Automated Flow Synthesis of Biomacromolecules

by

Alex Callahan

B.S., Chemistry; B.S., Biochemistry
University of Rochester (2018)

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

Doctor of Philosophy in Chemistry

at the

Massachusetts Institute of Technology

June 2023

Authored by: .....
Department of Chemistry
May 16, 2023

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

Accepted by: .....
Adam P. Willard
Associate Professor of Chemistry
Graduate Officer

This doctoral thesis has been examined by a committee of the

Department of Chemistry as follows:

Timothy F. Jamison .....
Thesis Committee Chair
Robert R. Taylor Professor of Chemistry

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

Ronald T. Raines .....
Thesis Committee Member
Roger and Georges Firmenich Professor of Chemistry

Automated Flow Synthesis of Biomacromolecules

by

Alex Callahan

Submitted to the Department of Chemistry on May 16, 2023
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy in Chemistry

Abstract

Chemically synthesized biomacromolecules are powerful tools that enable modern biological study, but certain classes remain under-investigated. Existing techniques to access phosphorodiamidate morpholino oligomers (PMOs), mirror image proteins (D proteins), and peptide nucleic acids (PNAs) are functional but require individualized attention and time-intensive effort. As a result, unlike easily accessible biopolymer classes, widespread application of PMOs, D proteins, and PNAs in biotechnological workflows remains rare.

In this work, we report the development of strategies for rapid, reliable access to these synthetically challenging classes of biomacromolecules. Existing synthetic pipelines face two primary complications: long timelines and unpredictable synthetic inefficiency. These issues arise from chemical properties inherent to the biopolymer backbone, for example, sequence-dependent aggregation of proteins and rapid side reactions of PNAs. This thesis reports the development of new production pipelines that address these issues with improvements to chemical synthesis techniques and purification and handling workflows.

Generalized production pipelines were developed for PMOs, proteins, and PNAs. We demonstrate the rapid synthesis of PMOs with a custom-designed automated fast-flow instrument. With this instrument, chemical and process variables of PMO synthesis were optimized, resulting in a platform that enables same-day PMO synthesis. No longer limited by week-long lead times, we demonstrated the on-demand perpetration of PMOs to treat underexplored disease areas, including rare Duchenne muscular dystrophy (DMD) subtypes and SARS-CoV-2. Continuing the investigation of biopolymers from automated flow peptide synthesis (AFPS) to enable therapeutic discovery, we developed an expedited mirror image phage display (MIPD) platform. MIPD is the premier technique for generating D-peptide ligands; however, the widespread application of this technique is hindered by the individualized attention required to prepare the mirror-image D-protein substrates. We demonstrate that automated flow protein synthesis addresses these challenges and delivers mirror image proteins for screening using a standardized, rapid format. These results show the value of rapid, general access to synthetic proteins, and we further refined their preparation with a novel purification scheme termed Folding Selection. We demonstrate that the minor chemical modifications present on synthetic side products result in substantially altered physical-chemical properties and that simple bio-purification techniques can separate them from the native protein in hours. With this strategy, we demonstrate the production of nine functional synthetic proteins in under ten hours each. Finally, we developed a streamlined AFPS pipeline for producing peptide-PNA conjugates without tedious individualized synthetic optimization. With this platform, we rapidly prepared PPNAs and showed their utility as therapeutics against SARS-CoV-2.

In summary, multidisciplinary improvements to biopolymer synthesis pipelines were explored. We envision that the routine production of biopolymers for study with AFPS pipelines stands to free up significant numbers of person-hours otherwise spent on troubleshooting challenging synthetic schemes.

Thesis Supervisor: Bradley L. Pentelute
Professor of Chemistry

Acknowledgements

I would first like to thank *Prof. Brad Pentelute* for his mentorship over the past five years. Brad's enthusiasm for tackling big scientific questions has been contagious since the very beginning, and I consider myself incredibly lucky to have learned so much from him. He helped to teach me how to determinedly pursue new scientific ideas all the while not taking yourself too seriously.

I am grateful that I had the chance to work with such talented and kind colleagues in the Pentelute group. In particular, I am forever grateful for the mentorship from *Prof. Chengxi Li* and *Prof. Nina Hartrampf* in the early stages of my Ph.D. Any acknowledgement of the group would be incomplete without special praise to *Dr. Andrei Loas*. His seemingly limitless energy and enthusiasm were essential to the lab, and I will always appreciate his friendship. I would also like to provide special thanks to those I worked closely with: *Jacob Rodriguez, Charlie Farquhar, Lia Lozano Salazar, Dr. Genwei Zhang, Prof. Yen-Chun Lee, Stephanie Hanna, Mackenzie Poskus, Dr. Carly Schissel, Dr. Satish Ghandesiri Dr. Heemal Dhanjee, Dr. Nick Truex, Dr. Giulio Fittolani, Dr. Ahmet Yesilcimen, and Dr. Aurelie Rondon*. The culture and scientific diversity of the Pentelute lab has always been an invaluable resource, and I would like to thank the many people who helped construct it: *Dr. Vlad Ahkmetov, Dr. John Albin, Sarah Antilla, Dr. Joe Brown, Dr. Stephen Byrne, Yehlin Cho, Amanda Cowfer, Dr. Dio Dippa-Matos, Dr. Max Distler, Dr. Zak Gates, Dr. Nathalie Grob, Hannah Grupe, Dr. Cam Hanna, Patrick Hauck, Dr. Rebecca Holden, Ching-Pei Hsu, Maysa Ilmanova, Prof. Muhammad Jbara, Dr. Katsushi Kitahara, Michael Lee, Dr. Alex Loftis, Dr. Aaron Mallek, Dr. Eva Maria Lopez Vidal, Dr. Rahi Masoom Reja, Dr. Ed Miller, Roman Misteli, Prof. Sebastian Pomplun, Dr. Anthony Quartatato, Dr. Simon Rossler, Dr. Azin Saebi, David Sarabia-Castillo, Prof. Kohei Sato, Adeline Schmitt, Daniel Sharygin, Dr. Elee Shimshoni, Prof. Chris Shugrue, Dr. Bente Somsen, Dr. Hiro Takeuchi, Dr. Xuyu Tan, Dr. Jason Tao, Dr. Suan Tuang, Prof. Nicholas Vecchiarello, Dr. Wayne Vuong, Dr. Jeff Wong, Dr. Tom Wood, Xiyun Ye, Gha Young Lee, and Dr. Peiyuan Zhang.* To my Pentelute Lab colleagues, I admire and respect all of you tremendously.

I would also like to thank my friends - their unending support and encouragement were enormously impactful on my journey. In particular, it is hard for me to imagine my graduate career without *Jacob Rodriguez* and *Charlie Farquhar* - thank you for all of the laughs, snacks, complaints, and general tomfoolery. This Žižek quote sums it up quite nicely, "*When I really love someone, I can only show it by making aggressive and bad-taste remarks*".

Finally, I would like to thank my family. To Ellen, my best friend and wife, this journey was only possible through your unending support. Your companionship and love mean more to me than I can express. To my parents, for cultivating my curiosity, enabling me to pursue my interests, and loving me unconditionally, I am unendingly grateful. To my sister, Emma, her husband, Tom, and the entirety of my family, thank you for your encouragement, your support was more helpful than you could have imagined.

Table of Contents

List of Figures

List of Tables

Background and Overview

Introduction

Reliable access to biomacromolecules with abiotic chemical synthesis has been a long-standing goal in chemical research. It was recognized in the early 19'th century that accurate models of biopolymer structure could only be found through the study of chemically defined model substrates (*1*). These requirements stimulated the development of chemical means to access model peptides and oligonucleotides. By the 1920s, reliable access to synthetic peptides with the newly-developed carbobenzoxy method of peptide synthesis (*2*) played a crucial role in developing the modern understanding of protein structure (*3*, *4*). Similarly, the development of synthetic techniques for producing oligonucleotides in the 1930s was essential for deciphering the codon code (*5*). In the time since these early discoveries, synthetic peptides and oligonucleotides have become universal tools to advance chemical and biological understanding (*6*). Modern research efforts require increasingly complicated biopolymer structures, thus necessitating the development of ever more efficient synthetic techniques.

Modern chemical methodology has further broadened access to increasingly large and diverse synthetic biopolymers. Once available only to experts, chemically synthesized peptides (*7*) and entire gene sequences hundreds of nucleobases long (*8*, *9*) are now readily commercially available. Critical to this commercialization was the development of numerous multi-disciplinary advances (*10*), including solid-phase synthesis techniques (*11*), efficient chromatographic strategies (*12*), and reliable high-resolution mass spectra instruments (*13*). However, these advances are not universally applicable. For example, backbones that require non-standard coupling chemistries or protein sequences over ~50 amino acids in length remain challenging to prepare (*14*) despite years of development. Such synthetic challenges limit the use of these biopolymers in advancing biological understanding.

New techniques are needed to improve access to synthetically challenging but valuable biopolymers. This chapter will discuss the applications and synthetic shortcomings of three such classes: full-length proteins, phosphorodiamidate morpholino oligomers, and peptide nucleic acids. The remaining chapters of this work will outline the development of new synthetic pipelines to prepare these biomacromolecules.

Synthetic Oligonucleotides

Modulation of gene expression with antisense oligonucleotides (ASOs) is an emerging therapeutic strategy. ASOs are short, synthetic oligomers that use base-pairing interactions to bind native RNA targets and modulate their activity. In contrast to traditional therapeutic approaches, ASOs derive target specificity through the ordering of their constituent monomers (*15*). Addressing different RNA targets is accomplished through simple reordering of the oligomer sequence to be the complement of the target sequence. As a result, these so-called “informational drugs” can address diverse diseases with minimal re-development (*16*). This property stimulated numerous drug discovery campaigns in the 1990s that were ultimately unsuccessful (*17*). Since then, new technological developments, including stabilizing chemical modifications and new delivery techniques, have made ASO drug development practical (*17*, *18*). With these improvements, renewed interest in informational drugs has led to the development of 18 FDA-approved nucleic acid therapeutics as of March 2023 (*19*).

Antisense oligonucleotides can be used in several mechanisms to modulate genetic information, and steric-blocking is an attractive therapeutic approach. In this mechanism of action, ASOs are engineered to tightly bind a substrate RNA at a region that would otherwise carry out some function (*20*). When bound, the RNA region is prevented from its function as the ASO sterically blocks certain active conformations. The nature of the impaired function is application-specific; for example, blocking intron-exon splice sites as for Duchenne Muscular Dystrophy (DMD) (*21*), or blocking transcription initiation sites as demonstrated for several viruses, including Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (*22*–*24*). The steric-blocking approach is not suitable for all disease areas, but it allows for significant chemical flexibility of the ASO backbone. In particular, several unique chemical modifications that impart increased selectivity for the correct RNA sequence and substantially improve resistance to endonucleases are compatible with this mechanism of action only. These properties are highly valuable, and as a result, the five FDA-approved steric-block ASOs (Exondys 51, Spinraza, Exondys 53, Viltepso, and Amondys 45) make extensive use of these chemical modifications (*19*).

Synthetic oligonucleotide classes with modified backbone architecture are valuable steric-block ASOs, but their synthesis remains challenging. Two such classes of backbone modification are found in phosphorodiamidate morpholino oligomers (PMOs) (*25*) and peptide nucleic acids (PNAs) (*26*). PMO backbones replace the 5-membered ribosyl ring with a 6-membered morpholinyl ring and phosphate linkages with uncharged phosphorodiamidates (*27*). These modifications impart significant nuclease resistance (*28*) and decreased immunogenicity while retaining binding affinity to RNA (*29*). Similarly, PNAs replace both the phosphate and ribose but instead replace them with a peptidic structure. These modifications also result in nuclease resistance, but in addition result in greatly increased specificity and affinity for its cognate RNA sequence (*30*). Because both of these modifications are of the backbone, they make these compounds incompatible with traditional oligonucleotide synthesis approaches. The remaining compatible synthetic strategies frequently require individualized optimization for each sequence. As a result, integrating these compound classes into drug development pipelines remains challenging.

The chemical synthesis of both PMOs and PNAs share similarities but have unique complications. Both backbones require total chemical synthesis using a solid phase support and share the same general synthesis scheme. Growing oligomer chains are covalently attached to a swellable solid polymer support, and monomers are added one at a time. After each monomer incorporation, a deprotection step reveals a chain growth site on the newly added monomer. Repetition of these steps affords full-length PMO or PNA chains that are then cleaved from the solid-phase resin and subsequently purified. Despite these similarities, both polymers have unique challenges. PMO synthesis presents two primary complications: monomer addition is slow, and the PMO chain is sensitive to forcing chemical conditions. As a result, compatible synthesis protocols require up to 180 mins for a single monomer incorporation, and entire sequence syntheses can take up to a week (*31*). PNAs face two separate challenges: aggregation of the growing chain and a number of competitive side reactions. Therefore, inefficient syntheses precluding certain PNA sequences' investigation are common (*32*). Generalized access to PMOs and PNAs faces significant chemical challenges.

Synthetic Proteins

Proteins carry out many processes essential to life, and their study is a cornerstone of modern biological research. Humans express more than 20,000 different proteins (*33*) to carry out numerous essential cellular functions, and the systematic characterization of these building blocks has driven fundamental insights into several disciplines, including the origins human disease (*34*). Among these techniques developed to prepare these essential protein substrates (*35*), chemical synthesis is unique in its ability to incorporate chemical modifications.

Protein total synthesis has emerged as a powerful tool to produce chemically modified proteins for the advancement of protein structure-function understanding. In contrast to recombinant expression where proteins are limited to the canonical 20 amino acids, synthetic proteins can incorporate any number of the numerous commercially available Fmoc-protected amino acid monomers. As a result, synthetic proteins can easily display modified functional groups that would be otherwise inaccessible (*36*, *37*). The ability of precisely defined modified proteins to inform biological understanding, for example in the characterization of lysine modifications in chromatin signaling (*38*, *39*), or in the role of backbone hydrogen bonding in phase separation (*40*) has stimulated interest in further enabling access to this technique. Despite growing interest, application of protein synthesis to new biological problems is limited by protein synthesis throughput. Current methods for the preparation of synthetic proteins require significant expertise, and as a result are only accessible to subject matter experts.

Protein synthesis is, in theory, applicable to a majority of small to medium proteins, but real-world execution of these techniques remains complicated and low throughput. After ~50 years of development, multi-fragment assembly is the preeminent strategy for protein synthesis (*14*, *41*). In this approach, protein sequences are separated into 30-50 amino acid lengths, which are then synthesized with Fmoc based solid-phase peptide synthesis (Fmoc-SPPS). Each fragment is individually purified and then joined with a neighboring fragment via dovetailing chemical moieties installed at the termini of the peptide fragments (*41*, *42*). The cycle of fragment ligation and purification is continued until the full-length protein is produced. With careful planning of fragment sites and ligation chemistries, proteins of enormous complexity can be produced, for example, in the

production of the highly glycosylated erythropoietin (*43*), or of the 53 kDa tetra-ubiquitin-α-globin construct (*44*).

In contrast to these feats of chemical engineering, more routine applications of synthetic proteins focus on smaller, more accessible protein targets under ~150 AA where synthetic planning is more straightforward. However, even for these small targets, typical multi-fragment assembly pathways require 2-3 ligations and 3-5 purification steps in total (*41*, *45*). Considering that each purification and ligation step is a potential site for non-ideal behavior (*36*), route optimization and troubleshooting are ubiquitous, even for small proteins. As a result, routine access to synthetic proteins requires years of hands-on experience and is a key limiter in protein synthesis' uptake by other scientific disciplines.

Thesis Overview

This introductory chapter has provided a primer on the use and preparation of synthetic biomacromolecules. **Chapter 1** will focus on the construction, evaluation, and application of an automated fast-flow instrument for the rapid production of phosphorodiamidate morpholino oligomers. **Chapter 2** outlines the development and application of a pipeline to prepare mirror image proteins to supply mirror image phage display. **Chapter 3** describes efforts to further refine synthetic protein production with the development of a new purification pipeline that enables single-day isolation. Finally, **Chapter 4** focuses on the construction of an automated fast-flow synthesis platform for the rapid production of peptide-peptide-nucleic-acid conjugates (PPNAs).

References

1. E. Fischer, E. Fourneau, Ueber einige Derivate des Glykocolls. *Berichte der deutschen chemischen Gesellschaft*. **34**, 2868–2877 (1901).

2. J. S. Fruton, The carbobenzoxy method of peptide synthesis. *Trends in Biochemical Sciences*. **7**, 37–39 (1982).

3. F. Sanger, "The Arrangement of Amino Acids in Proteins" in *Advances in Protein Chemistry*, M. L. Anson, K. Bailey, J. T. Edsall, Eds. (Academic Press, 1952; https://www.sciencedirect.com/science/article/pii/S0065323308600170), vol. 7, pp. 1–67.

4. J. S. FRUTON, Early Theories of Protein Structure. *Annals of the New York Academy of Sciences*. **325**, 1–20 (1979).

5. H. G. Khorana, Synthesis in the study of nucleic acids. The Fourth Jubilee Lecture. *Biochemical Journal*. **109**, 709–725 (1968).

6. J.-F. Lutz, M. Ouchi, D. R. Liu, M. Sawamoto, Sequence-Controlled Polymers. *Science*. **341**, 1238149 (2013).

7. M. Amblard, J.-A. Fehrentz, J. Martinez, G. Subra, "Fundamentals of Modern Peptide Synthesis" in *Peptide Synthesis and Applications*, J. Howl, Ed. (Humana Press, Totowa, NJ, 2005; https://doi.org/10.1385/1-59259-877-3:003), pp. 3–24.

8. S. Kosuri, G. M. Church, Large-scale de novo DNA synthesis: technologies and applications. *Nature Methods*. **11**, 499–507 (2014).

9. A. Hoose, R. Vellacott, M. Storch, P. S. Freemont, M. G. Ryadnov, DNA synthesis technologies to close the gene writing gap. *Nature Reviews Chemistry*. **7**, 144–161 (2023).

10. M. Stawikowski, G. B. Fields, *Current Protocols in Protein Science*, in press, doi:10.1002/0471140864.ps1801s69.

11. B. Merrifield, Solid Phase Synthesis. *Science*. **232**, 341–347 (1986).

12. F. E. Regnier, HPLC of proteins, peptides, and polynucleotides. *Anal. Chem.* **55**, 1298A-1306A (1983).

13. J. Griffiths, A Brief History of Mass Spectrometry. *Anal. Chem.* **80**, 5678–5683 (2008).

14. S. B. H. Kent, Total chemical synthesis of proteins. *Chem. Soc. Rev.* **38**, 338–351 (2009).

15. J. S. COHEN, Informational Drugs: A New Concept in Pharmacology. *Antisense Research and Development*. **1**, 191–193 (1991).

16. J. Lieberman, Tapping the RNA world for therapeutics. *Nature Structural & Molecular Biology*. **25** (2018), pp. 357–364.

17. A. Khvorova, J. K. Watts, The chemical evolution of oligonucleotide therapies of clinical utility. *Nature Biotechnology*. **35** (2017), pp. 238–248.

18. W. B. Wan, P. P. Seth, The Medicinal Chemistry of Therapeutic Oligonucleotides. *Journal of Medicinal Chemistry*. **59** (2016), pp. 9645–9667.

19. M. Egli, M. Manoharan, Chemistry, structure and function of approved oligonucleotide therapeutics. *Nucleic Acids Research*. **51**, 2529–2573 (2023).

20. G. F. Deleavey, M. J. Damha, Designing Chemically Modified Oligonucleotides for Targeted Gene Silencing. *Chemistry & Biology*. **19** (2012), pp. 937–954.

21. J. C. T. van Deutekom, G.-J. B. van Ommen, Advances in Duchenne muscular dystrophy gene therapy. *Nature Reviews Genetics*. **4** (2003), pp. 774–783.

22. C. Li, A. J. Callahan, M. D. Simon, K. A. Totaro, A. J. Mijalis, K.-S. Phadke, G. Zhang, N. Hartrampf, C. K. Schissel, M. Zhou, H. Zong, G. J. Hanson, A. Loas, N. L. B. Pohl, D. E. Verhoeven, B. L. Pentelute, Fully automated fast-flow synthesis of antisense phosphorodiamidate morpholino oligomers. *Nature Communications*. **12**, 4396 (2021).

23. K. L. Holden, D. A. Stein, T. C. Pierson, A. A. Ahmed, K. Clyde, P. L. Iversen, E. Harris, Inhibition of dengue virus translation and RNA synthesis by a morpholino oligomer targeted to the top of the terminal 3′ stem–loop structure. *Virology*. **344**, 439–452 (2006).

24. P. L. Iversen, T. K. Warren, J. B. Wells, N. L. Garza, D. V. Mourich, L. S. Welch, R. G. Panchal, S. Bavari, Discovery and early development of AVI-7537 and AVI-7288 for the treatment of Ebola virus and Marburg virus infections. *Viruses*. **4**, 2806–2830 (2012).

25. J. E. Summerton, "Invention and Early History of Morpholinos: From Pipe Dream to Practical Products" in *Morpholino Oligomers: Methods and Protocols*, H. M. Moulton, J. D. Moulton, Eds. (Springer New York, New York, NY, 2017; https://doi.org/10.1007/978-1-4939-6817-6_1), pp. 1–15.

26. M. Egholm, O. Buchardt, P. E. Nielsen, R. H. Berg, Peptide nucleic acids (PNA). Oligonucleotide analogs with an achiral peptide backbone. *Journal of the American Chemical Society*. **114** (1992), pp. 1895–1897.

27. J. D. Moulton, Y.-L. Yan, *Current Protocols in Molecular Biology*, in press, doi:10.1002/0471142727.mb2608s83.

28. R. M. HUDZIAK, E. BAROFSKY, D. F. BAROFSKY, D. L. WELLER, S.-B. HUANG, D. D. WELLER, Resistance of Morpholino Phosphorodiamidate Oligomers to Enzymatic Degradation. *Antisense and Nucleic Acid Drug Development*. **6**, 267–272 (1996).

29. J. SUMMERTON, D. WELLER, Morpholino Antisense Oligomers: Design, Preparation, and Properties. *Antisense and Nucleic Acid Drug Development*. **7** (1997), pp. 187–195.

30. J. Saarbach, P. M. Sabale, N. Winssinger, Peptide nucleic acid (PNA) and its applications in chemical biology, diagnostics, and therapeutics. *Current Opinion in Chemical Biology*. **52**, 112–124 (2019).

31. C. M. J. Fox, D. D. Weller, Method of synthesis of morpholino oligomers (2012), (available at https://patents.google.com/patent/US8299206/en?oq=8299206).

32. J. Tailhades, H. Takizawa, M. J. Gait, D. A. Wellings, J. D. Wade, Y. Aoki, F. Shabanpoor, Solid-Phase Synthesis of Difficult Purine-Rich PNAs through Selective Hmb Incorporation: Application to the Total Synthesis of Cell Penetrating Peptide-PNAs. *Frontiers in Chemistry*. **5** (2017), , doi:10.3389/fchem.2017.00081.

33. M. Zahn-Zabal, P.-A. Michel, A. Gateau, F. Nikitin, M. Schaeffer, E. Audot, P. Gaudet, P. D. Duek, D. Teixeira, V. Rech de Laval, K. Samarasinghe, A. Bairoch, L. Lane, The neXtProt knowledgebase in 2020: data, tools and usability improvements. *Nucleic Acids Research*. **48**, D328–D334 (2020).

34. M. Uhlen, C. Zhang, S. Lee, E. Sjöstedt, L. Fagerberg, G. Bidkhori, R. Benfeitas, M. Arif, Z. Liu, F. Edfors, K. Sanli, K. von Feilitzen, P. Oksvold, E. Lundberg, S. Hober, P. Nilsson, J. Mattsson, J. M. Schwenk, H. Brunnström, B. Glimelius, T. Sjöblom, P.-H. Edqvist, D. Djureinovic, P. Micke, C. Lindskog, A. Mardinoglu, F. Ponten, A pathology atlas of the human cancer transcriptome. *Science*. **357**, eaan2507 (2017).

35. S. Gräslund, P. Nordlund, J. Weigelt, B. M. Hallberg, J. Bray, O. Gileadi, S. Knapp, U. Oppermann, C. Arrowsmith, R. Hui, J. Ming, S. dhe-Paganon, H. Park, A. Savchenko, A. Yee, A. Edwards, R. Vincentelli, C. Cambillau, R. Kim, S.-H. Kim, Z. Rao, Y. Shi, T. C. Terwilliger, C.-Y. Kim, L.-W. Hung, G. S. Waldo, Y. Peleg, S. Albeck, T. Unger, O. Dym, J. Prilusky, J. L. Sussman, R. C. Stevens, S. A. Lesley, I. A. Wilson, A. Joachimiak, F. Collart, I. Dementieva, M. I. Donnelly, W. H. Eschenfeldt, Y. Kim, L. Stols, R. Wu, M. Zhou, S. K. Burley, J. S. Emtage, J. M. Sauder, D. Thompson, K. Bain, J. Luz, T. Gheyi, F. Zhang, S. Atwell, S. C. Almo, J. B. Bonanno, A. Fiser, S. Swaminathan, F. W. Studier, M. R. Chance, A. Sali, T. B. Acton, R. Xiao, L. Zhao, L. C. Ma, J. F. Hunt, L. Tong, K. Cunningham, M. Inouye, S. Anderson, H. Janjua, R. Shastry, C. K. Ho, D. Wang, H. Wang, M. Jiang, G. T. Montelione, D. I. Stuart, R. J. Owens, S. Daenke, A. Schütz, U. Heinemann, S. Yokoyama, K. Büssow, K. C. Gunsalus, Structural Genomics Consortium, Architecture et Fonction des Macromolécules Biologiques, Berkeley Structural Genomics Center, China Structural Genomics Consortium, Integrated Center for Structure and Function Innovation, Israel Structural Proteomics Center, Joint Center for Structural Genomics, Midwest Center for Structural Genomics, New York Structural GenomiX Research Center for Structural Genomics, Northeast Structural Genomics Consortium, Oxford Protein Production Facility, M. D. C. for M. M. Protein Sample Production Facility, RIKEN Structural Genomics/Proteomics Initiative, SPINE2-Complexes, Protein production and purification. *Nature Methods*. **5**, 135–146 (2008).

36. Y. Tan, H. Wu, T. Wei, X. Li, Chemical Protein Synthesis: Advances, Challenges, and Outlooks. *J. Am. Chem. Soc.* **142**, 20288–20298 (2020).

37. S. Kent, Chemical protein synthesis: Inventing synthetic methods to decipher how proteins work. *Bioorganic & Medicinal Chemistry*. **25**, 4926–4937 (2017).

38. M. M. Müller, T. W. Muir, Histones: At the Crossroads of Peptide and Protein Chemistry. *Chem. Rev.* **115**, 2296–2349 (2015).

39. R. K. McGinty, J. Kim, C. Chatterjee, R. G. Roeder, T. W. Muir, Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. *Nature*. **453**, 812–816 (2008).

40. X. Zhou, L. Sumrow, K. Tashiro, L. Sutherland, D. Liu, T. Qin, M. Kato, G. Liszczak, S. L. McKnight, Mutations linked to neurological disease enhance self-association of low-complexity protein sequences. *Science*. **377**, eabn5582.

41. A. C. Conibear, E. E. Watson, R. J. Payne, C. F. W. Becker, Native chemical ligation in protein synthesis and semi-synthesis. *Chem. Soc. Rev.* **47**, 9046–9068 (2018).

42. S. S. Kulkarni, J. Sayers, B. Premdjee, R. J. Payne, Rapid and efficient protein synthesis through expansion of the native chemical ligation concept. *Nature Reviews Chemistry*. **2**, 0122 (2018).

43. P. Wang, S. Dong, J.-H. Shieh, E. Peguero, R. Hendrickson, M. A. S. Moore, S. J. Danishefsky, Erythropoietin Derived by Chemical Synthesis. *Science*. **342** (2013), pp. 1357–1360.

44. H. Sun, A. Brik, The Journey for the Total Chemical Synthesis of a 53 kDa Protein. *Acc. Chem. Res.* **52**, 3361–3371 (2019).

45. T. Wei, H. Liu, H. Wu, F. Pu, X. Li, Serine/threonine ligation-assisted chemical synthesis of HMGA1a protein with site-specific post-translational modifications. *STAR Protocols*. **2**, 100777 (2021).

Discussed Publications

Chapter 1:

1. **Callahan, A.J.***; Li, C.*; 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. Automated Fast-Flow Synthesis of Antisense Phosphorodiamidate Morpholino Oligomers. *Nat. Comm*. **2021**, 12, 1, 1-8.

Chapter 2:

2. **Callahan, A.J.**; Gandhesiri, S.; Travaline, T.L., Lozano Salazar, L.; Hanna, S.; Lee, Y.-C.; Tokareva, O. S.; Swiecicki, J.-M.; Loas, A.; Verdine, G.L.; Pentelute, B.L. Single-Shot Flow Synthesis of D-Proteins for Mirror-Image Phage Display. *ChemRxiv*; **2023**; This content is a preprint and has not been peer-reviewed

Chapter 3:

3. **Callahan, A.J.**; Rondon, A.; Lozano Salazar, L.; Gandhesiri, S.; Rodriguez, J.; Loas, A.; Pentelute, B.L. Same Day Access to Proteins Enabled by Folding Selections. *ChemRxiv*; **2023**; This content is a preprint and has not been peer-reviewed.

Chapter 4:

4. 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 Central Science* 2022, 8, 2, 205-213

1. Fully Automated Fast-Flow Synthesis of Antisense Phosphorodiamidate Morpholino Oligomers

The work presented in this chapter has been reproduced from the following publication with the permission of Springer Nature:

Callahan, A.J.*; Li, C.*; 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. Automated Fast-Flow Synthesis of Antisense Phosphorodiamidate Morpholino Oligomers. *Nat. Comm*. **2021**, 12, 1, 1-8.

**These authors contributed equally to this work.*

1.1. Introduction

New and emerging pathogens, such as COVID-19, necessitate rapid drug development, and typical drug development pipelines are ill-suited to meet these demands. In contrast, antisense drugs' logical and rapid development makes them well-suited to this view (*1*–*3*). The strength of such a strategy is highlighted by the emergence of SARS-CoV-2, wherein oligonucleotide-based drugs were among the first treatment types to enter human trials as early as two months after the virus's first reports (*4*–*6*). Although these results are promising, rapid preclinical development of antisense drugs necessitates new technologies to enable high-throughput drug screening.

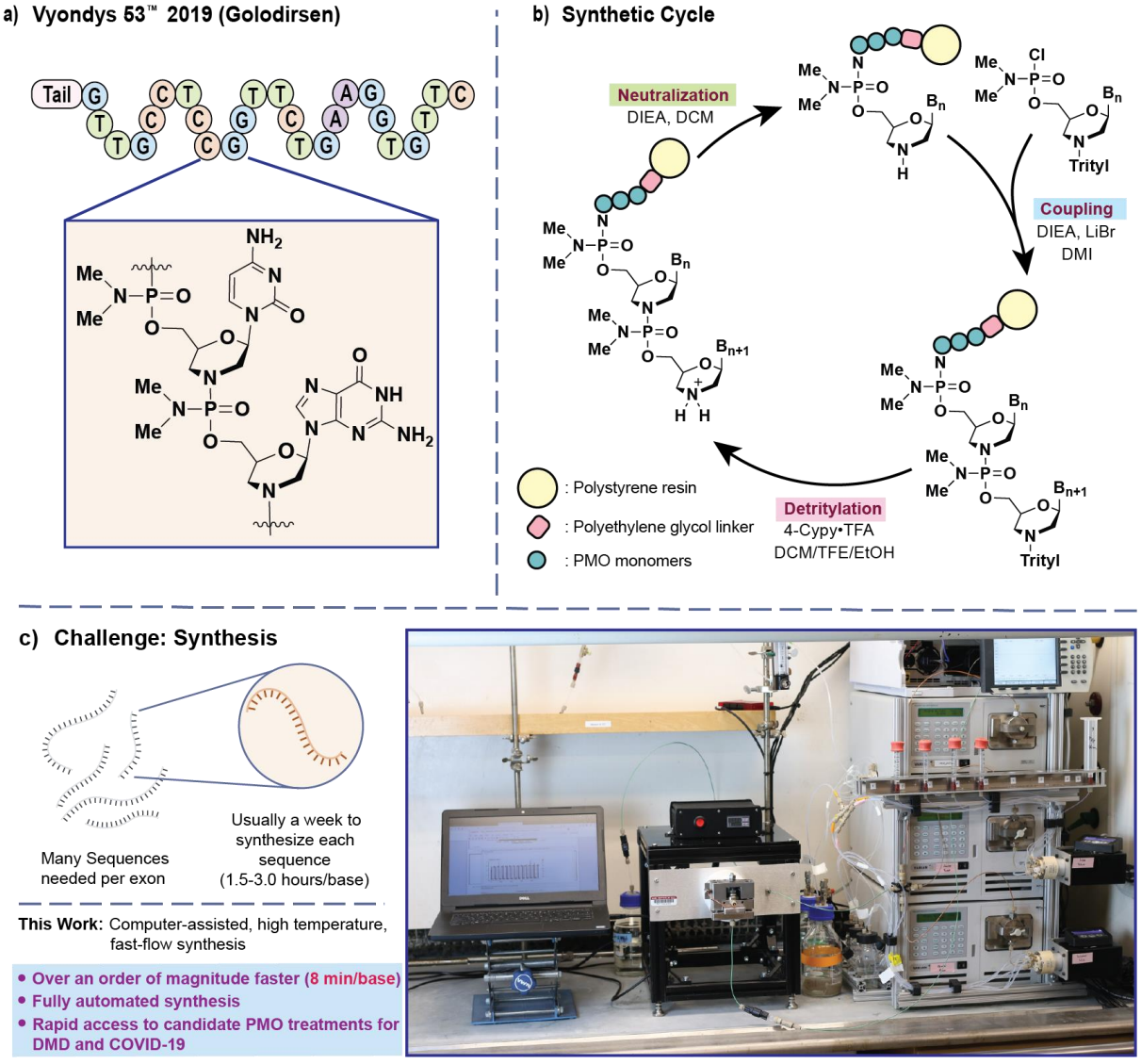


Figure 1.1 Automated fast-flow technology can produce on-demand customized PMO sequences.

a) The structure of PMO drug Golodirsen is shown. b) Each nucleotide is incorporated as a trityl-protected phosphoramido chloridate in 1,3-dimethyl-2-imidazolidinone (DMI). Detritylation frees the 3'-terminal amine with the conjugate acid of a non-nucleophilic heterocycle, typically 4-cyanopyridine in dichloromethane-trifluoroethanol (DCM/TFE) mixtures. Neutralization of the 3'-amine prepares the chain for the next coupling reaction using tertiary amines in DCM/isopropanol mixtures. c) Development of PMO drugs requires testing of many sequences and laborious production of sequential revisions. An automated fast-flow synthesizer allows for rapid manufacture of PMO sequences at 90 °C.

Antisense therapeutics are synthetic RNA mimetics that bind to mRNA via complementary base-pairing interactions to modulate transcription (*1*). Because their design is informational in nature, they can target virtually any genetic disease in a rapid and logical fashion (*7*). Among antisense compound types, phosphorodiamidate morpholino oligomers (*8–10*) (PMOs) are the most tested as antiviral agents, with clinical trials already underway for treatments against Dengue (*11*), Marburg (*12*), Ebola (*12*), Influenza (*13*), West Nile (*14*), and SARS-CoV-1 (*15*, *16*) viruses. PMO backbones are derived from RNA, wherein the five-membered ribosyl ring has been replaced with a six-membered morpholino ring, and phosphate linkages have been replaced with uncharged phosphorodiamidates (Fig. 1.1a). These modifications make PMOs resistant to nucleases (*9*), while retaining strong binding affinity for target RNA. Although PMOs are known to be effective antisense candidates, robust screening efforts for efficient transition to the clinic have remained elusive. Tedious synthesis protocols limit the production of screening libraries needed for sequence optimization, and development timelines remain long.

Changes to the chemistry of PMO synthesis are greatly needed to enable the rapid drug development. PMOs are synthesized from the 5′- to the 3′-end on a crosslinked polystyrene solid support (*8*, *17–19*) (Fig. 1.1b), with coupling times on the order of 180 min (*17*, *20*). With therapeutic PMO sequences on the order of 20 residues, synthesis times are on the order of weeks. Unsurprisingly, the production of screening libraries places a significant burden on the development of PMO drugs (Fig. 1.1c) due to lengthy protocols. To address this limitation, it is common to run synthesis reactions on automated systems (*17*). In theory, such instruments that combine automation with flow chemistry can mix and heat reagents with efficiencies that are unattainable by batch methods (*21–*

23). Still, application of these potential advantages to PMO synthesis has remained elusive. Further work is needed to improve access to PMO compounds.

Rapid production of PMOs will be required to expand their application to new disease types. With the emergence of SARSCoV-2 in late 2019, new techniques to develop drugs on a rapid timescale are greatly needed, and PMOs could find great application here. Accordingly, PMOs have been reported to treat SARS-CoV-1 (*15*, *16*), but progress at the time was too slow to justify further development. The use of the same strategy to inhibit SARS-CoV-2 (Fig. 1.2a) should be possible in principle, but long development timelines have left this modality underexplored. Even for well-characterized strategies where PMOs have been successful, development has been slow. PMO treatments for Duchenne muscular dystrophy (DMD) are well-understood (*24–26*) (Fig. 1.2b) and applicable to a majority of DMD subtypes (*27*, *28*), but commercial treatments are available to only a small fraction of these.

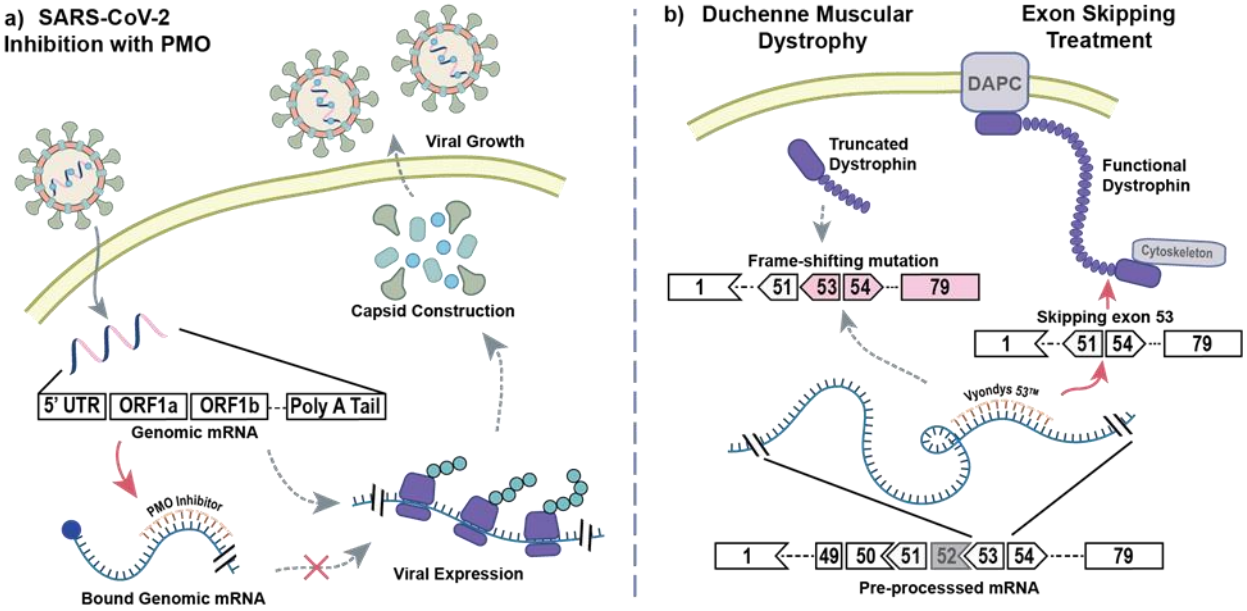


Figure 1.2 Antisense PMO therapeutics are adaptable to a range of disease types.

a) An antisense PMO binding to the 5' UTR of the SARS-CoV-2 genomic transcript can prevent expression of viral genes and subsequently inhibit viral growth. b) Dystrophin is an integral membrane protein that anchors the cytoskeleton to the muscle cell membrane via the dystrophin-associated protein complex (DAPC). Little to no natural dystrophin is produced in patients with Duchenne muscular dystrophy (DMD). The PMO Vyondys 53™ induces skipping of exon 53 to regain the proper reading frame, producing shorter but functional dystrophin.

In this work, we disclose an automated instrument that can expedite PMO synthesis by over an order of magnitude using chemistry done in flow. Enhanced reaction rates on the reported flow synthesizer result in nucleobase coupling in only 8 min. With this technique, we produce candidate PMOs for diseases with no current treatment in a matter of hours with our automated synthesis platform. A PMO designed to bind to the transcription regulatory stem-loop of the SARS-CoV-2 genomic mRNA to inhibit the production of viral proteins is synthesized in only 3.5 h. This antiviral compound demonstrates highly specific inhibition of the native SARS-CoV-2 virus in live cells. Further optimization of this compound could provide a greatly needed stop-gap therapeutic for the most severe cases of COVID-19. The optimized synthesis protocol can also accelerate the drug development timelines for established applications of PMOs. We synthesize three therapeutic candidates for DMD in a single day. The automation technology demonstrated here can open the door to enable rapid screening of PMO drug candidates for all possible DMD subtypes.

1.2. Results

1.2.1. Stability testing of PMO synthesis reagents

1.2.2. Design of an automated microscale flow synthesizer

A microscale instrument was designed to carry out the flow synthesis of PMOs. The instrument was constructed from commercially available components and a machined reaction vessel, using a design similar to a previously reported fast-flow peptide synthesizer (*29*, *30*). The base design consists of six modules connected in series (Fig. 1.3a). The first module is a collection of glass containers with liquid reagents stored under nitrogen. Two chemically inert valves compose the second module. Under computer control, each valve chooses its input from the available reagents in module 1, updating throughout the run. The third module is composed of two HPLC pumps, each connected to one of the outputs from module 2. Each pump is capable of supplying up to 2.5 mL per minute limited by reagent viscosity and pump configuration. The output streams from module 3 meet in a T-mixer and then travel to module 4, the reaction vessel module. Flow enters module 4 and is passed through a 90 cm long metal tube over a heated aluminum core bringing the solution to temperature in ~2 s. Module 4 holds the solid-phase resin in a removable reactor chamber 1 mL in volume at the desired temperature. The preheated

flow passes through the resin, reacting with the growing PMO chains. Module 5 is a UV–vis detector used to monitor the composition of the spent reagent solution inline. Module 6 is a computer that controls all other modules, using a modular script in the Mechwolf programming environment (*31*).

This design allows for precise control of reaction conditions for microscale PMO synthesis. The reactor body is designed for a 4.4 µmol-scale synthesis, the equivalent of 10 mg of a medium loading resin (0.39–0.43 mmol/g). Reagent delivery is encoded using pump strokes and flow rate. Each stroke of the HPLC pump carries 40 µL of solvent and it takes 12 strokes from both pumps (960 µL) to reach the resin. Flow rate is adjusted by increasing both the time to deliver strokes, and the time between strokes. Reaction time on the resin can be increased by increasing the number of pump strokes, or by decreasing the flow rate once the reagents hit the resin. Clearance of regents from the reservoir takes 20 pump strokes (1.6 mL), so reaction steps are separated by washes of at least 20 strokes of the appropriate wash solvent. The optimized synthesis sequence with controls for each module can be seen in Fig. 1.3b. A discussion about optimization of individual parts of the instrument is provided in the Section 1.4.7.

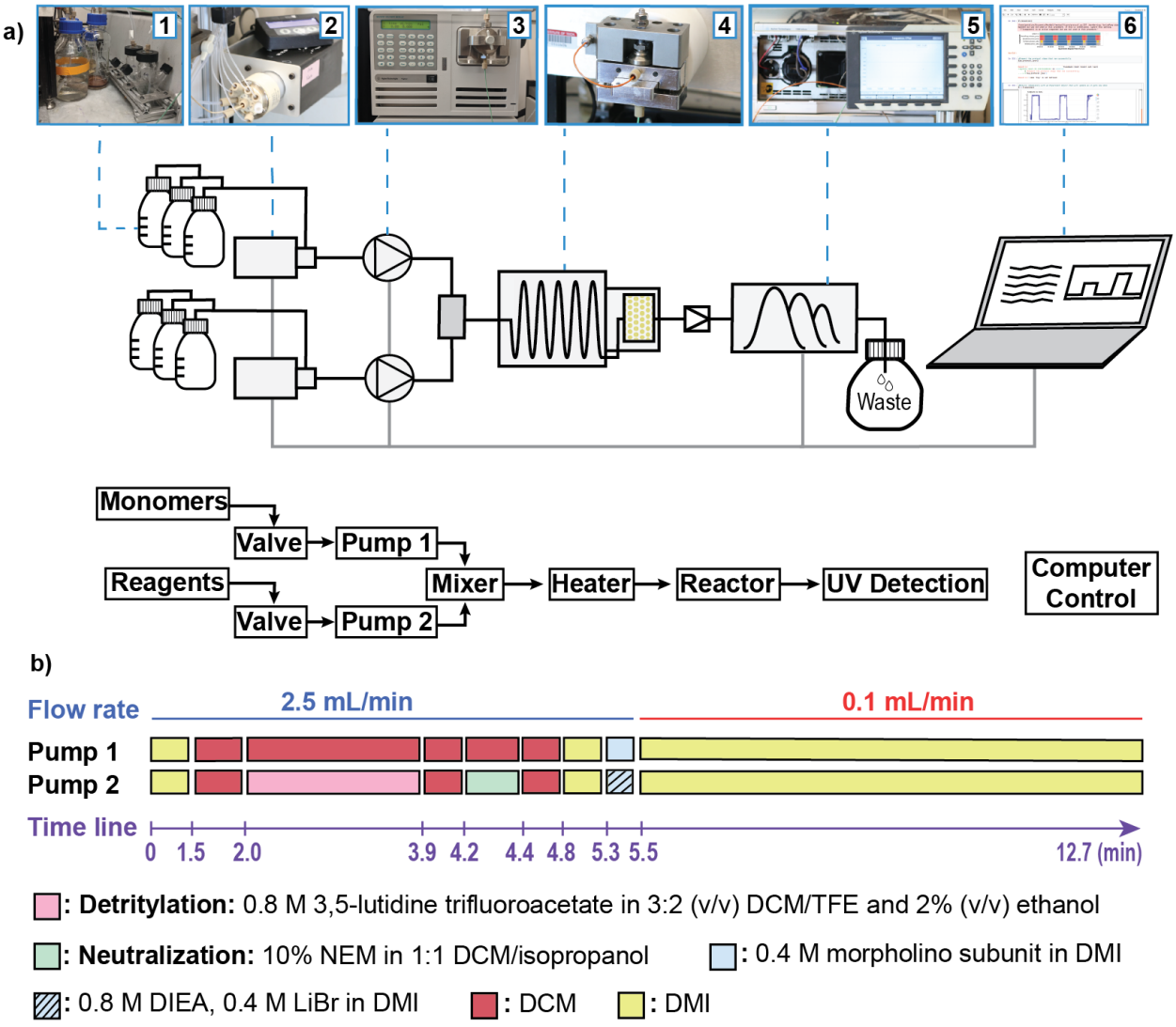


Figure 1.3 The automated fast-flow instrument features 6 modules that act to effect rapid PMO synthesis.

a) A schematic for the flow-path of the instrument is shown. Valves select the appropriate solutions from glass bottles under nitrogen pressure. Two HPLC pumps mix the reagents and flow them over a packed resin bed held at 90 °C. Effluent cools as it exits the reactor and passes through a back-pressure regulator to an in-line UV-Vis detector where reaction progress is monitored. A computer running an automated Python script controls the valves and pumps throughout the synthesis and tracks the instrument performance from the UV-Vis detector. b) The optimized protocol for PMO synthesis is shown. Instructions are delivered by the control program to the pumps and valves at each of the listed times. Bars represent activity of the two HPLC pumps, and colors indicate which reagents the valves are open to. Reagent stocks are prepared in double the concentration that is intended to hit the resin. Dilution from the second pump prepares the correct concentrations upon mixing.

1.2.3. Optimization of automated fast-flow synthesis

Iterative changes in flow synthesis variables were used to develop a flow recipe that can produce PMO sequences of similar purity during room temperature syntheses. Solid-phase PMO synthesis is sensitive to small variations in reaction efficiency, as the many reactions in series amplify off-target pathways. We took advantage of this amplification to optimize reactions in flow. For each variable of interest, we synthesized a 4-mer PMO, a process that involves twelve sequential reaction steps. Using LC–MS, we compared the crude purity of the products from each reaction sequence. Timings and reagents for each synthesis were as shown in Fig. 1.3b with the modifications listed in Fig. 1.4a. The actions that the Python script sends to modules 2` and 3 throughout the synthesis cycle are shown in Fig. 1.3b. The resulting resin-bound PMO product was cleaved from the solid support, the sample analyzed by LC–MS, and the relative levels of the product and high molecular weight side products were quantified, using a molecular feature extraction utility. Of special interest were side products arising from incomplete couplings, and we tracked their relative abundance separately.

a)

Entry	Deprotection Reagent	Neutralization Reagent	Coupling Base	Additive	Monomer Equivalents	Coupling Time (min)	Temperature (°C)	Crude Purity	A Deletion (%)	T Deletion (%)	C Deletion (%)	G Deletion (%)	Uncharacterized (%)
	Detritylation	*Neutralization*	*Coupling*				*Instrument*		*Side Products*				
1	4-Cypy•TFA	DIEA	DIEA	none	10	143	r.t.	**95%**	n.d.	n.d.	n.d.	n.d.	5.0
2	4-Cypy•TFA	DIEA	DIEA	none	10	17.4	70	**72%**	0.6	5.7	n.d.	9.9	11.8
3	4-Cypy•TFA	DIEA	DIEA	none	10	11.7	70	**74%**	0.7	5.6	0.3	11.3	8.1
4	4-Cypy•TFA	DIEA	DIEA	none	18	11.7	70	**85%**	n.d.	1.9	n.d.	3.6	9.5
5	4-Cypy•TFA	DIEA	DIEA	NMI	18	11.7	70	**87%**	n.d.	1.1	n.d.	1.3	10.6
6	4-Cypy•TFA	DIEA	DIEA	LiBr	18	11.7	70	**92%**	n.d.	0.9	n.d.	2.6	4.5
7	4-Cypy•TFA	DIEA	DIEA	LiBr	18	11.7	90	**82%**	n.d.	n.d.	n.d.	1.8	16.2
8	3,5-Lut•TFA	DIEA	DIEA	LiBr	18	12.6	90	**93%**	n.d.	n.d.	n.d.	1.5	5.5
9	3,5-Lut•TFA	DIEA	NEM	LiBr	18	12.6	90	**92%**	n.d.	n.d.	n.d.	1.5	6.5
10	3,5-Lut•TFA	NEM	DIEA	LiBr	18	12.6	90	**95%**	n.d.	n.d.	n.d.	1.3	3.7
11*	3,5-Lut•TFA	NEM	DIEA	LiBr	18	12.7	90	**99%**	n.d.	n.d.	n.d.	0.9	0.1
12*	3,5-Lut•TFA	NEM	DIEA	LiBr	10	22.1	90	**96%**	n.d.	n.d.	n.d.	1.5	2.5

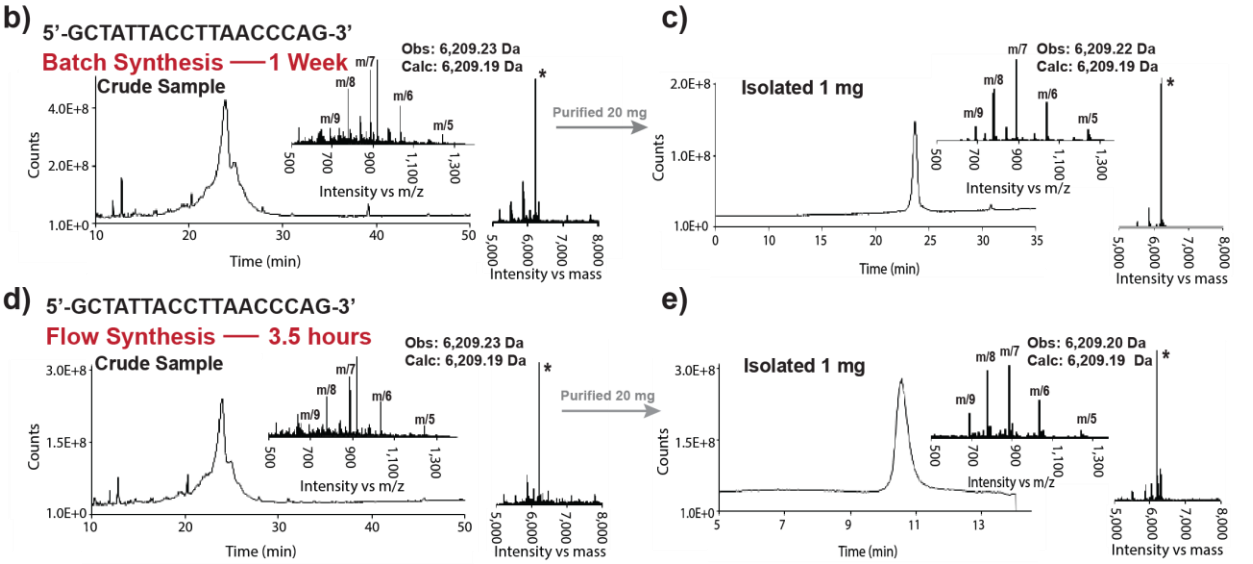


Figure 1.4 Optimized flow synthesis produces PMOs with similar purity to manual batch protocols.

*a) Synthesis variables were changed in sequence, and relative crude purities were tracked by comparison of the crude LC-MS traces. Crude purity and deletion sequences quantified as described in Section 1.4.6. *The Python script was edited to slowly deliver the monomers, reducing flow inaccuracies due to the viscous solution of the G monomer. N.d. = levels were below the sensitivity of detection with the described quantification method – not detected. Column colors indicate which category the variables correspond to: pink – detritylation, green – neutralization, blue – coupling, orange – instrument. Red lettering in line 11 denotes the best observed conditions. b) The total ion current chromatogram (TICC) of a batch-synthesized sample of the 18-mer PMO IVS2-654 is shown along with the mass spectrum and associated deconvoluted mass spectrum. c) The TICC of the batch-synthesized sample after purification by cation exchange*

chromatography is shown along with the mass spectrum and associated deconvoluted mass spectrum. d) The TICC, mass spectrum and deconvoluted mass spectrum are shown for the crude product from a flow synthesis of the same 18-mer sequence. e) The TICC of the flow-synthesized sample after purification along with the mass spectrum and deconvoluted mass spectrum. Note: Traces c and e were acquired using different LC-MS methods.

Use of the highest allowable temperature determined from monomer and resin-bound PMO stability studies, 90 °C, provided the cleanest crude PMO. Increasing the temperature of PMO synthesis significantly decreases reaction times, contingent on the stability of reagents. Temperatures of over 70 °C will accelerate both on-target and off-target reactions, with degradation of synthetic intermediates limiting the maximum possible synthesis temperature. To determine the maximum reaction temperature, we tracked the degradation of all synthetic components at a range of temperatures, and modified chemical variables to ensure stability of all components. Initial reaction screens were carried out at 70 °C, a milder temperature that enables use of the standard 4-cyanopyridine trifluoroacetate deprotection solution without the significant degradation found at 90 °C. Initial results from flow synthesis at 70 °C provided the desired material, but with a crude purity of 72%, lower than the benchmark 95% from room temperature syntheses (Fig. 1.4a, entries 1 vs 2). Further optimization with changes to the instrument command recipe, monomer equivalents, and coupling catalysts (*20*) improved crude purity to 92% (Fig. 1.4a, entries 3–6). Detectable levels of side products remained, so the improvement of the coupling reaction was still required. Although increasing monomer excess would likely increase crude purity, we capped this value at 18 equivalents, a generally accepted upper limit in academic and patent literature (*17*, *20*). We instead increased the temperature to 90 °C to improve coupling rates. Along with optimization of deprotection conditions, neutralization, and coupling bases (Fig. 1.4a, entries 7–10), the increase in temperature provided an optimized recipe that yields a crude purity of 99% (Fig. 1.4a, entry 11). This protocol was used for the production of the sequences reported in the remainder of this work.

The final experiment shown in Fig. 1.4a (entry 12) demonstrates that high temperature flow synthesis of PMOs does not necessarily require more monomer equivalents than room temperature protocols. The phosphoramido chloridate monomers are costly, and it is common to minimize the excess used. Using only ten equivalents of

monomer at high temperature is effective, but requires a longer coupling step with this hardware configuration.

Application of the optimized protocol enabled the rapid synthesis of an 18-mer PMO, which hybridizes to the β-thalassemia gene sequence, IVS2-654 (*32*, *33*). Using batch protocols, production of the PMO took one full week. The flow protocol enabled the production of the full PMO sequence in only 3.5 h, and the crude products were of comparable purity by LC–MS (Fig. 1.4b, d). From each synthesis 1 mg of PMO with >85% purity were obtained (10.1, 10.1, and 10.2% yields with respect to resin loading, 10 mg of 0.39–0.43 mmol/g; Fig. 1.4c, e).

1.2.4. Rapid microscale flow synthesis of potential DMD PMO therapeutics

We leveraged the platform for rapid production of three PMO sequences targeted to skip exon 46 of DMD pre-mRNA. Although exon skipping at this site has an important place in the history of splice alteration for DMD (*34*), no treatment options are available that target exon 46. Three sequences near splice acceptor and donor sites were chosen, as previously reported (*35*) (Fig. 1.5a). Using the automated fast-flow instrument, the three 20-mer sequences were synthesized in a single day in succession. In each case, after cleavage and purification, 1 mg of PMO material was isolated with >85% purity. Manufacturing these three sequences under batch conditions would take ~1 month if done sequentially (Fig. 1.5b).

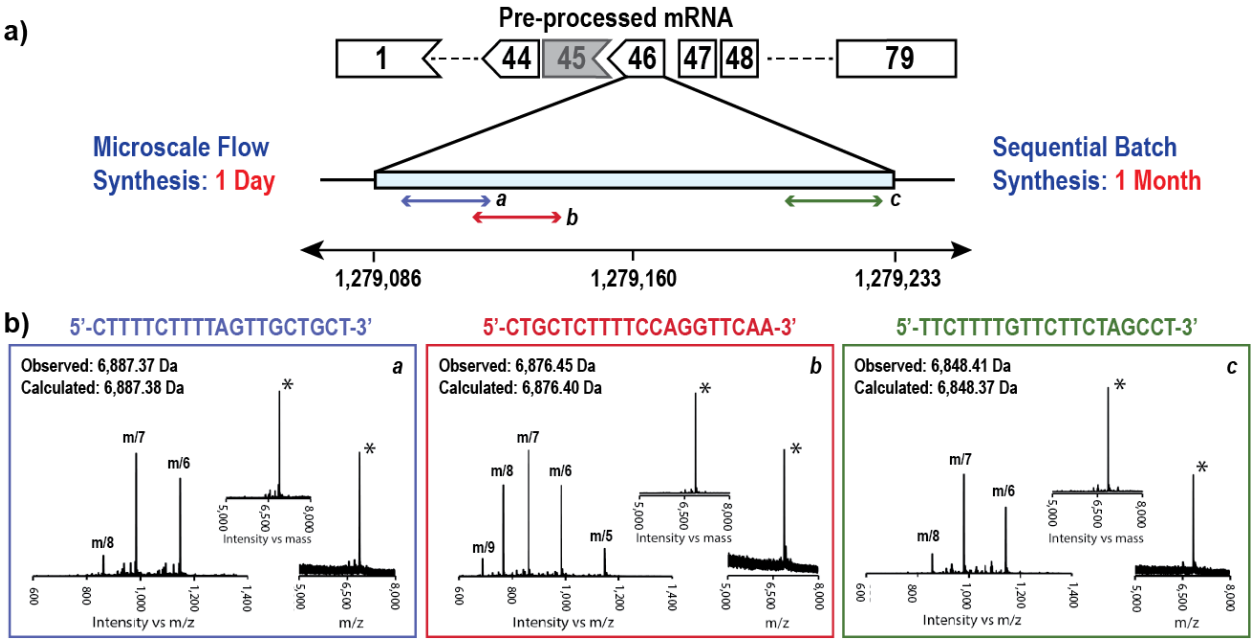


Figure 1.5 . Automated flow synthesis enables the rapid production of antisense PMO candidates to treat DMD.

a) Gene diagram showing the splicing sites where the three potential therapeutic sequences are targeted. Three sequences were chosen for targeting splice donor and acceptor sites of exon 46. b) The mass spectrum, deconvoluted mass spectrum and MALDI mass spectra of purified PMOs synthesized in flow is shown (see Supplementary Section 13.4).

1.2.5. Rapid synthesis and evaluation of anti-SARS-CoV-2 PMOs

PMOs used as steric block antisense compounds provide a potential route to inhibit the viral replication. We designed a PMO sequence to inhibit SARS-CoV-2 replication and synthesized it in 3.5 h. Using room temperature synthesis, manufacturing each new potential COVID-19 therapeutic would typically require at least a week. The 18-mer PMO was targeted to the transcription regulatory sequence (TRS) stem loop of the 5′ UTR of the genomic mRNA (Fig. 1.6a), a strategy that proved effective for treatment of the SARS-CoV-2 (*36*) and closely related SARS-CoV-1 (*15*, *16*). Synthesis and subsequent purification of one third of the crude sample yielded 0.5 mg of PMO with >90% purity (extrapolated yield of 5.5% with respect to resin loading, 10 mg of 0.39–0.43 mmol/g; Fig. 1.6b).

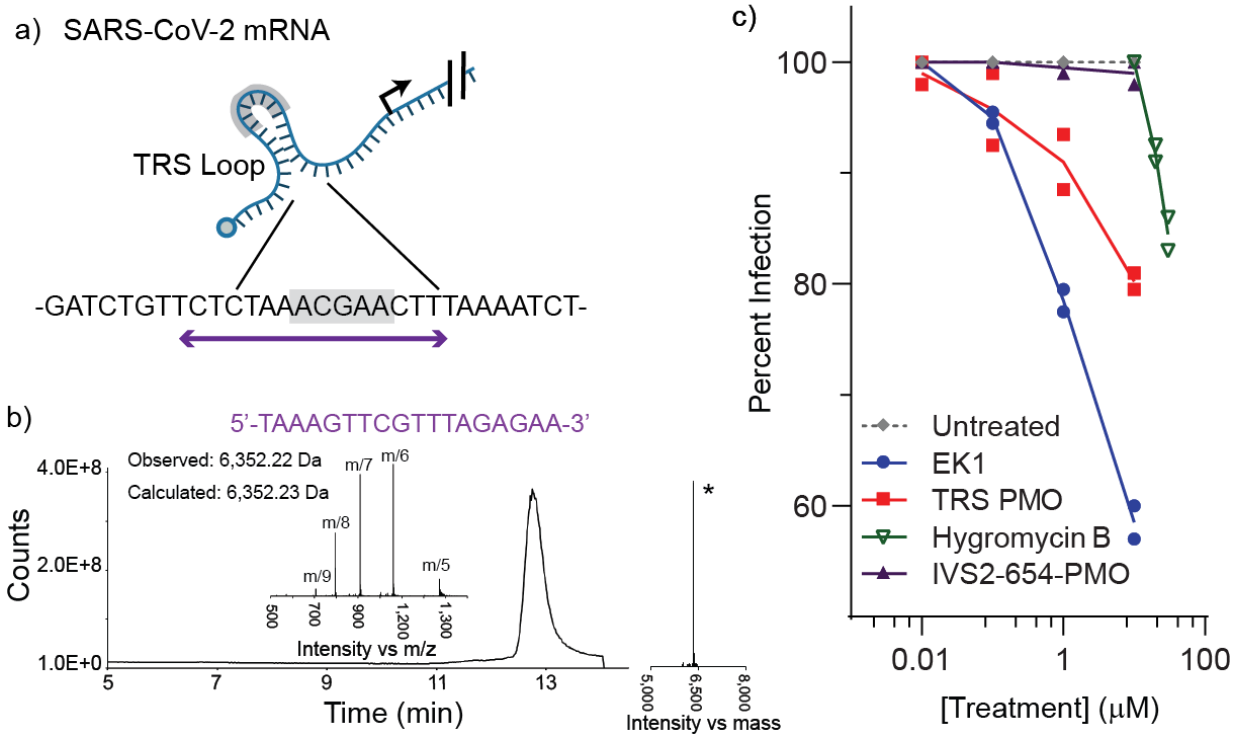


Figure 1.6 Synthetic PMO can be engineered to inhibit SARS-CoV-2 replication.

a) The 5' UTR TRS is conserved between coronaviruses, and is a target for antisense knockdown of viral replication. The TRS is shown in grey, and a sequence was chosen to encompass this region and surrounding bases, an effective strategy for targeting the SARS-CoV-1 mRNA. b) The TICC, mass spectrum, and deconvoluted mass spectrum of the purified PMO sequence is shown. c) Inhibition of native SARS-COV-2 by various treatments is shown. Percent inhibition represents observed viral reductions in culture supernatants as measured by qRT-PCR as compared to untreated control. Increasing concentrations of TRS-PMO results in inhibition of viral replication. Inhibition of viral growth is observed as previously reported for EK1 and hygromycin B. Measurement at each concentration is from n = 2 biological replicates, with each point averaging two technical replicates.

To test the antiviral activity of the synthesized PMO, we measured viral RNA levels after viral incubation with a low multiplicity of infection (MOI). Lower levels of viral RNA is consistent with slowed viral growth, and successful inhibition of viral protein production should lower the observed viral RNA. Prior to inoculation, Vero-E6 cells were incubated with either the TRS-PMO, IVS2-654-PMO as a negative control having no intracellular target present, or known inhibitors of viral growth as positive controls, hygromycin B15 or EK1 (*37*, *38*), a pancoronavirus spike protein peptide-based fusion inhibitor. Cells were inoculated at a MOI of 0.1 for 2 h and grown with their respective treatments. Subsequent viral RNA levels were measured 72 h after infection. We observed dose-dependent reduction of viral RNA copies with increasing concentrations of the TRSPMO, but observed no such inhibition with the same concentration of the IVS2-654-PMO (Fig. 1.6c). These results demonstrate specific inhibition of the SARS-CoV-2 virus with the synthetic PMO produced by our automated flow technology.

1.3. Discussion

Automated fast-flow synthesis is a potentially valuable tool that capitalizes on the recent successes of PMO antisense treatments (*24–26*) to expand the potential of PMOs to treat new diseases. With rapid flow synthesis, the production of novel PMO therapeutics will not be burdened by long lead optimization cycles. The fully automated fast-flow synthesizer developed here enables rapid preheating of reagents, and efficient heat and mass transfer within the resin bed on a scalable, mechanically robust platform. With these features, we adapted PMO synthesis to 90 °C and ultimately succeeded in decreasing coupling time from 3 h per nucleotide to 8 min.

To demonstrate the power of this flow platform, we synthesized three candidates for a new DMD treatment in a single day, and we anticipate that the development of new

PMO drugs using this platform will be similarly accelerated. Given the urgent circumstances around the global COVID-19 pandemic, such timeline reductions are needed, and we demonstrated the utility of fast PMO synthesis by producing a potential antisense antiviral treatment in only 3.5 h. The PMO therapeutic candidate described here demonstrates specific inhibition of the native SARSCoV-2 virus. Given the favorable toxicological profile of PMO drugs (*8*), further improvements in efficacy could result in an effective treatment for SARS-CoV-2.

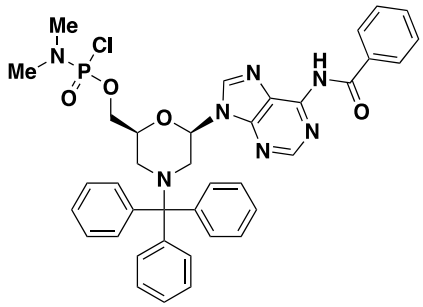
Overall, our results illustrate that machine control of flow chemistry can improve synthetic outcomes beyond what is possible with manual techniques. The strategy in this work is applicable to diverse polymer backbones, and we envision high temperature automated flow synthesis will enable the development of new on-demand biopolymers that may currently be inaccessible due to tedious, difficult, or impractical syntheses.

1.4.Materials & Methods

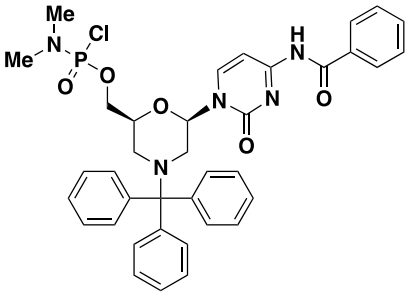
1.4.1. General Information

1.4.1.1. Materials

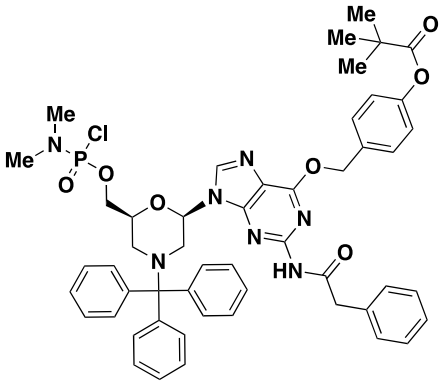
Activated morpholino subunits, moA, moC, moG, and moT (Figure 1.7), functionalized aminomethyl polystyrene resin (Figure 1.8), and functionalized polyethylene glycol linker ("Tail", Figure 1.9) were all provided by Sarepta Therapeutics. 1,3-dimethyl-2-imidazolidinone (DMI), N-methyl-2-pyrrolidone (NMP), dichloromethane (DCM), and N,N-diisopropylethylamine (DIEA) were obtained anhydrous from Sigma-Aldrich (St. Louis, MO). HPLC-grade Acetonitrile was purchased from VWR International (Philadelphia, PA) and LC-MS grade acetonitrile was purchased from Sigma-Aldrich (St. Louis, MO). Econo-Pac chromatography columns and accessories for solid-phase extraction were purchased from Bio-Rad (Hercules, CA). Water for HPLC was purified to 18.2MΩ·cm resistivity on a Millipore Milli-Q system. All other reagents and solvents were purchased from Sigma-Aldrich (St. Louis, MO) as the purest anhydrous grades available, and used without further purification. Unless specified otherwise, all solvents used were kept over activated 3 Å molecular sieves.



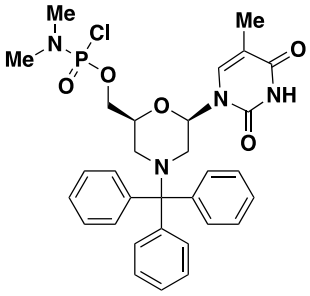
Activated, protected adenine monomer
"moA"



Activated, protected cytosine monomer
"moC"

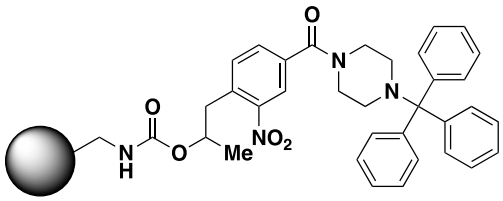


Activated, protected guanine monomer
"moG"



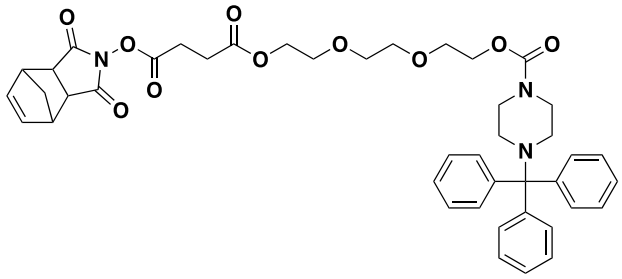
Activated, protected thymine monomer
"moT"

Figure 1.7 Structures of morpholino subunits



Functionalized, protected polystyrene resin
0.39-0.43 mmol/g

Figure 1.8 Structure of functionalized polystyrene resin



Activated, protected polyethylene glycol linker
"Tail"

Figure 1.9 Structure of the activated polyethylene glycol tail

1.4.1.2. Analytical Techniques

LC-MS:

Condition 1*:* Analysis was performed on an Agilent 1290 Infinity HPLC coupled to an Agilent 6520 ESI-Q-TOF mass spectrometer. MS was run in positive ionization mode, extended dynamic range (2GHz), and standard mass range (m/z in range 300 to 3000). 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 condition was used for PMO analysis. Column: Zorbax 300-SB C3 (5 µm, 150 x 2.1 mm, 300Å 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. Flow rate is 0.8 mL/min. MS data was acquired from 4 to 11 minutes.

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 (2GHz), and low mass range (m/z in range 100 to 1700). The solvent mixtures used were as above. Column: Phenomenex Luna C18 (2) (3 µm, 150 x 1 mm, 100 Å silica); Flow Rate: 0.05 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 was acquired from 4 to 14 minutes.

Condition 3*:* Agilent 6550 (1290 Infinity HPLC system with Dual Jet Stream ESI ionization followed by iFunnel Q-TOF MS. MS is run in positive ionization mode, extended dynamic range (2GHz), and standard mass range (m/z in range 300 to 3000). The solvent

mixtures used were as above Column: Phenomenex Kinetex PS C18 (100 Å, 2.6 µm, 2.1 x 100 mm); Flow Rate: 0.4 mL/min; Gradient: 0% hold for 3 minutes, 0-60% B over 67 minutes, MS on from 2-62 minutes.

Condition 4*:* Agilent 6550 (1290 Infinity HPLC system with Dual Jet Stream ESI ionization followed by iFunnel Q-TOF MS. MS is run in positive ionization mode, extended dynamic range (2 GHz), and standard mass range, m/z in range 300 to 3000). The solvent mixtures used were as above Column: Phenomenex Kinetex PS C18 (100 Å, 2.6 µm, 2.1 x 100 mm); Flow Rate: 0.4 mL/min; Gradient: 0% hold for 3 minutes, 0-60% B over 17 minutes, MS on from 2-15 minutes.

MALDI:

All MALDI spectra were collected on a Bruker Microflex II in linear mode, with positive ionization in low mass range (3000-8000 m/z range) and laser power under 50%. 1 µL of each sample was deposited onto a ground-steel MALDI target plate, mixed with 1 µL saturated sinapinic acid in 50% aqueous acetonitrile, and allowed to air dry.

HPLC:

Analytical HPLC was carried out on an Agilent 1200 series system with UV detection at 260 nm. Column: Zorbax CN SB, (150 x 4.6 mm, 3.6 µm, 100 Å silica); flow rate 0.8 mL/minute; Solvent System: A = water with 10 mM triethylamine acetate buffer pH 5.65, B = acetonitrile; Gradient: 3 minute hold 5% B, 5-65% B gradient over 60 minutes, 3 minute hold 65% B, 10 minute post run 5% B.

1.4.2. PMO Synthesis Reagents

The following reagent mixtures were used for PMO synthesis.

Batch: The following is a list of the stock solutions that were used for optimized manual PMO synthesis

Detritylation Mixture: 100 mM 4-cyanopyridine trifluoroacetate in 4:1 (v/v) DCM to trifluoroethanol (TFE) + 1% (v/v) ethanol. The solution was made by first mixing the DCM and TFE, followed by sequential addition of 4-cyanopyridine, trifluoroacetic acid (TFA), and ethanol. Solution was prepared fresh and not stored for more than 2 days.

Neutralization Mixture: 5% DIEA in 3:1 DCM/isopropanol (*i*PrOH) (v/v).

Coupling Solution: 0.2 M morpholino subunit, 0.4 M DIEA, and 0.2 M LiBr in DMI. The solution was made by first dissolving the subunit and LiBr in DMI. Base was added immediately prior to coupling.

Flow: The following is a list of the stock solutions that were used for optimized flow PMO synthesis. During automated flow synthesis, reaction mixtures were diluted two-fold prior to reaching the resin bed so the stocks were more concentrated than the corresponding batch synthesis solutions.

Detritylation Mixture for 70 ˚C synthesis: 400 mM 4-cyanopyridine trifluoroacetate in 3:2 (v/v) DCM/TFE + 2% (v/v) ethanol. The solution was made by first mixing the DCM and TFE, followed by sequential addition of 4-cyanopyridine, TFA, and ethanol. Solution was prepared fresh and stored under dry nitrogen pressure and used the same day.

Detritylation Mixture for 90 ˚C synthesis: 800 mM 3,5-lutidine trifluoroacetate in 3:2 (v/v) DCM/TFE + 2% (v/v) ethanol. The solution was made by first mixing the DCM and TFE, followed by sequential addition of 3,5-lutidine, TFA, and ethanol. Solution was prepared fresh and stored under dry nitrogen pressure and used the same day.

Neutralization Stock: 10% NEM in 1:1 DCM/*i*PrOH (v/v).

Monomer Stock: 0.4 M morpholino subunit in DMI.

Coupling Base: 0.8 M DIEA, and 0.4 M LiBr in DMI.

1.4.3. Manual PMO Synthesis Techniques

A procedure for the batch synthesis of PMOs was adapted from patent and academic reports (Figure 1.10) (*8*, *17–20*). Each step is described below (Table 1).

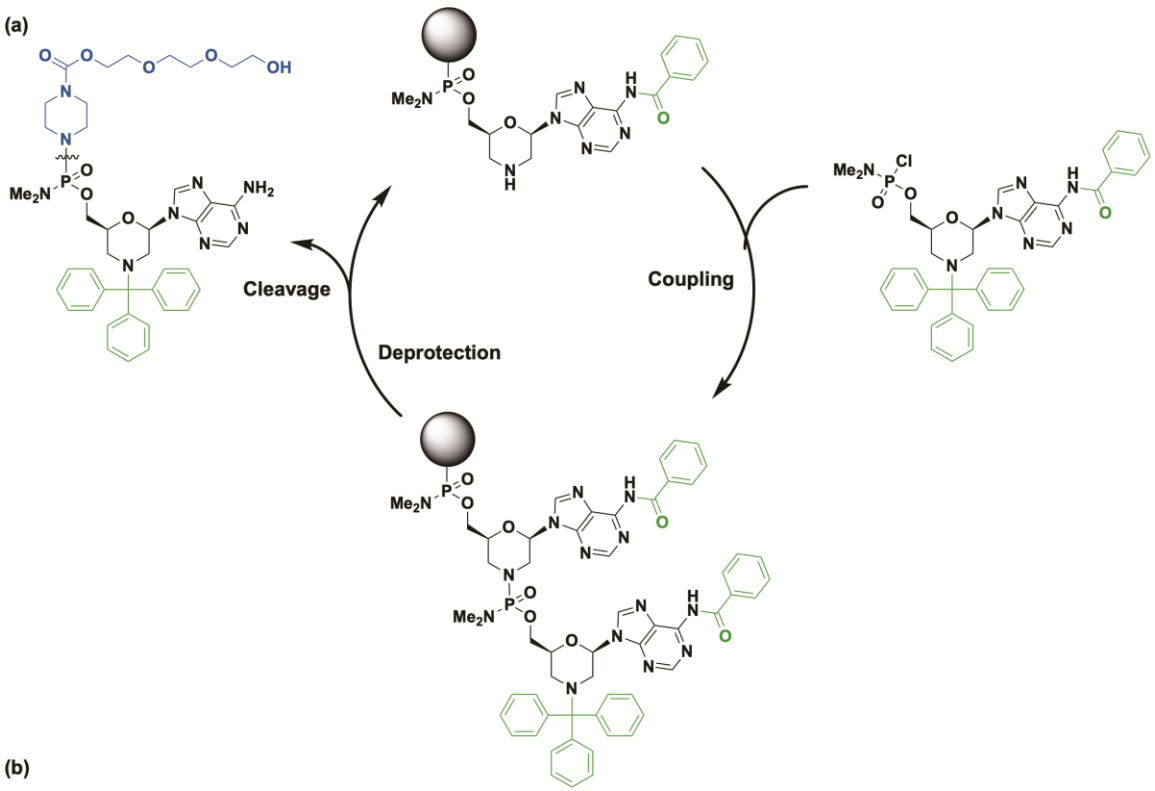


Figure 1.10 Manual PMO Synthesis Procedure Overview.

a) The chemical transformations that take place during PMO synthesis are shown. Green structures represent protecting groups that are lost upon cleavage. Blue structures represent the PEG$_3$ tail appended at the 5' end of each PMO sequence. Each nucleotide is incorporated as a trityl protected phosphoramido chloridate. Detritylation and neutralization regenerate a reactive 3' amine for the next coupling reaction

Aminomethyl polystyrene resin used in all studies had a loading of 0.39- 0.43 mmol/g and was functionalized with the "Tail" PEG linker bearing a piperazine terminus, according to literature procedres[1]. 100 mg of tail-loaded aminomethyl polystyrene resin (0.39-0.43 mmol/g loading) (39-43 µmol) was placed in a 20-mL disposable reaction vessel (Torviq) and swollen in NMP for at least 1 hour. Resin was washed with 30% TFE in DCM (2 x 7 mL) and placed in 7 mL of **Detritylating solution** for 15 minutes. This process was repeated 2 times. Deprotected resin was washed with DCM (2 x 7 mL) and 30% TFE in DCM (2 x 7 mL), and then placed in 7 mL of **Neutralization solution** for 5 minutes. This process was repeated 2 times. Neutralized resin was washed with DCM (2 x 7 mL), 30% TFE in DCM (2 x 7 mL), and DMI (2 x 7 mL). 10 eq. of monomer relative to the highest resin loading was added as **Coupling solution** and reacted for 90 minutes. Resin was then washed with DCM (3 x 7 mL) and then placed in 7 mL of **Neutralization**

solution for 5 minutes. This process was repeated 2 times. Finally, resin was washed with DCM (3 x 7 mL) and 30% TFE in DCM (3 x 7 mL).

Entry	Step	Volume	Time	Mixture
1	Wash	14 mL	15 sec	30% TFE:DCM
2	Detritylation	7 mL	15 min	100 mM 4-cyanopyridine 100 mM TFA in 4:1 DCM:TFE, 1% EtOH
3	Detritylation	7 mL	15 min	100 mM 4-cyanopyridine 100 mM TFA in 4:1 DCM:TFE, 1% EtOH
4	Wash	14 mL	15 sec	DCM
5	Wash	14 mL	15 sec	30% TFE:DCM
6	Neutralization	7 mL	5 min	5% DIEA in 3:1 DCM:*i*PrOH
7	Neutralization	7 mL	5 min	5% DIEA in 3:1 DCM:*i*PrOH
8	Wash	14 mL	15 sec	DCM
9	Wash	14 mL	15 sec	30% TFE:DCM
10	Wash	14 mL	15 sec	DMI
11	Coupling	2.6 mL	90 min	0.18 M monomer, 0.4 M DIEA, 0.2 M LiBr in DMI
12	Wash	21 mL	30 sec	DCM
13	Neutralization	7 mL	5 min	5% DIEA in 3:1 DCM:*i*PrOH
14	Neutralization	7 mL	5 min	5% DIEA in 3:1 DCM:*i*PrOH
15	Wash	21 mL	30 sec	DCM
16	Wash	21 mL	30 sec	30% TFE:DCM
Total	--	-	143 min	--

Table 1 Manual PMO Synthesis Protocol.

1.4.4. Cleavage Techniques

Completed PMO resins were washed with DCM and dried under vacuum prior to cleavage. Crude resin was placed in a 7 mL glass vial with a tight-fitting Teflon-lined screw cap and cleaved using one of the following methods:

Method 1: A 1:1 mixture of ammonia (sat. aq.) and methylamine (sat. aq.) was freshly prepared and 1 mL per 10 mg of crude resin was added. The vial was capped then kept at 65 °C for 15 minutes.

Method 2: A 1:1 mixture of ammonia (sat. aq.) and ethanol was freshly prepared and 1 mL per 10 mg of crude resin was added. The mixture vial was capped then kept at 65 °C for 36 hours.

The entire crude cleavage mixture was filtered through a 0.22 µm PTFE syringe filter and washed with 3 x 1 mL methanol. The filtrate was evaporated to dryness using nitrogen flow.

The residue was dissolved in 10 mL of Milli-Q water prior to SPE. Separately, SPE columns were prepared and conditioned following a previously reported procedure[1]. 20 mL Econo-Pac columns from Bio-Rad were charged with 3-4 mL of Amberchrome CG-300M resin and sealed with a frit. Then 8 mL of the following solutions were added to the column, in order, and drained before adding the next solution: 80% ACN in 1% NH_4OH (sat. aq.), 0.5 M NaOH in 20% EtOH, water, 50 mM H_3PO_4 in 80% ACN, water, 0.5 M NaOH in 20% EtOH, water, 1% NH_4OH (sat. aq.). When conditioning was complete, the column was stored in 8 mL of 1% NH_4OH (sat. aq.) at room temperature until used. The column was rinsed two times with 12 mL of water before loading the crude mixture onto the column. Then, the column was rinsed once with 3 mL of 1 M NaCl, followed by three rinses with 12 mL of water, and once with 3 mL of 10% acetonitrile in water. The PMO was then eluted with two 3 mL rinses of 50% acetonitrile in water. The eluent from the 50% acetonitrile wash was collected into a pre-weighed 15 mL conical centrifuge tube and lyophilized to afford the crude PMO as a white powder suitable for LC-MS analysis and purification.

1.4.5. Purification Techniques

Cation Exchange: Purification was carried out on an ÄKTA Pure using a 5 mL HiTrap SP HP column with UV detection at 260 nm; Flow rate: 5.0 mL/minute; Solvent system: C = 20% acetonitrile in water with 10 mM phosphoric acid, D = solvent C with 1 M potassium chloride; Gradient: 5 minute hold at 0% D, 0-10% D over 10 minutes, 10-40% D linear gradient over 120 minutes, 0% D for 10 minutes. Fractions collected from 10 to 120 minutes with 1-minute time slices.

Reverse Phase: Semi-preperative HPLC purification was carried out on an Agilent 1200 series system with UV detection at 260 nm. Column: Zorbax CN SB (150 x 4.6 mm, 3.6 µm, 100 Å silica); Flow rate: 0.8 mL/minute; Solvent system: E = water with 10 mM triethylamine acetate buffer, pH 5.65, F = acetonitrile; Gradient: 3 minute hold 5% F, 5-15% F linear gradient over 10 minutes, 15-35% F linear gradient over 100 minutes, 35-65% F linear gradient over 3 minutes, 3 minute hold 65% H, 10 minute post run 5% F; Flow rate: 0.8 mL/minute. Fractions collected from 15 to 115 minutes with 1-minute time slices.

Anion Exchange: Purification was carried out on an ÄKTA pure using a 5 mL HiTrap Q HP column with UV detection at 260 nm; Flow rate 5.0 mL/minute; Solvent system: G = water with 10 mM sodium hydroxide, H = water with 10 mM sodium hydroxide and 1 M sodium chloride; Gradient: 5 minute hold at 0% H, 0-10% H linear gradient over 10 minutes, 10-40% H linear gradient over 120 minutes, 0% H for 10 minutes. Fractions collected from 10 to 120 minutes with 1-minute time slices.

1.4.6.Quantification of flow synthesis results with LC-MS

Side products from PMO synthesis often co-elute on RP-HPLC, so quantification of sample purity was measured using total extracted ion counts. Extracted ion counts were tabulated using Agilent's Molecular Feature Extraction utility. For each sample, all tabulated ions that arose from the intended PMO sequence were summed. Small amounts of side products that didn't arise from the synthesis process, but instead arose from the cleavage process, or the purification method itself were included in the summations as pure product. For each species quantified, the abundances of the following identified compounds were summed: [M+14], [M+Bz protecting group], [M+Bz protecting group+14], and [M+G base elimination]. [M+14] is caused by methylation of

cytosine residues from cleavage in methylamine. [M+Bz protecting group] is the result of incomplete removal of benzoyl protecting groups on A and C residues. [M+G base elimination] is a result of loss of the G nucleobase that occurs under the acidic conditions in the LC-MS separation. Not included in the analyses are small molecules that are arise from the cleavage process. To ensure that these molecules were excluded from the calculation, the extraction utility ignored compounds with molecular weight less than 500 Da, and avoided peaks of the LC-MS separation that resulted from known cleavage molecules.

The following parameters were used in the Molecular Feature Extraction utility: Find by Molecular Feature; Restrict m/z to 500-2000; Restrict retention time to 6.5-8.0 minutes (this time range covers desired product and structure related side-products); No mass filter; No mass defect; Use peaks with height >= 100; Standard ion species; Limit max number of peaks to 100; Isotope spacing tolerance 0.0025 m/z plus 7.0 ppm; Isotope model: Common organic molecules; Limit assigned charge stated to a maximum of 3.

1.4.7. Flow Synthesizer Design

The automated synthesizer Tiny Tides is shown in Figure 1.11.

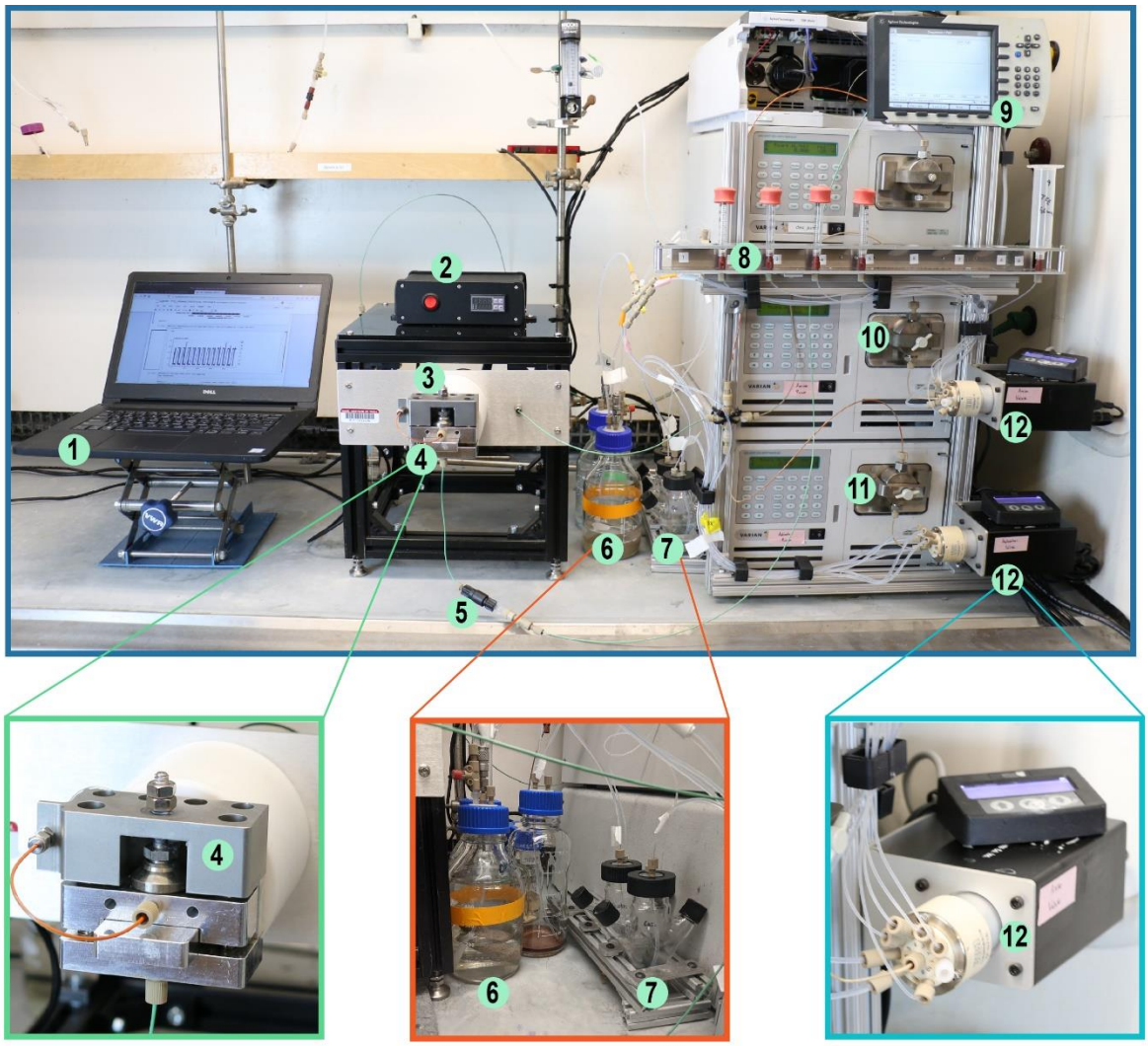


Figure 1.11 Tiny Tides is an automated instrument for the synthesis of PMOs.

Major parts are indicated with numerals, and a brief description is below. Only two HPLC pumps are labeled – parts 10 and 11 – the third was not used in this work.

(1) Computer – The instrument is controlled with an in-house Python script[10]. The programming environment is available on an open-source GitHub repository.

(2) Temperature Controller – A Misumi temperature controller (Part #: MTMNRD) was used to drive power for an OMEGA 120V 300W heater cartridge (Part#: HDC00059).

(3) Heating Coil – 3 feet of 316L stainless steel tubing (0.064” OD x 0.020” ID) wrapped around a heated aluminum core.

(4) Reactor – Custom aluminum and stainless steel reactor head. A screw-handle on top holds a cartridge containing resin in the solvent flow-path. See **Supplementary Section 10.2** for a detailed description.

(5) Back pressure regulator – 100 PSI IDEX back pressure regulator (Part #: P-607).

(6) Solvent bottles – DMI and DCM were stored over activated 3 Å molecular sieves in two 500 mL Schott brand media bottles. The bottles were fitted with machined polypropylene caps to supply nitrogen and maintain positive nitrogen head pressure. See **Supplementary Section 10.3** for more details.

(7) Reagent reservoirs – Deprotection, neutralization, and coupling base solutions were stored in 3-necked 100 mL stirrer flasks from Chem Glass (Product #: CLS-1401-100). Top caps were custom machined from polypropylene to enable maintenance of positive pressure of dry nitrogen. See **Supplementary Section 10.3** for more details.

