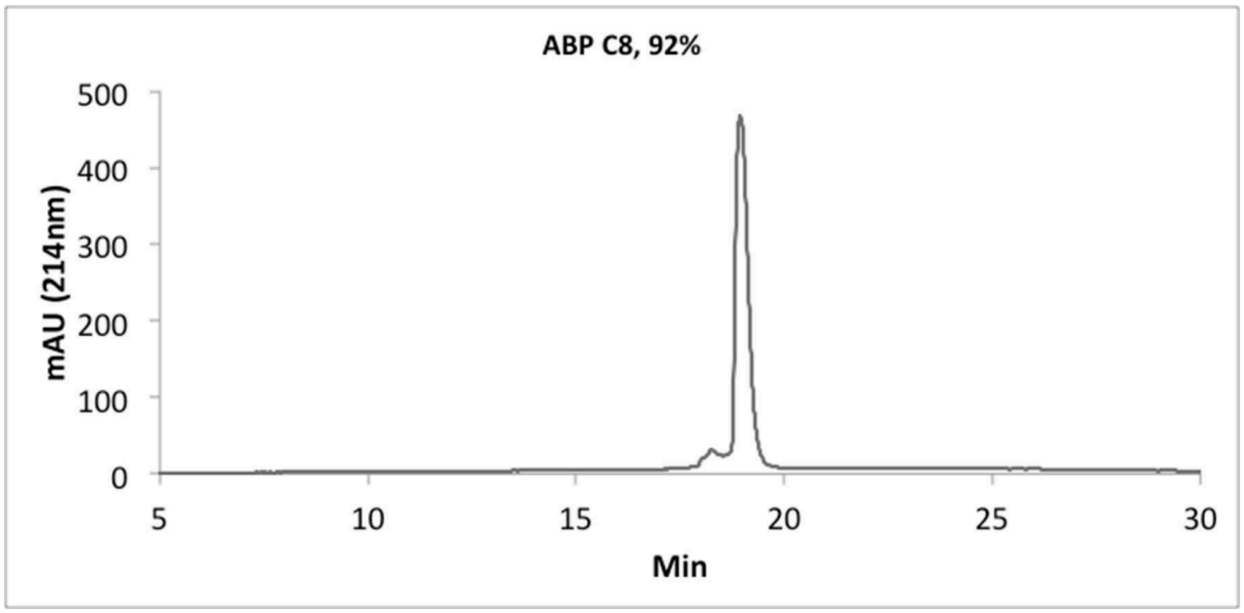

ABP C3, 91%
800
600
400
200
0
mAU (214nm)
5
10
15
20
25
30
Min

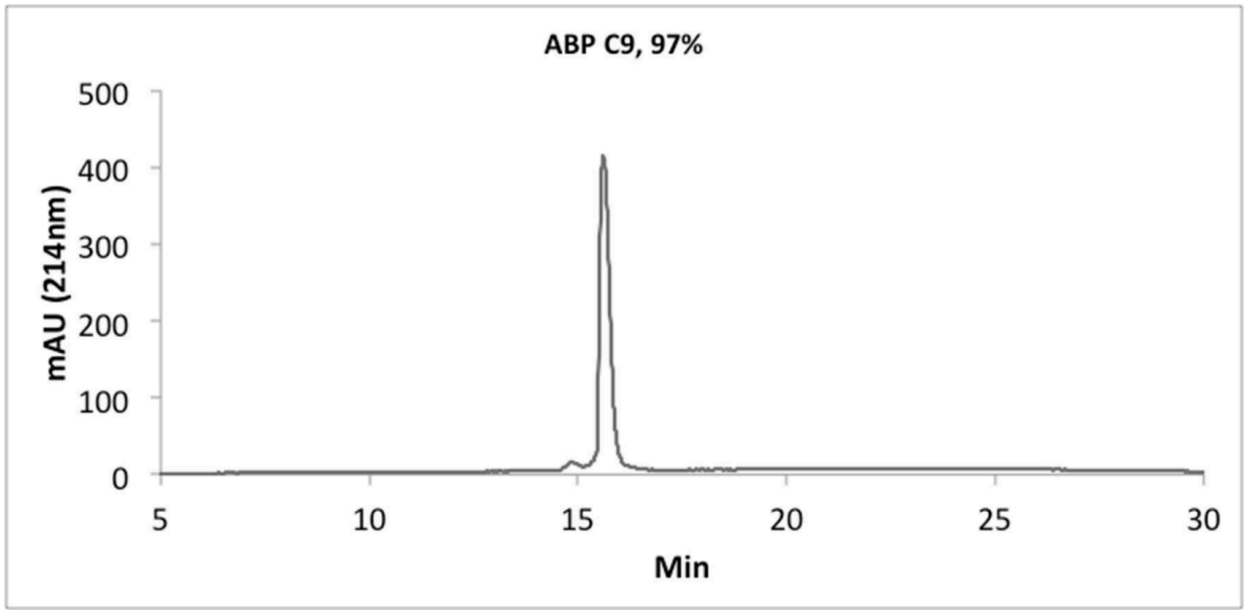

ABP C4, 91%
300
200
100
0
mAU (214nm)
5
10
15
20
25
30
Min


ABP C5, 91%
400
300
200
100
0
mAU (214nm)
5
10
15
20
25
30
Min

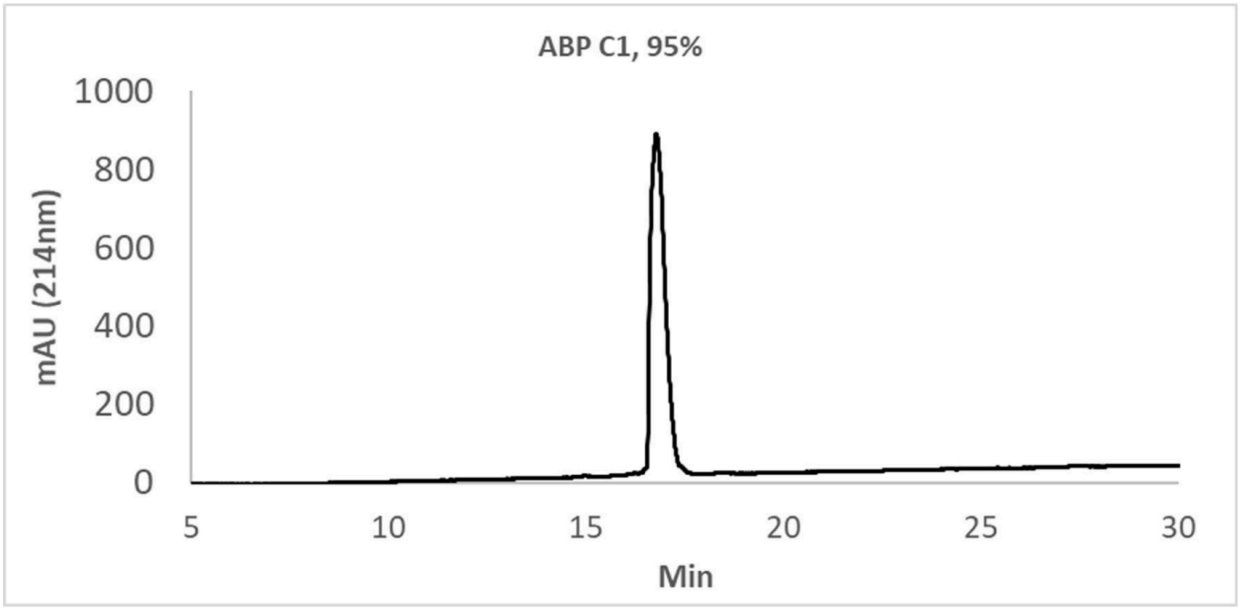

ABP C6, 90%
1200
1000
800
600
400
200
0
mAU (214nm)
5
10
15
20
25
30
Min

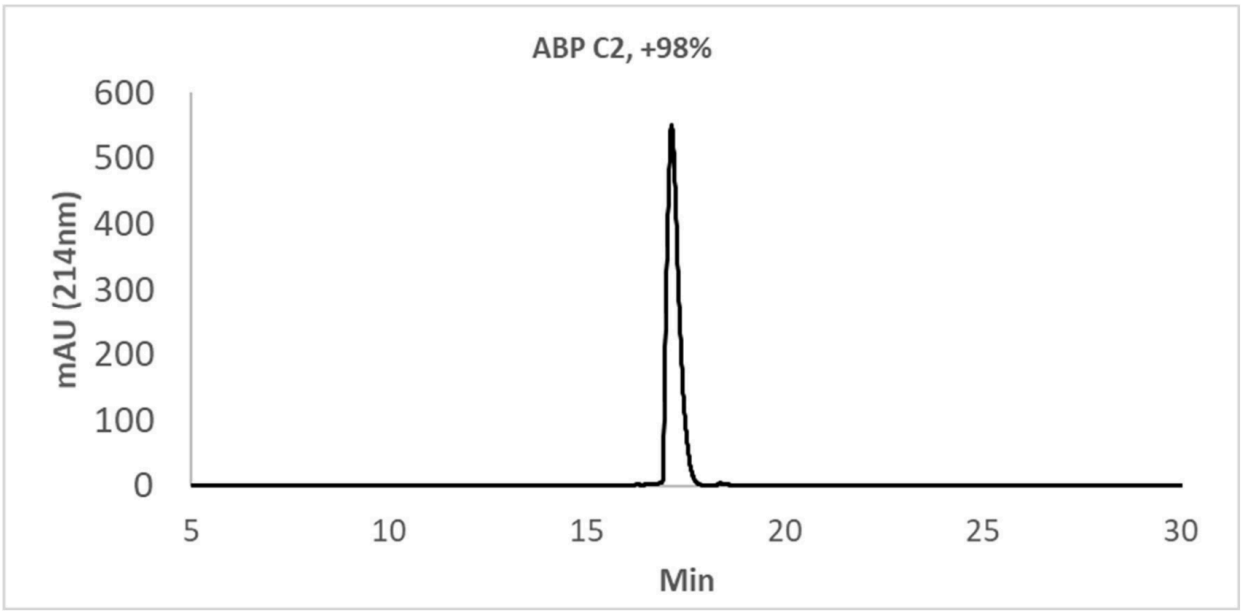

ABP C7, 95%
80
60
40
20
0
mAU (214nm)
5
10
15
20
25
30
Min

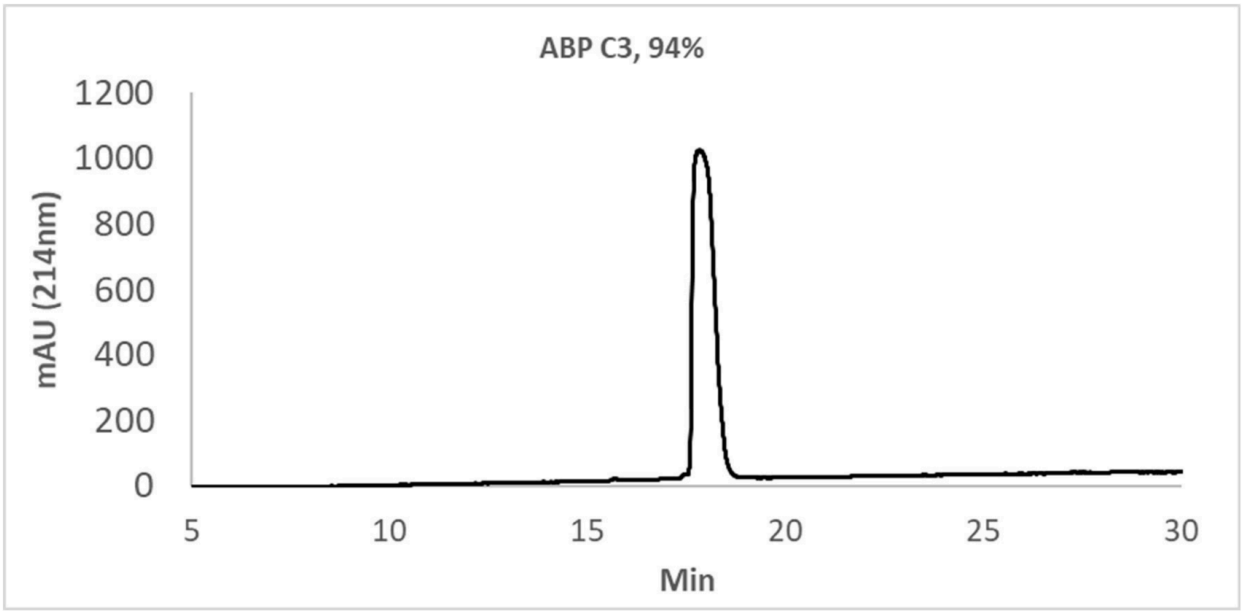

ABP C8, 92%
mAU (214nm)
500
400
300
200
100
0
5
10
15
20
25
30
Min

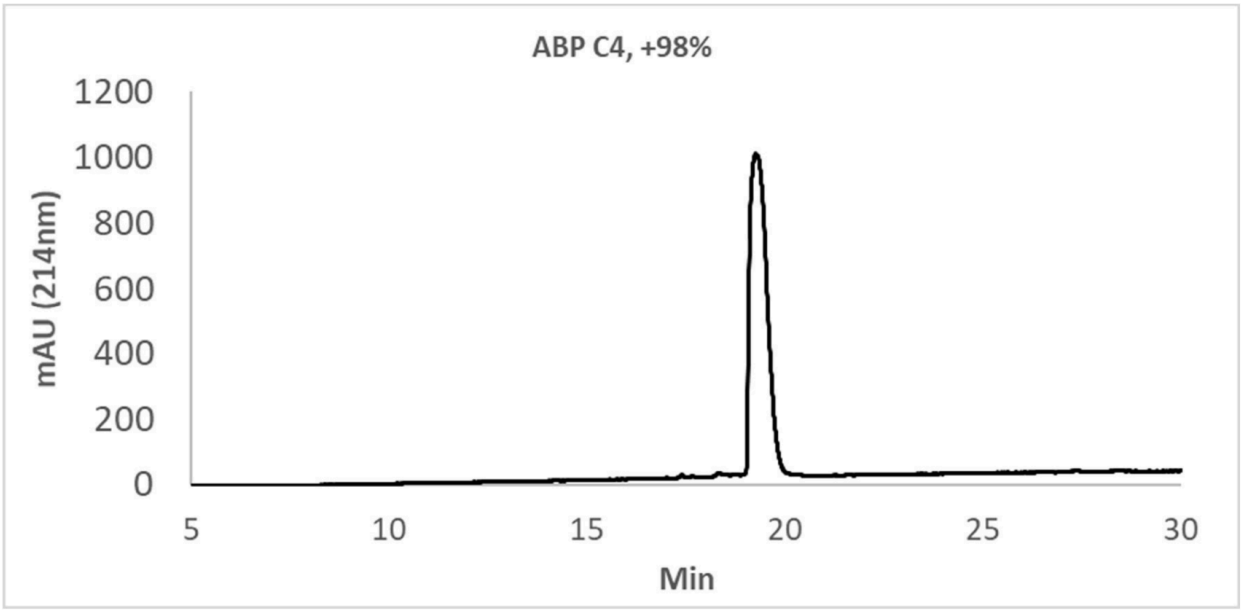

ABP C9, 97%
mAU (214nm)
500
400
300
200
100
0
5
10
15
20
25
30
Min

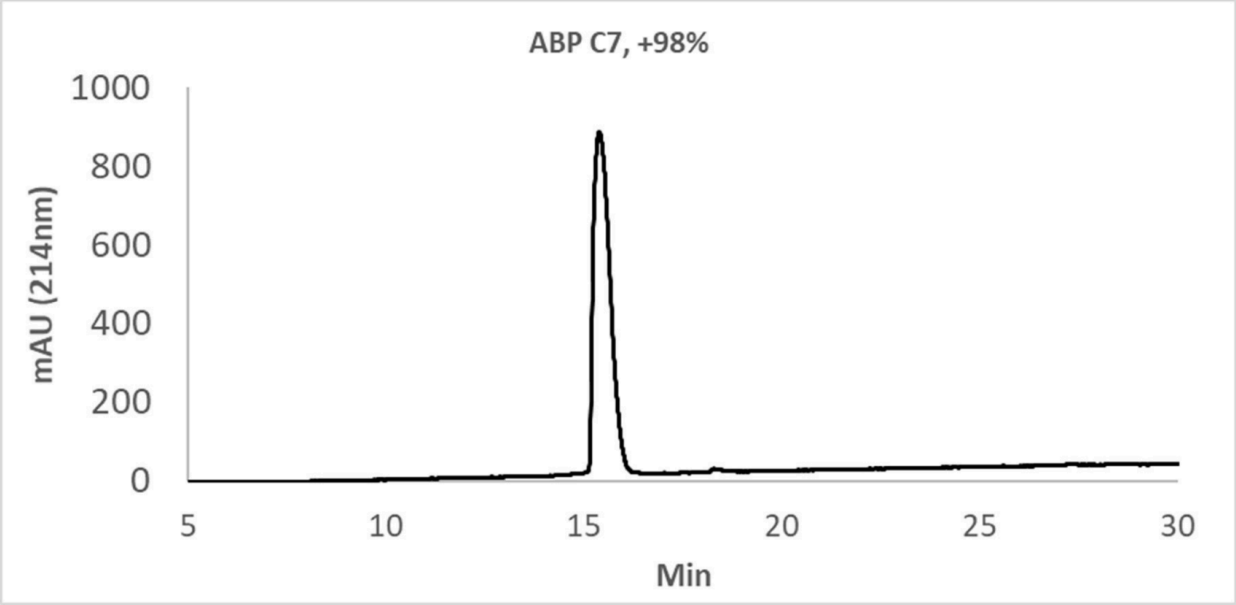
A2.7.4 HPLC trace of canonical peptides (non-biotinylated)