(8) Monomer reservoirs – Monomer solutions were stored in 5 mL polypropylene syringes under positive pressure of dry nitrogen supplied through custom polypropylene caps.

(9) UV Detector – Agilent 1260 UV detector (Part #: 1260DAD) fitted with a micro flow cell 6 mm path length (Part #:01200-90130 Rev. B). Effluent was monitored at 280 nm.

(10) Monomer HPLC Pump – Agilent 210 series HPLC pump fitted with a 5 mL SS pump head. One stroke delivers 40 µL of solution.

(11) Activator HPLC Pump – Agilent 210 series HPLC pump fitted with a 5 mL SS pump head. One stroke delivers 40 µL of solution.

(12) 10-Position valves – VICI brand EUHA body fitted with a 10-position Chem-Inert model 16P-0713L valve face.

Tubing: Tubing from the pump to the mixer T is PEEK 1/16” OD x 0.02” ID tubing (McMaster Part#: 51085K45). Tubing from the Reagents to the valves is PFA Nat 1/16” OD x 0.030” ID tubing (IDEX Part #: 1514L). All other tubing is PEEK 1/16” OD x 0.030” ID tubing (McMaster Part#: 51085K48).

Fittings: Unless otherwise specified, all fittings were IDEX brand Super Flangeless™ PEEK fittings.

The reactor chamber that holds the solid support is an important factor in flow instruments with solid supported reactants (*39*). Geometry, temperature flux, and reactor dead volume are significant factors that can affect reaction efficiency and economy (*40*). In this work, we designed a reactor that scaled down the optimized geometry previously reported for flow peptide synthesizers that utilize a similar solid support for solid phase peptide synthesis (*41*). The reactor used in our instrument, dubbed Tiny Tides, maintains the reported geometric ratios, but is scaled down by a factor of 10 to accommodate a 4.5 micromole synthesis (~10 mg resin in ~600 µL). This smaller reactor saved expensive reagents, while still producing enough PMO for initial biological studies.

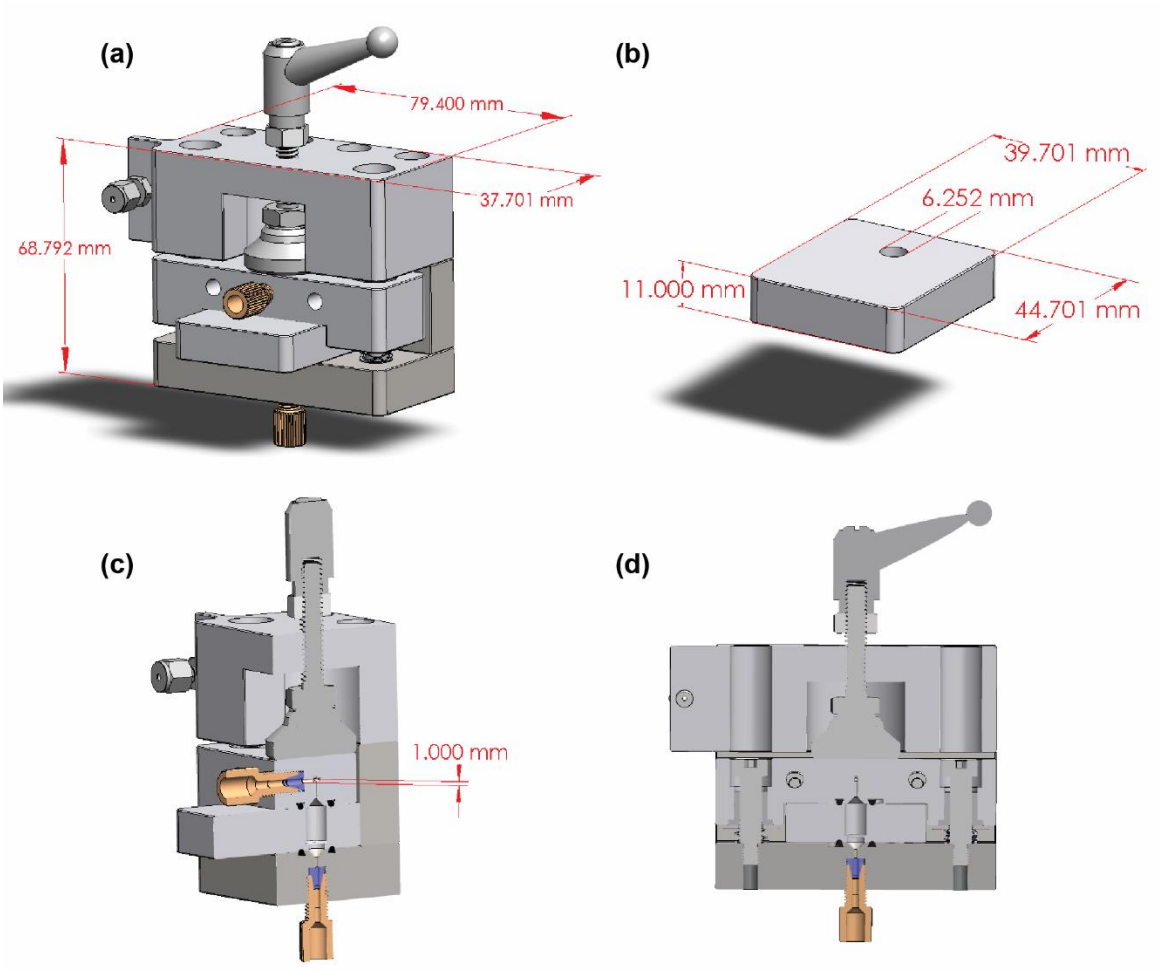


Figure 1.12. Computer Aided Design (CAD) Projection of the optimized Reactor Design for PMO synthesis

a) The reactor design is shown along with the cassette used to hold the resin. b) The inner cartridge that contains the resin is shown. Resin is retained in the thru-hole. c) Cross-section view of the reactor is shown, revealing the inner resin chamber. Flow enters

through the PEEK fitting in the centre, passes through the cartridge hole, and exits through the fitting in the bottom. The cartridge is held in place with the screw clamp from above. d) A cross-section view of the reactor from the front face is shown.

To ensure rapid turnover between syntheses and minimize maintenance down-time, a quick-release system was implemented to manage resin holders. Most importantly, the reactor must be stable to operation at high enough pressures to allow for the use of dichloromethane at 90 °C where its vapor pressure is about 5 bar. A cartridge-based system held resin in a reusable metal container that could be placed in the flow path and clamped down to withstand expected pressures. All pieces were initially machined from aluminum to ensure efficient heat transfer. A final revision of the reactor body changed the top half of the reactor to stainless steel to allow for larger clamping pressures (Figure 1.12).

The monomer solutions are sensitive to moisture, so we kept them under anhydrous nitrogen pressure. The slight positive pressure also helped eliminate cavitation in the pump heads. The plumbing system for dry nitrogen is shown in Figure 1.13. The containers used for monomer storage include an exit path to allow for continuous nitrogen flow, and flow to the monomers was controlled with an in-line air flow controller. Continuous nitrogen flow helps remove trace water that infiltrates the system of remains in the DMI used for monomer stock solution. This precaution is not required for the other reagents because they are less sensitive to moisture. The excess nitrogen flow from the intentionally leaky reagent containers would consume a large majority of the nitrogen flow if not regulated with the upstream flow limiter.

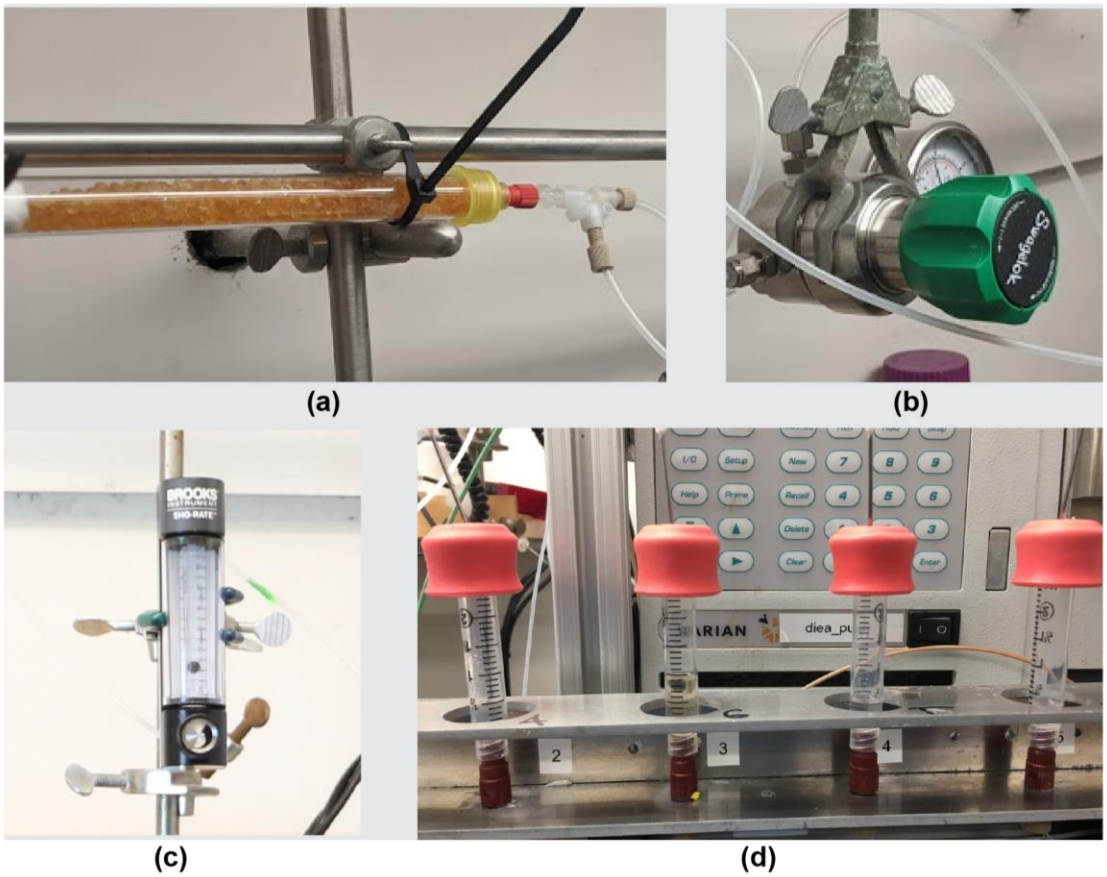


Figure 1.13 Nitrogen plumbing on the Tiny Tides allows for anhydrous storage of activated monomer solutions

a) Nitrogen from the hood manifold is plumbed through an activated silica gel desiccant column. b) An in-line pressure regulator allows for precise back pressure control over the entire system. c) A nitrogen flow limiter is placed before the monomer solutions but not the reagent lines to prevent excess nitrogen leakage from the monomer containers. d) Monomer solutions are stored in disposable syringes with nitrogen pressure plumbed in through the top.

Other reagents were held in glass containers under 5 PSI of nitrogen without continuous nitrogen flow. Nitrogen was supplied to GL-45 threaded bottles through a custom-machined polypropylene adapter for use with GL-45 caps with holes (Chemglass). These adaptors had three ¼-28 threaded ports. Two ports had 1/8" thru holes; the third had a 1/16" thru hole. The first hole was used for tubing to deliver reagent to the valves. The second contained a Swagelok® miniature Quick-Connect fitting for the anhydrous transfer of solvents, and the third was used for the nitrogen gas supply. The line used for reagent transfer was inserted into the bottom of the reagent reservoir and sealed in place with a Super Flangeless™ fitting (Idex XP-131). Nitrogen gas was

supplied by a 1/8" line seated against the 1/16" thru hole and sealed in with a Super Flangeless™ fitting. The fill port consisted of a thin wall 1/8" OD, 0.1" ID, stainless steel tube inserted to just below the bottom of the machined adaptor and sealed in place with a Super Flangeless™ fitting. The free side was fitted with a shut off Quick-Connect fitting (Swagelok SS-QM2-B-200KR) that was sealed when not in use but could be used to fill the reservoir from a second reservoir of anhydrous solvent under slightly higher pressure. Effective mixing of the streams from the two pumps is critical for effective PMO synthesis. Although there exists extensive research into more effective mixing T's (*42*), we found that use of a simple T union (Idex part #P-722) was adequate. Pump streams entered opposite each other at high flow leading to sufficient mixing for PMO synthesis (Figure 1.14).

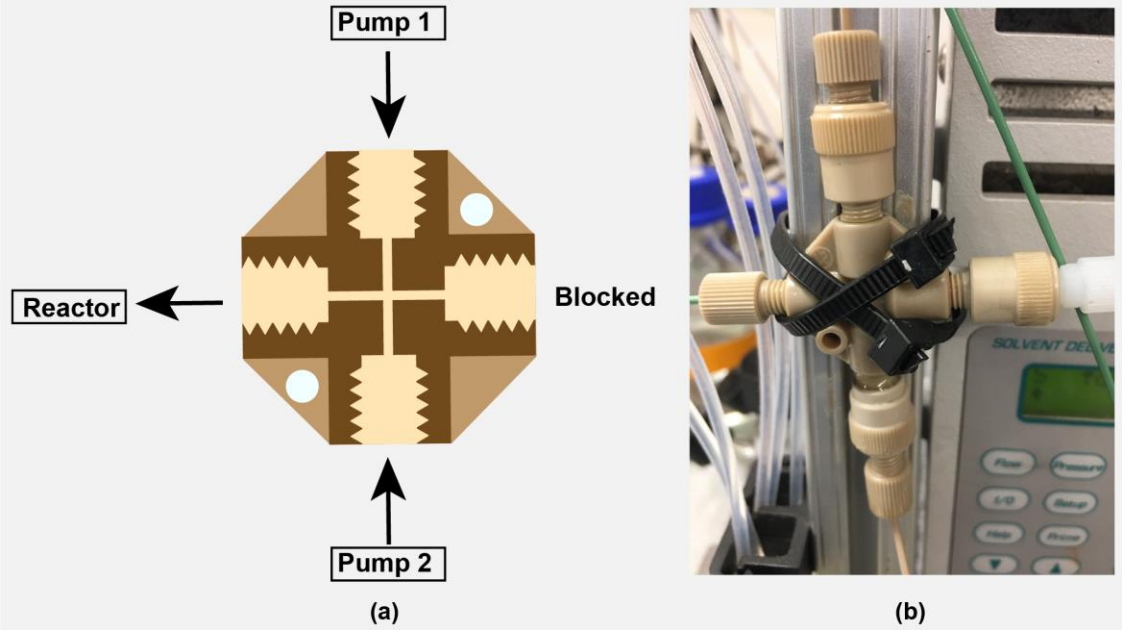


Figure 1.14 Pump stream mixing on Tiny Tides with a T mixer design.

a) The inner schematic for the T mixer used is shown. Flow from the two pumps pass through check valves and meet opposite each-other and exit through another channel. The remaining channel is not used and is blocked during synthesis. b) The streams from the two pumps meet from opposite directions in a PEEK cross-mixer. The arm shown on the right is not used for PMO synthesis.

An in-line UV-Vis monitor installed after the reactor was used to monitor the delivery of each reagent. Although the bright yellow trityl cation released during deprotection is an attractive way to monitor reaction progress, chemicals are included in

the optimized detritylation reagent rapidly quench the trityl cation and prevent it from reacting with the resin bound PMO. The quenched species have overlapping UV absorbance with the deprotection regent, restricting use of the UV detector to monitoring mechanical function of the instrument in routine operation. To rapidly optimize reaction parameters, deprotection mixtures without trityl cation quenchers were used to track the efficiency of PMO synthesis in real time. A modified synthesis protocol was developed to use 3% trichloroacetic acid in DCM (w/v) for deprotections. Although the resulting PMO showed significant impurities from acidolysis, monitoring trityl removal after coupling enabled rapid optimization of the coupling step in real time without cleavage and LCMS analysis.

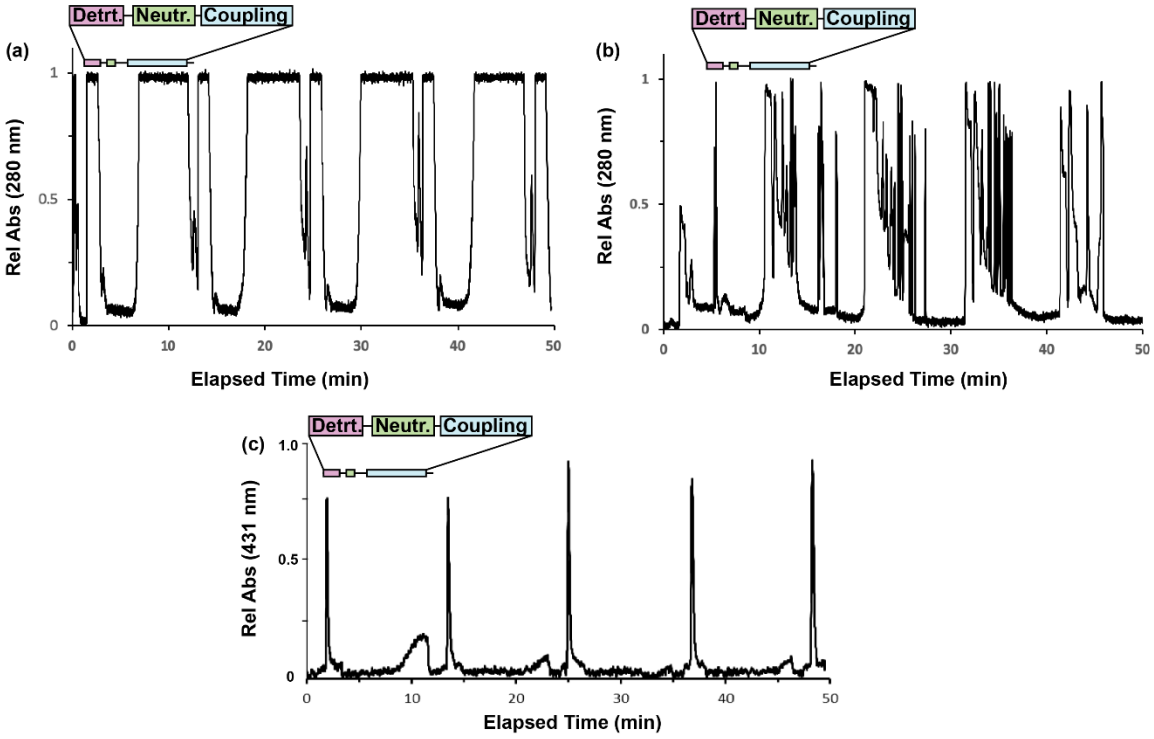


Figure 1.15 . In-line UV-Vis was used to monitor Tiny Tides instrument performance.

a) Reagent delivery during an efficient synthesis is consistent, as monitored by UV-Vis absorbance at 280 nm. b) System leaks cause a distinct, noisy UV-Vis trace. c) Use of a spectroscopically-silent deprotection mixture allows for tracking real-time reaction efficiency via monitoring trityl cation release at 431 nm.

In normal operation, however, we used the in-line UV-Vis detector to track real time instrument performance and ensure proper pump coordination, valve timings, software triggers, and other key parameters. A typical reaction trace along with one associated with improper reagent delivery is shown in Figure 1.15. The details of the traces were used to identify which part of the instrument was malfunctioning, greatly reducing repair and optimization times.

Unoptimized pump delivery cycles led to consistent under delivery of the PMO monomer solutions, with the default configuration delivering only 2/3 of the expected volume. Although there are many possible reasons for this behavior, including mechanical restrictions in the valve face, incorrect valve timings, or other software bugs, it only affected monomer solutions and some were more affected than others. This evidence suggested that the problem was caused by the composition of the monomer solutions.
We ultimately determined that cavitation in the pump head was responsible for the observed under-delivery, with under-delivery proportional to the viscosity of the solution. Monomer solutions are very concentrated (0.8 M) and as a result are very viscous. This problem is especially bad for moG, and less pronounced for moT, likely due to the much higher molecular weight of moG.

The Varian 210 HPLC pumps in the instrument are single piston reciprocating positive displacement pumps specifically selected for their very low non-displaced volume. The low non-displaced volume enables changeover of solvents and reagents in very few pump strokes (99% change in composition in about 3 strokes). The drawback of this pump design, however, is that the flow is discontinuous. There is a discrete delivery phase where the piston compresses fluid within the pump head and ejects it through the outlet, and a discrete withdraw phase where the piston withdraws and sucks fluid into the pump head. During the delivery phase, no new fluid is entering the pump head and during the withdraw phase there is zero outlet flow rate (Figure 1.16). This means that the pump must balance the withdraw and delivery rates to achieve the selected total flow rate.

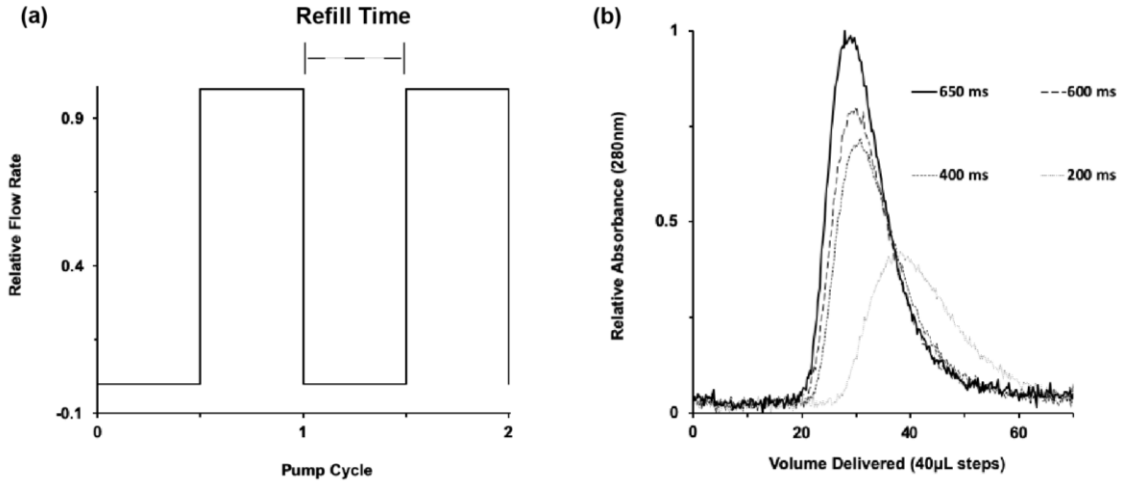


Figure 1.16 . Pump head cavitation can be alleviated with long refill times.

a) Schematic showing the flow rate as a function of time for the HPLC pumps used to deliver reagents. Increasing refill time results in higher relative flow rates with longer gaps in-between. b) Increasing the refill time results in better delivery of viscous reagent mixtures.

By default, the withdraw stroke is rapid to minimize discontinuities in the flow rate when pumping low viscosity solvents like water and acetonitrile. When connected to high viscosity solvents, however, there is no time for the viscous fluid to fill the pump head during the withdraw stroke. The piston will withdraw, creating a reasonable vacuum (~29”Hg), but the fluid cannot creep into the chamber fast enough. As a result, the delivery stroke collapses vacuum for part of its stroke, then delivers as normal. The result is a shorted delivery stroke that delivers less than the expected volume.

To confirm cavitation was the problem, the nitrogen head pressure on the monomers was increased. This fixed the cavitation and under-delivery of monomer, but was mechanically inconvenient to maintain. Instead, the pump withdraw time was increased from the default 200 ms to 650 ms to give the solution more time to fill the chamber on each stroke. To optimize the pump withdraw time, delivery of concentrated food dye in DMI was pumped, and the relative delivery of the dye solution was quantified by tracking with the in-line UV monitor (Figure 1.16). As refill time was increased, increasing amounts of monomer solution were delivered to the UV monitor. For all further studies, we chose the maximum refill time that the pumps can use when delivering 2.5 mL/min, 650 ms. It is likely that cavitation is still present in the system at this setting, but further increases in refill time would limit the maximum flow rate. We determined that the

cavitation at this setting is likely not significant, as the total consumed volume of the most viscous monomer solution, moG, after 5 consecutive couplings delivers 98% of the total theoretical volume (50 µL remaining of 1.6 mL to be delivered)

1.4.8. Optimized Synthesis Sequence

The synthesis order is characterized Table 2. For each chain elongation cycle, steps 1-8 are repeated, and an additional deprotection sequence (steps 2-6) is run at the end of synthesis to remove the final 3' trityl protecting group. Synthetic difficulty increases as the PMO sequence gets longer, so for PMO sequences between 10 and 20 nucleotides the number of pump strokes was increased from 5 to 6, and for sequences between 21 and 30 nucleosides 7 strokes of monomer solution were used to realize a highly efficient synthesis without significantly effecting the synthetic timeline.

Step #	Description	Monomer Pump	Reagent Pump	Time	Flow Rate
1	Wash	DMI	DMI	90 sec	2.5 mL/min x 2 pumps
2	Wash	DCM	DCM	30 sec	2.5 mL/min x 2 pumps
3	Deprotection	DCM	***Detritylation stock***	114 sec	2.5 mL/min x 2 pumps
4	Wash	DCM	DCM	18 sec	2.5 mL/min x 2 pumps
5	Neutralization	DCM	***Neutralization stock***	12 sec	2.5 mL/min x 2 pumps
6	Wash	DCM	DCM	24 sec	2.5 mL/min x 2 pumps
7	Wash	DMI	DMI	30 sec	2.5 mL/min x 2 pumps
8	Coupling	***Coupling base stock***	***Monomer stock***	5 sec	2 mL/min x 2 pumps
9	Wash	DMI	DMI	7 sec	2 mL/min x 2 pumps
10	Wash	DMI	DMI	420 sec	0.1 mL/min x 2 pumps

Table 2 Summary of coupling and deprotection steps performed during automated flow PMO synthesis.

1.4.9. Raw LC-MS Characterization

1.4.9.1. LCMS traces of 4-mer PMOs for condition optimization

A series of 4-mer PMOs were produced to optimize PMO synthesis in flow (Figure 1.4). Five mg of each resin bound PMO product was washed with DCM and dried under vacuum. After cleavage using Method 1, crude cleavage mixtures were analyzed with LC-MS using Condition 4. All the samples were quantified following the method described in Section 1.4.6.

Note 1: Peaks outside of 6.5 - 8.0 minutes were excluded from analysis. Peaks within 6.5 - 8.0 minutes cover the desired product and structurally related side-products while compounds outside of this range are small molecules generated during the cleavage process. These peaks corresponding to small molecules from cleavage were excluded to simplify purity interpretations.

Note 2: Purity is most accurately assessed using the method described in Supplementary Section 6. Although TICCs and mass spectra for certain conditions may contain many products, it is not necessarily indicative of poor synthesis quality, as side-products arise during the analytical characterization itself. This class of compounds includes cytosine residues methylated during the cleavage process, incomplete removal of benzoyl protecting groups, and guanosine nucleobases eliminated during the ionization process. Relative amounts of these compounds vary greatly from sample handling and preparation, and do not reflect changes in fast flow synthesis performance. Purity assessments made using the method described in Supplementary Section 6 do not count these compounds as side-products.

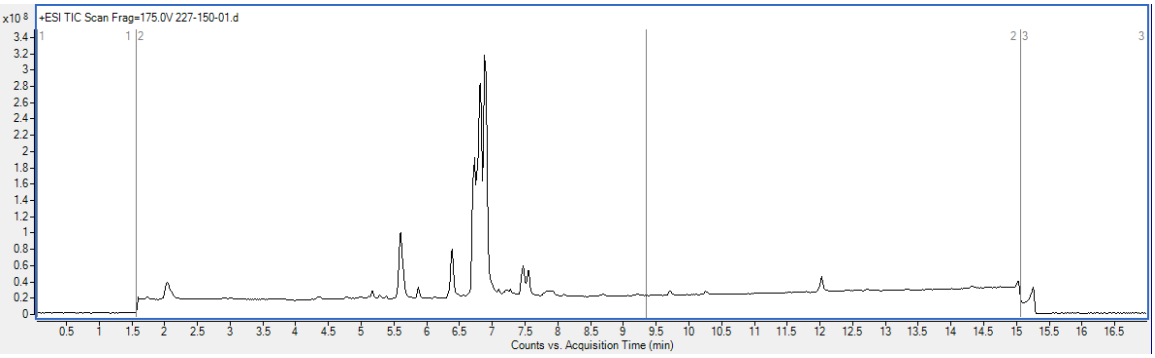
1. Fig. 1.4, Entry 1

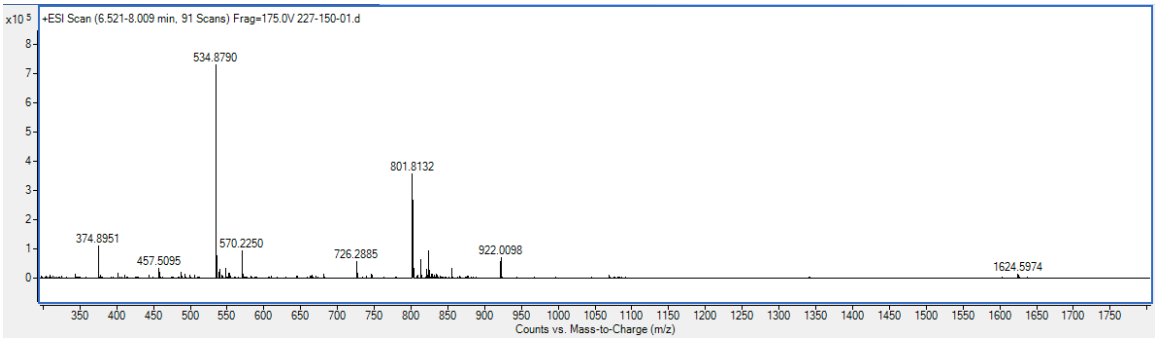
Sequence (5’ to 3’): Tail-ACGT (4-mer)

Condition: Batch synthesis (Section 1.4.3)

LCMS Method: Condition 4

Observed: 1601.63 Da Calculated: 1601.60 Da





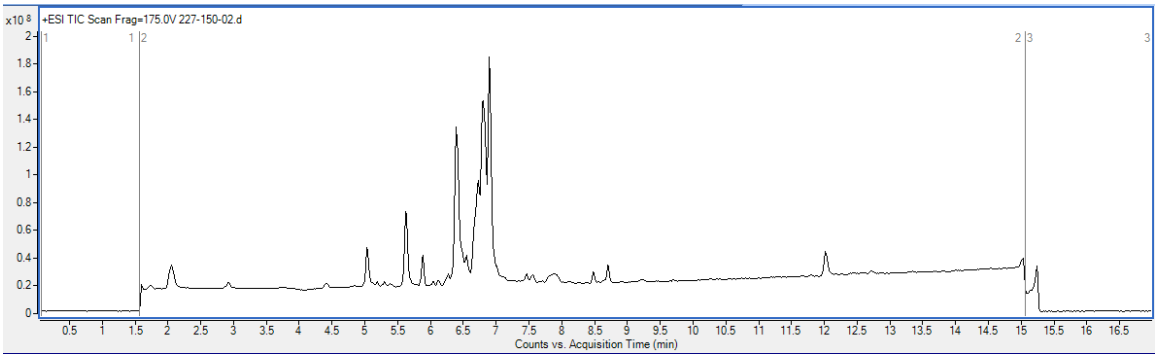
2. Fig. 1.4, Entry 2

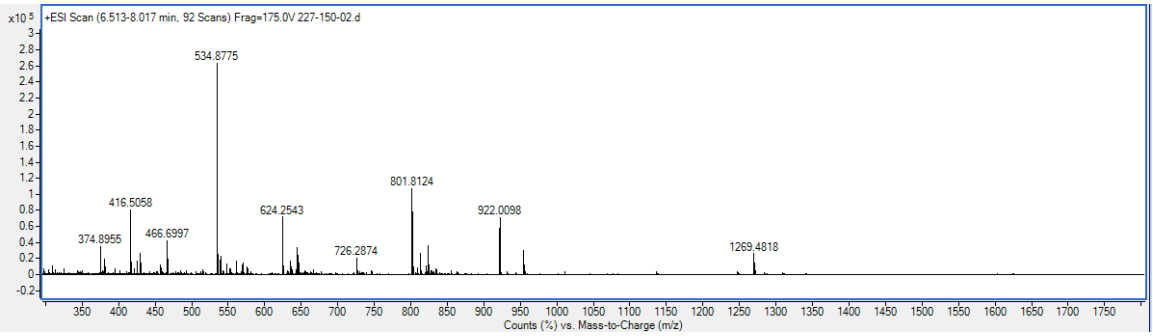
Sequence (5’ to 3’): Tail-ACGT (4-mer)

Condition: Tiny Tides synthesis, 70 °C, 4-cypy·TFA, DIEA neutr, DIEA base, no additive, 10 eq monomer, 17.4 min.

LCMS Method: Condition 4

Observed: 1601.62 Da Calculated: 1601.60 Da





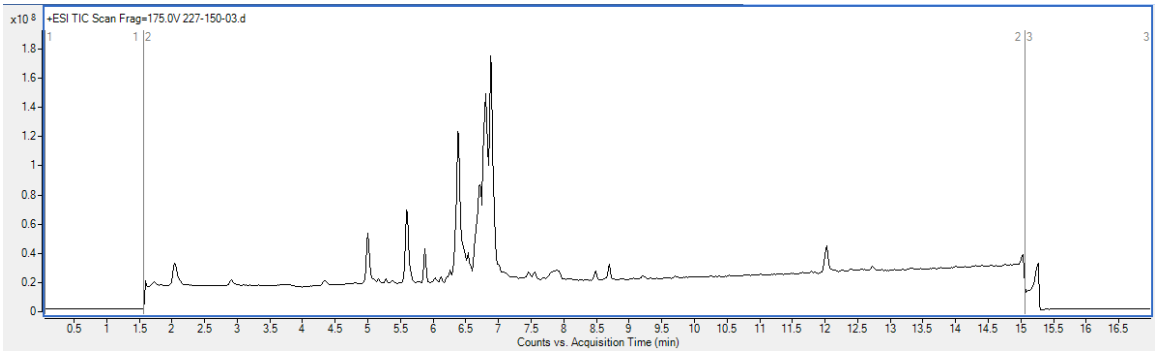
3. Fig. 1.4, Entry 3

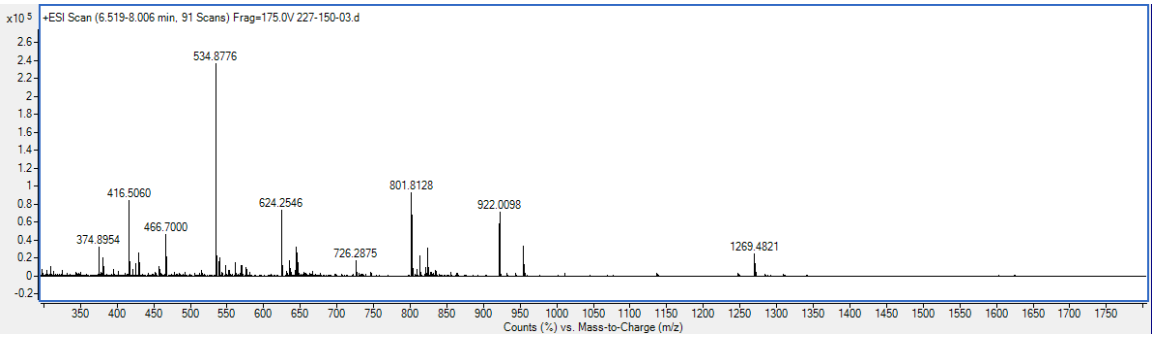
Sequence (5’ to 3’): Tail-ACGT (4-mer)

Condition: Tiny Tides synthesis, 70 °C, 4-cypy·TFA, DIEA neutr, DIEA base, no additive, 10 eq monomer, 11.7 min.

LCMS Method: Condition 4

Observed: 1601.63 Da Calculated: 1601.60 Da





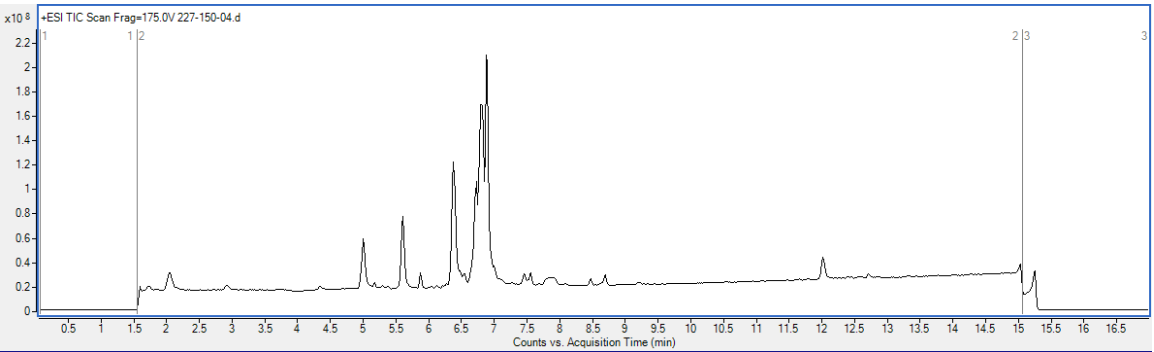
4. Fig. 1.4, Entry 4

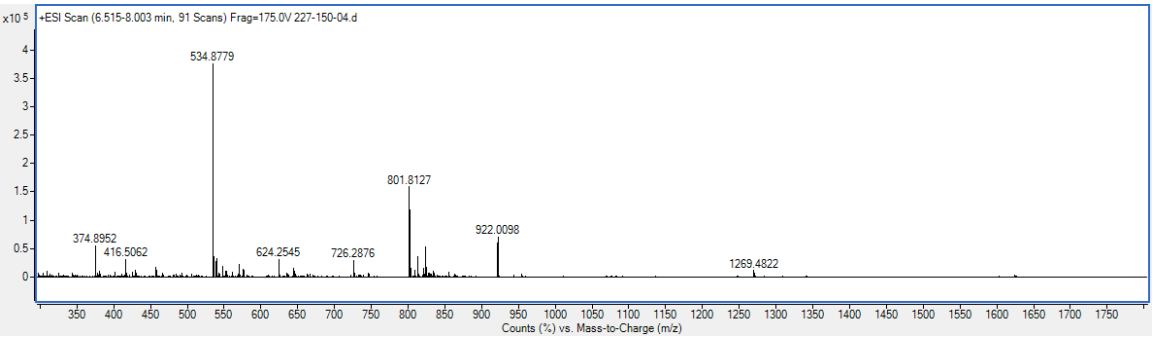
Sequence (5’ to 3’): Tail-ACGT (4-mer)

Condition: Tiny Tides synthesis, 70 °C, 4-cypy·TFA, DIEA neutr, DIEA base, no additive, 18 eq monomer, 11.7 min.

LCMS Method: Condition 4

Observed: 1601.63 Da Calculated: 1601.60 Da





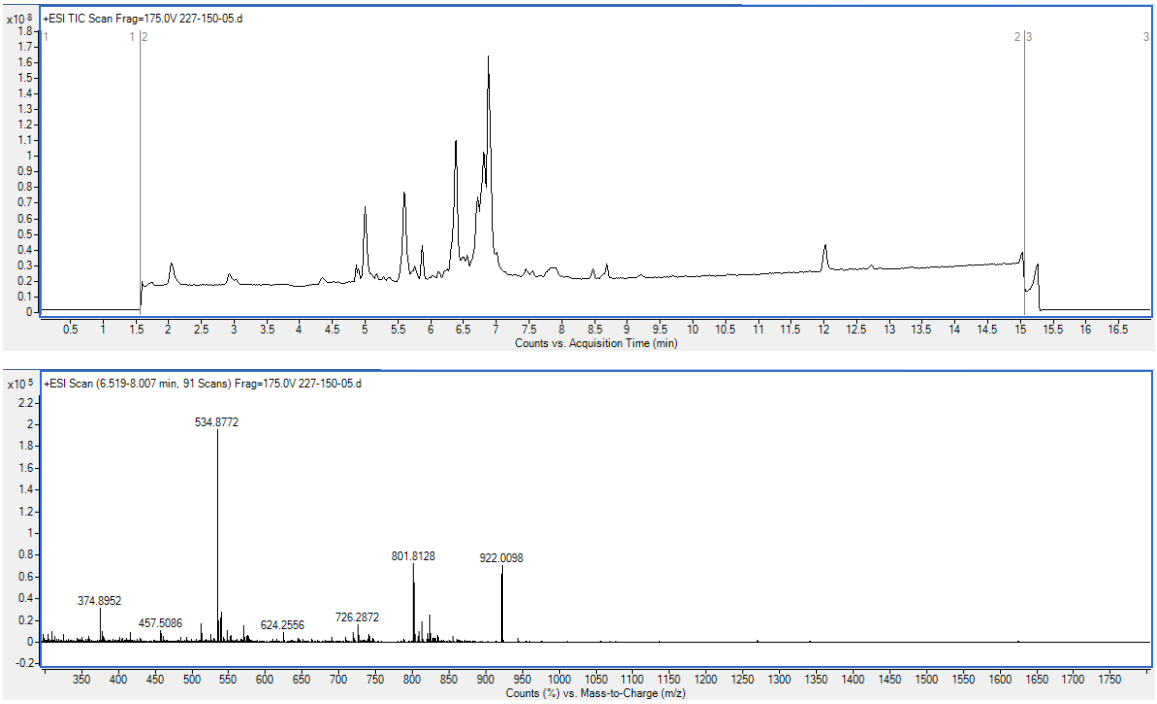
5. Fig. 1.4, Entry 5

Sequence (5’ to 3’): Tail-ACGT (4-mer)

Condition: Tiny Tides synthesis, 70 °C, 4-cypy·TFA, DIEA neutr, DIEA base, NMI 0.2 M, 18 eq monomer, 11.7 min.

LCMS Method: Condition 4

Observed: 1601.63 Da Calculated: 1601.60 Da



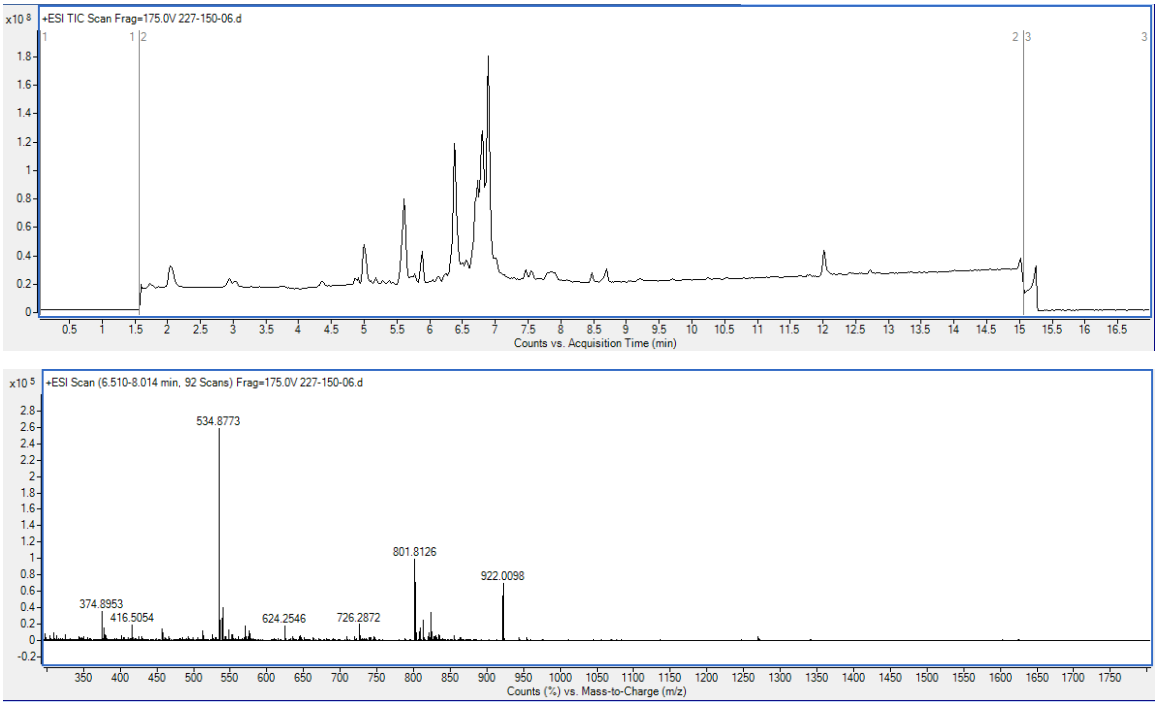
6. Fig. 1.4, Entry 6

Sequence (5’ to 3’): Tail-ACGT (4-mer)

Condition: Tiny Tides synthesis, 70 °C, 4-cypy·TFA, DIEA neutr, DIEA base, LiBr 0.2 M, 18 eq monomer, 11.7 min.

LCMS Method: Condition 4

Observed: 1601.63 Da Calculated: 1601.60 Da



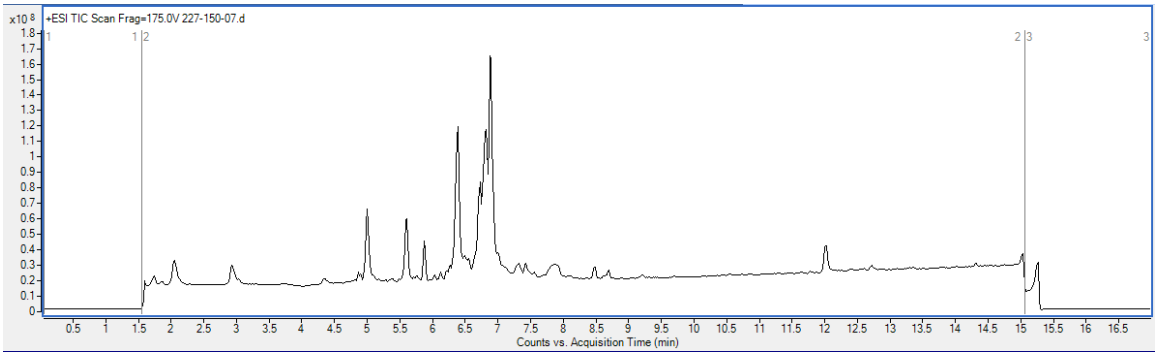
7. Fig. 1.4, Entry 7

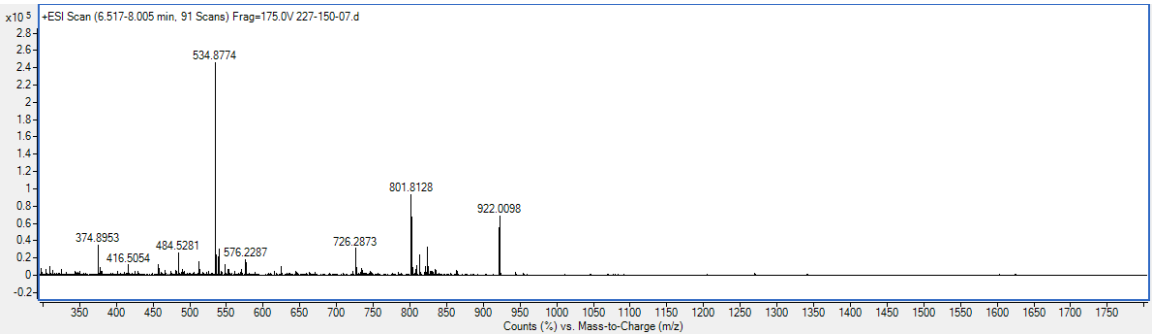
Sequence (5’ to 3’): Tail-ACGT (4-mer)

Condition: Tiny Tides synthesis, 90 °C, 4-cypy·TFA, DIEA neutr, DIEA base, LiBr 0.2 M, 18 eq monomer, 11.7 min.

LCMS Method: Condition 4

Observed: 1601.62 Da Calculated: 1601.60 Da





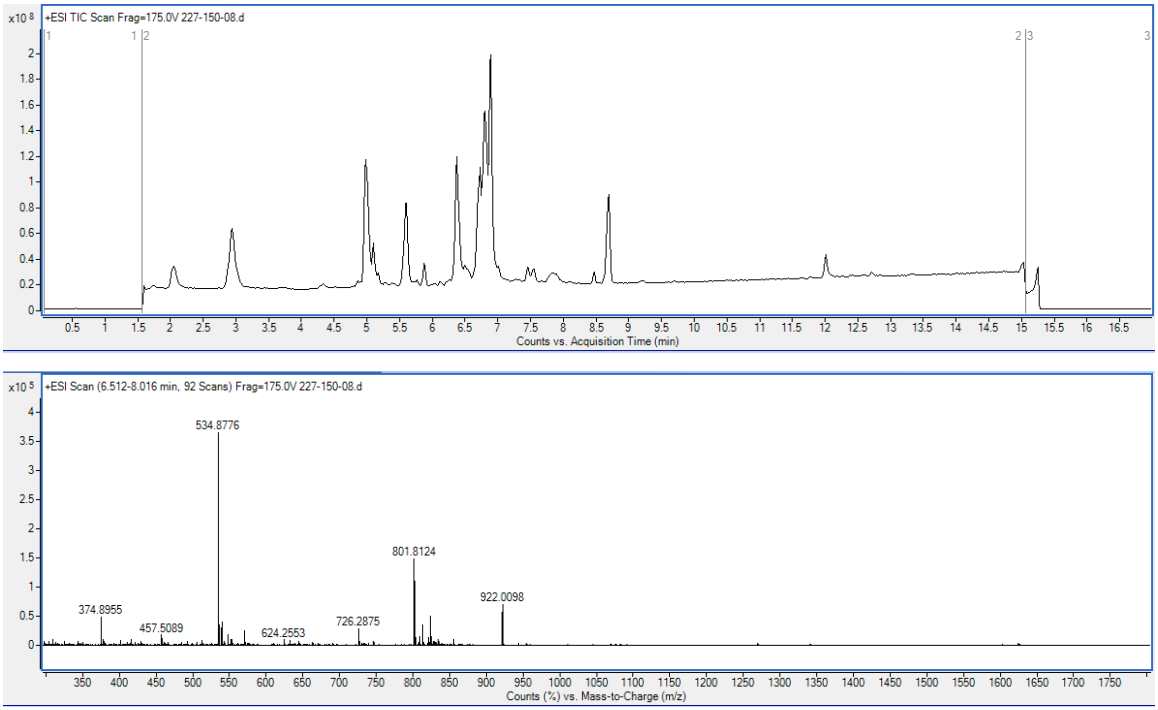
8. Fig. 1.4, Entry 8

Sequence (5’ to 3’): Tail-ACGT (4-mer)

Condition: Tiny Tides synthesis, 90 °C, 3,5-lut·TFA, DIEA neutr, DIEA base, LiBr 0.2 M, 18 eq monomer, 12.6 min.

LCMS Method: Condition 4

Observed: 1601.63 Da Calculated: 1601.60 Da



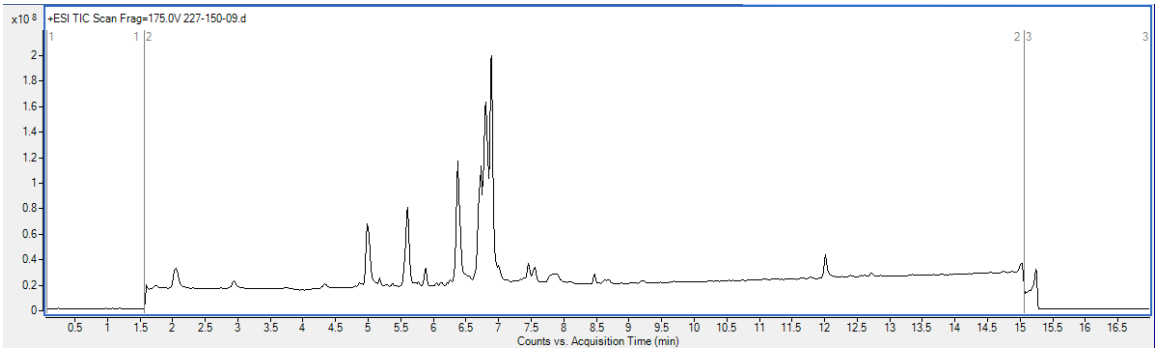
9. Fig. 1.4, Entry 9

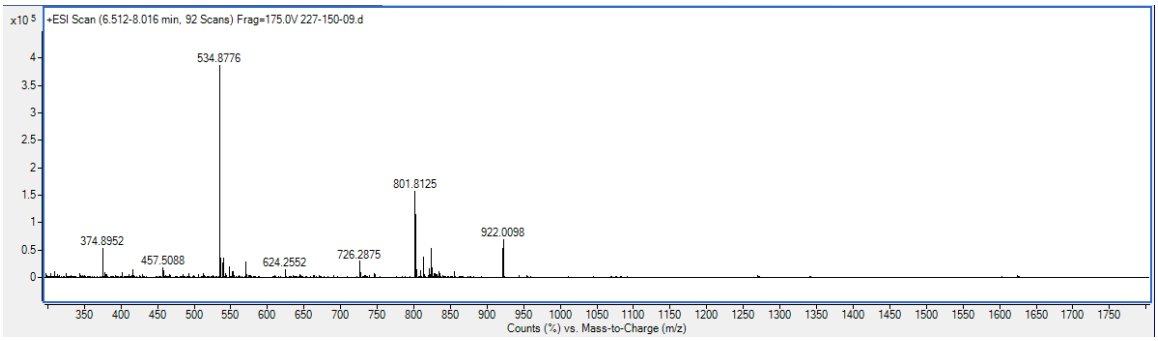
Sequence (5’ to 3’): Tail-ACGT (4-mer)

Condition: Tiny Tides synthesis, 90 °C, 3,5-lut·TFA, DIEA neutr, NEM base, LiBr 0.2 M, 18 eq monomer, 12.6 min.

LCMS Method: Condition 4

Observed: 1601.63 Da Calculated: 1601.60 Da





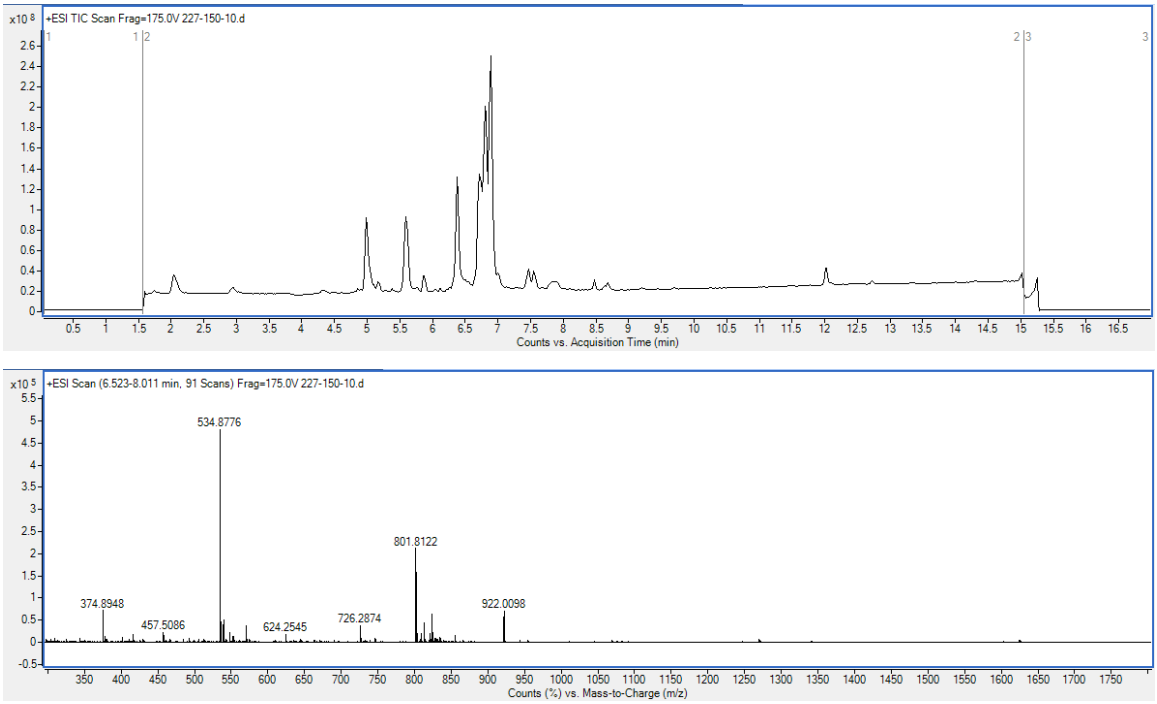
10. Fig. 1.4, Entry 10

Sequence (5’ to 3’): Tail-ACGT (4-mer)

Condition: Tiny Tides synthesis, 90 °C, 3,5-lut·TFA, NEM neutr, DIEA base, LiBr 0.2 M, 18 eq monomer, 12.6 min.

LCMS Method: Condition 4

Observed: 1601.62 Da Calculated: 1601.60 Da



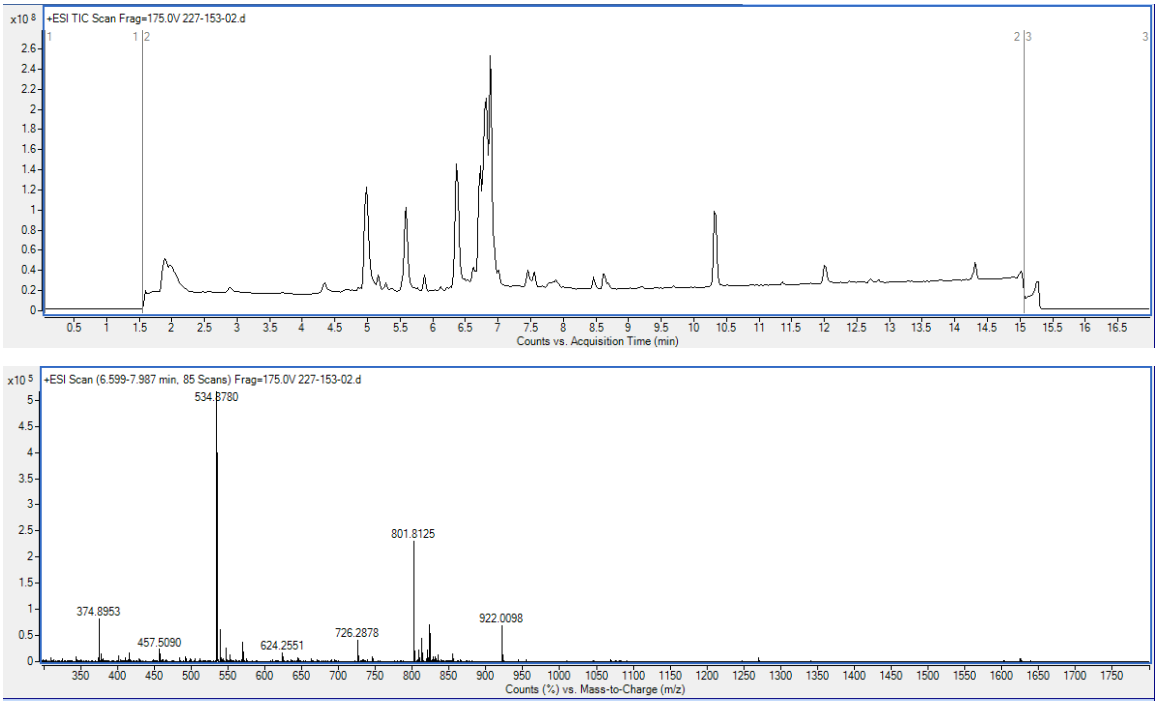
11. Fig. 1.4, Entry 11

Sequence (5’ to 3’): Tail-ACGT (4-mer)

Condition: Tiny Tides synthesis, 90 °C, 3,5-lut·TFA, NEM neutr, DIEA base, LiBr 0.2 M, 18 eq monomer, 12.7 min.

LCMS Method: Condition 4

Observed: 1601.63 Da Calculated: 1601.60 Da



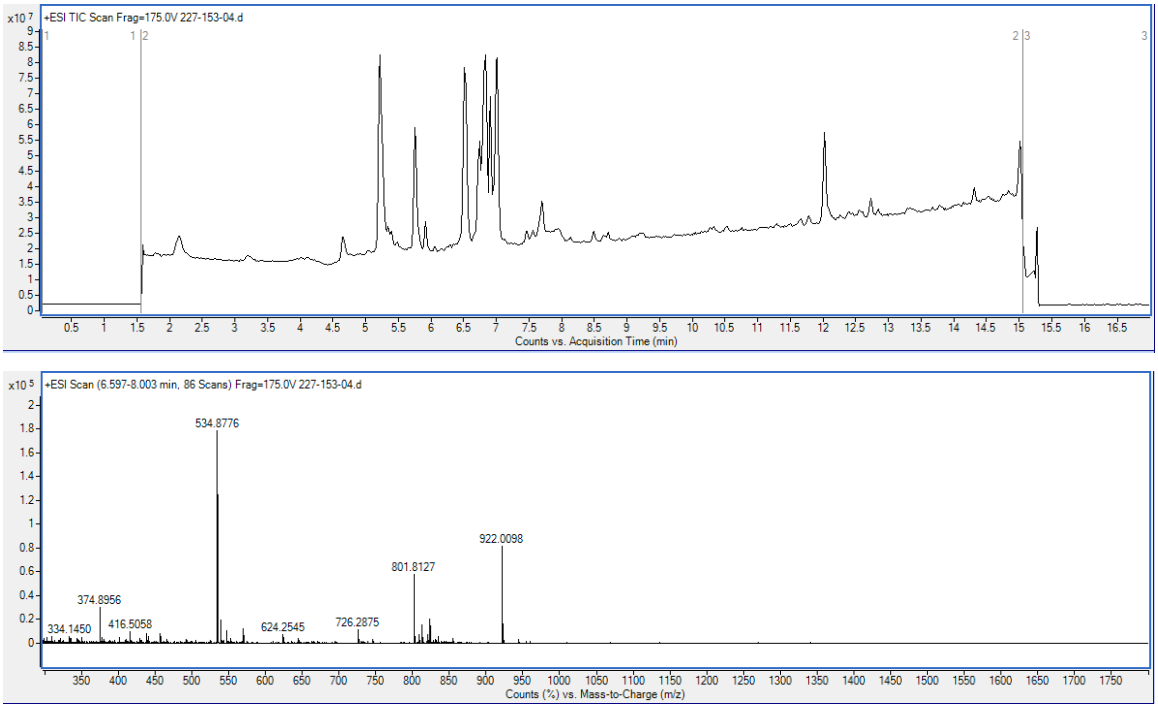
12. Fig. 1.4, Entry 12

Sequence (5’ to 3’): Tail-ACGT (4-mer)

Condition: Tiny Tides synthesis, 90 °C, 3,5-lut·TFA, NEM neutr, DIEA base, LiBr 0.2 M, 10 eq monomer, 22.1 min.

LCMS Method: Condition 4

Observed: 1601.63 Da Calculated: 1601.60 Da



1.4.9.2. LC-MS data of 12-mer PMO syntheses

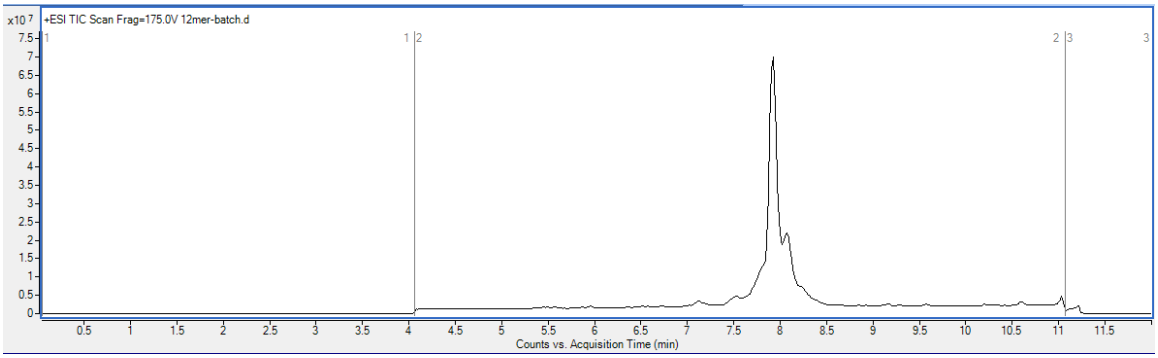
Batch synthesis of 12-mer PMO

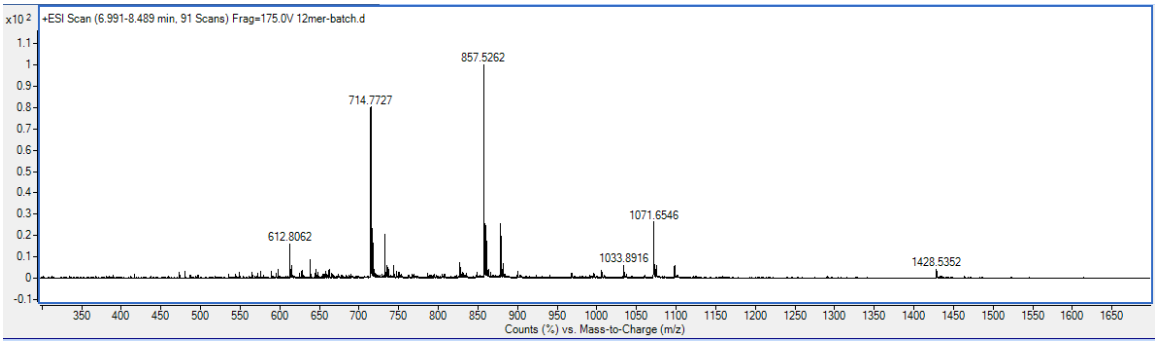
Sequence (5’ to 3’): Tail-ACGTACGTACGT (12-mer)

Condition: Batch synthesis (Section 1.4.3), 3-day synthesis.

LCMS Method: Condition 1

Observed: 4280.64 Da Calculated: 4280.52 Da





Tiny Tides synthesis of 12-mer PMO at 70˚C

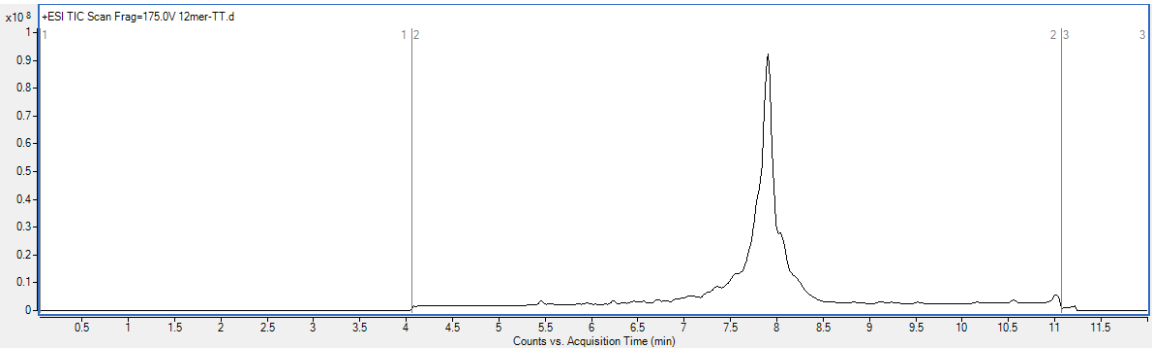
At 90 ˚C, 4-cyanopyridine·TFA solution will lead to acidolysis of the resin bound PMO. A 12-mer PMO was successfully synthesized with Tiny Tides at 70˚C using 4-cyanopyridine·TFA solution for detritylation, DIEA and LiBr were used for coupling, NEM was used for neutralization. Cleavage was performed using Method 1.

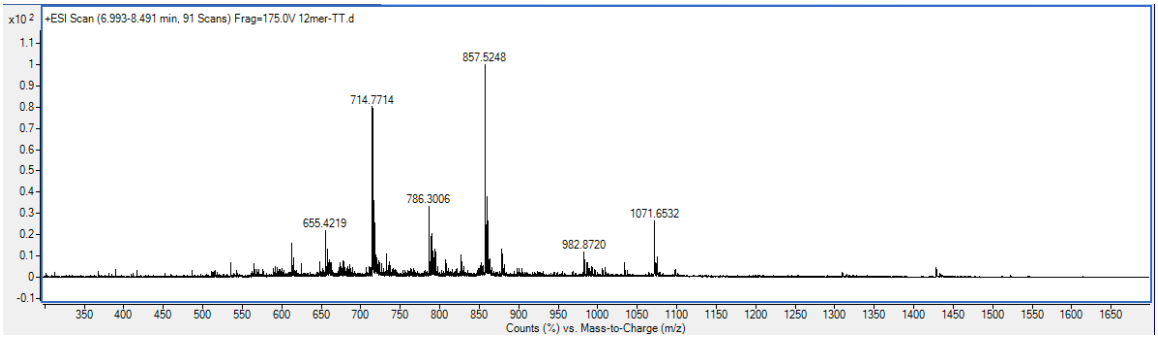
Sequence (5’ to 3’): Tail-ACGTACGTACGT (12-mer)

Condition: Tiny Tides synthesis, 70 °C, 4-cypy·TFA, NEM neutr, DIEA base, LiBr 0.2 M, 18 eq monomer, 2.3-hour synthesis.

LCMS Method: Condition 1

Observed: 4280.64 Da Calculated: 4280.52 Da





1.4.9.3.LC-MS data of 18-mer (IVS2-654) syntheses

The same 18-mer PMO was synthesized using both a batch protocol and rapid flow synthesis. Cleavage was performed using Method 2 for both samples. Purification of both samples was performed using reverse phase (Section 1.4.5).

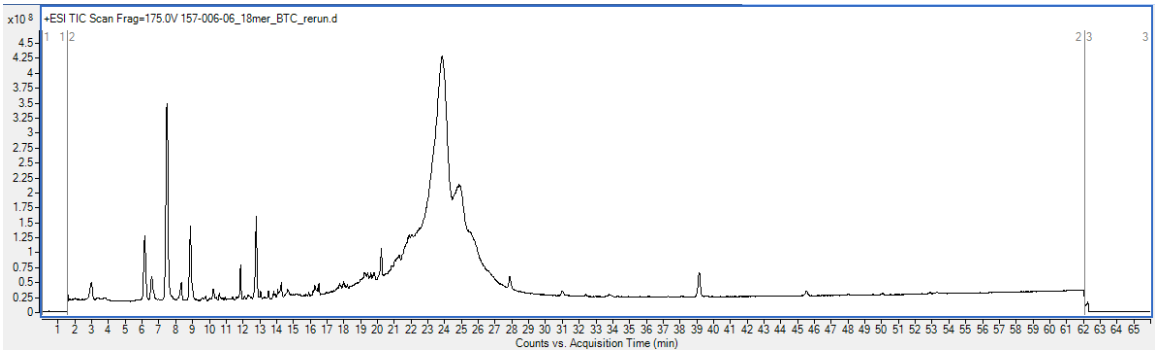
Batch synthesis of 18-mer PMO (IVS2-654) (crude)

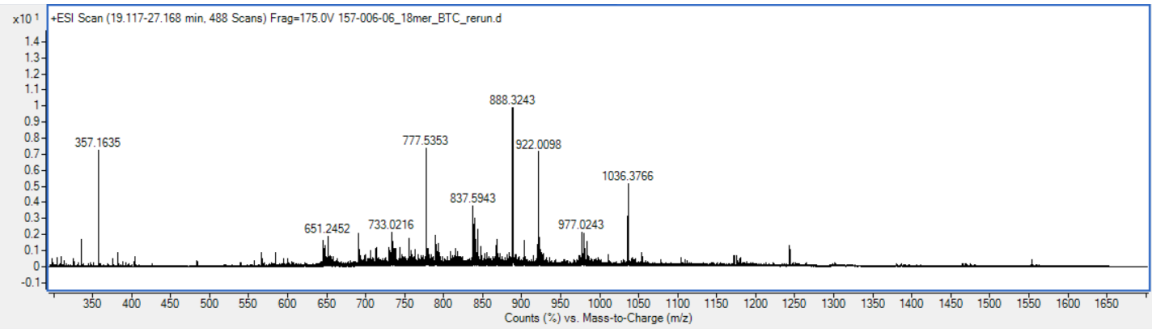
Sequence (5’ to 3’): Tail-GCTATTACCTTAACCCAG (18-mer)

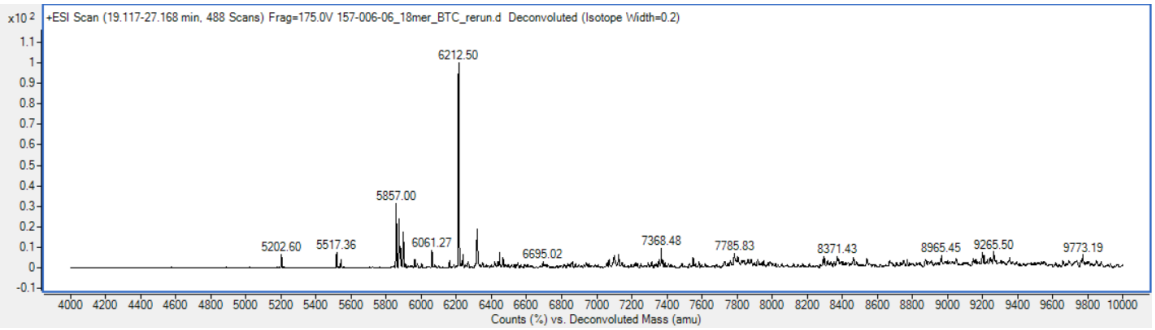
Condition: Batch synthesis (Supplementary Section 3), 1-week synthesis.

LCMS Method: Condition 3

Observed: 6209.23 Da Calculated: 6209.19 Da







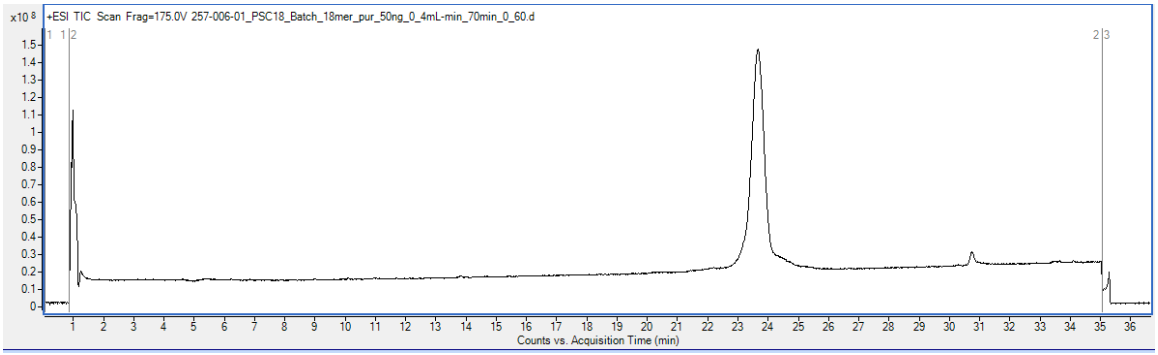
Batch synthesis of 18-mer PMO (IVS2-654) (post purification)

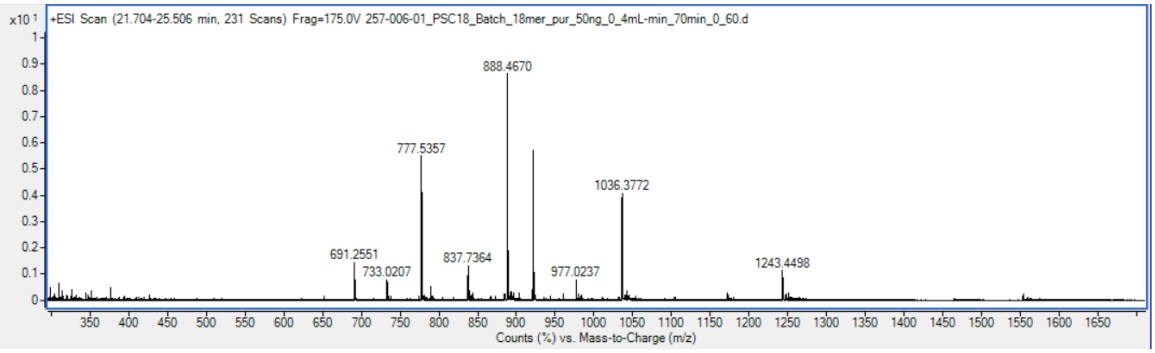
Sequence (5’ to 3’): Tail-GCTATTACCTTAACCCAG (18-mer)

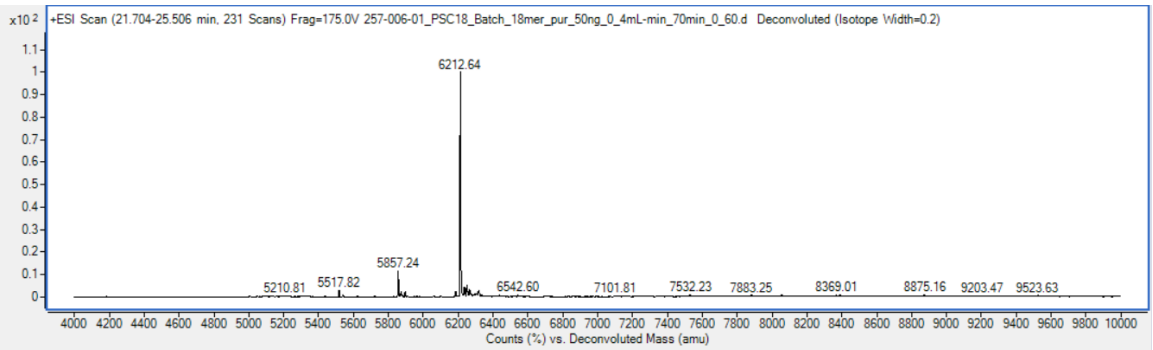
Condition: Batch synthesis (Section 1.4.3), 1-week synthesis.

LCMS Method: Condition 3

Observed: 6209.22 Da Calculated: 6209.19 Da







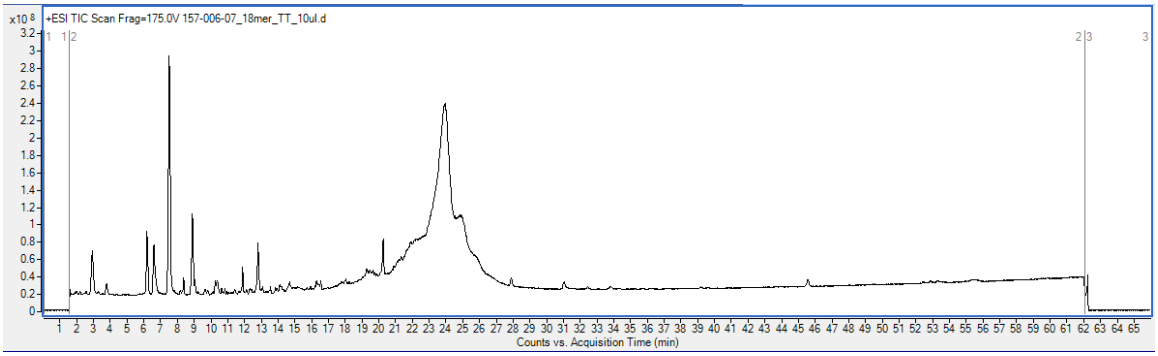
Tiny Tides synthesis of 18-mer PMO (IVS2-654) at 90˚C

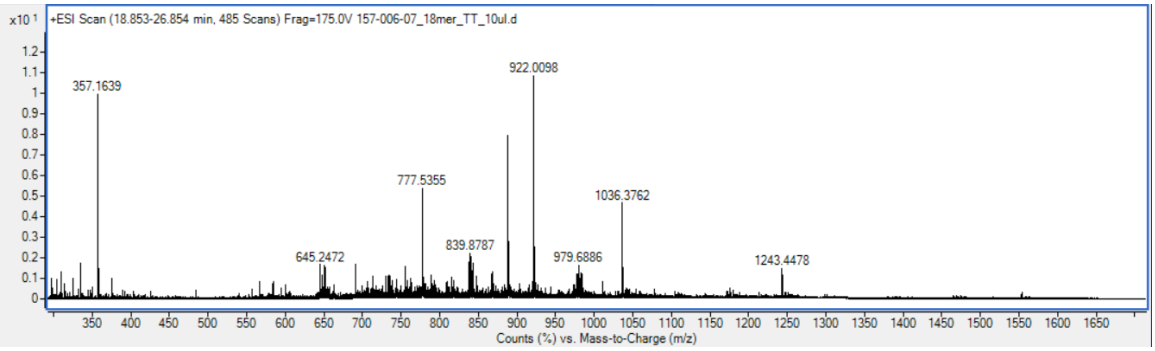
Sequence (5’ to 3’): Tail-GCTATTACCTTAACCCAG (18-mer)

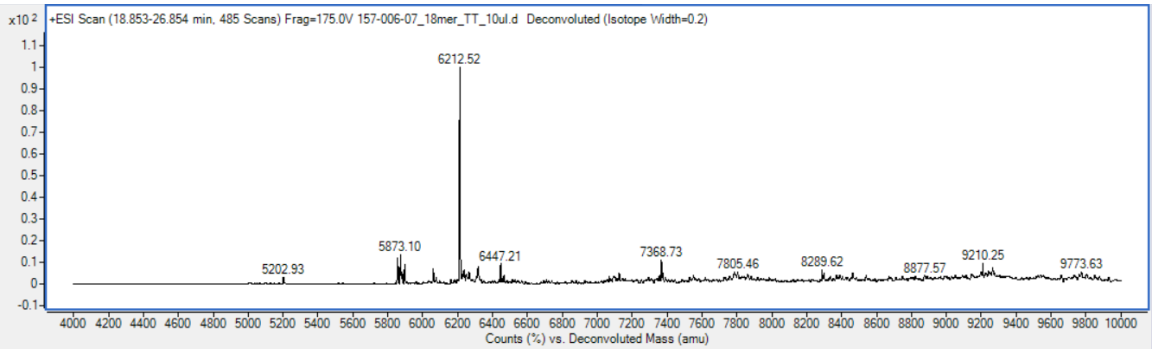
Condition: Tiny Tides synthesis, 90 °C, 3,5-lut·TFA, NEM neutr, DIEA base, LiBr 0.2 M, 18 eq monomer, 3.5 hours.

LCMS Method: Condition 3

Observed: 6209.23 Da Calculated: 6209.19 Da







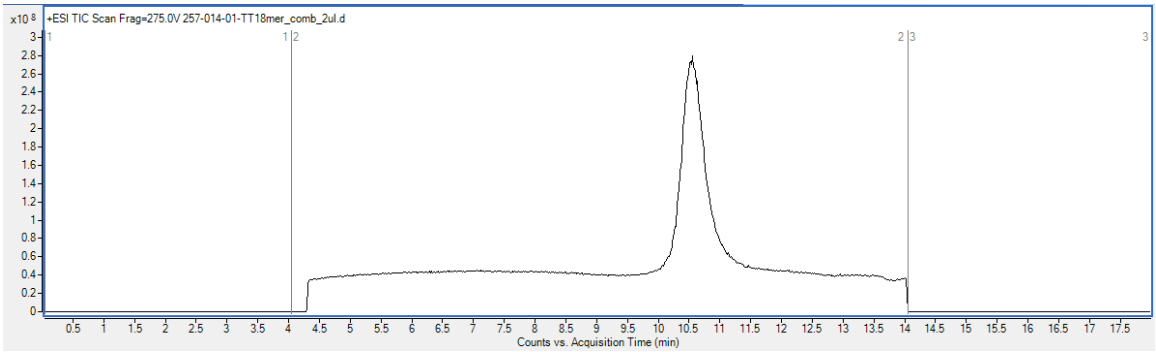
Tiny Tides synthesis of 18-mer PMO (IVS2-654) at 90˚C (post purification)

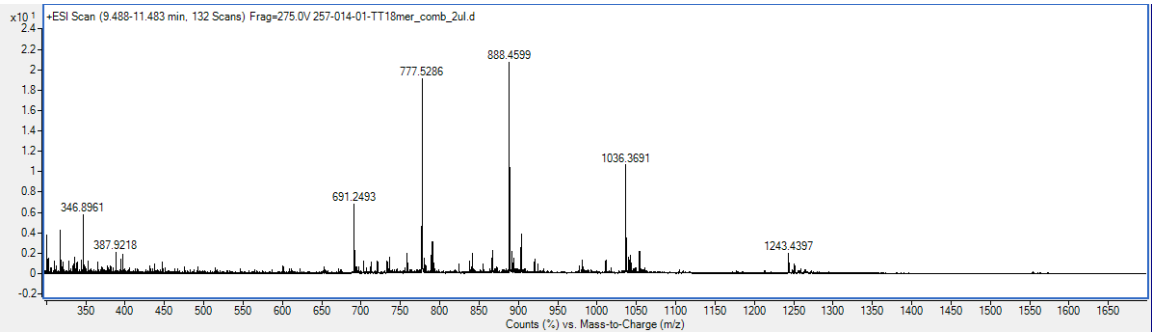
Sequence (5’ to 3’): Tail-GCTATTACCTTAACCCAG (18-mer)

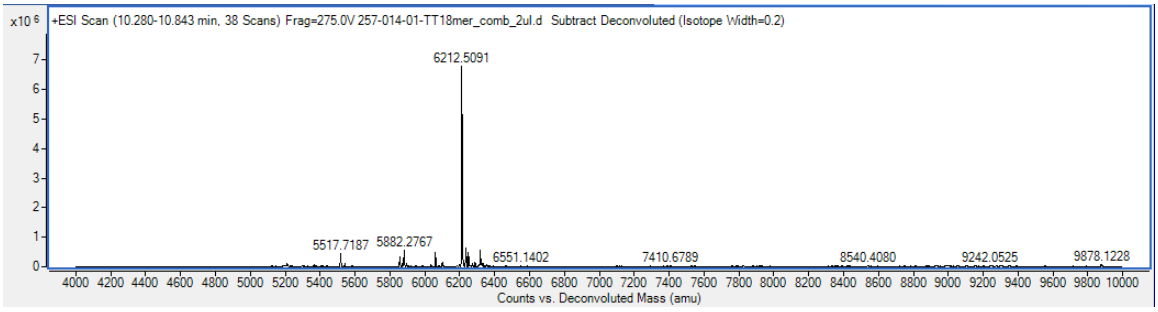
Condition: Tiny Tides synthesis, 90 °C, 3,5-lut·TFA, NEM neutr, DIEA base, LiBr 0.2 M, 18 eq monomer, 3.5 hours.

LCMS Method: Condition 2

Observed: 6209.20 Da Calculated: 6209.19 Da







1.4.9.4. LCMS data of exon 46 targeted 20-mer sequences

Three PMO sequences designed to skip exon 46 were synthesized using Tiny Tides. Starting from 10 mg of resin, after synthesis and purification, 1 mg of purified 20-mer sample was obtained for each sequence. Purification was performed in two stages using cation exchange followed by reverse phase (Section 1.4.5). Cleavage was performed using Method 2**.**

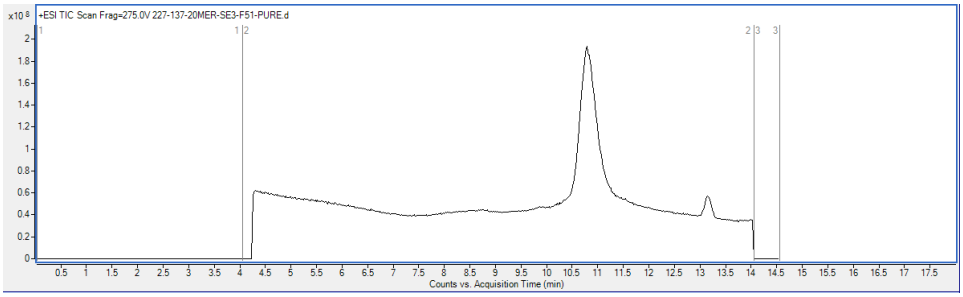
Tiny Tides synthesis of 20-mer PMO (seq. 1) at 90˚C

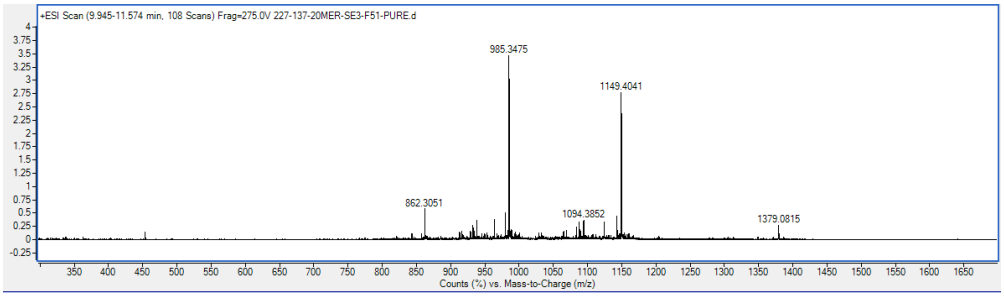
Sequence (5’ to 3’): Tail-CTTTTCTTTTAGTTGCTGCT (20-mer, seq. 1)

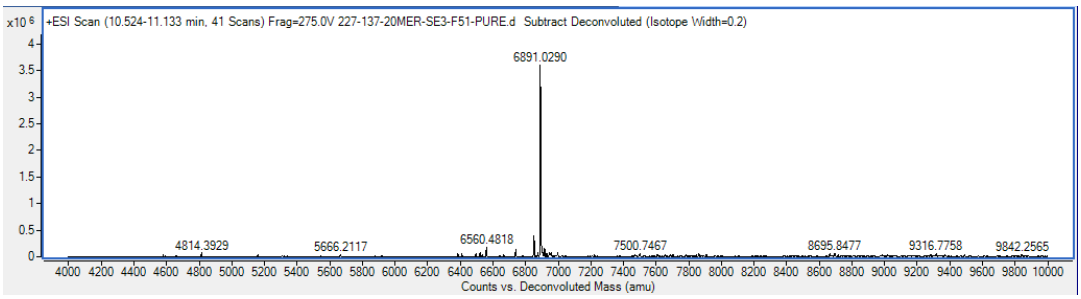
Condition: Tiny Tides synthesis, 90 °C, 3,5-lut·TFA, NEM neutr, DIEA base, LiBr 0.2 M, 18 eq monomer, 4.2 hours.

LCMS Method: Condition 2

Observed: 6887.37 Da Calculated: 6887.38 Da







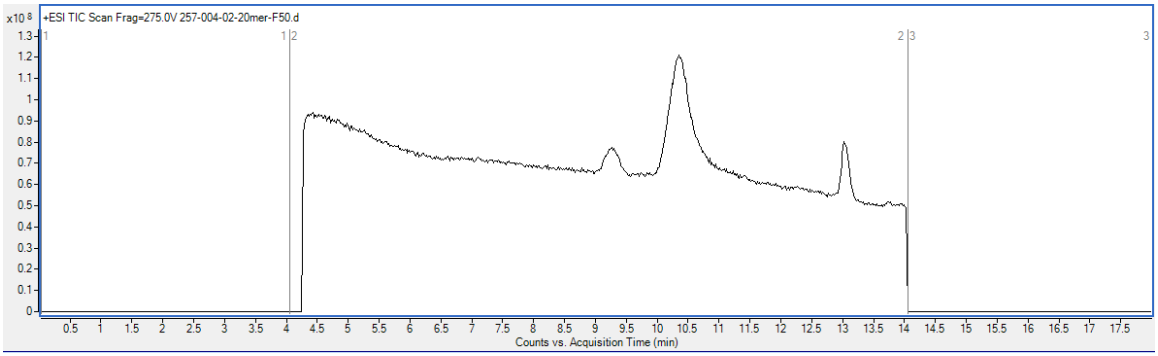
Tiny Tides synthesis of 20-mer PMO (seq. 2) at 90˚C

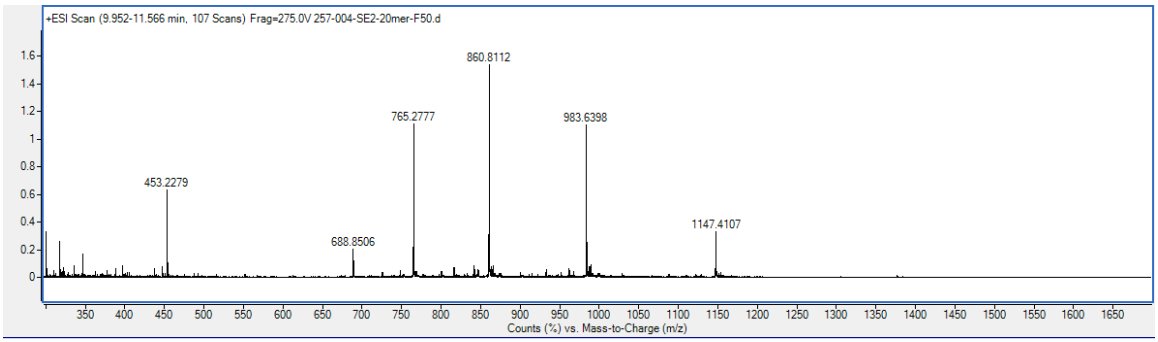
Sequence (5’ to 3’): Tail-CTGCTCTTTTCCAGGTTCAA (20-mer, seq. 2)

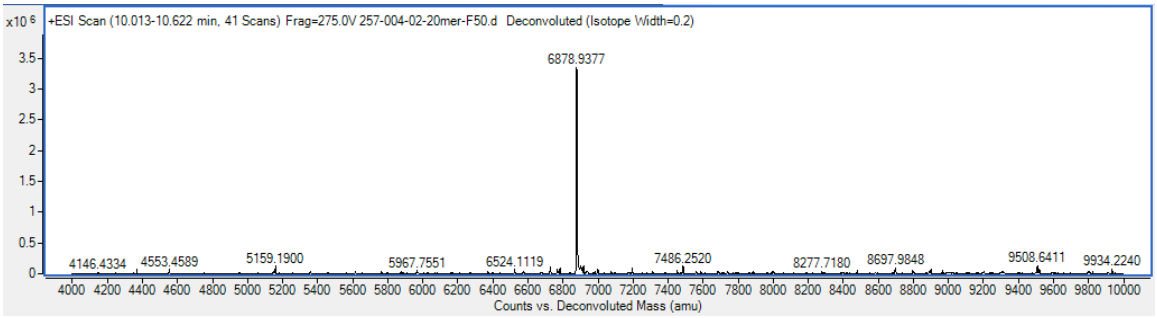
Condition: Tiny Tides synthesis, 90 °C, 3,5-lut·TFA, NEM neutr, DIEA base, LiBr 0.2 M, 18 eq monomer, 4.2 hours. Note that peaks at 9 and 13 mins are included in the background.

LCMS Method: Condition 2

Observed: 6876.45 Da Calculated: 6876.40 Da







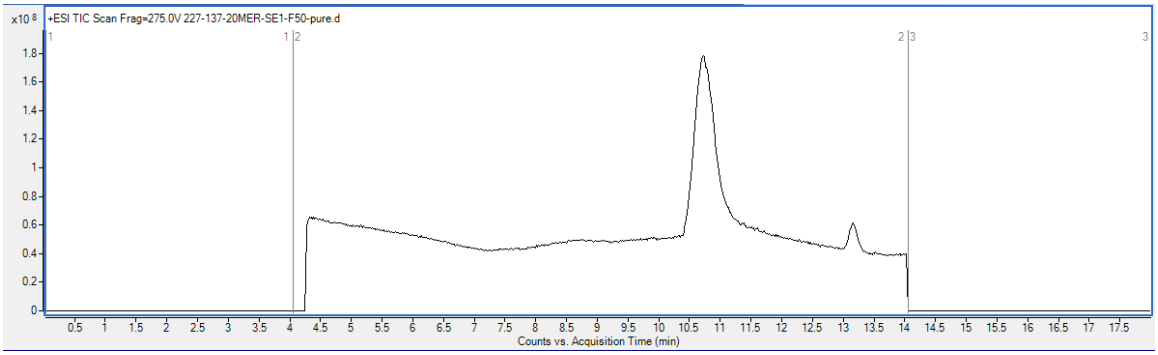
Tiny Tides synthesis of 20-mer PMO (seq. 3) at 90˚C

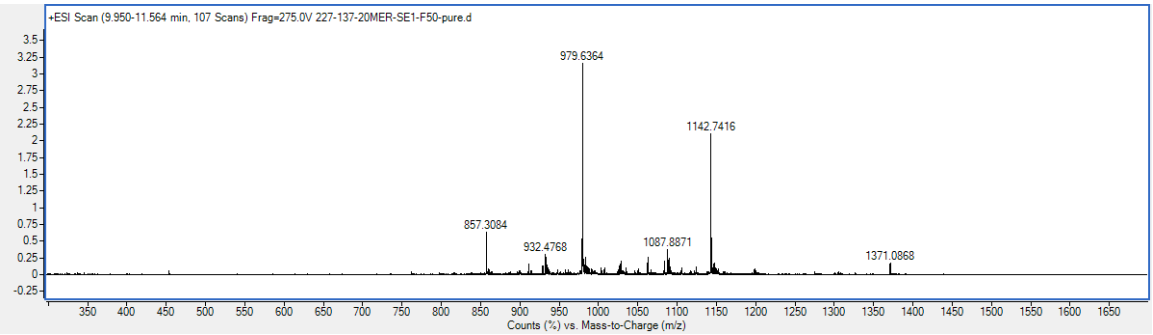
Sequence (5’ to 3’): Tail- TTCTTTTGTTCTTCTAGCCT (20-mer, seq. 3)

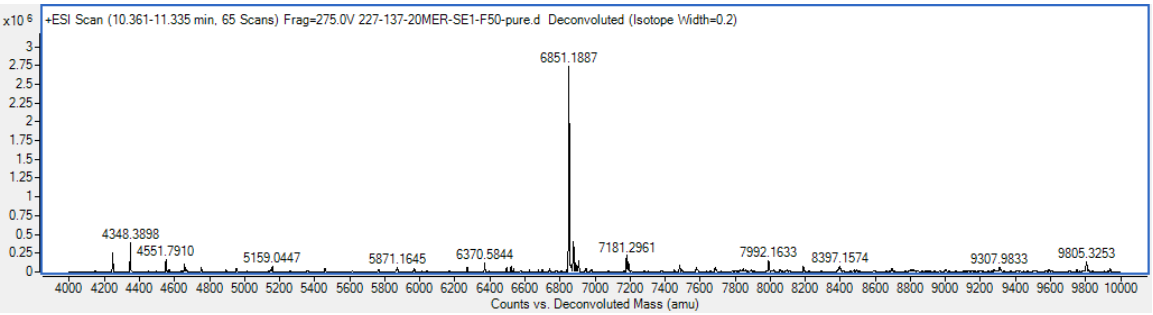
Condition: Tiny Tides synthesis, 90 °C, 3,5-lut·TFA, NEM neutr, DIEA base, LiBr 0.2 M, 18 eq monomer, 4.2 hours.

LCMS Method: Condition 2

Observed: 6848.41 Da Calculated: 6848.37 Da







1.4.9.5. LCMS data of SARS-CoV-2 5'UTR TRS targeted sequence

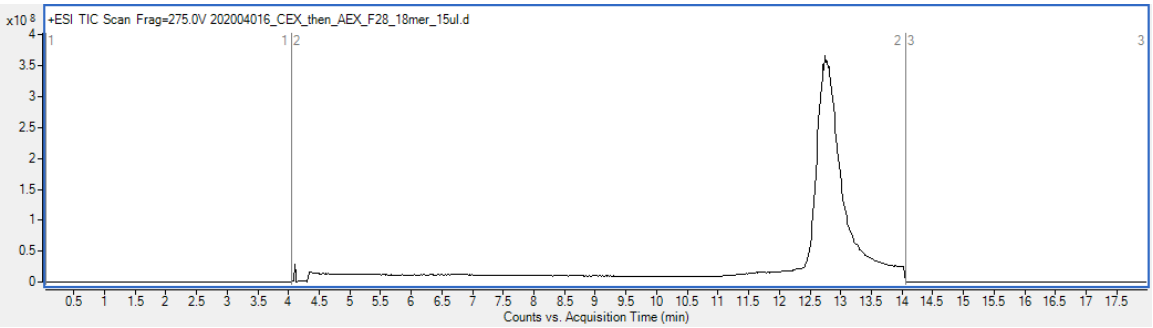
A PMO sequence designed to bind the TRS in the 5'UTR of the SARS-CoV-2 genomic mRNA was synthesized using Tiny Tides. Purification was carried out with CEX then AEX in sequence (Section 1.4.5).

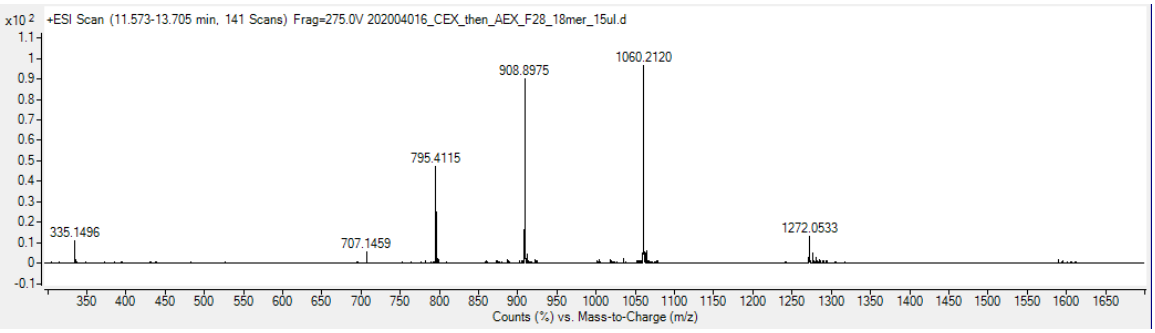
Sequence (5' to 3'): Tail-TAAAGTTCGTTTAGAGAA (18-mer)

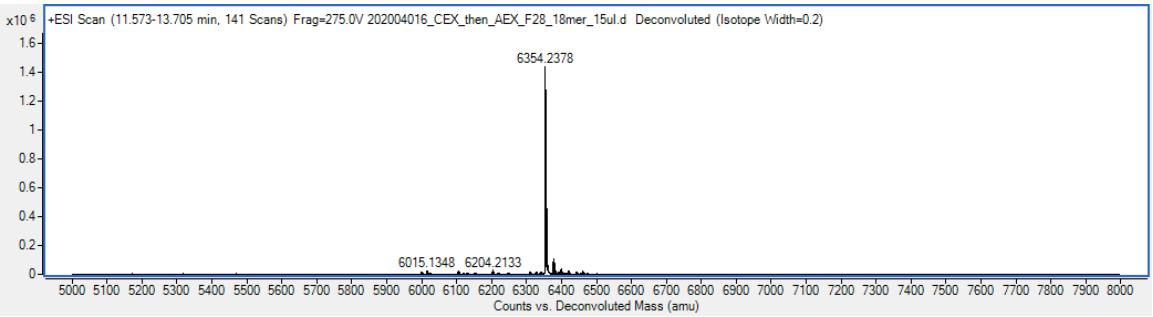
Condition: Tiny Tides synthesis, 90 °C, 3,5-lut·TFA, NEM neutr, DIEA base, LiBr 0.2 M, 20 eq monomer, 3.5 hours.

LCMS Method: Condition 2

Observed: 6352.22 Da Calculated: 6352.23 Da







1.4.9.6. LCMS data of purified fusion inhibitor EK1

A previously reported SARS-CoV-2 fusion inhibitor, EK1, was used as a positive control for the virus infection assay. Peptide was synthesized on a automated peptide synthesizer and purified with Biotage.

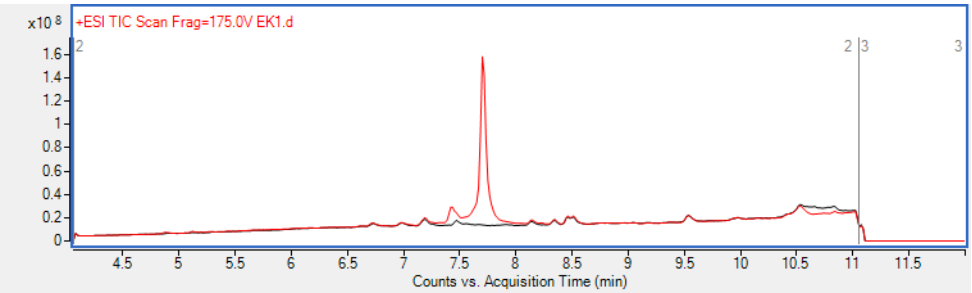
EK1 amino acid sequence: SLDQINVTFLDLEYEMKKLEEAIKKLEESYIDLKEL

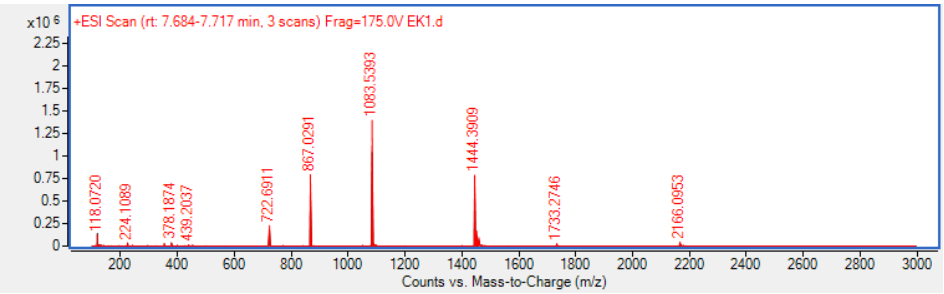
Synthesis condition: automated peptide synthesis, 90 °C, HATU (or PyAOP), DIEA base, 20% Piperidine, 28 eq monomer, 1.5 hours.

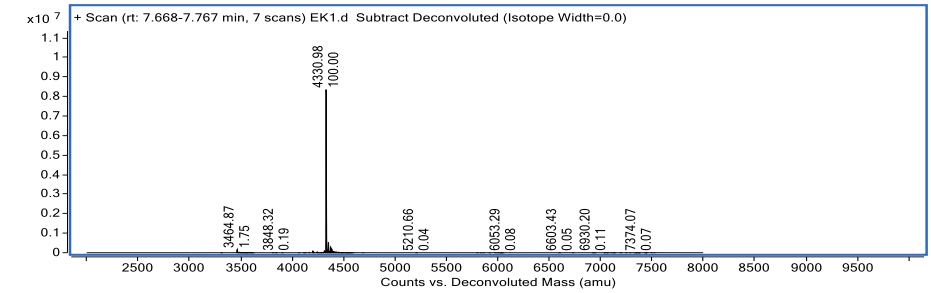
Purification condition: Biotage Sfär C18 D column (Duo 100 Å 30 µm, 12g), 5 to 60% acetonitrile over 12 column volume.

LCMS Method: Agilent Zorbax 300SB-C3 column, gradient: 1 to 91% acetonitrile over 15 mins.

Observed: 4328.17 Da Calculated: 4328.26 Da







1.5. References

1. J. Lieberman, Tapping the RNA world for therapeutics. *Nature Structural & Molecular Biology*. **25** (2018), pp. 357–364.

2. J. S. COHEN, Informational Drugs: A New Concept in Pharmacology. *Antisense Research and Development*. **1**, 191–193 (1991).

3. C. I. E. Smith, R. Zain, Therapeutic Oligonucleotides: State of the Art. *Annual Review of Pharmacology and Toxicology*. **59** (2019), pp. 605–630.

4. L. A. Jackson, E. J. Anderson, N. G. Rouphael, P. C. Roberts, M. Makhene, R. N. Coler, M. P. McCullough, J. D. Chappell, M. R. Denison, L. J. Stevens, A. J. Pruijssers, A. McDermott, B. Flach, N. A. Doria-Rose, K. S. Corbett, K. M. Morabito, S. O'Dell, S. D. Schmidt, P. A. Swanson, M. Padilla, J. R. Mascola, K. M. Neuzil, H. Bennett, W. Sun, E. Peters, M. Makowski, J. Albert, K. Cross, W. Buchanan, R. Pikaart-Tautges, J. E. Ledgerwood, B. S. Graham, J. H. Beigel, An mRNA Vaccine against SARS-CoV-2 — Preliminary Report. *N Engl J Med*. **383**, 1920–1931 (2020).

5. P. M. Folegatti, K. J. Ewer, P. K. Aley, B. Angus, S. Becker, S. Belij-Rammerstorfer, D. Bellamy, S. Bibi, M. Bittaye, E. A. Clutterbuck, C. Dold, S. N. Faust, A. Finn, A. L. Flaxman, B. Hallis, P. Heath, D. Jenkin, R. Lazarus, R. Makinson, A. M. Minassian, K. M. Pollock, M. Ramasamy, H. Robinson, M. Snape, R. Tarrant, M. Voysey, C. Green, A. D. Douglas, A. V. S. Hill, T. Lambe, S. C. Gilbert, A. J. Pollard, J. Aboagye, K. Adams, A. Ali, E. Allen, J. L. Allison, R. Anslow, E. H. Arbe-Barnes, G. Babbage, K. Baillie, M. Baker, N. Baker, P. Baker, I. Baleanu, J. Ballaminut, E. Barnes, J. Barrett, L. Bates, A. Batten, K. Beadon, R. Beckley, E. Berrie, L. Berry, A. Beveridge, K. R. Bewley, E. M. Bijker, T. Bingham, L. Blackwell, C. L. Blundell, E. Bolam, E. Boland, N. Borthwick, T. Bower, A. Boyd, T. Brenner, P. D. Bright, C. Brown-O'Sullivan, E. Brunt, J. Burbage, S. Burge, K. R. Buttigieg, N. Byard, I. Cabera Puig, A. Calvert, S. Camara, M. Cao, F. Cappuccini, M. Carr, M. W. Carroll, V. Carter, K. Cathie, R. J. Challis, S. Charlton, I. Chelysheva, J.-S. Cho, P. Cicconi, L. Cifuentes, H. Clark, E. Clark, T. Cole, R. Colin-Jones, C. P. Conlon, A. Cook, N. S. Coombes, R. Cooper, C. A. Cosgrove, K. Coy, W. E. M. Crocker, C. J. Cunningham, B. E. Damratoski, L. Dando, M. S. Datoo, H. Davies, H. De Graaf, T. Demissie, C. Di Maso, I. Dietrich, T. Dong, F. R. Donnellan, N. Douglas, C. Downing, J. Drake, R. Drake-Brockman, R. E. Drury, S. J. Dunachie, N. J. Edwards, F. D. L. Edwards, C. J. Edwards, S. C. Elias, M. J. Elmore, K. R. W. Emary, M. R. English, S. Fagerbrink, S. Felle, S. Feng, S. Field, C. Fixmer, C. Fletcher, K. J. Ford, J. Fowler, P. Fox, E. Francis, J. Frater, J. Furze, M. Fuskova, E. Galiza, D. Gbesemete, C. Gilbride, K. Godwin, G. Gorini, L. Goulston, C. Grabau, L. Gracie, Z. Gray, L. B. Guthrie, M. Hackett, S. Halwe, E. Hamilton, J. Hamlyn, B. Hanumunthadu, I. Harding, S. A. Harris, A. Harris, D. Harrison, C. Harrison, T. C. Hart, L. Haskell, S. Hawkins, I. Head, J. A. Henry, J. Hill, S. H. C. Hodgson, M. M. Hou, E. Howe, N. Howell, C. Hutlin, S. Ikram, C. Isitt, P. Iveson, S. Jackson, F. Jackson, S. W. James, M. Jenkins, E. Jones, K. Jones, C. E. Jones, B. Jones, R. Kailath, K. Karampatsas, J. Keen, S. Kelly, D. Kelly, D. Kerr, S. Kerridge, L. Khan, U. Khan, A. Killen, J. Kinch, T. B. King, L. King, J. King, L. Kingham-Page, P. Klenerman, F. Knapper, J. C. Knight, D. Knott, S. Koleva, A. Kupke, C. W. Larkworthy, J. P. J. Larwood, A. Laskey, A. M. Lawrie, A. Lee, K. Y. Ngan Lee, E. A. Lees, H. Legge, A. Lelliott, N.-M. Lemm, A. M. Lias, A. Linder, S. Lipworth, X. Liu, S. Liu, R. Lopez Ramon, M. Lwin, F. Mabesa, M. Madhavan, G. Mallett, K. Mansatta, I. Marcal, S. Marinou, E. Marlow, J. L. Marshall, J. Martin, J. McEwan, L. McInroy, G. Meddaugh, A. J. Mentzer, N. Mirtorabi, M. Moore, E. Moran, E. Morey, V.

Morgan, S. J. Morris, H. Morrison, G. Morshead, R. Morter, Y. F. Mujadidi, J. Muller, T. Munera-Huertas, C. Munro, A. Munro, S. Murphy, V. J. Munster, P. Mweu, A. Noé, F. L. Nugent, E. Nuthall, K. O'Brien, D. O'Connor, B. Oguti, J. L. Oliver, C. Oliveira, P. J. O'Reilly, M. Osborn, P. Osborne, C. Owen, D. Owens, N. Owino, M. Pacurar, K. Parker, H. Parracho, M. Patrick-Smith, V. Payne, J. Pearce, Y. Peng, M. P. Peralta Alvarez, J. Perring, K. Pfafferott, D. Pipini, E. Plested, H. Pluess-Hall, K. Pollock, I. Poulton, L. Presland, S. Provstgaard-Morys, D. Pulido, K. Radia, F. Ramos Lopez, J. Rand, H. Ratcliffe, T. Rawlinson, S. Rhead, A. Riddell, A. J. Ritchie, H. Roberts, J. Robson, S. Roche, C. Rohde, C. S. Rollier, R. Romani, I. Rudiansyah, S. Saich, S. Sajjad, S. Salvador, L. Sanchez Riera, H. Sanders, K. Sanders, S. Sapaun, C. Sayce, E. Schofield, G. Screaton, B. Selby, C. Semple, H. R. Sharpe, I. Shaik, A. Shea, H. Shelton, S. Silk, L. Silva-Reyes, D. T. Skelly, H. Smee, C. C. Smith, D. J. Smith, R. Song, A. J. Spencer, E. Stafford, A. Steele, E. Stefanova, L. Stockdale, A. Szigeti, A. Tahiri-Alaoui, M. Tait, H. Talbot, R. Tanner, I. J. Taylor, V. Taylor, R. Te Water Naude, N. Thakur, Y. Themistocleous, A. Themistocleous, M. Thomas, T. M. Thomas, A. Thompson, S. Thomson-Hill, J. Tomlins, S. Tonks, J. Towner, N. Tran, J. A. Tree, A. Truby, K. Turkentine, C. Turner, N. Turner, S. Turner, T. Tuthill, M. Ulaszewska, R. Varughese, N. Van Doremalen, K. Veighey, M. K. Verheul, I. Vichos, E. Vitale, L. Walker, M. E. E. Watson, B. Welham, J. Wheat, C. White, R. White, A. T. Worth, D. Wright, S. Wright, X. L. Yao, Y. Yau, Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. *The Lancet*. **396**, 467–478 (2020).

6. M. J. Mulligan, K. E. Lyke, N. Kitchin, J. Absalon, A. Gurtman, S. Lockhart, K. Neuzil, V. Raabe, R. Bailey, K. A. Swanson, P. Li, K. Koury, W. Kalina, D. Cooper, C. Fontes-Garfias, P.-Y. Shi, Ö. Türeci, K. R. Tompkins, E. E. Walsh, R. Frenck, A. R. Falsey, P. R. Dormitzer, W. C. Gruber, U. Şahin, K. U. Jansen, Phase I/II study of COVID-19 RNA vaccine BNT162b1 in adults. *Nature*. **586**, 589–593 (2020).

7. A. Khvorova, J. K. Watts, The chemical evolution of oligonucleotide therapies of clinical utility. *Nature Biotechnology*. **35** (2017), pp. 238–248.

8. J. SUMMERTON, D. WELLER, Morpholino Antisense Oligomers: Design, Preparation, and Properties. *Antisense and Nucleic Acid Drug Development*. **7** (1997), pp. 187–195.

9. R. M. HUDZIAK, E. BAROFSKY, D. F. BAROFSKY, D. L. WELLER, S.-B. HUANG, D. D. WELLER, Resistance of Morpholino Phosphorodiamidate Oligomers to Enzymatic Degradation. *Antisense and Nucleic Acid Drug Development*. **6**, 267–272 (1996).

10. V. Arora, D. C. Knapp, M. T. Reddy, D. D. Weller, P. L. Iversen, Bioavailability and Efficacy of Antisense Morpholino Oligomers Targeted to c-myc and Cytochrome P-450 3A2 Following Oral Administration in Rats. *Journal of Pharmaceutical Sciences*. **91**, 1009–1018 (2002).

11. K. L. Holden, D. A. Stein, T. C. Pierson, A. A. Ahmed, K. Clyde, P. L. Iversen, E. Harris, Inhibition of dengue virus translation and RNA synthesis by a morpholino oligomer targeted to the top of the terminal 3′ stem–loop structure. *Virology*. **344**, 439–452 (2006).

12. P. L. Iversen, T. K. Warren, J. B. Wells, N. L. Garza, D. V. Mourich, L. S. Welch, R. G. Panchal, S. Bavari, Discovery and early development of AVI-7537 and AVI-7288 for the treatment of Ebola virus and Marburg virus infections. *Viruses*. **4**, 2806–2830 (2012).

13. Q. Ge, M. Pastey, D. Kobasa, P. Puthavathana, C. Lupfer, R. K. Bestwick, P. L. Iversen, J. Chen, D. A. Stein, Inhibition of Multiple Subtypes of Influenza A Virus in Cell Cultures with Morpholino Oligomers. *Antimicrob. Agents Chemother.* **50**, 3724 (2006).

14. T. S. Deas, C. J. Bennett, S. A. Jones, M. Tilgner, P. Ren, M. J. Behr, D. A. Stein, P. L. Iversen, L. D. Kramer, K. A. Bernard, P.-Y. Shi, In Vitro Resistance Selection and In Vivo Efficacy of Morpholino Oligomers against West Nile Virus. *Antimicrob. Agents Chemother.* **51**, 2470 (2007).

15. B. W. Neuman, D. A. Stein, A. D. Kroeker, M. J. Churchill, A. M. Kim, P. Kuhn, P. Dawson, H. M. Moulton, R. K. Bestwick, P. L. Iversen, M. J. Buchmeier, Inhibition, Escape, and Attenuated Growth of Severe Acute Respiratory Syndrome Coronavirus Treated with Antisense Morpholino Oligomers. *Journal of Virology*. **79** (2005), p. 9665.

16. R. Burrer, B. W. Neuman, J. P. C. Ting, D. A. Stein, H. M. Moulton, P. L. Iversen, P. Kuhn, M. J. Buchmeier, Antiviral Effects of Antisense Morpholino Oligomers in Murine Coronavirus Infection Models. *Journal of Virology*. **81** (2007), pp. 5637–5648.

17. C. M. J. Fox, D. D. Weller, Method of synthesis of morpholino oligomers (2012), (available at https://patents.google.com/patent/US8299206/en?oq=8299206).

18. J. E. Summerton, D. D. Weller, Uncharged morpolino-based polymers having phosphorous containing chiral intersubunit linkages (1993), (available at https://patents.google.com/patent/US5185444A/en?oq=5185444).

19. T. Harakawa, H. Tsunoda, A. Ohkubo, K. Seio, M. Sekine, Development of an efficient method for phosphorodiamidate bond formation by using inorganic salts. *Bioorganic & Medicinal Chemistry Letters*. **22** (2012), pp. 1445–1447.

20. J. Bhadra, S. Pattanayak, S. Sinha, Synthesis of Morpholino Monomers, Chlorophosphoramidate Monomers, and Solid-Phase Synthesis of Short Morpholino Oligomers. *Current Protocols in Nucleic Acid Chemistry*. **62** (2015), p. 4.65.1-4.65.26, , doi:10.1002/0471142700.nc0465s62.

21. M. Schnölzer, P. Alewood, A. Jones, D. Alewood, S. B. H. Kent, In Situ Neutralization in Boc-chemistry Solid Phase Peptide Synthesis. *International Journal of Peptide Research and Therapeutics*. **13** (2007), pp. 31–44.

22. O. J. Plante, E. R. Palmacci, P. H. Seeberger, Automated Solid-Phase Synthesis of Oligosaccharides. *Science*. **291**, 1523–1527 (2001).

23. J. Yoshida, Y. Takahashi, A. Nagaki, Flash chemistry: flow chemistry that cannot be done in batch. *Chem. Commun.* **49**, 9896–9904 (2013).

24. Y.-A. Heo, Golodirsen: First Approval. *Drugs*. **80**, 329–333 (2020).

25. Y. Y. Syed, Eteplirsen: First Global Approval. *Drugs*. **76**, 1699–1704 (2016).

26. S. Dhillon, Viltolarsen: First Approval. *Drugs*. **80**, 1027–1031 (2020).

27. A. Aartsma-Rus, I. Fokkema, J. Verschuuren, I. Ginjaar, J. van Deutekom, G.-J. van Ommen, J. T. den Dunnen, Theoretic applicability of antisense-mediated exon skipping for Duchenne muscular dystrophy mutations. *Human Mutation*. **30**, 293–299 (2009).

28. J. C. T. van Deutekom, G.-J. B. van Ommen, Advances in Duchenne muscular dystrophy gene therapy. *Nature Reviews Genetics*. **4** (2003), pp. 774–783.

29. 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. *Nature Chemical Biology*. **13** (2017), p. 464.

30. N. Hartrampf, A. Saebi, M. Poskus, Z. P. Gates, A. J. Callahan, A. E. Cowfer, S. Hanna, S. Antilla, C. K. Schissel, A. J. Quartararo, X. Ye, A. J. Mijalis, M. D. Simon, A. Loas, S. Liu, C. Jessen, T. E. Nielsen, B. L. Pentelute, Synthesis of proteins by automated flow chemistry. *Science*. **368**, 980 (2020).

31. B. Lee, A. J. Mijalis, littleblackfish, N. M. Pohl, a-callahan/MechWolf_Pull 0.1.1. (2020), (available at doi:10.5281/zenodo.3774509.).

32. C. M. Fadzen, R. L. Holden, J. M. Wolfe, Z.-N. Choo, C. K. Schissel, M. Yao, G. J. Hanson, B. L. Pentelute, Chimeras of Cell-Penetrating Peptides Demonstrate Synergistic Improvement in Antisense Efficacy. *Biochemistry*. **58**, 3980–3989 (2019).

33. P. Sazani, S.-H. Kang, M. A. Maier, C. Wei, J. Dillman, J. Summerton, M. Manoharan, R. Kole, Nuclear antisense effects of neutral, anionic and cationic oligonucleotide analogs. *Nucleic Acids Research*. **29**, 3965–3974 (2001).

34. J. C. T. van Deutekom, M. Bremmer-Bout, A. A. M. Janson, I. B. Ginjaar, F. Baas, J. T. den Dunnen, G.-J. B. van Ommen, Antisense-induced exon skipping restores dystrophin expression in DMD patient derived muscle cells. *Human Molecular Genetics*. **10**, 1547–1554 (2001).

35. P. Sazani, R. Kole, Multiple exon skipping compositions for DMD (2016), (available at https://patents.google.com/patent/US9434948/en?oq=9434948).

36. K. Rosenke, S. Leventhal, H. M. Moulton, S. Hatlevig, D. Hawman, H. Feldmann, D. A. Stein, Inhibition of SARS-CoV-2 in Vero cell cultures by peptide-conjugated morpholino oligomers. *Journal of Antimicrobial Chemotherapy*. **76**, 413–417 (2021).

37. S. Xia, L. Yan, W. Xu, A. S. Agrawal, A. Algaissi, C.-T. K. Tseng, Q. Wang, L. Du, W. Tan, I. A. Wilson, S. Jiang, B. Yang, L. Lu, A pan-coronavirus fusion inhibitor targeting the HR1 domain of human coronavirus spike. *Sci Adv*. **5**, eaav4580 (2019).

38. S. Xia, M. Liu, C. Wang, W. Xu, Q. Lan, S. Feng, F. Qi, L. Bao, L. Du, S. Liu, C. Qin, F. Sun, Z. Shi, Y. Zhu, S. Jiang, L. Lu, Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. *Cell Research*. **30**, 343–355 (2020).

39. H. S. Folger, *Elements of chemical reaction engineering, fifth edition* (Prentice Hall, 2016).

40. O. Levenspiel, *Chemical Reaction Engineering, third edition* (Wiley, 1998).

41. 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).

42. K. D. Nagy, B. Shen, T. F. Jamison, K. F. Jensen, Mixing and Dispersion in Small-Scale Flow Systems. *Organic Process Research & Development*. **16** (2012), pp. 976–981.

1.6. Acknowledgements

Sarepta Therapeutics is gratefully acknowledged for providing financial support for this work and the reagents for PMO production. A.J.C. is supported by the MIT School of Science Fellowship in Cancer Research. We thank Adam Chevalier and Jason Gatlin at Sarepta Therapeutics for helpful suggestions on optimizing the PMO synthesis cycle in flow. N.L.B.P. is grateful to the Radcliffe Institute for Advanced Studies for the Edward, Frances, and Shirley B. Daniels Fellow position and to Benjamin Lee (both at Harvard University) for fostering the initial software development. We thank Andrew Wilson at Detailed Dynamic for his help in the design and production of the automated instrument.

1.7.Author Contributions

C.L. and A.J.C. built the updated PMO synthesizer, improved the Python program, optimized fast-flow synthesis conditions, synthesized, and purified the PMO sequences. M.D.S. and K.A.T. performed PMO stability studies. A.J.M. and N.L.B.P. designed the open-source synthesis software, Mechwolf. A.J.M. helped build the PMO synthesizer. K.-S.P. performed the SARS-CoV-2 viral infections. D.V. designed and supervised the SARS-CoV-2 inhibition assays, and performed the viral detections. G.Z. synthesized EK1, and helped plan the SARS-CoV-2 inhibition assays. N.H. and C.K.S. helped update the PMO synthesizer. M.Z., H.Z., and G.K.H. provided input on optimization of the PMO synthesis cycle. C.L., A.J.C., M.J.S., K.A.T., and B.L.P. conceptualized the research and designed the experiments. C.L., A.J.C., A.L., M.D.S., and B.L.P. wrote the manuscript with input from all coauthors.

2. Single-Shot Flow Synthesis of D-Proteins for Mirror-Image Biochemistry and Ligand Discovery

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

Callahan, A.J.; Gandhesiri, S.; Travaline, T.L., Lozano Salazar, L.; Hanna, S.; Lee, Y.-C.; Tokareva, O. S.; Swiecicki, J.-M.; Loas, A.; Verdine, G.L.; Pentelute, B.L. Single-Shot Flow Synthesis of D-Proteins for Mirror-Image Phage Display. *ChemRxiv*; **2023**; This content is a preprint and has not been peer-reviewed.

2.1. Introduction

The biopolymers that carry out the central dogma of all known life are generated from a set of homochiral building blocks: D-deoxyribonucleic acids for DNA, D-ribonucleic acids for RNA, and L-amino acids for proteins. When strung together, these chiral monomers form three-dimensional structures with biological activities that are inherently chiral. Alternatively, unnatural biopolymers made from building blocks of the opposite chirality form mirror-image three-dimensional structures. These compounds retain the activity of their natural counterpart in the mirror-image biochemical space, but are not recognized nor degraded by natural biomolecules (*4*–*7*). As a result, mirror-image biopolymers are useful tools to probe or perturb living systems. One such class of mirror-image compounds, D-peptides, are particularly valuable as therapeutic scaffolds, but remain under-developed.

Metabolic stability is a key bottleneck for the design of peptide-based therapeutics, and mirror-image peptides are uniquely resistant to degradation. The greater chiral and structural complexity of peptides make them valuable scaffolds to target protein surfaces not well-suited to disruption with small molecules (*8*). Despite these benefits, unmodified canonical peptides are rapidly degraded in vivo, and chemical modification to extend their biological half-life is critical in the design of peptide-based drugs. Modern techniques for peptide drug discovery integrate stabilizing modifications early into design cycles, but stability is constantly monitored throughout development (*9*). Mirror-image D-peptides, on the other hand, are not recognized by natural proteases, and therefore display long half-lives in biological systems on their own (*10*, *11*). As such, D-peptides are a privileged scaffold for peptide drug discovery, especially when combined with macrocyclic linkages or ‘staples’. Despite these benefits, modern discovery efforts rarely utilize mirror-image peptides, as the biological display techniques critical to modern peptide drug discovery efforts are not compatible with D-peptides. Development of mirror-image peptide-based therapeutics is limited to techniques that as of yet have not had sufficient throughput for modern drug discovery efforts.

Mirror-image phage display (MIPD) can generate high affinity D-peptide ligands to challenging protein interfaces, but its scope has remained limited. In MIPD, a library of L-

peptides displayed on phage surfaces are screened against a ‘bait’ mirror-image protein (*12*). By symmetry, the interaction of a D-peptide with an L-protein is identical to that of the screened L-peptide to a D-protein. As a result, the discovered L-peptide binders provide a blueprint for mirror-image D-peptide binders to the natural L-protein. This strategy has been employed to generate D-peptide ligands to therapeutically relevant protein targets, including mouse double minute 2 homolog (MDM2) (*13*), epidermal growth factor (EGF) (*14*), and amyloid beta (Aβ) (*15*). To the best of our knowledge, however, MIPD has only been applied to nine protein targets in total (*12*–*23*) during the 30 years since its introduction. Technical challenges in accessing the required D-protein baits mean only expert protein synthesis chemists are able to successfully apply this technique. Modern drug discovery efforts rely on simultaneous screening of many protein targets, on the understanding that successful translation into the clinic for any one variant is unlikely. Despite the value of a D-peptide scaffold, the low throughput of MIPD has stifled its widespread adoption in drug development.

A significant limitation to the broad application of MIPD relates to the challenges associated with accessing useful amounts of the required synthetic D-proteins. Modern chemical synthesis techniques have enabled access to synthetic proteins, but are not always practical nor dependable (*24*, *25*). Accessing sufficient amounts of folded synthetic protein to carry out phage display screening can require significant effort and expertise to define and optimize a suitable synthetic route. This outlook is further complicated by the fact that the requisite protected D-amino acids are on average ~10-times more expensive than their L-counterparts (*7*). Such considerations have driven the development of new methods to enhance access to synthetic proteins, but throughput remains low. We recently reported a protocol termed automated fast-flow peptide synthesis (AFPS, Fig. 2.1) (*26*, *27*), that utilizes highly optimized flow chemistry to produce long peptides (> 50 AAs) with unmatched speed and purity. We anticipated that this technology could overcome sequence-dependent complications that are ubiquitous in the production of novel protein targets for MIPD.

Here, we applied AFPS to a panel of 12 diverse proteins, eight of which had not been previously synthesized by chemical methods. We successfully isolated milligram

amounts of 12 proteins in both L- and D-forms (24 total proteins) (Fig. 2.1). To our knowledge, these syntheses represent a major fraction of the total number of D-proteins reported in the past five years (12 out of 36 reported targets (2–4, 11, 14, 15, 19, 20, 25–33, 33–39)).

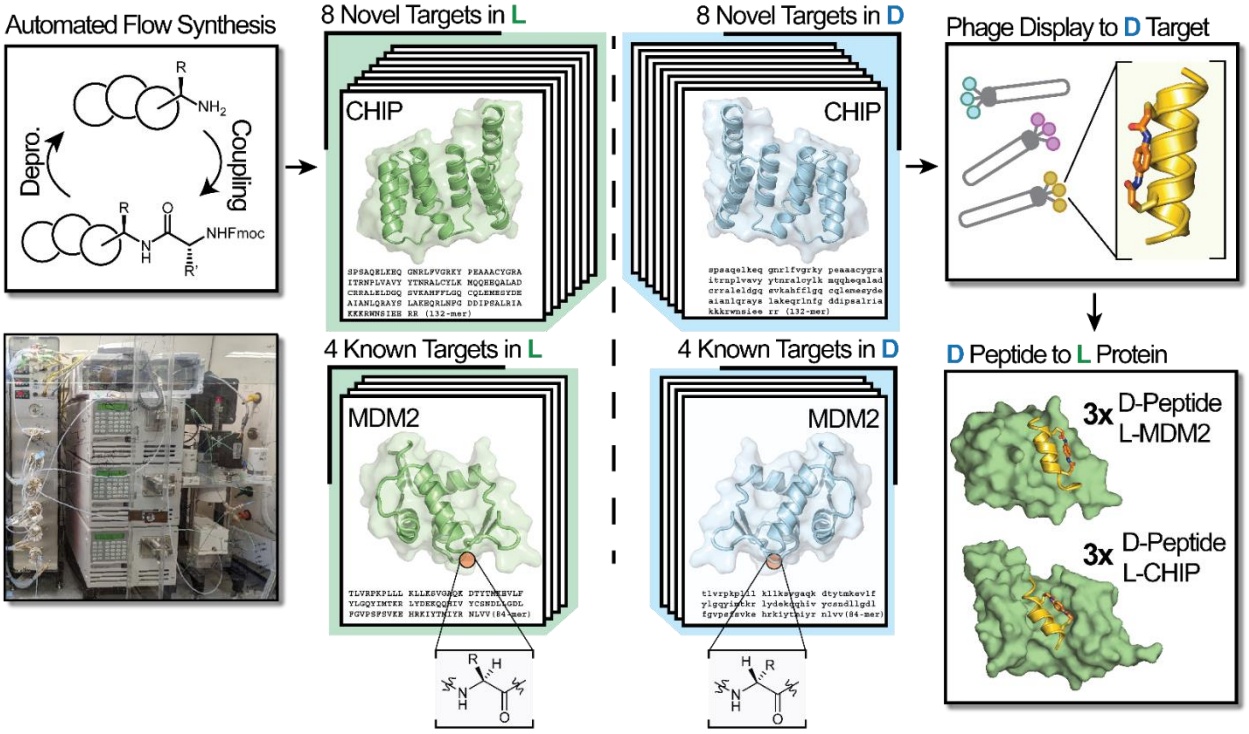


Figure 2.1 Rapid generation of D-peptide binders leverages automated flow protein synthesis coupled with mirror-image phage display.

Automated flow protein synthesis can rapidly manufacture proteins with L- or D-amino acids. Subsequent folding and biochemical purification of each polypeptide chain affords synthetic protein pairs which are mirror images of each other. Phage display screening of the D-protein enantiomer with stapled alpha-helical peptides can reveal sequences with low-micromolar binding affinity. Mirroring of the hit peptides from L- to D-chirality affords binders to the native L-protein with similar affinity.

With reliable access to synthetic D-proteins, we aimed to generate biologically stable macrocyclic D-peptide ligands to therapeutic protein targets. We utilized a recently reported screening platform based on phage display to generate high-affinity conformationally constrained α-helical peptides in a single screening pass (40). We used this platform to discover macrocyclic binders to two of the D-proteins we synthesized here, the established oncoprotein MDM2 and a novel E3 ubiquitin ligase, the C-terminus of Hsc70-interacting protein (CHIP). The resulting hits included a total of six distinct families of mirror-image binders, three to MDM2, and three to CHIP. Analysis of the X-ray

co-crystal structures between the discovered D-peptide binders and their protein targets revealed in some cases similar side-chain interactions to known binders that utilize different structural scaffolds, including loops, and α-helices with opposite handedness. These screening results represent a major advance in the throughput and general accessibility of D-peptide binder discovery. We anticipate that MIPD enabled by AFPS is poised to revitalize campaigns to generate mirror-image binders to existing and emerging protein targets.

2.2. Results

2.2.1. Synthesis of enantiomeric protein pairs with AFPS

We first defined a panel of 12 structurally diverse single-domain proteins for generation of synthetic L- and D-versions using AFPS (Fig. 2.2). Two of these proteins, MDM2 (*13*) and Barnase (*28*), were previously prepared in their all-D forms (see Sections 2.4.11.1 and 2.4.11.3), so can be used to assess the reliability and fidelity of the AFPS approach. We also chose two proteins that had previously only been synthesized in their native L-forms: Myc-Max, and Max-Max (*29*) (see Sections 2.4.11.17– 2.4.11.14). We chose an additional eight single-domain proteins ranging in length from 70-132 amino acids with no reported chemical synthesis: ERG (PNT domain), (see Section 2.4.11.2) IRAK2 (death domain) (see Section 2.4.11.4), CHIP (tetranucleotide repeat domain) (see Section 2.4.11.5), NEMO (coiled coil domain) (see Section 2.4.11.6), FKBP12 (see Section 2.4.11.7), BCL11a (Zn finger domains) (see Section 2.4.11.8), YAP1 (ww1-ww2 domains) (see Section 2.4.11.9) and NEMO_iZIP (coiled coil domain with iZip adapters (*30*)) (see Section 2.4.11.10). In all cases, we appended a biotin unit at the N-terminus of the protein through a PEG_{12} linker to facilitate phage display screening.

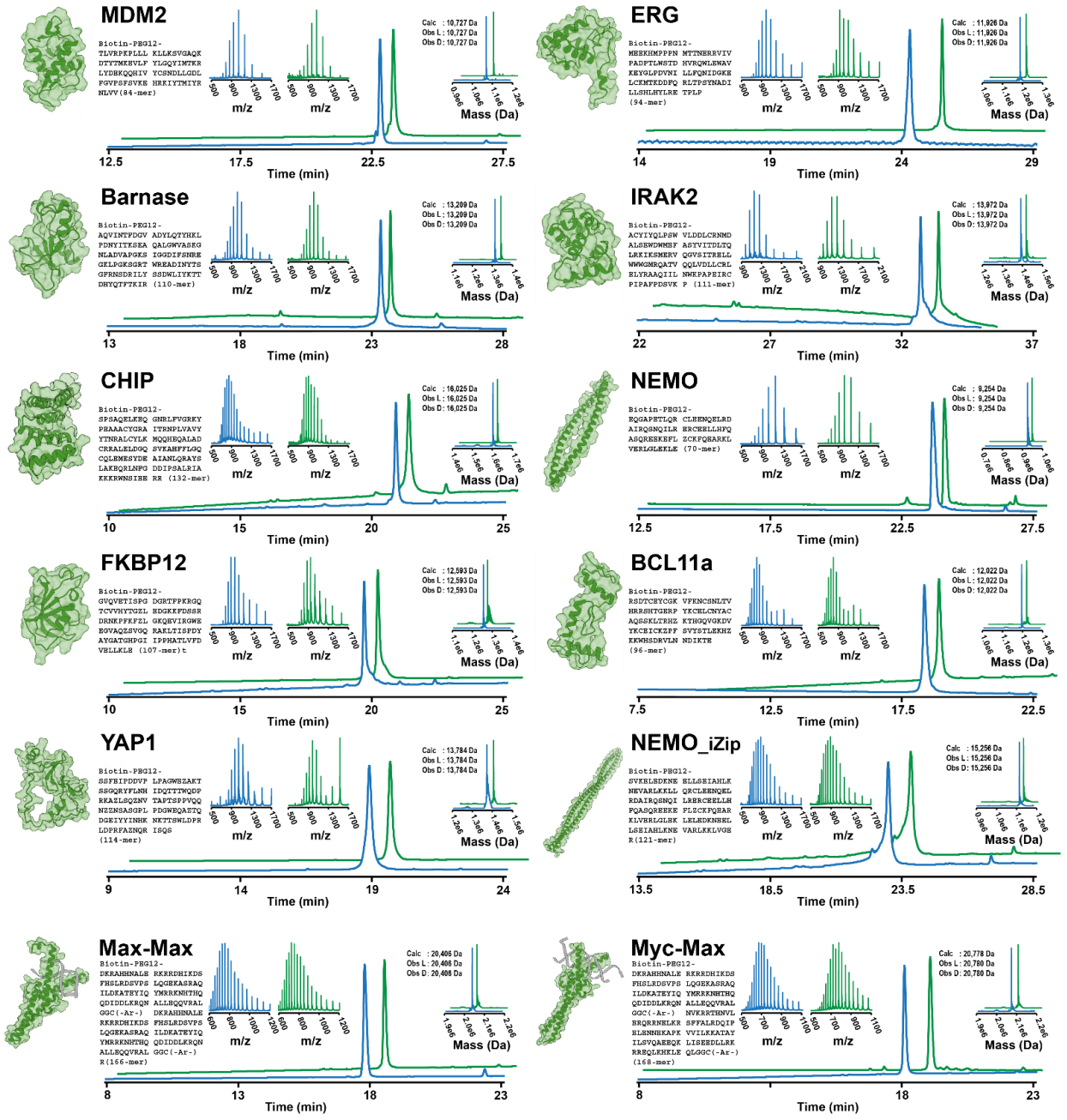


Figure 2.2 Automated flow synthesis delivers diverse single-domain protein chains in both enantiomeric forms.

Analytical characterization of reverse phase-purified L- and D-chains from AFPS. Green traces show characterization of L-proteins, and blue traces show characterization of D-proteins. For each protein target, the following data are shown: (1) analytical HPLC trace of purified material recorded at 214 nm (bottom overlayed chromatograms); (2) total ion current (TIC) post-injection on a Q-TOF LC-MS instrument (separate insets on the left); and (3) the deconvolution of the TIC traces of (2) shown as overlayed insets on the right. Observed masses from the deconvolution are shown along with the predicted values. Previously reported structures are depicted for each of the targets, sourced from prior X-

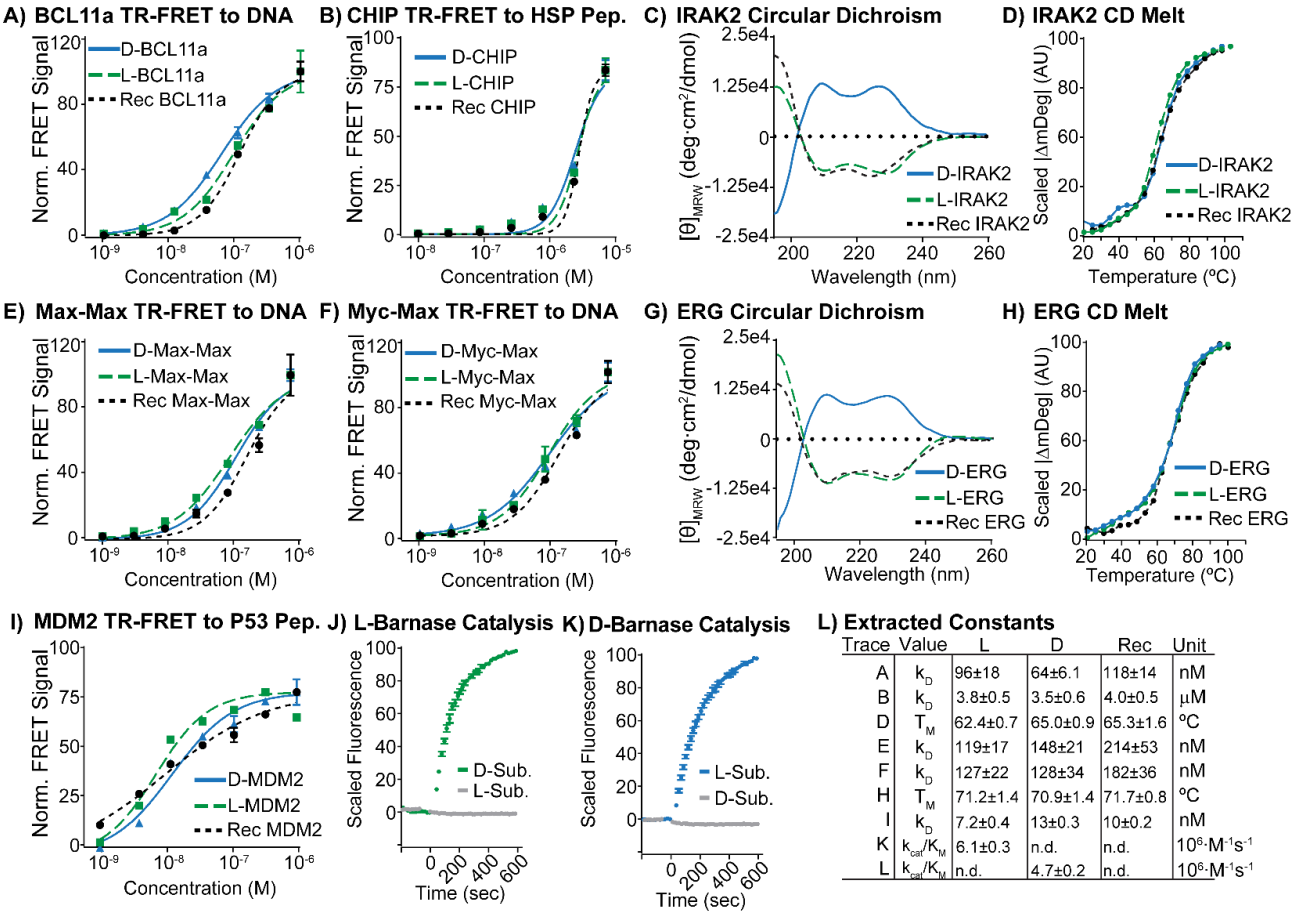
ray crystal structures, NMR structures, or Alphafold predictions (44, 45) (MDM2: 3FDO, ERG: 1SXE, Barnase: 1A2P, IRAK2: 3MOP CHIP: 4KBQ, NEMO: 3BRV, FKBP12: 2PPN, BCL11a: 6KI6, YAP1: Alphafold prediction of Uniprot ID P46937 region 63-276, NEMO_iZIP: 6MI3, Myc-Max: 1HLO, Max-Max: 1NKP).

We used AFPS to prepare each polypeptide chain in a stepwise fashion, with total synthesis times ranging from 4-7 hours, using three separate AFPS instruments. After we released the proteins from the H-Rink Amide solid support, we isolated them by precipitation with cold diethyl ether (see Sections 2.4.9 and 2.4.4). We used analytical reverse phase HPLC (RP-HPLC) and liquid chromatography–mass spectrometry (LC-MS) to analyze crude peptide mixtures and preparative mass-directed RP-HPLC to purify them. We typically obtained ~5 µmol of crude polypeptide powder from individual synthesis runs, and purifying ~3 µmol of this material afforded ~0.3 µmol of pure peptide. Gradient shape for the preparative purification was determined by a preliminary run at low protein loading on the same column used for purification (see Section 2.4.7). Typical yields are on the order of 1-10 mg (0.07–0.7 µmol, 0.3%–3% isolated yield based on resin loading). In all instances, a single synthesis experiment afforded sufficient material after folding for phage selections and biochemical validation.

2.2.2. Folding of diverse D-proteins with preparative size exclusion chromatography

Using a generalized folding protocol, each synthetic protein was folded to a homogeneous product as analyzed by size-exclusion chromatography (SEC). We set out to identify appropriate folding conditions for our 12 enantiomeric protein pairs and began by adopting the protocols that had been used to successfully isolate folded protein in aqueous solution for the few targets for which literature reports exist. Specifically, our group has previously isolated L/D Barnase (24) and L/D MDM2 (46) via stepwise dilution from a high concentration of guanidine hydrochloride. Subsequent purification using semi-preparative SEC achieves yields ranging from 4.5 to 10.6 nmol for all four targets. Here, using this technique, we were unable to identify any dilution conditions under which we could isolate folded BCL11a or IRAK2. Instead, we observed significant precipitation for both target pairs, with any remaining solubilized material eluting in the exclusion volume of the column, indicating an apparent molecular weight at least 10-fold over the expected value, consistent with the formation of soluble aggregates.

To address these complications, we adapted a refolding technique utilizing SEC to isolate the remaining protein targets. Purified proteins were dissolved in a denaturing buffer, typically a 10 mg / mL solution containing 6 M guanidine hydrochloride (Gdn HCl) buffered to pH 7.5 with 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) (HEPES) and including 50 mM dithiothreitol (DTT). The mixture was then submitted to semi-preparative SEC coupled to an HPLC instrument. The denaturant is removed as the protein progresses through the SEC column, and the resulting folded protein can be separated from off-target aggregates based on their different retention times. The running buffer of the SEC could be tuned to the requirements of each particular protein, but unless otherwise specified was 50 mM HEPES, 150 mM NaCl, 0.5 mM DTT, 5% glycerol (v/v), at pH 7.5. DTT was omitted in cases where the protein contained no free sulfhydryl groups (NEMO, NEMO_iZIP, Myc-Max, and Max-Max). For BCL11a, which contains three Zn-finger motifs, both the denaturing buffer and the running buffer were adjusted to contain $ZnCl_2$ with additional alterations to prevent precipitation of Zn salts. In each case, the folding protocol afforded material as major elution peaks with retention times consistent with the calculated molecular weights (Sections 2.4.13.1 to 2.4.13.12). This set includes BCL11a, and IRAK2 for which we were unable to isolate folded material by any other technique. Some proteins elute as peaks with apparent molecular weight larger than would be otherwise expected, including CHIP, NEMO, NEMO_iZIP, and BCL11a. In the case of NEMO, and NEMO_iZIP, this behavior has been established and has been attributed to their extended structures (43, 47).



L) Extracted Constants

Trace	Value	L	D	Rec	Unit
A	k_D	96±18	64±6.1	118±14	nM
B	k_D	3.8±0.5	3.5±0.6	4.0±0.5	μM
D	T_M	62.4±0.7	65.0±0.9	65.3±1.6	ºC
E	k_D	119±17	148±21	214±53	nM
F	k_D	127±22	128±34	182±36	nM
H	T_M	71.2±1.4	70.9±1.4	71.7±0.8	ºC
I	k_D	7.2±0.4	13±0.3	10±0.2	nM
K	k_{cat}/K_M	6.1±0.3	n.d.	n.d.	$10^6 \cdot M^{-1}s^{-1}$
L	k_{cat}/K_M	n.d.	4.7±0.2	n.d.	$10^6 \cdot M^{-1}s^{-1}$

Figure 2.3 Synthetic D-proteins display similar biological activity to their synthetic L- and recombinant counterparts.

Kinetic binding data for five synthetic protein targets to peptide ligands of the appropriate chirality are shown. Each of the shown D- and L-proteins display similar activity as their recombinant protein counterparts, and are consistent with literature reports of the same interactions. In all cases, experimental details are outlined in the Supporting Information. A) Synthetic D, L and recombinant BCL11a bind a model DNA oligonucleotide of the γ-globulin promoter(48) with similar affinity. The FAM-labeled oligonucleotide was incubated with varying concentrations of biotinylated BCCL11a and terbium-labeled streptavidin. The chirality of DNA substrate was adjusted to the chirality of the target (L-DNA to D-BCL11a, D-DNA to L-BCL11a). Binding was measured as an increase in FRET efficiency between the FAM and terbium streptavidin. B) Synthetic D, L and recombinant CHIP bind a peptide model of the heat shock protein 70 (Hsp70) C-terminus with similar affinity. The chirality of the peptide substrate was adjusted to the chirality of the target (D-peptide to D-CHIP, L-peptide to L-CHIP). The FAM-labeled HSP70 peptide was incubated with varying concentrations of biotinylated CHIP and terbium-labeled streptavidin, and binding was measured as described in A. C) Synthetic D, L and recombinant IRAK2 display similar proportions of secondary structure by near circular dichroism (CD). CD spectra were recorded from 195 to 260 nm over a 0.1 cm pathlength at a protein concentration of 0.1 mg/mL. D) Synthetic L, D and recombinant IRAK2 have similar melting temperatures (Tm) as determined by variable temperature CD. The absolute

intensity of the CD spectrum at 222 nm of each protein was monitored from 20 °C to 100 °C in 5 °C steps. E) Synthetic D, L and recombinant Max-Max bind a model DNA oligonucleotide of E-box DNA (42) with similar affinity. The FAM labeled DNA was incubated with varying concentrations of biotinylated Max-Max, and terbium-labeled streptavidin. Binding was measured as described in A. The chirality of DNA substrate was adjusted to the chirality of the target (L-DNA to D-Max-Max, D-DNA to L-Max-Max). F) Synthetic D, L and recombinant Myc-Max bind a model DNA oligonucleotide of E-box DNA(42) with similar affinity. The FAM labeled DNA was incubated with varying concentrations of biotinylated Myc-Max and terbium-labeled streptavidin. Binding was measured as described in A. The chirality of DNA substrate was adjusted to the chirality of the target (L-DNA to D-Myc-Max, D-DNA to L-Myc-Max). G) Synthetic D, L and recombinant ERG display similar proportions of secondary structural features by near circular dichroism (CD). CD spectra were recorded as described in C. H) Synthetic L, D and recombinant ERG have similar melting temperatures determined as described in D. I) Synthetic D, L and recombinant MDM2 bind a peptide model of the p53 protein with similar affinity. The FAM labeled p53 peptide was incubated with varying concentrations of biotinylated MDM2 and terbium-labeled streptavidin. Binding was measured as described in A. The chirality of peptide substrate was adjusted to the chirality of the target (D-peptide to D-MDM2, L-peptide to L-MDM2). J) and K) Synthetic L- and D-barnase, respectively, selectively catalyze the hydrolysis of a stereochemically matched RNA substrate. An RNA substrate labeled with FAM and TAMRA separated by a barnase cleavage site was prepared in both the L- and D-forms. Each substrate was incubated with L-barnase (J) and D-barnase (K), and catalytic activity was measured as an increase in fluorescence intensity after barnase addition. No cleavage for the mismatched substrate pairs (L-RNA to L-barnase and D-RNA to D-barnase) was observed. L) Table summarizing the experimentally determined binding constants, melting temperatures, and kinetic constants for the reported synthetic proteins (n.d. = not determined).

2.2.3. Synthetic D- and L-proteins display correct biochemical activity

For each enantiomeric pair generated using AFPS, we used biochemical assays to confirm that the activities of the synthetic proteins are similar to their recombinant versions. For proteins with known high-affinity binding partners, we measured the affinities of the synthetic proteins to fluorophore-tagged probes of the appropriate chirality using time-resolved fluorescence resonance energy transfer (TR-FRET). Binder substrates were modified with 5(6)-carboxyfluorescein (FAM), and the biotinylated target proteins were complexed with terbium-labeled streptavidin. Binding was measured as an increase in fluorescence quenching between the two fluorophores. For BCL11a binding to DNA (Fig. 2.3A and Section 2.4.14.1), CHIP binding to HSP peptide (Fig. 2.3B and Section 2.4.14.2), MDM2 binding to p53 peptide (Fig. 2.3I and Section 2.4.14.3), Max-Max binding to E-Box DNA (Fig. 2.3E and Section 2.4.14.4), and Myc-Max binding to E-Box DNA (Fig. 2.3F and Section 2.4.14.4), each synthetic chiral protein displayed

apparent dissociation constants (KD) for the L- and D-proteins to be consistent with those of the recombinant proteins (Fig. 2.3M). We next used surface plasmon resonance (SPR) binding assays with substrates of the matching chirality to asses binding of NEMO to iKKb peptide (see Section 2.4.14.7), NEMO_iZip binding to iKKb peptide (see Section 2.4.14.10), YAP1 binding to dendrin (see Section 2.4.14.9), and L-FKBP12 binding to Rapamycin followed by mTOR (see Section 2.4.14.8). Again, in all instances, the synthetic proteins displayed binding affinities similar to the recombinantly derived proteins, suggesting that the synthetic proteins derived from AFPS are folding into bioactive tertiary structures.

Synthetic L/D barnase from AFPS displays ribonucleolytic activity that is selective for the chirality of its nucleotide substrate. Barnase catalytically cleaves a fluorescently quenched reporter RNA substrate (49) composed of a tetranucleotide flanked with FAM and 6-carboxytetramethylrhodamine (6-TAMRA) labels on either end (FAM-DdA-DrU-DdA-DdA-TAMRA), where each nucleotide is specified as the L- or D-enantiomer and the ribose (r) or deoxyribose (d) sugar. We tracked the barnase-mediated cleavage of substrate by assessing the FRET between the FAM and TAMRA fluorophores, which is abolished upon cleavage. When the chirality of the substrate and protein were matched, as for L-barnase to the D-substrate FAM-DdA-DrU-DdA-DdA–TAMRA, and D-barnase to the L-substrate FAM-LdA-LrU-LdA-LdA–TAMRA, we observed rapid hydrolysis with extracted kinetic constants that are similar to previously reported values (Fig. 2.3 J, K) (41). In the mismatched cases, L-barnase to L-substrate and D-barnase to D-substrate, we observed no catalysis on the timescale of our observation (Fig. 2.3 J, K) (see Section 2.4.14.5). These results reconfirm the structural and biological integrity of the folded synthetic proteins, indicating that their enantiomeric purities are retained throughout the synthesis and purification processes.

For protein targets with no known binding partners and no available assays to assess enzymatic activity exists (ERG and IRAK2), we recorded circular dichroism (CD) spectra and compared them to recombinantly derived material. Both targets are reported to be mostly alpha-helical (50, 51), and we could confirm this for the recombinant proteins. The CD spectra for the synthetically derived L-proteins closely matched those of the recombinant versions, indicating that the synthetic material is forming secondary

structures in similar proportions (Fig. 2.3 C, G) (see Section 2.4.14.6). Both D-ERG and D-IRAK2 display a CD signal with similar absolute intensities as the corresponding synthetic L- and recombinant proteins, but with inverted sign, consistent with the formation of mirror-image secondary structures. Furthermore, we found that both synthetic protein pairs form folded structures with melting temperatures (Tm) close to those of the recombinantly derived protein, as shown by tracking the intensity of the α-helical signature as a function of temperature (Fig. 2.3 D, H). Together, these data suggest that the synthetically derived L/D ERG and L/D IRAK2 form tertiary structures that closely resemble the native proteins.

2.2.4. Generation of macrocyclic D-peptide ligands to MDM2 using mirror-image phage display

Having validated the structure and function of our synthetic D-proteins, we sought to benchmark our MIPD screening platform with a model target, mouse double minute 2 homolog (MDM2). MDM2 is an E3 ubiquitin ligase that recognizes the a FXXXWXXL motif present on its substrate, p53, a critical mediator of cell cycle arrest, senescence and apoptosis (52). As a result, p53 degradation driven by upregulation of MDM2 is found to accelerate growth in a variety of cancers (53). To disrupt p53 recognition, and thus cancer progression, a number of high-affinity MDM2 binders have been developed, and evaluated in the clinic (54). In particular, highly-potent biologically stable D-peptides to MDM2 have been generated with mirror image phage display (10). We set out to investigate if our MIPD platform could rediscover similar motifs, and perhaps generate new candidates for MDM2 inhibition. To that end, we successfully isolated synthetic mirror image MDM2, and carried it into our next-generation MIPD platform (40).

We identified three binding clusters selective for D-MDM2 with phage display screening (Fig. 2.4 A and Section 2.4.15.1). In brief, immobilized D-MDM2 from AFPS was screened against an unbiased phage library of 108 members displaying macrocyclic α-helical peptides. We sequenced phage particles that remained bound to the targets after washing using next-generation sequencing, and comparisons of sequencing reads to a spiked-in internal reference allowed for a semi-quantitative read-out of binding affinity as a function of protein bait concentration (see Section 2.4.8). The resulting hit sequences that displayed proper dose-response were clustered into binder families using hierarchical

statistical clustering. We identified three such clusters to D-MDM2, and Logo plots of their sequence overlaps are shown in Figure 2.4A (MDM2.C1-MDM2.C3). Consistent with previously described α-helical peptide binders to MDM2 (55, 56), all three clusters contain anchoring N-terminal tryptophan residues (dW2 in MDM2.C1, dW1 in MDM2.C2, and dW1 in MDM2.C3), and show conserved hydrophobic residues on the same α-helical face as the tryptophan (i,i+4/5 and i,i+8/9 relative to dW on each cluster). To validate the specificity of each binding cluster, we synthesized representative binding peptides in the D-form (Fig 2.4B, D-H101 from MDM2.C1, D-H102 from MDM2.C2, and D-H103 from MDM2.C3) and measured their affinity to recombinant MDM2 with SPR. We observed binding constants of 5.5 µM, 2.4 µM, and 0.88 µM, respectively (Fig. 2.4C and Section 2.4.15.2).

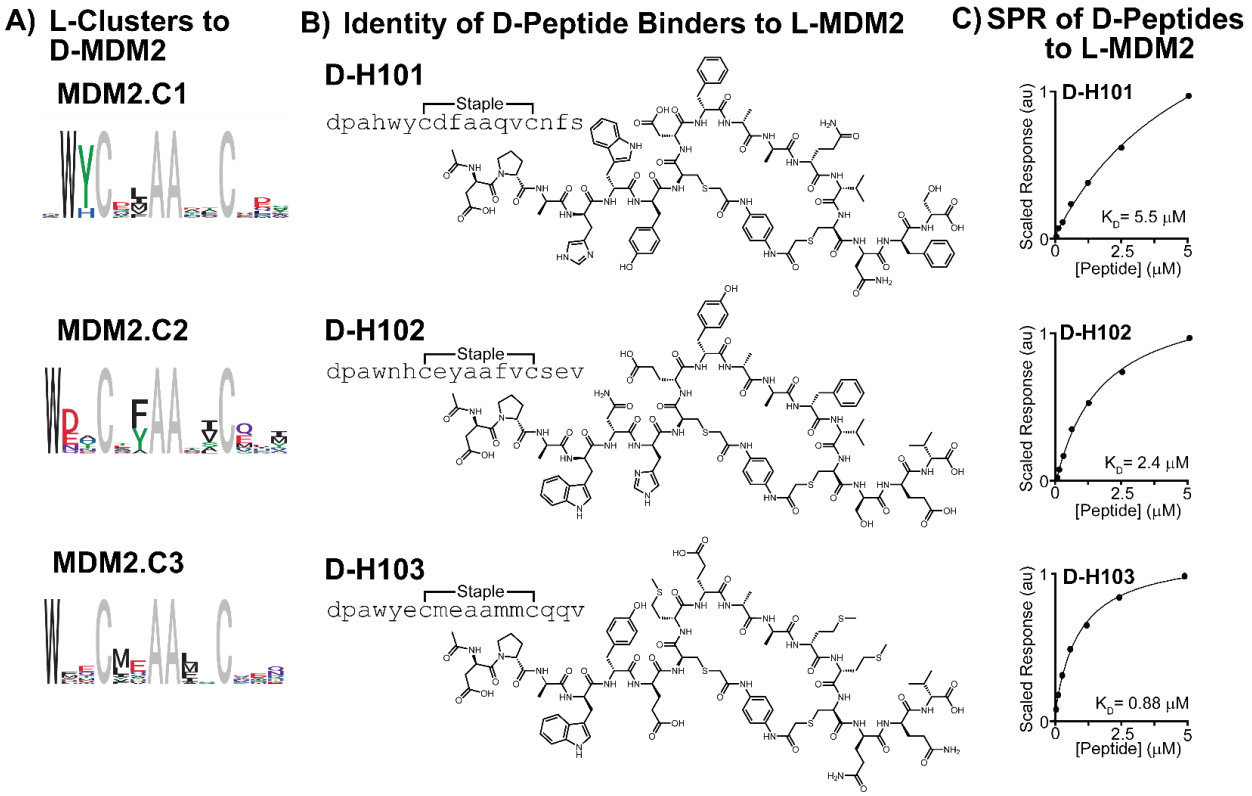


Figure.4 Mirror-image phage display of D-MDM2 generates novel macrocyclic D-peptide binders to L-MDM2.

Logo plots of the three clusters that bind to MDM2 are shown with fixed residues in grey. B) Identities and chemical structures of selected D macrocycles are shown. Selected peptide sequences are members of the cluster to their left. C) The discovered macrocyclic D-peptides bind recombinant MDM2. Steady-state SPR sensorgrams of the D-binders to D-MDM2 are shown with extracted binding affinities.

To further investigate the binding modes of each binding cluster, we solved the X-ray co-crystal structures of D-H101 (Fig. 2.5 A, 1.64 Å), D-H102 (Fig. 2.5 B, 1.30 Å), and D-H103 (Fig. 2.5 C, 1.90 Å) with recombinant MDM2 (see Sections 2.4.17.1 to 2.4.17.4). All three peptides form left-handed α-helices that engage the same hydrophobic groove on MDM2, but make use of different side chain interactions. In particular, D-H101 and D-H102 use similar amino acids to interact with MDM2 at its N-terminus (dW5/dW4, dY6/dH6, and dF9/dY9), but diverge towards the C-terminus where the side-chain conformations are altered (dQ12/dF12 and dE16/dF16) (Fig. 2.5 A, B). Beyond the conserved N-terminal aromatic residues, D-H103 makes use of altered side-chain to interact with MDM2 compared to either D-H101 or D-H102. (Fig. 2.5 C vs Fig. 2.5 A-B). Consistent with their sequence similarities, D-H101 utilizes similar side-chain interactions to two previously reported α-helical peptide binders to MDM2. Overlay of the known D-peptide MDM2 binder dPMI-δ[1-5, 9-12] (55) onto the D-H101 MDM2 co-crystal structure revealed that both peptides present into the same groove on MDM2 with the same axial direction (N->C), and make use of similar side chains (Figure 2.5 D). Chemically related side chains project with similar orientations for both peptides, perhaps due to the peptides' matched left-handed helices. Interestingly, matching helicity is not an exclusive requirement to make use of similar side chain interactions though. ATSP-7041 is a high-affinity L-peptide binder to MDM2 (56), and Figure 2.5E shows its structure overlaid onto the D-H101 MDM2 co-crystal structure. Despite ATSP-7041 forming a right-handed alpha helix, both peptides appear to use similar binding modes. Important side chain interactions of ATSP-7041 are closely overlapped with equivalent interactions for D-H101 (Fig. 2.5 C, dW5/F3, dY6/Y6, dF9/W7, and dV13/Cba10).

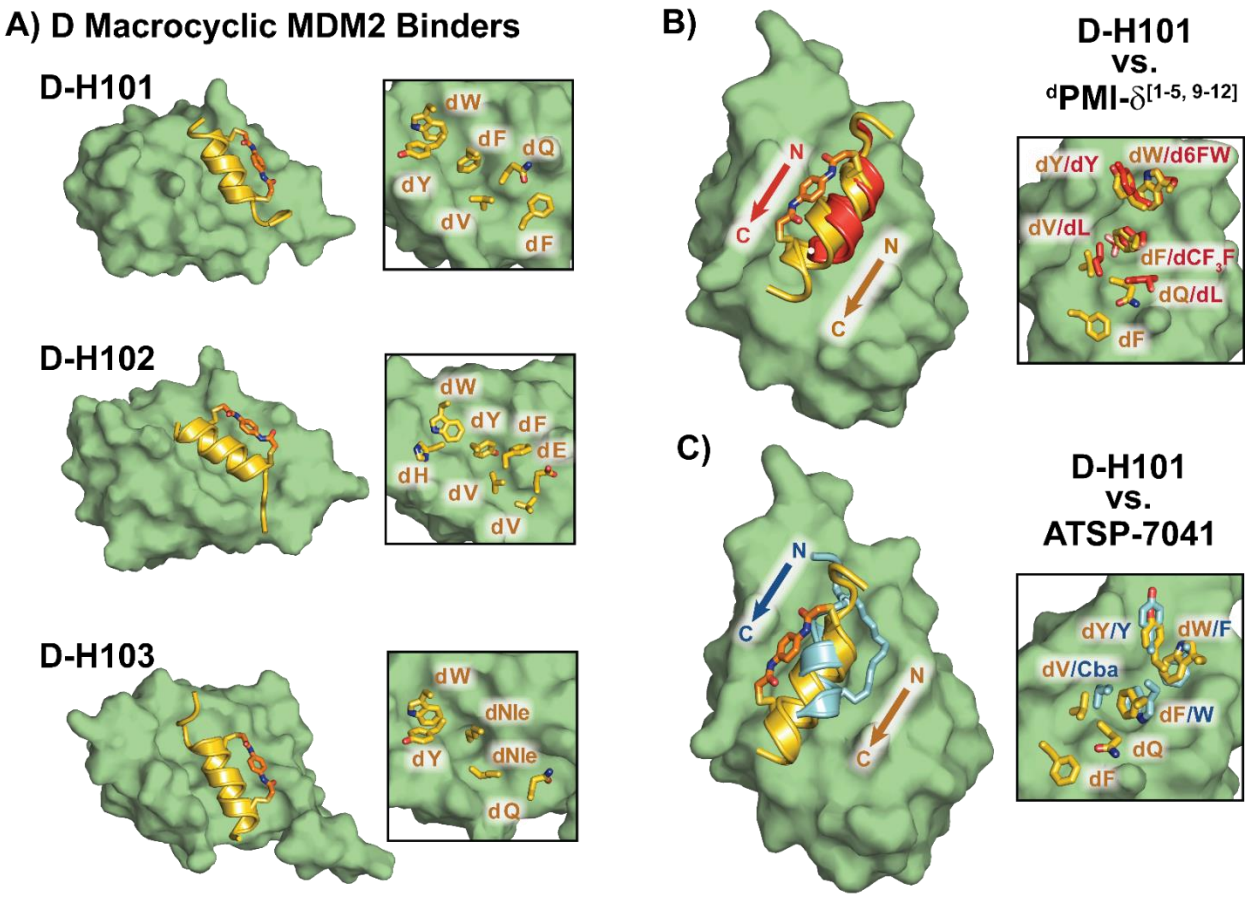


Figure 2.5 Discovered D macrocycles bind to a common hydrophobic groove on MDM2.

A) X-ray co-crystal structures of macrocyclic D-peptides from each cluster to L-MDM2 are shown. Side-chains projecting into the MDM2 surface are highlighted. B) The co-crystal structure of D-H101 with MDM2 is shown overlaid with known D-peptide binders dPMI-δ (18, 55) (PDB ID: 6KZU). Side-chains that project into the MDM2 surface are highlighted in the inset. C) The co-crystal structure of D-H101 with MDM2 is shown overlaid with known L-peptide binder and ATSP-7041 (56) (PDB ID: 4N5T). Side-chains that project into the MDM2 surface are highlighted in the inset.

2.2.5. Generation of macrocyclic D-peptide ligands to CHIP using mirror-image phage display

Not limited to known D-proteins, we nominated a novel target for MIPD screening, CHIP, and identified three macrocyclic D-peptide binders. CHIP is a member of the RING/U-Box family of E3 ligases that direct the ubiquitination of chaperones (57) and their bound client proteins (58, 59) (Fig. 2.6 A). This interaction is driven by a common (I/M)EEVD motif present on the C-terminus of heat shock protein 70 (Hsp70) and Hsp90 family members that is recognized by the TPR domain of CHIP(60) (Fig. 2.6 B). Degradation driven by CHIP is active in a diverse range of cellular processes from

monitoring protein quality control (59) to mediating interferon-γ signaling (61). Masking of the CHIP-TPR domain can in theory modulate these signaling pathways and as a result is an actively being pursued (60, 62). We hypothesized that this domain would be a valuable target for D-peptide discovery, and successfully isolated the protein in its mirror image form for MIPD screening.

From a single-round phage screen, we identified three clusters that displayed robust binding to CHIP (CHIP.C1 – CHIP.C3) (Section 2.4.15.3). All three clusters feature conserved terminal residues with carboxylate functionalities (dD3 in CHIP.C1, dD1/dE2 in CHIP.C2, and dD12 in CHIP.C3), consistent with the conserved terminal aspartic acid residues found on CHIP substrates. Hydrophobic residues are retained in the i,i+4/5 positions relative to the anchoring aspartic acid residue: dW6 in CHIP.C1, dW6 in CHIP.C2, and dA8 in CHIP.C3, presumably for interaction with the hydrophobic TPR groove. We validated binding of selected sequences from all three clusters to the recombinant CHIP-TRP domain (IC50 values of 0.47 μM, 0.55 μM, and 1.23 μM, respectively, Fig. 2.6 C) in a competition fluorescence polarization assay against FAM-Hsp70 loaded CHIP (Section 2.4.15.4).

X-ray co-crystal structures of each discovered D-macrocycle with recombinant CHIP revealed a common binding pocket despite variable side chain interactions. We solved the X-ray co-crystal structures of each peptide with recombinant CHIP (Fig. 2.6 C, 1.72 Å, 1.59 Å, and 1.76 Å for D-H201, D-H202, and D-H203 respectively) (see Sections 2.4.17.5 to 2.4.17.8). As predicted, all three peptides form left-handed α-helices to engage the same groove, though D-H202 binds in the reverse direction compared to D-201 and D-203 (Fig. 2.6 C). Despite varying residues, side chains projecting into the CHIP surface across all binders have significant spatial overlap (Fig. 2.6 D). Two loci for hydrophobic interactions on the CHIP surface in particular show strong side-chain overlap: dW7/dL6/dW7, and dF8/Staple/dM4 (of D-H201, D-H202, and D-H203, respectively). Additionally, the common terminal carboxylate-containing residues engage in similar hydrogen bonding networks with CHIP. These similarities extend to known biological ligands to CHIP, and an overlay of a co-crystal structure of the C-terminus of Hsc70 onto D-H201 is shown in Figure 6E. Biological ligands to the CHIP TPR domain are known to form distinct kinked backbone conformations (60), unlike the left-handed alpha-helices of

the discovered macrocycles. Despite this feature, the biological ligand utilizes similar side chain contacts to engage CHIP. Both hydrophobic side chain sites found in the discovered macrocycles (dF8/V4 and dW7/I1 in D-H201 and Hsc70, respectively) and the terminal carboxylate are also found in the biological CHIP ligand. These similarities indicate that the discovered macrocyclic D-peptides utilize a native binding mode translated into a novel left-handed alpha-helical interface.

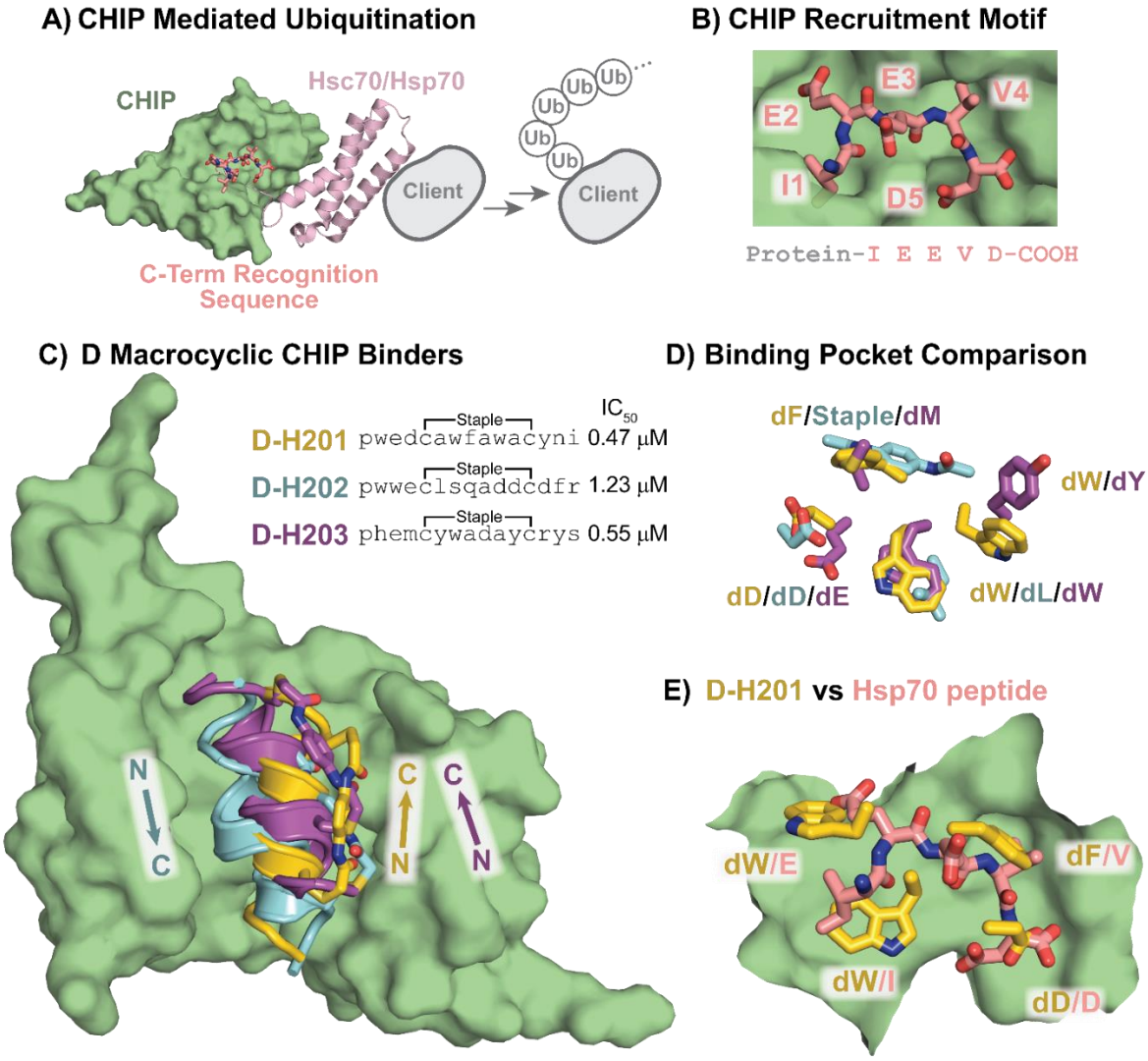


Figure 2.6 Novel D-macrocyclic peptides identified from MIPD are high affinity ligands to CHIP.

A) The E3 ubiquitin ligase CHIP marks chaperone bound client proteins for degradation through recognition of a C-terminal peptide motif. (PDB ID: 4KBQ) B) The TPR domain of CHIP recognizes a C-terminal -IEEVD- motif. C) Mirror-image phage display of the TPR domain of CHIP affords three D macrocyclic peptide binders to a common groove. X-ray co-crystal structures of the three identified sequences are shown projected onto the CHIP surface from D-H202, with annotated helical directions. D-H201 is shown in yellow, D-

H202 in blue, and D-H203 in purple. Binding affinities were determined from competition FP of each peptide to FAM-Hsp70-peptide bound recombinant CHIP. D) All three binders utilize different chemical contacts to engage the same pockets on CHIP, despite different helical directions. Interacting side chains from the X-ray co-crystal structures of each binder are shown overlaid. Side chain colorings are as in C. E) The discovered binders utilize unique side chains to engage contact sites of the native peptide substrate. The X-ray co-crystal structure of D-H201 is shown overlaid with the Hsp70 recognition peptide (PDB ID: 4KBQ). Overlapping side chain interactions are highlighted, with D-H201 in yellow and Hsp70 peptide in pink

2.3.Discussion

Peptide scaffolds are increasingly more often applied to new therapeutic targets, and D-peptides in particular have great potential to be broadly applicable. Native, canonical peptides are notoriously unstable in vivo, and this reputation has prevented their application as therapeutics (63). Increasingly effective and easily accessible chemical stabilization strategies have revitalized peptides as effective therapeutic scaffolds (6). With judicious optimization, newly reported peptide therapeutics routinely display long biological half-lives. For example, after optimization, the candidate PCSK9 inhibitor MK-0616 displays nearly compete stability against key biological proteases (64), and as a result is dosed orally in its Phase 2b clinical trial (65). However, such achievements still require a complex and meticulous drug development process. Mirror-image D-peptides on the other hand should display comparable metabolic stabilities per se. Routine implementation of screening strategies that afford D-peptide ligands stands to greatly simplify and accelerate peptide drug development.

We demonstrate for the first time a robust and scalable platform for D-peptide discovery using mirror-image phage display facilitated by automated flow synthesis technology. Among the main obstacles to the broad implementation of MIPD are the inconsistent and unpredictable outcomes of the total chemical synthesis of the requisite mirror-image proteins and their fragments. In this work, we demonstrate that AFPS technology can supply mirror-image proteins for MIPD in a single-shot with unmatched reliability and speed. We successfully isolated 12 mirror-image proteins with AFPS, 8 of which have no previously reported syntheses. Without AFPS, these efforts would likely require years of individualized synthesis route optimization and method troubleshooting. In our work, these mirror-image proteins were instead rapidly prepared for screening in a

standardized format. Continued improvement of D-protein synthesis quality is key to making MIPD practical for modern screening pipelines.

With practical and reliable access to mirror-image proteins, we describe the discovery of six novel macrocyclic D-peptide ligands to two protein targets. We first validated our discovery platform with screening of MDM2, a negative regulator of p53 known to be overexpressed in a variety of cancers (53). MDM2 is one of a few previously reported targets for mirror-image phage display (10), and we successfully rediscovered sequence motifs that are similar to these earlier reports. Furthermore, we recorded X-ray co-crystal structures of three novel mirror image binders to MDM2. Collectively, these results confirm the fidelity of our synthesis approach, and in addition provide new therapeutic leads for targeting MDM2. We further nominated CHIP for mirror image phage display screening, and again identified three novel macrocyclic D-peptide ligands. CHIP is an E3 ubiquitin ligase that directs the degradation of cellular chaperones and their bound client proteins (57). Inhibition of the CHIP ligand binding site is an emerging approach to modulate interferon-γ signaling (61). We anticipated that mirror-image peptide ligands could prove a valuable tool to engage this site. The three discovered peptide ligands all bound to recombinant CHIP and were able to displace a native CHIP ligand at low micromolar concentrations. Characterization of each mirror-image ligand with X-ray co-crystal structures to recombinant CHIP revealed a common binding site with native CHIP ligands, indicating their potential to disrupt this interaction. Collectively, the results reported here represent a significant fraction of the reported successful mirror-image phage display screens. Continued advances in the technologies that enable practical access to MIPD stand to revitalize campaigns to generate mirror-image binders to new and emerging protein targets.

2.4.Methods

2.4.1. General Information

Unless otherwise specified, all reactions with polypeptides, proteins, and protein oxidative addition complexes were set up on the bench top and carried out under ambient conditions. Unless stated otherwise, all small (≤ 1 mL) volumes were measured out using Eppendorf Research® plus, single-channel, variable, mechanical pipettes (referred to as mechanical pipettes). Universal low retention pipet tips (10 µL, 200 µL and 1000 µL sizes) were purchased from VWR International (Philadelphia, PA). Care was taken to use the appropriate mechanical pipette/tip combinations to ensure a dispensing error of ≤ 2% for volumes between 10 µL and 1000 µL and ≤ 4% for volumes between 1 µL and 10 µL. To avoid loss of protein due to non-specific adsorption, plastic tubes (Eppendorf Protein LoBind® tubes, 0.5, 1.5, and 2.0 mL) were used in all cases after folding. The weight of lyophilized powders of the peptides was directly measured using analytical scales (XS205DU Analytical Balance, Mettler-Toledo) with fixed with a SPI Westek Workstation Still Air Ionizer. All Fmoc-protected D-amino acids were lyophilized for at least 12 hours before use to remove trace solvents.

2.4.2. Materials for Peptide Synthesis

All reagents were purchased and used as received unless otherwise noted. Fmoc-protected L-amino acids (Fmoc-Ala-OH x H2O, Fmoc-Cys(Trt)-OH, Fmoc-Asp(OtBu)-OH, Fmoc-Glu(OtBu)-OH, Fmoc-Phe-OH, Fmoc-Gly-OH, Fmoc-Ile-OH, Fmoc-Lys(Boc)-OH, Fmoc-Leu-OH, Fmoc-Met-OH, Fmoc-Asn(Trt)-OH, Fmoc-Pro-OH, Fmoc-Gln(Trt)-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Ser(tBu)-OH, Fmoc-Thr(tBu), Fmoc-Val-OH, Fmoc-Trp(Boc)-OH, and Fmoc-Tyr(OtBu)-OH) were purchased from the Novabiochem line (Millipore-Sigma). L-Fmoc-His(Boc)-OH was purchased from Chem-Pep. L-Fmoc-Nle-OH, was purchased from Chem-Impex. Fmoc-protected D-amino acids (Fmoc-Ala-OH x H2O, Fmoc-Cys(Trt)-OH, Fmoc-Asp(OtBu)-OH, Fmoc-Glu(OtBu)-OH, Fmoc-Phe-OH, Fmoc-His(Boc)-OH, Fmoc-Ile-OH, Fmoc-Lys(Boc)-OH, Fmoc-Leu-OH, Fmoc-Met-OH, Fmoc-Asn(Trt)-OH, Fmoc-Pro-OH, Fmoc-Gln(Trt)-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Ser(tBu)-OH, Fmoc-Thr(tBu), Fmoc-Val-OH, Fmoc-Trp(Boc)-OH, and Fmoc-Tyr(OtBu)-OH) were purchased from both the Novabiochem-line from Millipore-Sigma and Chem-Impex. Fmoc-D-Nle-OH was purchased from Chem-Impex. O-(7-azabenzotriazol-1-yl)-

N,N,N',N'-tetramethyluronium hexafluorophosphate (HATU, ≥97.0%), (7-azabenzotriazol-1-yloxy) tripyrrolidinophosphonium hexafluorophosphate (PyAOP, ≥97.0%) were purchased from P3 Biosystems. Biosynthesis OmniSolv® grade N,N-dimethylformamide (DMF) was purchased from EMD Millipore, and stored over AldraAmine trapping agents (for 1000 – 4000 mL DMF, catalog number Z511706). Diisopropylethylamine (DIEA; 99.5%, biotech grade, catalog number 387649) was purchased from Sigma-Aldrich and purified by passage through an activated alumina column (Pure Process Technology solvent purification system). Piperidine (ACS reagent, ≥99.0%), trifluoroacetic acid (TFA, HPLC grade, ≥99.0%), triisopropylsilane (TIPS, ≥98.0%), acetonitrile (HPLC grade), formic acid (FA, ≥95.0%), phenol (ACS reagent, ≥99.0%), diethyl ether (Et2O, ≥99.7%, containing 1 ppm BHT as inhibitor), and 1,2-ethanedithiol (EDT, GC grade, ≥98.0%) were purchased from Sigma-Aldrich. H-Rink Amide (0.49 mmol/g and 0.18 mmol/g loading) resin was 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. 5 mL and 10 mL peptide synthesis reaction vessels were purchased from Torviq (catalog numbers SF-0500, and SF-1000 respectively). Biotin-PEG12-COOH (98%) was purchased from BroadPharm. Syringe tip caps were purchased from VWR (catalog number 97001-202).

2.4.3. Materials for Protein Folding

Guanidine HCl (Gdn HCl, molecular biology grade, ≥99%), 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid, N-(2-hydroxyethyl)piperazine-N′-(2-ethanesulfonic acid) (HEPES, ≥99.5%), tris(hydroxymethyl)aminomethane (TRIS, ≥99.9%), tris(hydroxymethyl)aminomethane hydrochloride (TRIS HCl, ≥99.0%), monopotassium phosphate (KPi, ≥99.0%), dipotassium phosphate (≥ 99.0%), sodium chloride (NaCl, BioXtra, ≥99.5%), glycerol (≥99.5%), sodium hydroxide (≥98%), and hydrochloric acid (36.5-38%) were purchased from Sigma Aldrich. A 0.5 M solution of tris(2-carboxyethyl)phosphine hydrochloride (TCEP, Bond-Breaker™, catalogue number 77720), was purchased from Thermo Fisher Scientific. 1,4-Dithio-DL-threitol (DTT, ≥99%) was purchased from Chem-Impex.

2.4.4. Cleavage Protocols

Following synthesis, peptidyl resins were washed with 5 volumes of dichloromethane (5 x 5 mL) and dried under an N2 stream. The peptidyl resin was transferred to a 10 mL reaction vessel and stored at 4 °C until cleavage. Special care was taken with peptidyl resins for sequences containing methionine to avoid prolonged storage.