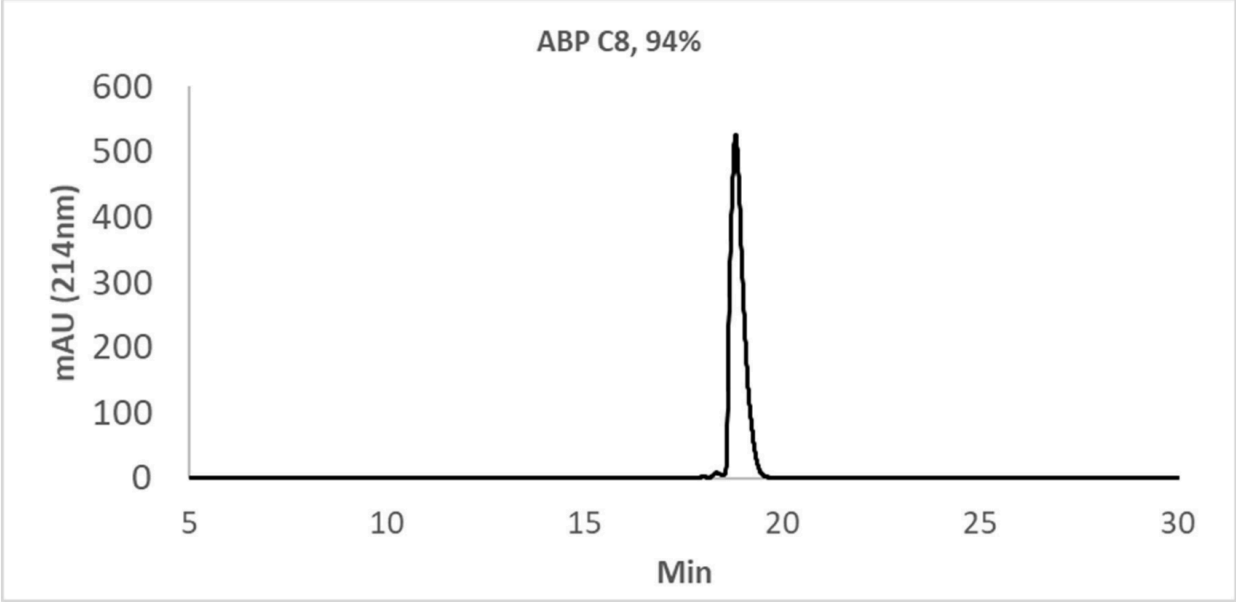


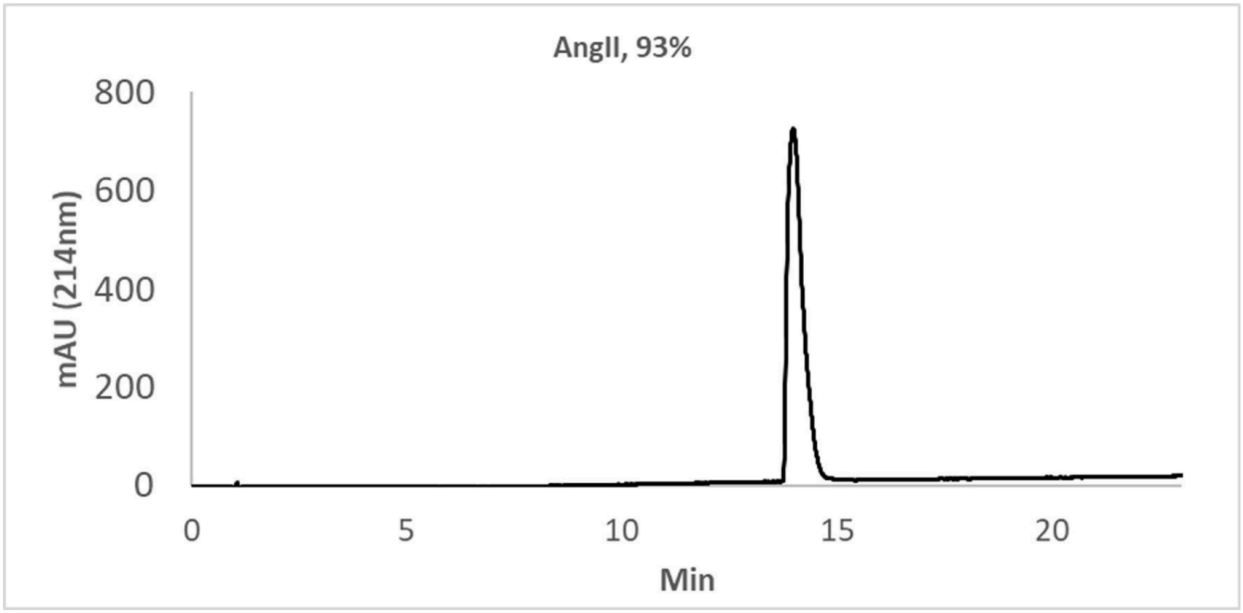



ABP C3, 94%
1200
1000
800
600
400
200
0
mAU (214nm)
5
10
15
20
25
30
Min


ABP C4, +98%
1200
1000
800
600
400
200
0
mAU (214nm)
5
10
15
20
25
30
Min


ABP C7, +98%
1000
800
600
400
200
0
mAU (214nm)
5
10
15
20
25
30
Min


ABP C8, 94%
600
500
400
300
200
100
0
mAU (214nm)
5
10
15
20
25
30
Min



A2.8 Binding affinity measurements

A2.8.1 BLI raw binding data of noncanonical peptides

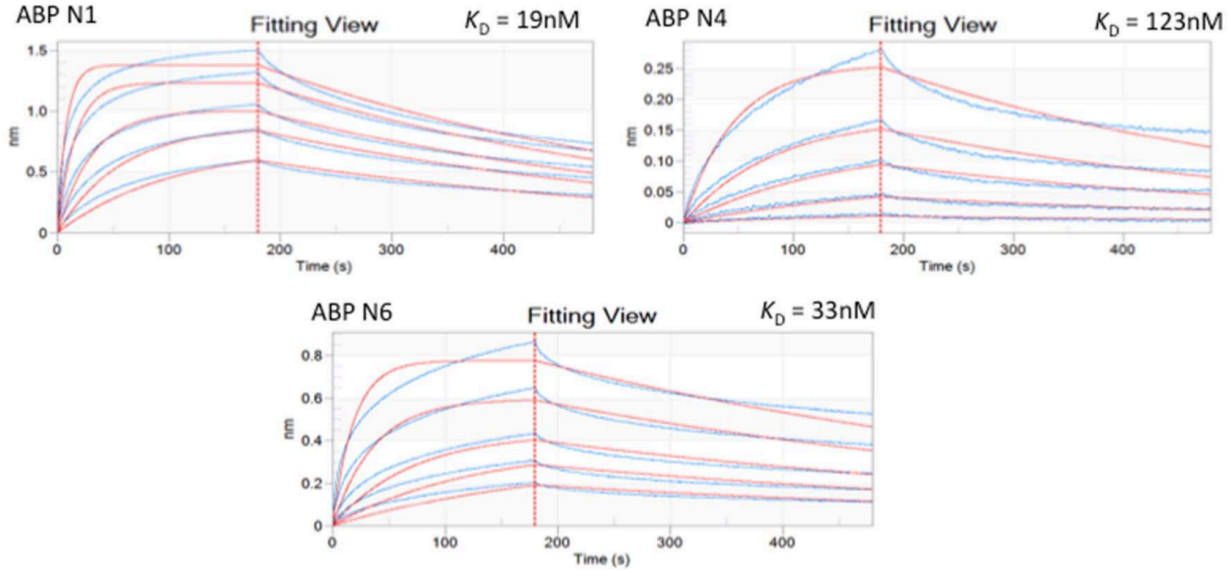


Figure A2.24. **The BLI binding curve fitting (red lines) of ABP N1, ABP N4 and ABP N6 from Library 2 AS-MS against ACE2.** The apparent binding affinity was reported using the kinetic fitting results (KD is calculated by divide the off rate by on rate). ABP N5 and ABP N8 showed no binding to ACE2. The ACE2 concentration titration (n = 5) is: 1000, 500, 250, 125, and 62.5 nM from top to bottom lines.

A2.8.2 BLI raw binding data of canonical peptides

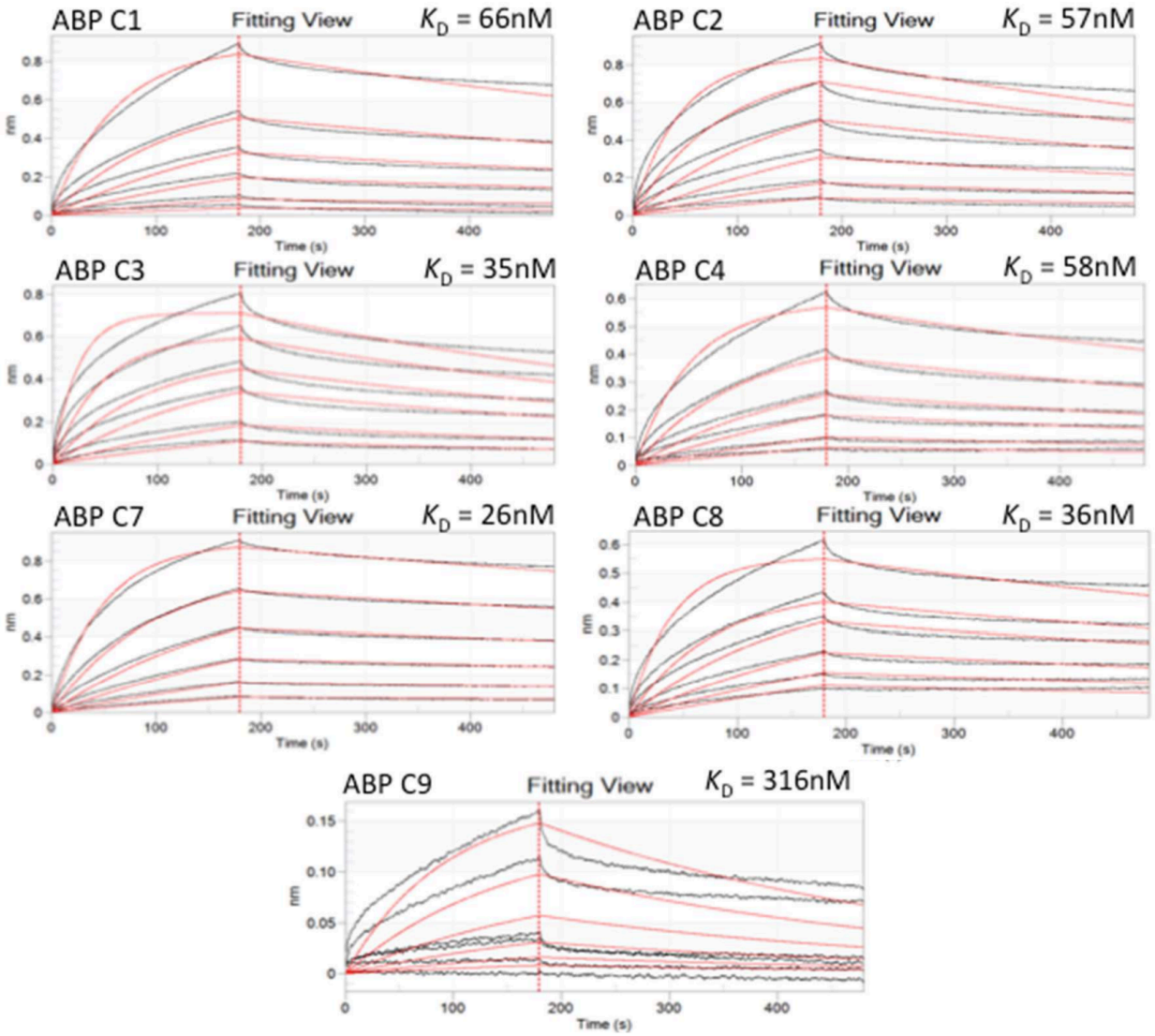


***Figure A2.25*. The BLI binding curve fitting (red lines) of ABP C1-C4 and ABP C7-C9 from Library 1 AS-MS against ACE2**. The apparent binding affinity was reported using the kinetic fitting results (KD is calculated by divide the off rate by on rate). ABP C5 and ABP C6 showed no binding to ACE2. The ACE2 concentration titration (n = 6) is: 1000, 500, 250, 125, 62.5, and 31.3 nM from top to bottom lines.

A2.8.3 The binding of scrambled ABP N1 peptides to ACE2

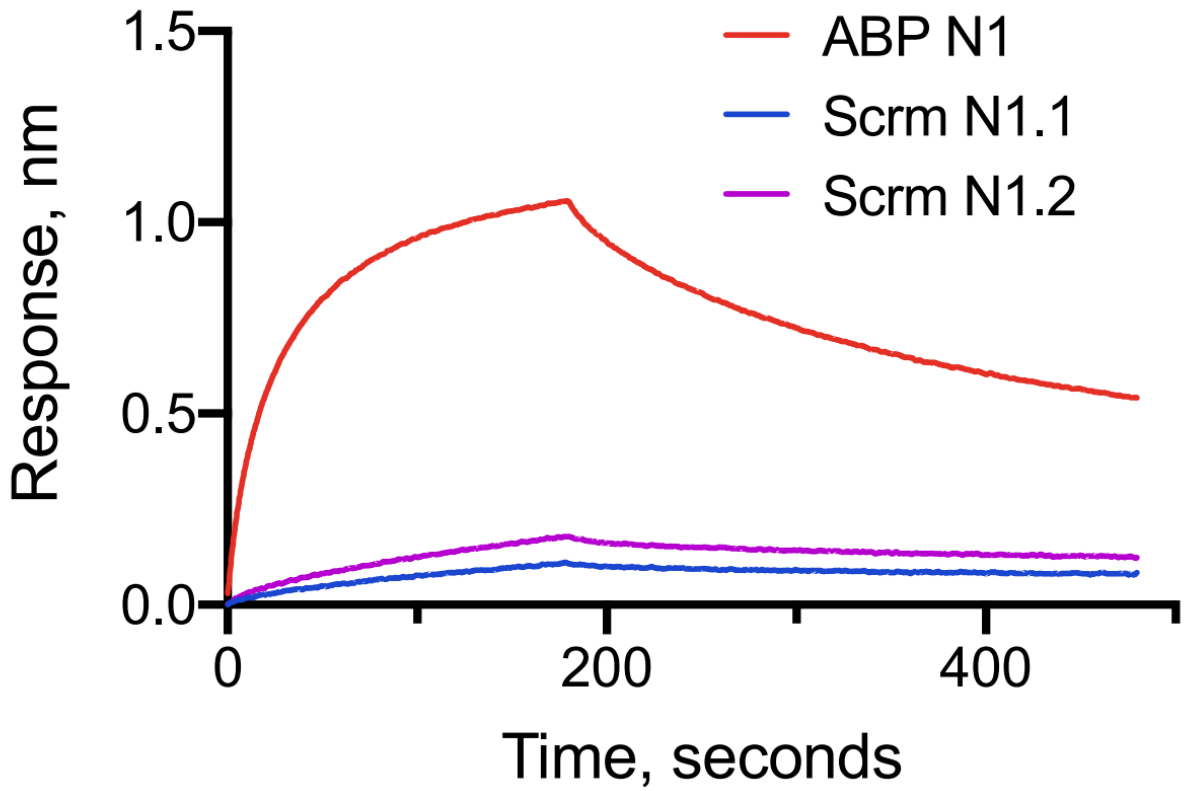


Figure A2.26. **A binding curve comparison of ABP N1 and its scrambled sequences, Scrm N1.1 and Scrm N1.2, measured by BLI.** The tested ACE2 concentration was 250 nM. No apparent binding was observed from neither scrambled peptides.

A2.8.4 The binding of ABP noncanonical peptides to an unrelated protein

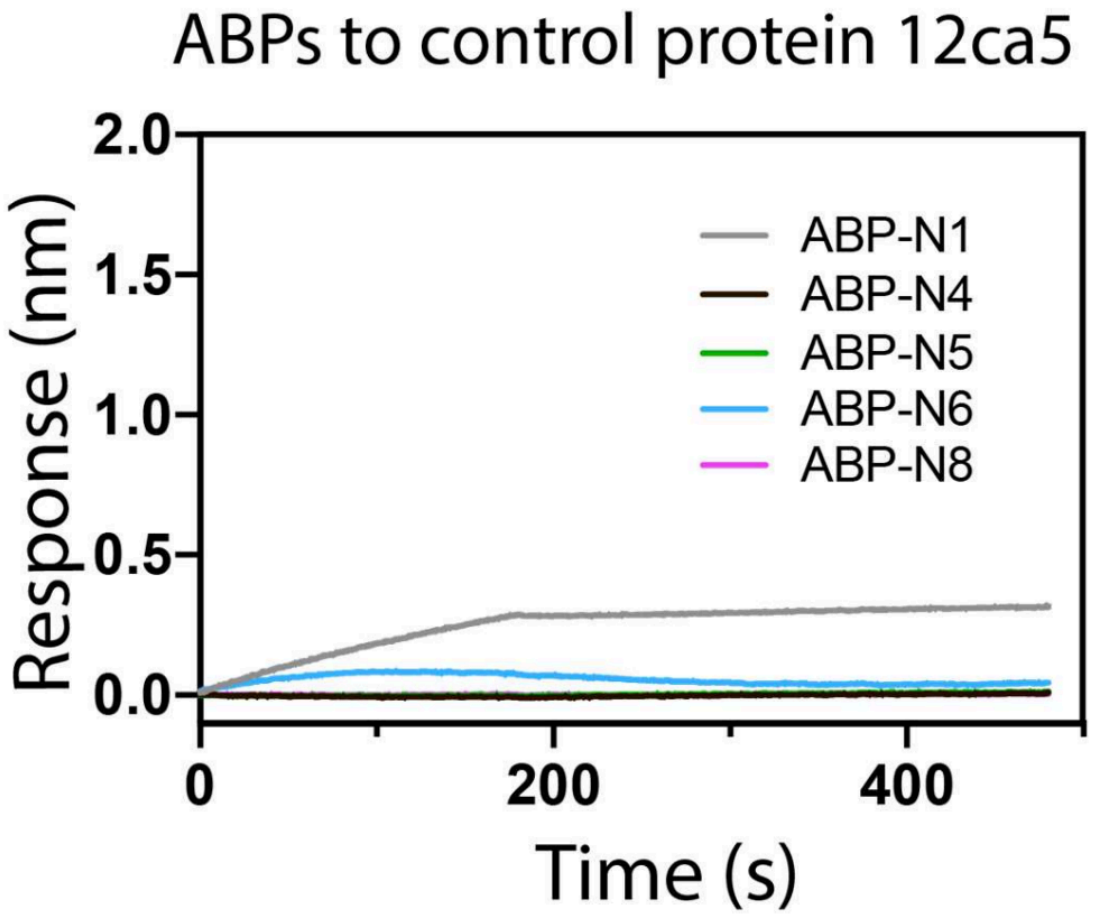


Figure A2.27. **BLI of ABPs to control protein 12ca5.** No apparent binding was observed for ABP N1, N4, N5, N6, and N8 to an unrelated protein, 12ca5, measured by BLI at concentration 500 nM.

A2.9 Binding competition of SARS-CoV-2 RBD-ACE2 interaction

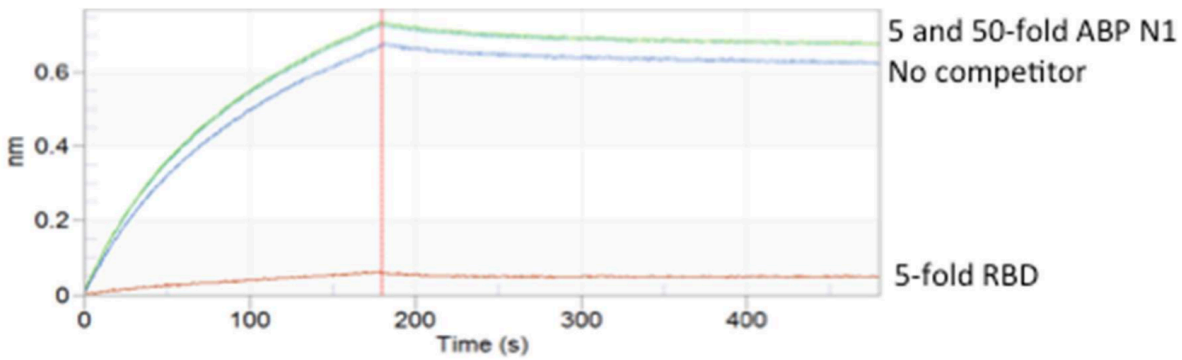


Figure A2.28. **ABP N1 does not interfere the binding of ACE2 and SARS-CoV-2 RBD**. A binding competition assay was performed using BLI. RBD protein was immobilized and dipped into either ACE2 or ACE2 mixed with RBD (the positive control), or ACE2 mixed with ABP N1 peptide. As shown from above binding curves, robust inhibition was observed by 5-fold (over ACE2) soluble RBD, but not by ABP N1 peptides at either 5-fold or 50-fold excess, indicating no direct inhibition from ABP N1 on the ACE2 and SARS-CoV-2 RBD interaction.

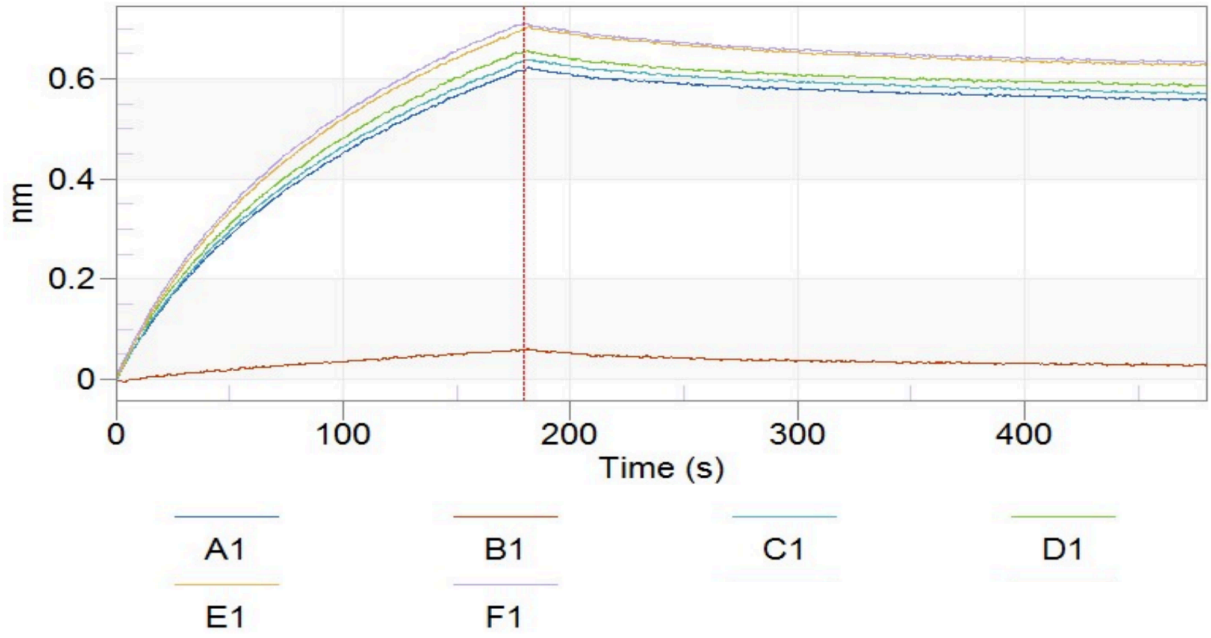


Figure A2.29. **ABP N4 and N6 do not interfere the binding of ACE2 and SARS-CoV-2 RBD**. A binding competition assay was performed using BLI. RBD protein was immobilized and dipped into either ACE2 (100 nM) or ACE2 mixed with RBD (the positive control, at either 0 or 500 nM), or ACE2 mixed with ABP N4 or N6 peptide (500 or 5000 nM). As shown from above binding curves, robust inhibition was observed by 5-fold (over ACE2) soluble RBD, but not by ABP N4 or N6 peptides at either 5-fold or 50-fold excess, indicating no direct inhibition from ABP N4 or N6 on the ACE2 and SARS-

CoV-2 RBD interaction. A1: 0 nM RBD; B1: 500 nM RBD; C1: 500 nM ABP N4; D1: 5000 nM ABP N4; E1: 500 nM ABP N6; F1: 5000 nM ABP N6.

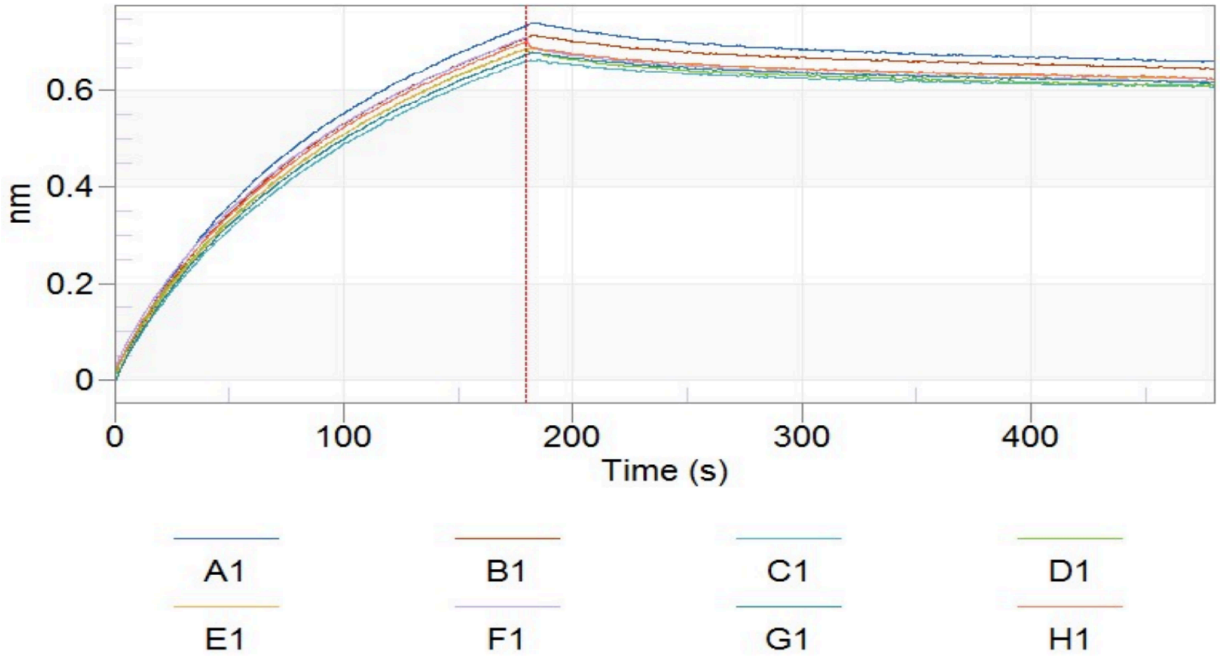


Figure A2.30. **ABP C1, C2, C3, and C4 do not interfere the binding of ACE2 and SARS-CoV-2 RBD.** A binding competition assay was performed using BLI. RBD protein was immobilized and dipped into either ACE2 (100 nM) or ACE2 mixed with ABP C1, C2, C3, or C4 peptides (at 500 or 5000 nM). As shown from above binding curves, no obvious inhibition was observed by ABP C1, C2, C3, or C4 peptides at either 5-fold or 50-fold excess, indicating no direct inhibition on the ACE2 and SARS-CoV-2 RBD interaction. A1: 500 nM ABP C1; B1: 5000 nM ABP C1; C1: 500 nM ABP C2; D1: 5000 nM ABP C2; E1: 500 nM ABP C3; F1: 5000 nM ABP C3; G1: 500 nM ABP C4; H1: 5000 nM ABP C4.