Method 1 – Reagent K

A fresh solution of a cleavage mixture was prepared according to the following recipe and used immediately: trifluoroacetic acid (82.5% v/v), water (5% v/v), phenol (5% v/v, melted in 60 °C water bath prior to use), thioanisole (5% v/v), and 1,2-ethanedithiol (2.5% v/v). A ratio of 1 mL of cleavage mixture per 2.5 µmol of protected peptide was used. The cleavage mixture was drawn-up with the reaction syringe containing the peptidyl resin, and the assembly capped with a luer-tip syringe tip cap. The assembly was placed on a nutating mixer and allowed to stand for 3.5 h. During this time, an aliquot of 50 mL of diethyl ether was prepared in a 50 mL Falcon tube and chilled on dry ice for at least 1 h. At the end of the incubation time, the cleavage supernatant in the reaction syringe was expelled into a fresh 50 mL Falcon tube, and the resin was retained by the internal frit of the reaction syringe. An additional aliquot of 2.5 mL of TFA was drawn up into the reaction syringe, the assembly gently mixed, and the aliquot expelled into the same Falcon tube containing the cleavage supernatant. The pre-chilled diethyl ether was added to the cleavage supernatant, and a strong white precipitate was generated. A fresh screw-cap was used to seal the Falcon tube containing the cleavage mixture, and then the assembly was centrifuged at 4 °C for 4 min at 3220 rcf while maintaining the temperature at 4 °C. The supernatant was decanted into the waste, and the solid pellet was retained for processing.

Method 2 – Standard Cleavage Mix

A fresh solution of a cleavage mixture was prepared according to the following recipe and used immediately: trifluoroacetic acid (94% v/v), water (2.5% v/v), 1,2-ethanedithiol (2.5% v/v), and triisopropylsilane (1% v/v). A ratio of 1 mL of cleavage mixture per 2.5 µmol of protected peptide was used. The cleavage mixture was drawn-up with the reaction syringe containing the peptidyl resin, and the assembly capped with

a luer-tip syringe tip cap. The assembly was placed on a nutating mixer and let stand for 2 h. During this time, an aliquot of 50 mL of diethyl ether was prepared in a 50 mL Falcon tube and chilled on dry ice for at least 1 h. At the end of the incubation time, the cleavage supernatant in the reaction syringe was expelled into a fresh 50 mL Falcon tube, and the resin was retained by the internal frit of the reaction syringe. An additional aliquot of 2.5 mL of trifluoroacetic acid was drawn up into the reaction syringe, the assembly gently mixed, and the aliquot expelled into the same Falcon tube containing the cleavage supernatant. The pre-chilled diethyl ether was added to the cleavage supernatant, and a strong white precipitate was generated. A fresh screw-cap was used to seal the Falcon tube containing the cleavage mixture, and then the assembly was centrifuged at 4 °C for 4 min at 3220 rcf while maintaining the temperature at 4 °C. The supernatant was decanted into the waste, and the solid pellet retained for processing.

After cleavage, the solid pellet was dried gently with N2 flow until no visible liquid remained. To the Falcon tube containing the dried peptide precipitate was added a volume equal to 3 times that of the precipitate of a 50:50 mixture of water/acetonitrile containing 0.1% TFA. To aid solubilization of the precipitate, the assembly was re-capped and vortexed. In most cases, the precipitate was fully solubilized. The mixture was left to incubate for 30 min at room temperature. The mixture was then flash-frozen with liquid nitrogen, and lyophilized for at least 16 h to provide the freeze-dried crude polypeptide as a white solid.

2.4.5. Analytical Reverse Phase High Pressure Liquid Chromatography Methods

Analysis of synthetic peptides was primarily carried out by evaluation of the chromatogram produced from the absorbance at 214 nm after separation on a reverse phase column. In each case, an appropriate amount of material was injected to produce peaks on the chromatogram that were within the dynamic range of the UV detector - typically between 100 and 2000 mAU (3-5 µg of the major product) for the major peak.

Method 1 – 5% to 65% in 30 min on Agilent 1200 HPLC

Analytical HPLC was carried out on an Agilent 1200 series system with UV detection at 214 nm. Column: Phenomenex Kinetex C18, (150 x 2.1 mm, 1.8 µm, 100 Å silica); flow rate 0.375 mL/minute; Solvent System: A = water with 0.1% TFA (v/v), B =

acetonitrile with 0.08% TFA (v/v); Gradient: 3 min hold 5% B, 5-65% B gradient over 30 min, 3 min hold 65% B, 10 min post run 5% B.

Method 2 - 5% to 65% in 15 min on Agilent 1200 HPLC

Analytical HPLC was carried out on an Agilent 1200 series system with UV detection at 214 nm. Column: Phenomenex Kinetex C18, (150 x 2.1 mm, 1.8 µm, 100 Å silica); flow rate 0.375 mL/minute; Solvent System: A = water with 0.1% TFA (v/v), B = acetonitrile with 0.08% TFA (v/v); Gradient: 3 min hold 5% B, 5-65% B gradient over 15 min, 3 min hold 65% B, 10 min post run 5% B.

Method 3 – 5% to 65% in 30 min on Agilent 1290 UHPLC

Analytical HPLC was carried out on an Agilent 1290 series system with UV detection at 214 nm. Column: Phenomenex Kinetex C18, (150 x 2.1 mm, 1.8 µm, 100 Å silica); flow rate 0.375 mL/minute; Solvent System: A = water with 0.1% TFA (v/v), B = acetonitrile with 0.08% TFA (v/v); Gradient: 3 min hold 5% B, 5-65% B gradient over 30 min, 3 min hold 65% B, 10 min post run 5% B.

2.4.6. Reverse Phase High Pressure Liquid Chromatography Mass Spectrometry Methods

Data were processed using Agilent MassHunter Workstation Qualitative Analysis Version B.10.00 Software or Agilent MassHunter BioConfirm Software B.10.00. Deconvoluted masses of proteins were obtained using a maximum entropy algorithm. Unless otherwise depicted, the following parameters were used for deconvolution: A range 3,000 Da less than the starting mass rounded down to the nearest 10,000 Da was used as a lower limit (not lower than 5,000 Da). For the higher limit, a mass to the nearest 10,000 Da rounding up was used. A limited m/z range from 600-3,000 was used with a baseline subtraction factor of 3 and a mass step of 1.

Method 1 – 6545 1-91 in 7 min

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 range 300 to 3000). 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 condition was used for analysis. Column: Zorbax 300-SB C3 (5 µm, 150 x 2.1 mm, 300Å silica); Flow

Rate: 0.6 mL/min; Gradient: linearly ramp from 1% B to 91% B 0 to 5 min, 91% B to 95% B 5 to 7 min. Post time is 1% B for 3 min. Flow rate is 0.8 mL/min. MS data was acquired from 1 to 5 min.

Method 2 – 6545 1-91 in 15 min

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 range 300 to 3000). 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 condition was used for analysis. Column: Zorbax 300-SB C3 (5 µm, 150 x 2.1 mm, 300 Å silica); Flow Rate: 0.6 mL/min; Gradient: 1% B 0-2 min, linearly ramp from 1% B to 91% B 2 to 12 min, 91% B to 95% B 12 to 15 min. Post time is 1% B for 3 min. Flow rate is 0.8 mL/min. MS data was acquired from 4 to 11 min.

Method 3 – 6550-1 5-95 in 10 min

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). The solvent mixtures used were as above. Column: Phenomenex Aeris C4 Widepore (5 µm, 150 x 2.1 mm, 300 Å silica); Flow Rate: 0.1 mL/min; Gradient: 5% B 0-2 min, linearly ramp from 5% B to 95% B 2 to 12 min, 95% B. Post time is 5% B for 3 min. MS data was acquired from 4 to 11 min.

Method 4 – 6550-2 1-91 in 15 min

Analysis was performed on an Agilent 1290 Infinity HPLC coupled to an Agilent 6550 ESI-Q-TOF mass spectrometer. MS was run in positive ionization mode, extended dynamic range (2 GHz), and standard mass range (m/z in range 300 to 3000). 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 condition was used for analysis. Column: Zorbax 300-SB C3 (5 µm, 150 x 2.1 mm, 300 Å silica); Flow Rate: 0.8 mL/min; Gradient: 1% B 0-2 min, linearly ramp from 1% B to 91% B 2 to 12 min, 91% B to 95% B 12 to 15 min. Post time is 1% B for 3 min. Flow rate is 0.8 mL/min. MS data was acquired from 4 to 11 min.

2.4.7.Preparative Reverse Phase Purification Methods

An appropriate amount of crude, lyophilized peptide powder was weighed and added to fresh 50 mL Falcon tube. A denaturing buffer consisting of 6 M Gdn HCl, 50 mM TRIS HCl, 50 mM DTT was freshly prepared, and the pH adjusted with 6 M NaOH and 6 M HCl to a value of 7.5 as measured with a pH electrode. An aliquot of the denaturing buffer was added to the lyophilized peptide to prepare a solution at 10 mg of peptide per mL of buffer. A fresh screw-cap was added to the Falcon tube, and the mixture sequentially vortexed and sonicated until no visible precipitate remained. Depending on the solubility of the peptide sequence, additional denaturing buffer was required. The pH was readjusted to 7.5 with 6 M NaOH and 6 M HCl as measured by pH paper, and the mixture was incubated for 30 min at room temperature. After incubation, the Falcon tube screw cap was removed, and the mixture was transferred to a 20 mL syringe affixed with a 0.22 µM nylon syringe filter. The mixture was forced through the syringe filter into a fresh Falcon tube. The Falcon tube originally containing the un-filtered peptide solution was rinsed with an additional 2 mL of the denaturing buffer that was then filtered using the same filtration apparatus into the Falcon tube containing the filtered peptide solution. Any bubbles generated were removed by centrifugation of the falcon tube containing the dissolved peptide at 3220 rcf for 5 min. The clarified peptide mixture was drawn-up with a fresh 20 mL syringe and added to the purification column via multiple injections with a manual injection valve affixed with a 10 mL loading loop.

Method 1 – 2-Stage Semiprep

Column: Agilent Zorbax 300 SB C3 (5 µm, 9.4 x 250 mm, 300 Å pore size) with an Agilent C3 Zorbax SB 300 guard column.

Loading: < 50 mg

Flow Rate: 4 mL/min.

Column Temperature: 60 °C

Instrument: Agilent mass directed purification system (1260 Infinity LC and 6130 Single Quad MS), affixed with a Timberline Instrument TL105 HPLC column heater. Mobile phases used for LC-MS analysis were Solution A (0.1% v/v TFA in water) and Solution B (0.1% v/v TFA in acetonitrile).

Purification consisted of two stages: gradient generation, and preparative separation.

Gradient Generation

The purification column was equilibrated to 5% B, and an aliquot of the freshly prepared filtered protein solution accounting for approximately 1 mg of crude protein was injected. The column rinsed with 5% B until the denaturing buffer components were rinsed from the column, and the absorbance at 214 nm retuned to baseline. The B percentage was then linearly raised from 5% B to 65% B over 60 min. The resulting chromatogram was then analyzed to identify the mobile phase composition that began elution of the desired protein peak (% B can be calculated from the gradient and elution time) – this value is termed “C”.

Preparative Separation

The purification column was equilibrated to 5% B, and the remainder of the filtered protein sample was injected onto the column. The column rinsed with 5% B until the denaturing buffer components were rinsed from the column, and the absorbance at 214 nm retuned to baseline. The B percentage was then linearly raised from 5% B to (C - 10) % B at a rate of 1% B / min. The B percentage was then linearly raised from (C - 10) % B to (C + 10) % B over 100 minutes, with 1-minute fractions. The column was then washed with a gradient from (C + 10) % B to 65% B at a rate of 1% B / min.

Method 2 – 2-Stage Prep

Column: Agilent Zorbax 300 SB C3 (5 µm, 21 x 250 mm, 300 Å pore size) with an Agilent C3 Zorbax SB 300 guard column.

Loading: < 250 mg

Flow Rate: 20 mL/min.

Column Temperature: 60 °C

Instrument: Agilent mass directed purification system (1260 Infinity LC and 6130 Single Quad MS), affixed with a Timberline Instrument TL105 HPLC column heater. Mobile phases used for LC-MS analysis were Solution A (0.1% v/v TFA in water) and Solution B (0.1% v/v TFA in acetonitrile).

Purification consisted of two stages: gradient generation, and preparative separation.

Gradient Generation

The purification column was equilibrated to 5% B, and an aliquot of the freshly prepared filtered protein solution accounting for approximately 5 mg of crude protein was injected. The column rinsed with 5% B until the denaturing buffer components were rinsed from the column, and the absorbance at 214 nm retuned to baseline. The B percentage was then linearly raised from 5% B to 65% B over 60 min The resulting chromatogram was then analyzed to identify the mobile phase composition that began elution of the desired protein peak (% B can be calculated from the gradient and elution time) – this value is termed “C”.

Preparative Separation

The purification column was equilibrated to 5% B, and the remainder of the filtered protein sample was injected onto the column. The column rinsed with 5% B until the denaturing buffer components were rinsed from the column, and the absorbance at 214 nm retuned to baseline. The B percentage was then linearly raised from 5% B to (C - 10) % B at a rate of 1% B / min. The B percentage was then linearly raised from (C - 10) % B to (C + 10) % B over 100 min, with 0.75-min fractions. The column was then washed with a gradient from (C + 10) % B to 65% B at a rate of 1% B / min.

Method 3 – 1-Stage Semiprep

Column: Agilent Zorbax 300 SB C3 (5 µm, 9.4 x 150 mm, 300 Å pore size) with an Agilent C3 Zorbax SB 300 guard column.

Loading: < 15 mg

Flow Rate: 8 mL/min.

Column Temperature: 60 °C

Instrument: Agilent mass-directed purification system (1260 Infinity LC and 6130 Single Quad MS), affixed with a Timberline Instrument TL105 HPLC column heater. Mobile phases used for LC-MS analysis were Solution A (0.1% v/v TFA in water) and Solution B (0.1% v/v TFA in acetonitrile).

The purification column was equilibrated to 15% B, and the freshly prepared filtered crude protein solution was injected. The column rinsed with 5% B until the denaturing buffer components were rinsed from the column, and the absorbance at 214 nm retuned to baseline. The B percentage was then linearly raised from 15% B to 65% B over 80 min with 1-min fractions.

2.4.8. Phage Display Protocol

2.4.8.1. Phage Library Construction and Sequencing

Stapled phage-displayed peptide libraries were constructed using the previously described methods (40). Briefly, Peptide Display Cloning System kit from New England Biolabs is used to construct M13KE-containing stapled phage libraries (New England Biolabs, Ipswich, MA). Library oligonucleotides are chemically synthesized using a mix of trimer phosphoramides (Glen Research, Sterling, VA) without codons encoding cysteine, lysine, proline, and glycine, annealed, extended, and ligated into a digested M13KE vector. All DNA products are purified using Monarch PCR and DNA cleanup kit (New England Biolabs, Ipswich, MA). The resulted library-containing phage vector is transformed into E. coli strain ER2738 (Lucigen, Middleton, WI) by electroporation and amplified by adding the post-rescue electroporated cells to a 500 mL E. coli culture at early-lag phase (OD600 = 0.01). Phage propagation, purification, and stapling are conducted according to a previously described method(1). The quality of the final phage libraries is assessed by Sanger and Next Generation Sequencing.

To conduct phage library screening, we followed our previously described procedure (40). Briefly, peptide-displayed phage libraries are incubated with streptavidin magnetic beads for 1 h at room temperature in a blocking buffer containing 5% w/v nonfat milk. For each screening condition, biotinylated protein is captured with streptavidin-coated magnetic beads at room temperature for 15 min, the supernatant is removed using a plate magnet and the beads are resuspended in 50 µL of the blocking buffer. 150 µL of the depleted phage library is added to each well for 200 µL final volume, plates are sealed, and the screening reactions are incubated at room temperature for 45 min, with rotation to maintain beads in solution. Following binding, beads are washed 5 times with ice-cold washing buffer (1X TBS, 1 mM MgCl2, 1% (w/v) BSA, 0.1% Tween-20, 0.02% (w/v)

sodium azide, 2% (w/v) glycerol), beads containing protein-bound phage are collected and directly processed for NGS.

2.4.8.2. Phage Next Generation Sequencing, Hit ID and Clustering

Next Generation Sequencing is performed according to a previously described procedure (40). Briefly, phage particles are denatured from magnetic beads at 95 ºC for 15 min with an added spike-in sequence (not a library member that is used to enable cross-well normalization of sequence reads), followed by a two-step low-cycled PCR to introduce Illumina adaptors and 10bp TruSeq DNA UD Indexes (Illumina, San Diego, CA) according to an Illumina's 16S Metagenomic Sequencing Library Preparation protocol. The NGS library is sequenced by an Illumina NovaSeq platform using a 2x150 bp high-output kit (Illumina, San Diego, CA).

Hit ID and clustering is performed according to a previously described(40). Briefly, NGS reads are trimmed for quality and filtered for sequences that matched the design of the phage library. Counts for each unique sequence are tallied, and then normalized by the counts of the spike-in sequence added to each sample. A metric called Hit Strength is computed for each sequence as the fold change between the normalized counts in the highest target concentration sample and the normalized counts in the blank samples (averaged across experimental replicates). When 0 counts are observed for a sequence in blank samples, a count of 0.5 is used to prevent dividing by zero.

2.4.9. Peptide Synthesis Techniques

2.4.9.1.Automated Flow Peptide Synthesis Conditions

All peptides were synthesized on one of three automated fast-flow systems, which were built in the Pentelute lab: the "Amidator", the "Automatide" system, or the "Peptidator". In all cases, synthesis conditions were identical to the optimized protein synthesis conditions as previously described (24). Synthesis variables are summarized in the following table.

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

Coupling Step	0.40 M amino acid stock in DMF 0.38 M activator stock in DMF Coupling conditions: HATU (13 pump strokes) except S and A with HATU (26 strokes) and C, H, N, Q, V, R, T with PyAOP (26 strokes)
Deprotection Step	40% piperidine in DMF with 2% FA (13 pump strokes)
Washing Steps	DMF (40 pump strokes)

Table 3 Summary of Automated Flow Synthesis Variables

2.4.9.2. Manual Peptide Synthesis Conditions

Unless otherwise specified, all manual coupling reactions were carried out in 10 mL disposable peptide synthesis vessels from Torviq affixed to vacuum manifolds. To the reaction vessel was added an appropriate amount of peptide synthesis resin, and DMF was added to completely fill the vessel. The mixture was stirred with a glass rod to remove clumps and fully suspend the resin in DMF. The finely divided resin was incubated in DMF for 30 min to fully swell.

To a 500 mL volumetric flask (VWR, part number 10124-384) was weighed 72.24 g (0.19 mol) of O-(7-azabenzotriazol-1-yl)-N,N,N’,N’-tetramethyluronium hexafluorophosphate (HATU) and DMF was added to prepare a 0.38 M stock solution of HATU in DMF. The HATU stock solution was transferred to a 500 mL Corning Pyrex glass media bottle (Mfr # 1395-500) and stored for not more than 1 week. To a 20 mL glass scintillation vial was weighed 5 eq (relative to resin loading) of Fmoc-protected amino acid. To the glass scintillation vial was added an appropriate volume of HATU stock to produce a 0.4 M solution of Fmoc-protected amino acid. The amino acid was fully solubilized with sonication and stored for not more than 24 h. In cases where the amino acid was not fully solubilized, DMF was added to adjust amino acid concentration to 0.2 M.

After the incubation time, the DMF was drained from the reaction vessel. To the glass scintillation vial with the solubilized amino acid and HATU was added 15 eq of DIEA (relative to resin loading), the screw cap affixed, and the solution mixed rapidly. After a 30 s incubation, the solution containing the now activated amino acid was added to the reaction vessel. The resin was stirred with the same glass rod to produce a homogeneous

mixture, and the apparatus was allowed to sit for 15 min. The reaction mixture was drained from the reaction vessel, and the resin rinsed with 10 mL DMF three times.

After washing the resin, the Fmoc protecting group was removed from the terminal amine. Separately, a stock solution of 20% piperidine (v/v) in DMF was prepared in a 500 mL Corning Pyrex glass media bottle. To the reaction vessel was added enough of the deprotection mixture to fully cover the resin, and let sit for 15 min. The deprotection mixture was drained from the reaction vessel and the resin rinsed with 10 mL DMF three times.

2.4.10. Protein Folding Techniques

2.4.10.1. Determination of Protein Concentration

The concentration of proteins in aqueous solutions was determined using the solution's absorbance at 280 nm (A280), or the Quick Start™ Bradford Protein Assay:

A_{280} Measurement:

All measurements were recorded on a BioTec Epoch instrument with a Take3 Micro-Volume Plate. For each sample or buffer blank to be measured, 2 µL was added to each read position. The sample buffer baseline was recorded as an average of at least 8 spots recorded simultaneously. Solution containing the protein of interest was added to no fewer than 4 spots as 2 µL aliquots, and the absorbance at 280 nm recorded. The molar extinction coefficient of each protein sequence was estimated based on the sequence of the protein via ExPASy Swiss Institute of Bioinformatics - Bioinformatics Resource Portal. Protein concentration was calculated using Beer's law

Bradford Protein Assay

For proteins where the calculated extinction coefficient was too low to provide acceptable signal intensity at the expected protein concentrations, the Quick Start™ Bradford protein assay from BioRad (Catalog number 500-0201) was used. Bovine Serum Albumin standards were prepared according to manufacturer directions. To the bottom of a black, flat bottomed, 96 plate was added 5 µL of the calibrant or solution to be titrated. Each sample was prepared three times on the same plate (3 technical replicates). Working quickly, to each sample was added 250 µL of the working dye solution equilibrated at room temperature. The plate was incubated for 5 min before the absorbance at 495 nm was measured using a BioTec Epoch instrument.

2.4.10.2. Protein Folding via Size Exclusion Chromatography

Semi-preparative size exclusion chromatography for the folding of synthetic proteins was carried out on an Agilent InfinityLab LC Series 1260 Infinity II Bio-Inert LC System with a Superdex 75 Increase 10/300 GL size exclusion chromatography (SEC) column. In all cases, the flow rate used was 0.8 mL/min. Sample application was accomplished with the use of an Agilent 1200 series manual injection valve fitted with either a 0.5 mL or 0.25 mL loading loop placed before the entrance to the SEC column. An appropriate running buffer was freshly prepared and filtered through a 0.22 µm PES bottle top filter (Corning, catalog number 431118) and the SEC column equilibrated in the running buffer for at least 2 column volumes before sample application. Unless otherwise specified, the running buffer was 50 mM HEPES, 150 mM NaCl, 0.5 mM DTT, 5% glycerol (v/v) adjusted to pH 7.5 with 6 M NaOH or 6 M HCl as required.

An appropriate amount of lyophilized, purified protein (typically 0.1 – 1.0 mg) was weighed into a low-protein-binding 2 mL Eppendorf tube. An appropriate denaturing buffer was freshly prepared. Unless otherwise specified, the denaturing buffer consisted of 6 M Gdn HCl, 50 mM TRIS HCl, and 50 mM DTT adjusted to pH 7.5 with 6 M NaOH or 6 M HCl as required. A sufficient amount of denaturing buffer was added to the protein to achieve a final concentration of 10 mg/mL. Additional buffer was added as required to achieve a completely clear solution with no precipitate, but the total volume of the sample was kept below half the volume of the loading loop. The sample was clarified by centrifugation at 21k rcf for 15 min. The sample was applied to the SEC column, and eluted over 2 column volumes with manual fractionation into 2 mL low-protein-binding Eppendorf tubes. Fractions corresponding to the folded protein were collected, and checked by LC-MS. The resulting mixture was separated into 1 nmol aliquots, flash frozen with liquid nitrogen, and stored at -80 °C until further use.

2.4.10.3. Protein Folding via Dilution

An appropriate amount of lyophilized, purified protein (typically 0.1 – 1.0 mg) was weighed into a low-protein-binding 5 mL Eppendorf tube. Appropriate volumes of a denaturing buffer and a dilution buffer were freshly prepared. Unless otherwise specified, the denaturing buffer consisted of 6 M Gdn HCl, 50 mM TRIS HCl, and 50 mM DTT adjusted to pH 7.5 with 6 M NaOH or 6 M HCl as required, and the dilution buffer consisted

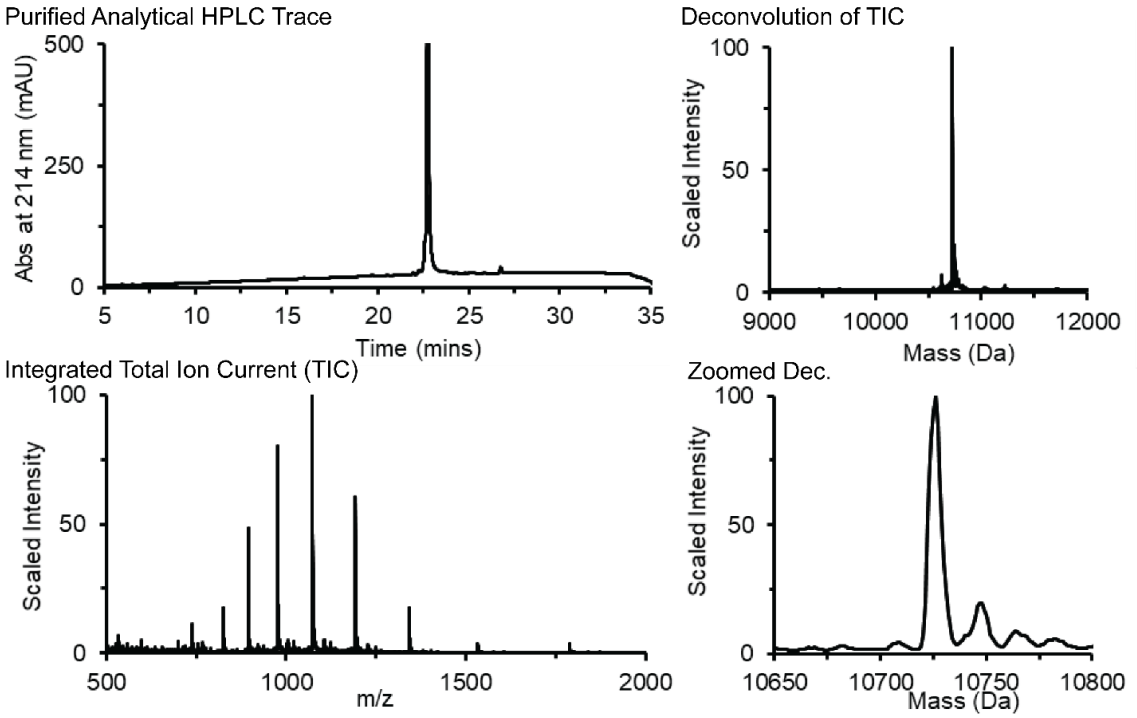
of 50 mM HEPES, 150 mM NaCl, 0.5 mM DTT, 5% glycerol (v/v) adjusted to pH 7.5 with 6 M NaOH or 6 M HCl as required. Enough denaturing buffer was added to the protein to achieve a final concentration of 10 mg/mL. Protein concentration was measured using methods 4.1.1 or 4.1.2, and the stock concentration adjusted with additional denaturing buffer to 150 µM. Dilution buffer was added to the protein in denaturing buffer to afford a final protein concentration of 25 µM. The resulting solution was kept at room temperature for 1 h before further use.

2.4.11. Protein Synthesis Data

2.4.11.1. MDM2 from AFPS

Protein	Uniprot ID	Start	End	Length	Modifications
MDM2	Q00987	26	109	84	N-Term PEG_{12}Biotin, C-Term carboxamide
TLVRPKPLLL KLLKSVGAQK DTYTMKEVLF YLGQYIMTKR LYDEKQQHIV YCSNDLLGDL FGVPSFSVKE HRKIYTMIYR NLVV					

L MDM2 Synthesis Results



Starting Resin	Crude Resin	Cleaved Resin	Isolated Mass	Theoretical Mass	Purified Mass	Isolated after Purification	Extrapolated Yield
130 mg	428.8 mg	213.7 mg	65.4 mg	137.0 mg	35.0 mg	1.3 mg	4.9 mg
0.18 mmol / g (23.4 µmol)	50% of total resin		48% Yield		54% of Isolated Peptide	4% Purification Yield	1.8% Overall Yield

Cleavage Method: Method 1

Purification Method: Method 1

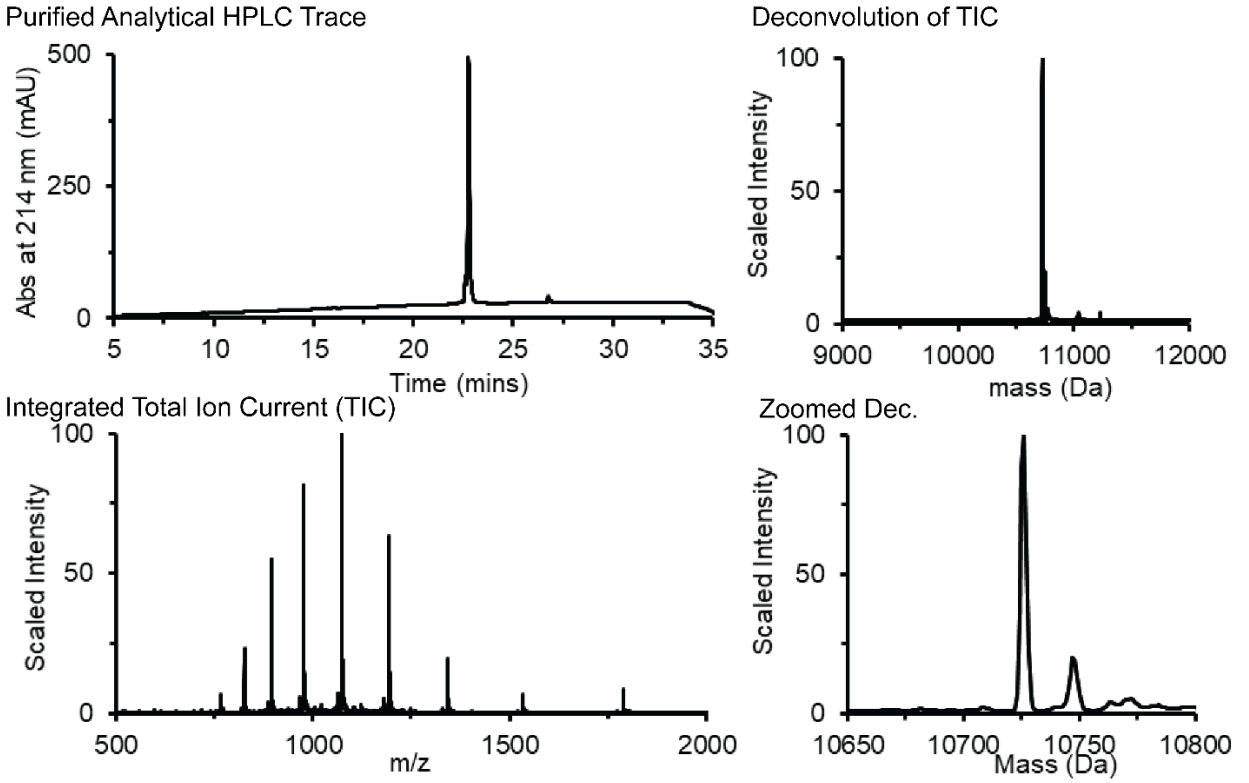
Purified HPLC Method: Method 1

Purified LCMS Method: Method 1

Calculated Mass: 10727 Da

Observed Mass: 10727 Da

D MDM2 Synthesis Results



Starting Resin	Crude Resin	Cleaved Resin	Isolated Mass	Theoretical Mass	Purified Mass	Isolated after Purification	Extrapolated Yield
130 mg	402.3 mg	202.7 mg	53.0 mg	138.5 mg	26.5 mg	2.5 mg	9.9 mg
0.18 mmol / g (23.4 µmol)	50% of total resin		38% Yield		50% of Isolated Peptide	9% Purification Yield	3.6% Overall Yield

Cleavage Method: Method 1

Purification Method: Method 1

Purified HPLC Method: Method 1

Purified LCMS Method: Method 1

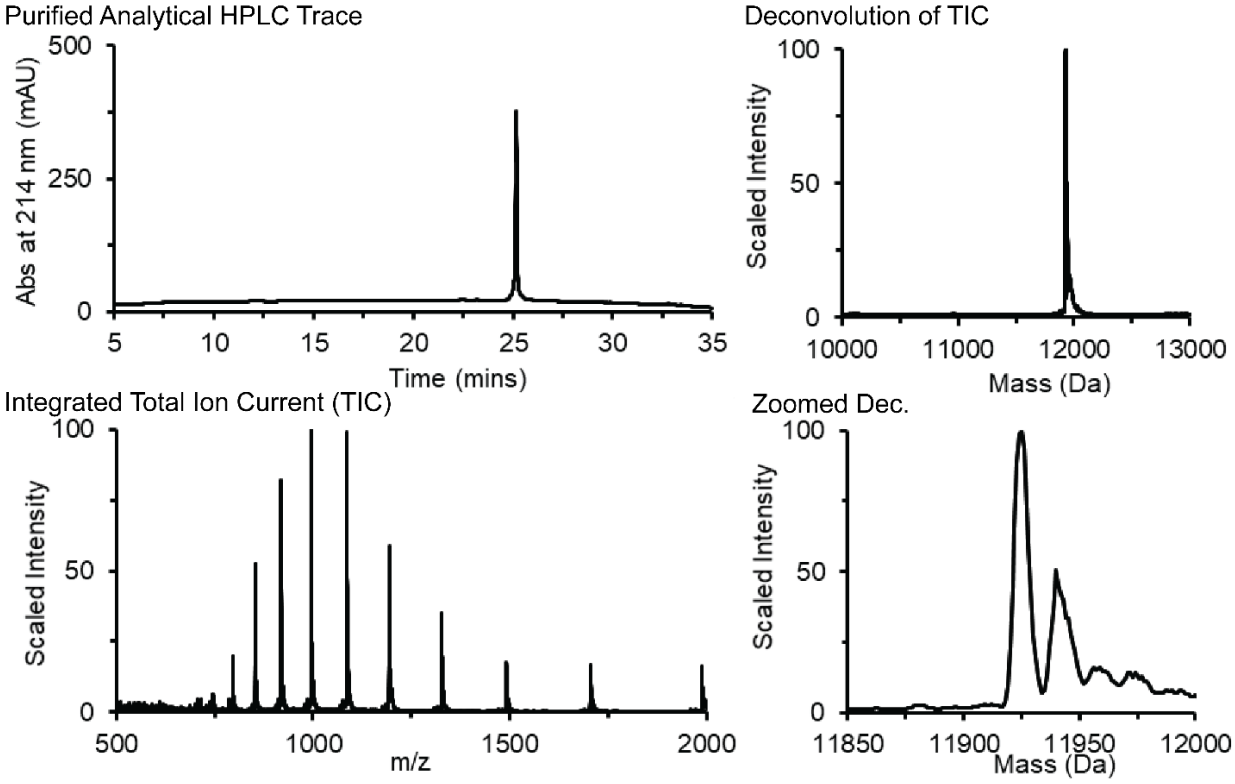
Calculated Mass: 10727 Da

Observed Mass: 10727 Da

2.4.11.2. ERG From AFPS

Protein	Uniprot ID	Start	End	Length	Modifications
ERG	P11308	108	201	94	N-Term PEG_{12}Biotin, C-Term carboxamide
MEEKHMPPPN MTTNERRVIV PADPTLWSTD HVRQWLEWAV KEYGLPDVNI LLFQNIDGKE LCKMTKDDFQ RLTPSYNADI LLSHLHYLRE TPLP					

L ERG Synthesis Results



Starting Resin	Crude Resin	Cleaved Resin	Isolated Mass	Theoretical Mass	Purified Mass	Isolated after Purification	Extrapolated Yield
130 mg	377.4 mg	188.7 mg	60.2 mg	149.4 mg	60.2 mg	6.3 mg	12.6 mg
0.18 mmol / g (23.4 µmol)	50% of total resin		40% Yield		100% of Isolated Peptide	10% Purification Yield	4.2% Overall Yield

Cleavage Method: Method 1

Purification Method: Method 1

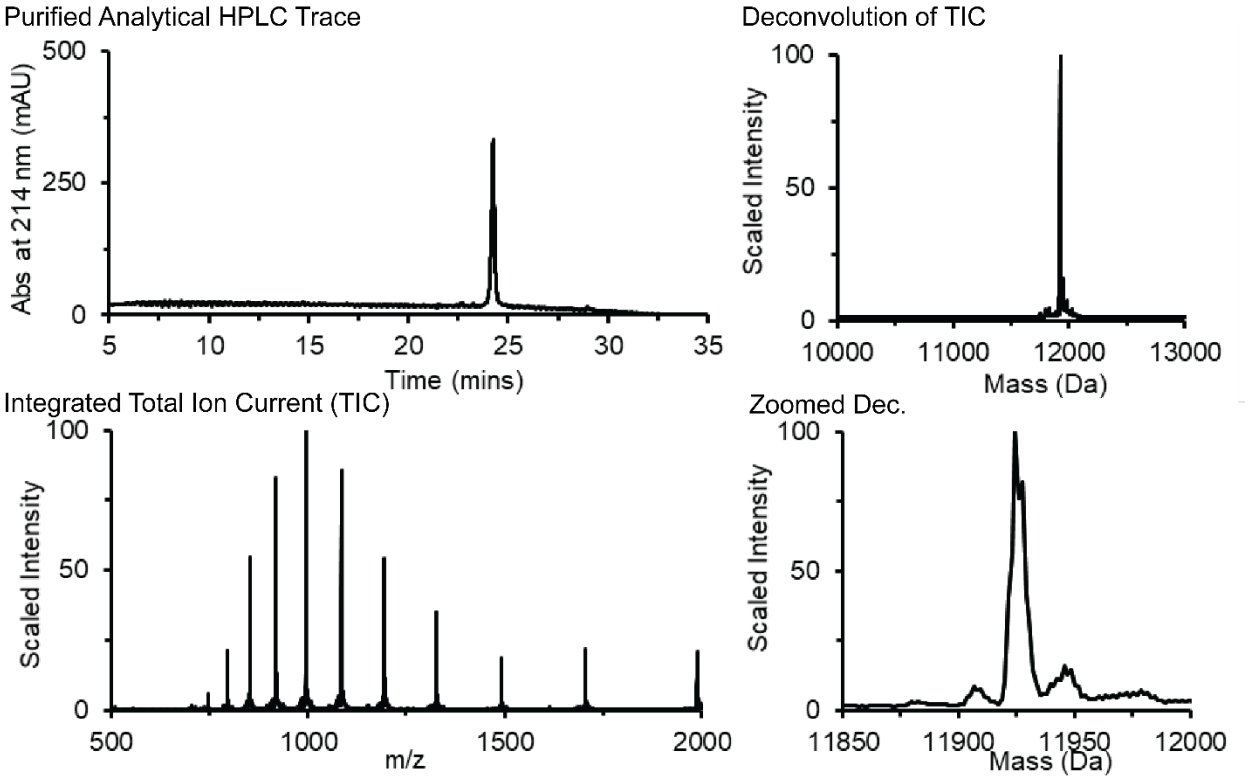
Purified HPLC Method: Method 1

Purified LCMS Method: Method 1

Calculated Mass: 11926 Da

Observed Mass: 11926 Da

D ERG Synthesis Results



Starting Resin	Crude Resin	Cleaved Resin	Isolated Mass	Theoretical Mass	Purified Mass	Isolated after Purification	Extrapolated Yield
130 mg	384.8 mg	192.8 mg	73.2 mg	149.7 mg	73.2 mg	6.9 mg	13.9 mg
0.18 mmol / g (23.4 µmol)	50% of total resin		49% Yield		100% of Isolated Peptide	9% Purification Yield	4.6% Overall Yield

Cleavage Method: Method 1

Purification Method: Method 1

Purified HPLC Method: Method 1

Purified LCMS Method: Method 1

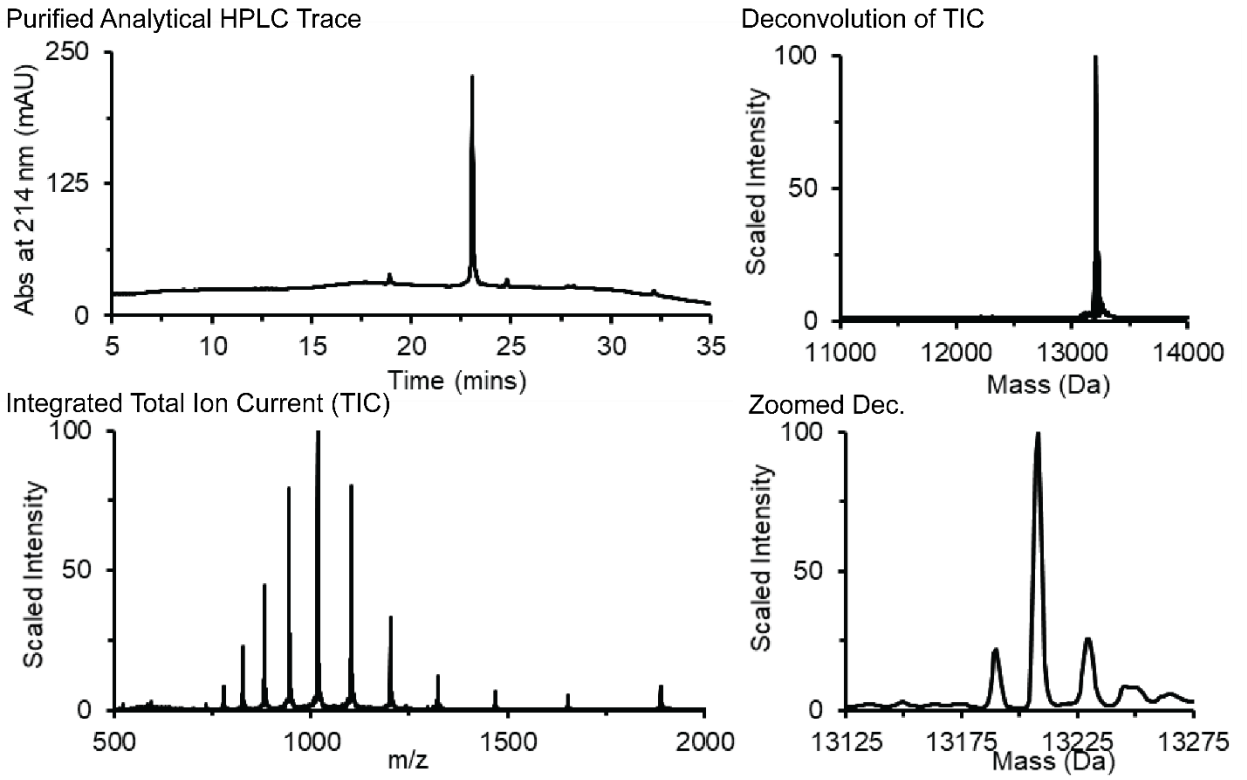
Calculated Mass: 11926 Da

Observed Mass: 11926 Da

2.4.11.3. Barnase from AFPS

Protein	Uniprot ID	Start	End	Length	Modifications
Barnase	P00648	48	157	110	N-Term PEG_{12}Biotin, C-Term carboxamide
AQVINTFDGV ADYLQTYHKL PDNYITKSEA QALGWVASKG NLADVAPGKS IGGDIFSNRE GKLPGKSGRT WREADINYTS GFRNSDRILY SSDWLIYKTT DHYQTFTKIR					

L Barnase Synthesis Results



Starting Resin	Crude Resin	Cleaved Resin	Isolated Mass	Theoretical Mass	Purified Mass	Isolated after Purification	Extrapolated Yield
130 mg	407.2 mg	196.4 mg	31.1 mg	160.2 mg	31.1 mg	1.6 mg	3.3 mg
0.18 mmol / g (23.4 µmol)	48% of total resin		19% Yield		100% of Isolated Peptide	5% Purification Yield	1.0% Overall Yield

Cleavage Method: Method 1

Purification Method: Method 1

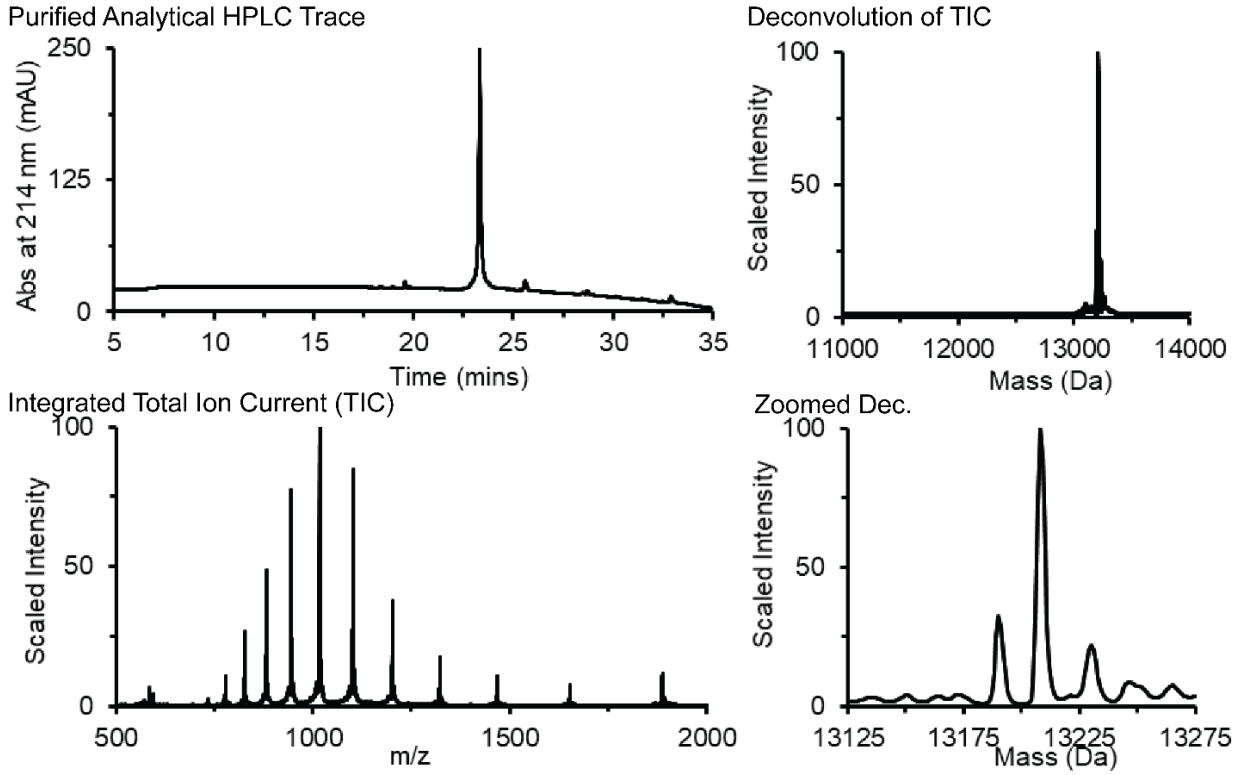
Purified HPLC Method: Method 1

Purified LCMS Method: Method 1

Calculated Mass: 13209 Da

Observed Mass: 13209 Da

D Barnase Synthesis Results



Starting Resin	Crude Resin	Cleaved Resin	Isolated Mass	Theoretical Mass	Purified Mass	Isolated after Purification	Extrapolated Yield
130 mg	384.4 mg	190.9 mg	32.0 mg	165.0 mg	32.0 mg	2.3 mg	4.6 mg
0.18 mmol / g (23.4 µmol)	50% of total resin		19% Yield		100% of Isolated Peptide	7% Purification Yield	1.4% Overall Yield

Cleavage Method: Method 1

Purification Method: Method 1

Purified HPLC Method: Method 1

Purified LCMS Method: Method 1

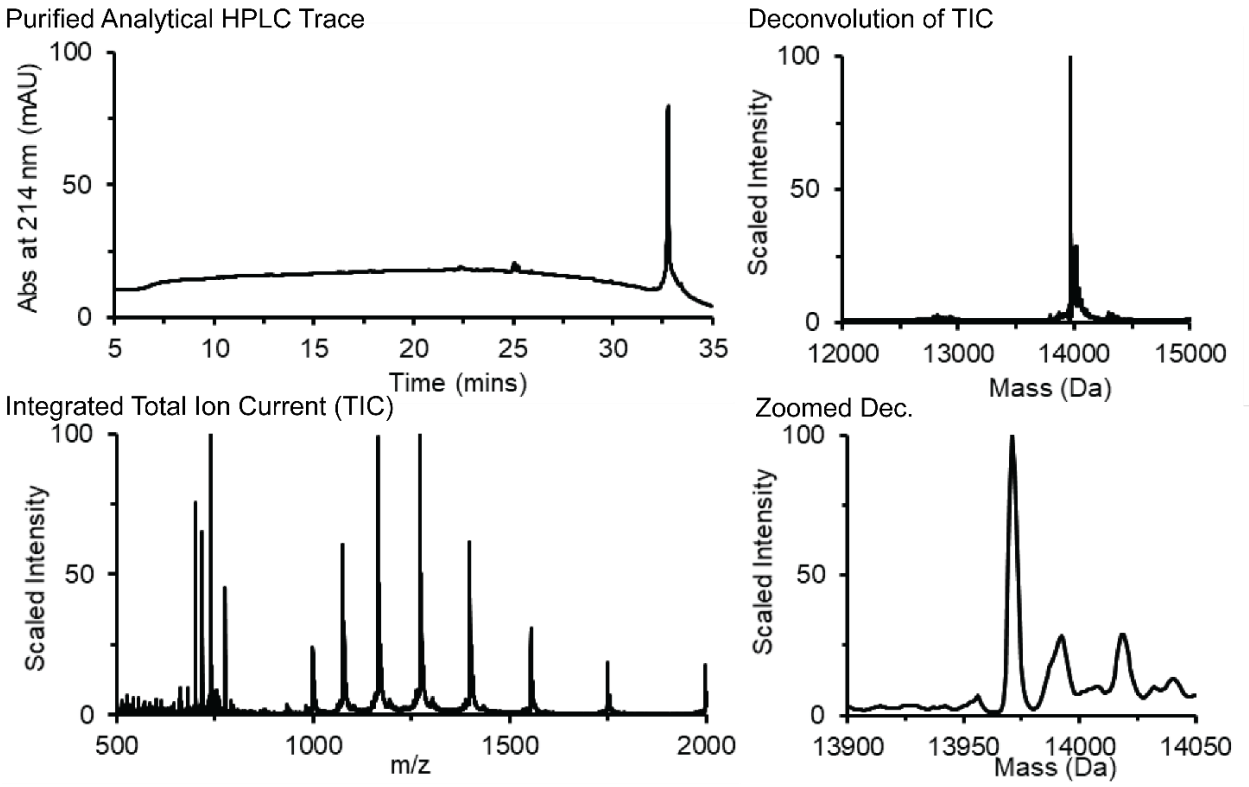
Calculated Mass: 13209 Da

Observed Mass: 13209 Da

2.4.11.4.IRAK2 from AFPS

Protein	Uniprot ID	Start	End	Length	Modifications
IRAK2	O43187	2	112	111	N-Term PEG_{12}Biotin, C-Term carboxamide
ACYIYQLPSW VLDDLCRNMD ALSEWDWMEF ASYVITDLTQ LRKIKSMERV QGVSITRELL WWWGMRQATV QQLVDLLCRL ELYRAAQIIL NWKPAPEIRC PIPAFPDSVK P					

L IRAK2 Synthesis Results



Starting Resin	Crude Resin	Cleaved Resin	Isolated Mass	Theoretical Mass	Purified Mass	Isolated after Purification	Extrapolated Yield
130 mg	402.0 mg	201.0 mg	22.4 mg	169.5 mg	22.4 mg	2.1 mg	4.1 mg
0.18 mmol / g (23.4 µmol)	50% of total resin		26% Yield		100% of Isolated Peptide	9% Purification Yield	1.2% Overall Yield

Cleavage Method: Method 1

Purification Method: Method 1

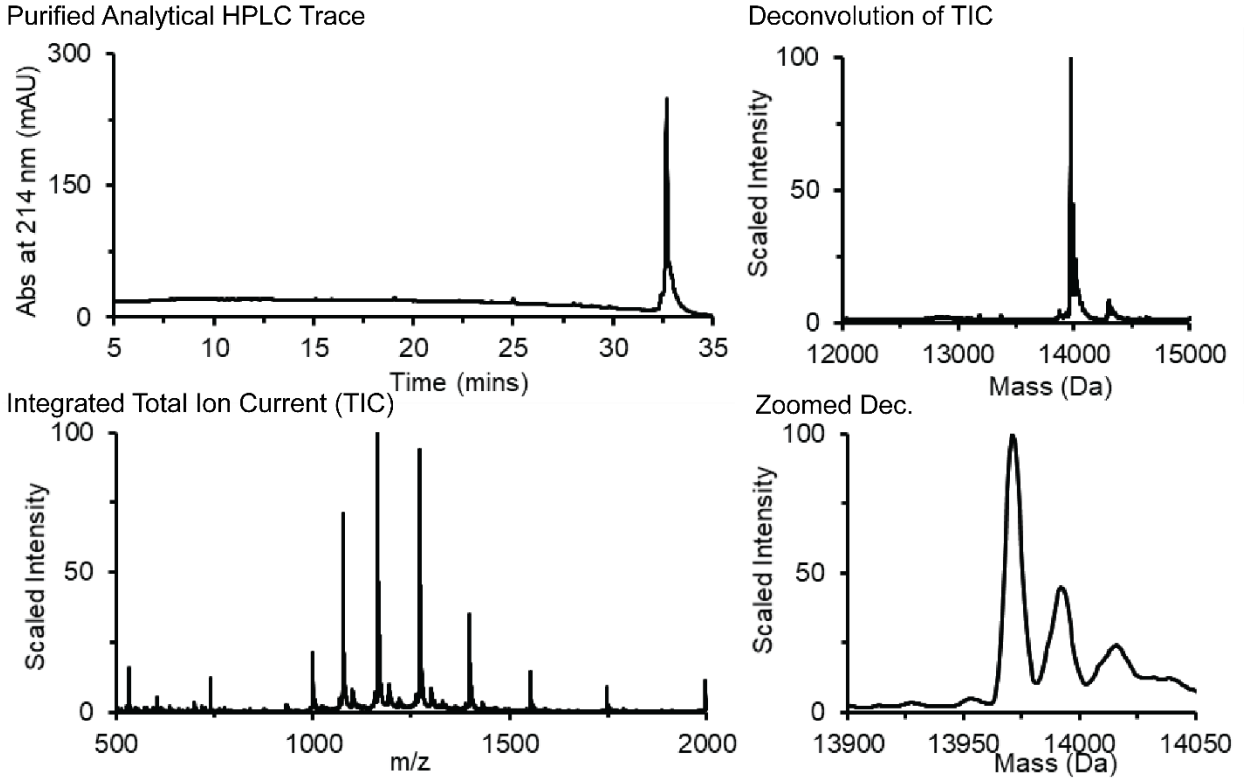
Purified HPLC Method: Method 1

Purified LCMS Method: Method 1

Calculated Mass: 13972 Da

Observed Mass: 13972 Da

D IRAK2 Synthesis Results



Starting Resin	Crude Resin	Cleaved Resin	Isolated Mass	Theoretical Mass	Purified Mass	Isolated after Purification	Extrapolated Yield
130 mg	410.3 mg	204.3 mg	73.0 mg	169.5 mg	38.4 mg	2 mg	7.6 mg
0.18 mmol / g (23.4 µmol)	50% of total resin		13% Yield		53% of Isolated Peptide	5% Purification Yield	2.1% Overall Yield

Cleavage Method: Method 1

Purification Method: Method 1

Purified HPLC Method: Method 1

Purified LCMS Method: Method 1

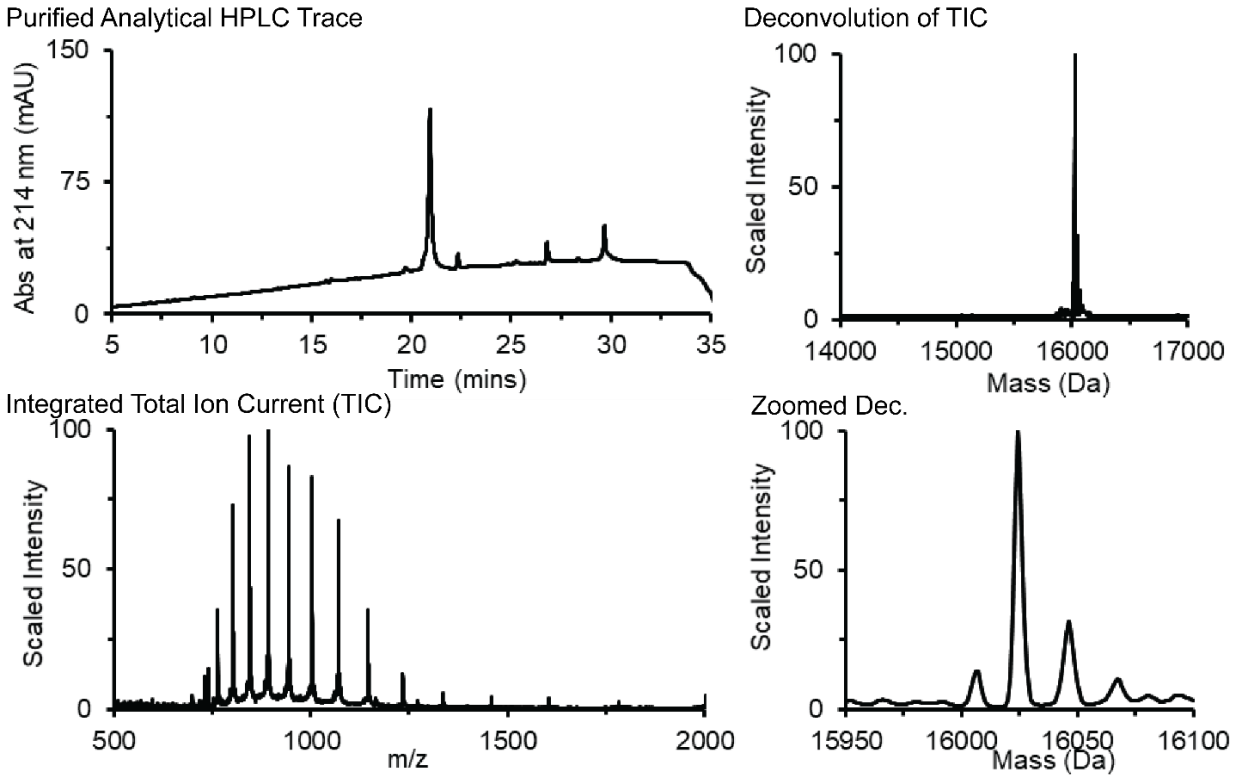
Calculated Mass: 13972 Da

Observed Mass: 13972 Da

2.4.11.5. CHIP from AFPS

Protein	Uniprot ID	Start	End	Length	Modifications
CHIP	Q9UNE7	23	153	132	N-Term PEG_{12}Biotin, C-Term carboxamide
SPSAQELKEQ GNRLFVGRKY PEAAACYGRA ITRNPLVAVY YTNRALCYLK MQQHEQALAD CRRALELDGQ SVKAHFFLGQ CQLEMESYDE AIANLQRAYS LAKEQRLNFG DDIPSALRIA KKKRWNSIEE RR					

L CHIP Synthesis Results



Starting Resin	Crude Resin	Cleaved Resin	Isolated Mass	Theoretical Mass	Purified Mass	Isolated after Purification	Extrapolated Yield
130 mg	587.0 mg	295.6 mg	45.4 mg	207.5 mg	29.4 mg	0.6 mg	1.8 mg
0.18 mmol / g (23.4 µmol)	100% of total resin		22% Yield		65% of Isolated Peptide	2% Purification Yield	0.4% Overall Yield

Cleavage Method: Method 1

Purification Method: Method 1

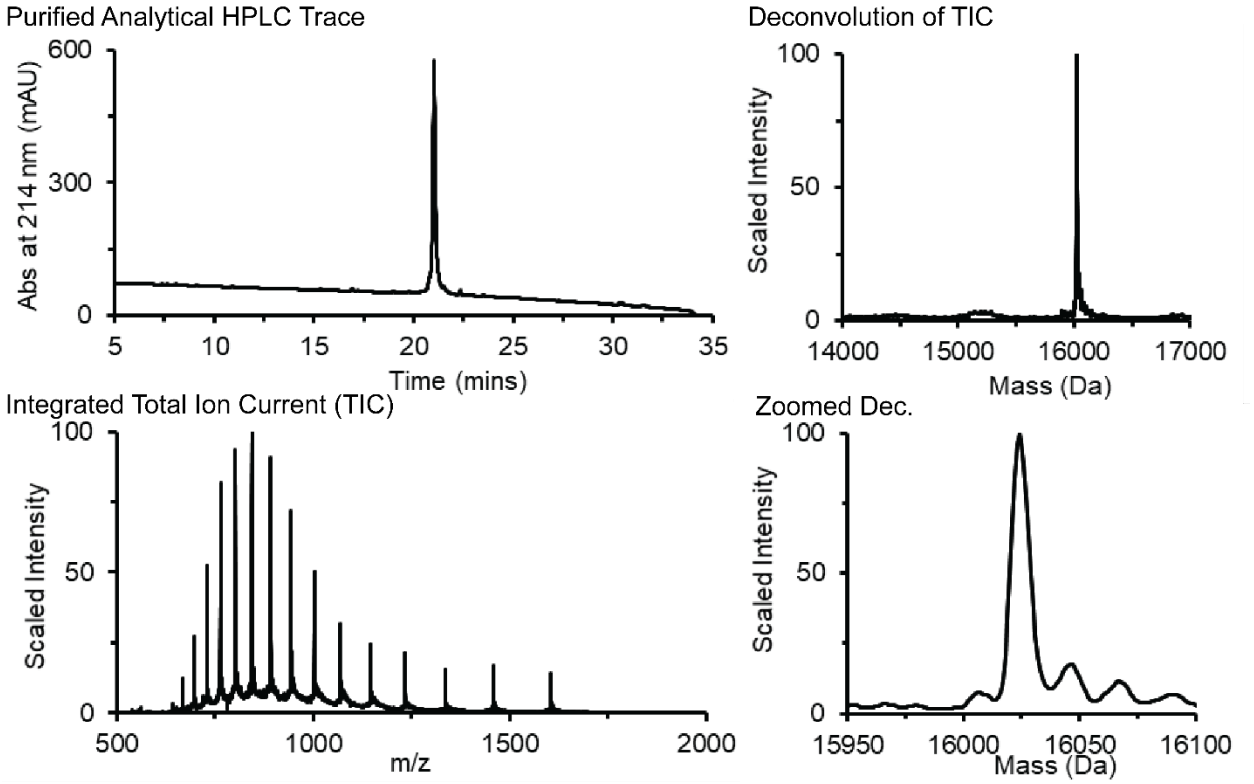
Purified HPLC Method: Method 1

Purified LCMS Method: Method 1

Calculated Mass: 16025 Da

Observed Mass: 16025 Da

D CHIP Synthesis Results



Starting Resin	Crude Resin	Cleaved Resin	Isolated Mass	Theoretical Mass	Purified Mass	Isolated after Purification	Extrapolated Yield
130 mg	n.d.	n.d.	63.1 mg	412.1 mg	63.1 mg	1.6 mg	1.6 mg
0.18 mmol / g (23.4 µmol)	100% of total resin		15% Yield		100% of Isolated Peptide	3% Purification Yield	0.4% Overall Yield

Cleavage Method: Method 1

Purification Method: Method 1

Purified HPLC Method: Method 1

Purified LCMS Method: Method 1

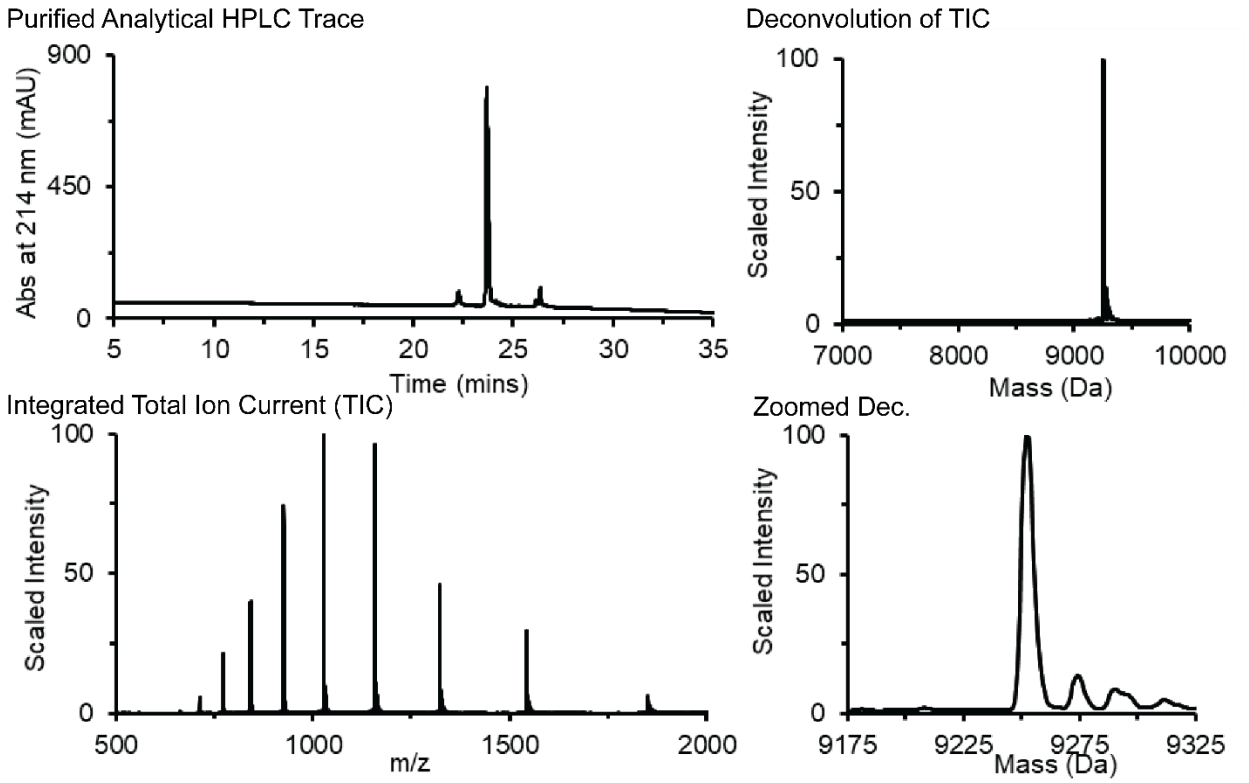
Calculated Mass: 16025 Da

Observed Mass: 16025 Da

2.4.11.6. NEMO from AFPS

Protein	Uniprot ID	Start	End	Length	Modifications
NEMO	Q9Y6K9	44	112	70	N-Term PEG_{12}Biotin, C-Term carboxamide, Nle in place of Met
EQGAPETLQR CLEENQELRD AIRQSNQILR ERCEELLHFQ ASQREEKEFL vCKFQEARKL VERLGLEKLE					

L NEMO Synthesis Results



Starting Resin	Crude Resin	Cleaved Resin	Isolated Mass	Theoretical Mass	Purified Mass	Isolated after Purification	Extrapolated Yield
130 mg	n.d.	n.d.	70.5 mg	237.3 mg	20.0 mg	1.6 mg	5.6 mg
0.18 mmol / g (23.4 µmol)	100% of total resin		30% Yield		28% of Isolated Peptide	8% Purification Yield	2.4% Overall Yield

Cleavage Method: Method 1

Purification Method: Method 1

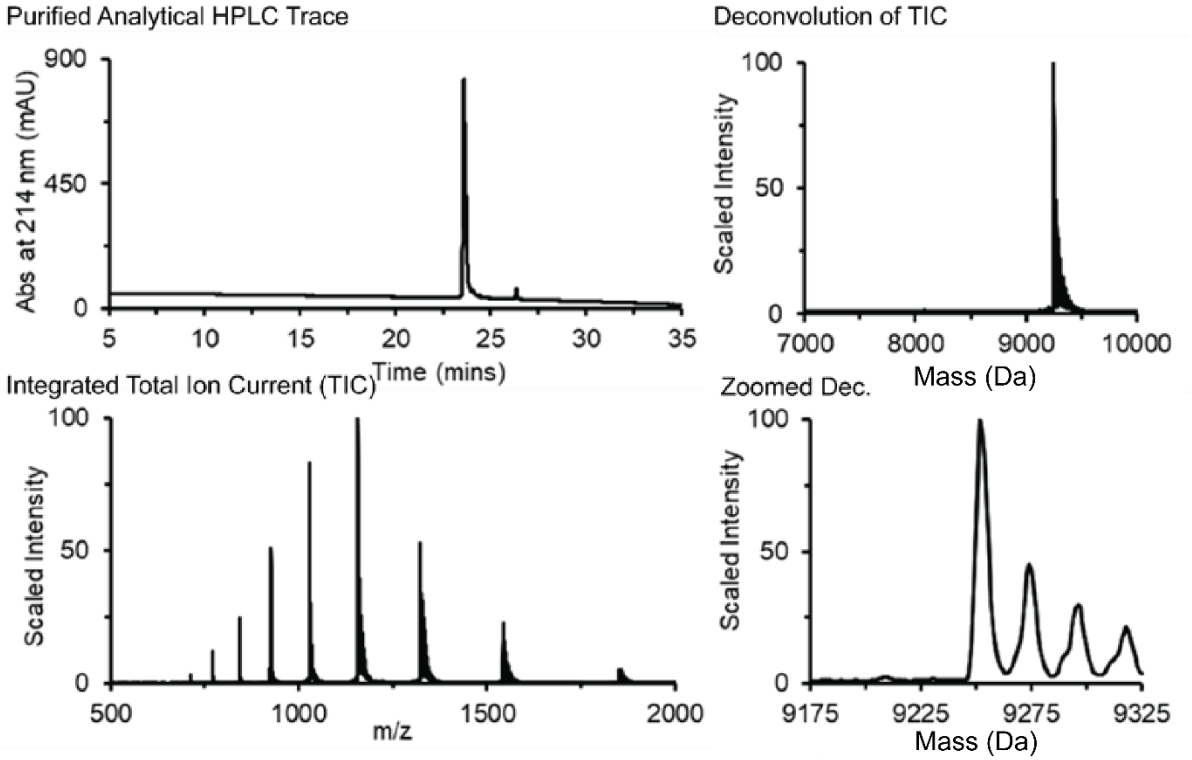
Purified HPLC Method: Method 1

Purified LCMS Method: Method 1

Calculated Mass: 9254 Da

Observed Mass: 9254 Da

D NEMO Synthesis Results



Starting Resin	Crude Resin	Cleaved Resin	Isolated Mass	Theoretical Mass	Purified Mass	Isolated after Purification	Extrapolated Yield
130 mg	n.d.	n.d.	29.7 mg	118.6 mg	29.7 mg	6.5 mg	13.0 mg
0.18 mmol / g (23.4 µmol)	50% of total resin		25% Yield		100% of Isolated Peptide	22% Purification Yield	5.5% Overall Yield

Cleavage Method: Method 1

Purification Method: Method 1

Purified HPLC Method: Method 1

Purified LCMS Method: Method 4

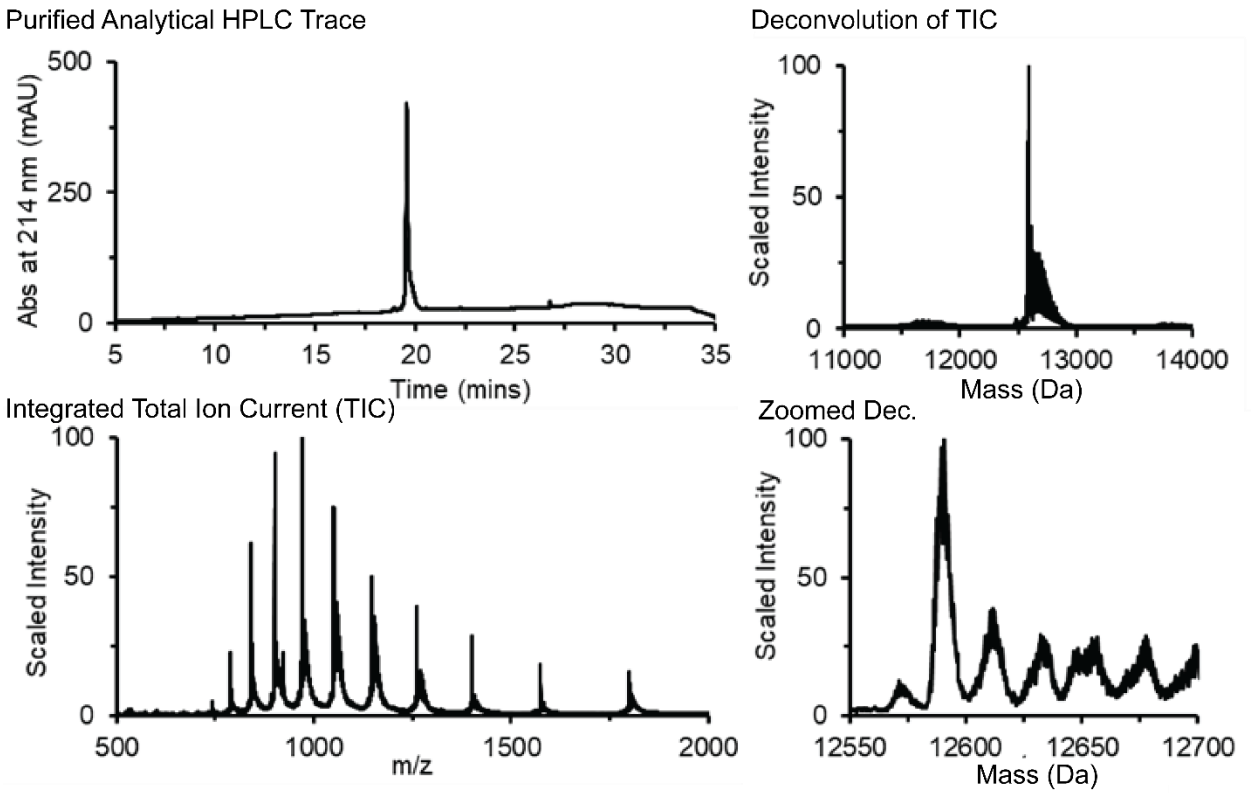
Calculated Mass: 9254 Da

Observed Mass: 9254 Da

2.4.11.7. FKBP12 from AFPS

Protein	Uniprot ID	Start	End	Length	Modifications
FKBP12	P62942	2	108	107	N-Term PEG_{12}Biotin, C-Term carboxamide
GVQVETISPG DGRTFPKRGQ TCVVHYTGML EDGKKFDSSR DRNKPFKFML GKQEVIRGWE EGVAQMSVGQ RAKLTISPDY AYGATGHPGI IPPHATLVFD VELLKLE					

L FKBP12 Synthesis Results



Starting Resin	Crude Resin	Cleaved Resin	Isolated Mass	Theoretical Mass	Purified Mass	Isolated after Purification	Extrapolated Yield
130 mg	n.d.	n.d.	95.3 mg	321.1 mg	30.6 mg	2.4 mg	7.5 mg
0.18 mmol / g (23.4 µmol)	100% of total resin		30% Yield		32% of Isolated Peptide	8% Purification Yield	2.3% Overall Yield

Cleavage Method: Method 1

Purification Method: Method 1

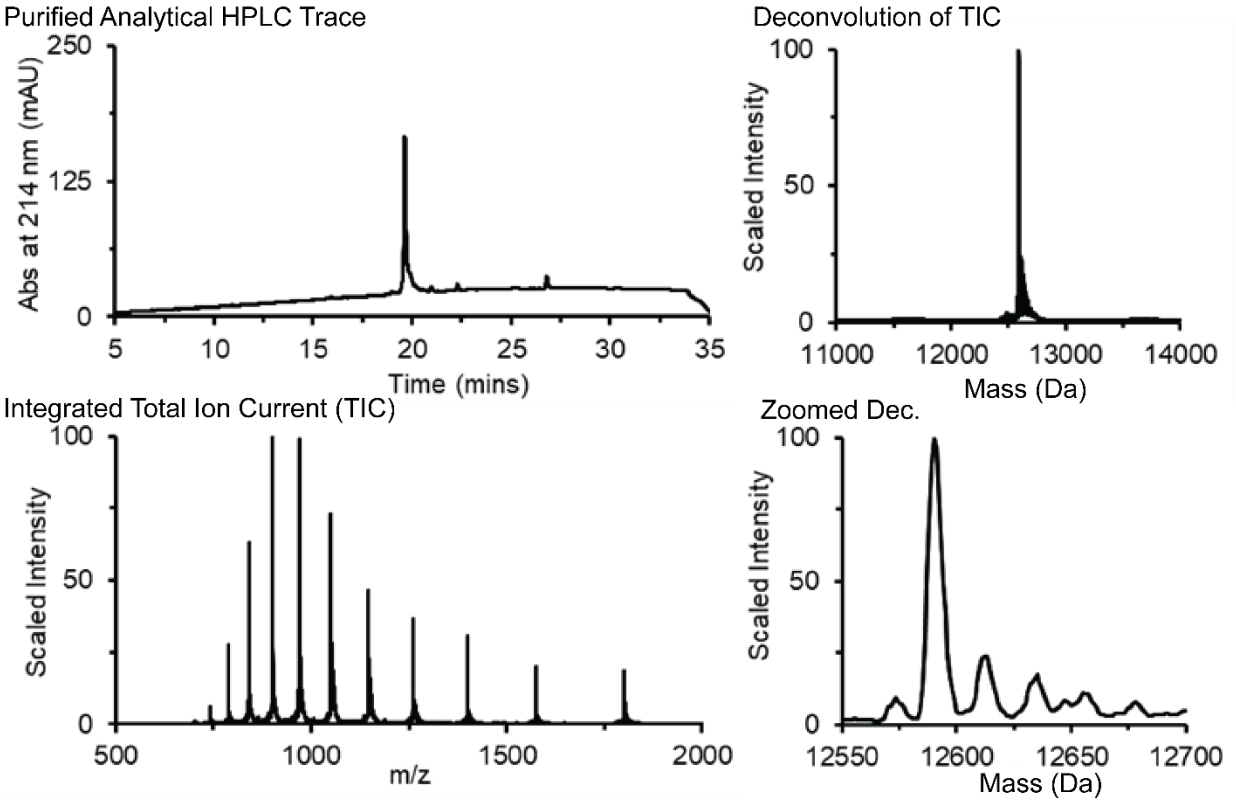
Purified HPLC Method: Method 1

Purified LCMS Method: Method 4

Calculated Mass: 12593 Da

Observed Mass: 12593 Da

D FKBP12 Synthesis Results



Starting Resin	Crude Resin	Cleaved Resin	Isolated Mass	Theoretical Mass	Purified Mass	Isolated after Purification	Extrapolated Yield
130 mg	n.d.	n.d.	89.5 mg	321.1 mg	35.8 mg	2.2 mg	6 mg
0.18 mmol / g (23.4 µmol)	100% of total resin		28% Yield		40% of Isolated Peptide	6% Purification Yield	1.7% Overall Yield

Cleavage Method: Method 1

Purification Method: Method 1

Purified HPLC Method: Method 1

Purified LCMS Method: Method 1

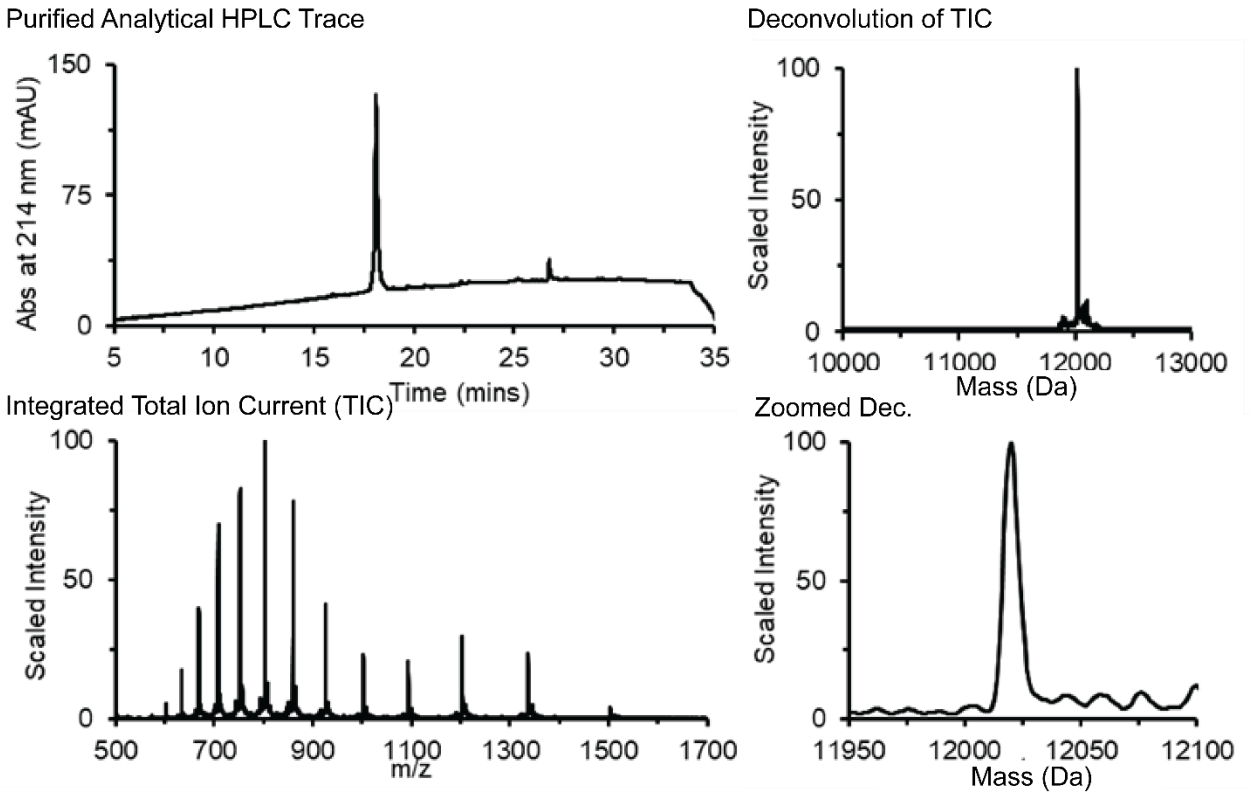
Calculated Mass: 12593 Da

Observed Mass: 12593 Da

2.4.11.8. BCL11A from AFPS

Protein	Uniprot ID	Start	End	Length	Modifications
BCL11a	Q9H165	740	835	96	N-Term PEG_{12}Biotin, C-Term carboxamide, Nle in place of Met
RSDTCEYCGK VFKNCSNLTV HRRSHTGERP YKCELCNYAC AQSSKLTRHM KTHGQVGKDV YKCEICKMPF SVYSTLEKHM KKWHSDRVLN NDIKTE					

L BCL11A Synthesis Results



Starting Resin	Crude Resin	Cleaved Resin	Isolated Mass	Theoretical Mass	Purified Mass	Isolated after Purification	Extrapolated Yield
130 mg	n.d.	n.d.	31.0 mg	163.1 mg	31.0 mg	5.2 mg	10.4 mg
0.18 mmol / g (23.4 µmol)	50% of total resin		19% Yield		100% of Isolated Peptide	17% Purification Yield	3.2% Overall Yield

Cleavage Method: Method 1

Purification Method: Method 1

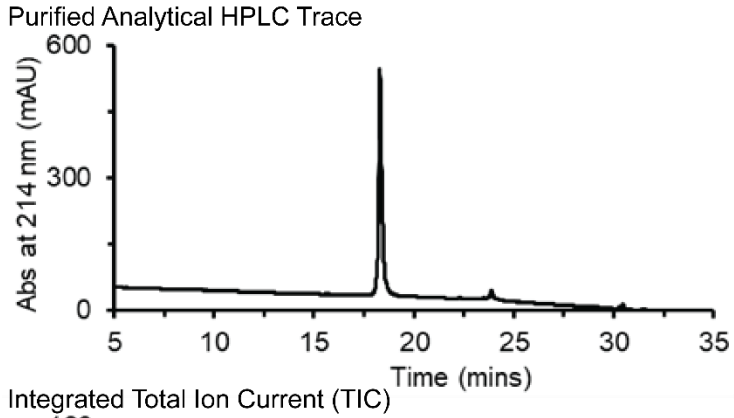
Purified HPLC Method: Method 1

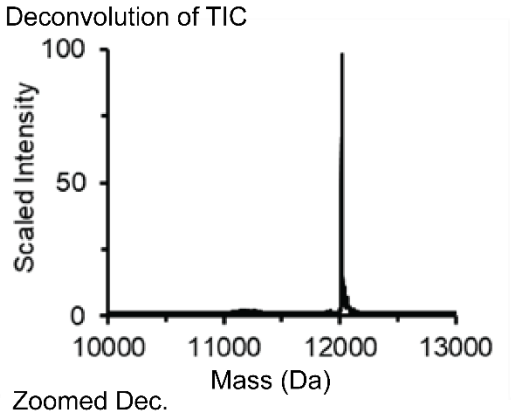
Purified LCMS Method: Method 3

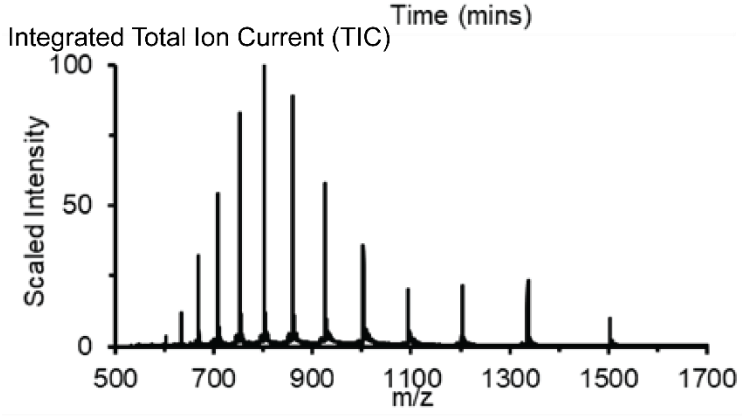
Calculated Mass: 12022 Da

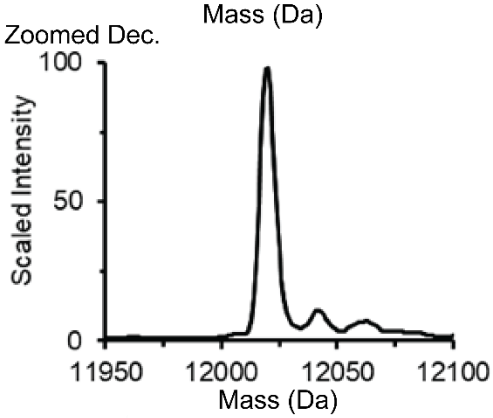
Observed Mass: 12022 Da

D BCL11A Synthesis Results









Starting Resin	Crude Resin	Cleaved Resin	Isolated Mass	Theoretical Mass	Purified Mass	Isolated after Purification	Extrapolated Yield
130 mg	n.d.	n.d.	60.0 mg	326.1 mg	60.0 mg	5.6 mg	5.6 mg
0.18 mmol / g (23.4 µmol)	100% of total resin		18% Yield		100% of Isolated Peptide	9% Purification Yield	1.7% Overall Yield

Cleavage Method: Method 1

Purification Method: Method 1

Purified HPLC Method: Method 1

Purified LCMS Method: Method 3

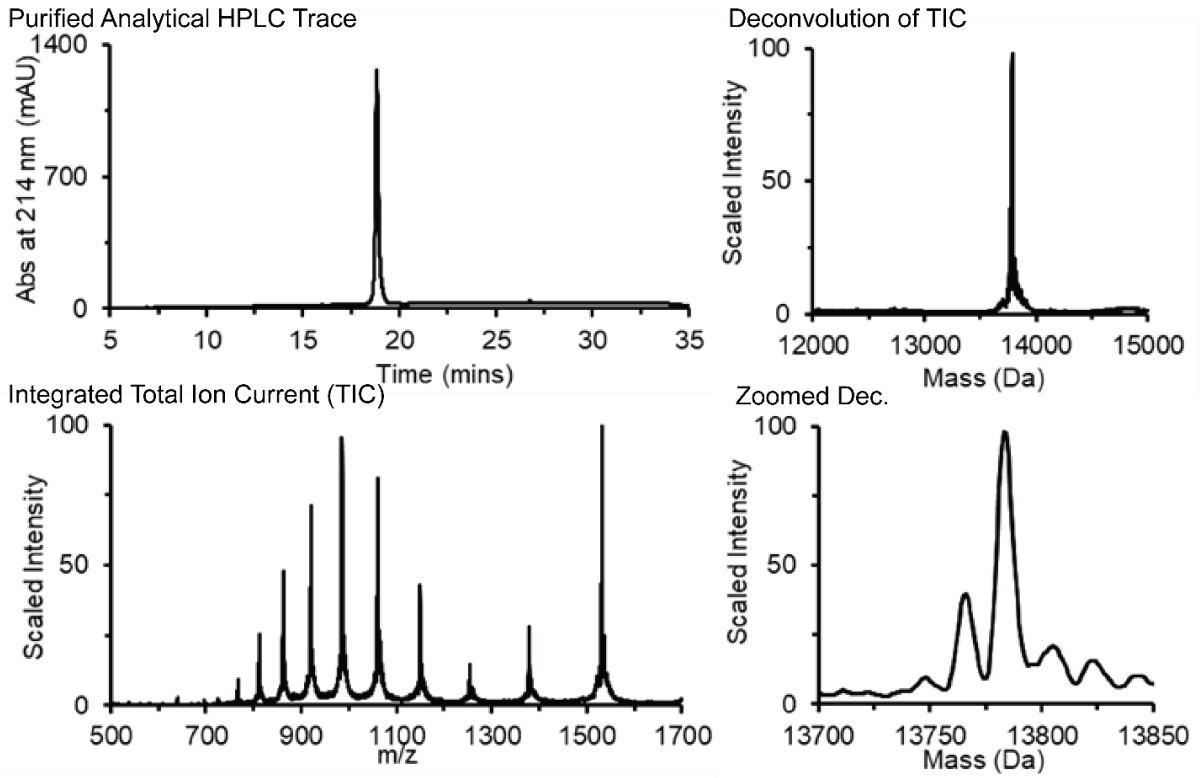
Calculated Mass: 12022 Da

Observed Mass: 12022 Da

2.4.11.9. YAP1 from AFPS

Protein	Uniprot ID	Start	End	Length	Modifications
Yap1	P46937	163	276	114	N-Term PEG_{12}Biotin, C-Term carboxamide, Nle in place of Met
SSFEIPDDVP LPAGWEMAKT SSGQRYFLNH IDQTTTWQDP RKAMLSQMNV TAPTSPPVQQ NMMNSASGPL PDGWEQAMTQ DGEIYYINHK NKTTSWLDPR LDPRFAMNQR ISQS					

L YAP1 Synthesis Results



Starting Resin	Crude Resin	Cleaved Resin	Isolated Mass	Theoretical Mass	Purified Mass	Isolated after Purification	Extrapolated Yield
130 mg	n.d.	n.d.	126.3 mg	332.4 mg	50.2 mg	1.1 mg	2.8 mg
0.18 mmol / g (23.4 µmol)	100% of total resin		38% Yield		40% of Isolated Peptide	2% Purification Yield	0.8% Overall Yield

Cleavage Method: Method 1

Purification Method: Method 1

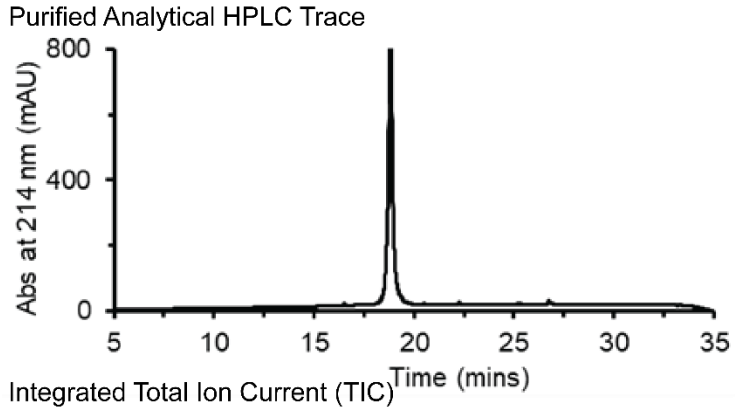
Purified HPLC Method: Method 1

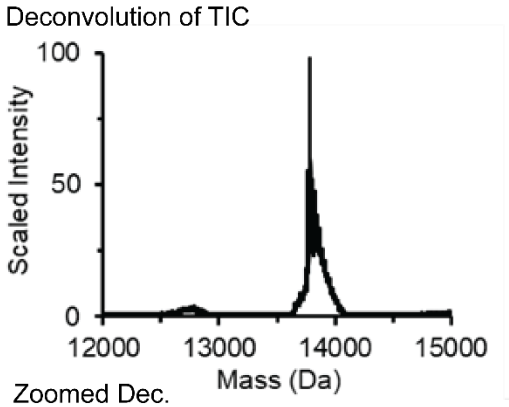
Purified LCMS Method: Method 3

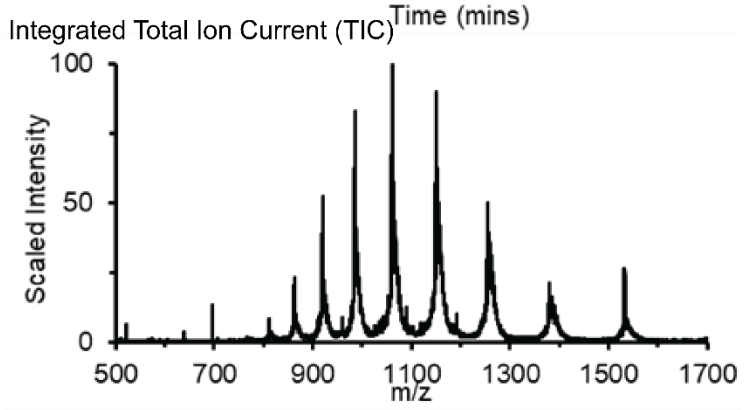
Calculated Mass: 13784 Da

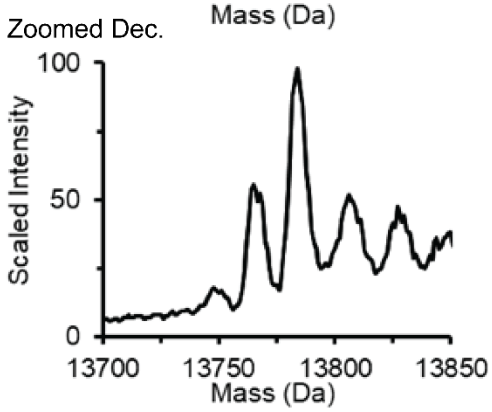
Observed Mass: 13784 Da

D YAP1 Synthesis Results









Starting Resin	Crude Resin	Cleaved Resin	Isolated Mass	Theoretical Mass	Purified Mass	Isolated after Purification	Extrapolated Yield
130 mg	n.d.	n.d.	90.4 mg	332.4 mg	50.2 mg	2.6 mg	6.1 mg
0.18 mmol / g (23.4 µmol)	100% of total resin		27% Yield		43% of Isolated Peptide	7% Purification Yield	1.8% Overall Yield

Cleavage Method: Method 1

Purification Method: Method 1

Purified HPLC Method: Method 1

Purified LCMS Method: Method 3

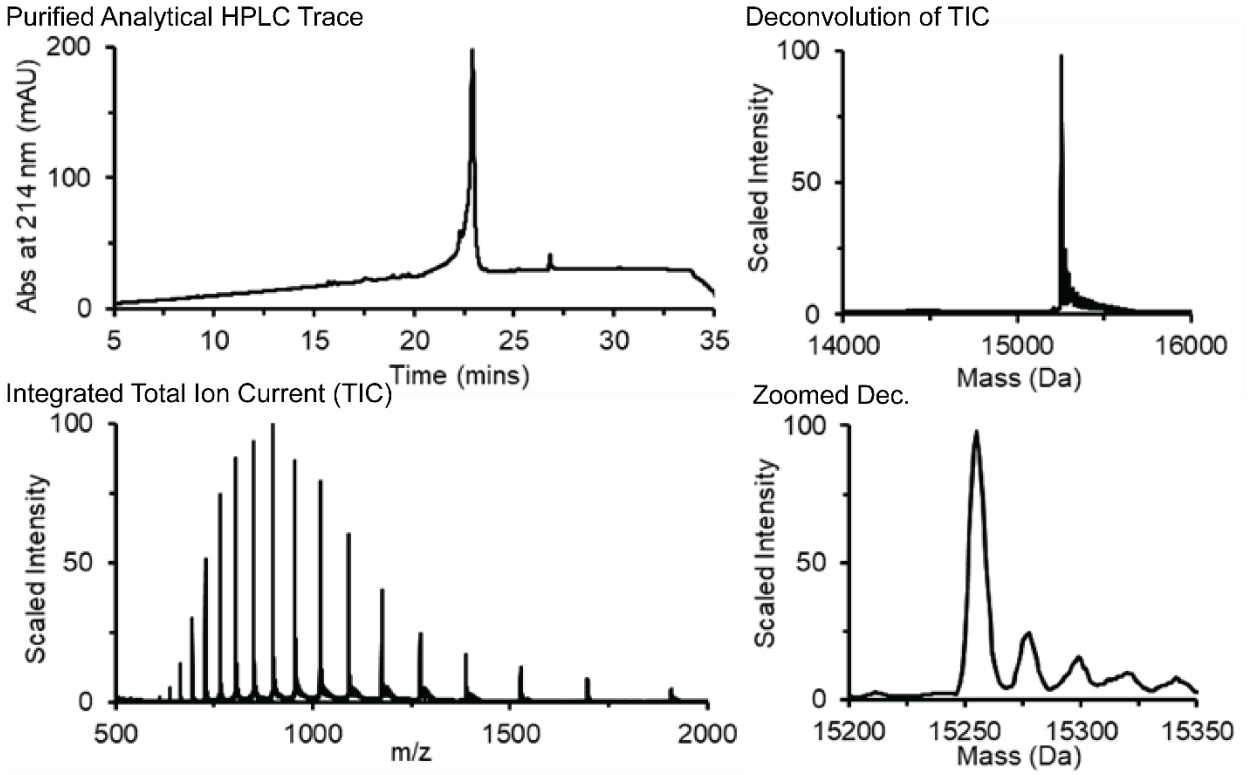
Calculated Mass: 13784 Da

Observed Mass: 13784 Da

2.4.11.10. NEMI_iZIP from AFPS

Protein	Uniprot ID	Start	End	Length	Modifications
Nemo_iZip	*6MI4	4	124	121	N-Term PEG_{12}Biotin, C-Term carboxamide, Nle in place of Met
SVKELEDKNE ELLSEIAHLK NEVARLKKLL QRCLEENQEL RDAIRQSNQI LRERCEELLH FQASQREEKE FLMCKFQEAR KLVERLGLEK LELEDKNEEL LSEIAHLKNE VARLKKLVGE R					

L NEMO_iZIP Synthesis Results



Starting Resin	Crude Resin	Cleaved Resin	Isolated Mass	Theoretical Mass	Purified Mass	Isolated after Purification	Extrapolated Yield
130 mg	n.d.	n.d.	40.3 mg	202.6 mg	40.3 mg	0.9 mg	1.8 mg
0.18 mmol / g (23.4 µmol)	50% of total resin		20% Yield		100% of Isolated Peptide	2% Purification Yield	0.4% Overall Yield

Cleavage Method: Method 1

Purification Method: Method 1

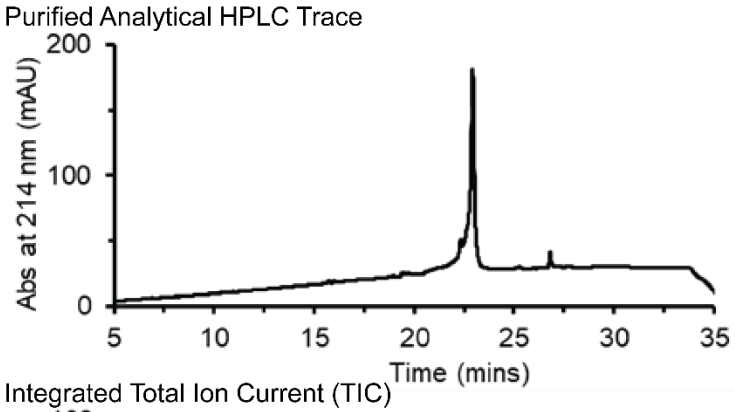
Purified HPLC Method: Method 1

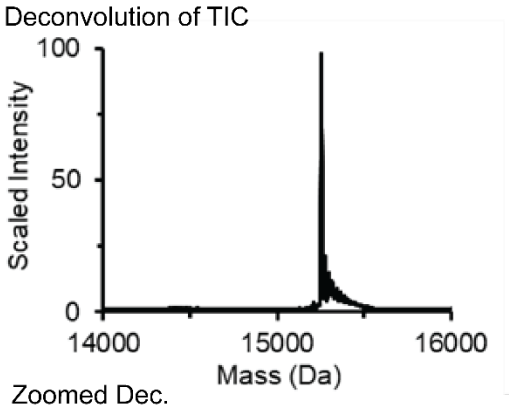
Purified LCMS Method: Method 4

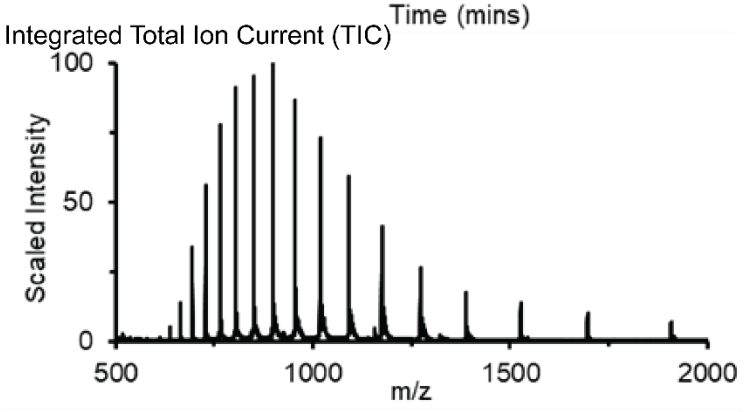
Calculated Mass: 15256 Da

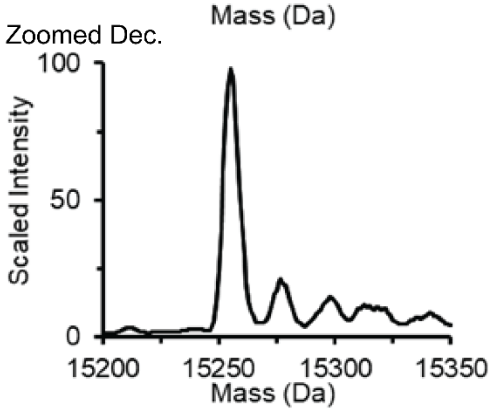
Observed Mass: 15256 Da

D NEMO_iZIP Synthesis Results









Starting Resin	Crude Resin	Cleaved Resin	Isolated Mass	Theoretical Mass	Purified Mass	Isolated after Purification	Extrapolated Yield
130 mg	n.d.	n.d.	52.4 mg	202.6 mg	52.4 mg	2.0 mg	4.0 mg
0.18 mmol / g (23.4 µmol)	50% of total resin		26% Yield		100% of Isolated Peptide	4% Purification Yield	1.0% Overall Yield

Cleavage Method: Method 1

Purification Method: Method 1

Purified HPLC Method: Method 1

Purified LCMS Method: Method 4

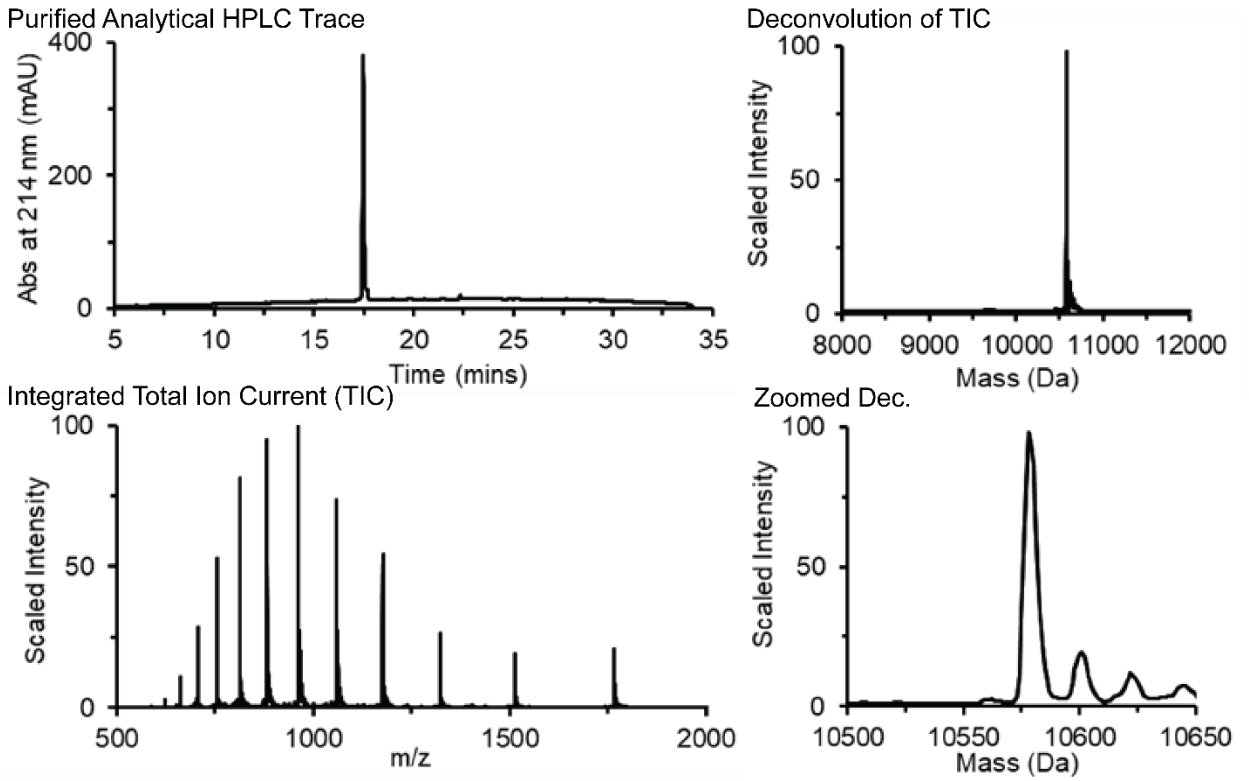
Calculated Mass: 15256 Da

Observed Mass: 15256 Da

2.4.11.11. Max from AFPS

Protein	Uniprot ID	Start	End	Length	Modifications
Max	P61244	23	102	83	N-Term PEG_{12}Biotin, C-Term +GGC + carboxamide
DKRAHHNALE RKRRDHIKDS FHSLRDSVPS LQGEKASRAQ ILDKATEYIQ YMRRKNHTHQ QDIDDLKRQN ALLEQQVRAL GGC					

L Max Synthesis Results



Starting Resin	Crude Resin	Cleaved Resin	Isolated Mass	Theoretical Mass	Purified Mass	Isolated after Purification	Extrapolated Yield
85 mg	nd	nd	208.7 mg	550 mg	161.8 mg	25.8 mg	33.0 mg
0.49 mmol / g (41.6 µmol)	100% of total resin		35% Yield		78% of Isolated Peptide	16% Purification Yield	6.1% Overall Yield

Cleavage Method: Method 2

Purification Method: Method 2

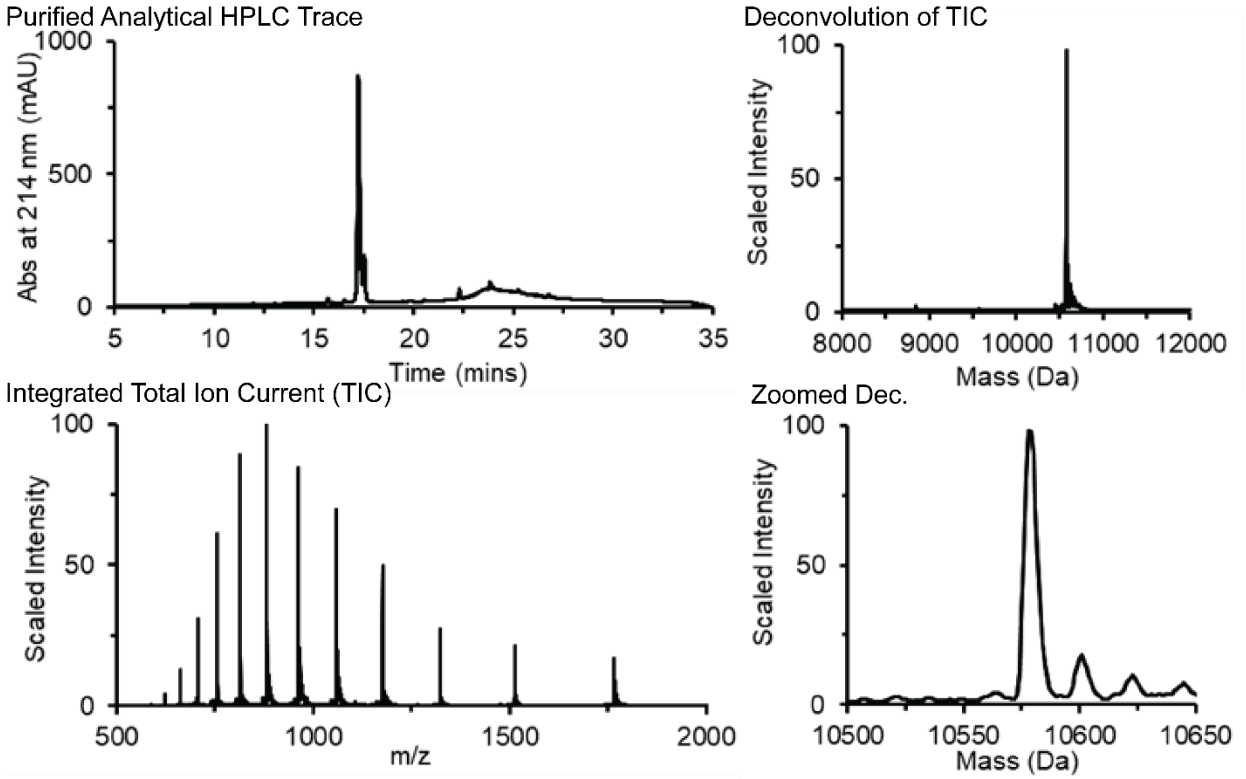
Purified HPLC Method: Method 1

Purified LCMS Method: Method 1

Calculated Mass: 10580 Da

Observed Mass: 10580 Da

D Max Synthesis Results



Starting Resin	Crude Resin	Cleaved Resin	Isolated Mass	Theoretical Mass	Purified Mass	Isolated after Purification	Extrapolated Yield
85 mg	nd	nd	208.7 mg	550 mg	161.8 mg	25.8 mg	33.0 mg
0.49 mmol / g (41.6 µmol)	100% of total resin		35% Yield		78% of Isolated Peptide	16% Purification Yield	6.1% Overall Yield

Cleavage Method: Method 1

Purification Method: Method 1

Purified HPLC Method: Method 1

Purified LCMS Method: Method 1

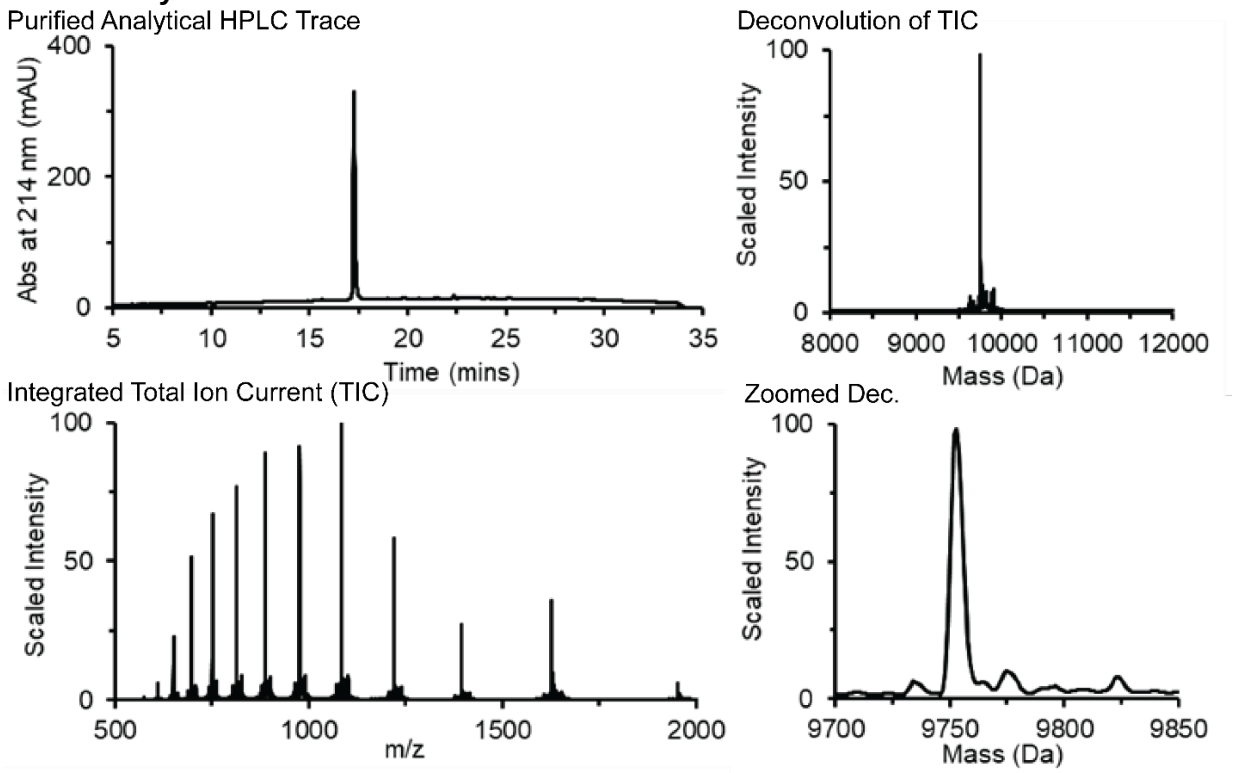
Calculated Mass: 10580 Da

Observed Mass: 10580 Da

2.4.11.12. Max-nb from AFPS

Protein	Uniprot ID	Start	End	Length	Modifications
Max-nb	P61244	23	102	83	C-Term +GGC + carboxamide
DKRAHHNALE RKRRDHIKDS FHSLRDSVPS LQGEKASRAQ ILDKATEYIQ YMRRKNHTHQ QDIDDLKRQN ALLEQQVRAL GGC					

L Max-nb Synthesis Results



Starting Resin	Crude Resin	Cleaved Resin	Isolated Mass	Theoretical Mass	Purified Mass	Isolated after Purification	Extrapolated Yield
85 mg	n.d.	n.d.	229.4 mg	515 mg	148.7 mg	30.4 mg	46.9 mg
0.49 mmol / g (41.6 µmol)	100% of total resin		45% Yield		65% of Isolated Peptide	20% Purification Yield	9.1% Overall Yield

Cleavage Method: Method 2

Purification Method: Method 2

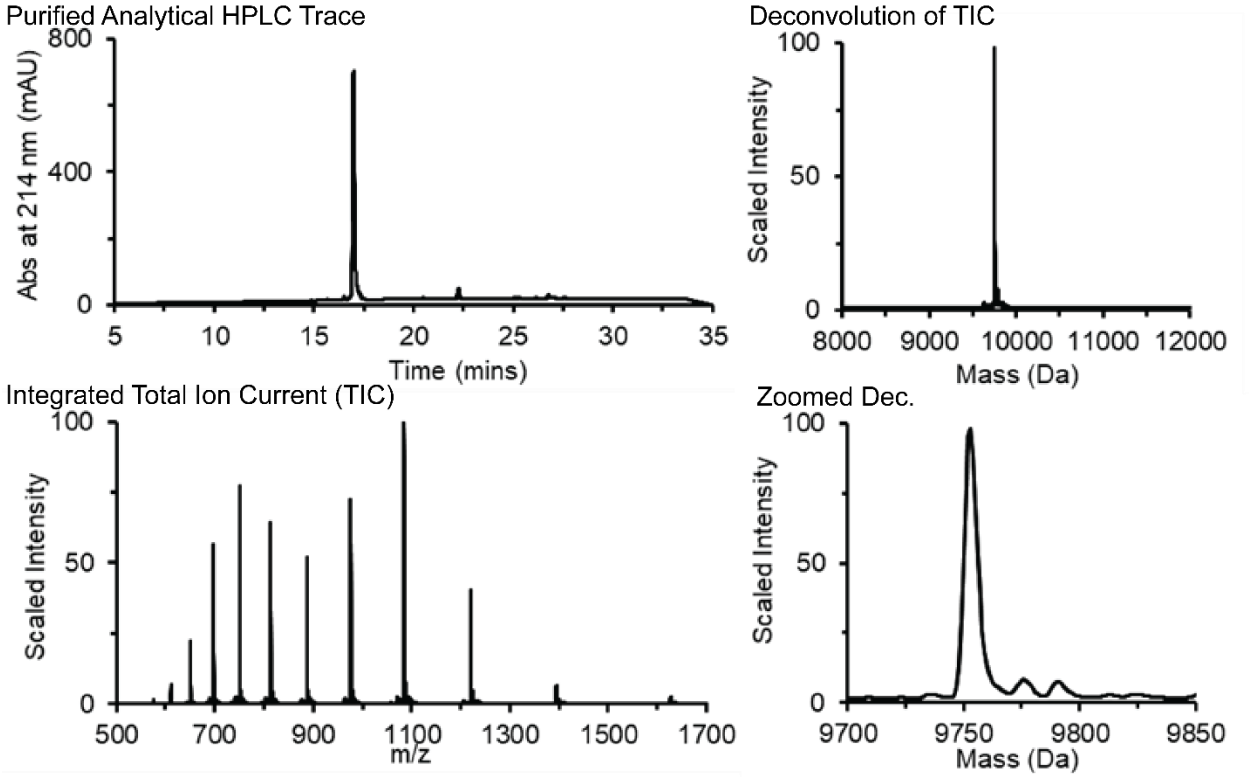
Purified HPLC Method: Method 1

Purified LCMS Method: Method 1

Calculated Mass: 9753 Da

Observed Mass: 9753 Da

D Max nb Synthesis Results



Starting Resin	Crude Resin	Cleaved Resin	Isolated Mass	Theoretical Mass	Purified Mass	Isolated after Purification	Extrapolated Yield
130 mg	n.d.	n.d.	76.8 mg	258.1 mg	76.8 mg	7.3 mg	14.6 mg
0.18 mmol / g (23.4 µmol)	50% of total resin		30% Yield		100% of Isolated Peptide	10% Purification Yield	2.8% Overall Yield

Cleavage Method: Method 1

Purification Method: Method 1

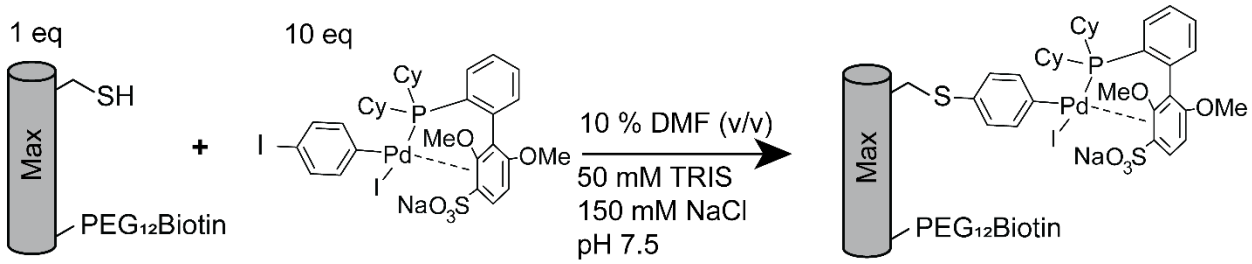
Purified HPLC Method: Method 1

Purified LCMS Method: Method 3

Calculated Mass: 9753 Da

Observed Mass: 9753 Da

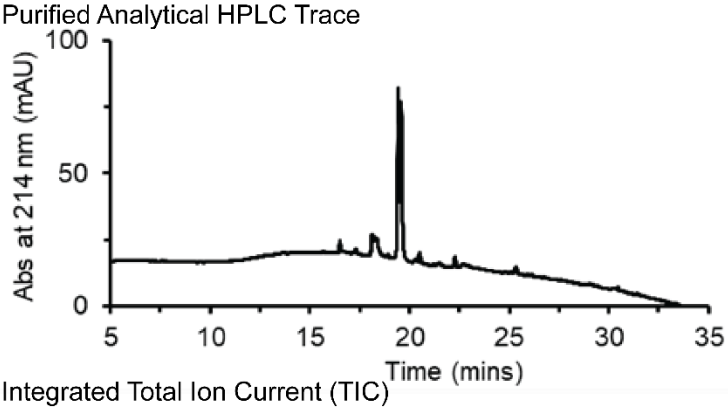
2.4.11.13. Max Oxidative Addition Complex Synthesis



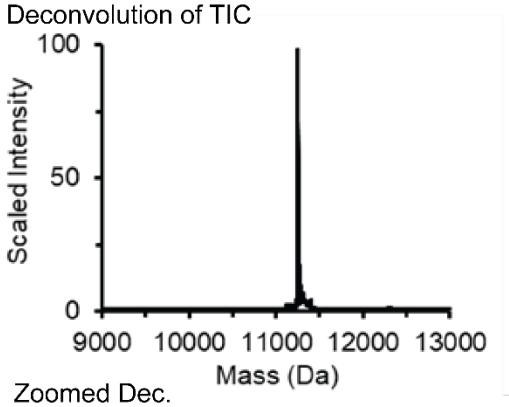
Synthesis of Max-OAC. The protocol for OAC generation was adapted from an earlier report(42). To a 50 mL Falcon tube was added the purified Max as a lyophilized powder and sufficient reaction buffer (50 mM TRIS HCl, 150 mM NaCl, pH 7.5) was added to prepare a solution at 2 mg / mL (205 µM). Protein concentration was measured with Method 4.1.1, and the concentration of the protein stock adjusted to 1.1 mg/mL (113 µM) with addition of reaction buffer. Separately, to a 2.0 mL Eppendorf tube was added 10 equiv of sSPhos reinsertion reagent relative to the protein, and dissolved in DMF to 4.8 mg / mL (6 mM). The sSPhos reinsertion reagent in DMF was added to the protein stock, the Falcon tube quickly recapped, and the apparatus vortexed. After incubation for 5 min, the Falcon tube screw cap was removed, and the mixture was transferred to a 20 mL syringe affixed with a 0.22 µM nylon syringe filter. The mixture was forced through the syringe filter into a fresh Falcon tube. The Falcon tube originally containing the un-filtered peptide solution was rinsed with an additional 2 mL of the reaction buffer that was then filtered using the same filtration apparatus into the Falcon tube containing the filtered peptide solution. Any bubbles generated were removed by centrifugation at 3220 rcf. The peptide mixture was then purified.

L Max Oxidative Addition Complex Results

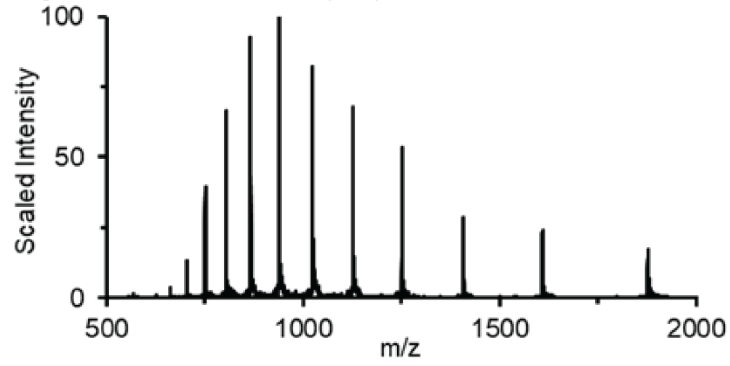
Purified Analytical HPLC Trace



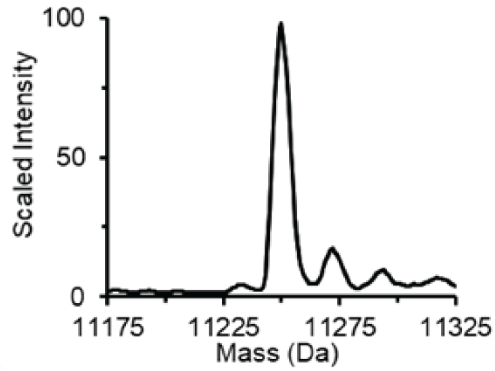
Deconvolution of TIC



Integrated Total Ion Current (TIC)



Zoomed Dec.



Starting Mass	Purified Mass	Isolated after Purification
11.5 mg	100% of reaction mix	2.2 mg
45% of iso L-Max		18.2% Yield

Purification Method: Method 3

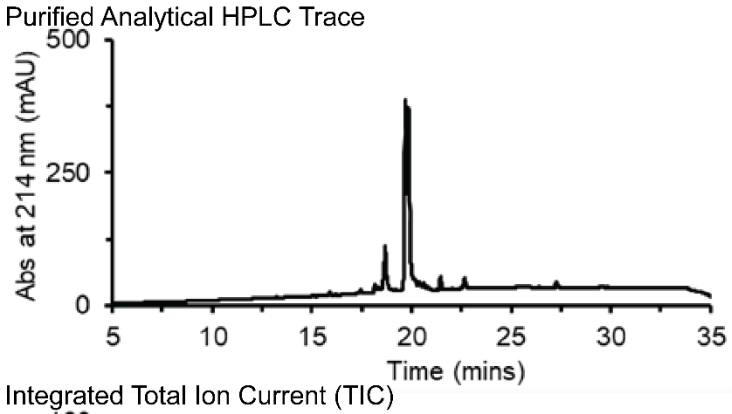
Purified HPLC Method: Method 3

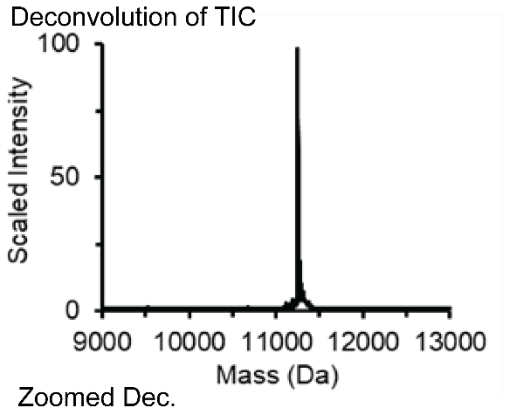
Purified LCMS Method: Method 1

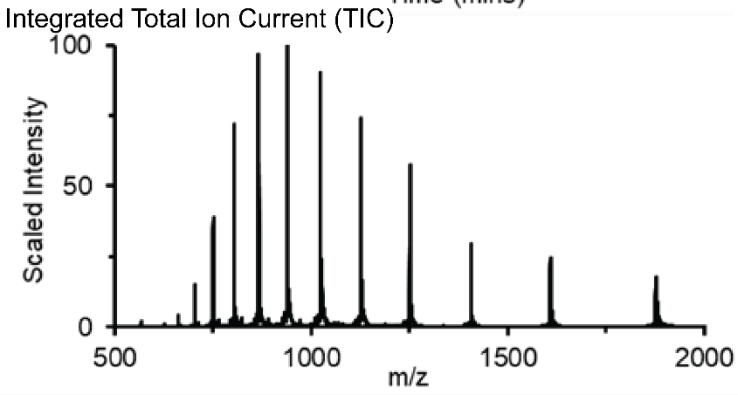
Calculated Mass: 11253 Da (Halide ligand to Pd is lost on ionization)

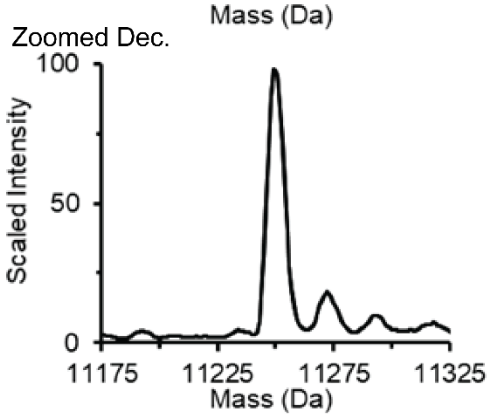
Observed Mass: 11253 Da

D Max Oxidative Addition Complex Results









Starting Mass	Purified Mass	Isolated after Purification
6.9 mg	100% of reaction mix	1.1 mg
100% of iso D-Max		15.2% Yield

Purification Method: Method 3

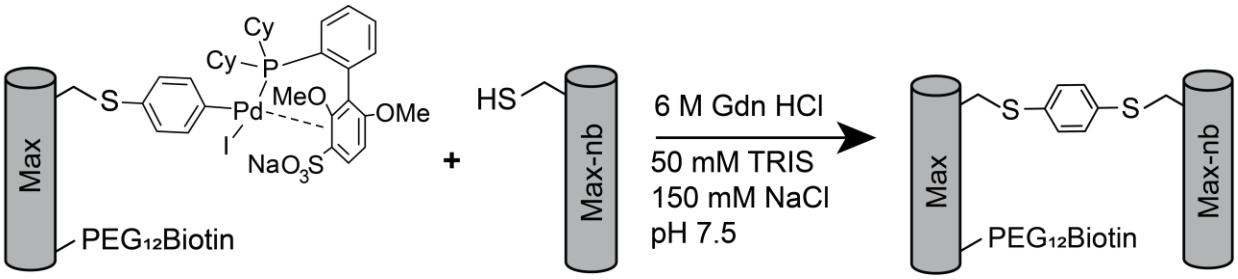
Purified HPLC Method: Method 3

Purified LCMS Method: Method 1

Calculated Mass: 11253 Da (Halide ligand to Pd is lost on ionization)

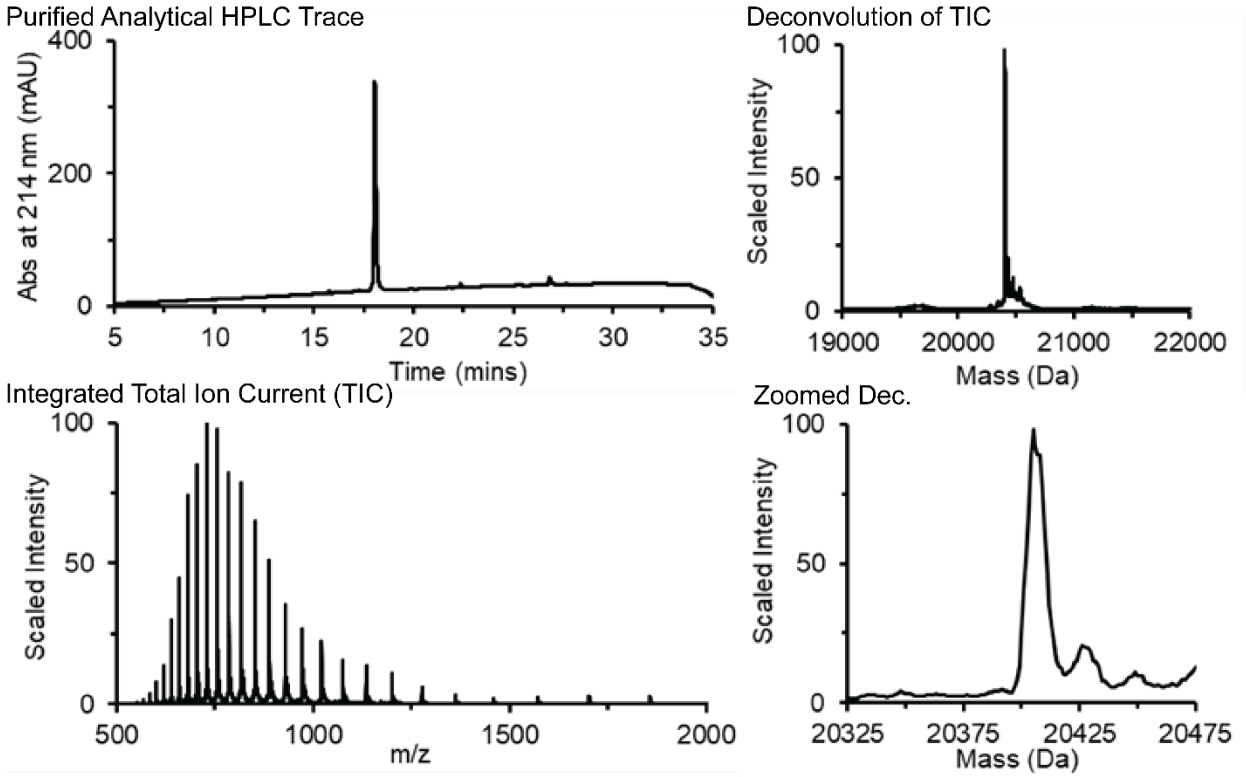
Observed Mass: 11253 Da

2.4.11.14. Max-Max-nb Conjugation



Synthesis of Max-Max. The protocol for protein dimerization was adapted from an earlier report(42). To a 2.0 mL low-protein binding Eppendorf tube was added 1 equiv of the Max-Pd complex as a lyophilized powder. Separately, to a 2.0 mL low-protein binding Eppendorf tube was added 1 equiv of the Max-nb as a lyophilized powder. Sufficient denaturing buffer (6 M Gdn HCl, 50 mM TRIS HCl, 150 mM NaCl, pH 7.5) was added to the Eppendorf tube containing the Max-nb to produce a solution at 10 mg/mL (1.03 mM). The solution containing the Max-nb was quickly transferred to the Eppendorf tube containing the Max-Pd complex. The now-empty Eppendorf rube was rinsed with 15 µL of denaturing buffer, and the solution added to the reaction mixture in the other Eppendorf tube. After incubation for 30 min, the Eppendorf tube containing the reaction mixture was clarified by centrifugation at 21,000 rcf for 10 min, and the supernatant purified.

L Max-Max-nb Dimerization Results



Starting Mass	Purified Mass	Isolated after Purification
2.2 mg	100% of reaction mix	1.2 mg
100% of iso L-Max-OAC		29.5% Yield

Purification Method: Method 3

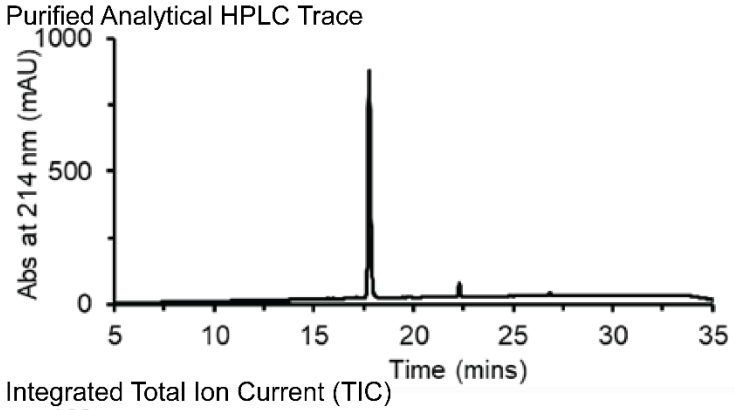
Purified HPLC Method: Method 3

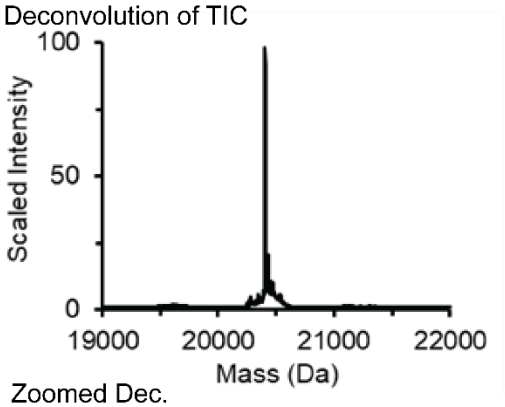
Purified LCMS Method: Method 1

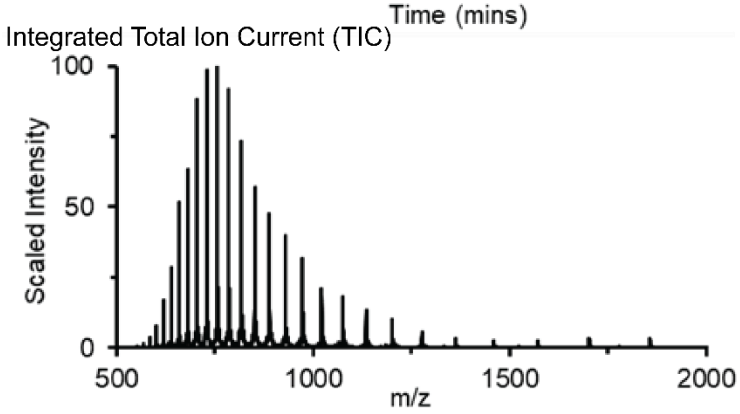
Calculated Mass: 20406 Da

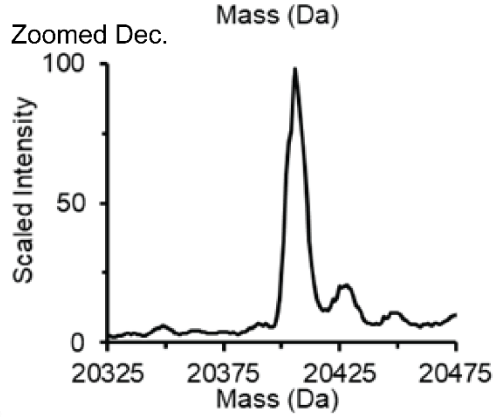
Observed Mass: 20406 Da

D Max-Max-nb Dimerization Results









Starting Mass	Purified Mass	Isolated after Purification
1.1 mg	100% of reaction mix	0.7 mg
100% of iso D-Max-OAC		34.4% Yield

Purification Method: Method 3

Purified HPLC Method: Method 3

Purified LCMS Method: Method 1

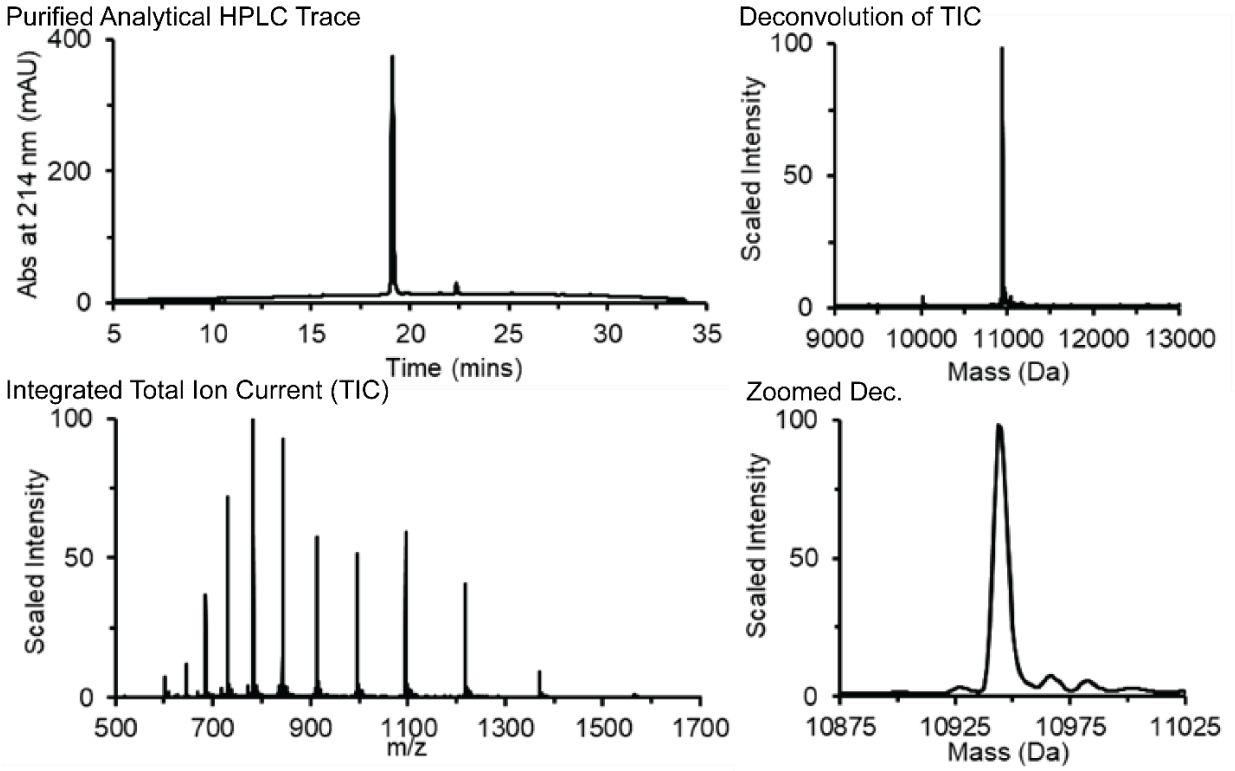
Calculated Mass: 20406 Da

Observed Mass: 20406 Da

2.4.11.15.Myc from AFPS

Protein	Uniprot ID	Start	End	Length	Modifications
Myc	P01106	353	427	85	N-Term PEG_{12}Biotin, C-Term +GGC + carboxamide
NVKRRTHNVL ERQRRNELKR SFFALRDQIP ELENNEKAPK VVILKKATAY ILSVQAEEQK LISEEDLLRK RREQLKHKLE QLGGC					

L Myc Synthesis Results



Starting Resin	Crude Resin	Cleaved Resin	Isolated Mass	Theoretical Mass	Purified Mass	Isolated after Purification	Extrapolated Yield
85 mg	n.d.	n.d.	163.7 mg	560 mg	163.7 mg	14.4 mg	14.4
0.49 mmol / g (41.6 µmol)	100% of total resin		29% Yield		100% of Isolated Peptide	9% Purification Yield	2.6% Overall Yield

Cleavage Method: Method 2

Purification Method: Method 2

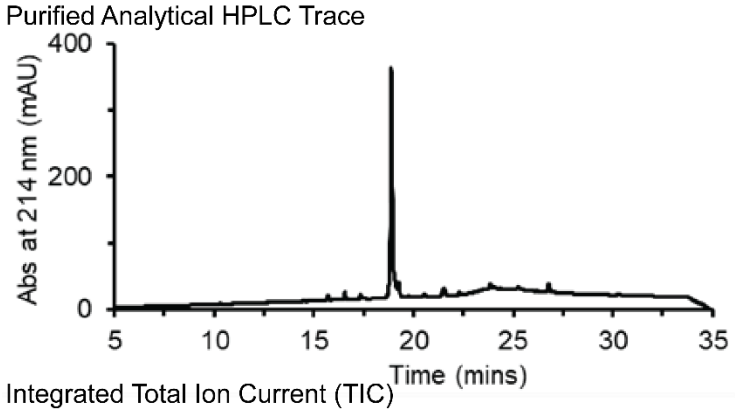
Purified HPLC Method: Method 3

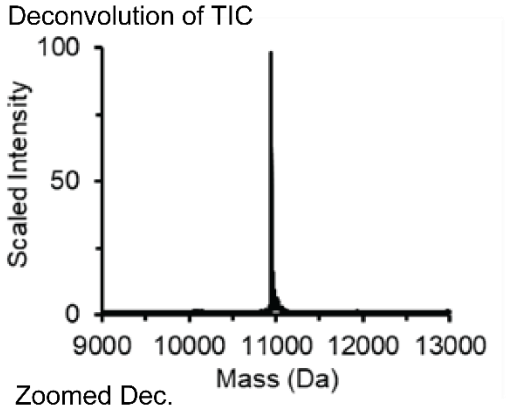
Purified LCMS Method: Method 3

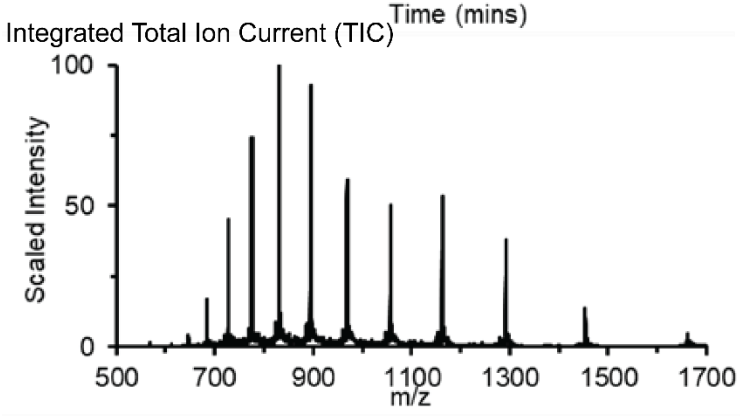
Calculated Mass: 10945 Da

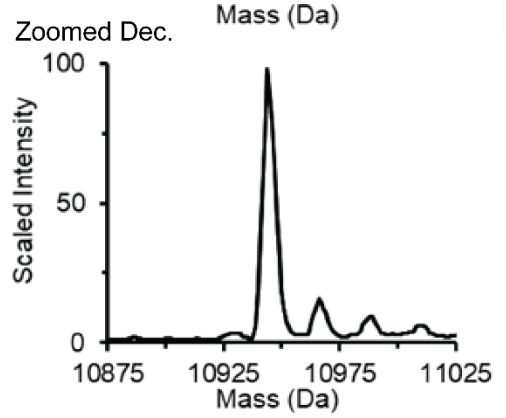
Observed Mass: 10945 Da

D Myc Synthesis Results









Starting Resin	Crude Resin	Cleaved Resin	Isolated Mass	Theoretical Mass	Purified Mass	Isolated after Purification	Extrapolated Yield
130 mg	n.d.	n.d.	93.8 mg	280.2 mg	93.8 mg	7.1 mg	14.2 mg
0.18 mmol / g (23.4 µmol)	50% of total resin		33% Yield		100% of Isolated Peptide	8% Purification Yield	2.5% Overall Yield

Cleavage Method: Method 1

Purification Method: Method 1

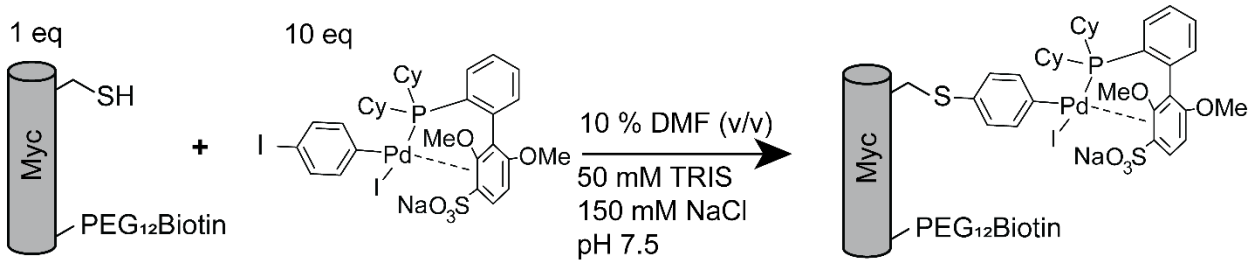
Purified HPLC Method: Method 3

Purified LCMS Method: Method 3

Calculated Mass: 10945 Da

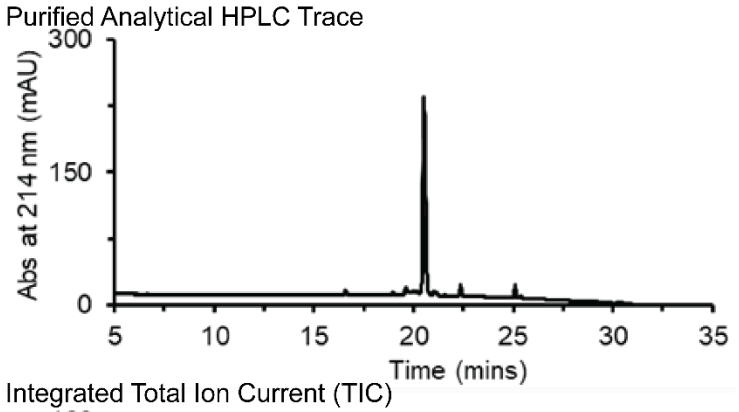
Observed Mass: 10945 Da

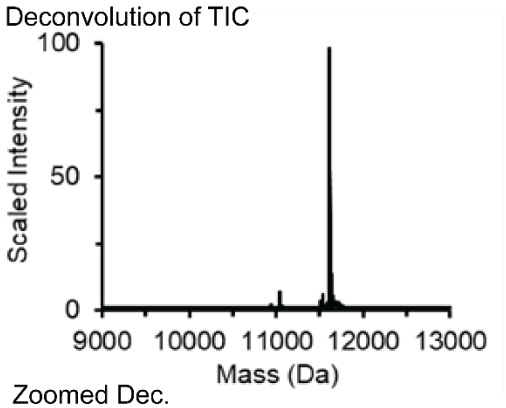
2.4.11.16. Myc Oxidative Addition Complex Synthesis

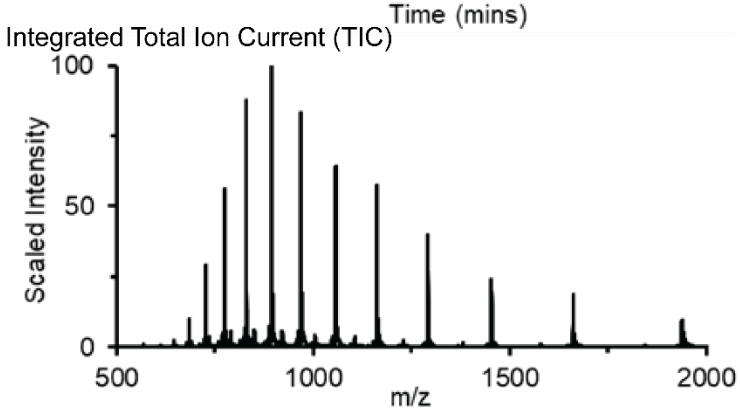


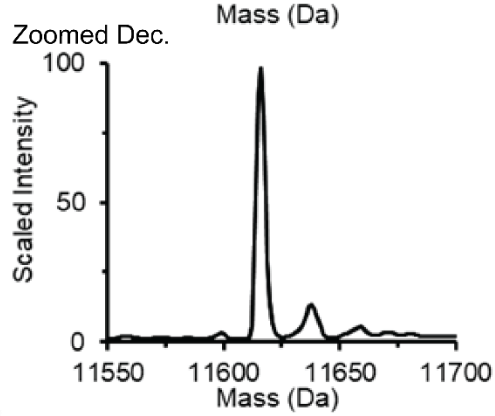
Synthesis of Myc-OAC. The protocol for OAC generation was adapted from an earlier report(42). To a 50 mL Falcon tube was added the purified Myc as a lyophilized powder and sufficient reaction buffer (50 mM TRIS HCl, 150 mM NaCl, pH 7.5) was added to prepare a solution at 2 mg/mL (205 µM). Protein concentration was measured with Method 4.1.1, and the concentration of the protein stock adjusted to 1.1 mg/mL (113 µM) with addition reaction buffer. Separately, to a 2.0 mL Eppendorf tube was added 10 eq of sSPhos reinsertion reagent relative to the protein, and dissolved in DMF to 4.8 mg/mL (6 mM). The sSPhos reinsertion reagent in DMF was added to the protein stock, the Falcon tube quickly recapped, and the apparatus vortexed. After incubation for 5 min, the Falcon tube screw cap was removed, and the mixture was transferred to a 20 mL syringe affixed with a 0.22 µM nylon syringe filter. The mixture was forced through the syringe filter into a fresh Falcon tube. The Falcon tube originally containing the un-filtered peptide solution was rinsed with an additional 2 mL of the reaction buffer that was then filtered using the same filtration apparatus into the Falcon tube containing the filtered peptide solution. Any bubbles generated were removed by centrifugation at 3220 rcf. The peptide mixture was then purified.

L Myc Oxidative Addition Complex Synthesis Results









Starting Mass	Purified Mass	Isolated after Purification
10.0 mg	100% of reaction mix	1.2 mg
69% of iso L-Myc		11.5% Yield

Purification Method: Method 3

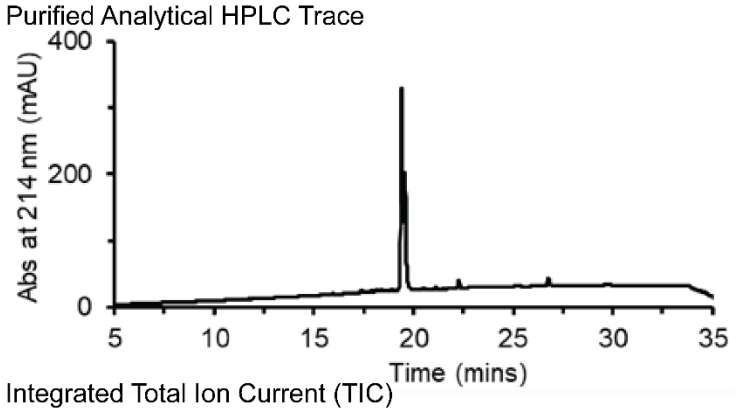
Purified HPLC Method: Method 3

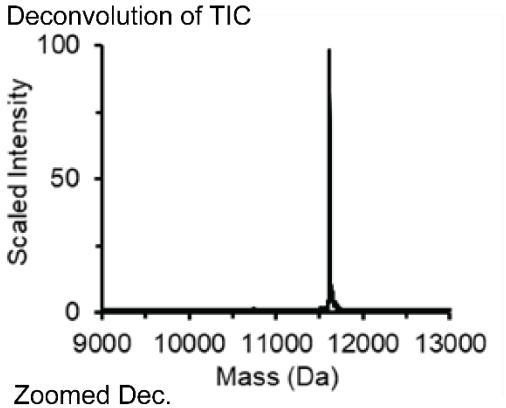
Purified LCMS Method: Method 1

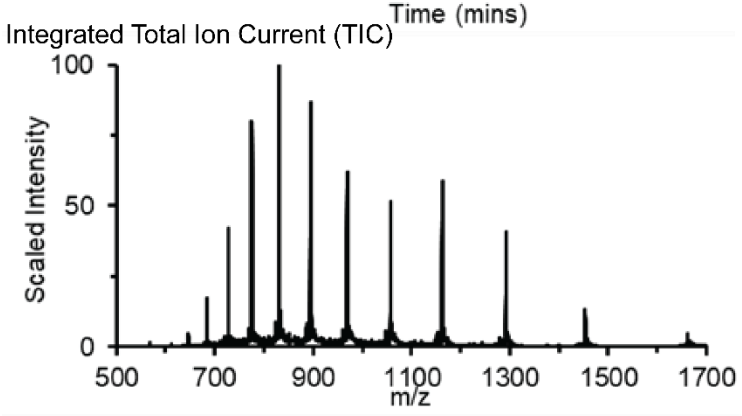
Calculated Mass: 11618 Da (Halide ligand to Pd is lost on ionization)

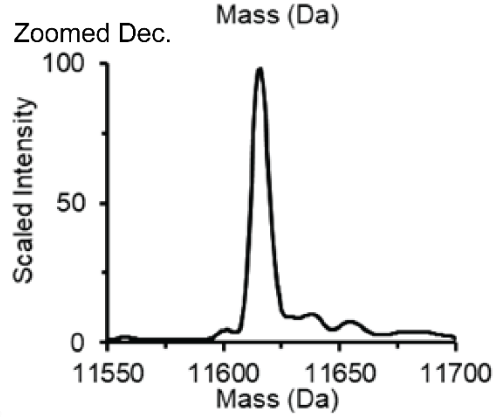
Observed Mass: 11618 Da

D Myc Oxidative Addition Complex Synthesis Results









Starting Mass	Purified Mass	Isolated after Purification
7.1 mg	100% of reaction mix	2.1 mg
100% of iso D-Myc		28.2% Yield

Purification Method: Method 3

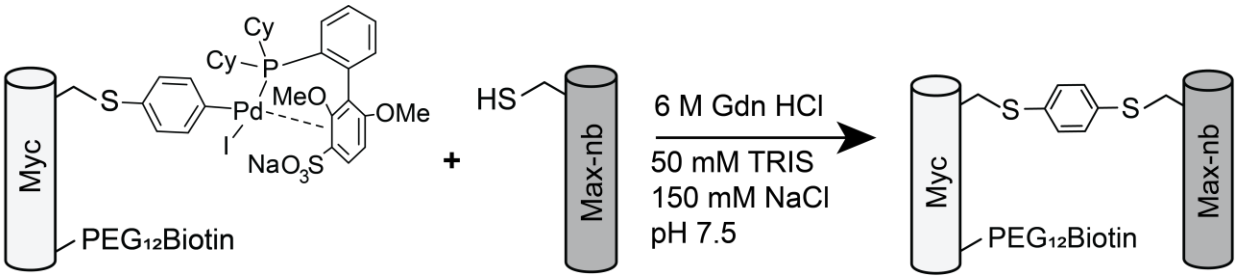
Purified HPLC Method: Method 3

Purified LCMS Method: Method 3

Calculated Mass: 11618 Da (Halide ligand to Pd is lost on ionization)

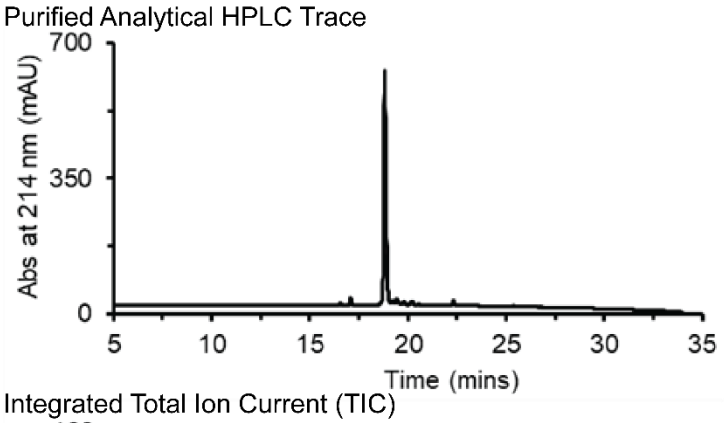
Observed Mass: 11618 Da

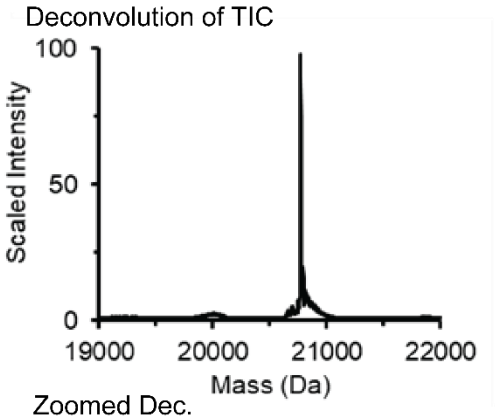
2.4.11.17. Myc Max-nb Conjugation

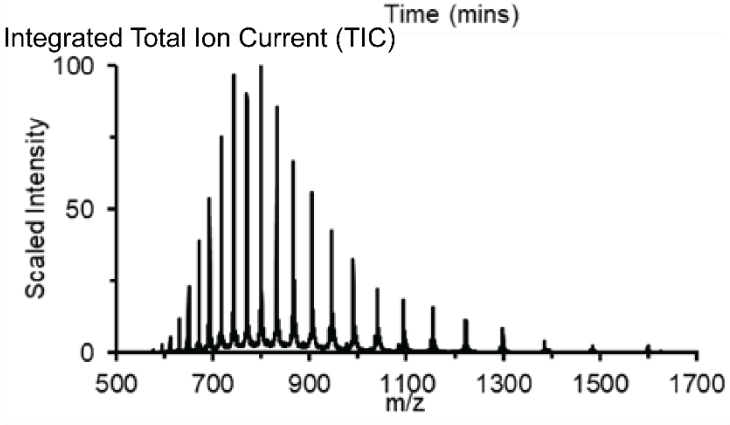


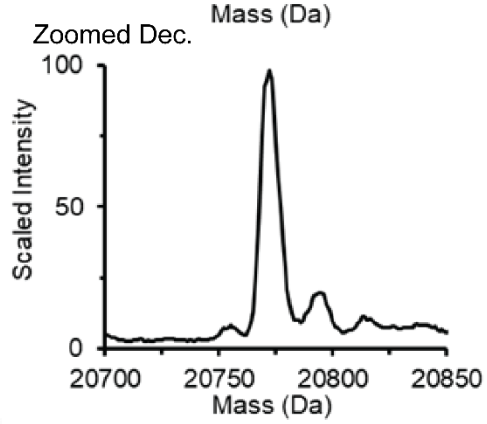
Synthesis of Myc-Max. The protocol for protein dimerization was adapted from an earlier report(42). To a 2.0 mL low-protein binding Eppendorf tube was added 1 equiv of the Myc-Pd complex as a lyophilized powder. Separately, to a 2.0 mL low-protein binding Eppendorf tube was added 1 equiv of the Max-nb as a lyophilized powder. Sufficient denaturing buffer (6 M Gdn HCl, 50 mM TRIS HCl, 150 mM NaCl, pH 7.5) was added to the Eppendorf tube containing the Max-nb to produce a solution at 10 mg/mL (1.03 mM). The solution containing the Max-nb was quickly transferred to the Eppendorf tube containing the Myc-Pd complex. The now-empty Eppendorf rube was rinsed with 15 µL of denaturing buffer, and the solution added to the reaction mixture in the other Eppendorf tube. After incubation for 30 min, the Eppendorf tube containing the reaction mixture was clarified by centrifugation at 21,000 rcf for 10 min, and the supernatant purified.

L Myc-Max-nb Dimerization Results









Starting Mass	Purified Mass	Isolated after Purification
1.2 mg	100% of reaction mix	0.6 mg
100% of iso L-Myc-OAC		27% Yield

Purification Method: Method 3

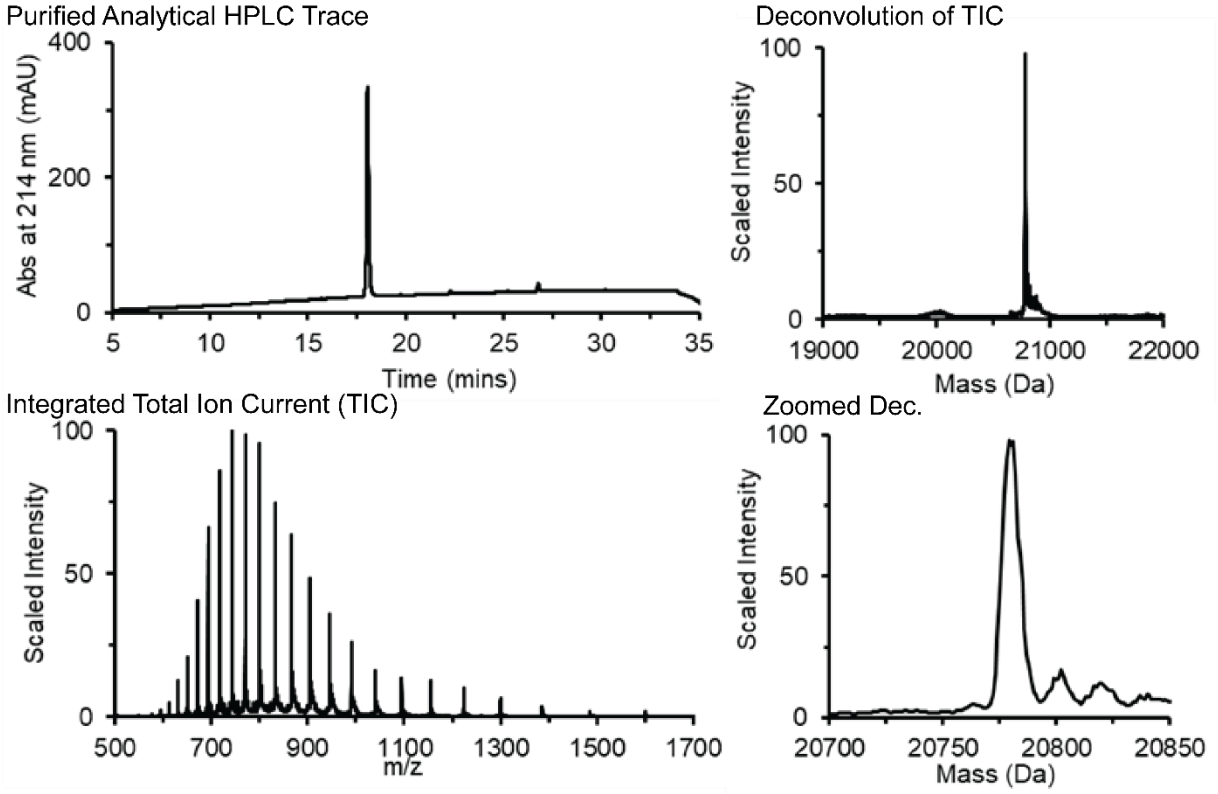
Purified HPLC Method: Method 3

Purified LCMS Method: Method 3

Calculated Mass: 20778 Da

Observed Mass: 20780 Da

D Myc-Max-nb Dimerization Results



Starting Mass	Purified Mass	Isolated after Purification
2.1 mg	100% of reaction mix	1.3 mg
100% of iso D-Myc-OAC		33.8% Yield

Purification Method: Method 3

Purified HPLC Method: Method 3

Purified LCMS Method: Method 3

Calculated Mass: 20778 Da

Observed Mass: 20780 Da

2.4.12. Collected List of Protein Sequences

Protein	Uniprot ID	Start	End	Length	Modifications
MDM2	Q00987	26	109	84	N-Term PEG_{12}Biotin, C-Term carboxamide
TLVRPKPLLL KLLKSVGAQK DTYTMKEVLF YLGQYIMTKR LYDEKQQHIV YCSNDLLGDL FGVPSFSVKE HRKIYTMIYR NLVV					
ERG	P11308	108	201	94	N-Term PEG_{12}Biotin, C-Term carboxamide
MEEKHMPPPN MTTNERRVIV PADPTLWSTD HVRQWLEWAV KEYGLPDVNI LLFQNIDGKE LCKMTKDDFQ RLTPSYNADI LLSHLHYLRE TPLP					
Barnase	P00648	48	157	110	N-Term PEG_{12}Biotin, C-Term carboxamide
AQVINTFDGV ADYLQTYHKL PDNYITKSEA QALGWVASKG NLADVAPGKS IGGDIFSNRE GKLPGKSGRT WREADINYTS GFRNSDRILY SSDWLIYKTT DHYQTFTKIR					
IRAK2	O43187	2	112	111	N-Term PEG_{12}Biotin, C-Term carboxamide
ACYIYQLPSW VLDDLCRNMD ALSEWDWMEF ASYVITDLTQ LRKIKSMERV QGVSITRELL WWWGMRQATV QQLVDLLCRL ELYRAAQIIL NWKPAPEIRC PIPAFPDSVK P					
CHIP	Q9UNE7	23	153	132	N-Term PEG_{12}Biotin, C-Term carboxamide
SPSAQELKEQ GNRLFVGRKY PEAAACYGRA ITRNPLVAVY YTNRALCYLK MQQHEQALAD CRRALELDGQ SVKAHFFLGQ CQLEMESYDE AIANLQRAYS LAKEQRLNFG DDIPSALRIA KKKRWNSIEE RR					
NEMO	Q9Y6K9	44	112	70	N-Term PEG_{12}Biotin, C-Term carboxamide, Nle in place of Met
EQGAPETLQR CLEENQELRD AIRQSNQILR ERCEELLHFQ ASQREEKEFL MCKFQEARKL VERLGLEKLE					
FKBP12	P62942	2	108	107	N-Term PEG_{12}Biotin, C-Term carboxamide
GVQVETISPG DGRTFPKRGQ TCVVHYTGML EDGKKFDSSR DRNKPFKFML GKQEVIRGWE EGVAQMSVGQ RAKLTISPDY AYGATGHPGI IPPHATLVFD VELLKLE					
BCL11a	Q9H165	740	835	96	N-Term PEG_{12}Biotin, C-Term carboxamide
RSDTCEYCGK VFKNCSNLTV HRRSHTGERP YKCELCNYAC AQSSKLTRHM KTHGQVGKDV YKCEICKMPF SVYSTLEKHM KKWHSDRVLN NDIKTE					
Yap1	P46937	163	276	114	N-Term PEG_{12}Biotin, C-Term carboxamide, Nle in place of Met
SSFEIPDDVP LPAGWEMAKT SSGQRYFLNH IDQTTTWQDP RKAMLSQMNV TAPTSPPVQQ NMMNSASGPL PDGWEQAMTQ DGEIYYINHK NKTTSWLDPR LDPRFAMNQR ISQS					
Nemo_iZip	*6MI4	4	124	121	N-Term PEG_{12}Biotin, C-Term carboxamide, Nle in place of Met
SVKELEDKNE ELLSEIAHLK NEVARLKKLL QRCLEENQEL RDAIRQSNQI LRERCEELLH FQASQREEKE FLMCKFQEAR KLVERLGLEK LELEDKNEEL LSEIAHLKNE VARLKKLVGE R					
Max	P61244	23	102	83	N-Term PEG_{12}Biotin, C-Term +GGC + carboxamide
DKRAHHNALE RKRRDHIKDS FHSLRDSVPS LQGEKASRAQ ILDKATEYIQ YMRRKNHTHQ QDIDDLKRQN ALLEQQVRAL GGC					
Myc	P01106	353	427	85	N-Term PEG_{12}Biotin, C-Term +GGC + carboxamide
NVKRRTHNVL ERQRRNELKR SFFALRDQIP ELENNEKAPK VVILKKATAY ILSVQAEEQK LISEEDLLRK RREQLKHKLE QLGGC					
Max-nb	P61244	23	102	83	C-Term +GGC + carboxamide
DKRAHHNALE RKRRDHIKDS FHSLRDSVPS LQGEKASRAQ ILDKATEYIQ YMRRKNHTHQ QDIDDLKRQN ALLEQQVRAL GGC					

Table 4 Collected List of Proteins Under Investigation

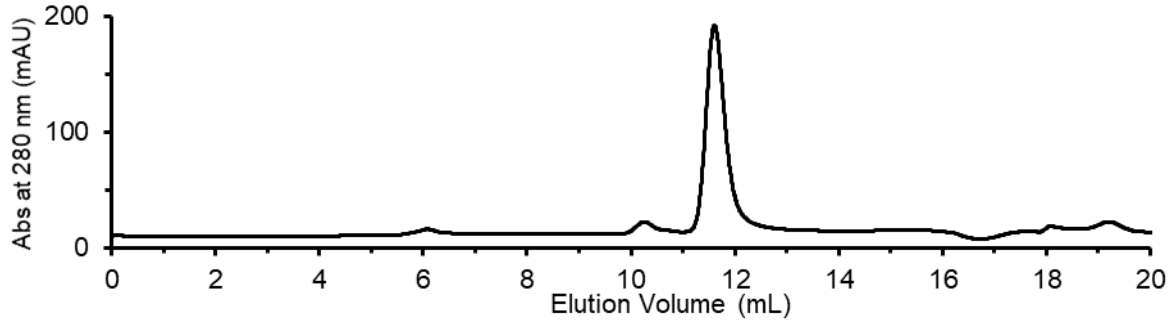
**PDB ID referenced in place of Uniprot ID.*

2.4.13. Protein Folding Results

2.4.13.1. MDM2 Folding

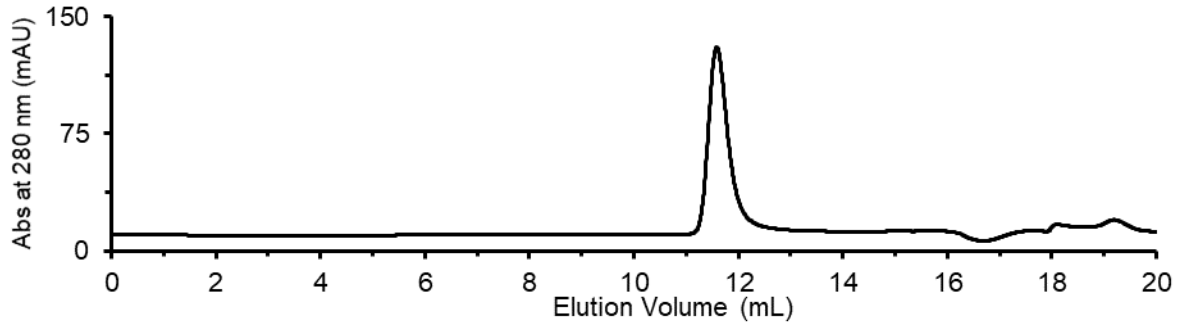
Synthetic MDM2 was folded in a similar manner to previous reports using ***Method 4.3*** *(24)*. After folding, the protein solution was transferred to an Amicon Ultra 15 mL 3K MWCO spin filtration unit, and the apparatus spun at 3220 rcf until the volume of the retentate was approximately 1 mL. The concentrated protein solution was transferred to an Amicon Ultra 0.5 mL 3K MWCO spin filtration unit, and the apparatus spun at 14,000 rcf until the volume of the retentate was approximately 100 µL. Separately, a semi-preparative SEC purification apparatus was prepared in the same manner as ***Method 4.2***, and the concentrated protein solution injected in total. The peak corresponding to the major product was collected, and concentrated to 100 µL protein with an Amicon Ultra 0.5 mL 3K MWCO spin filtration unit. The concentration of the solution was measured with ***Method 4.1.1***. The protein stock was separated into 1 nmol aliquots, flash frozen, and stored at -80 °C.

L MDM2 Folding Results



Starting Weight	Starting mol	Isolated Volume	Isolated [Protein] (mg/mL)	Isolated [Protein] (µM)	Isolated mol	Extrapolated Yield
1.2 mg	93.2 nmol	950 µL	0.11 mg / mL	10.3 µM	9.7 nmol	48.7 nmol
100% of Isolated Peptide	10.5% Folding Yield					

D MDM2 Folding Results

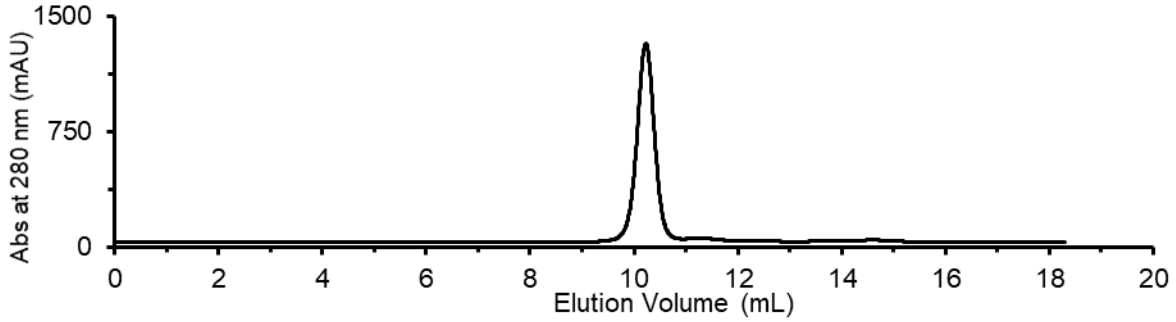


Starting Weight	Starting mol	Isolated Volume	Isolated [Protein] (mg/mL)	Isolated [Protein] (µM)	Isolated mol	Extrapolated Yield
0.6 mg	48 nmol	100 µL	0.49 mg / mL	45.5 µM	4.5 nmol	75.2 nmol
24% of Isolated Peptide	9.4% Folding Yield					

2.4.13.2. ERG Folding

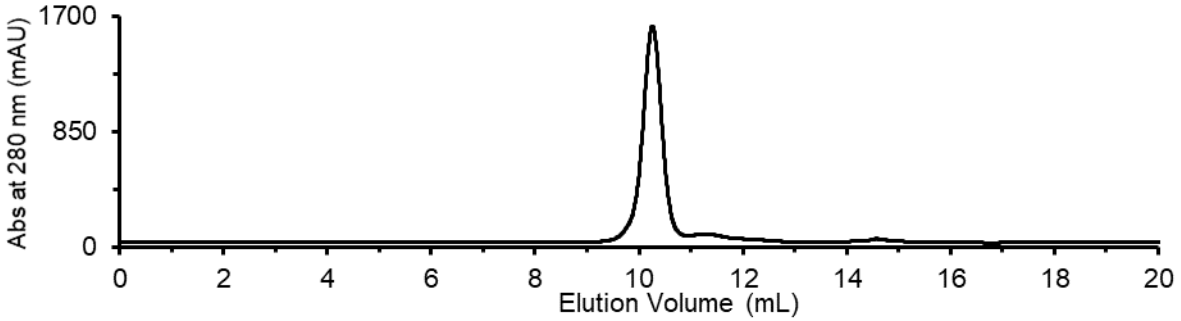
Synthetic ERG was folded according to ***Method 4.3***, and the resulting folded protein concentration measured with ***Method 4.1.1***. The protein stock was separated into 1 nmol aliquots, flash frozen, and stored at -80 °C.

L ERG Folding Results



Starting Weight	Starting mol	Isolated Volume	Isolated [Protein] (mg/mL)	Isolated [Protein] (µM)	Isolated mol	Extrapolated Yield
1.0 mg	74.0 nmol	938 µL	0.4 mg / mL	30.6 µM	28.7 nmol	361.5 nmol
15.9% of Isolated Peptide	38.8% Folding Yield					

D ERG Folding Results

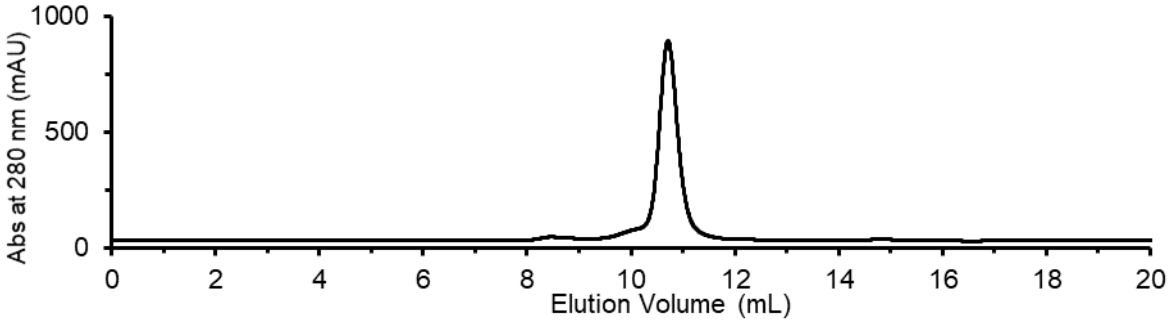


Starting Weight	Starting mol	Isolated Volume	Isolated [Protein] (mg/mL)	Isolated [Protein] (µM)	Isolated mol	Extrapolated Yield
1.0 mg	74.0 nmol	839 µL	0.5 mg / mL	44.9 µM	37.6 nmol	521.5 nmol
14.4% of Isolated Peptide	50.9% Folding Yield					

2.4.13.3. Barnase Folding

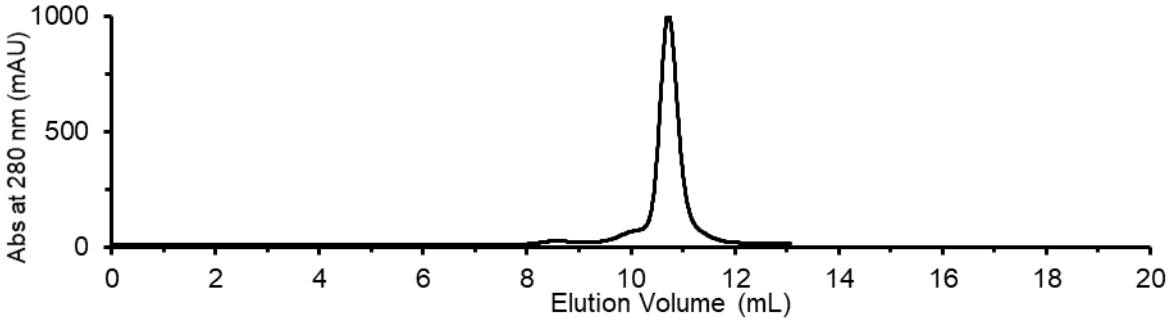
Synthetic Barnase was folded according to ***Method 4.3***, and the major peak corresponding to the major product was collected, and concentrated to 100 µL protein with an Amicon Ultra 0.5 mL 3K MWCO spin filtration unit. The concentration was measured with ***Method 4.1.1***. The protein stock was separated into 1 nmol aliquots, flash frozen, and stored at -80 °C.

L Barnase Folding Results



Starting Weight	Starting mol	Isolated Volume	Isolated [Protein] (mg/mL)	Isolated [Protein] (µM)	Isolated mol	Extrapolated Yield
0.6 mg	39.9 nmol	100 µL	0.8 mg / mL	58.2 µM	5.8 nmol	32.2 nmol
37.5% of Isolated Peptide	14.6% Folding Yield					

D Barnase Folding Results

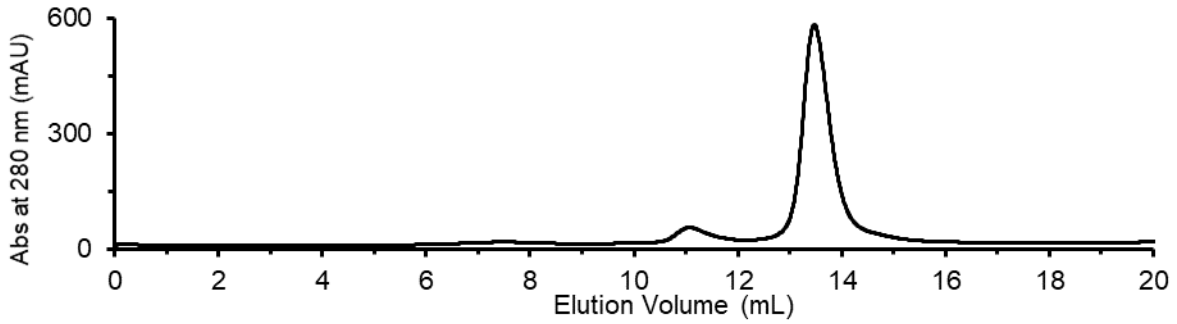


Starting Weight	Starting mol	Isolated Volume	Isolated [Protein] (mg/mL)	Isolated [Protein] (µM)	Isolated mol	Extrapolated Yield
0.6 mg	39.9 nmol	100 µL	1.4 mg / mL	105.9 µM	10.6 nmol	81.8 nmol
26.1% of Isolated Peptide	26.5% Folding Yield					

2.4.13.4. IRAK2 Folding

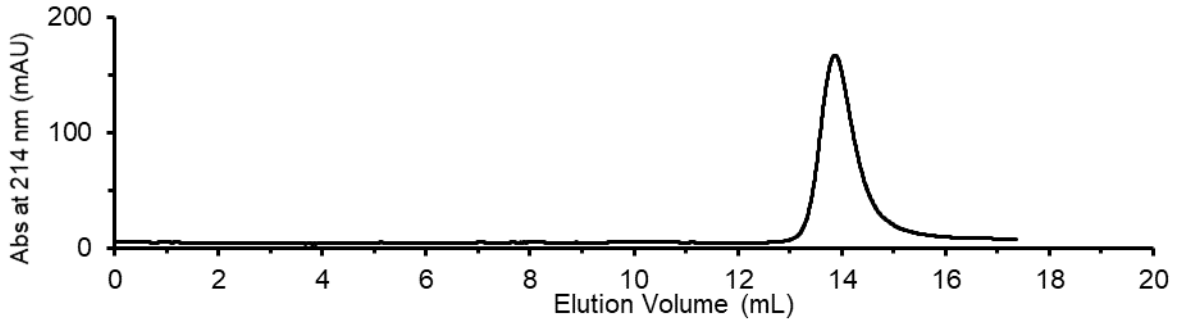
Synthetic IRAK2 was folded according to ***Method 4.3***, and the resulting folded protein concentration measured with ***Method 4.1.1***.

L IRAK2 Folding Results

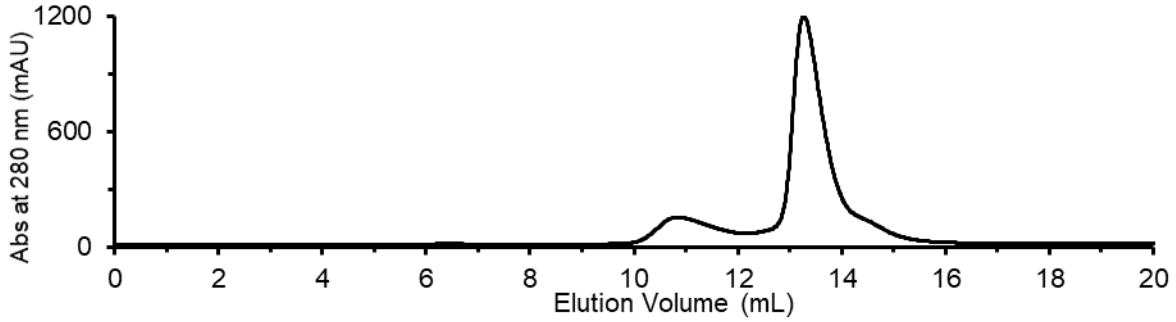


Starting Weight	Starting mol	Isolated Volume	Isolated [Protein] (mg/mL)	Isolated [Protein] (µM)	Isolated mol	Extrapolated Yield
0.6 mg	39.1 nmol	838 µL	0.1 mg / mL	8.4 µM	7.0 nmol	48.2 nmol
29.3% of Isolated Peptide	17.9% Folding Yield					

As there was a larger molecular weight shoulder that was removed upon the SEC purification, additional analytical SEC of the purified SEC fraction was recorded to verify its removal. The SEC purification performed as described in Method 4.3, with minor modifications: the injected sample was 1 nmol of folded L-IRAK2 not denatured, the chromatogram was recorded at 214 nm for increased sensitivity, and the running buffer was 50 mM TRIS, 150 mM NaCl, 5% Glycerol (v/v), pH 7.5.

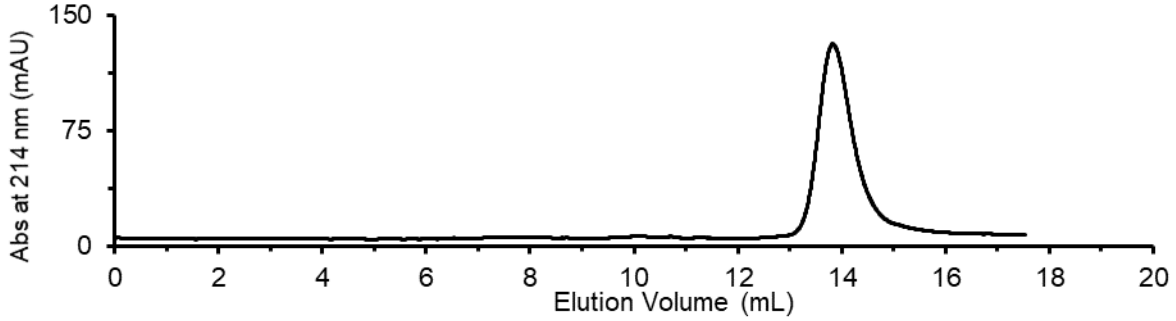


D IRAK2 Folding Results



Starting Weight	Starting mol	Isolated Volume	Isolated [Protein] (mg/mL)	Isolated [Protein] (µM)	Isolated mol	Extrapolated Yield
0.6 mg	39.1 nmol	1.15 mL	0.2 mg / mL	15.2 µM	17.4 nmol	66.6 nmol
100% of Isolated Peptide	44.6% Folding Yield					

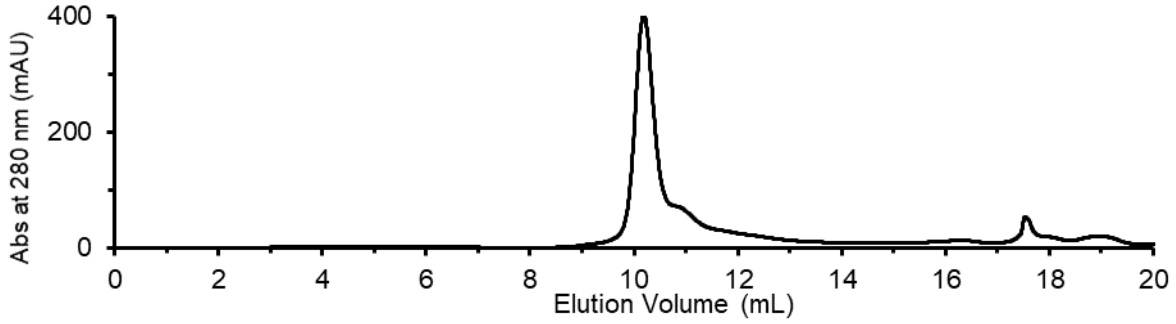
As there was a larger molecular weight shoulder that was removed upon the SEC purification, additional analytical SEC of the purified SEC fraction was recorded to verify its removal. The SEC purification performed as described in Method 4.3, with minor modifications: the injected sample was 1 nmol of folded D-IRAK2 not denatured, the chromatogram was recorded at 214 nm for increased sensitivity, and the running buffer was 50 mM TRIS, 150 mM NaCl, 5% Glycerol (v/v), pH 7.5.