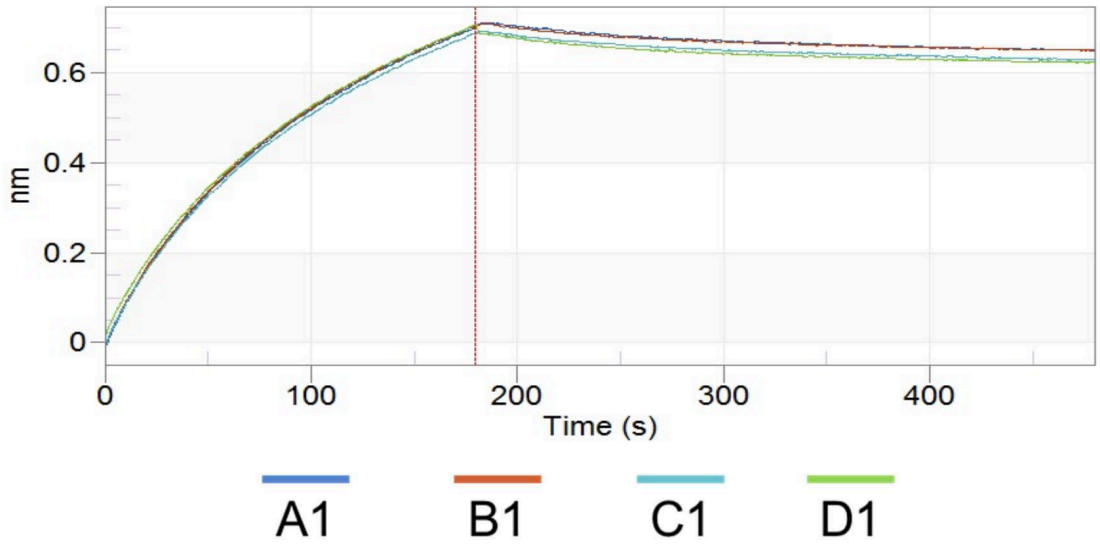


***Figure A2.31*. ABP C7 and C8 do not interfere the binding of ACE2 and SARS-CoV-2 RBD.** A binding competition assay was performed using BLI. RBD protein was immobilized and dipped into either ACE2 (100 nM) or ACE2 mixed with ABP C7 or C8 peptides (at 500 or 5000 nM). As shown from above binding curves, no obvious inhibition was observed by ABP C7 or C8 peptides at either 5-fold or 50-fold excess, indicating no direct inhibition on the ACE2 and SARS-CoV-2 RBD interaction. A1: 500 nM ABP C7; B1: 5000 nM ABP C7; C1: 500 nM ABP C8; D1: 5000 nM ABP C8.

A2.10 Binding competition of known ACE2 inhibitors to ABPs-ACE2 interaction

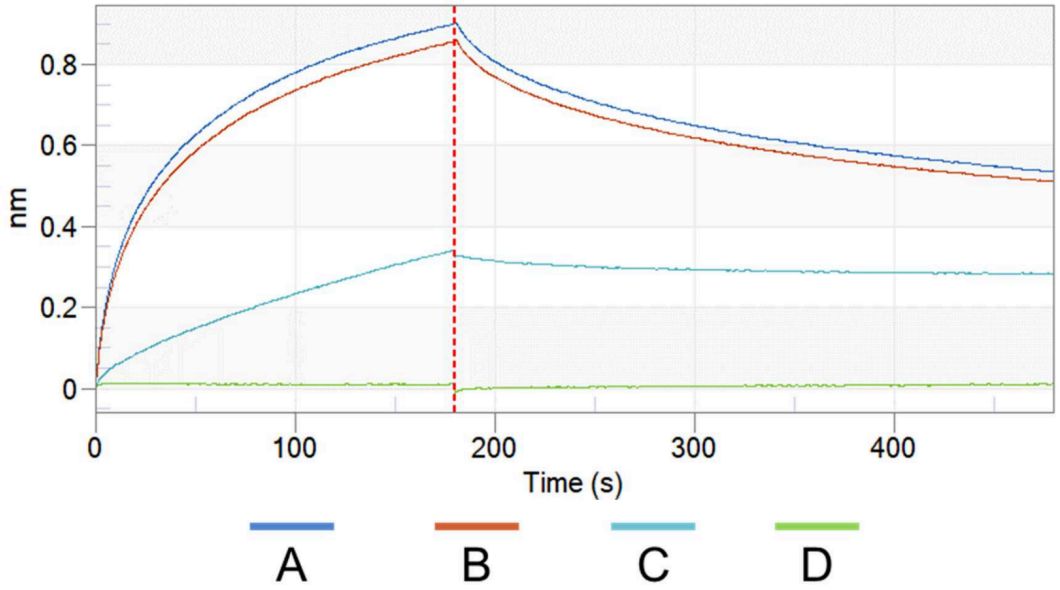


***Figure A2.32*. The binding of ABP N1 is partially inhibited by ACE2 inhibitor MLN-4760 and not inhibited by AngII.** A binding competition assay was performed using

BLI. Biotinylated peptide was immobilized dipped into either ACE2 (500 nM, “A” above in blue) or ACE2 mixed with AngII in 10-fold excess (5 uM, “B” in red) or ACE2 mixed with MLN-4760 in 10-fold excess (5 uM, “C” in cyan). A no protein control (0 nM ACE) is shown in “D” in green to observed any nonspecific BSA binding to the immobilized peptide. As shown from above binding curves, no obvious inhibition was observed by the AngII peptide at 10- fold excess. However, MLN-4760 decreased the binding of ACE2 by more than half, indicating that ABP N1 binding is inhibited. Immobilized: ABP N1; A: 500 nM ACE2; B: 500 nM ACE2 + 5 uM AngII; C: 500 nM ACE2 + 5 uM MLN-4760; D: 0 nM ACE2.

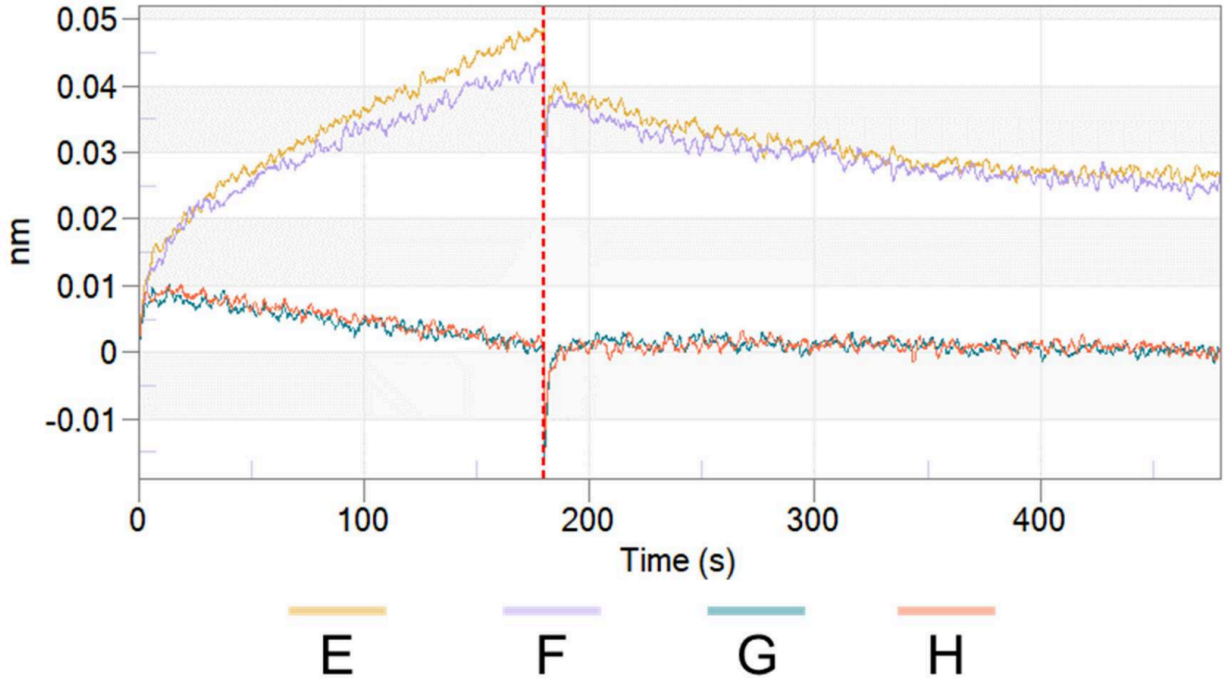


Figure A2.33. **The binding of ABP N4 is completely inhibited by ACE2 inhibitor MLN-4760 and not inhibited by AngII.** A binding competition assay was performed using BLI. Biotinylated peptide was immobilized dipped into either ACE2 (500 nM, “E” above in gold) or ACE2 mixed with AngII in 10-fold excess (5 uM, “F” in lavender) or ACE2 mixed with MLN-4760 in 10-fold excess (5 uM, “G” in teal). A no protein control (0 nM ACE) is shown in “H” in orange to observed any nonspecific BSA binding to the immobilized peptide. As shown from above binding curves, no obvious inhibition was observed by the AngII peptide at 10- fold excess. However, MLN-4760 decreased the binding of ACE2 completely, indicating that ABP N4 binding is inhibited. Immobilized: ABP N4; E: 500 nM ACE2; F: 500 nM ACE2 + 5 uM AngII; G: 500 nM ACE2 + 5 uM MLN-4760; H: 0 nM ACE2.

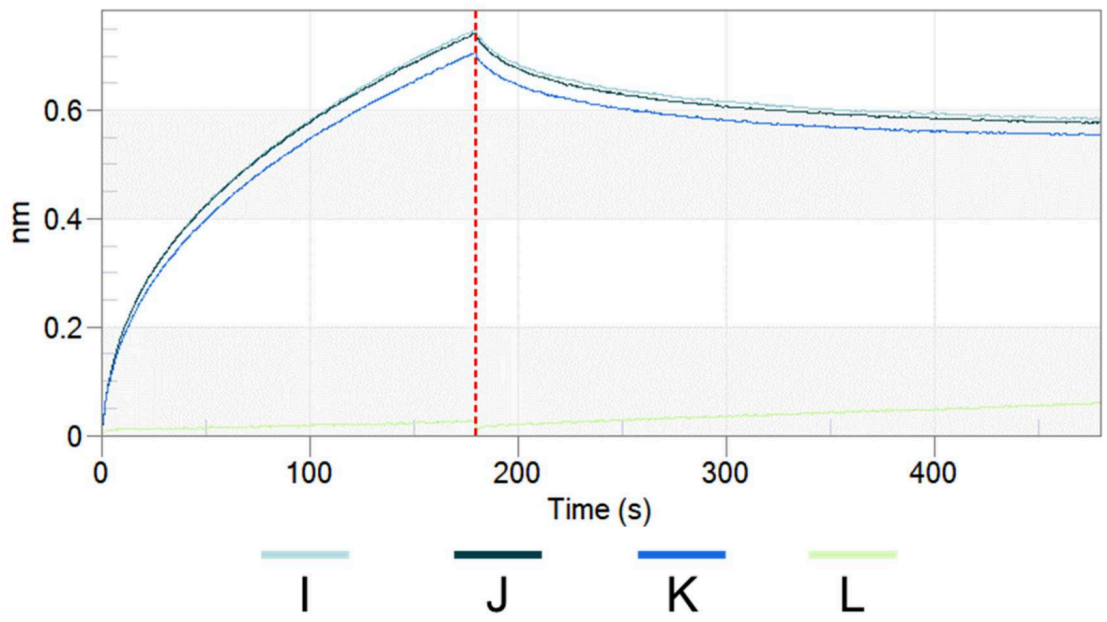


Figure A2.34. **The binding of ABP N6 is not inhibited by ACE2 inhibitor MLN-4760 or AngII.** A binding competition assay was performed using BLI. Biotinylated peptide was immobilized dipped into either ACE2 (500 nM, "I" above in aqua) or ACE2 mixed with AngII in 10-fold excess (5 uM, "J" in dark green) or ACE2 mixed with MLN-4760 in 10-fold excess (5 uM, "K" in royal blue). A no protein control (0 nM ACE) is shown in "L" in light green to observed any nonspecific BSA binding to the immobilized peptide. As shown from above binding curves, no obvious inhibition was observed by the AngII peptide or MLN4760 at 10-fold excess. Immobilized: ABP N6; I: 500 nM ACE2; J: 500 nM ACE2 + 5 uM AngII; K: 500 nM ACE2 + 5 uM MLN-4760; L: 0 nM ACE2.

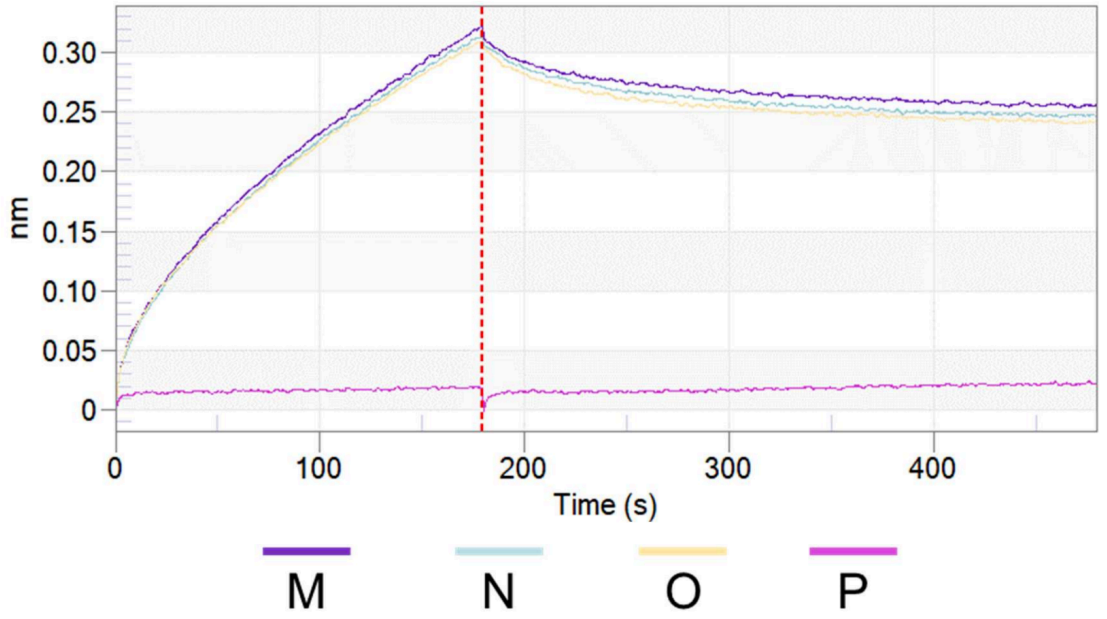


Figure A2.35. **The binding of ABP C1 is not inhibited by ACE2 inhibitor MLN-4760 or AngII.** A binding competition assay was performed using BLI. Biotinylated peptide

was immobilized dipped into either ACE2 (500 nM, “M” above in purple) or ACE2 mixed with AngII in 10-fold excess (5 uM, “N” in aqua) or ACE2 mixed with MLN-4760 in 10-fold excess (5 uM, “O” in light yellow). A no protein control (0 nM ACE) is shown in “P” in magenta to observed any nonspecific BSA binding to the immobilized peptide. As shown from above binding curves, no obvious inhibition was observed by the AngII peptide or MLN4760 at 10-fold excess. Immobilized: ABP C1; M: 500 nM ACE2; N: 500 nM ACE2 + 5 uM AngII; O: 500 nM ACE2 + 5 uM MLN-4760; P: 0 nM ACE2.

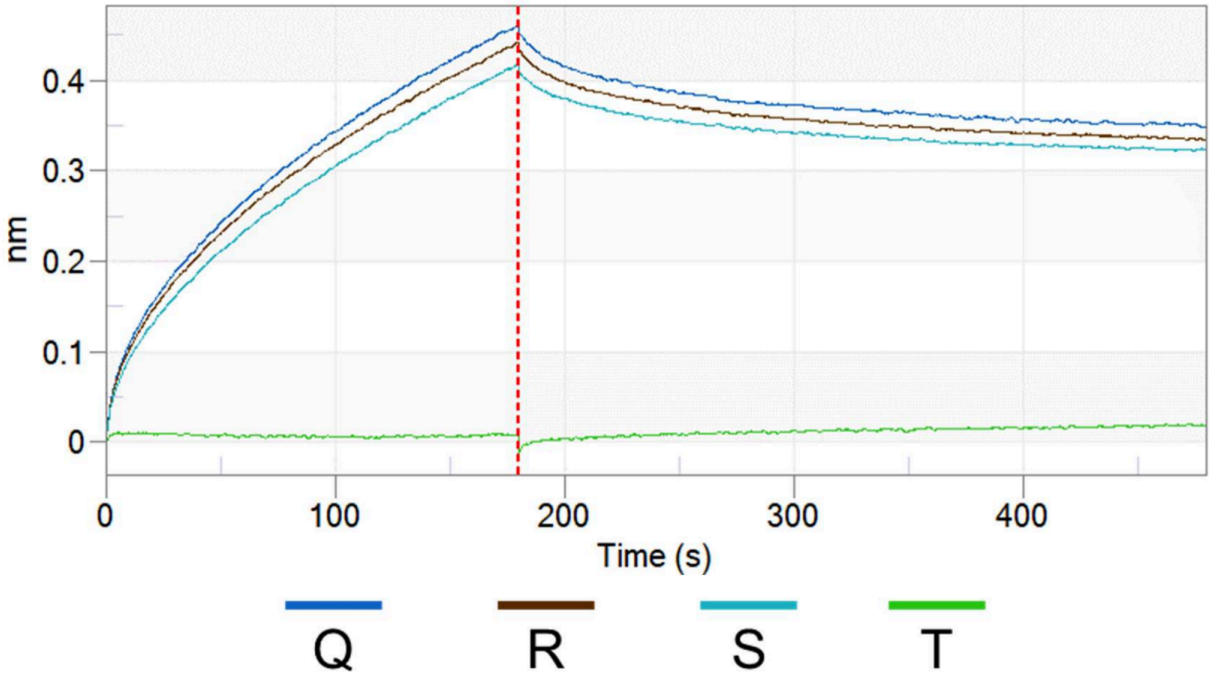


Figure A2.36. **The binding of ABP C2 is not inhibited by ACE2 inhibitor MLN-4760 or AngII.** A binding competition assay was performed using BLI. Biotinylated peptide was immobilized dipped into either ACE2 (500 nM, “Q” above in blue) or ACE2 mixed with AngII in 10-fold excess (5 uM, “R” in brown) or ACE2 mixed with MLN-4760 in 10-fold excess (5 uM, “S” in aqua). A no protein control (0 nM ACE) is shown in “T” in green to observed any nonspecific BSA binding to the immobilized peptide. As shown from above binding curves, no obvious inhibition was observed by the AngII peptide or MLN-4760 at 10-fold excess. Immobilized: ABP C2; Q: 500 nM ACE2; R: 500 nM ACE2 + 5 uM AngII; S: 500 nM ACE2 + 5 uM MLN-4760; T: 0 nM ACE2.

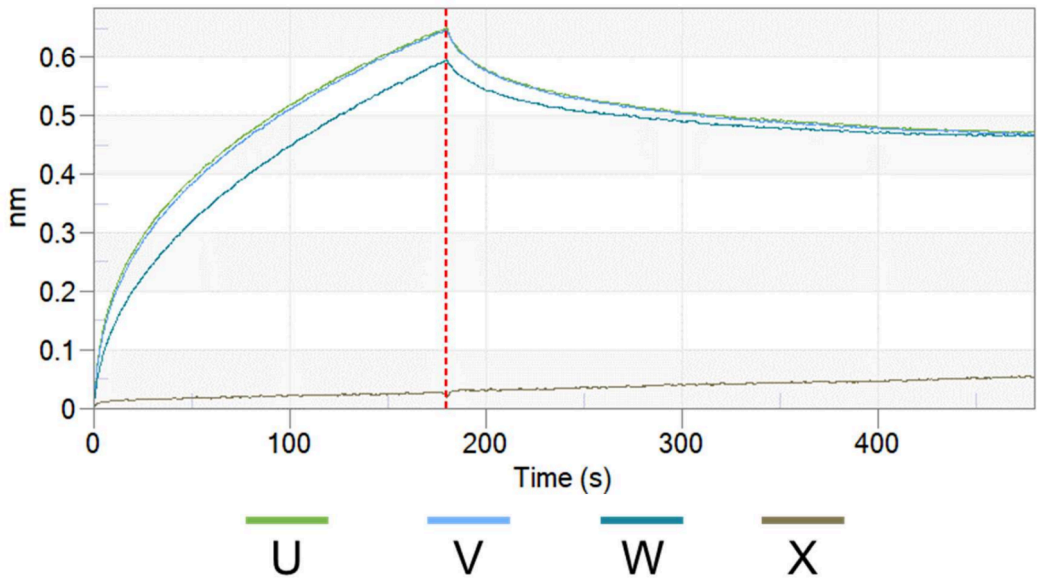


***Figure A2.37*. The binding of ABP C3 is not inhibited by ACE2 inhibitor MLN-4760 or AngII.** A binding competition assay was performed using BLI. Biotinylated peptide was immobilized dipped into either ACE2 (500 nM, "U" above in green) or ACE2 mixed with AngII in 10-fold excess (5 uM, "V" in light blue) or ACE2 mixed with MLN-4760 in 10-fold excess (5 uM, "W" in teal). A no protein control (0 nM ACE) is shown in "X" in brown to observed any nonspecific BSA binding to the immobilized peptide. As shown from above binding curves, no obvious inhibition was observed by the AngII peptide or MLN-4760 at 10-fold excess. Immobilized: ABP C3; U: 500 nM ACE2; V: 500 nM ACE2 + 5 uM AngII; W: 500 nM ACE2 + 5 uM MLN-4760; X: 0 nM ACE2.