2.4.13.5. CHIP Folding

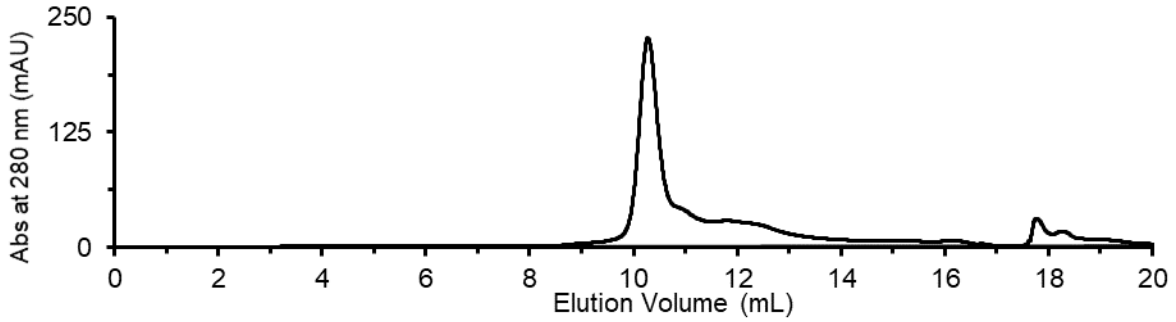
Synthetic CHIP was folded according to ***Method 4.3***, and the resulting folded protein concentration measured with ***Method 4.1.1***.

L CHIP Folding Results



Starting Weight	Starting mol	Isolated Volume	Isolated [Protein] (mg/mL)	Isolated [Protein] (µM)	Isolated mol	Extrapolated Yield
0.6 mg	32.2 nmol	1012 µL	0.2 mg / mL	14.2 µM	14.3 nmol	44.0 nmol
100% of Isolated Peptide	44.6% Folding Yield					

D CHIP Folding Results

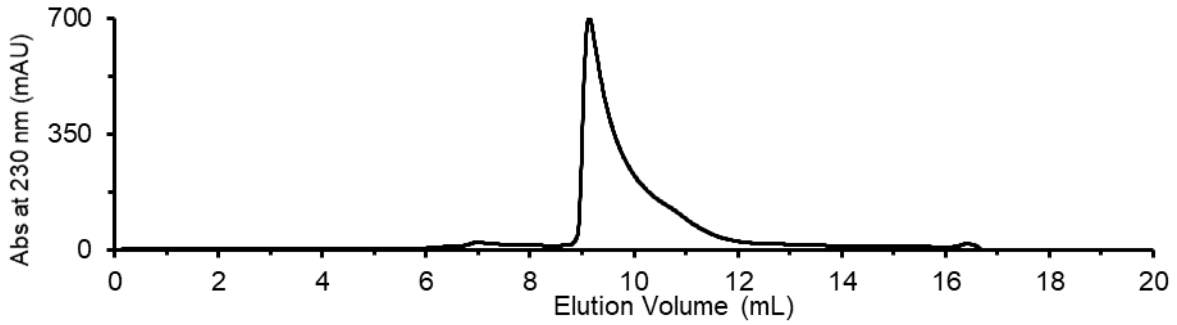


Starting Weight	Starting mol	Isolated Volume	Isolated [Protein] (mg/mL)	Isolated [Protein] (µM)	Isolated mol	Extrapolated Yield
1.6 mg	85.8 nmol	896 µL	0.4 mg / mL	23.5 µM	21.0 nmol	21.0 nmol
100% of Isolated Peptide	24.5% Folding Yield					

2.4.13.6. NEMO Folding

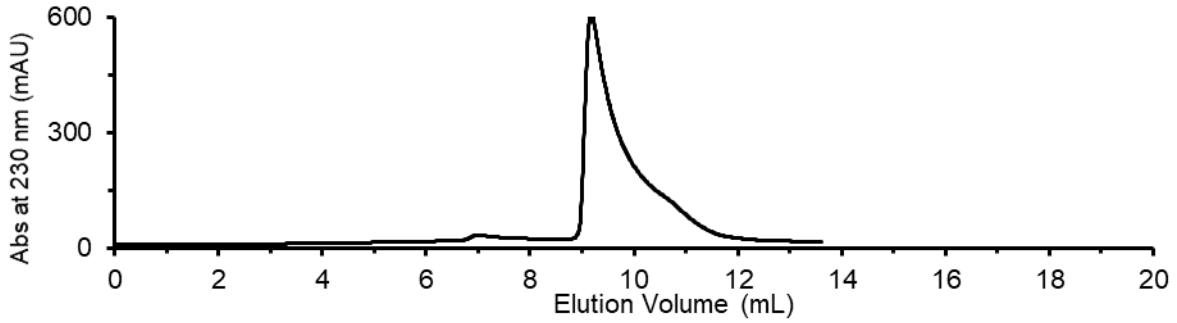
Synthetic NEMO was folded according to ***Method 4.3***, and the resulting folded protein concentration measured with ***Method 4.1.2***. Because of NEMO's low extinction coefficient at 280 nm, the SEC chromatogram was monitored at 230 nm.

L NEMO Folding Results



Starting Weight	Starting mol	Isolated Volume	Isolated [Protein] (mg/mL)	Isolated [Protein] (µM)	Isolated mol	Extrapolated Yield
1.6 mg	149.0 nmol	633 µL	0.4 mg / mL	41.5 µM	26.3 nmol	92.6 nmol
100.0% of Isolated Peptide	17.6% Folding Yield					

D NEMO Folding Results

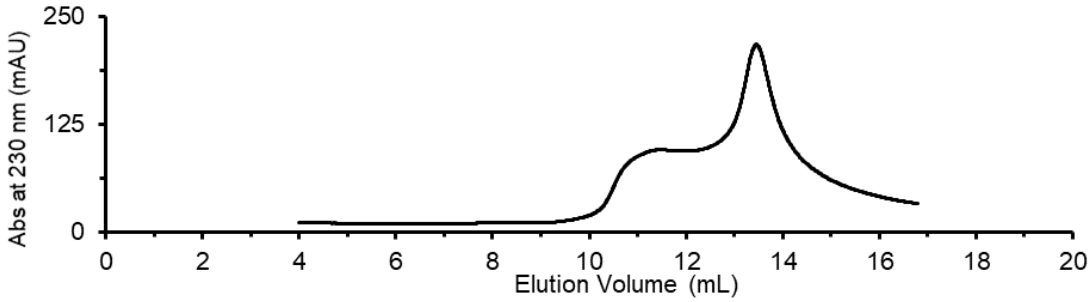


Starting Weight	Starting mol	Isolated Volume	Isolated [Protein] (mg/mL)	Isolated [Protein] (µM)	Isolated mol	Extrapolated Yield
2.5 mg	232 nmol	1200 µL	0.3 mg / mL	30.7 µM	36.8 nmol	191.5 nmol
38.5% of Isolated Peptide	15.8% Folding Yield					

2.4.13.7. FKBP12 Folding

Synthetic FKBP12 was folded according to ***Method 4.3***, and the resulting folded protein concentration measured with ***Method 4.1.1***.

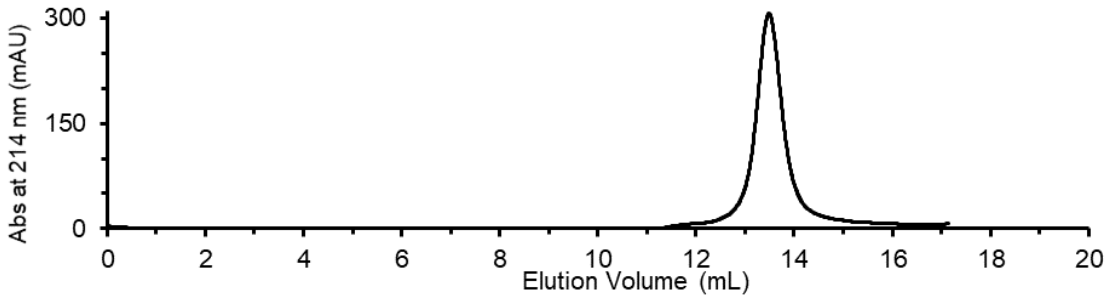
L FKBP12 Folding Results



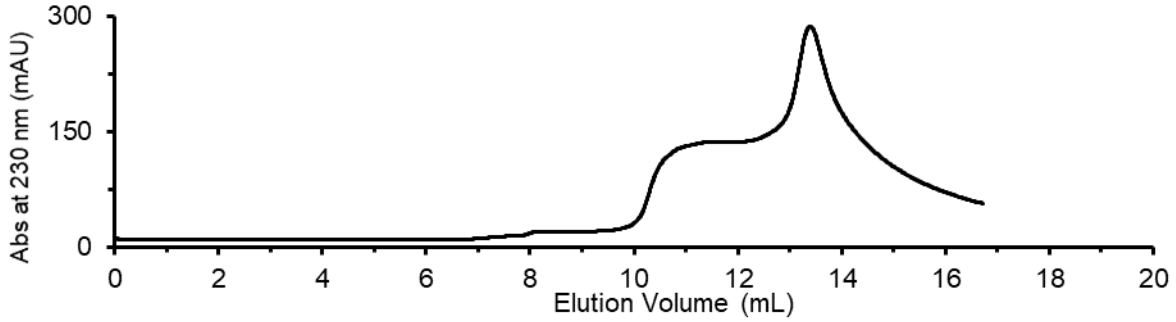
Starting Weight	Starting mol	Isolated Volume	Isolated [Protein] (mg/mL)	Isolated [Protein] (µM)	Isolated mol	Extrapolated Yield
2.4 mg	165.2 nmol	1000 µL	0.3 mg / mL	20.1 µM	20.1 nmol	62.6 nmol
100% of Isolated Peptide	12.2% Folding Yield					

*An error occurred during initiation of the UV recorder, and data was only collected an unknown amount of time after the run had started. As a result, the absolute elution time of the above peak is unknown. An additional buffer of 4 mL is added to the beginning of the run time to account for the recorder error. This value of 4 mL was estimated as the offset to account for the time required to restart the analysis software, and results in a peak elution time that is in good agreement with the subsequent analytical SEC trace.

As there was a larger molecular weight shoulder that was removed upon the SEC purification, additional analytical SEC of the purified SEC fraction was recorded to verify its removal. The SEC purification performed as described in Method 4.3, with minor modifications: the injected sample was 1 nmol of folded L-FKBP12 not denatured, the chromatogram was recorded at 214 nm for increased sensitivity, and the running buffer was 50 mM TRIS, 150 mM NaCl, 5% Glycerol (v/v), pH 7.5.

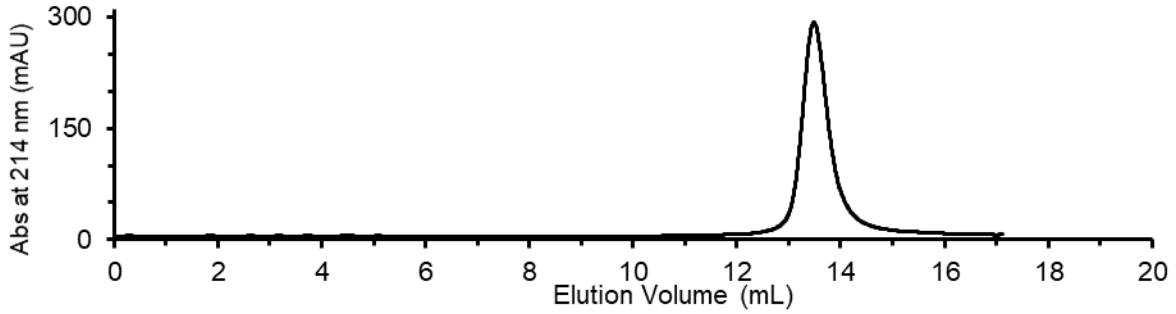


D FKBP12 Folding Results



Starting Weight	Starting mol	Isolated Volume	Isolated [Protein] (mg/mL)	Isolated [Protein] (µM)	Isolated mol	Extrapolated Yield
2.2 mg	151.4 nmol	870 µL	0.4 mg / mL	29.2 µM	25.4 nmol	63.6 nmol
100% of Isolated Peptide	16.8% Folding Yield					

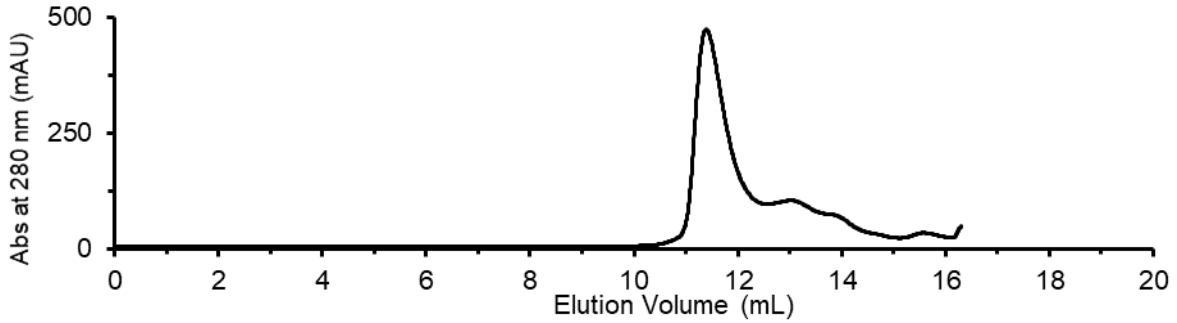
As there was a larger molecular weight shoulder that was removed upon the SEC purification, additional analytical SEC of the purified SEC fraction was recorded to verify its removal. The SEC purification performed as described in Method 4.3, with minor modifications: the injected sample was 1 nmol of folded D-FKBP12 not denatured, the chromatogram was recorded at 214 nm for increased sensitivity, and the running buffer was 50 mM TRIS, 150 mM NaCl, 5% Glycerol (v/v), pH 7.5.



2.4.13.8. BCL11A Folding

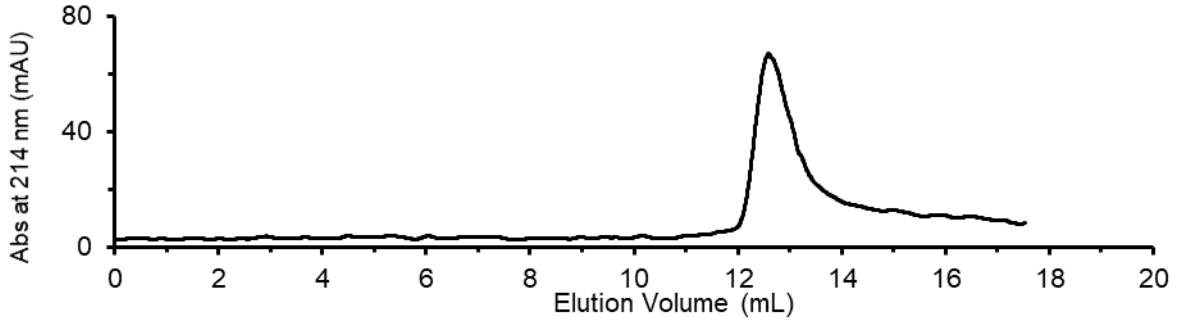
Synthetic BCL11a was folded according to ***Method 4.3***, and the resulting folded protein concentration measured with ***Method 4.1.1***. The denaturing buffer and running buffer were adjusted to include $ZnCl_2$. Denaturing buffer: 8 M urea, 100 mM TRIS HCl, 150 mM NaCl, 5 mM $ZnCl_2$, pH 7.5 Running Buffer: 100 mM TRIS HCl, 150 mM NaCl, 0.5 mM $ZnCl_2$, pH 7.5.

L BCL11A Folding Results

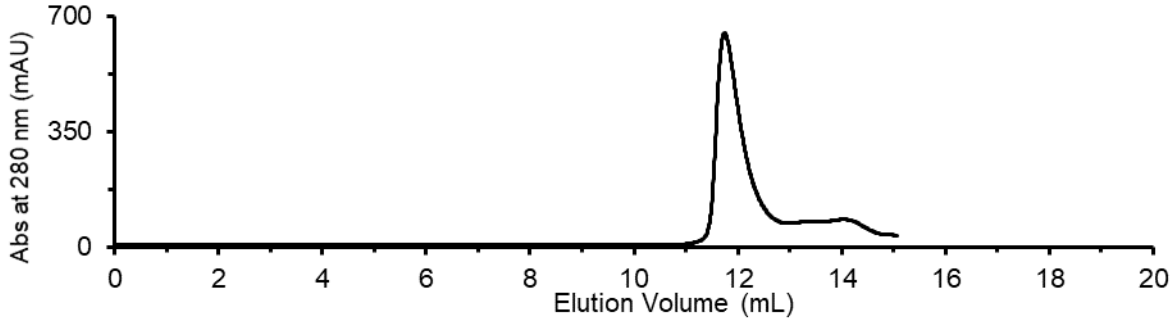


Starting Weight	Starting mol	Isolated Volume	Isolated [Protein] (mg/mL)	Isolated [Protein] (µM)	Isolated mol	Extrapolated Yield
2.5 mg	169.4 nmol	782 µL	0.7 mg / mL	55.3 µM	43.3 nmol	180.1 nmol
48.1% of Isolated Peptide	25.5% Folding Yield					

As there was a lower molecular weight shoulder that was removed upon the SEC purification, additional analytical SEC of the purified SEC fraction was recorded to verify its removal. The SEC purification performed as described in Method 4.3, with minor modifications: the injected sample was 1 nmol of folded L-BCL11a not denatured, the chromatogram was recorded at 214 nm for increased sensitivity, and the running buffer was 100 mM TRIS, 150 mM NaCl, 0.5 mM ZnCl2, pH 7.5.

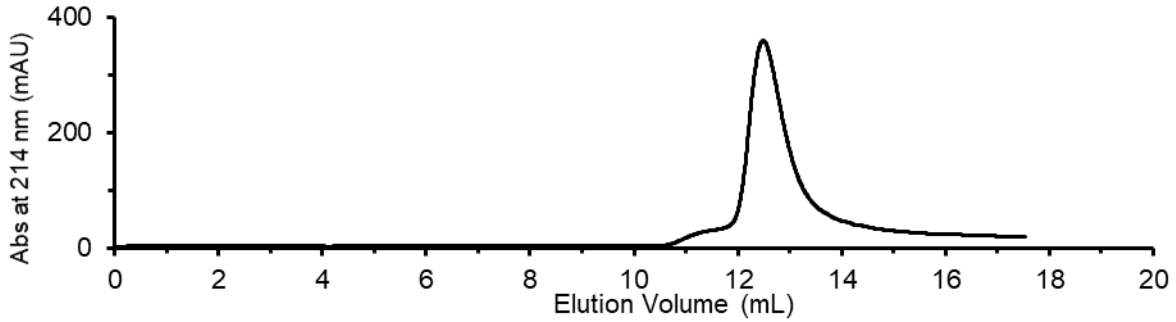


D BCL11A Folding Results



Starting Weight	Starting mol	Isolated Volume	Isolated [Protein] (mg/mL)	Isolated [Protein] (µM)	Isolated mol	Extrapolated Yield
2.4 mg	162.6 nmol	785 µL	0.4 mg / mL	36.4 µM	28.6 nmol	66.7 nmol
42.9% of Isolated Peptide	17.6% Folding Yield					

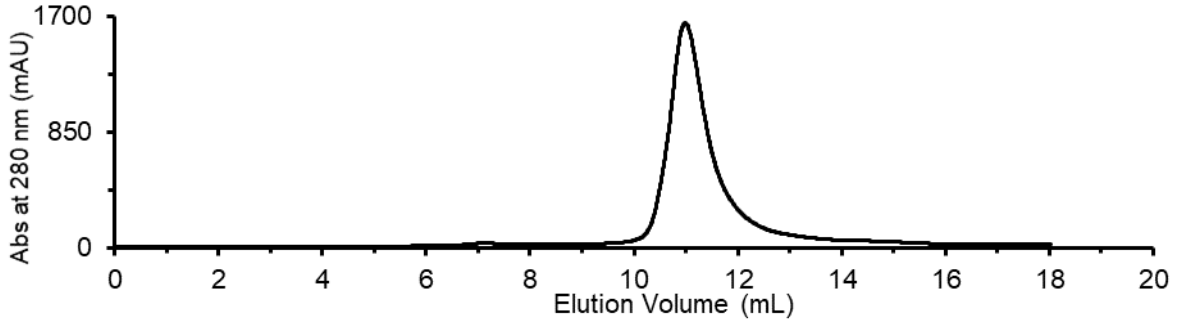
As there was a lower molecular weight shoulder that was removed upon the SEC purification, additional analytical SEC of the purified SEC fraction was recorded to verify its removal. The SEC purification performed as described in Method 4.3, with minor modifications: the injected sample was 1 nmol of folded D-BCL11a not denatured, the chromatogram was recorded at 214 nm for increased sensitivity, and the running buffer was 100 mM TRIS, 150 mM NaCl, 0.5 mM ZnCl2, pH 7.5.



2.4.13.9.YAP1 Folding

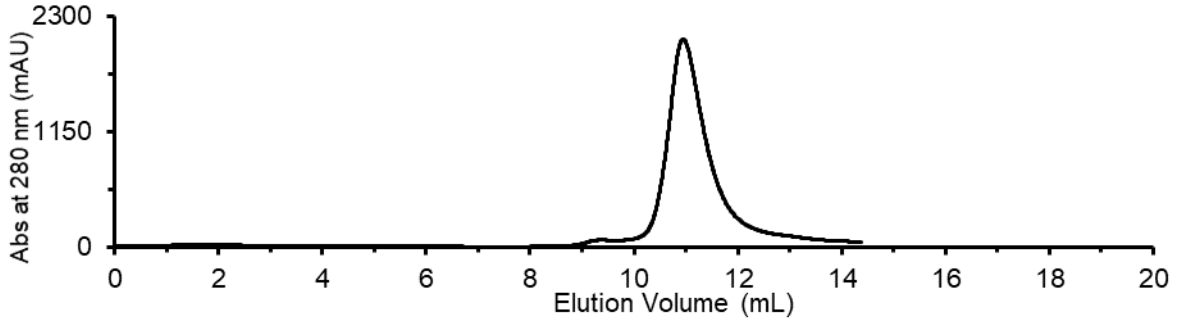
Synthetic YAP1 was folded according to ***Method 4.3***, and the resulting folded protein concentration measured with ***Method 4.1.1***.

L YAP1 Folding Results



Starting Weight	Starting mol	Isolated Volume	Isolated [Protein] (mg/mL)	Isolated [Protein] (µM)	Isolated mol	Extrapolated Yield
1.1 mg	73.1 nmol	786 µL	0.7 mg / mL	54.0 µM	42.4 nmol	106.7 nmol
100% of Isolated Peptide	58.0% Folding Yield					

D YAP1 Folding Results

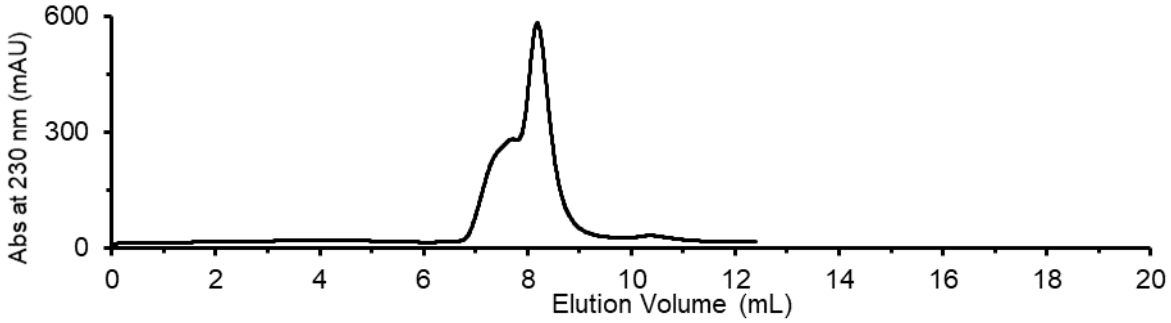


Starting Weight	Starting mol	Isolated Volume	Isolated [Protein] (mg/mL)	Isolated [Protein] (µM)	Isolated mol	Extrapolated Yield
2.6 mg	172.9 nmol	650 µL	1.0 mg / mL	72.1 µM	46.9 nmol	109.5 nmol
100% of Isolated Peptide	27.1% Folding Yield					

2.4.13.10. NEMO_iZIP Folding

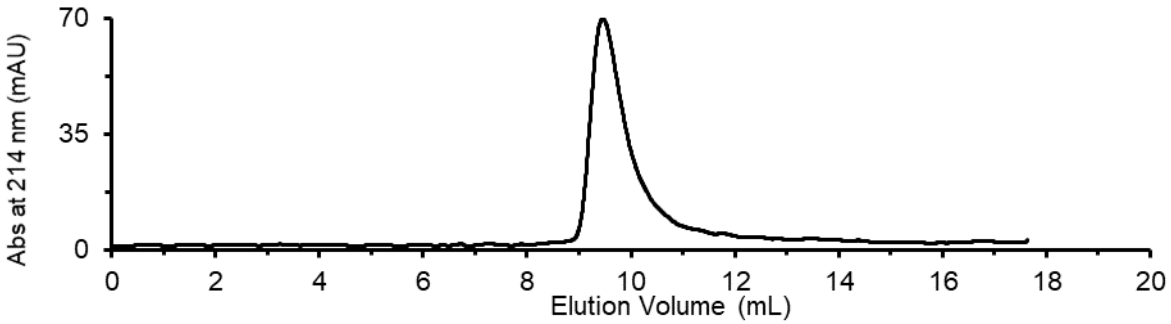
Synthetic NEMO_iZIP was folded according to ***Method 4.3***, and the resulting folded protein concentration measured with ***Method 4.1.2***. Because of NEMO_iZIP's low extinction coefficient at 280 nm, the SEC chromatogram was monitored at 230 nm.

L NEMO_iZIP Folding Results

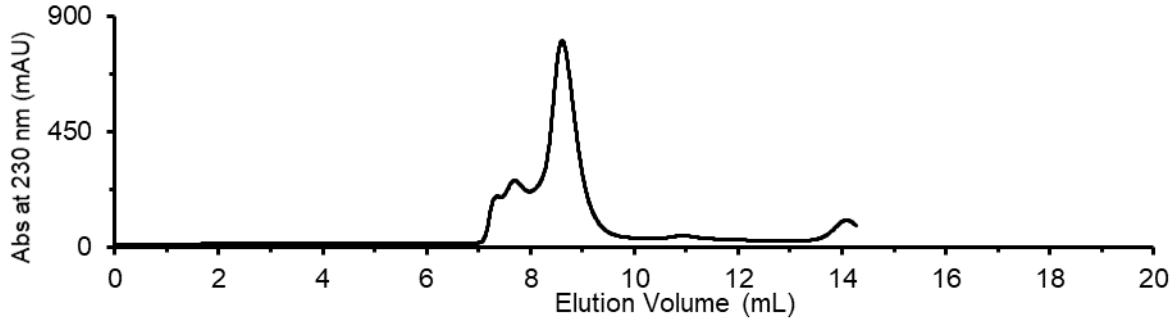


Starting Weight	Starting mol	Isolated Volume	Isolated [Protein] (mg/mL)	Isolated [Protein] (µM)	Isolated mol	Extrapolated Yield
0.9 mg	49.1 nmol	468 µL	0.2 mg / mL	15.7 µM	7.3 nmol	14.7 nmol
100% of Isolated Peptide	14.9% Folding Yield					

As there was a larger molecular weight shoulder that was removed upon the SEC purification, additional analytical SEC of the purified SEC fraction was recorded to verify its removal. The SEC purification performed as described in ***Method 4.3***, with minor modifications: the injected sample was 1 nmol of folded NEMO_iZIP not denatured, the chromatogram was recorded at 214 nm for increased sensitivity, and the running buffer was 50 mM TRIS, 150 mM NaCl, 5% Glycerol (v/v), pH 7.5.

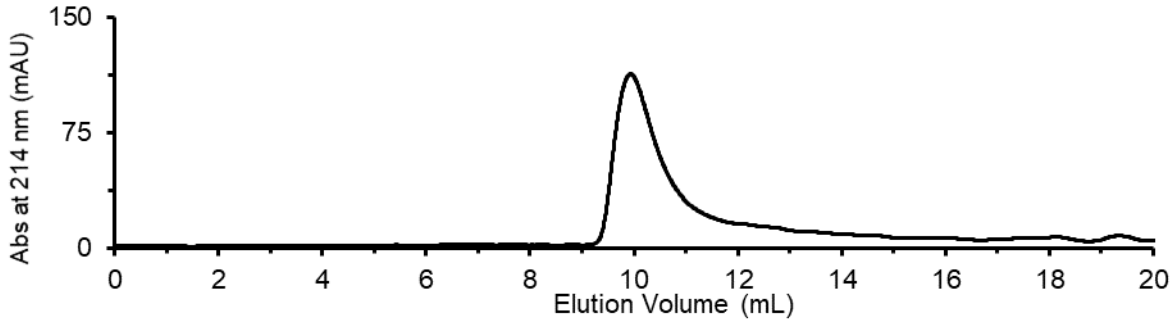


D NEMO_iZIP Folding Results



Starting Weight	Starting mol	Isolated Volume	Isolated [Protein] (mg/mL)	Isolated [Protein] (µM)	Isolated mol	Extrapolated Yield
2.0 mg	109.1 nmol	688 µL	0.2 mg / mL	15.5 µM	10.6 nmol	21.3 nmol
100% of Isolated Peptide	9.8% Folding Yield					

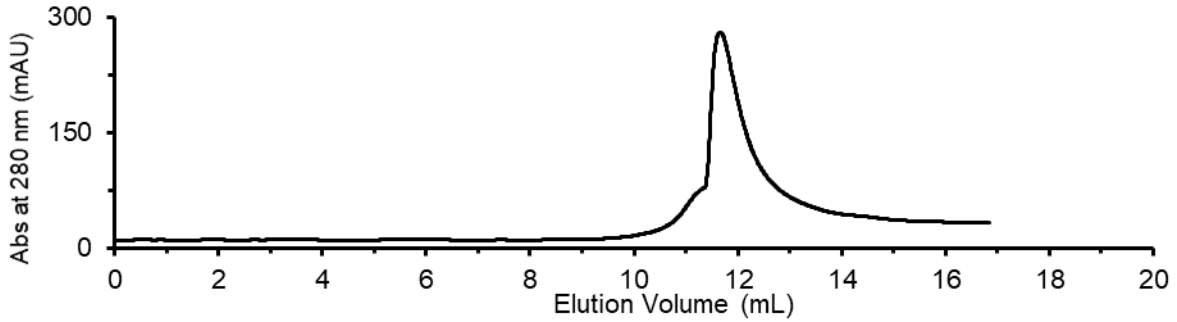
As there was a larger molecular weight shoulder that was removed upon the SEC purification, additional analytical SEC of the purified SEC fraction was recorded to verify its removal. The SEC purification performed as described in ***Method 4.3***, with minor modifications: the injected sample was 1 nmol of folded NEMO_iZIP not denatured, the chromatogram was recorded at 214 nm for increased sensitivity, and the running buffer was 50 mM TRIS, 150 mM NaCl, 5% Glycerol (v/v), pH 7.5.



2.4.13.11. Max-Max-nb Folding

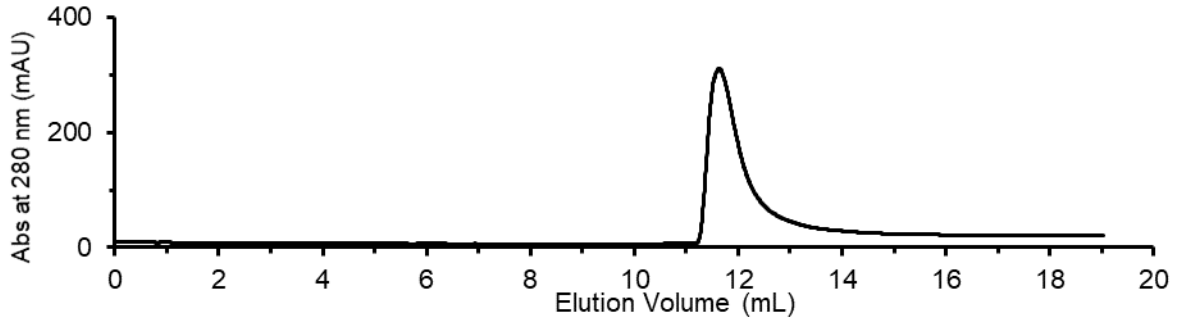
Synthetic Max-Max was folded according to ***Method 4.3***, and the resulting folded protein concentration measured with ***Method 4.1.1***.

L Max-Max-nb Folding Results



Starting Weight	Starting mol	Isolated Volume	Isolated [Protein] (mg/mL)	Isolated [Protein] (µM)	Isolated mol	Extrapolated Yield
1.2 mg	46.8 nmol	1074 µL	0.15 mg / mL	7.4 µM	7.9 nmol	
100% of Isolated Peptide	16.9% Folding Yield					

D Max-Max-nb Folding Results

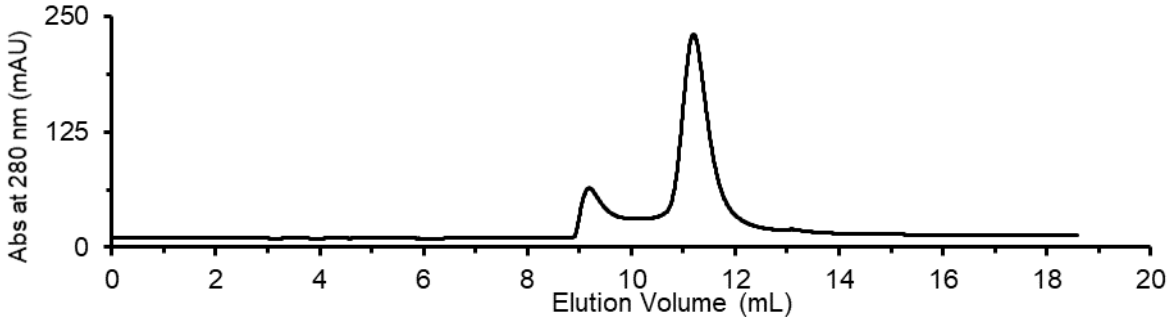


Starting Weight	Starting mol	Isolated Volume	Isolated [Protein] (mg/mL)	Isolated [Protein] (µM)	Isolated mol	Extrapolated Yield
0.7 mg	27.3 nmol	1237 µL	0.16 mg / mL	7.8 µM	9.6 nmol	
100% of Isolated Peptide	35.3% Folding Yield					

2.4.13.12. Myc-Max-nb Folding

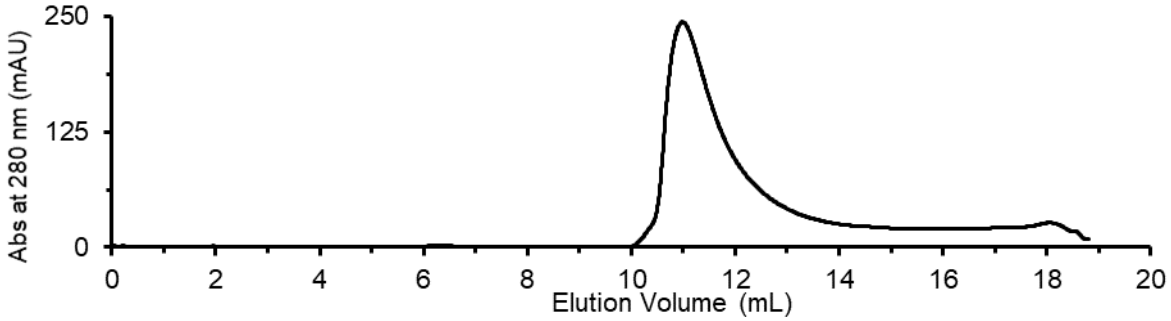
Synthetic Max-Max was folded according to ***Method 4.3***, and the resulting folded protein concentration measured with ***Method 4.1.1***.

L Myc-Max-nb Folding Results



Starting Weight	Starting mol	Isolated Volume	Isolated [Protein] (mg/mL)	Isolated [Protein] (µM)	Isolated mol	Extrapolated Yield
0.6 mg	23.2 nmol	866 µL	0.3 mg / mL	14.5 µM	12.6 nmol	
100% of Isolated Peptide	54.4% Folding Yield					

D Myc-Max-nb Folding Results



Starting Weight	Starting mol	Isolated Volume	Isolated [Protein] (mg/mL)	Isolated [Protein] (µM)	Isolated mol	Extrapolated Yield
1.3 mg	50.2 nmol	1128 µL	0.38 mg / mL	18.0 µM	20.4 nmol	
100% of Isolated Peptide	40.6% Folding Yield					

2.4.14. Biological Assay Techniques

2.4.14.1. BCL11A binding to DNA via TR-FRET

[DNA] (M)	Recombinant BCL11a to D-DNA		Synthetic L-BCL11a to D-DNA		Synthetic D-BCL11a to L-DNA	
	Replicate 1	Replicate 2	Replicate 1	Replicate 2	Replicate 1	Replicate 2
1.00E-06	0.957	1.04	1.09	0.91	0.99	1.01
3.33E-07	0.769	0.778	0.773	0.778	0.817	0.856
1.11E-07	0.495	0.496	0.542	0.555	0.603	0.651
3.70E-08	0.156	0.158	0.219	0.217	0.378	0.362
1.20E-08	0.0275	0.0362	0.151	0.141	0.153	0.155
4.00E-09	0.00618	0.00831	0.0412	0.0404	0.0565	0.052
1.00E-09	-0.00178	0.00209	0.0111	0.0113	0.0172	0.0137
0.00E+00	-0.00051	0.00051	0.00112	-0.00112	0.00087	-0.00087

Table 5 Normalized TR-TRET signal for BCL11A binding to DNA

For TR-FRET measurement of BCL11a binding to DNA, biotinylated, L, D or recombinant BCL11a was diluted to 100 nM, Alexa Fluor™ 488 labeled L- and D-DNA was diluted to 2 µM and Terbium-labeled streptavidin (Cis-Bio) was diluted to 20 nM in assay buffer (10 mM HEPES, 2.5 mM MgCl2, 0.05 mM EDTA, 10% glycerol, 50 mM NaCl, 50 ng/mL poly(dI-dC), 0.1% Triton X-100, 1 mM dithiothreitol, pH 7.5). DNA was serially diluted in assay buffer and 20 µL was added to a black 384-well plate (Costar) containing 20 µL BCL11a protein (final DNA top concentration of 1 µM). Samples were incubated for 30 min at room temperature. TR-FRET signal was determined using a LanthaScreen™ filter on a PheraStar (BMG Biotech) plate reader (Ex: 337 nm; Em1: 490 nM; Em2: 520 nM). The ratio of Em520 to Em490 was calculated, multiplied by 10,000, and plotted against compound concentration. Resulting data was normalized and fit to a 1:1 binding curve.

2.4.14.2. CHIP Binding to HSP peptide via TR-FRET

[HSP Peptide] (M)	Recombinant CHIP to L-P53 Peptide		Synthetic L-CHIP to L-HSP Peptide		Synthetic D-CHIP to D-HSP Peptide	
	Replicate 1	Replicate 2	Replicate 1	Replicate 2	Replicate 1	Replicate 2
1.00E-05	1.03	0.974	1.05	0.947	1.04	0.956
3.33E-06	0.331	0.308	0.385	0.368	0.411	0.43
1.11E-06	0.113	0.0987	0.152	0.155	0.173	0.157
3.70E-07	0.0406	0.037	0.0651	0.0665	0.0782	0.0759
1.23E-07	0.0136	0.0132	0.0285	0.0275	0.0337	0.0303
4.10E-08	0.00441	0.00468	0.012	0.012	0.0118	0.0141
1.40E-08	0.00137	0.00371	0.00471	0.00345	0.00365	0.00656
0.00E+00	-3.34E-06	0.00000334	0.000057	-0.000057	0.0000161	-0.0000161

Table 6 Normalized TR-FRET signal for CHIP binding to HSP peptide

For TR-FRET measurement of CHIP binding to HSP peptide, biotinylated, L, D or recombinant CHIP was diluted to 20 nM, L- and D-FAM-labeled HSP peptide was diluted to 20 μM and Terbium-labeled streptavidin (Cis-Bio) was diluted to 20 nM in assay buffer (10 mM HEPES pH 7.5, 150 mM NaCl, 0.05% Tween 20, 1 mM DTT). HSP peptides were serially diluted in assay buffer (final top concentration of 10 μM) and 20 μL was added to a black 384-well plate (Costar) containing 20 μL of CHIP protein. Samples were incubated for 30 min at room temperature. TR-FRET signal was determined using a LanthaScreen™ filter on a PheraStar (BMG Biotech) plate reader (Ex: 337 nm; Em1: 490 nM; Em2: 520 nM). The ratio of Em520 to Em490 was calculated, multiplied by 10,000, and plotted against compound concentration. Resulting data was normalized and fit to a 1:1 binding curve.

2.4.14.3. MDM2 binding to P53 peptide via TR-FRET

[P53 Peptide] (M)	Recombinant MDM2 to L-P53 Peptide		Synthetic L-MDM2 to L-P53 Peptide		Synthetic D-MDM2 to D-P53 Peptide	
	Replicate 1	Replicate 2	Replicate 1	Replicate 2	Replicate 1	Replicate 2
1.00E-06	1.01	0.989	0.837	0.856	1.06	0.944
3.33E-07	0.848	0.879	1.01	0.988	0.949	0.94
1.11E-07	0.707	0.769	0.893	0.89	0.77	0.84
3.70E-08	0.682	0.672	0.82	0.825	0.727	0.734
1.20E-08	0.565	0.558	0.705	0.718	0.578	0.561
4.00E-09	0.378	0.381	0.319	0.295	0.213	0.19
1.00E-09	0.191	0.19	0.0852	0.0842	0.0527	0.0478
0.00E+00	-0.00208	0.00208	0.000229	-0.000229	0.000353	-0.000353

Table 7 Normalized TR-FRET signal for MDM2 binding to P53 peptide

For TR-FRET measurement of MDM2 binding to P53 peptide, biotinylated, L, D or recombinant MDM2 was diluted to 20 nM, L- and D-FAM-labeled P53 peptide was diluted to 2 µM and Terbium-labeled streptavidin (Cis-Bio) was diluted to 20 nM in assay buffer (10 mM HEPES pH 7.5, 150 mM NaCl, 0.05% Tween 20, 1 mM DTT). P53 peptides were serially diluted in assay buffer (final top concentration of 1 µM) and 20 µL was added to a black 384-well plate (Costar) containing 20 µL of MDM2 protein. Samples were incubated for 30 minutes at room temperature. TR-FRET signal was determined using a LanthaScreen™ filter on a PheraStar (BMG Biotech) plate reader (Ex: 337 nm; Em1: 490 nM; Em2: 520 nM). The ratio of Em520 to Em490 was calculated, multiplied by 10,000, and plotted against compound concentration. Resulting data was normalized and fit to a 1:1 binding curve.

2.4.14.4. Max-Max-nb and Myc-Max-nb binding to DNA via TR-FRET

[DNA] (M)	Recombinant-Myc-Max to D-DNA		Synthetic L-Myc-Max to D-DNA		Synthetic D-Myc-Max to L-DNA	
	Replicate 1	Replicate 2	Replicate 1	Replicate 2	Replicate 1	Replicate 2
3.00E-06	0.928	1.07	0.927	1.07	0.928	1.07
1.00E-06	0.987	0.892	0.727	0.671	0.791	0.76
3.33E-07	0.67	0.708	0.534	0.578	0.561	0.601
1.11E-07	0.479	0.466	0.422	0.349	0.43	0.373
3.70E-08	0.226	0.245	0.203	0.197	0.21	0.201
1.20E-08	0.0834	0.0992	0.0842	0.0884	0.0825	0.0883
4.00E-09	0.0266	0.0335	0.0292	0.0328	0.0308	0.0342
0.00E+00	0.0022	-0.0022	-0.000192	0.000192	0.000293	-0.000293

Table 8 Normalized TR-FRET signal for Myc-Max-nb binding to DNA

[DNA] (M)	Recombinant Max-Max to D-DNA		Synthetic L-Max-Max to D-DNA		Synthetic D-Max-Max to L-DNA	
	Replicate 1	Replicate 2	Replicate 1	Replicate 2	Replicate 1	Replicate 2
1.00E-06	1.09	0.911	0.992	1.01	1.03	0.974
3.33E-07	0.543	0.603	0.708	0.687	0.674	0.69
1.11E-07	0.282	0.285	0.464	0.454	0.383	0.394
3.70E-08	0.139	0.174	0.257	0.246	0.19	0.2
1.23E-08	0.0588	0.0724	0.124	0.0875	0.0948	0.0747
4.12E-09	0.0217	0.0206	0.052	0.0475	0.0374	0.0422
1.37E-09	0.0313	0.00322	0.0253	0.0138	0.0176	0.00563
0.00E+00	-0.00766	0.00766	0.0018	-0.0018	-0.00931	0.00931

Table 9 Normalized TR-FRET signal for Max-Max-nb binding to DNA

For TR-FRET measurement of Myc-Max and Max-Max binding to E-box DNA, biotinylated, L, D or recombinant Myc-Max or Max-Max was diluted to 20 nM, Alexa Fluor™ 488 labeled L- and D-DNA was diluted to 6 μM and Terbium-labeled streptavidin (Cis-Bio) was diluted to 20 nM in assay buffer (10 mM HEPES, 2.5 mM MgCl2, 0.05 mM EDTA, 10% glycerol, 50 mM NaCl, 50 ng/mL poly(dI-dC), 0.1% Triton X-100, 1 mM dithiothreitol, pH 7.5). DNA was serially diluted in assay buffer and 20 μL was added to a black 384-well plate (Costar) containing 20 μL Myc/Max or Max/Max protein (final DNA top concentration of 3 μM). Samples were incubated for 30 minutes at room temperature. TR-FRET signal was determined using a LanthaScreen™ filter on a PheraStar (BMG Biotech) plate reader (Ex: 337 nm; Em1: 490 nM; Em2: 520 nM). The ratio of Em520 to Em490 was calculated, multiplied by 10,000, and plotted against compound concentration. Resulting data was normalized and fit to a 1:1 binding curve.

2.4.14.5. Barnase catalysis FRET assay

The relative catalytic activity of synthetic L- and D-Barnase was measured in a similar manner to previous work(*49*). A fluorogenic substrate of the appropriate chirality (6-FAM-dA^{D}rG^{D}dA^{D}dA^{D}-6-TAMRA for L-barnase, and 6-FAM-dA^{L}rG^{L}dA^{L}dA^{L}-6-TAMRA for D-barnase) undergoes an increase in fluorescence intensity upon cleavage at the rG-dA junction by barnase. The resulting time course of fluorescence growth can be fit to a first order rate equation to extract observed k_{cat}/K_M values.

Both fluorogenic substrates (6-FAM-dA^{D}rG^{D}dA^{D}dA^{D}-6-TAMRA for L-barnase, and 6-FAM-dA^{L}rG^{L}dA^{L}dA^{L}-6-TAMRA for D-barnase) for the assay was purchased from ChemGenes as HPLC-purified solid and dissolved in assay buffer solution (100 mM MES, 100 mM NaCl, pH 6.0).

Fluorescence measurements were recorded on a Tecan Spark® plate reader with excitation at 495 nm and monitoring emission at 515 nm. Assays were carried out in flat bottomed, black 96-well plates. The fluorogenic substrate of the appropriate chirality was diluted in assay buffer, and the baseline absorbance recorded for at least 150 sec. To initiate hydrolysis, L or D barnase was added in a single aliquot and the fluorescence monitored every 15 s until sufficient data was collected for analysis. Reactions were carried out in a total of 250 µL with 200 nM substrate, and 2 nM barnase. The experiment was performed in duplicate. Blank runs were performed in parallel with the experimental runs.

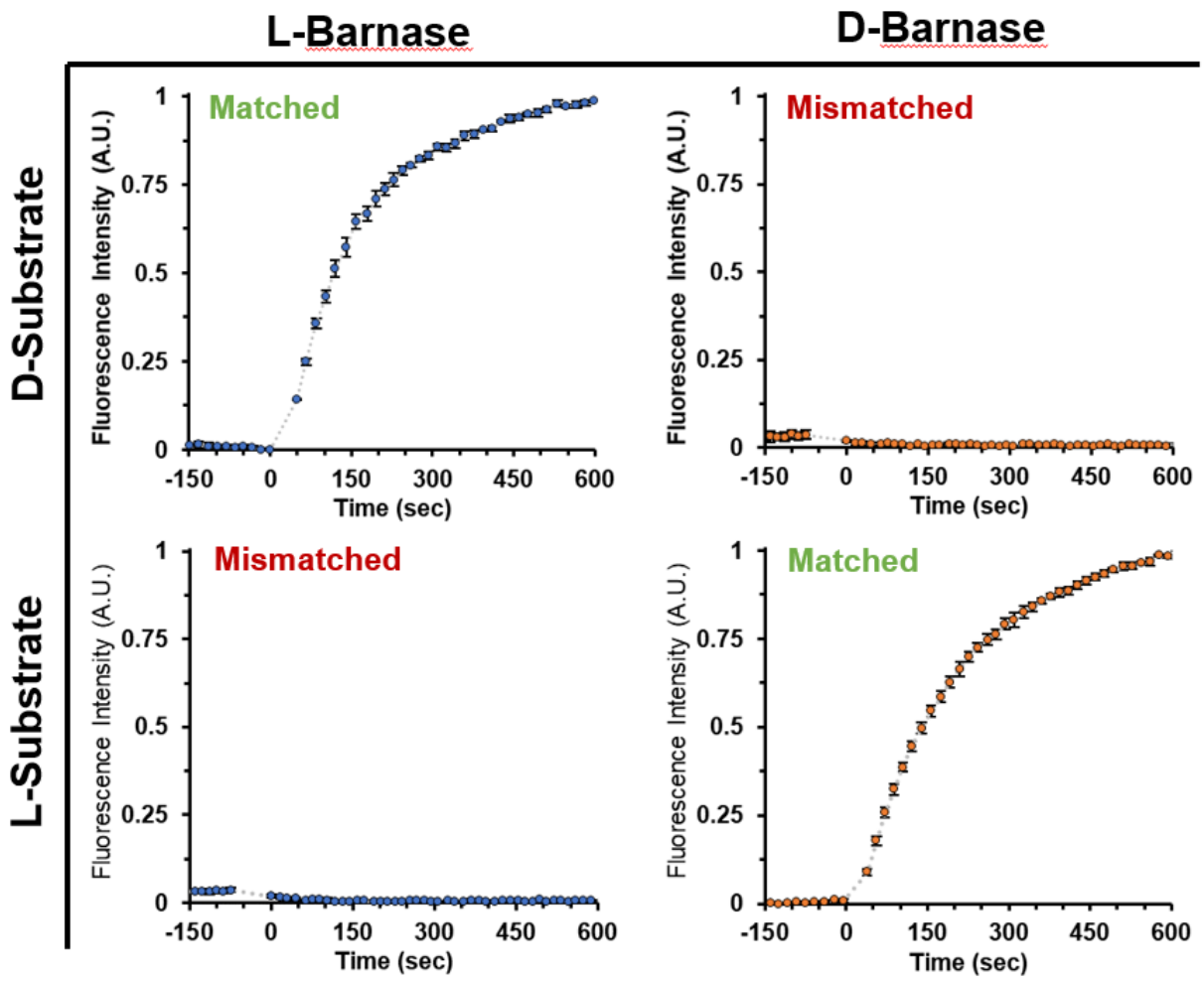


Figure 2.7 Raw time course for the Barnase catalyzed hydrolysis of a fluorgenic RNA substrate.

L-Barnase						D-Barnase					
L-Substrate			D-Substrate			L-Substrate			D-Substrate		
t (s)	Mean	SD	T (s)	Mean	SD	t (s)	Mean	SD	t (s)	Mean	SD
-426.468	0.027	0.001	-183.389	0.030	0.002	-221.057	0.014	0.015	-426.468	0.118	0.018
-411.632	0.026	0.001	-165.393	0.021	0.003	-204.640	0.005	0.010	-411.632	0.028	0.019
-396.771	0.029	0.002	-149.154	0.013	0.003	-188.408	0.005	0.004	-396.771	0.032	0.015
-383.786	0.032	0.006	-132.930	0.014	0.010	-172.179	0.004	0.005	-383.786	0.030	0.015
-370.817	0.027	0.004	-113.866	0.010	0.011	-155.798	0.001	0.002	-370.817	0.027	0.017
-357.828	0.028	0.006	-97.482	0.010	0.004	-139.567	0.001	0.003	-357.828	0.027	0.017
-344.823	0.027	0.003	-81.188	0.008	0.006	-123.370	0.000	0.002	-344.823	0.027	0.015
-331.795	0.028	0.004	-64.988	0.006	0.009	-107.156	0.001	0.001	-331.795	0.027	0.012
-316.803	0.030	0.002	-48.775	0.009	0.005	-90.951	0.005	0.002	-316.803	0.030	0.014
-303.813	0.032	0.003	-32.544	0.006	0.003	-74.680	0.002	0.004	-303.813	0.028	0.015
-288.706	0.029	0.003	-16.251	0.000	0.003	-58.414	0.005	0.003	-288.706	0.029	0.016
-273.731	0.032	0.005	0.000	0.000	0.002	-40.000	0.006	0.003	-273.731	0.032	0.019
-260.788	0.032	0.003	49.496	0.141	0.002	-21.653	0.010	0.002	-260.788	0.034	0.017
-243.888	0.028	0.004	66.586	0.250	0.010	-5.471	0.007	0.005	-243.888	0.027	0.012
-230.361	0.031	0.004	85.235	0.357	0.014	38.839	0.089	0.008	-230.361	0.029	0.015
-217.364	0.032	0.006	103.611	0.434	0.017	56.000	0.179	0.013	-217.364	0.030	0.013

-204.364	0.031	0.004	119.893	0.513	0.022	72.234	0.259	0.014	-204.364	0.030	0.010
-191.399	0.031	0.005	140.626	0.573	0.026	88.447	0.324	0.015	-191.399	0.033	0.016
-178.367	0.032	0.004	158.709	0.646	0.021	104.696	0.386	0.013	-178.367	0.031	0.017
-165.328	0.032	0.003	179.439	0.668	0.021	121.051	0.445	0.015	-165.328	0.032	0.015
-152.332	0.033	0.005	195.666	0.710	0.023	139.015	0.497	0.016	-152.332	0.034	0.013
-139.326	0.031	0.004	211.910	0.738	0.017	157.058	0.546	0.017	-139.326	0.032	0.016
-126.278	0.032	0.003	228.227	0.764	0.020	175.101	0.586	0.016	-126.278	0.030	0.010
-113.248	0.033	0.007	244.502	0.790	0.014	191.346	0.628	0.015	-113.248	0.031	0.014
-100.233	0.034	0.007	260.808	0.804	0.007	209.863	0.664	0.020	-100.233	0.038	0.012
-87.215	0.033	0.005	277.115	0.822	0.009	226.206	0.699	0.015	-87.215	0.032	0.012
-74.172	0.035	0.006	293.451	0.832	0.012	242.451	0.725	0.013	-74.172	0.038	0.013
0.000	0.018	0.005	309.667	0.857	0.010	260.865	0.748	0.015	0.000	0.019	0.002
16.636	0.015	0.004	326.059	0.854	0.012	277.202	0.762	0.013	16.636	0.013	0.006
29.823	0.013	0.004	342.317	0.866	0.012	293.592	0.792	0.016	29.823	0.013	0.008
45.792	0.012	0.003	358.705	0.889	0.012	309.923	0.804	0.020	45.792	0.012	0.005
63.259	0.007	0.002	377.513	0.892	0.011	328.054	0.827	0.017	63.259	0.010	0.004
76.362	0.010	0.005	393.915	0.904	0.004	344.402	0.841	0.013	76.362	0.014	0.003
89.470	0.008	0.003	410.330	0.908	0.008	360.814	0.857	0.008	89.470	0.011	0.005
102.718	0.008	0.005	426.644	0.928	0.004	377.078	0.870	0.008	102.718	0.011	0.002
117.900	0.004	0.003	442.988	0.936	0.012	393.460	0.881	0.013	117.900	0.004	0.001
131.872	0.004	0.003	459.364	0.940	0.007	409.958	0.886	0.012	131.872	0.010	0.000
145.105	0.003	0.002	475.823	0.948	0.006	426.566	0.902	0.012	145.105	0.005	0.003
158.836	0.006	0.003	494.625	0.953	0.010	442.926	0.913	0.011	158.836	0.008	0.003
171.955	0.005	0.002	512.269	0.961	0.008	459.566	0.925	0.009	171.955	0.006	0.003
188.720	0.004	0.002	531.111	0.979	0.009	475.912	0.935	0.008	188.720	0.009	0.006
201.843	0.004	0.003	547.465	0.972	0.005	492.350	0.945	0.006	201.843	0.010	0.002
214.917	0.004	0.002	566.153	0.976	0.009	511.040	0.956	0.012	214.917	0.007	0.000
228.057	0.003	0.002	582.623	0.980	0.009	527.681	0.957	0.009	228.057	0.012	0.003
241.318	0.004	0.002	599.020	0.986	0.003	544.146	0.965	0.004	241.318	0.009	0.003
254.527	0.005	0.001	615.545	0.992	0.002	560.545	0.969	0.010	254.527	0.004	0.003
267.665	0.006	0.000	632.306	0.991	0.013	576.936	0.986	0.003	267.665	0.006	0.002
280.981	0.007	0.001	648.728	1.000	0.003	593.302	0.984	0.004	280.981	0.006	0.003
294.056	0.002	0.002	665.120	1.000	0.007	609.689	0.987	0.011	294.056	0.006	0.002
307.213	0.003	0.002	681.651	0.999	0.011	626.580	1.000	0.005	307.213	0.003	0.003
324.741	0.005	0.002	698.066	1.003	0.013	643.081	1.006	0.004	324.741	0.012	0.003
337.942	0.004	0.002	714.529	1.009	0.013	659.607	1.010	0.006	337.942	0.010	0.002
353.025	0.003	0.003	731.121	1.017	0.005	676.232	1.020	0.002	353.025	0.009	0.003
369.192	0.006	0.003	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	369.192	0.009	0.002
382.349	0.005	0.003	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	382.349	0.009	0.004
395.516	0.004	0.002	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	395.516	0.007	0.002
412.313	0.004	0.003	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	412.313	0.005	0.002
425.436	0.006	0.003	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	425.436	0.006	0.001
440.949	0.005	0.002	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	440.949	0.008	0.003
454.055	0.006	0.004	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	454.055	0.005	0.003
467.178	0.004	0.001	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	467.178	0.006	0.003
480.358	0.004	0.001	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	480.358	0.010	0.001
493.561	0.008	0.003	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	493.561	0.006	0.004
506.672	0.005	0.001	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	506.672	0.005	0.002
519.870	0.006	0.000	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	519.870	0.011	0.000
533.089	0.005	0.002	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	533.089	0.008	0.001
546.211	0.003	0.001	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	546.211	0.007	0.006
559.438	0.005	0.002	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	559.438	0.008	0.003
572.716	0.006	0.001	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	572.716	0.009	0.003
587.317	0.005	0.001	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	587.317	0.005	0.001
600.570	0.008	0.001	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	600.570	0.010	0.003
615.698	0.008	0.003	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	615.698	0.005	0.001
631.188	0.008	0.002	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	631.188	0.008	0.002
648.861	0.005	0.004	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	648.861	0.009	0.002
666.454	0.010	0.001	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	666.454	0.008	0.002

679.545	0.010	0.003	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	679.545	0.007	0.001
692.819	0.008	0.002	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	692.819	0.009	0.003
705.939	0.010	0.001	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	705.939	0.008	0.004
719.131	0.007	0.003	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	719.131	0.006	0.000
732.281	0.007	0.002	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	732.281	0.009	0.002
745.434	0.007	0.003	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	745.434	0.007	0.002
758.543	0.008	0.001	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	758.543	0.007	0.004
771.719	0.008	0.003	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	771.719	0.006	0.003
785.040	0.006	0.001	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	785.040	0.009	0.002
800.502	0.006	0.002	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	800.502	0.004	0.001
813.681	0.005	0.003	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	813.681	0.010	0.003
826.847	0.008	0.001	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	826.847	0.011	0.003

Table 10 Averaged data from fluorogenic RNase activity assay was used for the determination of kcat/KM values.

2.4.14.6. ERG and IRAK2 Circular Dichroism

Before analysis, all proteins were exchanged into CD buffer (10 mM KP*i*, 5% glycerol v/v, pH 7.4). Fast protein liquid chromatography (FPLC) purification was performed using GE Life Sciences columns on an ÄKTA Pure FPLC system. The ÄKTA Pure, FPLC columns, and all buffers were stored and subsequently used in a cold room maintained at 4 °C.

HiTrap™ Desalting 5 mL: Before desalting, the column was equilibrated with 20 mL of the CD buffer at 2 mL/min or until the conductance reading from the ÄKTA Pure stabilizes, whichever is the greater volume. Up to 1.5 mL of sample was loaded into the column using a syringe fitted with a compatible male-end screw tip, and then the loaded column was attached to the ÄKTA Pure. An isocratic flow of buffer (2 CV) at 1 mL/min provided the desalted, buffer-exchanged protein. The concentration of the resulting protein solution was determined with ***Method 4.1.1***. Protein concentration was adjusted to 0.1 mg / mL with additions of CD buffer.

All CD spectra were recorded on an AVIV Model 420 Circular Dichroism Spectrometer, with a 1 mm quartz cuvette (Hellma Analytics, Article No.: 100-1-40). Before all readings, cuvettes were cleaned with Hellmanex™ 3 concentrate according to manufacturer directions. Far circular dichroism spectra were recorded as the average of three scans between 190 nm to 260 nm with 1 nm steps, 2 nm scanning windows, and a 3 s averaging time for each scan at 25 °C.

For variable temperature CD, CD signal was monitored at 225 nm with 2 nm bandwidth and 8 second averaging time. Spectra were recorded between 0 and 110 °C with 5 °C steps and 2-minute incubation time at each step.

2.4.14.7. NEMO binding to IKKb peptide with SPR

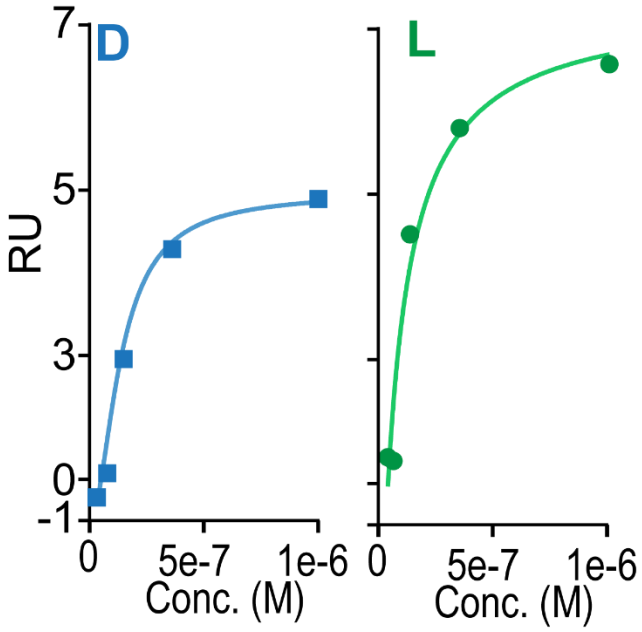


Figure 2.8 SPR of Synthetic L, D and Recombinant NEMO to IKKB peptide

Sensorgrams of IKKB peptide binding to immobilized NEMO are shown.

SPR experiments were performed on a Biacore S200 (Cytiva) instrument at 25 °C in 1x HBS-P+ buffer (Cytiva) with 1% DMSO and 0.5 mM TCEP. A CAP Series S sensor chip was docked and rehydrated overnight with water. On the day of the experiment, the instrument was primed with running buffer. CAP reagent was diluted 5X in water and injected over the CAP surface at 2 µL/min for 120 seconds for a final capture level of ~200 RU. Synthetic L, D, or recombinant NEMO was diluted to 200 nM in running buffer and captured to channels 2, 3, or 4, respectively, at 5 µL/min for 60 seconds for a final immobilization level of ~30 RU. The iKKB peptide was diluted to 1 µM in running buffer then serially diluted 3-fold for a total of 8 concentrations with one blank. Compounds were injected over the immobilized and reference surfaces at 30 µL/min for 180 seconds and then allowed to dissociate for 360 seconds. The surface was regenerated with CAP regeneration solution at 10 µL/min for 120 seconds. Sensorgrams were double-referenced and fit to a 1:1 steady state affinity model.

2.4.14.8. FKBP12 binding to Rapamycin/mTor with SPR

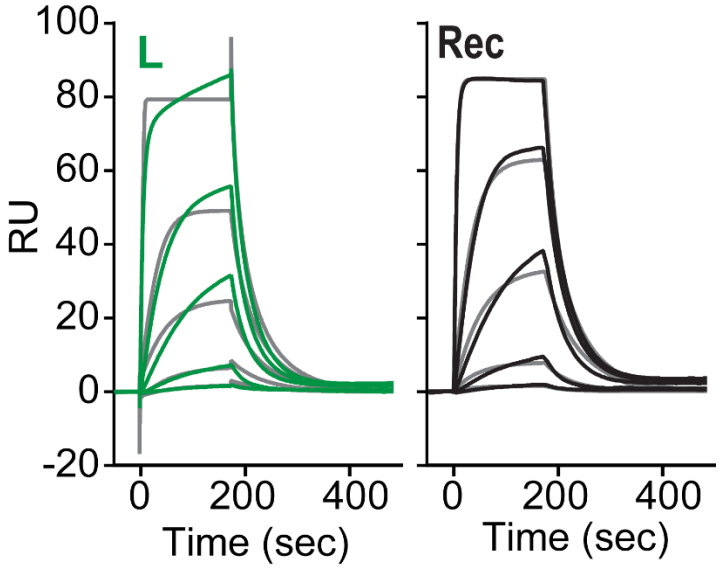


Figure 2.9 SPR of Synthetic L and Recombinant FKBP12 to Rapamycin/mTOR.

Sensorgrams of Rapamycin/mTOR binding to immobilized FKBP12 are shown.

SPR experiments were performed on a Biacore S200 (Cytiva) instrument at 25 °C in 1x HBS-P+ buffer (Cytiva) with 1% DMSO and 0.5 mM TCEP. A CAP Series S sensor chip was docked and rehydrated overnight with water. On the day of the experiment, the instrument was primed with running buffer. CAP reagent was diluted 5X in water and injected over the CAP surface at 2 µL/min for 120 seconds for a final capture level of ~200 RU. Synthetic L- and recombinant FKBP12 were diluted to 200 nM in running buffer and captured to channels 2, or 4, respectively, at 5 µL/min for 60 seconds for a final immobilization level of ~200 RU. The Rapamycin and mTOR were each diluted to a final concentration of 500 nM in running buffer then serially diluted 3-fold for a total of 8 concentrations with one blank. The complex was injected over the immobilized and reference surfaces at 30 µL/min for 180 seconds and then allowed to dissociate for 360 seconds. The surface was regenerated with CAP regeneration solution at 10 µL/min for 120 seconds. Sensorgrams were double-referenced and fit to a 1:1 binding model.

2.4.14.9. YAP1 binding to Dendrin with SPR

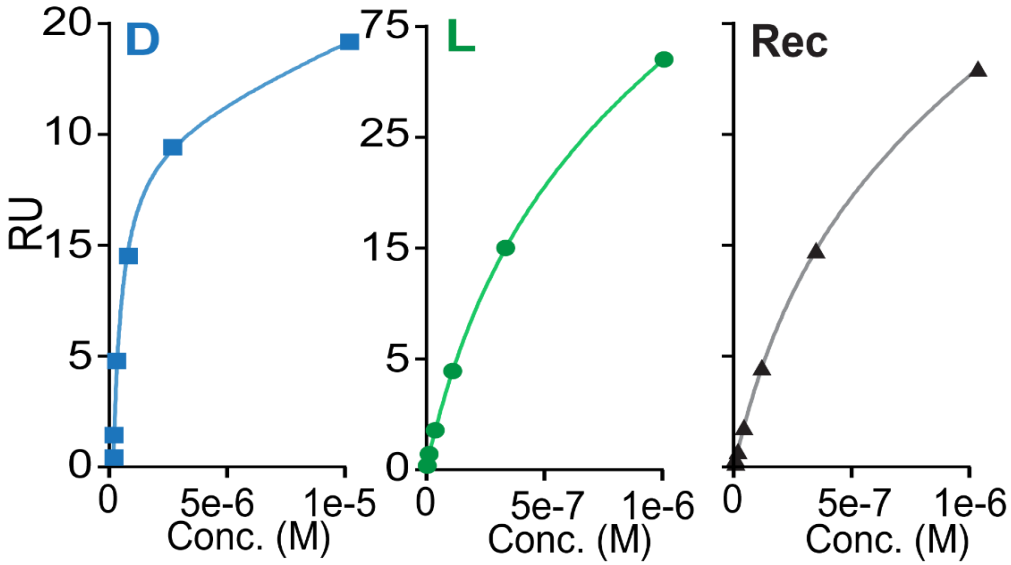


Figure 2.10 SPR of Synthetic L, D and Recombinant YAP1 to Dendrin peptide

Sensorgrams of Dendrin peptide binding to immobilized YAP1 are shown.

SPR experiments were performed on a Biacore S200 or 8K (Cytiva) instrument at 25 °C in 1x HBS-P+ buffer (Cytiva) with 1% DMSO and 0.5 mM TCEP using the CAP system (D-YAP1) or SA capture (L and recombinant YAP1). For CAP experiment, a CAP Series S sensor chip was docked and rehydrated overnight with water. On the day of the experiment, the instrument was primed with running buffer. CAP reagent was diluted 5X in water and injected over the CAP surface at 2 µL/min for 120 seconds for a final capture level of ~200 RU. Synthetic D- and recombinant YAP1 were diluted to 400 nM in running buffer and captured to channels 2, and 4, respectively, at 5 µL/min for 60 seconds for a final immobilization level of ~170 RU. L- and D-Dendrin peptide were each diluted to 10 µM in running buffer then serially diluted 4-fold for a total of 6 concentrations with one blank. The peptides were injected over the immobilized and reference surfaces at 30 µL/min for 180 seconds and then allowed to dissociate for 360 seconds. The surface was regenerated with CAP regeneration solution at 10 µL/min for 120 seconds. Sensorgrams were double-referenced and fit to a 1:1 steady state affinity model. For SA experiments, a SA Series S sensor chip was docked and pre-conditioned with 3 injections of 50 mM NaOH/1 M NaCl to remove unbound streptavidin from the surface. Synthetic L, or

recombinant YAP1 was diluted to 1 µg/mL in running buffer and immobilized to channels 1 and 2, respectively, at 1 µL/min for 300 seconds for a final immobilization level of ~1500 RU. L-Dendrin peptide was diluted to 10 µM in running buffer then serially diluted 3-fold for a total of 7 concentrations with one blank. The peptide was injected over the immobilized and reference surfaces at 30 µL/min for 180 seconds and then allowed to dissociate for 360 seconds. The surface was regenerated with a 15 second injection of 1 M sodium chloride. Sensorgrams were double-referenced and fit to a 1:1 steady state affinity model.

2.4.14.10. NEMO_iZIP binding to IKKb with SPR

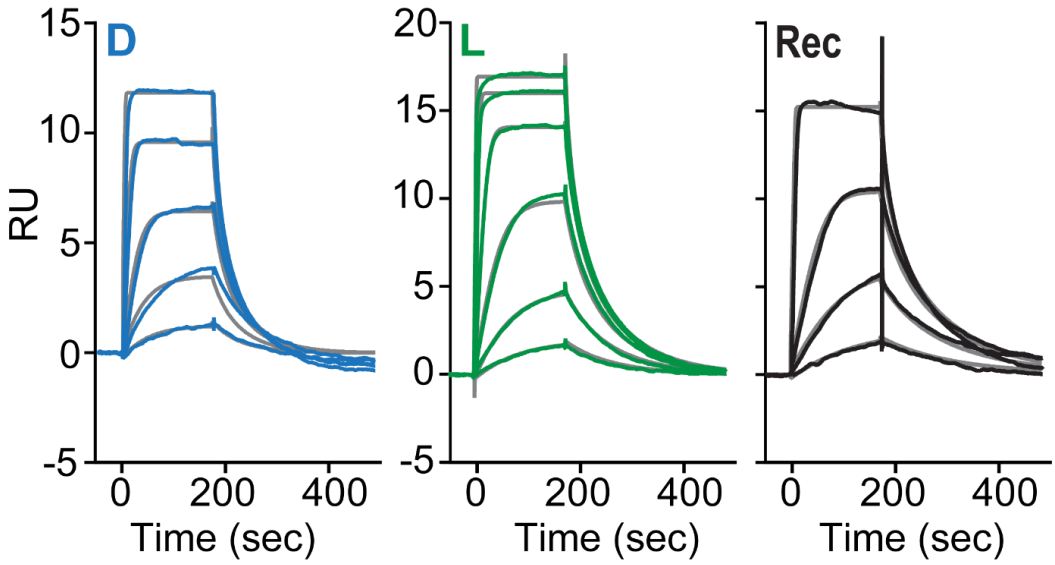


Figure 2.11 SPR of Synthetic L, D and Recombinant NEMO_iZIP SPR to IKKB peptide.

Sensorgrams of IKKB peptide binding to immobilized NEMO_iZIP are shown.

SPR experiments were performed on a Biacore S200 (Cytiva) instrument at 25 °C in 1x HBS-P+ buffer (Cytiva) with 1% DMSO and 0.5 mM TCEP. A CAP Series S sensor chip was docked and rehydrated overnight with water. On the day of the experiment, the instrument was primed with running buffer. CAP reagent was diluted 5X in water and injected over the CAP surface at 2 µL/min for 120 seconds for a final capture level of ~200 RU. Synthetic L- and recombinant FKBP12 were diluted to 200 nM in running buffer and captured to channels 2, 3 and 4, respectively, at 5 µL/min for 60 seconds for a final immobilization level of ~60 RU. L- and D-IKKB peptides were diluted to 500 nM in running buffer then serially diluted 3-fold for a total of 8 concentrations with one blank. The peptides were injected over the immobilized and reference surfaces at 30 µL/min for 180 seconds and then allowed to dissociate for 360 seconds. The surface was regenerated with CAP regeneration solution at 10 µL/min for 120 seconds. Sensorgrams were double-referenced and fit to a 1:1 binding model.

2.4.15. Peptide Clusters from Phage Display

2.4.15.1. MDM2 Screening Results

Cluster	ID	Sequence	Measured K_D
MDM2.C1	H101‡	Ac-dpahwycdfaaqvcnfs-NH_2	5.5* µM
MDM2.C2	H102‡	Ac-dpawnhceyaafvcsev-NH_2	2.7 µM
MDM2.C3	H103‡	Ac-dpawyecmeaammcqqv-NH_2	0.88 µM
MDM2.C4	H104	Ac-dpalvecnlaadmchfy-NH_2	1.8 µM
MDM2.C5	H105	Ac-dpaantciwaaiecsmy-NH_2	16* µM
MDM2.C6	H106	Ac-dpawatcmdaalnclqm-NH_2	1.8 µM
MDM2.C7	H107	Ac-dpasrncqwaahlceyw-NH_2	1.1 µM
MDM2.C8	H108	Ac-dpalaqcrwaawrcdfe-NH_2	12* µM
MDM2.C9	H109	Ac-dpaswqcvmaamdcvld-NH_2	2.4 µM

Table 11 Binders to D-MDM2 identified from mirror image phage display

*Listed sequences were cyclized between C7 and C14 with N,N'-(1,4-phenylene)bis(2-bromoacetamide). *Listed KD values are beyond tested concentrations. ‡Sequences were selected for co-crystallization with MDM2.*

2.4.15.2. MDM2 Binder SPR Validation

SPR experiments were performed on a Biacore 8K (Cytiva) instrument at 25 °C in 1x HBS-P+ buffer (Cytiva) with 1% DMSO. An SA Series S sensor chip was docked and pre-conditioned with three injections of 50 mM NaOH/1 M NaCl to remove unbound streptavidin from the surface. Biotinylated $MDM2^{25-109}$ were diluted to 2 µg/mL in running buffer and immobilized to channels 1 through 8 at 5 µL/min for 71 s for a final immobilization level of ~390 RU. Peptides were diluted to 5 µM in running buffer then serially diluted 2-fold for a total of seven concentrations with one blank (7-point two-fold peptide dilution series with top concentration = 5 µM and bottom concentration = 78 nM). Compounds were injected over the immobilized and reference surfaces at 30 µL/min for 60 s and then allowed to dissociate for 180 s without surface regeneration (N = 1). Data was analyzed using Biacore Insight Evaluation software (Cytiva). Sensorgrams were double referenced, with most of them fitted to a 1:1 steady state affinity model, with a few fitted with both steady state affinity model and kinetics model.

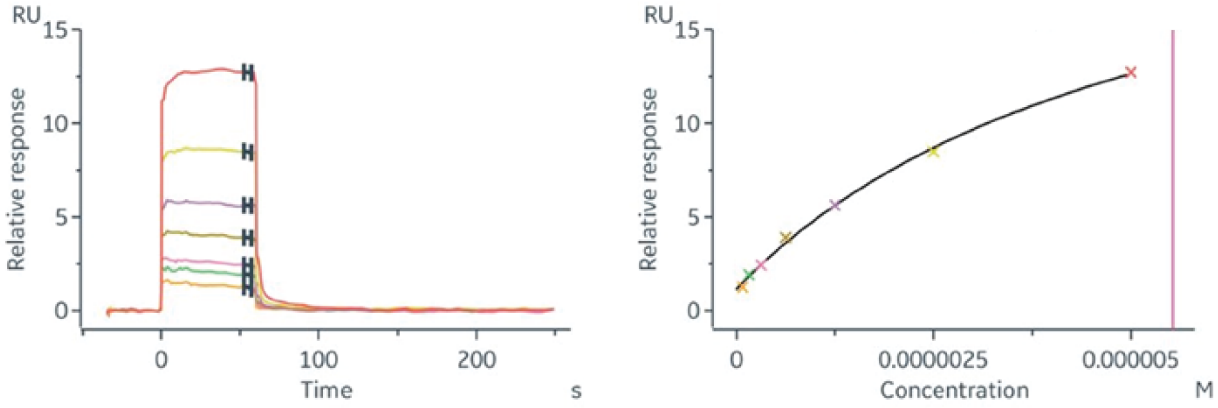


Figure 2.12 D-H101 SPR to Recombinant MDM2

Sensorgrams and steady state affinity plots of D-H101 binding to recombinant MDM2 are shown.

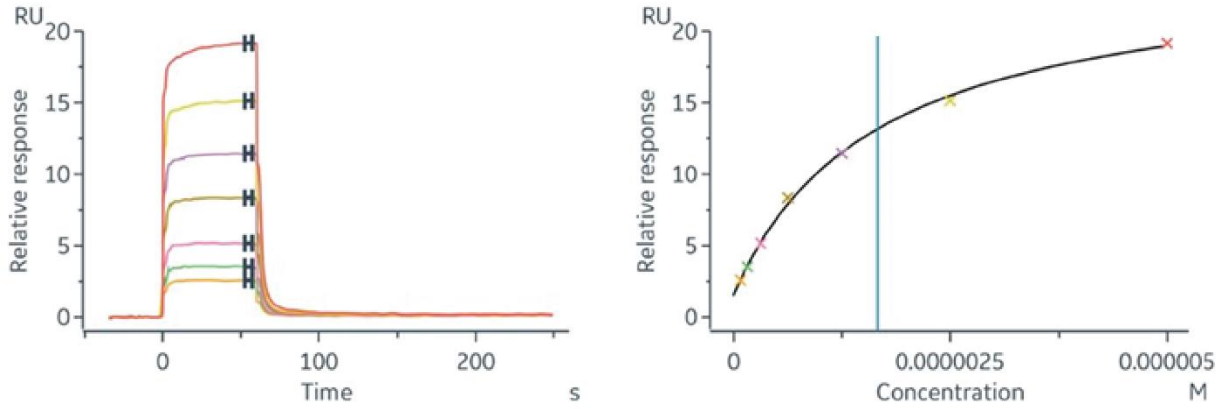


Figure 2.13 D-H102 SPR to Recombinant MDM2

Sensorgrams and steady state affinity plots of D-H102 binding to recombinant MDM2 are shown.

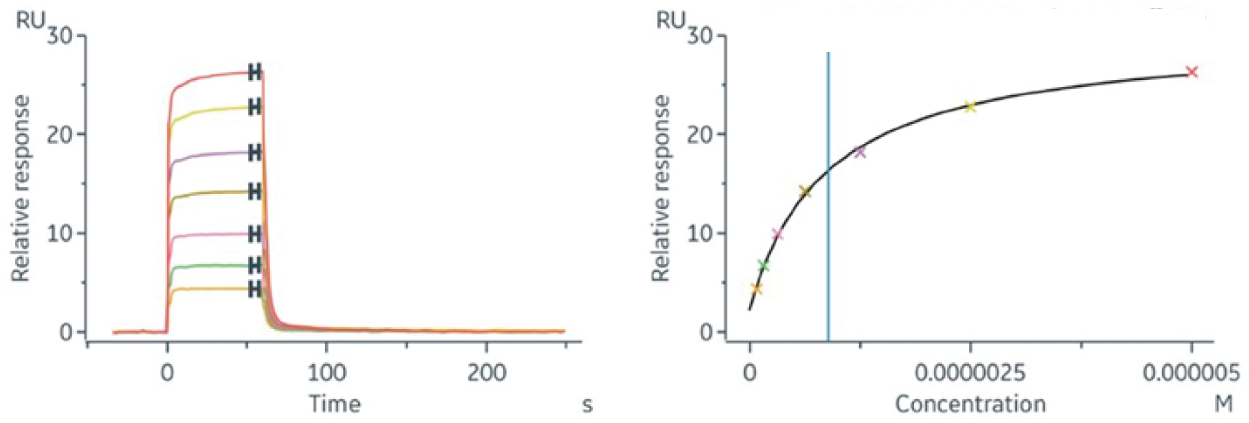


Figure 2.14 D-H103 SPR to Recombinant MDM2

Sensorgrams and steady state affinity plots of D-H103 binding to recombinant MDM2 are shown.

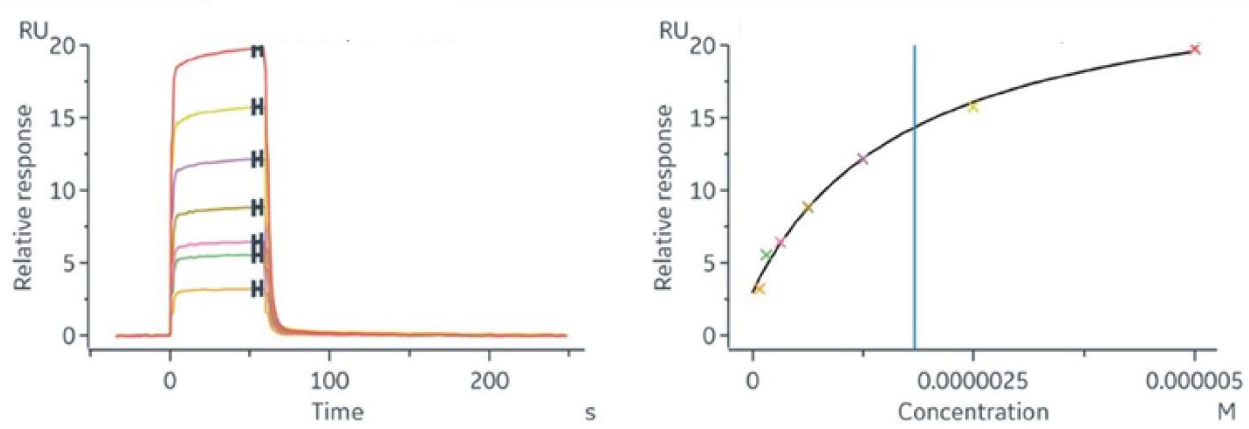


Figure 2.15 D-H104 SPR to Recombinant MDM2

Sensorgrams and steady state affinity plots of D-H104 binding to recombinant MDM2 are shown.

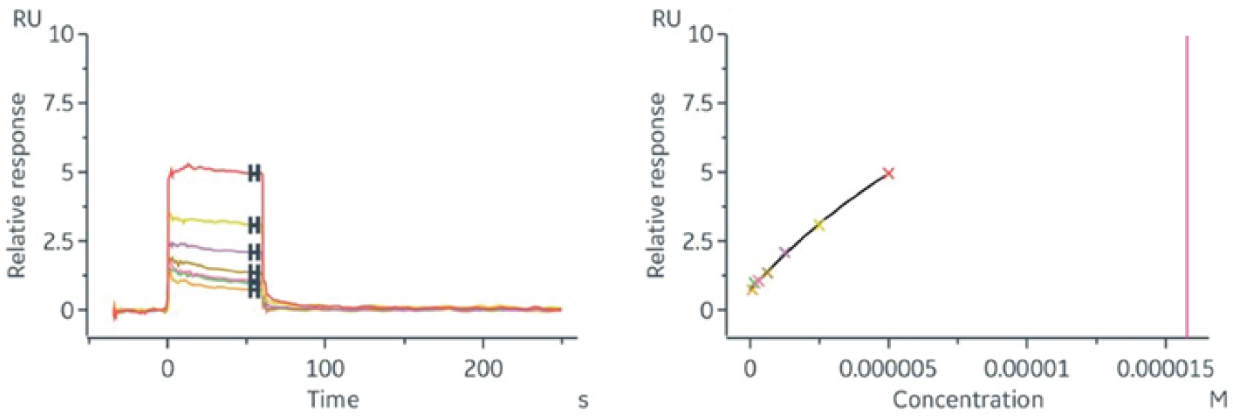


Figure 2.16 D-H105 SPR to Recombinant MDM2

Sensorgrams and steady state affinity plots of D-H105 binding to recombinant MDM2 are shown.

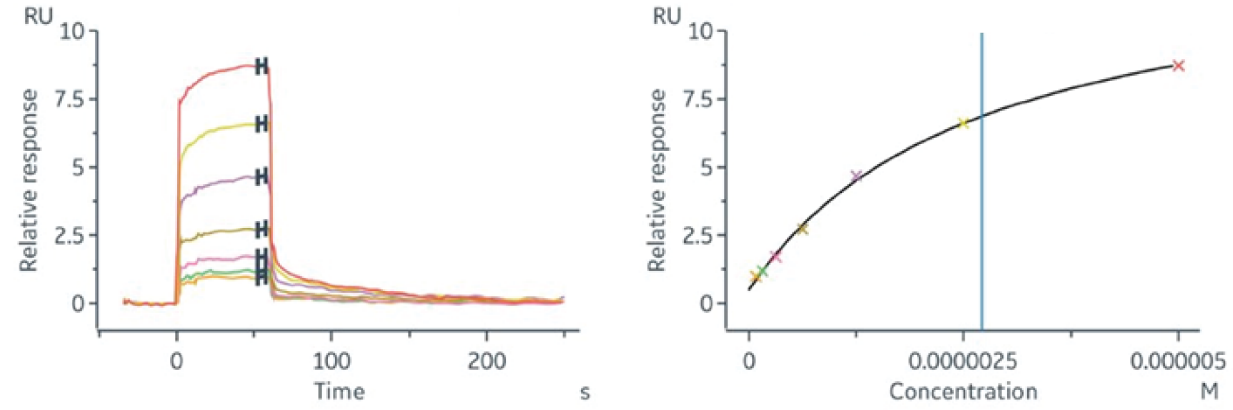


Figure 2.17 D-H106 SPR to Recombinant MDM2

Sensorgrams and steady state affinity plots of D-H106 binding to recombinant MDM2 are shown.

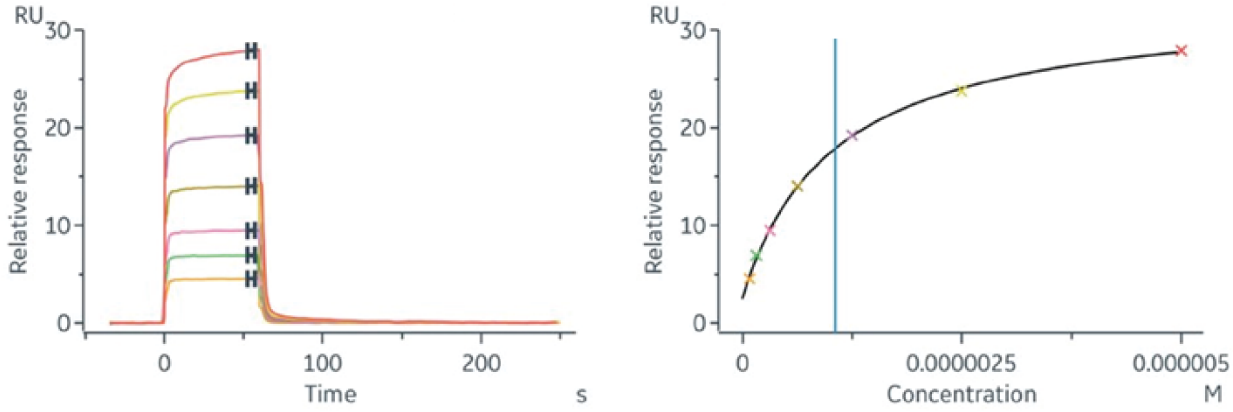


Figure 2.18 D-H107 SPR to Recombinant MDM2

Sensorgrams and steady state affinity plots of D-H107 binding to recombinant MDM2 are shown.

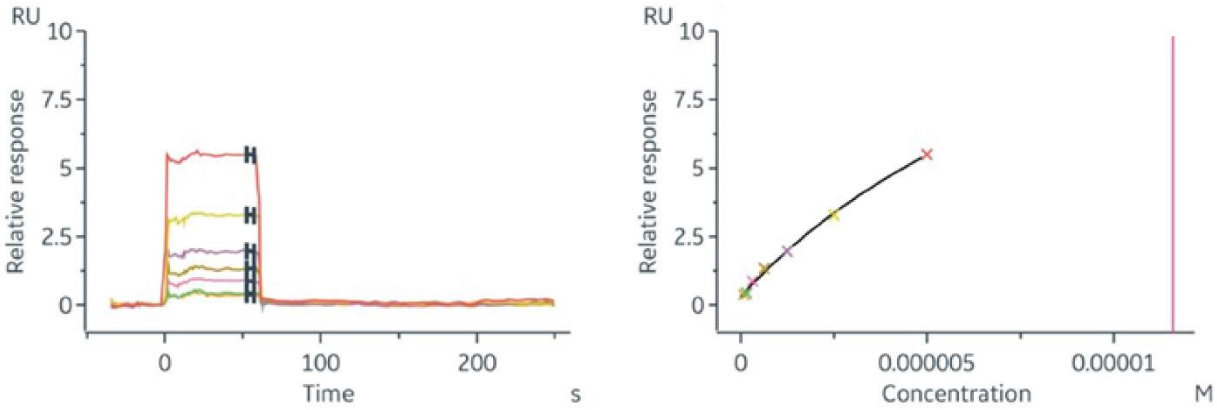


Figure 2.19 D-H108 SPR to Recombinant MDM2

Sensorgrams and steady state affinity plots of D-H108 binding to recombinant MDM2 are shown.

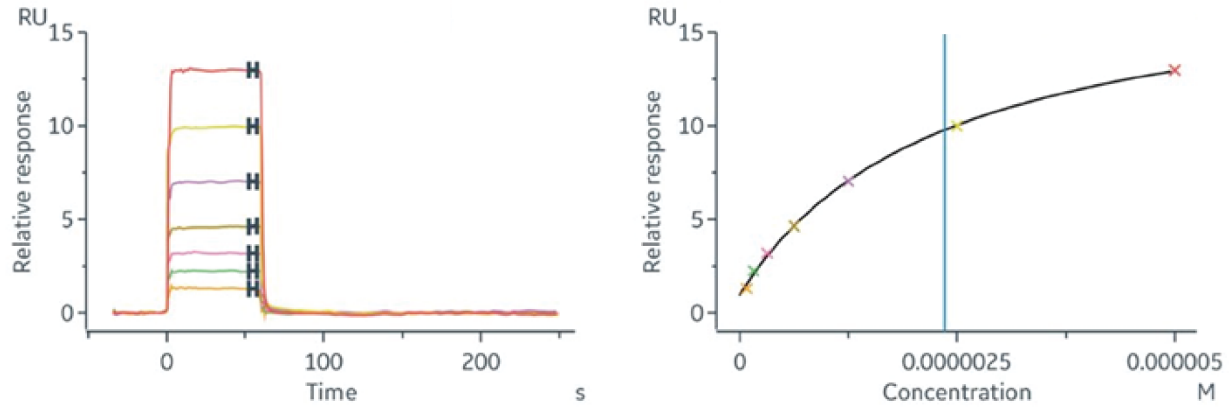


Figure 2.20 D-H109 SPR to Recombinant MDM2

Sensorgrams and steady state affinity plots of D-H109 binding to recombinant MDM2 are shown.