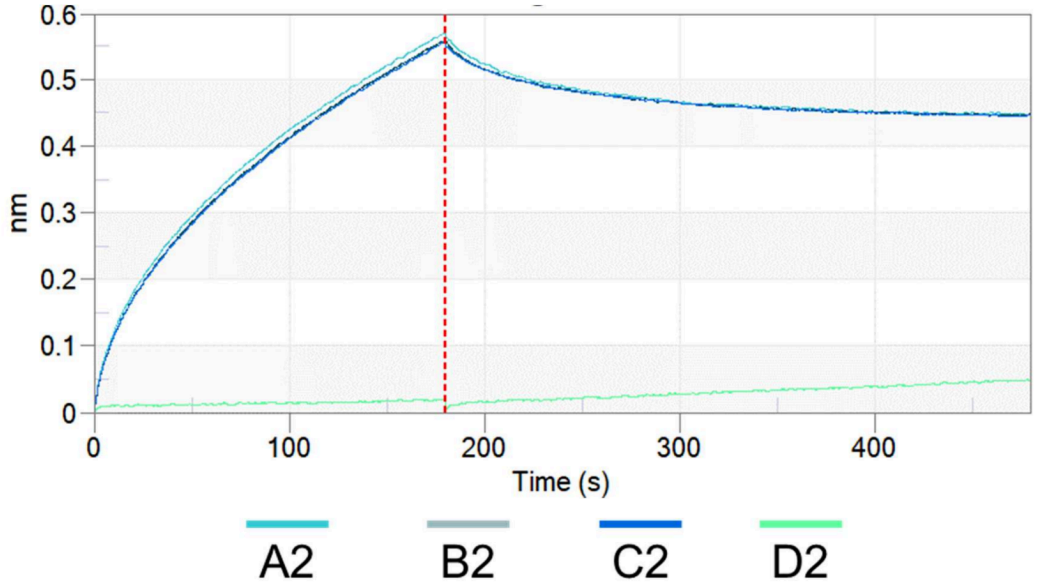


***Figure A2.38*. The binding of ABP C4 is not inhibited by ACE2 inhibitor MLN-4760 or AngII.** A binding competition assay was performed using BLI. Biotinylated peptide was immobilized dipped into either ACE2 (500 nM, "A2" above in cyan) or ACE2 mixed

with AngII in 10-fold excess (5 uM, “B2” in gray) or ACE2 mixed with MLN-4760 in 10-fold excess (5 uM, “C2” in blue). A no protein control (0 nM ACE) is shown in “D2” in lime green to observed any nonspecific BSA binding to the immobilized peptide. As shown from above binding curves, no obvious inhibition was observed by the AngII peptide or MLN4760 at 10-fold excess. Immobilized: ABP C4; A2: 500 nM ACE2; B2: 500 nM ACE2 + 5 uM AngII; C2: 500 nM ACE2 + 5 uM MLN-4760; D2: 0 nM ACE2.

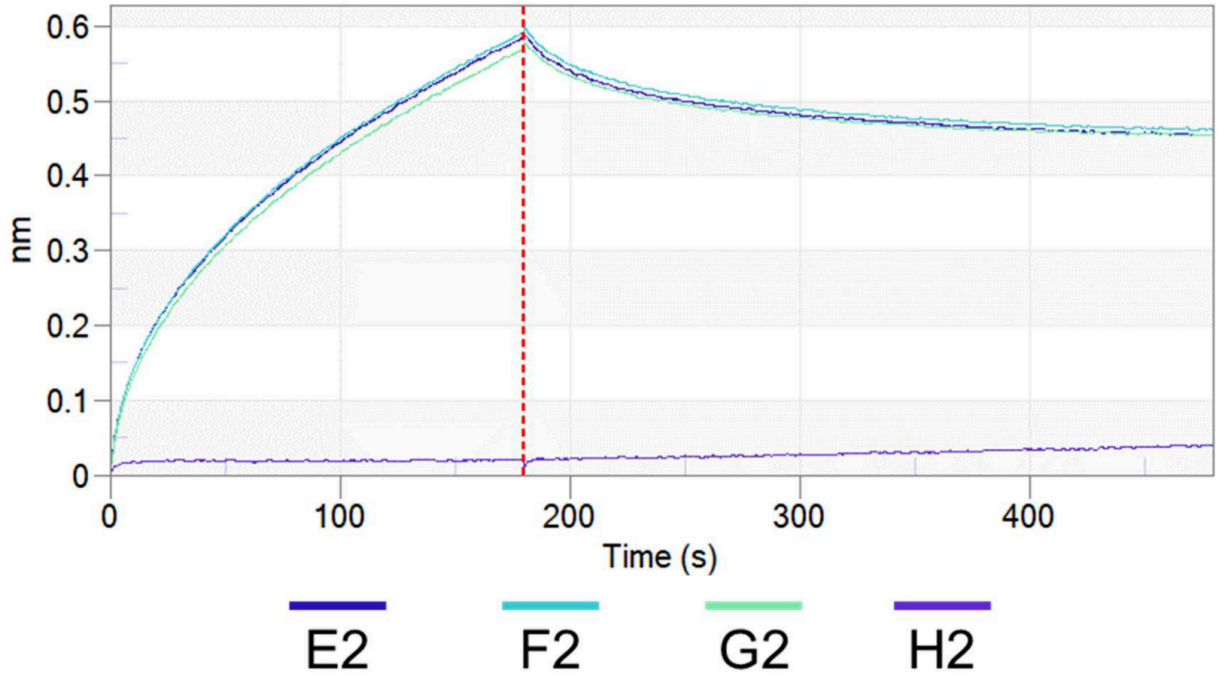


Figure A2.39. **The binding of ABP C7 is not inhibited by ACE2 inhibitor MLN-4760 or AngII.** A binding competition assay was performed using BLI. Biotinylated peptide was immobilized dipped into either ACE2 (500 nM, “E2” above in dark purple) or ACE2 mixed with AngII in 10-fold excess (5 uM, “F2” in cyan) or ACE2 mixed with MLN-4760 in 10-fold excess (5 uM, “G2” in lime green). A no protein control (0 nM ACE) is shown in “H2” in purple to observed any nonspecific BSA binding to the immobilized peptide. As shown from above binding curves, no obvious inhibition was observed by the AngII peptide or MLN4760 at 10-fold excess. Immobilized: ABP C7; E2: 500 nM ACE2; F2: 500 nM ACE2 + 5 uM AngII; G2: 500 nM ACE2 + 5 uM MLN-4760; H2: 0 nM ACE2.

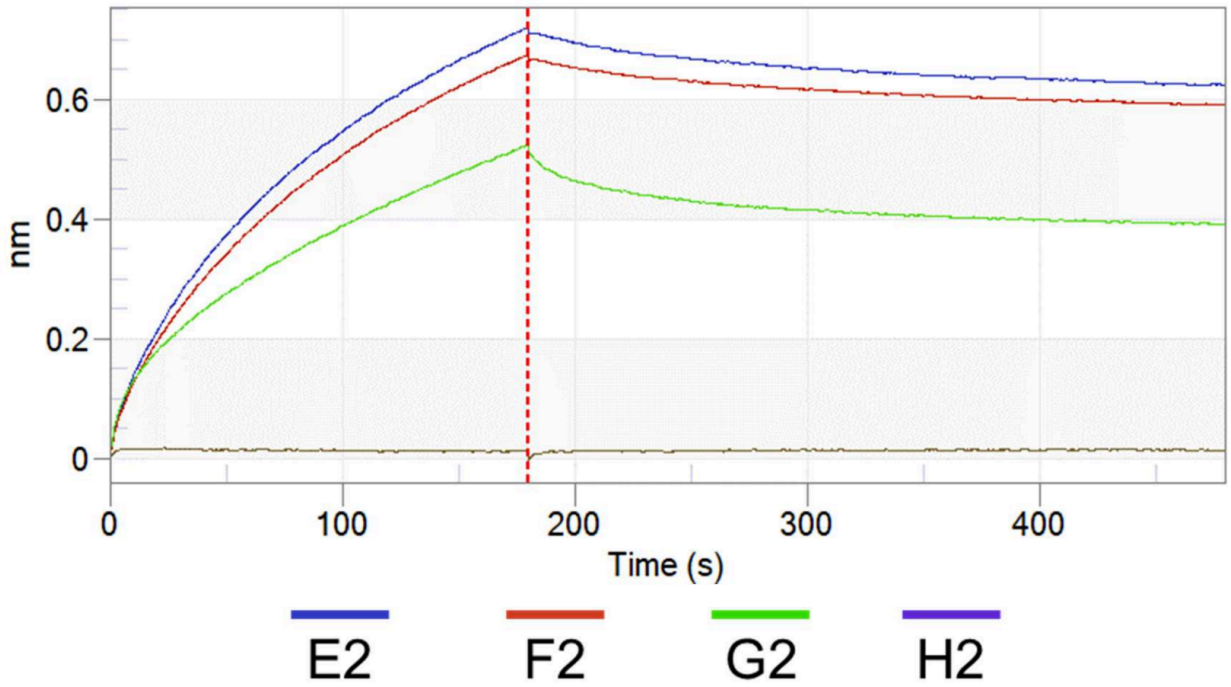


Figure A2.40. **The binding of ABP C8 is partially inhibited by ACE2 inhibitor MLN-4760 and not inhibited by AngII.** A binding competition assay was performed using BLI. Biotinylated peptide was immobilized dipped into either ACE2 (500 nM, “E2” above in dark blue) or ACE2 mixed with AngII in 10-fold excess (5 uM, “F2” in red) or ACE2 mixed with MLN-4760 in 10-fold excess (5 uM, “G2” in green). A no protein control (0 nM ACE) is shown in “H2” in purple to observed any nonspecific BSA binding to the immobilized peptide. As shown from above binding curves, no obvious inhibition was observed by the AngII peptide at 10-fold excess. However, MLN-4760 decreased the binding of ACE2 by less than half, indicating that ABP C8 binding is inhibited. Immobilized: ABP C8; E2: 500 nM ACE2; F2: 500 nM ACE2 + 5 uM AngII; G2: 500 nM ACE2 + 5 uM MLN-4760; H2: 0 nM ACE2.