2.4.15.3. CHIP Screening Results

Cluster	ID	Sequence	Measured IC_{50}
CHIP.C1	H201‡	Ac-pwedcawfawacyni-NH_2	470 nM
CHIP.C3	H202‡	Ac-pwweclsqaddcdfr-NH_2	1230 M
CHIP.C2	H203‡	Ac-phemcywadaycrys-NH_2	550 nM
CHIP.C2	H204‡	Ac-pldlcywaslhcivs-NH_2	770 nM

Table 12 Binders to D-CHIP identified from mirror image phage display

Listed sequences were cyclized between C5 and C12 with N,N'-(1,4-phenylene)bis(2-bromoacetamide). ‡Sequences were selected for co-crystallization with CHIP.

2.4.15.4. CHIP Binder FP Validation

For the competition FP of CHIP, Helicons at 10 mM in DMSO were serially diluted 1:3 in DMSO for a total of 11 concentrations using a Mosquito LV (SPT Labtech), then diluted 1000-fold in buffer (1 x HBS-P+, Cytiva) in duplicate by the Mosquito LV (SPT Labtech) into a black polystyrene 384-well plate (Corning) (11-point three-fold peptide dilution series with top concentration = 10 µM). The assay was performed with 400 nM CHIP23-303 recombinant protein as target, and 20 nM CHIP-binding peptide as probe (5FAM-bAla-SSGPTIEEVD, derived from HSP70). The plate was incubated and protected from light for 1 hour at room temperature prior to reading. Reads were performed on a CLARIOstar plate reader (BMG Labtech) with excitation at 485 nm, emission at 525 nm,

and cutoff at 504 nm. Data were fitted to a 1:1 binding model with Hill slope using an in-house script.

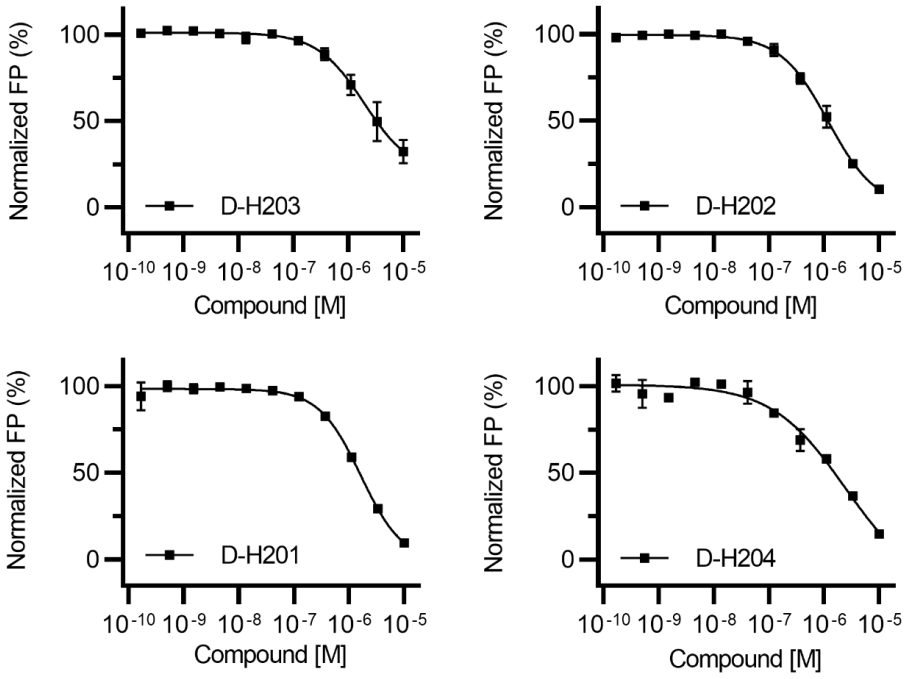


Figure 2.21 Validation of D-macrocyclic binders to recombinant CHIP with FP.

Normalized FP signal of synthetic D macrocyclic binders to recombinant CHIP are shown.

2.4.16. Recombinant Protein Expression Techniques

2.4.16.1. MDM2

For protein used in phage screens and biochemical assays: MDM2 (residues 25-109) with an N-terminal 6xHis–yBBr–TEV tag was recombinantly expressed in *E. coli* BL21 CodonPlus cells (Agilent) from pET-derived expression vectors (Novagen). The cells were induced at OD600 = 0.6 with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 4 h at 37 °C, then harvested and resuspended in 25 mM Tris-HCl pH 7.5, 300 mM NaCl, 10 % glycerol, 1 mM PMSF. For purification, the pellet was lysed with a tip sonicator, and pelleted at 22,000 x g for 30 min at 4 °C. The pellets were washed with 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 M urea, 1.0% triton X-100 three times, and the inclusion body was dissolved in in 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 8 M urea, and 2 mM ß-mercaptoethanol (ß-me). The supernatant was purified with Ni-NTA resin

(Qiagen), eluting with 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 8 M urea, 2 mM ß-me, and 250 mM imidazole. Protein elutes were diluted to ~ 0.1 mg/mL and dialyzed into buffers containing 10 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM β-me, with 4, 2 ,1, or 0 M urea, at 4 °C for 8 h for each urea gradient. Urea-free proteins were concentrated with Amicon spin filters (Millipore Sigma) to ~1 mg/mL and biotinylated via the yBBr reaction according to standard procedures(66). Biotinylated proteins were pooled, concentrated, and loaded onto a Superdex™ 10/300 75 pg (Cytiva) SEC column and eluted in 20 mM HEPES pH 7.0, 200 mM NaCl, 5% glycerol, 0.5 mM tris(2-carboxyethyl) phosphine (TCEP). Fractions containing pure protein were collected, pooled, concentrated to ~1 mg/mL and stored at -80°C.

For protein used in crystallography: MDM2 (residues 17-111, C17S) with an N-terminal 6xHis–TEV tag was recombinantly expressed in E. coli BL21 (DE3) cells (Agilent) from pET-derived expression vectors (Novagen). The cells were induced at OD600 = 0.6 with 0.15 mM IPTG for 16 h at 37 °C, then harvested and resuspended in 50 mM Tris, pH 8.0, 200 mM NaCl, 10 % glycerol, 1 mM TCEP, and 20 mM imidazole. For purification, the pellet was lysed with a tip sonicator with power level set at 200 W, 3 seconds on and 3 seconds off for 20 min, pelleted at 22,000 x g for 30 min at 4 °C. The supernatant was purified with Ni-NTA resin (Qiagen), eluting with 50 mM Tris, pH 8.0, 200 mM NaCl, 10 % glycerol, 1 mM TCEP, and 250 mM imidazole. Eluted proteins were pooled, concentrated, and TEV-cleaved by adding TEV protease at a ratio of 1:10 protease to protein and incubated overnight at 4 °C. TEV-cleaved proteins were loaded onto a Superdex™ 10/300 75 pg (Cytiva) SEC column and eluted in 50 mM Tris, pH 8.0, 200 mM NaCl, 10% glycerol, and 1 mM TCEP. Fractions containing pure protein were collected, pooled, concentrated to 8 mg/mL and stored at -80 °C.

2.4.16.2. ERG

ERG (115-208) with an N-terminal 6xhis-TEV-yBBR-Thrombin tag was recombinantly expressed in One Shot™ BL21 Star™ (DE3) Chemically Competent *E. coli* (ThermoFisher) from pET-derived expression vectors (ThermoFisher) in Luria broth supplemented with 0.1 mg/mL ampicillin. The cells were induced at OD600 = 0.8 with 1

mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 8 h at 25 °C then overnight at 15 °C, then harvested and resuspended in lysis buffer (50 mM HEPES, 200 mM NaCl, 10 mM imidazole, 0.5 mM TCEP, 2 mM ATP, 10 mM MgCl2, 1% NP40, 5% glycerol, Roche complete inhibitors, pH 7.5). For purification, the pellet was lysed with a tip sonicator then pelleted at 20,000 x g for 30 min at 4 °C. Supernatant was collected and purified with cOmplete™ His-Tag Purification Columns (Roche) eluting with 300 mM imidazole. Protein was TEV cleaved by adding TEV protease at a ratio of 1:10 protease to protein and incubated overnight at 25 °C. For biochemical assays and phage screening, cleaved protein was then biotin labeled via yBBR according to established methods(66). Cleaved protein was then concentrated and injected onto a Superdex® 75 10/300 GL gel filtration column equilibrated with formulation buffer (50 mM HEPES, 200 mM NaCl, 0.5 mM TCEP, 5% glycerol, pH 7.5). Protein fractions were collected, concentrated, aliquoted and frozen.

2.4.16.3. IRAK2

IRAK2 Death domain (residues 2-96, C3S) with an N-terminal 6xHis–MBP–3C–YBBR–TEV tag was recombinantly expressed in *E. coli* BL21 CodonPlus RIPL cells (Agilent) from pET-derived expression vectors (Novagen). The cells were induced at OD600 = 0.6 with 0.15 mM IPTG for 16 hours at 16°C, then harvested and resuspended in 20 mM Tris pH 8.0, 200 mM NaCl, 5% glycerol, and 20 mM imidazole. For purification, the pellet was lysed with a tip sonicator and pelleted at 22,000 x g for 30 minutes at 4 °C. The supernatant was purified with Ni-NTA resin (Qiagen), eluted with 20 mM Tris, pH 8.0, 200 mM NaCl, 5 % glycerol, and 250 mM imidazole. Protein was 3C-cleaved by adding Pierce™ HRV 3C protease (ThermoFisher) at a ratio of 1:10 protease to protein and incubated for 4 hours at 4 °C, and biotinylated via the yBBr reaction according to standard procedures(66). Biotinylated proteins were diluted into buffer 5 mM Tris, pH 8.0, 10 % glycerol, and 1 mM DTT, and loaded onto a Q-HP (Cytiva) column. Proteins were eluted with a gradient from 50 mM to 1000 mM NaCl. Protein elutes were pooled, concentrated, and loaded onto a Superdex™ 10/300 75 pg (Cytiva) SEC column, and eluted in 25 mM Tris pH 8.0, 200 mM NaCl, 5 % glycerol, and 1mM TCEP. Fractions containing pure protein were collected, pooled, concentrated to ~8 mg/mL and stored at -80 °C.

2.4.16.4. CHIP

For protein used in phage screen and biochemical assays: CHIP (STUB1, residues 23-154) with an N-terminal 6xHis–yBBr–TEV tag was recombinantly expressed in *E. coli* BL21 CodonPlus cells (Agilent) from pET-derived expression vectors (Novagen). The cells were induced at OD600 = 0.6 with 1 mM IPTG for 4 hours at 37 °C, then harvested and resuspended in 50 mM Tris-HCl pH 8.0, 500 mM NaCl ,10 mM imidazole, 10% glycerol, and 10 mM ß-me. For purification, the pellet was lysed with a tip sonicator and pelleted at 22,000 x g for 30 minutes at 4 °C. The supernatant was purified with Ni-NTA resin (Qiagen), eluted with 50 mM Tris-HCl pH 8.0, 500 mM NaCl, 10% glycerol, 10 mM ß-me and 250 mM imidazole, and biotinylated via the yBBr reaction according to standard procedures(66). Biotinylated proteins were pooled, concentrated, and loaded onto a Superdex™ 10/300 75 pg (Cytiva) SEC column, and eluted in 20 mM HEPES pH 7.0, 150 mM NaCl, 10% glycerol, 2 mM DTT. Fractions containing pure protein were collected, pooled, concentrated to ~1.2 mg/mL and stored at -80°C.

For protein used in competition FP and crystallography: CHIP (residues 23-303 for competition FP and 23–154 for crystallography) with an N-terminal 6xHis–TEV tag was recombinantly expressed in E. coli BL21 CodonPlus cells (Agilent) from a pET21b-derived expression vector (Novagen). The cells were induced at OD600 = 0.6 with 1 mM IPTG for 4 h at 37 °C, then harvested and resuspended in 50 mM Tris-HCl pH 8.0, 500 mM NaCl,10 mM imidazole, 10% glycerol, and 10 mM ß-me. For purification, the pellet was lysed with a tip sonicator with power level set at 400 W, 3 seconds on and 3 seconds off for 20 min, pelleted at 22,000 x g for 30 min at 4 °C. The supernatant was purified with Ni-NTA resin (Qiagen), eluting with 250 mM imidazole. Eluted proteins were pooled, concentrated, and TEV-cleaved by adding TEV protease at a ratio of 1:10 protease to protein and incubated overnight at 4 °C. TEV-cleaved proteins were loaded onto a Superdex™ 10/300 75 pg (Cytiva) SEC column, and eluted in 50 mM HEPES, pH 8.0, 150 mM NaCl, 10% glycerol, 2 mM DTT. Fractions containing pure protein were collected, pooled, concentrated to 30 mg/mL and stored at -80 °C.

2.4.16.5. NEMO

NEMO (IKBKG, residues 48-112) with an N-terminal 6xHis–yBBr–TEV tag was recombinantly expressed in *E. coli* BL21 CodonPlus RIPL cells (Agilent) from pET-derived expression vectors (Novagen). The cells were induced at OD600 = 0.6 with 1 mM IPTG for 4 h at 37 °C, then harvested and resuspended in 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 1M urea, 1.0% Triton X-100. For purification, the pellet was lysed with a tip sonicator, and pelleted at 22,000 x g for 30 min at 4 °C. The pellets were washed three times and the inclusion body was dissolved in 20 mM Tris-HCl pH 8.0, 150 mM NaCl, and 8 M urea. The supernatant was purified with Ni-NTA resin (Qiagen), eluting with 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 8 M urea, and 250 mM imidazole. Protein elutes were diluted dropwise diluted into 50 mM Tris-HCl pH 8.0, 200 mM NaCl, 100 mM $MgCl_2$, 100 mM $CaCl_2$, 200 mM L-Arginine, 1 mM DTT, 0.5 mM reduced glutathione (GSH), 0.05 mM oxidized glutathione (GSSG), and 10% glycerol, and then dialyzed into buffer 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM DTT, 0.5 mM GSH, 0.05 mM GSSG, and 10 % glycerol. Proteins were biotinylated via the yBBr reaction according to standard procedures (*66*). Biotinylated proteins were pooled, concentrated, and loaded onto a Superdex™ 10/300 200pg (Cytiva) SEC column and eluted in 50 mM Tris pH 7.5, 200 mM NaCl, 10% glycerol, and 1 mM DTT. Fractions containing pure protein were collected, pooled, concentrated to ~2.5 mg/mL and stored at -80 °C.

2.4.16.6. FKBP12

FKBP12 (residues 2-108) with an N-terminal 6xHis–yBBr–TEV tag was recombinantly expressed in E. coli BL21 CodonPlus RIPL cells (Agilent) from pET-derived expression vectors (Novagen). The cells were induced at OD600 = 0.6 with 0.15 mM IPTG for 16 h at 16 °C, then harvested and resuspended in 10 mM Tris-HCl pH 8.8, 200 mM NaCl, 10% glycerol, and 1 mM phenylmethylsulfonyl fluoride (PMSF). For purification, the pellet was lysed with a tip sonicator and pelleted at 22,000 x g for 30 min at 4 °C. The supernatant was purified with Ni-NTA resin (Qiagen), eluted with 20 mM Tris-HCl pH 8.8, 200 mM NaCl, 10% glycerol, 50 mM and 250 mM imidazole, and biotinylated via the yBBr reaction according to standard procedures(66). Biotinylated proteins were diluted into buffer 20 mM Tris-HCl pH 8.8, 10% glycerol, 1mM DTT, and loaded onto a Q-HP (Cytiva) column.

Proteins were eluted with a gradient from 50 mM to 1000 mM NaCl. Protein elutes were pooled, concentrated, and loaded onto a Superdex™ 10/300 75 pg (Cytiva) SEC column, and eluted in 20 mM Tris-HCl pH 7.4, 200 mM NaCl, 10% glycerol, and 1 mM DTT. Fractions containing pure protein were collected, pooled, concentrated to ~5.4 mg/mL and stored at -80°C.

2.4.16.7. BCL11A

BCL11a (740-835) with an N-terminal GST-3C-yBBR-TEV tag was recombinantly expressed in One Shot™ BL21 Star™ (DE3) Chemically Competent E. coli (ThermoFisher) from pET-derived expression vectors (ThermoFisher) in Luria broth supplemented with 0.1 mg/mL ampicillin and 100 µM zinc chloride. The cells were induced at OD600 = 0.2 with 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 16 h at 18 °C, then harvested and resuspended in lysis buffer (50 mM HEPES, 500 mM NaCl, 1 mM DTT, 2 mM ATP, 10 mM MgCl2, 1% NP40, 5% glycerol, Roche complete inhibitors, 100 µM ZnCl2, pH 7.5). For purification, the pellet was lysed with a tip sonicator then pelleted at 20,000 x g for 30 min at 4 °C. Supernatant was collected and purified with GST resin (Cytiva), eluting with 10 mM GSH. Protein was 3C cleaved by adding HRV 3C protease at a ratio of 1:10 protease to protein and incubated overnight at 25 °C. Cleaved protein was then concentrated and diluted into formulation buffer (50 mM HEPES, 200 mM NaCl, 5% glycerol, 1mM DTT, pH 7.5) and labeled with biotin via yBBR according to established methods (66). Labeled protein was injected onto a Superdex® 75 10/300 GL gel filtration column equilibrated with formulation buffer. Protein fractions were collected, concentrated, aliquoted and frozen.

2.4.16.8. YAP1-WW1-WW2

YAP1-WW1-WW2 (residues 164-266) with an N-terminal 6xhis-SUMO-yBBR tag was recombinantly expressed in One Shot™ BL21 Star™ (DE3) Chemically Competent *E. coli* (ThermoFisher) from pET-derived expression vectors (ThermoFisher) in Luria broth supplemented with 0.1 mg/mL ampicillin. The cells were induced at OD600 = 0.5 with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 8 h at 25 °C then overnight at 15 °C, then harvested and resuspended in lysis buffer (50 mM HEPES, 300 mM NaCl, 2 mM

ATP, 10 mM $MgCl_2$, 1% NP40, 5% glycerol, Roche complete inhibitors, pH 7.5). For purification, the pellet was lysed with a tip sonicator then pelleted at 20,000 x g for 30 min at 4 °C. Supernatant was collected and purified with cOmplete™ His-Tag Purification Columns (Roche), eluting with 300 mM imidazole. Protein was SUMO cleaved by adding SUMO protease at a ratio of 1:100 protease to protein and incubated overnight at 25 °C. Cleaved protein was then concentrated and diluted into formulation buffer (50 mM HEPES, 150 mM NaCl, 5% glycerol, pH 7.5) and labeled with biotin via yBBR according to established methods. Labeled protein was injected onto a Superdex® 75 10/300 GL gel filtration column equilibrated with formulation buffer. Protein fractions were collected, concentrated, aliquoted and frozen.

2.4.16.9. NEMO_iZIP

NEMO_iZIP (IKBKG, residues 51-112 with both N- and C-terminal coiled-coil based on GCN4(*67*) with an N-terminal 6xHis–yBBr–TEV tag was recombinantly expressed and refolded identical to NEMO, with pure protein collected, pooled, concentrated to ~1.0 mg/mL and stored at -80°C.

2.4.17. Crystal Structure Data Collection

To obtain the structures of the protein-peptide complexes, briefly, 10 mM peptides stock in 90% DMSO were added to the protein stocks to a final 1:1.25 protein: peptide molar ratio and screened against commercially available crystallization screens. Crystals were obtained by sitting hanging drop vapor diffusion methods at room temperature, with their crystallization conditions detailed in the Data Collection and Refinement Statistics tables below. Crystals were cryo-protected with glycerol or ethylene glycol followed by flash-freezing in liquid nitrogen. Diffraction datasets were collected at 100 K at a variety of sources as described in the tables. Data was processed in XDS(68), autoPROC(69), and AIMLESS(70). Molecular replacement solutions were obtained using PHASER(*71*) with previously deposited high resolution PDB structures as search models. Complete models were built through iterative cycles of manual model building in COOT(*72*) and structure refinement using REFMAC(*73*) and PHENIX(*74*). All the structure model figures in the paper were prepared using PyMOL (The PyMOL Molecular Graphics System, Version 2.4, Schrödinger, LLC), and ChimeraX(*75*). The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org.

2.4.17.1. MDM2-H101 Data Collection and Refinement Statistics

Title	**MDM2_D-H101**
PDB Entry	**8F0Z**
Description	Structure of the MDM2 P53 binding domain in complex with D-H101, an all-D Helicon Polypeptide
Wavelength (Å)	1.03319
Resolution range (Å)	40.87-1.61 (1.64-1.61)
Space group	P 21 21 2
Unit cell: a,b,c (Å), α, β, γ (°)	40.87 81.36 34.98 90 90 90
Total reflections	200195(10013)
Unique reflections	15696(774)
Multiplicity	12.8(12.9)
Completeness (%)	99.6(100.0)
Mean I/sigma(I)	19.3(2.1)
Wilson B-factor ($Å^2$)	27.45
R-merge	0.067(1.18)
R-meas	0.072(1.28)
R-pim	0.028(0.48)
CC1/2	0.998(0.774)
Reflections used in refinement	15649 (1542)
Reflections used for R-free	756 (74)
R-work	0.2209 (0.3047)
R-free	0.2423 (0.3188)
Number of non-hydrogen atoms	1006
macromolecules	923
ligands	16
solvent	67
Protein residues	104
RMS(bonds) (Å)	0.009
RMS(angles) (°)	1.33
Ramachandran favored (%)	100
Ramachandran allowed (%)	0
Ramachandran outliers (%)	0
Rotamer outliers (%)	4.40
Clashscore	7.97
Average B-factor ($Å^2$)	35.27
macromolecules	34.61
ligands	31.76
solvent	45.08

Title	**MDM2_D-H101**
PDB Entry	**8F0Z**
Crystallization conditions	3.5 M Sodium Formate pH 7.0
Beamline and Acknowledgements	APS, 23-ID-B: GM/CA@APS has been funded by the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006, P30GM138396). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357.

2.4.17.2. MDM2-D-H102 Data Collection and Refinement Statistics

Title	**MDM2_D-H102**
PDB Entry	**8F10**
Description	Structure of the MDM2 P53 binding domain in complex with D-H102, an all-D Helicon Polypeptide
Wavelength (Å)	0.97933
Resolution range (Å)	42.17-1.28 (1.30-1.28)
Space group	P 1 21 1
Unit cell: a,b,c (Å), α, β, γ (°)	28.15 39.92 42.36 90 95.44 90
Total reflections	99690(439)
Unique reflections	20614(333)
Multiplicity	4.8(1.3)
Completeness (%)	85.3(29.1)
Mean I/sigma(I)	18.2(2.1)
Wilson B-factor ($Å^2$)	10.14
R-merge	0.039(0.30)
R-meas	0.049(0.42)
R-pim	0.029(0.30)
CC1/2	0.999(0.909)
Reflections used in refinement	20091 (736)
Reflections used for R-free	968 (38)
R-work	0.1587 (0.2479)
R-free	0.1708 (0.2064)
Number of non-hydrogen atoms	1062
macromolecules	908
ligands	49
solvent	105
Protein residues	102
RMS(bonds) (Å)	0.013
RMS(angles) (°)	1.33
Ramachandran favored (%)	97.65
Ramachandran allowed (%)	2.35
Ramachandran outliers (%)	0
Rotamer outliers (%)	0
Clashscore	10.35
Average B-factor ($Å^2$)	17.25
macromolecules	15.16
ligands	26.08
solvent	31.17

Title	**MDM2_D-H102**
PDB Entry	**8F10**
Crystallization conditions	0.01 M tri-Sodium citrate, 33 % w/v PEG 6000
Beamline and Acknowledgements	NSLS II, 17-ID-2 FMX: The FMX beamline is part of the Center for BioMolecular Structure (CBMS) which is primarily supported by the NIH, NIGMS through a Center Core P30 grant (grant no. P30GM133893), and by the US Department of Energy (DOE), Office of Biological and Environmental Research (contract no. KP1607011). NSLS-II is supported in part by the US DOE, Office of Science, Office of Basic Energy Sciences Program (contract nos. DE-SC0012704; KC0401040).

2.4.17.3. MDM2-D-H103 Data Collection and Refinement Statistics

Title	**MDM2_D-H103**
PDB Entry	**8F12**
Description	Structure of the MDM2 P53 binding domain in complex with D-H103, an all-D Helicon Polypeptide
Wavelength (Å)	1.03317
Resolution range (Å)	49.11-1.86(1.90-1.86)
Space group	P 43 21 2
Unit cell: a,b,c (Å), α, β, γ (°)	60.31 60.31 84.59 90 90 90
Total reflections	354507(19344)
Unique reflections	13715(817)
Multiplicity	25.8(23.7)
Completeness (%)	99.8(98.2)
Mean I/sigma(I)	18.0(2.2)
Wilson B-factor ($Å^2$)	25.46
R-merge	0.113(1.58)
R-meas	0.117(1.64)
R-pim	0.031(0.46)
CC1/2	0.999(0.851)
Reflections used in refinement	13668 (1318)
Reflections used for R-free	661 (54)
R-work	0.2172 (0.2821)
R-free	0.2422 (0.3280)
Number of non-hydrogen atoms	1073
macromolecules	971
ligands	25
solvent	77
Protein residues	111
RMS(bonds) (Å)	0.014
RMS(angles) (°)	1.38
Ramachandran favored (%)	96.77
Ramachandran allowed (%)	3.23
Ramachandran outliers (%)	0
Rotamer outliers (%)	1.03
Clashscore	4.01
Average B-factor ($Å^2$)	37.11
macromolecules	35.83
ligands	54.29
solvent	47.68

Title	**MDM2_D-H103**
PDB Entry	**8F12**
Crystallization conditions	3.2 M Ammonium sulfate, 0.1 M Citrate pH 5.0
Beamline and Acknowledgements	APS, 23-ID-B: GM/CA@APS has been funded by the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006, P30GM138396). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357.

2.4.17.4. MDM2-D-H103-Alt Data Collection and Refinement Statistics

Title	**MDM2_D-H103-Alt**
PDB Entry	**8F13**
Description	Structure of the MDM2 P53 binding domain in complex with D-H103, an all-D Helicon Polypeptide, alternative C-terminus
Wavelength (Å)	1.03319
Resolution range (Å)	42.45-1.40(1.42-1.40)
Space group	P 43 21 2
Unit cell: a,b,c (Å), α, β, γ (°)	60.29 60.29 84.90 90 90 90
Total reflections	645852(13067)
Unique reflections	31020(1334)
Multiplicity	20.8(9.8)
Completeness (%)	98.4(86.0)
Mean I/sigma(I)	23.0(2.3)
Wilson B-factor ($Å^2$)	19.45
R-merge	0.083(1.27)
R-meas	0.087(1.42)
R-pim	0.025(0.61)
CC1/2	0.996(0.771)
Reflections used in refinement	30774 (2630)
Reflections used for R-free	1526 (121)
R-work	0.1744 (0.3414)
R-free	0.2078 (0.3765)
Number of non-hydrogen atoms	1175
macromolecules	974
ligands	51
solvent	150
Protein residues	112
RMS(bonds) (Å)	0.009
RMS(angles) (°)	1.44
Ramachandran favored (%)	97.85
Ramachandran allowed (%)	2.15
Ramachandran outliers (%)	0
Rotamer outliers (%)	6.25
Clashscore	7.27
Average B-factor ($Å^2$)	29.64
macromolecules	26.74
ligands	37.87
solvent	45.69

Title	**MDM2_D-H103-Alt**
PDB Entry	**8F13**
Crystallization conditions	0.2 M Potassium sodium tartrate, 0.1 M tri-Sodium citrate pH 5.6, 2.0 M Ammonium sulfate
Beamline and Acknowledgements	APS, 23-ID-B: GM/CA@APS has been funded by the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006, P30GM138396). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357.

2.4.17.5. CHIP-D-H201 Data Collection and Refinement Statistics

Title	**CHIP06_D-H201**
PDB Entry	**8F14**
Description	Structure of the STUB1 TPR domain in complex with D-H201, an all-D Helicon Polypeptide
Wavelength (Å)	0.89685
Resolution range (Å)	92.33-1.69(1.72-1.69)
Space group	C 2 2 21
Unit cell: a,b,c (Å), α, β, γ (°)	46.16 90.12 92.33 90 90 90
Total reflections	296534(14797)
Unique reflections	21997(1103)
Multiplicity	13.5(13.4)
Completeness (%)	100.0(100.0)
Mean I/sigma(I)	20.2(2.2)
Wilson B-factor ($Å^2$)	23.75
R-merge	0.069(1.22)
R-meas	0.075(1.32)
R-pim	0.028(0.50)
CC1/2	0.999(0.850)
Reflections used in refinement	21945 (2153)
Reflections used for R-free	1106 (116)
R-work	0.1967 (0.2927)
R-free	0.2351 (0.3368)
Number of non-hydrogen atoms	1422
macromolecules	1256
ligands	32
solvent	134
Protein residues	149
RMS(bonds) (Å)	0.011
RMS(angles) (°)	1.20
Ramachandran favored (%)	98.46
Ramachandran allowed (%)	1.54
Ramachandran outliers (%)	0
Rotamer outliers (%)	2.63
Clashscore	2.78
Average B-factor ($Å^2$)	34.98
macromolecules	34.11
ligands	39.38
solvent	42.10

Title	**CHIP06_D-H201**
PDB Entry	**8F14**
Crystallization conditions	0.2 M Lithium Sulfate, 0.1 M Bis-Tris pH 6.5, 25% w/v PEG 3350
Beamline and Acknowledgements	Diamond I03: Diamond Light Source for beamtime, and the staff of beamlines I03 for assistance with crystal testing and data collection.

2.4.17.6. CHIP-D-H202 Data Collection and Refinement Statistics

Title	**CHIP06_D-H202**
PDB Entry	**8F15**
Description	Structure of the STUB1 TPR domain in complex with D-H202, an all-D Helicon Polypeptide
Wavelength (Å)	1.00000
Resolution range (Å)	49.02-1.73(1.76-1.73)
Space group	P 43
Unit cell: a,b,c (Å), α, β, γ (°)	50.56 50.56 199.85 90 90 90
Total reflections	721070(40585)
Unique reflections	51998(2848)
Multiplicity	13.9(14.3)
Completeness (%)	100.0(100.0)
Mean I/sigma(I)	16.6(2.2)
Wilson B-factor ($Å^2$)	17.46
R-merge	0.102(1.24)
R-meas	0.110(1.33)
R-pim	0.042(0.50)
CC1/2	0.999(0.785)
Reflections used in refinement	51994 (5201)
Reflections used for R-free	2477 (287)
R-work	0.1800 (0.2395)
R-free	0.2096 (0.2729)
Number of non-hydrogen atoms	4113
macromolecules	3702
ligands	79
solvent	332
Protein residues	448
RMS(bonds) (Å)	0.014
RMS(angles) (°)	1.43
Ramachandran favored (%)	98.73
Ramachandran allowed (%)	1.27
Ramachandran outliers (%)	0
Rotamer outliers (%)	3.78
Clashscore	3.91
Average B-factor ($Å^2$)	27.58
macromolecules	26.55
ligands	30.81
solvent	38.32

Title	**CHIP06_D-H202**
PDB Entry	**8F15**
Crystallization conditions	0.1 M TRIS pH 8, 30% v/v PEG 400
Beamline and Acknowledgements	Spring8 BL45XU: RIKEN Structural Biology beamline I BL45XU Spring-8 in Japan.

2.4.17.7. CHIP-D-H203 Data Collection and Refinement Statistics

Title	**CHIP06_D-H203**
PDB Entry	**8F16**
Description	Structure of the STUB1 TPR domain in complex with D-H203, an all-D Helicon Polypeptide
Wavelength (Å)	1.18057
Resolution range (Å)	45.41-1.56(1.59-1.56)
Space group	P 21 21 21
Unit cell: a,b,c (Å), α, β, γ (°)	50.56 66.88 103.28 90 90 90
Total reflections	315958(16165)
Unique reflections	50341(2511)
Multiplicity	6.3(6.4)
Completeness (%)	99.5(100.0)
Mean I/sigma(I)	22.4(2.2)
Wilson B-factor ($Å^2$)	22.78
R-merge	0.030(0.68)
R-meas	0.036(0.80)
R-pim	0.019(0.43)
CC1/2	1.000(0.893)
Reflections used in refinement	50091 (5005)
Reflections used for R-free	2436 (257)
R-work	0.1979 (0.2629)
R-free	0.2302 (0.3036)
Number of non-hydrogen atoms	2942
macromolecules	2556
ligands	60
solvent	326
Protein residues	301
RMS(bonds) (Å)	0.011
RMS(angles) (°)	1.25
Ramachandran favored (%)	99.23
Ramachandran allowed (%)	0.77
Ramachandran outliers (%)	0
Rotamer outliers (%)	3.39
Clashscore	5.83
Average B-factor ($Å^2$)	28.91
macromolecules	27.74
ligands	34.66
solvent	37.03

Title	**CHIP06_D-H203**
PDB Entry	**8F16**
Crystallization conditions	0.1 M BICINE pH 8.5, 15% w/v PEG 1,500
Beamline and Acknowledgements	CLS, 08B1: Part or all of the research described in this paper was performed using beamline CMCF-BM at the Canadian Light Source, a national research facility of the University of Saskatchewan, which is supported by the Canada Foundation for Innovation (CFI), the Natural Sciences and Engineering Research Council (NSERC), the National Research Council (NRC), the Canadian Institutes of Health Research (CIHR), the Government of Saskatchewan, and the University of Saskatchewan.

2.4.17.8. CHIP-D-H204 Data Collection and Refinement Statistics

Title	**CHIP06_D-H204**
PDB Entry	**8F17**
Description	Structure of the STUB1 TPR domain in complex with D-H204, an all-D Helicon Polypeptide
Wavelength (Å)	1.18057
Resolution range (Å)	46.01-2.21(2.28-2.21)
Space group	P 1 21 1
Unit cell: a,b,c (Å), α, β, γ (°)	38.03 72.71 59.79 90 96.31 90
Total reflections	53489(4781)
Unique reflections	16124(1388)
Multiplicity	3.3(3.4)
Completeness (%)	99.0(98.6)
Mean I/sigma(I)	16.9(2.3)
Wilson B-factor ($Å^2$)	38.8
R-merge	0.032(0.40)
R-meas	0.045(0.56)
R-pim	0.032(0.40)
CC1/2	0.999(0.883)
Reflections used in refinement	15189 (1493)
Reflections used for R-free	764 (84)
R-work	0.2166 (0.2997)
R-free	0.2523 (0.3046)
Number of non-hydrogen atoms	2426
macromolecules	2338
ligands	36
solvent	52
Protein residues	292
RMS(bonds) (Å)	0.003
RMS(angles) (°)	0.51
Ramachandran favored (%)	97.64
Ramachandran allowed (%)	2.36
Ramachandran outliers (%)	0
Rotamer outliers (%)	1.40
Clashscore	4.70
Average B-factor ($Å^2$)	50.27
macromolecules	50.33
ligands	50.91
solvent	47.19

Title	**CHIP06_D-H204**
PDB Entry	**8F17**
Crystallization conditions	0.15 M Sodium chloride, 0.1 M TRIS pH 8, 8% w/v PEG 6000
Beamline and Acknowledgements	CLS, 08B1: Part or all of the research described in this paper was performed using beamline CMCF-BM at the Canadian Light Source, a national research facility of the University of Saskatchewan, which is supported by the Canada Foundation for Innovation (CFI), the Natural Sciences and Engineering Research Council (NSERC), the National Research Council (NRC), the Canadian Institutes of Health Research (CIHR), the Government of Saskatchewan, and the University of Saskatchewan.

2.5. References

1. R. C. Milton, S. C. Milton, S. B. Kent, Total chemical synthesis of a D-enzyme: the enantiomers of HIV-1 protease show reciprocal chiral substrate specificity [corrected]. *Science*. **256**, 1445–1448 (1992).

2. Z. Wang, W. Xu, L. Liu, T. F. Zhu, A synthetic molecular system capable of mirror-image genetic replication and transcription. *Nature Chemistry*. **8**, 698–704 (2016).

3. Y. Xu, T. F. Zhu, Mirror-image T7 transcription of chirally inverted ribosomal and functional RNAs. *Science*. **378**, 405–412 (2022).

4. C. Fan, Q. Deng, T. F. Zhu, Bioorthogonal information storage in l-DNA with a high-fidelity mirror-image Pfu DNA polymerase. *Nature Biotechnology*. **39**, 1548–1555 (2021).

5. L. Wang, N. Wang, W. Zhang, X. Cheng, Z. Yan, G. Shao, X. Wang, R. Wang, C. Fu, Therapeutic peptides: current applications and future directions. *Signal Transduction and Targeted Therapy*. **7**, 48 (2022).

6. M. Muttenthaler, G. F. King, D. J. Adams, P. F. Alewood, Trends in peptide drug discovery. *Nature Reviews Drug Discovery*. **20**, 309–325 (2021).

7. Z. Feng, B. Xu, Inspiration from the mirror: D-amino acid containing peptides in biomedical approaches. *Biomolecular Concepts*. **7**, 179–187 (2016).

8. T. Kremsmayr, A. Aljnabi, J. B. Blanco-Canosa, H. N. T. Tran, N. B. Emidio, M. Muttenthaler, *J. Med. Chem.*, *in press*, doi:10.1021/acs.jmedchem.2c00094.

9. T. N. M. Schumacher, L. M. Mayr, D. L. Minor, M. A. Milhollen, M. W. Burgess, P. S. Kim, Identification of d-Peptide Ligands Through Mirror-Image Phage Display. *Science*. **271**, 1854–1857 (1996).

10. M. Liu, M. Pazgier, C. Li, W. Yuan, C. Li, W. Lu, A Left-Handed Solution to Peptide Inhibition of the p53–MDM2 Interaction. *Angewandte Chemie International Edition*. **49**, 3649–3652 (2010).

11. C. Díaz-Perlas, M. Varese, S. Guardiola, M. Sánchez-Navarro, J. García, M. Teixidó, E. Giralt, Protein Chemical Synthesis Combined with Mirror-Image Phage Display Yields d-Peptide EGF Ligands that Block the EGF–EGFR Interaction. *ChemBioChem*. **20**, 2079–2084 (2019).

12. K. Wiesehan, K. Buder, R. P. Linke, S. Patt, M. Stoldt, E. Unger, B. Schmitt, E. Bucci, D. Willbold, Selection of D-Amino-Acid Peptides That Bind to Alzheimer's Disease Amyloid Peptide Aβ1–42 by Mirror Image Phage Display. *ChemBioChem*. **4**, 748–753 (2003).

13. D. M. Eckert, V. N. Malashkevich, L. H. Hong, P. A. Carr, P. S. Kim, Inhibiting HIV-1 Entry: Discovery of D-Peptide Inhibitors that Target the gp41 Coiled-Coil Pocket. *Cell*. **99**, 103–115 (1999).

14. P. S. Marinec, K. E. Landgraf, M. Uppalapati, G. Chen, D. Xie, Q. Jiang, Y. Zhao, A. Petriello, K. Deshayes, S. B. H. Kent, D. Ault-Riche, S. S. Sidhu, A Non-immunogenic Bivalent d-Protein Potently Inhibits Retinal Vascularization and Tumor Growth. *ACS Chem. Biol.* **16**, 548–556 (2021).

15. X. Zhou, C. Zuo, W. Li, W. Shi, X. Zhou, H. Wang, S. Chen, J. Du, G. Chen, W. Zhai, W. Zhao, Y. Wu, Y. Qi, L. Liu, Y. Gao, A Novel d-Peptide Identified by Mirror-Image Phage Display Blocks TIGIT/PVR for Cancer Immunotherapy. *Angewandte Chemie International Edition*. **59**, 15114–15118 (2020).

16. H.-N. Chang, B.-Y. Liu, Y.-K. Qi, Y. Zhou, Y.-P. Chen, K.-M. Pan, W.-W. Li, X.-M. Zhou, W.-W. Ma, C.-Y. Fu, Y.-M. Qi, L. Liu, Y.-F. Gao, Blocking of the PD-1/PD-L1 Interaction by a D-Peptide Antagonist for Cancer Immunotherapy. *Angewandte Chemie International Edition*. **54**, 11760–11764 (2015).

17. K. Mandal, M. Uppalapati, D. Ault-Riché, J. Kenney, J. Lowitz, S. S. Sidhu, S. B. H. Kent, Chemical synthesis and X-ray structure of a heterochiral {D-protein antagonist plus vascular endothelial growth factor} protein complex by racemic crystallography. *Proceedings of the National Academy of Sciences*. **109**, 14779–14784 (2012).

18. M. Liu, C. Li, M. Pazgier, C. Li, Y. Mao, Y. Lv, B. Gu, G. Wei, W. Yuan, C. Zhan, W.-Y. Lu, W. Lu, D-peptide inhibitors of the p53–MDM2 interaction for targeted molecular therapy of malignant neoplasms. *Proceedings of the National Academy of Sciences*. **107**, 14321–14326 (2010).

19. L. Huang, J. Xie, Q. Bi, Z. Li, S. Liu, Q. Shen, C. Li, Highly Selective Targeting of Hepatic Stellate Cells for Liver Fibrosis Treatment Using a d-Enantiomeric Peptide Ligand of Fn14 Identified by Mirror-Image mRNA Display. *Mol. Pharmaceutics*. **14**, 1742–1753 (2017).

20. Z. Li, J. Xie, S. Peng, S. Liu, Y. Wang, W. Lu, J. Shen, C. Li, Novel Strategy Utilizing Extracellular Cysteine-Rich Domain of Membrane Receptor for Constructing d-Peptide Mediated Targeted Drug Delivery Systems: A Case Study on Fn14. *Bioconjugate Chem*. **28**, 2167–2179 (2017).

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

22. P. E. Dawson, S. B. H. Kent, Synthesis of Native Proteins by Chemical Ligation. *Annu. Rev. Biochem*. **69**, 923–960 (2000).

23. 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. *Nature Chemical Biology*. **13** (2017), p. 464.

24. N. Hartrampf, A. Saebi, M. Poskus, Z. P. Gates, A. J. Callahan, A. E. Cowfer, S. Hanna, S. Antilla, C. K. Schissel, A. J. Quartararo, X. Ye, A. J. Mijalis, M. D. Simon, A. Loas, S. Liu, C. Jessen, T. E. Nielsen, B. L. Pentelute, Synthesis of proteins by automated flow chemistry. *Science*. **368**, 980 (2020).

25. C. Zuo, B. Zhang, M. Wu, D. Bierer, J. Shi, G.-M. Fang, Chemical synthesis and racemic crystallization of rat C5a-desArg. *Chinese Chemical Letters*. **31**, 693–696 (2020).

26. J.-J. Ling, C. Fan, H. Qin, M. Wang, J. Chen, P. Wittung-Stafshede, T. F. Zhu, Mirror-Image 5S Ribonucleoprotein Complexes. *Angewandte Chemie International Edition*. **59**, 3724–3731 (2020).

27. C. Zuo, W.-W. Shi, X.-X. Chen, M. Glatz, B. Riedl, I. Flamme, E. Pook, J. Wang, G.-M. Fang, D. Bierer, L. Liu, Chimeric protein probes for C5a receptors through fusion of the anaphylatoxin C5a core region with a small-molecule antagonist. *Science China Chemistry*. **62**, 1371–1378 (2019).

28. J. Weidmann, M. Schnölzer, P. E. Dawson, J. D. Hoheisel, Copying Life: Synthesis of an Enzymatically Active Mirror-Image DNA-Ligase Made of D-Amino Acids. *Cell Chemical Biology*. **26**, 645-651.e3 (2019).

29. K. Shu, N. Iwamoto, K. Honda, Y. Kondoh, H. Hirano, H. Osada, H. Ohno, N. Fujii, S. Oishi, Development of Mirror-Image Screening Systems for XIAP BIR3 Domain Inhibitors. *Bioconjugate Chem*. **30**, 1395–1404 (2019).

30. M. Wang, W. Jiang, X. Liu, J. Wang, B. Zhang, C. Fan, L. Liu, G. Pena-Alcantara, J.-J. Ling, J. Chen, T. F. Zhu, Mirror-Image Gene Transcription and Reverse Transcription. *Chem*. **5**, 848–857 (2019).

31. Q.-Y. Guo, L.-H. Zhang, C. Zuo, D.-L. Huang, Z. A. Wang, J.-S. Zheng, C.-L. Tian, Channel activity of mirror-image M2 proton channel of influenza A virus is blocked by achiral or chiral inhibitors. *Protein & Cell*. **10**, 211–216 (2019).

32. L. De Rosa, R. Di Stasi, L. D. D'Andrea, Total chemical synthesis by native chemical ligation of the all-D immunoglobulin-like domain 2 of Axl. *Tetrahedron*. **75**, 894–905 (2019).

33. C.-C. Chen, S. Gao, H.-S. Ai, Q. Qu, C.-L. Tian, Y.-M. Li, Racemic X-ray structure of L-type calcium channel antagonist Calciseptine prepared by total chemical synthesis. *Science China Chemistry*. **61**, 702–707 (2018).

34. A. M. Levinson, J. H. McGee, A. G. Roberts, G. S. Creech, T. Wang, M. T. Peterson, R. C. Hendrickson, G. L. Verdine, S. J. Danishefsky, Total Chemical Synthesis and Folding of All-l and All-d Variants of Oncogenic KRas(G12V). *Journal of the American Chemical Society*. **139** (2017), pp. 7632–7639.

35. A. Pech, J. Achenbach, M. Jahnz, S. Schülzchen, F. Jarosch, F. Bordusa, S. Klussmann, A thermostable d-polymerase for mirror-image PCR. *Nucleic Acids Research*. **45**, 3997–4005 (2017).

36. W. Xu, W. Jiang, J. Wang, L. Yu, J. Chen, X. Liu, L. Liu, T. F. Zhu, Total chemical synthesis of a thermostable enzyme capable of polymerase chain reaction. *Cell Discovery*. **3**, 17008 (2017).

37. T. Noguchi, H. Ishiba, K. Honda, Y. Kondoh, H. Osada, H. Ohno, N. Fujii, S. Oishi, Synthesis of Grb2 SH2 Domain Proteins for Mirror-Image Screening Systems. *Bioconjugate Chem*. **28**, 609–619 (2017).

38. K. Shu, T. Noguchi, K. Honda, Y. Kondoh, H. Osada, H. Ohno, N. Fujii, S. Oishi, Synthesis of the Src SH2 domain and its application in bioassays for mirror-image screening. *RSC Adv*. **7**, 38725–38732 (2017).

39. C. Guo, L.-Q. Zhang, W. Jiang, Biodegrading plastics with a synthetic non-biodegradable enzyme. *Chem*, doi:10.1016/j.chempr.2022.09.008.

40. K. Li, O. S. Tokareva, T. M. Thomson, S. C. T. Wahl, T. L. Travaline, J. D. Ramirez, S. K. Choudary, S. Agarwal, W. G. Walkup, T. J. Olsen, M. J. Brennan, G. L. Verdine, J. H. McGee, De novo mapping of α-helix recognition sites on protein surfaces using unbiased libraries. *Proceedings of the National Academy of Sciences*. **119**, e2210435119 (2022).

41. A. A. Vinogradov, E. D. Evans, B. L. Pentelute, Total synthesis and biochemical characterization of mirror image barnase. *Chem. Sci*. **6**, 2997–3002 (2015).

42. M. Jbara, S. Pomplun, C. K. Schissel, S. W. Hawken, A. Boija, I. Klein, J. Rodriguez, S. L. Buchwald, B. L. Pentelute, Engineering Bioactive Dimeric Transcription Factor Analogs via Palladium Rebound Reagents. *Journal of the American Chemical Society*. **143**, 11788–11798 (2021).

43. B. Guo, C. O. Audu, J. C. Cochran, D. F. Mierke, M. Pellegrini, Protein Engineering of the N-Terminus of NEMO: Structure Stabilization and Rescue of IKKβ Binding. *Biochemistry*. **53**, 6776–6785 (2014).

44. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature*. **596**, 583–589 (2021).

45. M. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Žídek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, E. Clancy, R. Green, A. Vora, M. Lutfi, M. Figurnov, A. Cowie, N. Hobbs, P. Kohli, G. Kleywegt, E. Birney, D. Hassabis, S. Velankar, AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*. **50**, D439–D444 (2022).

46. 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. *Journal of the American Chemical Society*. **134**, 6855–6864 (2012).

47. M. Rushe, L. Silvian, S. Bixler, L. L. Chen, A. Cheung, S. Bowes, H. Cuervo, S. Berkowitz, T. Zheng, K. Guckian, M. Pellegrini, A. Lugovskoy, Structure of a NEMO/IKK-Associating Domain Reveals Architecture of the Interaction Site. *Structure*. **16**, 798–808 (2008).

48. Y. Yang, Z. Xu, C. He, B. Zhang, Y. Shi, F. Li, Structural insights into the recognition of γ-globin gene promoter by BCL11A. *Cell Research*. **29**, 960–963 (2019).

49. B. R. Kelemen, T. A. Klink, M. A. Behike, S. R. Eubanks, P. A. Leland, R. T. Raines, Hypersensitive substrate for ribonucleases. *Nucleic Acids Research*. **27**, 3696–3701 (1999).

50. C. D. Mackereth, M. Schärpf, L. N. Gentile, S. E. MacIntosh, C. M. Slupsky, L. P. McIntosh, Diversity in Structure and Function of the Ets Family PNT Domains. *Journal of Molecular Biology*. **342**, 1249–1264 (2004).

51. S.-C. Lin, Y.-C. Lo, H. Wu, Helical assembly in the MyD88–IRAK4–IRAK2 complex in TLR/IL-1R signalling. *Nature*. **465**, 885–890 (2010).

52. L. T. Vassilev, B. T. Vu, B. Graves, D. Carvajal, F. Podlaski, Z. Filipovic, N. Kong, U. Kammlott, C. Lukacs, C. Klein, N. Fotouhi, E. A. Liu, In Vivo Activation of the p53 Pathway by Small-Molecule Antagonists of MDM2. *Science*. **303**, 844–848 (2004).

53. P. Chène, Inhibiting the p53–MDM2 interaction: an important target for cancer therapy. *Nature Reviews Cancer*. **3**, 102–109 (2003).

54. K. H. Khoo, C. S. Verma, D. P. Lane, Drugging the p53 pathway: understanding the route to clinical efficacy. *Nature Reviews Drug Discovery*. **13**, 217–236 (2014).

55. S. Kannan, P. G. A. Aronica, S. Ng, D. T. Gek Lian, Y. Frosi, S. Chee, J. Shimin, T. Y. Yuen, A. Sadruddin, H. Y. K. Kaan, A. Chandramohan, J. H. Wong, Y. S. Tan, Z. W. Chang, F. J. Ferrer-Gago, P. Arumugam, Y. Han, S. Chen, L. Rénia, C. J. Brown, C. W. Johannes, B. Henry, D. P. Lane, T. K. Sawyer, C. S. Verma, A. W. Partridge, Macrocyclization of an all-d linear α-helical peptide imparts cellular permeability. *Chem. Sci*. **11**, 5577–5591 (2020).

56. Y. S. Chang, B. Graves, V. Guerlavais, C. Tovar, K. Packman, K.-H. To, K. A. Olson, K. Kesavan, P. Gangurde, A. Mukherjee, T. Baker, K. Darlak, C. Elkin, Z. Filipovic, F. Z. Qureshi, H. Cai, P. Berry, E. Feyfant, X. E. Shi, J. Horstick, D. A. Annis, A. M. Manning, N. Fotouhi, H. Nash, L. T. Vassilev, T. K. Sawyer, Stapled α−helical peptide drug development: A potent dual inhibitor of

MDM2 and MDMX for p53-dependent cancer therapy. *Proceedings of the National Academy of Sciences*. **110**, E3445–E3454 (2013).

57. S.-B. Qian, H. McDonough, F. Boellmann, D. M. Cyr, C. Patterson, CHIP-mediated stress recovery by sequential ubiquitination of substrates and Hsp70. *Nature*. **440**, 551–555 (2006).

58. J. Jiang, C. A. Ballinger, Y. Wu, Q. Dai, D. M. Cyr, J. Höhfeld, C. Patterson, CHIP Is a U-box-dependent E3 Ubiquitin Ligase: IDENTIFICATION OF Hsc70 AS A TARGET FOR UBIQUITYLATION *. *Journal of Biological Chemistry*. **276**, 42938–42944 (2001).

59. P. Connell, C. A. Ballinger, J. Jiang, Y. Wu, L. J. Thompson, J. Höhfeld, C. Patterson, The co-chaperone CHIP regulates protein triage decisions mediated by heat-shock proteins. *Nature Cell Biology*. **3**, 93–96 (2001).

60. M. Ravalin, P. Theofilas, K. Basu, K. A. Opoku-Nsiah, V. A. Assimon, D. Medina-Cleghorn, Y.-F. Chen, M. F. Bohn, M. Arkin, L. T. Grinberg, C. S. Craik, J. E. Gestwicki, Specificity for latent C termini links the E3 ubiquitin ligase CHIP to caspases. *Nature Chemical Biology*. **15**, 786–794 (2019).

61. G. Apriamashvili, D. W. Vredevoogd, O. Krijgsman, O. B. Bleijerveld, M. A. Ligtenberg, B. de Bruijn, J. Boshuizen, J. J. H. Traets, D. D'Empaire Altimari, A. van Vliet, C.-P. Lin, N. L. Visser, J. D. Londino, R. Sanchez-Hodge, L. E. Oswalt, S. Altinok, J. C. Schisler, M. Altelaar, D. S. Peeper, Ubiquitin ligase STUB1 destabilizes IFNγ-receptor complex to suppress tumor IFNγ signaling. *Nature Communications*. **13**, 1923 (2022).

62. S. Ng, A. C. Brueckner, S. Bahmanjah, Q. Deng, J. M. Johnston, L. Ge, R. Duggal, B. Habulihaz, B. Barlock, S. Ha, A. Sadruddin, C. Yeo, C. Strickland, A. Peier, B. Henry, E. C. Sherer, A. W. Partridge, Discovery and Structure-Based Design of Macrocyclic Peptides Targeting STUB1. *J. Med. Chem*. **65**, 9789–9801 (2022).

63. A. Henninot, J. C. Collins, J. M. Nuss, The Current State of Peptide Drug Discovery: Back to the Future? *J. Med. Chem*. **61**, 1382–1414 (2018).

64. T. J. Tucker, M. W. Embrey, C. Alleyne, R. P. Amin, A. Bass, B. Bhatt, E. Bianchi, D. Branca, T. Bueters, N. Buist, S. N. Ha, M. Hafey, H. He, J. Higgins, D. G. Johns, A. D. Kerekes, K. A. Koeplinger, J. T. Kuethe, N. Li, B. Murphy, P. Orth, S. Salowe, A. Shahripour, R. Tracy, W. Wang, C. Wu, Y. Xiong, H. J. Zokian, H. B. Wood, A. Walji, A Series of Novel, Highly Potent, and Orally Bioavailable Next-Generation Tricyclic Peptide PCSK9 Inhibitors. *J. Med. Chem*. **64**, 16770–16800 (2021).

65. Ballantyne Christie M., Banka Puja, Mendez Gustavo, Garcia Raymundo, Rosenstock Julio, Rodgers Anthony, Mendizabal Geraldine, Mitchel Yale, Catapano Alberico L., Efficacy and safety of the oral PCSK9 inhibitor MK-0616: a phase 2b randomized controlled trial. *Journal of the American College of Cardiology*. **0**, doi:10.1016/j.jacc.2023.02.018.

66. J. Yin, P. D. Straight, S. M. McLoughlin, Z. Zhou, A. J. Lin, D. E. Golan, N. L. Kelleher, R. Kolter, C. T. Walsh, Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proceedings of the National Academy of Sciences*. **102**, 15815–15820 (2005).

67. A. H. Barczewski, M. J. Ragusa, D. F. Mierke, M. Pellegrini, The IKK-binding domain of NEMO is an irregular coiled coil with a dynamic binding interface. *Scientific Reports*. **9**, 2950 (2019).

68. W. Kabsch, XDS. *Acta Crystallographica Section D*. **66**, 125–132 (2010).

69. M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson, Overview of the ıt CCP4 suite and current developments. *Acta Crystallographica Section D*. **67**, 235–242 (2011).

70. P. R. Evans, G. N. Murshudov, How good are my data and what is the resolution? *Acta Crystallographica Section D*. **69**, 1204–1214 (2013).

71. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, Phaser crystallographic software. *Journal of Applied Crystallography*. **40**, 658–674 (2007).

72. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallographica Section D*. **66**, 486–501 (2010).

73. G. N. Murshudov, P. Skubák, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, A. A. Vagin, REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D*. **67**, 355–367 (2011).

74. D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkóczi, V. B. Chen, T. I. Croll, B. Hintze, L.-W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams, P. D. Adams, Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallographica Section D*. **75**, 861–877 (2019).

75. T. D. Goddard, C. C. Huang, E. C. Meng, E. F. Pettersen, G. S. Couch, J. H. Morris, T. E. Ferrin, UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science*. **27**, 14–25 (2018).

2.6. Acknowledgements

We thank Mirella Bucci and Jacob Rodriguez for their valuable comments during manuscript preparation. Please refer to the Supplementary Information Section 11 for acknowledgments related to X-ray crystallography data collection.

2.7. Author Contributions

A.J.C, G.L.V., J.H.M., and B.L.P. conceptualized the research. A.J.C., S.G., L.L.S., S.H., and Y.C.L. carried out protein synthesis and A.J.C. carried out protein folding. T.L.T. and A.J.C. carried out protein biochemical validation. O.S.T., K.L., and J.M.S. carried out phage screening and binder validation. K.L. carried out protein crystallization and structure refinement. A.J.C. and K.L. generated figures. A.J.C., A.L., J.H.M., and B.L.P. wrote the manuscript with input from all authors.

3. Same Day Access to Proteins Enabled by Folding Selections

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

Callahan, A.J.; Rondon, A.; Lozano Salazar, L.; Gandhesiri, S.; Rodriguez, J.; Loas, A.; Pentelute, B.L. Same Day Access to Proteins Enabled by Folding Selections. *ChemRxiv*; **2023**; This content is a preprint and has not been peer-reviewed.

3.1.Introduction

The standardized production of proteins with heterologous expression has revolutionized the study and manipulation of biological systems. Proteins carry out many processes essential to life, with humans expressing more than 20,000 different proteins (*1*). Moreover, studying individual proteins informs fundamental biological understanding, such as the origins of human disease (*2*) or the mechanisms underpinning consciousness (*3*). Modern recombinant expression techniques are responsible for obtaining almost every protein studied in the laboratory (*4*–*7*). However, despite their widespread implementation in industry and academia, carrying out a complete protein production cycle from DNA plasmid design to purification can take up to several weeks (*8*). In these techniques, DNA sequences encoding the protein target of interest are synthesized or amplified from cDNA libraries and copied into an expression vector. This vector is then over-expressed in a host cell, with subsequent purification affording the encoded protein sequence (Figure 1). New technologies that enable access to folded proteins in hours are thus highly desirable.

Alternatively, chemical protein synthesis can prepare entire unfolded sequences in hours; however, downstream handling and purification take weeks (9, *10*). Fmoc-based solid-phase peptide synthesis (Fmoc-SPPS) is a chemical process for assembling component amino acids into peptide chains (Figure 1). When paired with multi-fragment ligation techniques, SPPS can afford synthetic proteins of significant complexity (*11*), for example, in the decade-long campaigns to produce the highly glycosylated erythropoietin (*12*) and the 53 kDa tetra-Ub-α-globin construct (*13*). Despite this scope, simple SPPS targets require days of synthesis, with total production and purification taking weeks to many months. To accelerate this process, we recently developed a rapid synthesis technique, termed single-shot automated fast-flow protein synthesis (AFPS) (*9*, *14*). AFPS rapidly provides denatured, single-domain proteins of up to ~200 amino acids (*15*) in hours, however, the necessary downstream steps remain unchanged from traditional workflows resulting in weeks of work.

Recombinant proteins are routinely isolated from highly complex cell-lysate inputs based on folded structure alone (*16*, *17*), and we hypothesized that a similar approach could be applied to crude mixtures from protein chemical synthesis. Minor modifications

in a protein's chemical structure can enact significant conformational changes. For example, deleting or inserting a single amino acid within a secondary structure unit frequently results in structural accommodations that significantly disrupt protein activity (*18–21*). In addition, our previous work has shown that random amino acid substitutions for an unnatural chemical structure (D-alanine) frequently result in substantially less-stable folds (*22*). Considering that most synthetic side products feature unnatural chemical modifications we hypothesized that, when folded, these structures would be well-differentiated from the native folded protein. Based on these insights, we folded crude protein mixtures from AFPS and discovered that a suite of rapid biological purification protocols effectively selected for and provided the correct biologically functional protein sequence from synthetic side products.

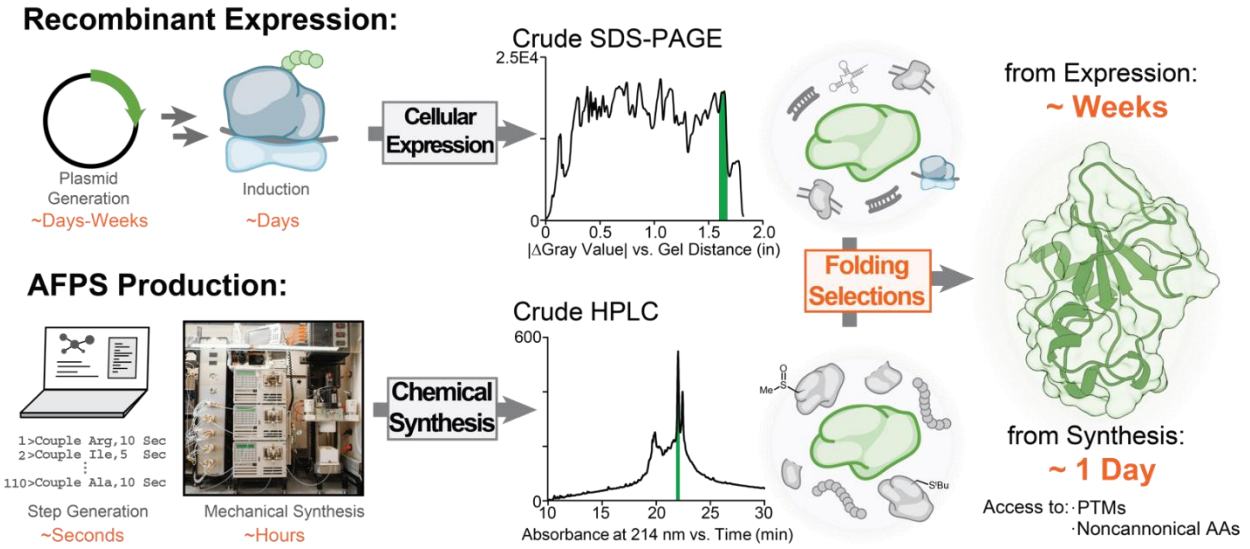


Figure 3.1 Folding selections allow for the same-day production of synthetic proteins

Proteins produced from recombinant expression or chemical synthesis must be separated from a complex mixture of contaminating biomolecules. A schematic outline for the production of a representative protein, barnase, with either recombinant expression or AFPS is outlined with the upper and lower paths, respectively. Analysis of the resulting crude mixtures from each strategy is shown (SDS-PAGE for Recombinant Expression and RP-HPLC for AFPS Production) with the peaks corresponding to the desired protein highlighted in green. Cartoons diagrams representing typical side products are shown for each strategy (e.g., nucleic acids, lipids, and cellular proteins for Recombinant Expression, and chemical adducts, amino acid deletions, and protein fragments for AFPS Production). Side products from either approach can be removed with folding selections to afford purified barnase (PDB: 1A2P). AAs: amino acids; AFPS: automated flow peptide synthesis, PDB: protein data bank, PTM: post-translational modifications; RP-HPLC:

reverse-phase high-performance liquid chromatography, SDS-PAGE: sodium dodecyl sulfate-polyacrylamide gel electrophoresis.

We leverage these insights to develop a rapid protein purification pipeline, termed folding selection (Figure 3.1). First, the protein sequence is prepared with AFPS (*9*). The amino acid sequence of the desired protein is translated into a series of amino acid coupling and Fmoc deprotection steps. Next, the resulting protocol is carried out on a protein synthesis machine, affording a crude protein mixture after cleavage from its solid-phase support. The denatured crude protein mixture is then solubilized and folded with a suitable refolding technique. Following this, a bio-purification tuned to the desired protein is chosen, and with it, the protein is separated from a host of misfolded synthetic side products. With this strategy, proteins can be produced in a single working day; a representative protein of 100 amino acids requires 4.5 hours of synthesis,1.5 hours of cleavage, and 1-3 hours for folding and purification.

We used this pipeline for the on-demand production of nine target proteins. As a proof of concept, we generated folding selection workflows for five natural protein targets ranging from enzymes to co-chaperones (the pointed domain of ETS-related gene: ERG, barnase, thioredoxin-1: Trx-1, dihydrofolate reductase type II: DHFR, and stress induced phosphoprotein 1: STIP1), two proteins with non-natural amino acids (ubiquitin, and the DNA binding domain of B-cell lymphoma/leukemia 11A: BCL11A), and two proteins with post-translational modifications (STIP1 acetylated at K8 and phosphorylated at S16). Only possible with this pipeline, we isolated each folded biologically functional protein in under ten working hours. Rapid biological discovery requires on-demand access to customized proteins, and the folding selection workflows described here are the fastest reported to date.

3.2. Results

3.2.1. Folding Selection Overview

We selected nine protein targets to study with folding selections (Extended Table 1). The following are the primary steps for this process: 1) protein synthesis with AFPS, 2) cleavage of the protein and side products from the synthetic solid-phase support, 3) solubilization of the cleaved mixture in 8 M guanidium hydrochloride (Gdn-HCl), 4) folding

of the crude protein mixture via desalting, and 5) a bio-purification suitable for each protein. In the following sections, we detail the preparation and analysis of each protein.

3.2.2. Synthetic side products misfold

To investigate the behavior of crude synthetic protein mixtures, we prepared the pointed (PNT) domain of ERG as a model substrate. The 94-mer protein was prepared with AFPS, and analysis of the crude material by liquid chromatography-mass spectrometry (LC-MS) is shown in Figure 3.2 A. The observed elution profile indicates the presence of numerous side products with similar chemical structures, as there is only a slight chromatographic separation between different components. The resulting mass spectra of these co-eluting regions were complex, lacking identifiable protein ionization envelopes, thus precluding individual assignment. Despite this complexity, Fmoc-SPPS results in well-known side-product profiles (*9*), so the material is presumedly composed of partial protein fragments, sequences with single amino acid deletions, and chemical adducts with small molecules (*23*). The folding behavior of these synthetic side products is uncharacterized, and we began our investigation by evaluating their behavior with different bio-purification techniques.

After folding, side products from the ERG mixture can be selectively removed on the basis of surface hydrophobicity. We first generated a folded mixture of ERG and its synthetic side products by desalting the denatured crude mixture into a folding buffer. We hypothesized that the random structural rearrangements of synthetic side products would likely result in increased surface hydrophobicity. To test the basic distribution of surface hydrophobicity of the folded mixture, we carried out increasingly stringent ammonium sulfate cuts. Kosmotropic salts, such as ammonium sulfate, strip hydrating water molecules from exposed hydrophobic patches, and, as a result, increasing kosmotrope concentration selectively precipitates proteins (*16*). This behavior, termed salting out, is leveraged in ammonium sulfate cuts, where an added kosmotropic salt, typically $(NH_4)_2SO_4$, sets a hydrophobicity limit to select for a desired protein; more-hydrophobic proteins are removed with low $[(NH_4)_2SO_4]$, and less-hydrophobic proteins remain in solution at high $[(NH_4)_2SO_4]$ (*25*). We hypothesized that synthetic side products with altered folds would precipitate at decreased $[(NH_4)_2SO_4]$. Upon adding $(NH_4)_2SO_4$, we observed the formation of precipitate, with more precipitate at higher $[(NH_4)_2SO_4]$.

Components that remained in solution were analyzed by reverse-phase high-performance liquid chromatography (RP-HPLC) (Figure 3.2 B). Consistent with our expectations, increasing [$(NH_4)_2SO_4$] selectively precipitated synthetic side products suggesting that most side products have a substantially altered surface hydrophobicity. However, concentration higher than 0.8 M $(NH_4)_2SO_4$ led to nonselective precipitation of all remaining components in solution, indicating that some side products form near-native folds.

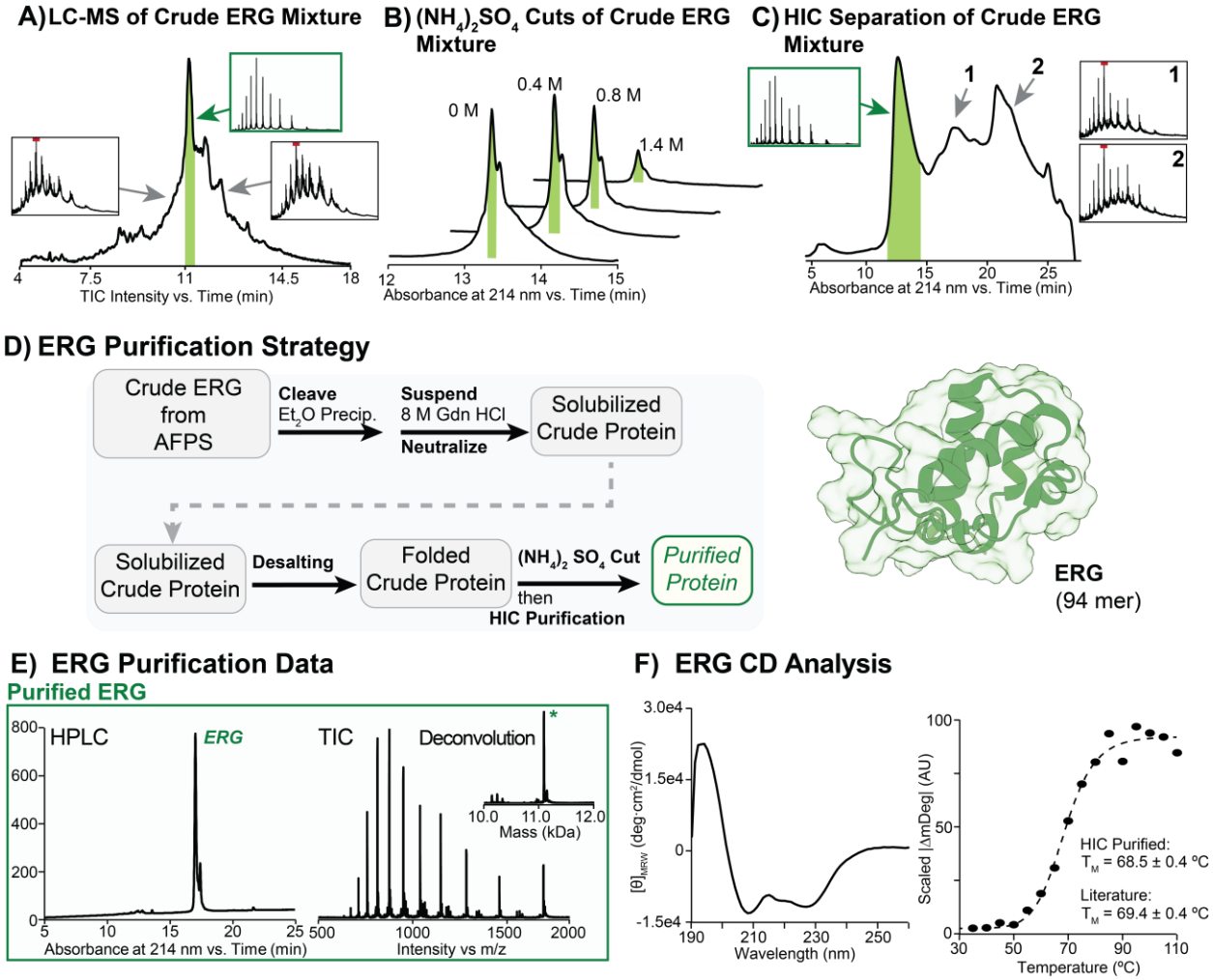


Figure 3.2 Isolation of ERG in a single day with Folding Selections. A

A) Side products elute before and after ERG in high-resolution LC-MS separation. The peak of the desired protein is highlighted in green, and integrated mass spectra are shown for both it (green box), and the impure shoulders (black boxes). Red marks indicate cut m/z values from the internal MS standard. B) Ammonium sulfate fractionation selects against synthetic side products. The synthetic ERG mixture was incubated with different concentrations of ammonium sulfate, resulting in increased precipitation. Analysis of the

supernatant by RP-HPLC is shown with the peak corresponding to the desired protein highlighted in green. C) After folding, native ERG is the most hydrophilic component by HIC. The HIC separation of the ERG mixture is shown, with integrated mass spectra of specific regions highlighted (green box: native ERG, black boxes: side products). Red marks indicate cut m/z values from the internal MS standard, and the peak of the desired protein is highlighted in green. D) A simplified diagram highlighting the protocol used to purify the ERG mixture is shown (PDB: 1SXE). E) Folding selections afford purified ERG. Analytical characterization by RP-HPLC and LC-MS of ERG from the folding selection protocol in D is shown. LC-MS analysis includes the integrated total ion current (TIC) chromatogram and its deconvolution, with the green asterisk indicating the correct observed mass (see Section 3.4.21). F) ERG from folding selection displays the expected circular dichroism (CD) spectra. A CD spectrum was recorded from 195 to 260 nm of ERG from Folding Selection, displaying an alpha-helical structure similar to that of recombinant ERG (24). Points from the average of three technical replicates are plotted (see Table 14). ERG from folding selection also displays a melting temperature (TM) determined by variable temperature CD that closely matches the literature-reported value for recombinant protein24. Points are from the result of a single replicate (see Table 15). CD: circular dichroism; ERG: the pointed domain of ETS-related gene; Gdn-HCl: guanidinium hydrochloride; HIC: hydrophobic interaction chromatography; LC-MS: liquid chromatography-mass spectrometry; PDB: protein data bank; RP-HPLC: reverse-phase high-performance liquid chromatography.

Using analytical hydrophobic interaction chromatography (HIC), we observed that native ERG is the most hydrophilic component of the crude synthetic mixture. Knowing that most synthetic side products are separable by salting out, we investigated whether we could select against the remaining soluble species with HIC. In HIC separations, instead of precipitating under salting out conditions, proteins adhere onto hydrophobically modified resins, and individual components are eluted with a gradual drop in kosmotrope concentration (*26–28*). The most hydrophilic components elute first followed by more hydrophobic species as the buffer composition changes. The separation of crude synthetic biotinylated ERG is shown in Figure 3.2 C. Consistent with the precipitation behavior, the native protein elutes under the first major peak, followed by increasingly hydrophobic synthetic side products. However, species not separated by ammonium sulfate cuts are separated by HIC, as the major ERG peak is free synthetic side products by LC-MS. We hypothesized that the observed selectivity could translate to a preparative scale purification where a crude ammonium sulfate cut removes the majority of side products, followed by semi-preparative HIC to separate the few near-native side products.

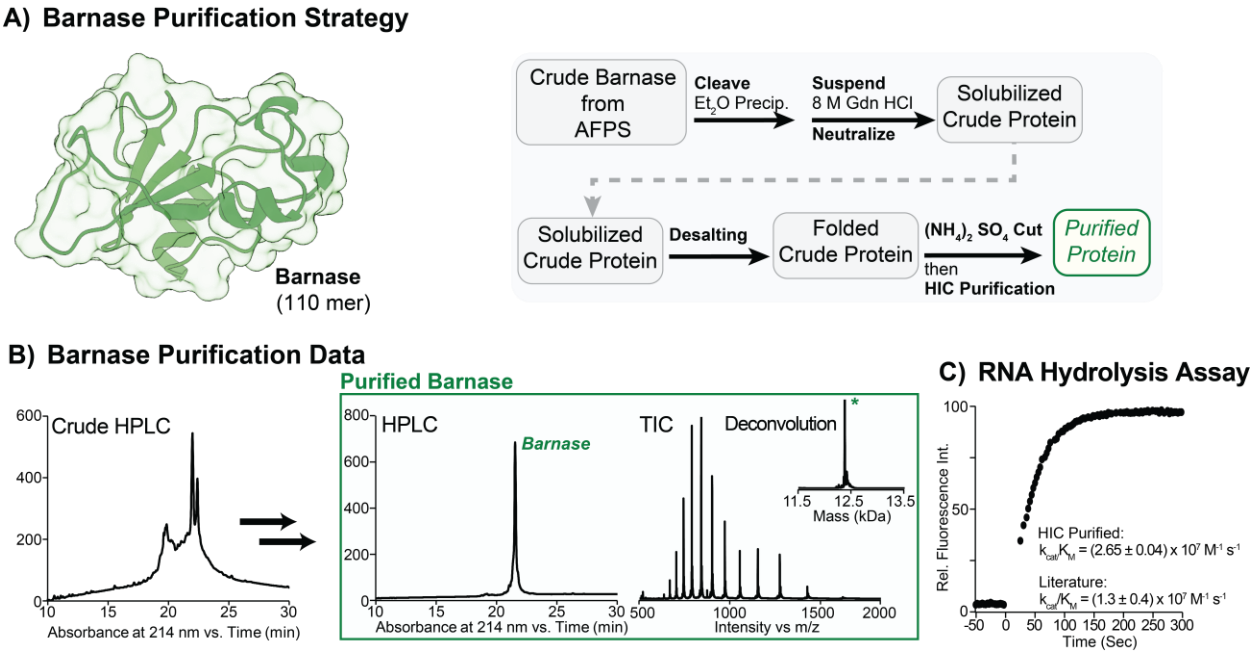


Figure 3.3 Isolation of barnase in a single day with Folding Selections

Enzymatically active barnase can be produced in a single day with Folding Selection. A) A simplified diagram of the folding selection protocol used to isolate barnase is shown (PDB: 1A2P). B) Folding selection affords purified barnase. RP-HPLC separation of the barnase mixture from AFPS is shown in the left column. Analytical characterization by RP-HPLC and LC-MS of barnase from the folding selection protocol in A is shown in the highlighted box. The green asterisk indicates the correct observed mass (see Section 3.4.22). C) Barnase purified using folding selection catalytically cleaves a model RNA substrate with appropriate specific activity comparable to the reported recombinant protein value. An RNA substrate labeled with 5(6)-carboxyfluorescein (FAM) and 6-carboxytetramethylrhodamine (6-TAMRA) separated by a barnase cleavage site was incubated with synthetic barnase, and catalytic activity was measured as an increase in fluorescence intensity (29). Points from the average of three technical replicates are plotted, and error bars are omitted as they are smaller than the symbol size (see Table 16). Extracted kinetic constants of the reaction progress curve closely match that of recombinant material. AFPS: automated-flow peptide synthesis; Dec.: deconvoluted; HPLC: high-performance liquid chromatography; Gdn-HCl: guanidinium-hydrochloride; PDB: protein data bank; TIC: total ion chromatogram.

3.2.3. Folding selections afford purified synthetic proteins

We developed a preparative folding selection protocol that selects for surface hydrophobicity and purified two synthetic proteins with it: ERG and barnase. ERG was AFPS-synthesized, cleaved, and folded, and a portion of the total material was purified with an ammonium sulfate cut and semi-preparative HIC (see Section 3.4.21). Consistent

with the analytical separations, ERG was eluted first by HIC and was free from synthetic side products. Purification of the material from a single synthesis would afford 0.5 mg of purified ERG, although we purified a fraction of the total material and isolated 79 μg (Section 3.4.21). An overview of the purification strategy is shown in Figure 3.2 D, and the total operational time was under ten hours. Analytical RP-HPLC and LC-MS spectra of the purified protein are shown in Figure 3.2 E. The circular dichroism spectrum of the synthetic ERG matches that of ERG from recombinant expression and previously reported AFPS synthesis (*24*) (Figure 3.2 F), suggesting it is the native structure.

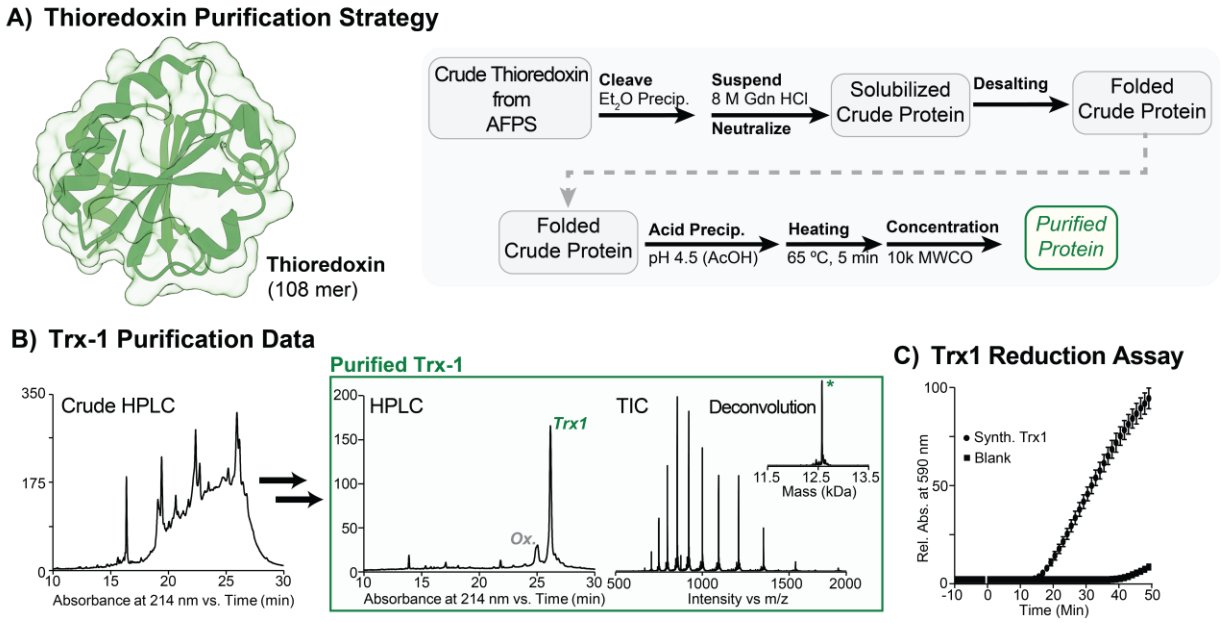


Figure 3.4 Folding selection of thioredoxin without chromatography.

A) A simplified diagram is shown highlighting the folding selection protocol used to isolate Trx-1 (PDB: TRX2) with acid and heat precipitations. B) Analytical characterization of Trx-1 before and after purification with the protocol from A. The early-eluting peak by RP-HPLC corresponds to the oxidized (+16 Da) product, and the green asterisk indicates the correct observed mass (see Section 3.4.23). C) Trx-1 from folding selection catalyzes the disulfide exchange of insulin. Bovine insulin was incubated with Trx-1 and either with or without 1,4-dithiothreitol (DTT), as previously reported32. Absorbance at 590 nm was monitored to track the precipitation of disulfide-exchanged insulin. Points from the average of two technical replicates are plotted, and error bars represent standard deviation (see Table 17).

Using the same folding selection protocol, we successfully isolated a structurally unrelated protein, barnase, and verified its catalytic activity. Barnase has been the subject of AFPS (*27*) and multi-fragment synthesis campaigns (*28*, *52*) and is a benchmark for

evaluating new protein synthesis protocols. Barnase was synthesized with AFPS and then purified with the same strategy with adjusted kosmotrope concentration (Figure 3A, see Section 3.4.7). Analytical HPLC and LC-MS spectra of the purified protein are shown in Figure 3B, and total operational time was again under 10 hours, affording 0.24 mg of folded barnase (extrapolating to 1 mg from processing the full synthesis, Section 3.4.22). Furthermore, we observed catalytic activity from the folded synthetic barnase that was indistinguishable from that of recombinant material (*29*) (Figure 3.3 C), again indicating that folding selection with HIC can afford properly folded synthetic protein.

When differences in solubility between the protein and side products are sufficiently large, folding selections can afford purified protein without chromatography. First isolated in 1964 (*33*), thioredoxin-1 (Trx-1) can be purified from other cellular components based on its stability at low pH and temperatures up to 80 ºC (*34*). We adapted these classical biochemical purification protocols and generated a folding selection strategy that can be benchmarked against known multi-fragment ligation-based assembly of synthetic Trx-1 (*35*, *36*) (Figure 3.4 A). The AFPS synthesis of thioredoxin-1 was less efficient than those of ERG or barnase, but acid and heat precipitations afforded 66 µg of the purified material shown in Figure 3.4 B (extrapolating to 0.3 mg from processing the full synthesis, Section 3.4.23). Furthermore, the purified Trx-1 demonstrated proper oxidoreductase activity in the catalytic reduction of bovine insulin with DTT (*32*) (Figure 3.4 C), again indicating that the material from Folding selections can generate biochemically active synthetic proteins.

3.2.4. Diversification of folding selection protocols affords a variety of functional synthetic proteins

With the addition of two new biochemical purification techniques, we isolated three synthetic proteins: BCL11A, Ubiquitin, and DHFR (Type II). The DNA binding domain of BCL11A comprises three tandem C2H2 zinc fingers (*37*), and we anticipated that it would be a challenging target for folding selection due to its structural complexity. We synthesized the protein with AFPS with its three methionine residues substituted for the non-natural norleucine (Nle), as previously reported (*24*), due to their sensitivity to oxidation. Despite including non-natural amino acids, a HIC protocol outlined in Figure 5A afforded 20 µg of purified BCL11A (Figure 3.5 A, extrapolating to 0.2 mg from processing

the full synthesis, Section 3.4.24). Based on this success, we investigated whether different chromatographic methodologies would be suitable for Folding Selections. We anticipated that mixed-mode (MM) purifications would offer favorable selectivity for folding selections and nominated ubiquitin as a test substrate. Ubiquitin is a small protein responsible for regulating protein degradation (*38*) and is a classical target for protein synthesis (*39*). We prepared ubiquitin with AFPS and isolated 9.0 mg of purified protein from a single post-synthesis MM purification (Figure 3.5 B, extrapolating to 34 mg from processing the full synthesis, Section 3.4.25). In addition, we selected dihydrofolate reductase (Type II) (DHFR) as it assembles a tetramer upon folding (*40*). We hypothesized that synthetic side products would be unable to assemble, and a selection based on apparent molecular weight could remove these misfolded impurities. After AFPS synthesis, crude DHFR was purified by size exclusion chromatography (SEC), and the peak corresponding to the approximate MW of the tetramer was collected, affording 0.9 mg of purified DHFR (Figure 3.5 C, extrapolating to 1.9 mg from processing the full synthesis, Figure 5C, Section 3.4.26). Together, these results indicate that folding selection is compatible with a range of biochemical purification protocols and can be tuned based on the properties of the individual protein target.

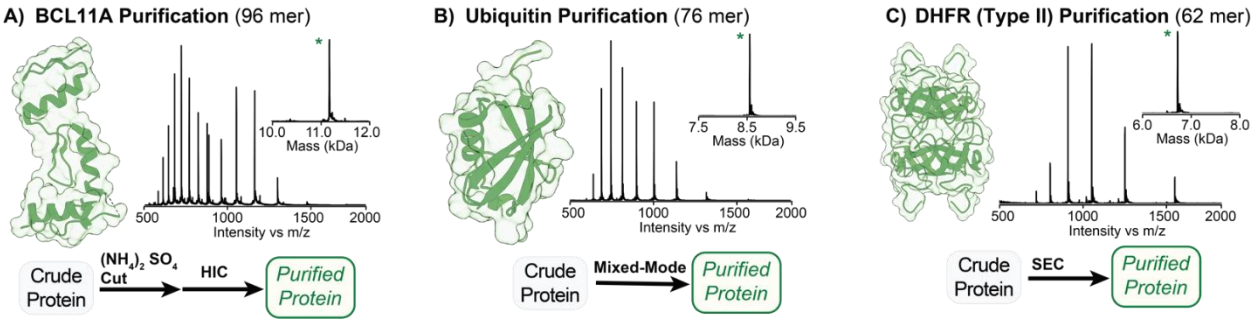


Figure 3.5 Preparation of three synthetic proteins with diversified folding selection protocols.

Three synthetic proteins were prepared with AFPS and purified with Folding Selections. LC-MS traces after purification are included with a diagram of key selection steps for A) BCL11a, B) Ubiquitin, and C) DHFR (PDBs: 6KI6, 1 UBQ, and 5UFE, respectively). The green asterisks indicate the correct observed mass (see Sections 3.4.24, 3.4.25, and 3.4.26).

3.2.5. Post-translationally modified proteins from folding selections

Folding selections of synthetic proteins can access post-translationally modified (PTM) protein targets that would be difficult to prepare with recombinant expression. To

demonstrate this capability of our folding selection platform, we synthesized and isolated two post-translational modification variants of the tetranucleotide repeat 1 (TPR1) domain of stress induced phosphoprotein 1 (STIP1). STIP1 is an important mediator of the chaperone activity of heat shock proteins HSP70-HSP90 (*41*) and is known to be modified in vivo with post-translational modifications (*42*). However, the biological consequences of these modifications are under-characterized partly because they are difficult to access with recombinant expression techniques (*43*). In contrast, proteins from chemical synthesis can precisely incorporate any suitable number of PTMs by selecting appropriate commercially available Fmoc-protected amino acid monomers. We nominated two STIP1 PTM sites for preparation and isolation with Folding Selections: acetylation of lysine 8 (*44*) and phosphorylation of serine 16 (*45*). We first prepared the native TPR1 domain of STIP1 with HIC and successfully isolated 0.24 mg of purified wild-type (WT) STIP1 (Figure 3.5 A, extrapolating to 0.47 mg from processing the full synthesis, Section 3.4.27). Both PTM STIP1 targets were prepared by AFPS with substitution with either Fmoc-Lys(Ac)-OH or Fmoc-Ser(PO(OBzl)OH)-OH. Purification of both sequences with the same HIC folding selection strategy afforded 0.53 mg and 0.26 mg of purified STIP1-K8-Ac, and STIP1-S16-Phos, respectively (Figure 3.5 B-C, extrapolating to 1.0 mg and 0.5 mg from processing the full syntheses, respectively, Section 3.4.27). We anticipate that the rapid generation of PTM proteins could enable rapid production and characterization of unknown or understudied modifications.

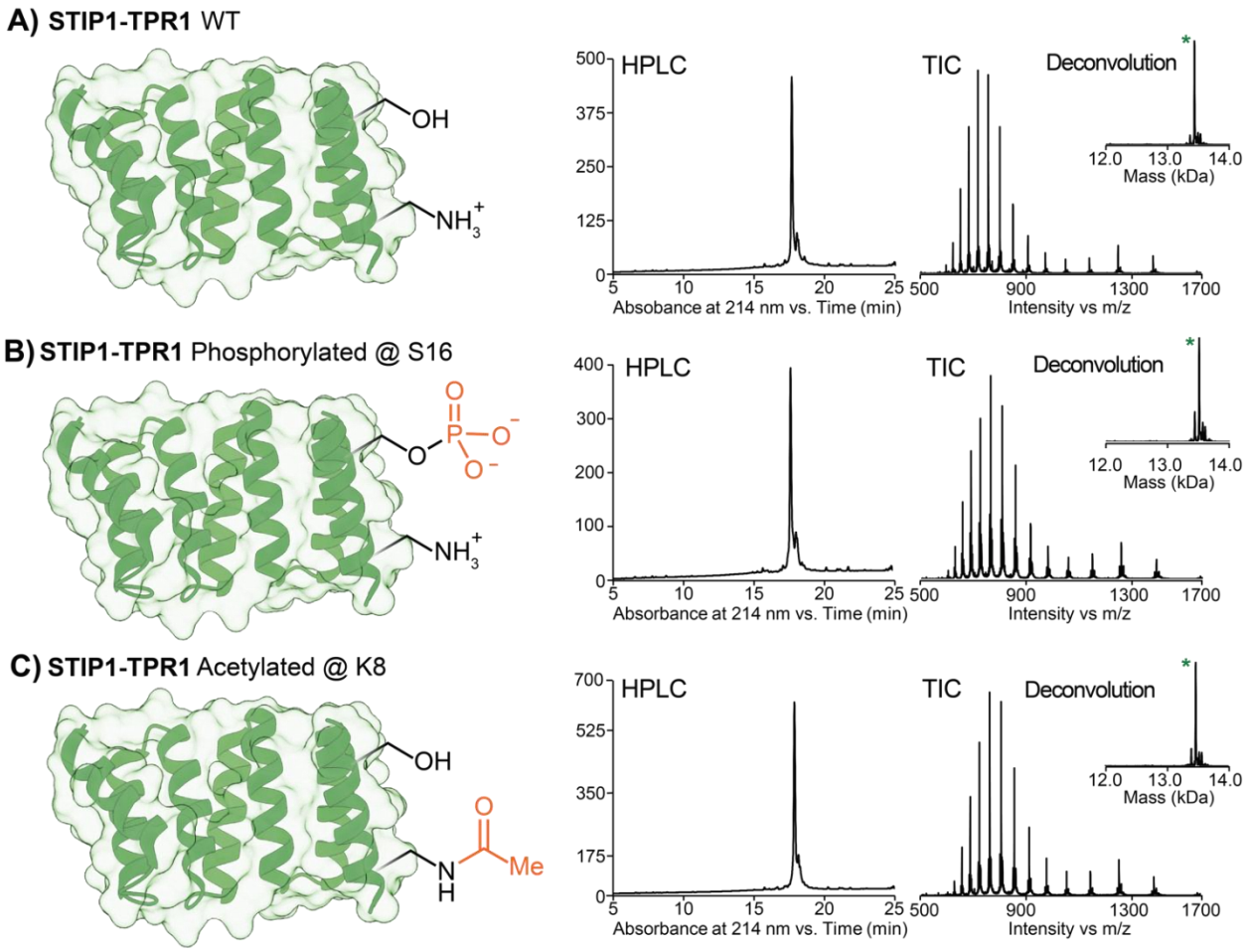


Figure 3.6 Preparation of post-translationally modified STIP1 with Folding Selections.