A2.11 References

1. Frearson, J. A. & Collie, I. T. HTS and hit finding in academia—from chemical genomics to drug discovery. Drug Discov. Today 14, 1150–1158 (2009).

2. Uhlén, M. et al. A human protein atlas for normal and cancer tissues based on antibody proteomics. Mol. Cell. Proteomics 4, 1920–1932 (2005).

3. Taussig, M. J. et al. ProteomeBinders: Planning a European resource of affinity reagents for analysis of the human proteome. Nat. Methods 4, 13–17 (2007).

4. Valeur, E. et al. New modalities for challenging targets in drug discovery. Angew. Chem. Int. Ed. 56, 10294–10323 (2017).

5. Modell, A. E., Blosser, S. L. & Arora, P. S. Systematic targeting of protein–protein interactions. Trends Pharmacol. Sci. 37, 702–713 (2016).

6. Benson, M. D., Ngo, D., Ganz, P. & Gerszten, R. E. Emerging affinity reagents for high throughput proteomics. Circulation 140, 1610–1612 (2019).

7. Uhlen, M. et al. Tissue-based map of the human proteome. Science 347, 1260419–1260419 (2015).

8. Thul, P. J. et al. A subcellular map of the human proteome. Science 356, eaal3321 (2017).

9. Edwards, A. M. et al. Too many roads not taken. Nature 470, 163–165 (2011).

10. Colwill, K. et al. A roadmap to generate renewable protein binders to the human proteome. Nat. Methods 8, 551–561 (2011).

11. Li, F., Vijayasankaran, N., Shen, A., Kiss, R. & Amanullah, A. Cell culture processes for monoclonal antibody production. MAbs 2, 466–479 (2010).

12. Goodnow, R. A., Dumelin, C. E. & Keefe, A. D. DNA-encoded chemistry: Enabling the deeper sampling of chemical space. Nat. Rev. Drug Discov. 16, 131–147 (2017).

13. Eidam, O. & Satz, A. L. Analysis of the productivity of DNA encoded libraries. Medchemcomm 7, 1323–1331 (2016).

14. Dunn, M. R., Jimenez, R. M. & Chaput, J. C. Analysis of aptamer discovery and technology. Nat. Rev. Chem. 1, 1–16 (2017).

15. Marx, V. Calling the next generation of affinity reagents. Nat. Methods 10, 829–833 (2013).

16. Fosgerau, K. & Hoffmann, T. Peptide therapeutics: current status and future directions. Drug Discov. Today 20, 122–128 (2015).

17. Nielsen, T. E. & Schreiber, S. L. Towards the optimal screening collection: A synthesis strategy. Angew. Chem. Int. Ed. 47, 48–56 (2008).

18. Schreiber, S. L. Target-oriented and diversity-oriented organic synthesis in drug discovery. Science 287, 1964–1969 (2000).

19. Gebauer, M. & Skerra, A. Engineered protein scaffolds as next-generation therapeutics. Annu. Rev. Pharmacol. Toxicol. 60, 391–415 (2020).

20. Ståhl, S. et al. Affibody molecules in biotechnological and medical applications. Trends Biotechnol. 35, 691–712 (2017).

21. Vinogradov, A. A., Yin, Y. & Suga, H. Macrocyclic peptides as drug candidates: Recent progress and remaining challenges. J. Am. Chem. Soc. 141, 4167–4181 (2019).

22. Harvey, A. L., Edrada-Ebel, R. & Quinn, R. J. The re-emergence of natural products for drug discovery in the genomics era. Nat. Rev. Drug Discov. 14, 111–129 (2015).

23. Tsomaia, N. Peptide therapeutics: Targeting the undruggable space. Eur. J. Med. Chem. 94, 459–470 (2015).

24. Bird, G. H. et al. Biophysical determinants for cellular uptake of hydrocarbonstapled peptide helices. Nat. Chem. Biol. 12, 845–853 (2016).

25. Chow, H. Y., Zhang, Y., Matheson, E. & Li, X. Ligation technologies for the synthesis of cyclic peptides. Chem. Rev. 119, 9971–10001 (2019).

26. Rogers, J. M., Passioura, T. & Suga, H. Nonproteinogenic deep mutational scanning of linear and cyclic peptides. Proc. Natl Acad. Sci. USA 115, 10959–10964 (2018).

27. Lau, J. et al. Discovery of the once-weekly glucagon-like peptide-1 (GLP-1) analogue semaglutide. J. Med. Chem. 58, 7370–7380 (2015).

28. Clackson, T. & Wells, J. A. In vitro selection from protein and peptide libraries. Trends Biotechnol. 12, 173–184 (1994).

29. Kale, S. S. et al. Cyclization of peptides with two chemical bridges affords large scaffold diversities. Nat. Chem. 10, 715–723 (2018).

30. Josephson, K., Ricardo, A. & Szostak, J. W. MRNA display: From basic principles to macrocycle drug discovery. Drug Discov. Today 19, 388–399 (2014).

31. Tharp, J. M. et al. Initiation of protein synthesis with non-canonical amino acids in vivo. Angew. Chem. Int. Ed. 59, 3122–3126 (2020).

32. Lee, J. et al. Expanding the limits of the second genetic code with ribozymes. Nat. Commun. 10, 1–12 (2019).

33. Katoh, T., Tajima, K. & Suga, H. Consecutive elongation of D-amino acids in translation. Cell Chem. Biol. 24, 46–54 (2017).

34. Johnson, J. A., Lu, Y. Y., Van Deventer, J. A. & Tirrell, D. A. Residue-specific incorporation of non-canonical amino acids into proteins: Recent developments and applications. Curr. Opin. Chem. Biol. 14, 774–780 (2010).

35. Zuckermann, R. N., Kerr, J. M., Siani, M. A., Banville, S. C. & Santi, D. V. Identification of highest-affinity ligands by affinity selection from equimolar peptide mixtures generated by robotic synthesis. Proc. Natl Acad. Sci. USA 89, 4505–4509 (1992).

36. Kaur, S., McGuire, L., Tang, D., Dollinger, G. & Huebner, V. Affinity selection and mass spectrometry-based strategies to identify lead compounds in combinatorial libraries. J. Protein Chem. 16, 505–511 (1997).

37. Kościuczuk, E. M. et al. Cathelicidins: Family of antimicrobial peptides. A review. Mol. Biol. Rep. 39, 10957–10970 (2012).

38. Touti, F., Gates, Z. P., Bandyopdhyay, A. & Lautrette, G. In-solution enrichment identifies peptide inhibitors of protein–protein interactions protein–protein interactions. Nat. Chem. Biol. 15, 410–418 (2019).

39. Gates, Z. P. et al. Xenoprotein engineering via synthetic libraries. Proc. Natl Acad. Sci. USA 115, 201722633 (2018).

40. Weiss, G. A., Watanabe, C. K., Zhong, A., Goddard, A. & Sidhu, S. S. Rapid mapping of protein functional epitopes by combinatorial alanine scanning. Proc. Natl Acad. Sci. USA 97, 8950–8954 (2000).

41. Ye, X. et al. Deep alanine scanning reveals potent multi-alanine-substituted protein–protein interaction inhibitors. ChemRxiv https://doi.org/10.26434/chemrxiv.13609499.v1 (2021).

42. Vinogradov, A. A. et al. Library design-facilitated high-throughput sequencing of synthetic peptide libraries. ACS Comb. Sci. 19, 694–701 (2017).

43. Quartararo, A. J. et al. Ultra-large chemical libraries for the discovery of highaffinity peptide binders. Nat. Commun. 11, 3183 (2020).

44. Pomplun, S. et al. De novo discovery of high-affinity peptide binders for the SARS-CoV-2 spike protein. ACS Cent. Sci. 7, 156–163 (2021).

45. Pomplun, S., Gates, Z. P., Zhang, G., Quartararo, A. J. & Pentelute, B. L. Discovery of nucleic acid binding molecules from combinatorial biohybrid nucleobase peptide libraries. J. Am. Chem. Soc. 142, 19642–19651 (2020).

46. Fernández-Ruiz, I. ACE2 level as a marker of CVD. Nat. Rev. Cardiol. 17, 759–759 (2020).

47. Narula, S. et al. Plasma ACE2 and risk of death or cardiometabolic diseases: A case-cohort analysis. Lancet 396, 968–976 (2020).

48. Wang, Q. et al. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. Cell 181, 894–904.e9 (2020).

49. Lan, J. et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 581, 215–220 (2020).

50. Ma, B. et al. PEAKS: powerful software for peptidede novo sequencing by tandem mass spectrometry. Rapid Commun. Mass Spectrom. 17, 2337–2342 (2003).

51. Vickers, C. et al. Hydrolysis of biological peptides by human angiotensinconverting enzyme-related carboxypeptidase. J. Biol. Chem. 277, 14838–14843 (2002).

52. Towler, P. et al. ACE2 X-ray structures reveal a large hinge-bending motion important for inhibitor binding and catalysis *. J. Biol. Chem. 279, 17996–18007 (2004).

53. Dales, N. A. et al. Substrate-based design of the first class of angiotensinconverting enzyme-related carboxypeptidase (ACE2) inhibitors. J. Am. Chem. Soc. 124, 11852–11853 (2002).

54. Huang, L. et al. Novel peptide inhibitors of angiotensin-converting enzyme 2. J. Biol. Chem. 278, 15532–15540 (2003).

55. Sama, I. E. et al. Circulating plasma concentrations of angiotensin-converting enzyme 2 inmen and women with heart failure and effects of reninangiotensin-aldosterone inhibitors. Eur. Heart J. 41, 1810–1817 (2020).

56. Ramchand, J. et al. Plasma ACE2 activity predicts mortality in aortic stenosis and is associated with severe myocardial fibrosis. JACC Cardiovasc. Imaging 13, 655–664 (2020).

57. Ramchand, J., Patel, S. K., Srivastava, P. M., Farouque, O. & Burrell, L. M. Elevated plasma angiotensin converting enzyme 2 activity is an independent predictor of major adverse cardiac events in patients with obstructive coronary artery disease. PLoS One 13, e0198144 (2018).

58. Walters, T. E. et al. Angiotensin converting enzyme 2 activity and human atrial fibrillation: Increased plasma angiotensin converting enzyme 2 activity is

associated with atrial fibrillation and more advanced left atrial structural remodelling. Europace 19, 1280–1287 (2017).

59. Epelman, S. et al. Soluble angiotensin-converting enzyme 2 in human heart failure: Relation with myocardial function and clinical outcomes. J. Card. Fail. 15, 565–571 (2009).

60. Hartrampf, N. et al. Synthesis of proteins by automated flow chemistry. Science 368, 980–987 (2020).

61. Mijalis, A. J. et al. A fully automated flow-based approach for accelerated peptide synthesis. Nat. Chem. Biol. 13, 464–466 (2017).

62. Touti, F. et al. Antibody-bactericidal macrocyclic peptide conjugates to target gram-negative bacteria. ChemBioChem 19, 2039–2044 (2018).