Three variants of STIP1 were prepared with AFPS and purified with Folding Selections. Post-translational modification sites are highlighted on the protein diagram, with purified HPLC and LC-MS traces to the right for A) unmodified STIP1, B) STIP1 phosphorylated at serine 16, and C) STIP1 acetylated at lysine 8 (PDB: 1ELW). The green asterisks indicate the correct observed mass (see Section 3.4.27).

3.3. Discussion

We present a new framework to understand protein synthesis side products based on their resultant fold. Chemical protein synthesis generates a complex mixture of side products that must be separated from the desired protein. By traditional frameworks, purifications are increasingly unable to purify the desired peptide as sequences approach the length of entire protein domains. This model relies on the postulate that small chemical modifications lead to only small chromatographic differences. We show that this assumption does not hold when entire protein domains are prepared and folded. Protein

folding is sensitive to even minor changes in chemical makeup, and the modified chemical structures of synthetic side cause them to misfold. The biophysical properties of these misfolded impurities are unconnected to the biophysical properties of the native fold and are easily removed with rapid bio-purification techniques. This folding model is a fundamentally new way of conceptualizing synthetic protein purifications and enables a rapid purification pipeline, termed *folding selection*.

Our results show that new post-folding purification strategies informed by this framework can access chemically synthesized proteins in a single day. We selected nine protein targets and isolated each in under ten hours with a panel of folding selection strategies. New synthetic, molecular, and chemical biology research is driven by reliable access to chemically defined proteins, and our folding selection approach can uniquely accelerate these efforts. Folding selection should be generalizable to a wide range of single-domain proteins, including many targets inaccessible with other production techniques. Proteins from folding selection can incorporate an unrestricted number of non-natural chemical modifications, enabling the incorporation of novel catalytic apparatus in designed enzymes or crucial stabilizing mutations in designed protein-protein interfaces. Furthermore, the ability to rapidly prepare a panel of homogeneously post-translationally modified proteins, shown here by the preparation of STIP1 variants (Fig. 3.6), should help further refine our understanding of these modifications' structural consequences. More generally, further refinement of chemical protein synthesis and folding selection strategies could ultimately arrive at a routine production pipeline that provides sequence-defined proteins for rapid study in all branches of biological science

3.4. Methods

3.4.1. Proteins from AFPS

All proteins were synthesized on one of three automated fast-flow systems, which were built in the Pentelute lab: the “Amidator”, the “Automatide” system, or the “Peptidator”. In all cases, synthesis conditions were identical to the optimized protein synthesis conditions as previously described (*9*).

Following synthesis, peptidyl resins were washed with 5 volumes of dichloromethane (5 x 5 mL) and dried under an N_2 stream. To the synthesis resin, was added a ratio of 1 mL of cleavage mixture (trifluoroacetic acid: 94% v/v, water: 2.5% v/v,

1,2-ethanedithiol: 2.5% v/v, and triisopropylsilane: 1% v/v, freshly prepared) per 2.5 µmol of protected protein, and the container placed on a nutating mixer. After 2 hours of incubation, the supernatant was expelled into a fresh 50 mL Flacon tube and the crude peptide precipitated with the addition of 40 mL of diethyl ether prechilled to -80 ºC. The peptide was then pelleted by centrifugation at 4 °C for 4 min at 3220 rcf. The supernatant was decanted into waste, and the solid pellet containing the protein retained for processing.

Protein	Sequence	MW (kDa)	pI*	Uniprot ID
ERG	MEEKHMPPPN MTTNERRVIV PADPTLWSTD HVRQWLEWAV KEYGLPDVNI LLFQNIDGKE LCKMTKDDFQ RLTPSYNADI LLSHLHYLRE TPLP	11.09	5.44	P11308
Barnase	AQVINTFDGV ADYLQTYHKL PDNYITKSEA QALGWVASKG NLADVAPGKS IGGDIFSNRE GKLPGKSGRT WREADINYTS GFRNSDRILY SSDWLIYKTT DHYQTFTKIR	12.38	8.88	P00648
Thioredoxin-1	SDKIIHLTDD SFDTDVLKAD GAILVDFWAE WCGPCKMIAP ILDEIADEYQ GKLTVAKLNI DQNPGTAPKY GIRGIPTLLL FKNGEVAATK VGALSKGQLK EFLDANLA	11.67	4.67	P0AA25
DHFR	VFPSNATFGM GDRVRKKSGA AWQGQIVGWY CTNLTPEGYA VESEAHPGSV QIYPVAALER IN	6.73	6.73	P00383
Ubiquitin	MQIFVKTLTG KTITLEVEPS DTIENVKAKI QDKEGIPPDQ QRLIFAGKQL EDGRTLSDYN IQKESTLHLV LRLRGG	8.54	6.56	P0CG48
BCL11A	RSDTCEYCGK VFKNCSNLTV HRRSHTGERP YKCELCNYAC AQSSKLTRHM KTHGQVGKDV YKCEICKMPF SVYSTLEKHM KKWHSDRVLN NDIKTE	11.19	9.18	Q9H165
STIP1**	MEQVNELKEK GNKALSVGNI DDALQCYSEA IKLDPHNHVL YSNRSAAYAK KGDYQKAYED GCKTVDLKPD WGKGYSRKAA ALEFLNRFEE AKRTYEEGLK HEANNPQLKE GLQNMEAR	13.46	6.31	P31948

Table 13 Sequences and characteristics of the proteins synthesized using AFPS for Folding Selection.

*AFPS: automated-flow peptide synthesis. MW: molecular weight. ERG: transcriptional regulator ERG; DHFR: dihydrofolate reductase; BCL11A: B-cell lymphoma/leukemia 11A; STIP1: stress-inducible protein 1.*Theoretical. **Plus, two STIP1 variants, acetylation at K8 and phosphorylation at S16.*

3.4.2. Folding by Desalting

The cleaved protein pellet was dissolved with 5 mL of 8 M guanidine hydrochloride (Gdn HCl) and incubated at 37 ºC for 15 min. To the solution was then added 1 mL of 1.0 M tris(hydroxymethyl)aminomethane (TRIS) followed by additions 6 M NaOH to a final pH of 7.0. The denatured protein solution was then diluted to a theoretical concentration of 10 mg / ml with additional 8 M Gdn HCl for a final volume under 15 mL. Separately, a Cytiva 26/10 Hi-Prep desalting column was affixed to an ÄKTA Pure FPLC and prepared in folding buffer (50 mM TRIS, 150 mM NaCl, 5 mM DTT, 5% glycerol v/v, pH 7.0). The solubilized crude protein mixture was added by manual injection to the desalting column, and fractions containing folded protein without eluting Gdn HCl were collected and verified by LC-MS.

3.4.3. Ammonium sulfate cut of ERG from AFPS

Crude ERG from AFPS was folded as described and adjusted to 1 mg / mL as determined by the Bradford protein assay (See SI Section 2.5.2). To a series of 2.0 mL lo-bind Eppendorf tubes was added 0.2 mL each of folded ERG. To each vial was added increasing volumes of a 3.0 M $(NH_4)_2SO_4$ stock prepared in the same buffer (50 mM TRIS, 150 mM NaCl, 5 mM DTT, 5 % glycerol, pH 7.0). The samples were capped, gently mixed by inversion, and incubated at room temperature for 30 mins. The samples were clarified by centrifugation at 21k rcf for 5 min, and the supernatant analyzed by RP-HPLC and LC-MS (see SI sections 2.3.1 and 2.4.2).

3.4.4.Analytical HIC of ERG from SS-AFPS

A Proteomix HIC butyl column (5 µm, 4.6 x 35 mm) was attached to an Agilent 1290 series system, and equilibrated at 100% A (50 mM sodium phosphate, 3.0 M M $(NH_4)_2SO_4$, pH 7.0), 0% B (50 mM sodium phosphate, pH 7.0), and 0 % C (isopropanol) at 1 mL / min. Crude ERG from AFPS with an N-terminal PEG_{12}Biotin cap was dissolved after cleavage as described earlier, and approximately 0.1 mg injected onto the HIC column. A gradient from (100% A, 0 % B, 0% C) to (50 % A, 40 % B, 10% C) was carried

out over 40 mins with UV detection at 214 nm. Fractions corresponding to separate peak areas were manually collected, and subsequently analyzed by LC-MS (See SI section 2.4.2).

3.4.5.HIC Based Folding Selection of ERG

ERG was synthesized, cleaved and folded as described earlier. The folded ERG solution was made to 0.7 M (NH4)2SO4 with the addition of a sufficient amount of a 3.2 M (NH4)2SO4 stock prepared in the same buffer (50 mM TRIS, 150 mM NaCl, 5 mM DTT, 5 % glycerol, pH 7.0). The sample was incubated at room temperature for 30 mins and clarified by centrifugation at 3220 rcf for 15 min. Separately, a Cytiva HiTrap Butyl HP HIC column was affixed to an ÄKTA Pure FPLC and prepared in 100% A (0.7 M (NH4)2SO4, 50 mM sodium phosphate, pH 7.0), 0% B (50 mM sodium phosphate, pH 7.0) at 0.7 mL / min. The supernatant after centrifugation was added by manual injection to the HIC column, and the sample eluted with a linear gradient from 0 to 100% B in 100 minutes. Fractions were analyzed by LC-MS, collected, and concentrated with a 10 K MWCO spin-concentrator to 1 mL. The concentrated ERG solution was desalted into storage buffer (50 mM TRIS, 150 mM NaCl, 5 mM DTT, pH 7.0) with a Cytiva 5 mL HiTrap Desalting column affixed to an ÄKTA Pure FPLC and concentrated to 50 µL with a 10 K MWCO spin-concentrator (See SI Section 5.1).

3.4.6.ERG CD Analysis

ERG exchanged into CD buffer (10 mM KPi, 5% glycerol v/v, pH 7.4) with a 5 mL HiTrap Desalting column affixed to an ÄKTA Pure FPLC. Fractions corresponding to protein free of loaded salt as monitored by buffer conductivity were collected and concentrated with a 10 K MWCO spin-concentrator to 0.05 mg / mL as determined by A280 (see SI section 2.5.1). CD spectra were recorded on an AVIV Model 420 Circular Dichroism Spectrometer, with a 1 mm quartz cuvette (Hellma Analytics, Article No.: 100-1-40). Before all readings, cuvettes were cleaned with Hellmanex™ 3 concentrate according to manufacturer directions.

Far circular dichroism spectra were recorded as the average of three scans between 190 nm to 260 nm with 1 nm steps, 2 nm scanning windows, and a 3 s averaging time for each scan at 25 °C (See SI Section 6.1.1).

For variable temperature CD, CD signal was monitored at 225 nm with 2 nm bandwidth and 8 second averaging time. Spectra were recorded between 10 and 110 °C with 5 °C steps and 2-minute incubation time at each step. Data were fit to a sigmoidal curve to extract the observed melting temperature (See SI Section 6.1.2).

Wavelength (nm)	Rep 1 (mDeg)	Rep 2 (mDeg)	Rep 3 (mDeg)	Baseline Subtracted MRW (deg cm^2/dmol)	SD
260	15.796	15.624	15.003	0.0	667.44
259	15.889	15.226	14.976	-88.5	754.90
258	16.132	15.086	15.196	-60.6	919.67
257	16.009	15.392	15.309	4.0	611.91
256	16.031	15.113	15.363	37.6	759.35
255	15.845	15.377	15.121	88.8	587.42
254	16.117	15.138	15.543	45.6	787.08
253	15.664	15.33	15.392	13.9	284.26
252	15.982	14.911	15.376	68.3	859.27
251	15.958	15.591	15.202	-60.5	604.89
250	15.651	14.955	15.226	-106.3	561.32
249	15.659	15.297	15.087	-105.6	462.95
248	15.96	15.04	15.274	-275.2	765.04
247	15.673	14.969	14.837	-332.0	719.09
246	15.601	14.94	14.865	-529.9	648.03
245	15.508	14.341	14.71	-812.9	954.40
244	15.262	14.645	14.244	-1095.7	820.49
243	15.244	13.818	14.296	-1426.4	1161.27
242	14.884	13.952	14.09	-1813.6	804.81
241	14.569	13.571	13.515	-2231.1	948.84
240	14.38	13.021	13.619	-2648.7	1089.80
239	13.886	12.953	13.387	-3233.2	747.00
238	13.625	12.171	12.746	-3868.5	1171.64
237	12.991	11.911	11.988	-4684.3	964.07
236	12.515	11.2	11.187	-5652.3	1220.79
235	11.683	10.244	11.039	-6752.9	1153.31
234	10.906	9.619	9.762	-7906.8	1128.64
233	9.883	8.997	9.356	-9022.1	713.03
232	9.448	8.413	8.676	-10014.9	860.73
231	9.233	8.161	8.126	-10910.5	1006.83
230	8.431	7.535	7.604	-11609.9	797.73
229	8.464	7.545	6.982	-12094.4	1196.95
228	8.303	7.363	7.237	-12363.5	932.00
227	8.21	7.849	7.443	-12476.0	613.95
226	8.395	7.356	6.975	-12444.9	1175.95
225	8.49	7.455	7.279	-12276.1	1046.89
224	8.951	8.009	7.209	-12062.8	1395.14
223	8.955	8.288	7.859	-11783.3	883.66
222	8.331	8.34	8.151	-11449.9	170.59
221	9.034	8.455	8.236	-11342.3	659.70
220	9.153	8.034	7.789	-11231.6	1163.48
219	8.982	8.915	8.035	-11198.1	845.56
218	8.818	8.139	8.116	-11123.3	638.12
217	9.144	8.28	8.862	-10889.9	704.95
216	9.666	9.085	7.976	-10384.8	1373.82
215	9.093	13.24	7.387	-10011.5	4816.22
214	9.088	10.897	7.888	-10145.1	2423.58
213	8.901	9.242	7.141	-10590.7	1804.06
212	8.111	8.128	7.146	-11622.1	899.38
211	7.74	7.509	6.735	-12568.3	842.21
210	7.141	7.506	6.13	-13268.5	1140.52
209	7.282	7.124	5.626	-13708.7	1462.24
208	7.559	6.714	5.811	-13801.2	1398.66

207	7.563	7.193	6.534	-13322.4	833.95
206	8.613	8.825	6.93	-12393.5	1661.29
205	9.313	9.672	8.014	-10613.2	1395.65
204	11.649	11.611	10.176	-8513.9	1343.49
203	13.047	12.624	11.364	-5873.6	1400.80
202	15.231	14.79	14.149	-2704.0	870.52
201	17.299	17.292	16.18	712.3	1030.47
200	20.155	20.098	18.805	4864.8	1221.60
199	23.285	22.402	22.492	9131.5	777.45
198	25.421	25.637	25.112	12819.3	422.19
197	26.505	25.528	26.397	16051.2	857.00
196	28.133	27.433	27.731	18629.6	562.06
195	30.327	28.892	28.298	20658.3	1669.03
194	31.01	30.399	29.976	21811.5	831.75
193	28.034	27.557	31.488	21744.0	3432.26
192	33.417	26.437	22.937	21523.2	8536.69
191	26.898	34.522	24.441	17916.8	8410.50
190	5.811	37.281	21.245	2306.3	25177.54

Table 14 Raw Far CD Scan of ERG from Crude Refolding

Temperature (ºC)	CD Signal (mDeg)	Scaled CD Signal
10.002	-1.061	9.6
15.006	-1.158	5.6
20.012	-1.296	0.0
24.988	-1.147	6.1
29.988	-1.156	5.7
35.007	-1.162	5.5
39.988	-1.157	5.7
44.981	-1.102	7.9
49.982	-1.122	7.1
54.993	-0.957	13.9
60.011	-0.767	21.7
64.994	-0.472	33.7
70.018	0.066	55.8
75.009	0.489	73.1
79.994	0.743	83.5
84.98	1.069	96.8
90	0.75	83.7
94.999	1.147	100.0
99.986	1.076	97.1
104.997	1.029	95.2
109.984	0.849	87.8

Table 15 Raw CD Melt data of ERG from Crude Refolding

3.4.7.HIC Based Folding Selection of Barnase

Barnase was synthesized, cleaved and folded as described earlier. The folded Barnase solution was made to 1.8 M (NH4)2SO4 with the addition of a sufficient amount of solid (NH4)2SO4. The sample was incubated at room temperature for 30 mins and clarified by centrifugation at 3220 rcf for 15 min. Separately, a Cytiva 1 mL Phe Source column was was affixed to an ÄKTA Pure FPLC and prepared in 100% A (1.8 M (NH4)2SO4, 50 mM sodium phosphate, pH 7.0), 0% B (50 mM sodium phosphate, pH 7.0) at 2 mL / min. The supernatant after centrifugation was added by manual injection to the HIC column, and the sample eluted with a linear gradient from 0 to 100% B in 15 minutes. Fractions were analyzed by LC-MS, collected, and concentrated with a 10 K MWCO spin-concentrator to 1 mL. The concentrated Barnase solution was desalted into storage buffer (50 mM TRIS, 150 mM NaCl, 5 mM DTT, pH 7.0) with a Cytiva 5 mL HiTrap Desalting column affixed to an ÄKTA Pure FPLC and concentrated to 50 µL with a 10 K MWCO spin-concentrator (See SI Section 5.2).

3.4.8.Barnase Activity Assay

The relative catalytic activity of synthetic Barnase was measured with a previously reported assay3,29. Barnase cleaves a fluorogenic oligonucleotide substrate (6-FAM-dA-rG-dA-dA-6-TAMRA) at its rG-dA junction, quenching the FRET between the FAM and TAMRA fluorophores. The resultant increase in fluorescence can be fit to a first order rate equation to extract observed kcat/KM values. Fluorescence measurements were recorded on a Tecan Spark® plate reader with excitation at 495 nm and monitoring emission at 515 nm. Assays were carried out in flat bottomed, black 96-well plates. The fluorogenic substrate of the was diluted in assay buffer, and the baseline absorbance recorded for at least 150 sec. Barnase was added to initiate hydrolysis and the fluorescence monitored every 15 sec until sufficient data was collected for analysis. The experiment was performed in triplicate. Blank runs were performed in parallel with the experimental runs (See SI section 6.2).

t (s)	Mean (RFU)	SD	t (s)	Mean (RFU)	SD	t (s)	Mean (RFU)	SD
-53.321	0.006	0.005	85.604	0.847	0.013	210.323	0.983	0.003
-50.549	0.012	0.004	89.165	0.851	0.002	213.285	0.982	0.003
-47.720	0.008	0.001	92.152	0.876	0.012	216.090	0.984	0.008
-44.887	0.013	0.005	97.332	0.887	0.005	219.174	0.983	0.005
-42.113	0.010	0.005	100.150	0.889	0.014	222.163	0.989	0.007
-39.312	0.011	0.004	102.954	0.902	0.015	225.137	0.986	0.005
-36.439	0.014	0.007	105.853	0.905	0.009	228.117	0.982	0.006
-33.647	0.011	0.004	108.721	0.914	0.014	231.036	0.990	0.009
-30.739	0.010	0.007	111.630	0.915	0.005	233.891	0.985	0.006
-27.894	0.014	0.004	114.468	0.920	0.005	236.850	0.989	0.005
-25.100	0.012	0.008	117.499	0.927	0.008	239.739	0.990	0.004
-22.289	0.019	0.003	120.380	0.937	0.006	242.499	0.993	0.007
-19.609	0.014	0.004	123.425	0.940	0.002	245.301	0.987	0.003
-14.883	0.011	0.001	127.913	0.943	0.004	250.446	0.993	0.002
-12.116	0.010	0.008	130.747	0.950	0.005	256.063	0.992	0.002
-9.246	0.012	0.004	133.640	0.955	0.004	259.010	0.991	0.006
-5.715	0.013	0.001	136.708	0.954	0.007	261.915	0.984	0.006
-2.833	0.011	0.003	141.937	0.961	0.005	265.773	0.996	0.002
0.000	0.008	0.005	147.219	0.959	0.007	270.117	0.984	0.006
27.620	0.333	0.020	150.382	0.966	0.005	276.006	0.992	0.004
33.169	0.411	0.028	153.159	0.966	0.007	279.057	0.979	0.005
38.162	0.451	0.023	158.089	0.972	0.004	281.885	0.990	0.003
41.051	0.496	0.023	161.059	0.966	0.005	284.735	0.984	0.004
43.929	0.534	0.022	163.976	0.974	0.005	287.640	0.986	0.001
46.757	0.567	0.023	166.953	0.971	0.009	290.519	0.986	0.013
49.589	0.602	0.022	169.865	0.976	0.004	293.485	0.984	0.006
52.439	0.627	0.021	172.928	0.971	0.018	296.312	0.987	0.007
55.301	0.656	0.019	175.804	0.984	0.005	299.337	0.987	0.009
58.183	0.681	0.015	178.827	0.980	0.005	302.233	0.981	0.005
61.238	0.706	0.015	181.705	0.979	0.001	n.d.	n.d.	n.d.
64.084	0.746	0.015	189.207	0.983	0.003	n.d.	n.d.	n.d.
69.322	0.758	0.006	196.663	0.985	0.010	n.d.	n.d.	n.d.

72.215	0.779	0.013	201.700	0.983	0.003	n.d.	n.d.	n.d.
75.178	0.802	0.012	204.580	0.984	0.007	n.d.	n.d.	n.d.
78.084	0.831	0.011	207.432	0.987	0.001	n.d.	n.d.	n.d.

Table 16 Averaged Fluorescence Data from RNase activity assay of Barnase

3.4.9.Solubility Based Folding Selection of Trx-1

Trx-1 was synthesized, cleaved and folded as described earlier, with omission of DTT in the desalting buffer. The folded Trx-1 solution was acidified to pH 4.5 with glacial acetic acid, incubated for 10 mins at room temperature, and the solution was clarified by centrifugation at 21,000 rcf for 5 min. The acidified protein solution was transferred to a series of 0.2 mL PCR strip tubes, and placed in a PCR thermal cycler that was heated to 65 °C for 5 min. The heated tubes were cooled to rt, pooled into a single 1.5 mL lo-bind Eppendorf tube, and clarified by centrifugation at 21,000 rcf for 5 min. The supernatant was neutralized to pH 7.0 with ammonium hydroxide and concentrated with a 10 K MWCO spin-concentrator to 0.5 mL. The solution was then diluted to 2 mL with folding buffer and re-concentrated to 0.5 mL. Serial dilution and re-concentration was performed a total of three times with the final concentration step continuing until the sample reached a volume of 20 µL (See SI Section 5.3).

3.4.10.Trx-1 Insulin Reduction Assay

The relative catalytic activity of thioredoxin was measured with a previously reported assay32,35. In the presence of DTT, Trx-1 catalysis the exchange of disulfide bonds on bovine insulin, resulting in precipitation that can be monitored by absorbance at 650 nm.

Absorbance measurements were recorded on a Tecan Spark® plate reader at 650 nm. Assays were carried out in glass bottomed, flat, black 96-well plates. Assay buffer consisted of 1x PBS with 2.2 mM EDTA, and total well volumes were 200 µL. Each well was made 0.13 mM in bovine insulin, with test wells also containing 2 µM Trx-1. The assay was initiated with the addition of DTT to a final concentration of 0.33 mM in each well. Abosrbance at 650 nm was monitored once every 15 sec for 50 mins. The experiment was performed in duplicate. Blank runs containing DTT, but no Trx-1 were performed in parallel with the experimental runs. Data points were down-sampled by a factor of three by averaging three sample points.

Time (min)	+ Trx-1 Rep 1	+ Trx-1 Rep 2	- Trx-1 Rep 1	- Trx-1 Rep 2	Time (min)	+ Trx-1 Rep 1	+ Trx-1 Rep 2	- Trx-1 Rep 1	- Trx-1 Rep 2
-10.36	0.037	0.037	0.037	0.037	22.02	0.163	0.138	0.037	0.037
-9.11	0.037	0.037	0.037	0.037	23.27	0.190	0.164	0.037	0.037
-7.86	0.037	0.037	0.037	0.037	24.52	0.221	0.192	0.037	0.037
-6.61	0.037	0.037	0.037	0.037	25.77	0.252	0.220	0.037	0.037
-5.36	0.037	0.037	0.037	0.037	27.02	0.284	0.249	0.037	0.037
-4.10	0.037	0.037	0.037	0.037	28.27	0.315	0.279	0.037	0.037
-2.85	0.037	0.037	0.037	0.037	29.52	0.344	0.307	0.037	0.037
-1.38	0.037	0.037	0.037	0.037	30.77	0.373	0.336	0.037	0.037
0.75	0.037	0.037	0.037	0.037	32.02	0.407	0.364	0.037	0.037
2.00	0.037	0.037	0.037	0.037	33.27	0.435	0.393	0.037	0.037
3.25	0.037	0.037	0.037	0.037	34.52	0.464	0.419	0.037	0.037
4.50	0.037	0.037	0.037	0.037	35.77	0.491	0.446	0.037	0.037
5.75	0.037	0.037	0.037	0.037	37.02	0.518	0.473	0.037	0.037
7.01	0.037	0.037	0.037	0.037	38.28	0.544	0.497	0.037	0.038
8.26	0.037	0.037	0.037	0.037	39.53	0.570	0.520	0.038	0.040
9.51	0.037	0.037	0.037	0.037	40.78	0.593	0.545	0.039	0.043
10.76	0.037	0.037	0.037	0.037	42.03	0.616	0.566	0.042	0.047
12.01	0.037	0.037	0.037	0.037	43.28	0.638	0.587	0.045	0.053
13.26	0.037	0.037	0.037	0.037	44.53	0.660	0.609	0.050	0.059
14.51	0.041	0.038	0.037	0.037	45.78	0.680	0.625	0.056	0.066
15.76	0.051	0.044	0.037	0.037	47.03	0.700	0.645	0.062	0.074
17.01	0.067	0.056	0.037	0.037	48.28	0.718	0.662	0.069	0.082
18.26	0.087	0.072	0.037	0.037	49.53	0.736	0.681	0.077	0.091
19.51	0.111	0.091	0.037	0.037	50.78	0.754	0.696	0.085	0.100
20.77	0.137	0.114	0.037	0.037	n.d.	n.d.	n.d.	n.d.	n.d.

Table 17 Raw Absorbance Data of Trx-1 Catalyzed Reduction of Bovine Insulin

3.4.11.HIC Based Folding Selection of BCL11a

BCL11a was synthesized, cleaved and the resulting protein pellet dissolved in 10 mL of 1:1 H2O:acetonitrile + 0.1% trifluoroacetic acid (v/v) and lyophilized overnight. Solid lyophilized BCL11a was dissolved in a denaturing buffer (8 M Urea, 15 mM ZnCl2,0.1 mM TRIS, 5 mM TCEP, 0.1 M NaCl, pH 7.6) and folded on a Cytiva HiPrep 26/10 delsating column, with the folding buffer modified to include ZnCl2 (0.1 mM Tris, 0.1 M NaCl, 0.5 mM TCEP, 0.5 mM ZnCl¬2 pH 7.6). The folded BCL11a solution was made to 2.3 M (NH4)2SO4 with the addition of a sufficient amount of solid (NH4)2SO4. The sample was then incubated at room temperature for 30 mins and clarified by centrifugation at 3220 rcf for 15 min. Separately, a Cytiva 1 mL Iso Source column was affixed to an ÄKTA Pure FPLC and prepared in 100% A (2.3 M (NH4)2SO4, 50 mM TRIS, pH 7.0), 0% B (50 mM TRIS, pH 7.0) at 2 mL / min. The supernatant after centrifugation was added by manual injection to the HIC column, and the sample eluted with a linear gradient from 0 to 100% B in 15 minutes. Fractions were analyzed by LC-MS, collected, and concentrated with a 10 K MWCO spin-concentrator to 1 mL. The concentrated

BCL11a solution was desalted into storage buffer (50 mM TRIS, 150 mM NaCl, pH 7.0) with a 5 mL Cytiva HiTrap Desalting column affixed to an ÄKTA Pure FPLC and concentrated to 20 µL with a 10 K MWCO spin-concentrator (See SI Section 5.4).

3.4.12.MM Based Folding Selection of Ubiquitin

Ubiquitin was synthesized and cleaved as described earlier. After solubilization with 6 M Gdn HCl, the solution was neutralized to pH 7.0 with sequential additions of 1 M TRIS pH 8.0. To the crude protein mixture was added a sufficient volume of 16 mM beta-cyclodextrin in 1x PBS to achieve a final concentration of 8 mM beta-cyclodextrin. The solution was incubated for 30 minutes and then centrifuged at 16,000 rpm to remove insoluble material. The solution was desalted with a Cytiva 5 mL HiTrap Desalting column affixed to an ÄKTA Pure FPLC prepared in folding buffer (50 mM TRIS, 75 mM NaCl, 5 mM MgCl2, 10% glycerol v/v, pH 8.0 and fractions containing folded protein without eluting Gdn HCl were collected and verified by LC-MS. Separately, a Cytiva Capto MMC Imp Res column was prepared in 1x PBS pH 7.4 at 1 mL / min. The desalted crude protein was added by manual injection to the MMC column and eluted isocratically. Fractions were analyzed by LC-MS, collected, and concentrated with a 3 K MWCO spin-concentrator to 1.5 mL (See SI Section 5.5).

3.4.13.SEC Based Folding Selection of DHFR

DHFR was synthesized, cleaved and folded as described earlier. The folded DHFR solution was concentrated to 0.5 mL with a 10 K MWCO spin-concentrator. Separately, a Cytiva SuperDex 75 Increase 10/300 GL column was affixed to an ÄKTA Pure FPLC and equilibrated in SEC buffer (50 mM TRIS, 150 mM NaCl, 5 mM DTT, pH 7.0) at 0.8 mL/min. The concentrated crude DHFR solution was added to a 1 mL loading loop and injected onto the prepared SEC column. Fractions from the peak corresponding to the approximate MW of the tetrameric structure were collected and concentrated with a 10 K MWCO spin-concentrator to 1 mL.

3.4.14.HIC Based Folding Selection of STIP1 and its Variants

STIP1 and its variants was synthesized, cleaved and folded as described earlier. For acetylated STIP1, lysine 8 was incorporated as Fmoc-Lys(Ac)-OH, and for phosphorylated STIP1, serine 16 was manually incorporated as Fmoc-Ser(PO(OBzl)OH)-OH. The folded STIP1 solutions were made to 2.5 M (NH4)2SO4 with the addition of solid

(NH4)2SO4, incubated at room temperature for 30 mins and clarified by centrifugation at 3220 rcf for 15 min. Separately, a 1 mL Iso Source column was affixed to an ÄKTA Pure FPLC and prepared in 100% A (2.0 M (NH4)2SO4, 50 mM sodium phosphate, pH 7.0), 0% B (50 mM sodium phosphate, pH 7.0) at 2 mL / min. The supernatants after centrifugation was added by manual injection to the HIC column, and eluted with a linear gradient from 0 to 100% B in 15 minutes. Fractions were analyzed by LC-MS, collected, and concentrated with a 10 K MWCO spin-concentrator to 1 mL. The concentrated STIP1 solutions were desalted into storage buffer (50 mM TRIS, 150 mM NaCl, pH 7.0) with a 5 mL HiTrap Desalting column affixed to an ÄKTA Pure FPLC (See SI Section 5.7).

3.4.15. General information

Unless otherwise specified, all reactions with polypeptides, proteins, and protein oxidative addition complexes were set up on the bench top and carried out under ambient conditions. Unless stated otherwise, all small (≤ 1 mL) volumes were measured out using Eppendorf Research® plus, single-channel, variable, mechanical pipettes (referred to as mechanical pipettes). Universal low retention pipet tips (10 µL, 200 µL and 1000 µL sizes) were purchased from VWR International (Philadelphia, PA). Care was taken to use the appropriate mechanical pipette/tip combinations to ensure a dispensing error of ≤ 2% for volumes between 10 µL and 1000 µL and ≤ 4% for volumes between 1 µL and 10 µL. To avoid loss of protein due to non-specific adsorption, plastic tubes (Eppendorf Protein LoBind® tubes, 0.5, 1.5, and 2.0 mL) were used in all cases after folding. The weight of lyophilized powders of the peptides was directly measured using analytical scales (XS205DU Analytical Balance, Mettler-Toledo) with fixed with a SPI Westek Workstation Still Air Ionizer.

3.4.16. Materials for peptide and protein synthesis

All reagents were purchased and used as received unless otherwise noted. Fmoc-protected L-amino acids (Fmoc-Ala-OH x H_2O, Fmoc-Cys(Trt)-OH, Fmoc-Asp(OtBu)-OH, Fmoc-Glu(OtBu)-OH, Fmoc-Phe-OH, Fmoc-Gly-OH, Fmoc-Ile-OH, Fmoc-Lys(Boc)-OH, Fmoc-Leu-OH, Fmoc-Met-OH, Fmoc-Asn(Trt)-OH, Fmoc-Pro-OH, Fmoc-Gln(Trt)-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Ser(tBu)-OH, Fmoc-Thr(tBu), Fmoc-Val-OH, Fmoc-Trp(Boc)-OH, and Fmoc-Tyr(OtBu)-OH) were purchased from the Novabiochem line (Millipore-Sigma). L-Fmoc-His(Boc)-OH was purchased from Chem-Pep. L-Fmoc-Nle-OH, was

purchased from Chem-Impex. *O*-(7-azabenzotriazol-1-yl)-*N,N,N',N'*-tetramethyluronium hexafluorophosphate (HATU, ≥97.0%), (7-azabenzotriazol-1-yloxy) tripyrrolidinophosphonium hexafluorophosphate (PyAOP, ≥97.0%) were purchased from P3 Biosystems. Biosynthesis OmniSolv® grade *N*,*N*-dimethylformamide (DMF) was purchased from EMD Millipore, and stored over AldraAmine trapping agents (for 1000 – 4000 mL DMF, catalog number Z511706). Diisopropylethylamine (DIEA; 99.5%, biotech grade, catalog number 387649) was purchased from Sigma-Aldrich and purified by passage through an activated alumina column (Pure Process Technology solvent purification system). Piperidine (ACS reagent, ≥99.0%), trifluoroacetic acid (TFA, HPLC grade, ≥99.0%), triisopropylsilane (TIPS, ≥98.0%), acetonitrile (HPLC grade), formic acid (FA, ≥95.0%), phenol (ACS reagent, ≥99.0%), diethyl ether (Et_2O, ≥99.7%, containing 1 ppm BHT as inhibitor), and 1,2-ethanedithiol (EDT, GC grade, ≥98.0%) were purchased from Sigma-Aldrich. H-Rink Amide (0.49 mmol/g and 0.18 mmol/g loading) resin was purchased from PCAS Biomatrix. Water was deionized using a Milli-Q Reference water purification system (Millipore). 5 mL and 10 mL peptide synthesis reaction vessels were purchased from Torviq (catalog numbers SF-0500, and SF-1000 respectively). Biotin-PEG_{12}-COOH (98%) was purchased from BroadPharm. Syringe tip caps were purchased from VWR (catalog number 97001-202).

Guanidine HCl (Gdn HCl, molecular biology grade, ≥99%), tris(hydroxymethyl)aminomethane (TRIS, ≥99.9%), tris(hydroxymethyl)aminomethane hydrochloride (TRIS HCl, ≥99.0%), monopotassium phosphate (KPi, ≥99.0%), dipotassium phosphate (≥ 99.0%), sodium chloride (NaCl, BioXtra, ≥99.5%), glycerol (≥99.5%), sodium hydroxide (≥98%), hydrochloric acid (36.5-38%), urea (BioReagent grade), zinc chloride (≥99.995% trace metals basis), magnesium chloride hexahydrate (BioXtra, ≥99%), ethylenediaminetetraacetic acid (EDTA, BioUltra, ≥99%), and □-cyclodextrin (BioReagent grade) were purchased from Sigma Aldrich. 1,4-Dithio-DL-threitol (DTT, ≥99%) and ammonium sulfate (ACS Reagent grade, ≥99%) was purchased from Chem-Impex. A 0.5 M solution of tris(2-carboxyethyl)phosphine hydrochloride (TCEP, Bond-Breaker™, catalogue number 77720), was purchased from Thermo Fisher Scientific.

3.4.17. Analytical Reverse Phase High Pressure Liquid Chromatography Methods

Analysis of synthetic peptides was primarily carried out by evaluation of the chromatogram produced from the absorbance at 214 nm after separation on a reverse phase column. In each case, an appropriate amount of material was injected to produce peaks on the chromatogram that were within the dynamic range of the UV detector - typically between 100 and 2000 mAU (3-5 μμg of the major product) for the major peak.

Method 1 - 5% to 65% in 15 min on Agilent 1290 HPLC

Analytical HPLC was carried out on an Agilent 1290 series system with UV detection at 214 nm. Column: Phenomenex Kinetex C18, (150 x 2.1 mm, 1.8 μm, 100 Å silica); flow rate 0.375 mL/minute; Solvent System: A = water with 0.1% TFA (v/v), B = acetonitrile with 0.08% TFA (v/v); Gradient: 3 min hold 5% B, 5-65% B gradient over 15 min, 3 min hold 65% B, 10 min post run 5% B.

Method 2 - 5% to 65% in 30 min on Agilent 1290 UHPLC

Analytical HPLC was carried out on an Agilent 1290 series system with UV detection at 214 nm. Column: Phenomenex Kinetex C18, (150 x 2.1 mm, 1.8 μm, 100 Å silica); flow rate 0.375 mL/minute; Solvent System: A = water with 0.1% TFA (v/v), B = acetonitrile with 0.08% TFA (v/v); Gradient: 3 min hold 5% B, 5-65% B gradient over 30 min, 3 min hold 65% B, 10 min post run 5% B.

3.4.18. Reverse Phase High Pressure Liquid Chromatography Mass Spectrometry Methods

Data were processed using Agilent MassHunter Workstation Qualitative Analysis Version B.10.00 Software or Agilent MassHunter BioConfirm Software B.10.00. Deconvoluted masses of proteins were obtained using a maximum entropy algorithm. Unless otherwise depicted, the following parameters were used for deconvolution: a 3,000 Da window centered on the calculated mass was used (not lower than 5,000 Da), and a limited *m/z* range from 600-3,000 was used with a baseline subtraction factor of 3 and a mass step of 1.

Method - 1 Agilent 6545 1-91 in 15 min Zorbax C3

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), standard mass range (m/z in range 300 to 3000), with scan rate of 1.00

spectra/sec. The following instrument parameters were used: gas temperature 350 ºC, gas flow 11 L/min, nebulizer pressure 60 psig, sheath gas temperature 350 ºC, sheath gas flow 11 L/min. The following scan source parameters were used: VCap: 3500, nozzle voltage 1000 V, fragmentor 175, skimmer1 65, OctopoleRFPeak 750. 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 condition was used for analysis. Column: Zorbax 300-SB C3 (5 µm, 150 x 2.1 mm, 300 Å silica); Flow Rate: 0.8 mL/min; Gradient: 1% B 0-2 min, linearly ramp from 1% B to 91% B 2 to 11 min, 91% B to 95% B 11 to 12 min. Post time is 1% B for 3 min. LC stream directed to the MS in 4-11 minutes.

Method - 2 Agilent 6550 1-91 in 15 min Zorbax C3

Analysis was performed on an Agilent 1290 Infinity HPLC coupled to an Agilent 6550 ESI-Q-TOF mass spectrometer. MS was run in positive ionization mode, extended dynamic range (2 GHz), standard mass range (m/z in range 300 to 3000), with scan rate of 1.00 spectra/sec. The following instrument parameters were used: gas temperature 200 ºC, gas flow 14 L/min, nebulizer pressure 55 psig, sheath gas temperature 350 ºC, sheath gas flow 11 L/min. The following scan source parameters were used: VCap: 3500, nozzle voltage 1000 V, fragmentor 175, skimmer1 65, OctopoleRFPeak 750. 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 condition was used for analysis. Column: Zorbax 300-SB C3 (5 µm, 150 x 2.1 mm, 300 Å silica); Flow Rate: 0.5 mL/min; Gradient: 1% B 0-2 min, linearly ramp from 1% B to 91% B 2 to 12 min, 91% B to 95% B 12 to 13 min. Post time is 1% B for 2 min. LC stream directed to the MS in 4-12 minutes.

Method - 3 Agilent 6550 1-91 in 10 min Zorbax C3

Analysis was performed on an Agilent 1290 Infinity HPLC coupled to an Agilent 6550 ESI-Q-TOF mass spectrometer. MS was run in positive ionization mode, extended dynamic range (2 GHz), standard mass range (m/z in range 300 to 3000), with scan rate of 1.00 spectra/sec. The following instrument parameters were used: gas temperature 200 ºC, gas flow 14 L/min, nebulizer pressure 55 psig, sheath gas temperature 350 ºC, sheath gas flow 11 L/min. The following scan source parameters were used: VCap: 3500, nozzle

voltage 1000 V, fragmentor 175, skimmer1 65, OctopoleRFPeak 750. 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 condition was used for analysis. Column: Zorbax 300-SB C3 (5 µm, 150 x 2.1 mm, 300 Å silica); Flow Rate: 0.5 mL/min; Gradient: 1% B 0-1 min, linearly ramp from 1% B to 91% B 1 to 7 min, 91% B to 95% B 7 to 8 min. Post time is 1% B for 2 min. LC stream directed to the MS in 1-7 minutes.

Method - 4 Agilent 6545 1-91 in 10 min Zorbax C3

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), standard mass range (m/z in range 300 to 3000), with scan rate of 1.00 spectra/sec. The following instrument parameters were used: gas temperature 350 ºC, gas flow 11 L/min, nebulizer pressure 60 psig, sheath gas temperature 350 ºC, sheath gas flow 11 L/min. The following scan source parameters were used: VCap: 3500, nozzle voltage 1000 V, fragmentor 175, skimmer1 65, OctopoleRFPeak 750. 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 condition was used for analysis. Column: Zorbax 300-SB C3 (5 µm, 150 x 2.1 mm, 300 Å silica); Flow Rate: 0.8 mL/min; Gradient: 1% B 0-2 min, linearly ramp from 1% B to 91% B 2 to 7 min, 91% B to 95% B 7 to 8 min. Post time is 1% B for 2 min. LC stream directed to the MS in 2-7 minutes.

Method - 5 Agilent 6550 1-91 in 18 min Aeris C4

Analysis was performed on an Agilent 1290 Infinity HPLC coupled to an Agilent 6550 ESI-Q-TOF mass spectrometer. MS was run in positive ionization mode, extended dynamic range (2 GHz), low mass range (m/z in range 100 to 1700), with scan rate of 1.50 spectra/sec. The following instrument parameters were used: gas temperature 200 ºC, gas flow 14 L/min, nebulizer pressure 55 psig, sheath gas temperature 350 ºC, sheath gas flow 11 L/min. The following scan source parameters were used: VCap: 3500, nozzle voltage 1000 V, fragmentor 175, skimmer1 0, OctopoleRFPeak 750. 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 condition was used for

analysis. Column: Phenomenex Aeris C4 Widepore (5 µm, 150 x 2.1 mm, 300 Å silica); Flow Rate: 0.2 mL/min; Gradient: 1% B 0-2 min, linearly ramp from 1% B to 91% B 2 to 15 min, 91% B to 95% B 15 to 15.01 min, 95% B 15.01 to 18 min. Post time is 1% B for 2 min. LC stream directed to the MS in 3-15 minutes.

3.4.19. Determination of Protein Concentration

All measurements were recorded on a BioTec Epoch instrument with a Take3 Micro-Volume Plate. For each sample or buffer blank to be measured, 2 µL was added to each read position. The sample buffer baseline was recorded as an average of at least 8 spots recorded simultaneously. Solution containing the protein of interest was added to no fewer than 4 spots as 2 µL aliquots, and the absorbance at 280 nm recorded. The molar extinction coefficient of each protein sequence was estimated based on the sequence of the protein via ExPASy Swiss Institute of Bioinformatics - Bioinformatics Resource Portal. Protein concentration was calculated using Beer's law

3.4.20. Crude LC-MS of ERG from AFPS

Crude ERG from AFPS (SI section 5.1) was dissolved in 6M Gdn HCl, 50 mM TRIS, 50 mM DTT, pH 8.0 and approximately 50 ng was separated on an optimized LC-MS method. Analysis was performed on an Agilent 1290 Infinity HPLC coupled to an Agilent 6550 ESI-Q-TOF mass spectrometer. MS was run in positive ionization mode, extended dynamic range (2 GHz), standard mass range (m/z in range 300 to 3000), with scan rate of 1.00 spectra/sec. The following instrument parameters were used: gas temperature 290 ºC, gas flow 14 L/min, nebulizer pressure 55 psig, sheath gas temperature 350 ºC, sheath gas flow 11 L/min. The following scan source parameters were used: VCap: 5500, nozzle voltage 2000 V, fragmentor 380, skimmer1 65, OctopoleRFPeak 750. 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 condition was used for analysis. Column: Phenomenex Kinetex C18, (150 x 2.1 mm, 1.8 µm, 100 Å silica); Flow Rate: 0.375 mL/min; Gradient: 25% B 0-2 min, linearly ramp from 25% B to 45% B 2 to 19 min, 45% B to 65% B 19 to 19.1 min, 65% B 19.1 to 20 min. Post time is 1% B for 2 min. LC stream directed to the MS in 3-18 minutes.

TIC intensity corresponding to individual regions (main peak: 9.6 to 10.9, early shoulder: 11 to 11.3, and late shoulder: 11.4 to 14.0) were integrated.

3.4.21. Folding Selection of ERG

Protein	Uniprot ID	Start	End	Length	Modifications
ERG	P11308	108	201	94	N-Term PEG_{12}Biotin, C-Term carboxamide
MEEKHMPPPN MTTNERRVIV PADPTLWSTD HVRQWLEWAV KEYGLPDVNI LLFQNIDGKE LCKMTKDDFQ RLTPSYNADI LLSHLHYLRE TPLP					

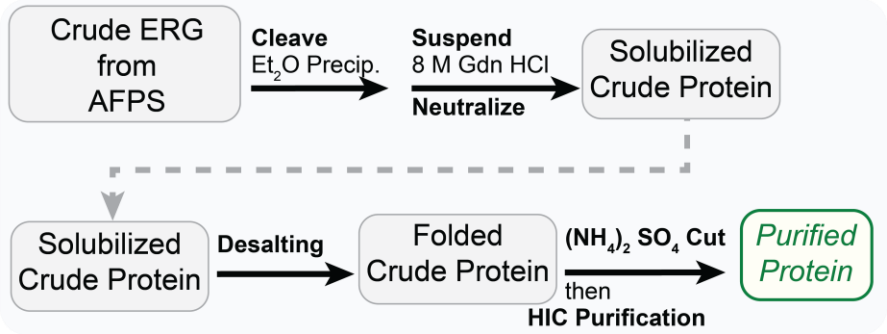
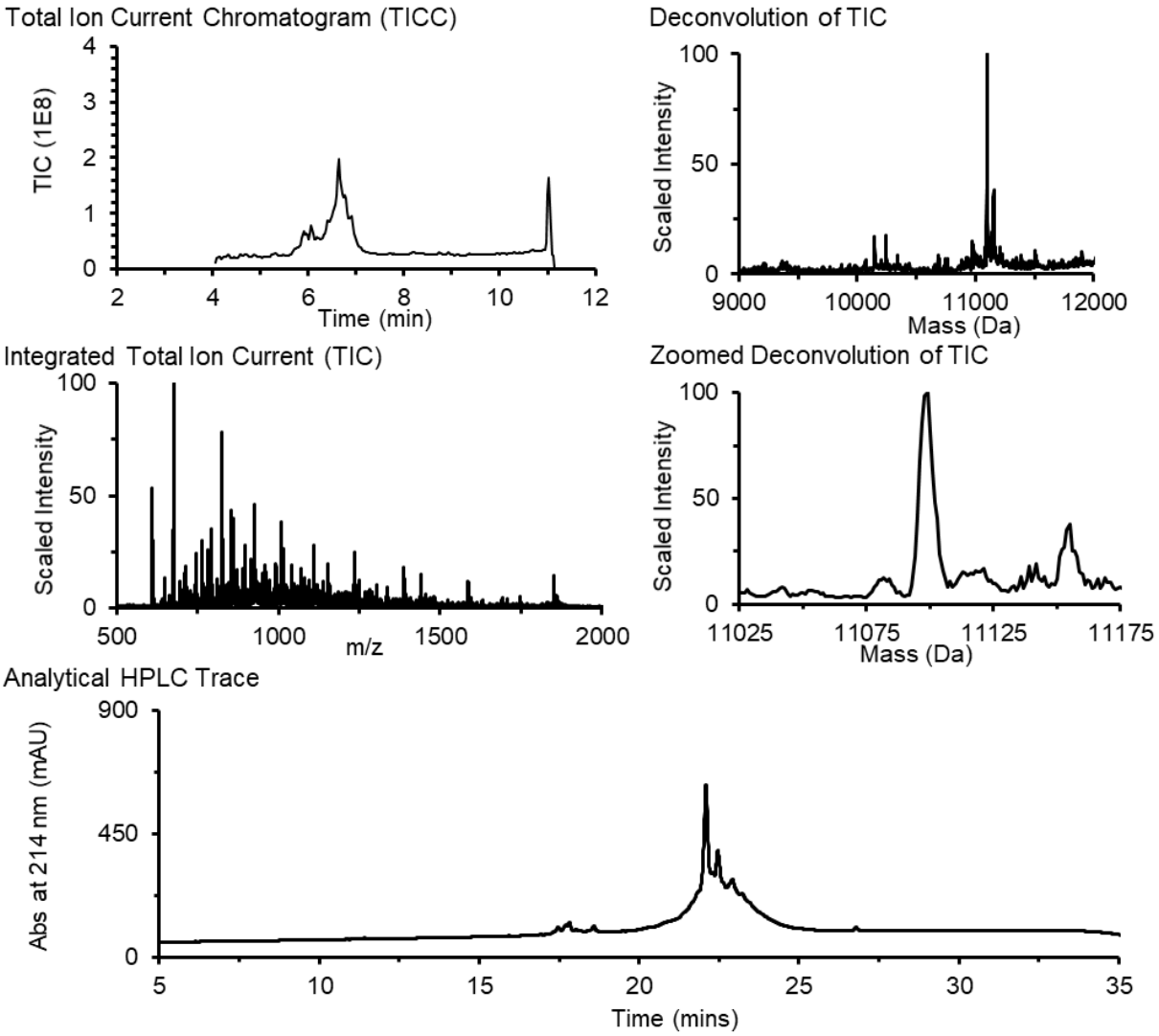


Figure 3.7 Scheme for the folding selection workflow for ERG.

Crude Analytical Characterization



Synthesis Scale: 23.4 µmol (130 mg, 0.18 mmol/g resin)

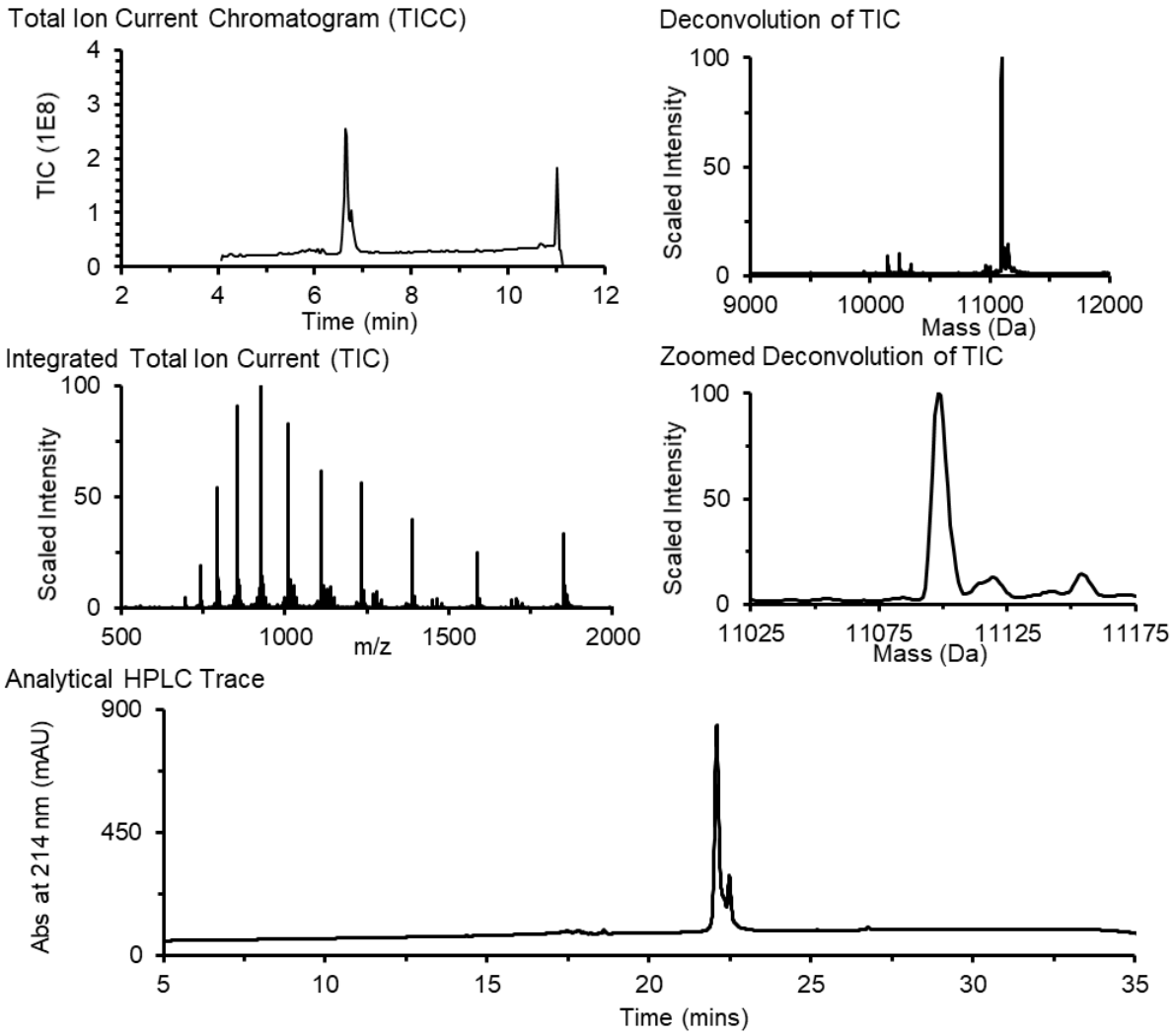
Crude RP-HPLC Method: Method 2

Crude LC-MS Method: Method 1

Calculated Mass: 11098 Da

Observed Mass: 11098 Da

Purified Analytical Characterization



Purified RP-HPLC Method: Method 2

Purified LC-MS Method: Method 1

Calculated Mass: 11098 Da

Observed Mass: 11098 Da

Percentage Cleaved: 50% (of the synthesis resin)

Percentage Folded: 29% (of the cleaved material)

Percentage Purified: 100% (of the folded material)

Isolated Amount: 7.1 nmol (50 µL, 142 µM)

Isolated Yield: 0.2% (relative to resin loading)

Extrapolated Amount: 49.8 nmol (0.6 mg)

3.4.22. Folding Selection of Barnase

Protein	Uniprot ID	Start	End	Length	Modifications
Barnase	P00648	48	157	110	C-Term carboxamide
AQVINTFDGV ADYLQTYHKL PDNYITKSEA QALGWVASKG NLADVAPGKS IGGDIFSNRE GKLPGKSGRT WREADINYTS GFRNSDRILY SSDWLIYKTT DHYQTFTKIR					

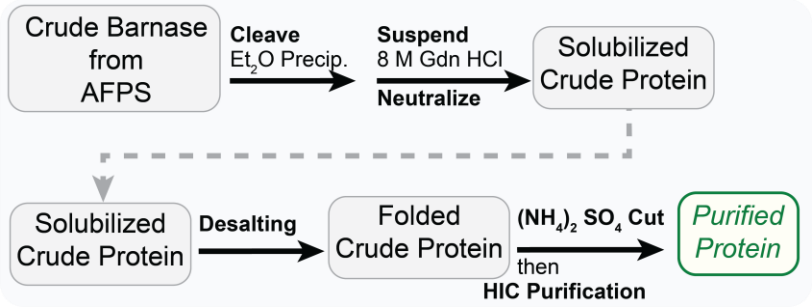
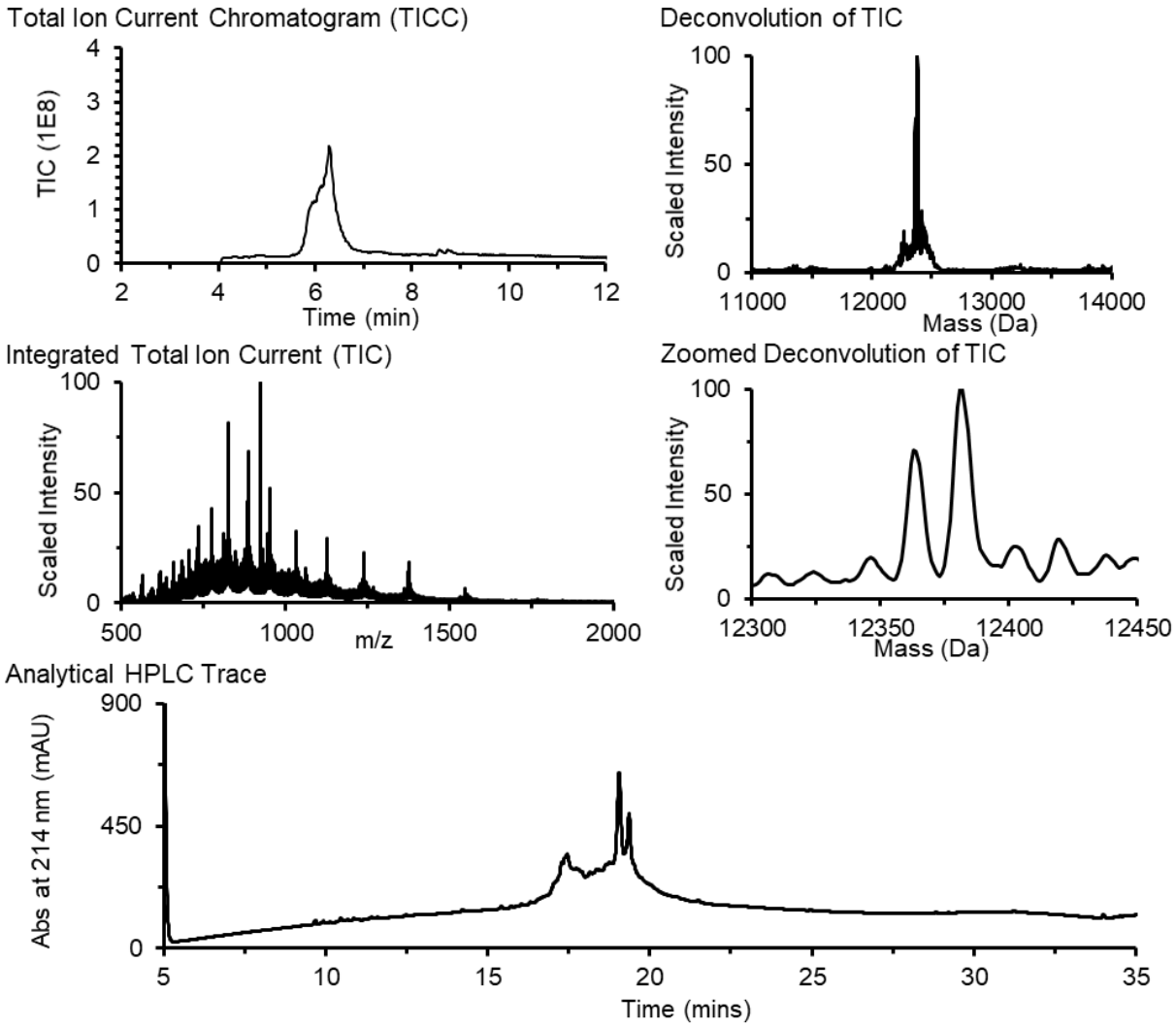


Figure 3.8 Scheme for the folding selection workflow for barnase.

Crude Analytical Characterization



Synthesis Scale: 23.4 µmol (130 mg, 0.18 mmol/g resin)

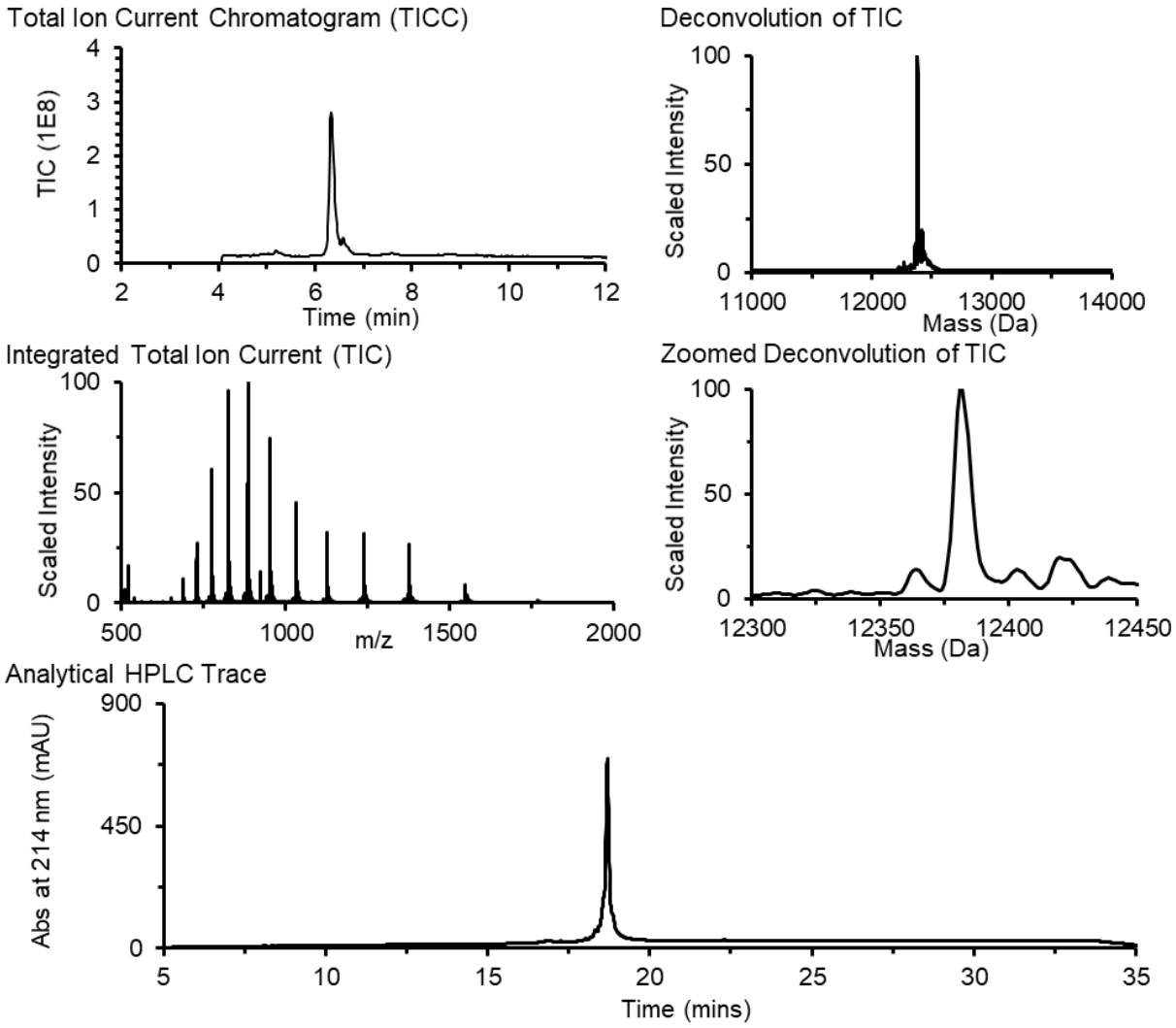
Purified RP-HPLC Method: Method 2

Purified LC-MS Method: Method 2

Calculated Mass: 12382 Da

Observed Mass: 12382 Da

Purified Analytical Characterization



Purified RP-HPLC Method: Method 2

Purified LC-MS Method: Method 2

Calculated Mass: 12382 Da

Observed Mass: 12382 Da

Percentage Cleaved: 50% (of the synthesis resin)

Percentage Folded: 100% (of the cleaved material)

Percentage Purified: 50% (of the folded material)

Isolated Amount: 19 nmol (400 µL, 48.5 µM)

Isolated Yield: 0.3% (relative to resin loading)

Extrapolated Amount: 77.5 nmol (0.96 mg)

3.4.23.Folding Selection of Trx-1

Protein	Uniprot ID	Start	End	Length	Modifications
Thioredoxin-1	P0AA25	2	109	108	C-Term carboxamide
SDKIIHLTDD SFDTDVLKAD GAILVDFWAE WCGPCKMIAP ILDEIADEYQ GKLTVAKLNI DQNPGTAPKY GIRGIPTLLL FKNGEVAATK VGALSKGQLK EFLDANLA					

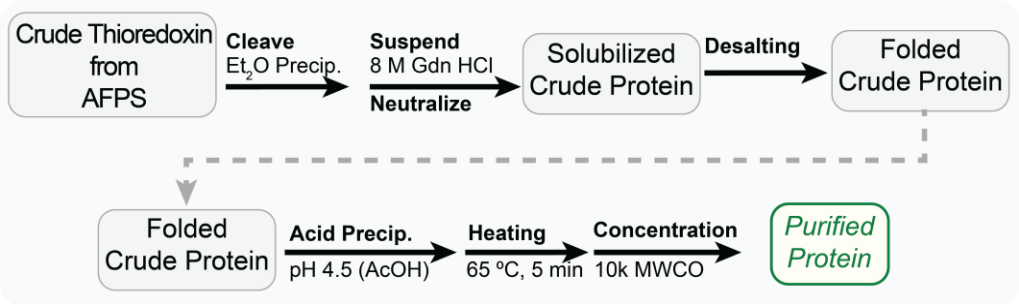
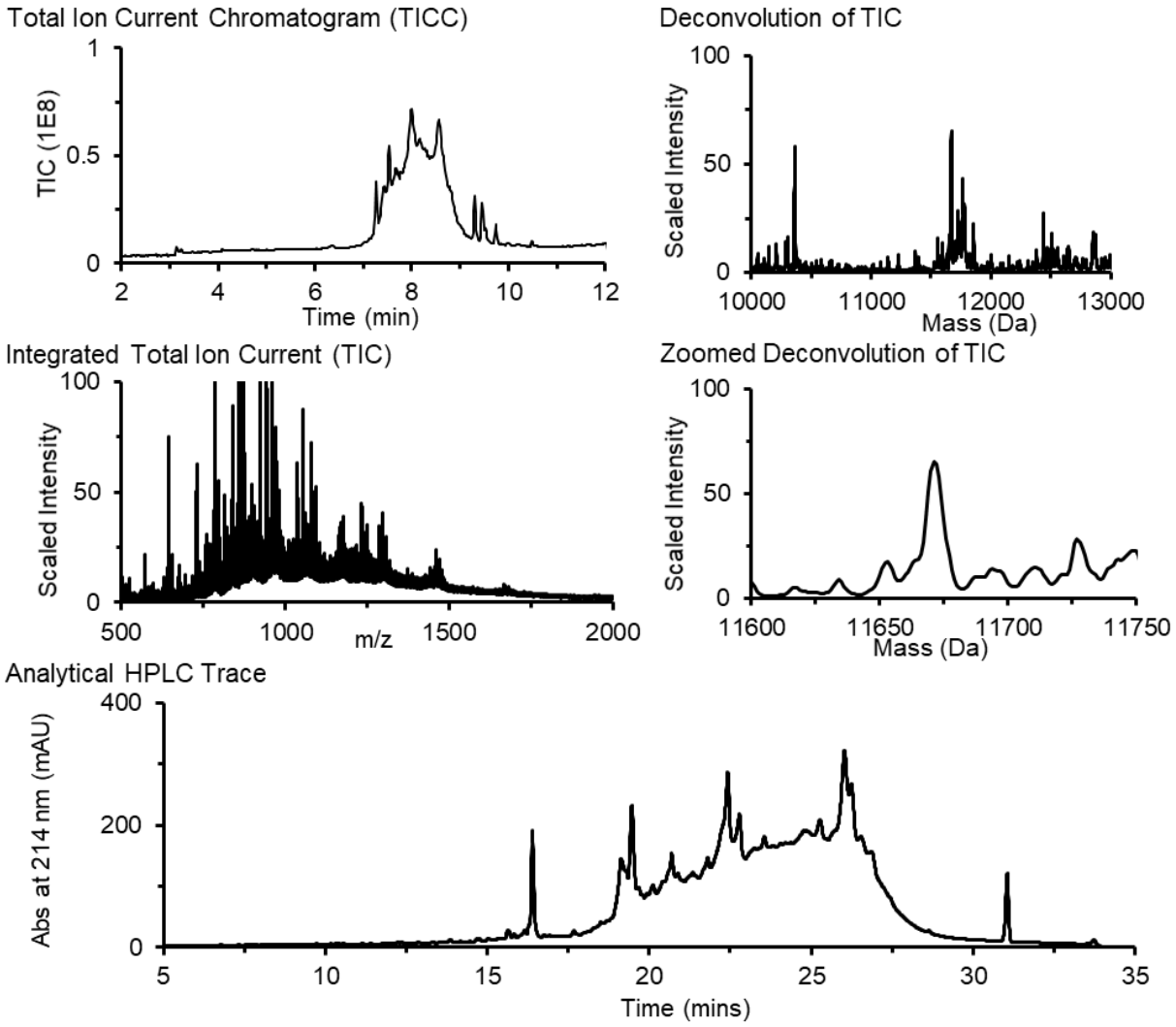


Figure 3.9 Scheme for the folding selection workflow for Trx-1.

Crude Analytical Characterization



Synthesis Scale: 23.4 µmol (130 mg, 0.18 mmol/g resin)

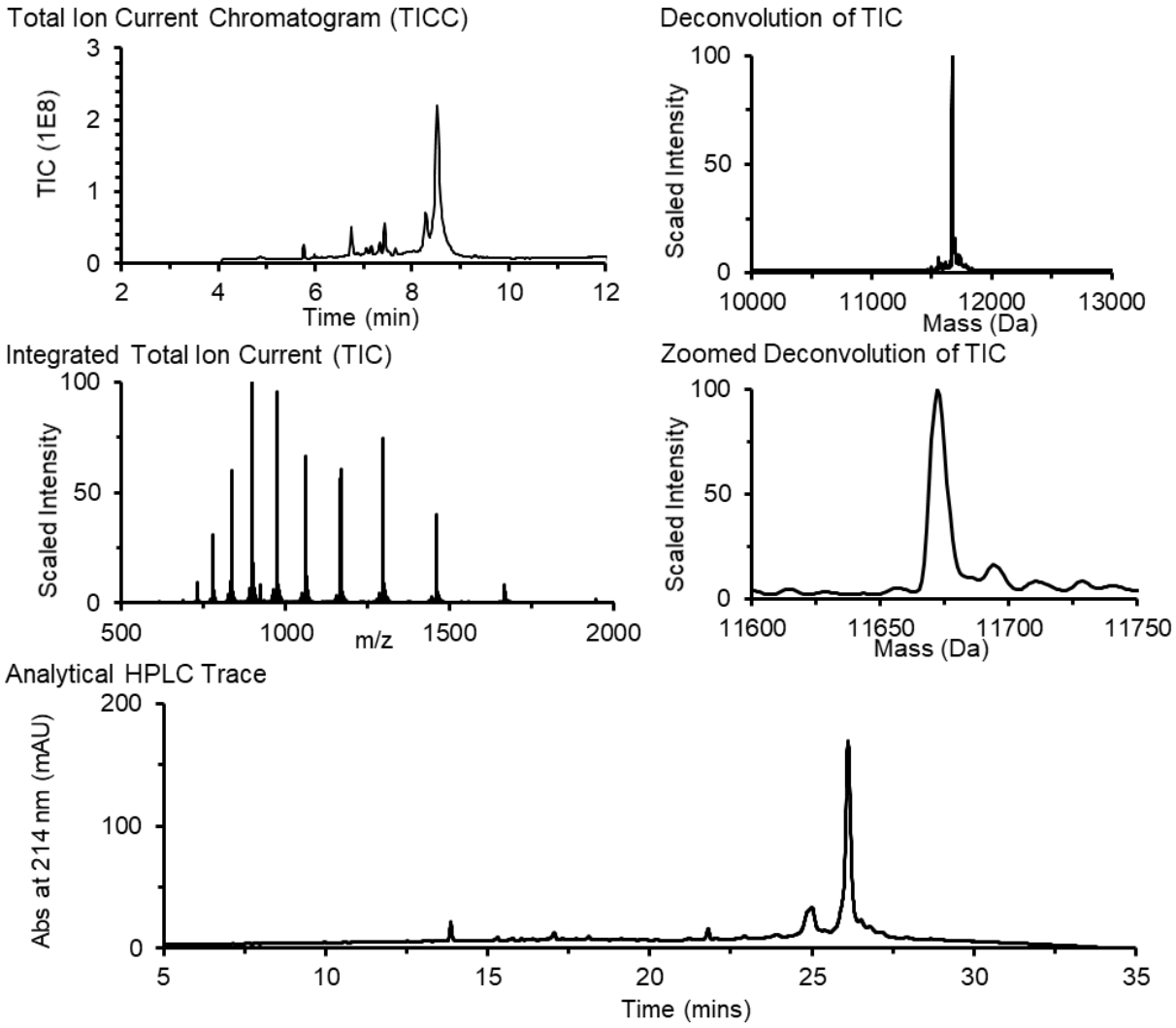
Crude RP-HPLC Method: Method 2

Crude LC-MS Method: Method 2

Calculated Mass: 11675 Da

Observed Mass: 11675 Da

Purified Analytical Characterization



Purified RP-HPLC Method: Method 2

Purified LC-MS Method: Method 2

Calculated Mass: 11673 Da

Observed Mass: 11673 Da

Percentage Cleaved: 100% (of the synthesis resin)

Percentage Folded: 100% (of the cleaved material)

Percentage Purified: 20% (of the folded material)

Isolated Amount: 5.7 nmol (20 µL, 283 µM)

Isolated Yield: 0.1% (relative to resin loading)

Extrapolated Amount: 28.3 nmol (0.3 mg)

3.4.24. Folding Selection of BCL11A

Protein	Uniprot ID	Start	End	Length	Modifications
BCL11A	Q9H165	740	835	96	C-Term carboxamide, Nle in place of Met
RSDTCEYCGK VFKNCSNLTV HRRSHTGERP YKCELCNYAC AQSSKLTRHM KTHGQVGKDV YKCEICKMPF SVYSTLEKHM KKWHSDRVLN NDIKTE					

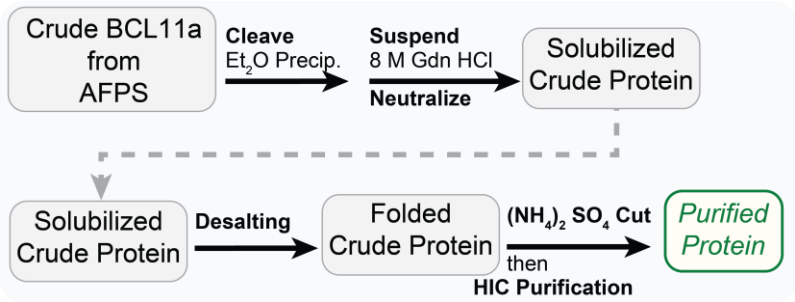
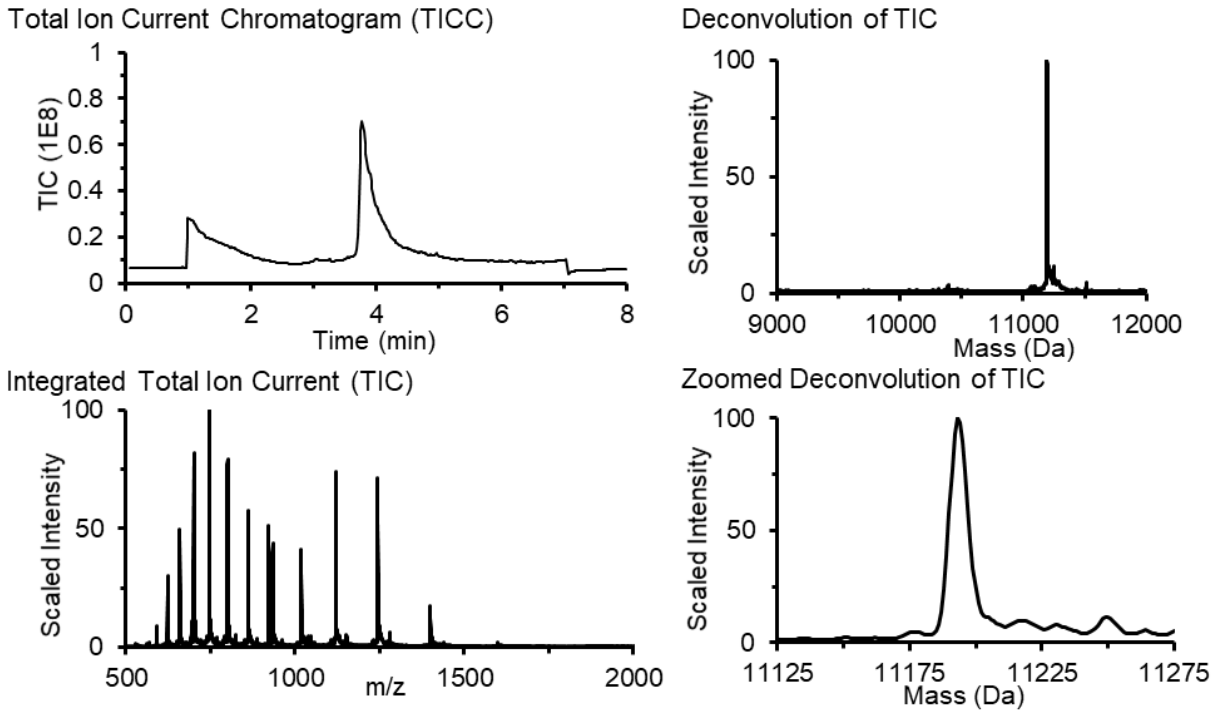


Figure 3.10 Scheme for the folding selection workflow for BCL11A.

Purified Analytical Characterization



Purified RP-HPLC Method: Method 2

Purified LC-MS Method: Method 3

Calculated Mass: 11193 Da

Observed Mass: 11193 Da

Synthesis Scale: 23.4 µmol (130 mg, 0.18 mmol/g resin)

Percentage Cleaved: 50% (of the synthesis resin)

Percentage Folded: 20% (of the cleaved material)

Percentage Purified: 100% (of the folded material)

Isolated Amount: 1.8 nmol (200 µL, 8.9 µM)

Isolated Yield: 0.08% (relative to resin loading)

Extrapolated Amount: 17.9 nmol (0.2 mg)

3.4.25. Folding Selection of Ubiquitin

Protein	Uniprot ID	Start	End	Length	Modifications
Ubiquitin	P0CG48	1	76	76	C-Term carboxamide, Nle in place of Met
MQIFVKTLTG KTITLEVEPS DTIENVKAKI QDKEGIPPDQ QRLIFAGKQL EDGRTLSDYN IQKESTLHLV LRLRGG					

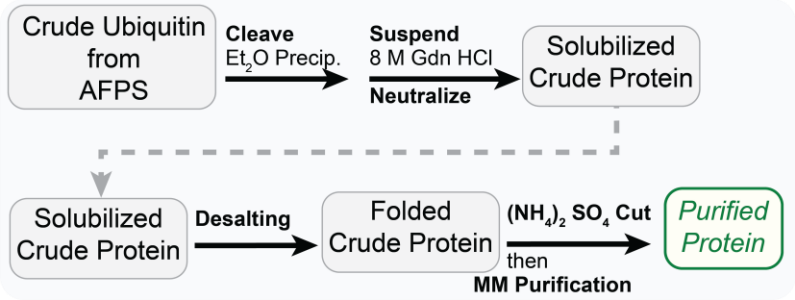
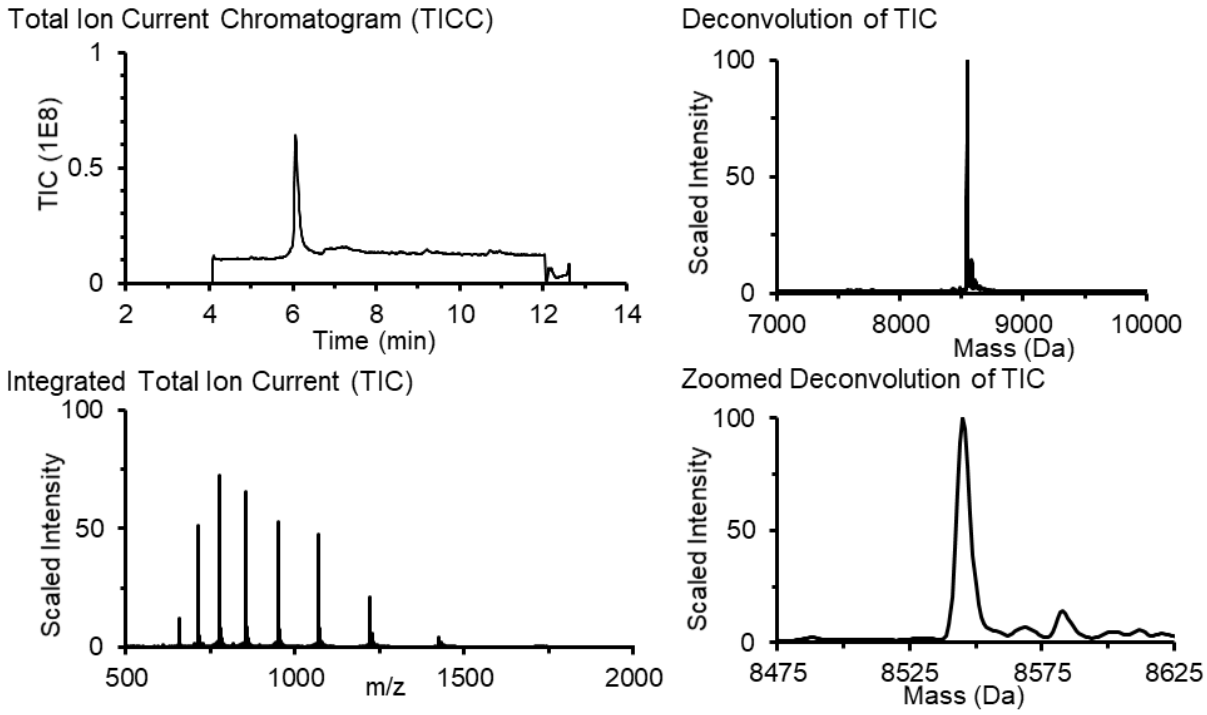


Figure 3.11 Scheme for the folding selection workflow for ubiquitin.

Purified Analytical Characterization



Purified LC-MS Method: Method 2

Calculated Mass: 8545 Da

Observed Mass: 8545 Da

Synthesis Scale: 23.4 µmol (130 mg, 0.18 mmol/g resin)

Percentage Cleaved: 15%

Percentage Folded: 100%

Percentage Purified: 100%

Isolated Amount: 0.6 µmol (1.5 mL, 0.40 mM)

Isolated Yield: 17.0% (relative to resin loading)

Extrapolated Amount: 3.9 µmol (34 mg)

3.4.26.Folding Selection of DHFR

Protein	Uniprot ID	Start	End	Length	Modifications
DHFR (Type II)	P00383	17	87	62	C-Term carboxamide
VFPSNATFGM GDRVRKKSGA AWQGQIVGWY CTNLTPEGYA VESEAHPGSV QIYPVAALER IN					

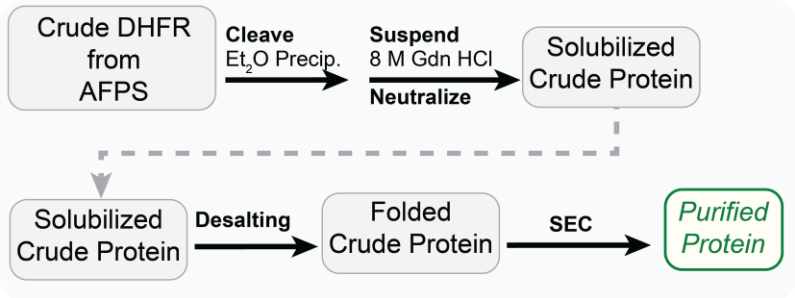
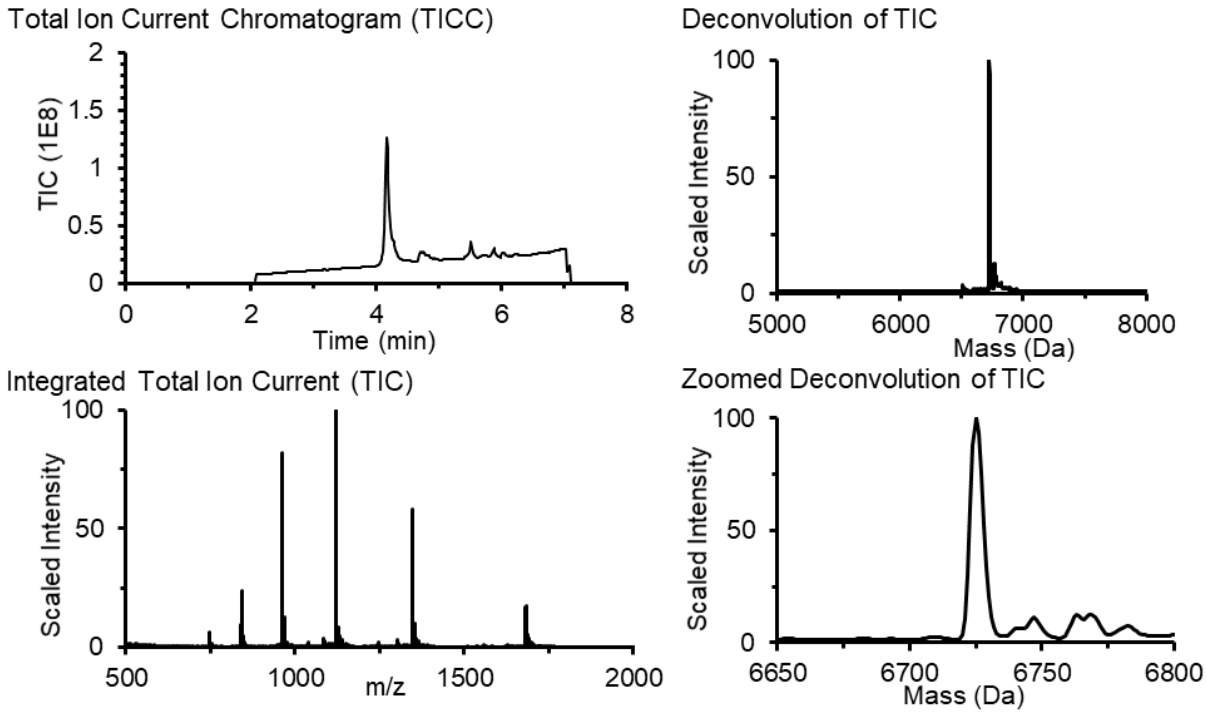


Figure 3.12 Scheme for the folding selection workflow for DHFR.

Purified Analytical Characterization



Purified RP-HPLC Method: Method 2

Purified LC-MS Method: Method 4

Calculated Mass: 6725 Da

Observed Mass: 6725 Da

Percentage Cleaved: 50% (of the synthesis resin)

Percentage Folded: 100% (of the cleaved material)

Percentage Purified: 100% (of the folded material)

Isolated Amount: 158 nmol (2 mL, 79 µM)

Isolated Yield: 1.3% (relative to resin loading)

Extrapolated Amount: 315.5 nmol (1.98 mg)

3.4.27.Folding Selection of STIP1 Variants

Protein	Uniprot ID	Start	End	Length	Modifications
STIP1-TPR1	P31948	1	118	118	C-Term carboxamide
MEQVNELKEK GNKALSVGNI DDALQCYSEA IKLDPHNHVL YSNRSAAYAK KGDYQKAYED GCKTVDLKPD WGKGYSRKAA ALEFLNRFEE AKRTYEEGLK HEANNPQLKE GLQNMEAR					

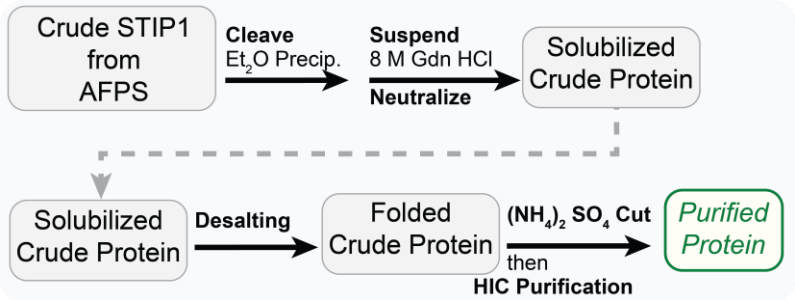
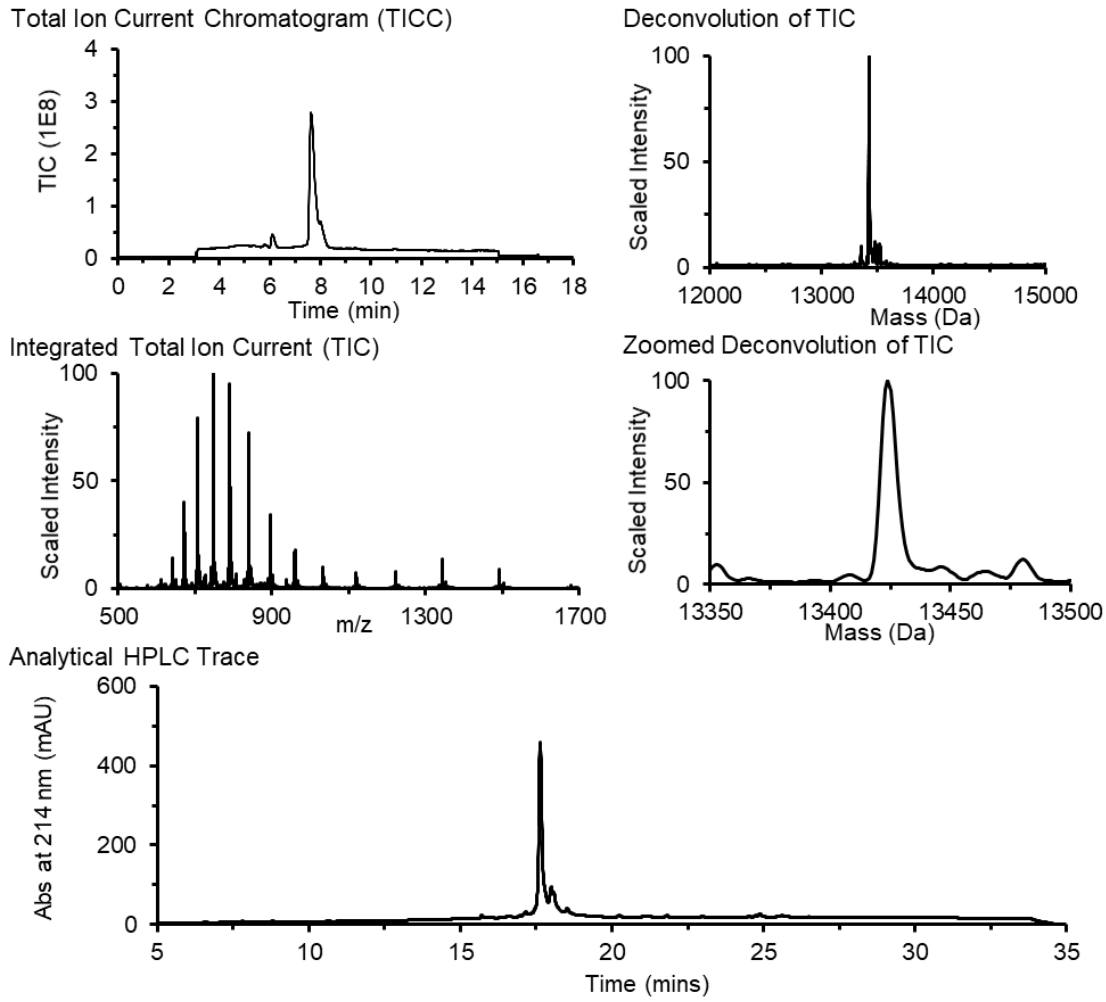


Figure 3.13 Scheme for the folding selection workflow for STIP1 and its variants.

Purified Analytical Characterization of STIP1-WT



Purified RP-HPLC Method: Method 2

Purified LC-MS Method: Method 5

Calculated Mass: 13424 Da

Observed Mass: 13424 Da

Synthesis Scale: 14 µmol (70 mg, 0.2 mmol/g resin)

Percentage Cleaved: 100% (of the synthesis resin)

Percentage Folded: 100% (of the cleaved material)

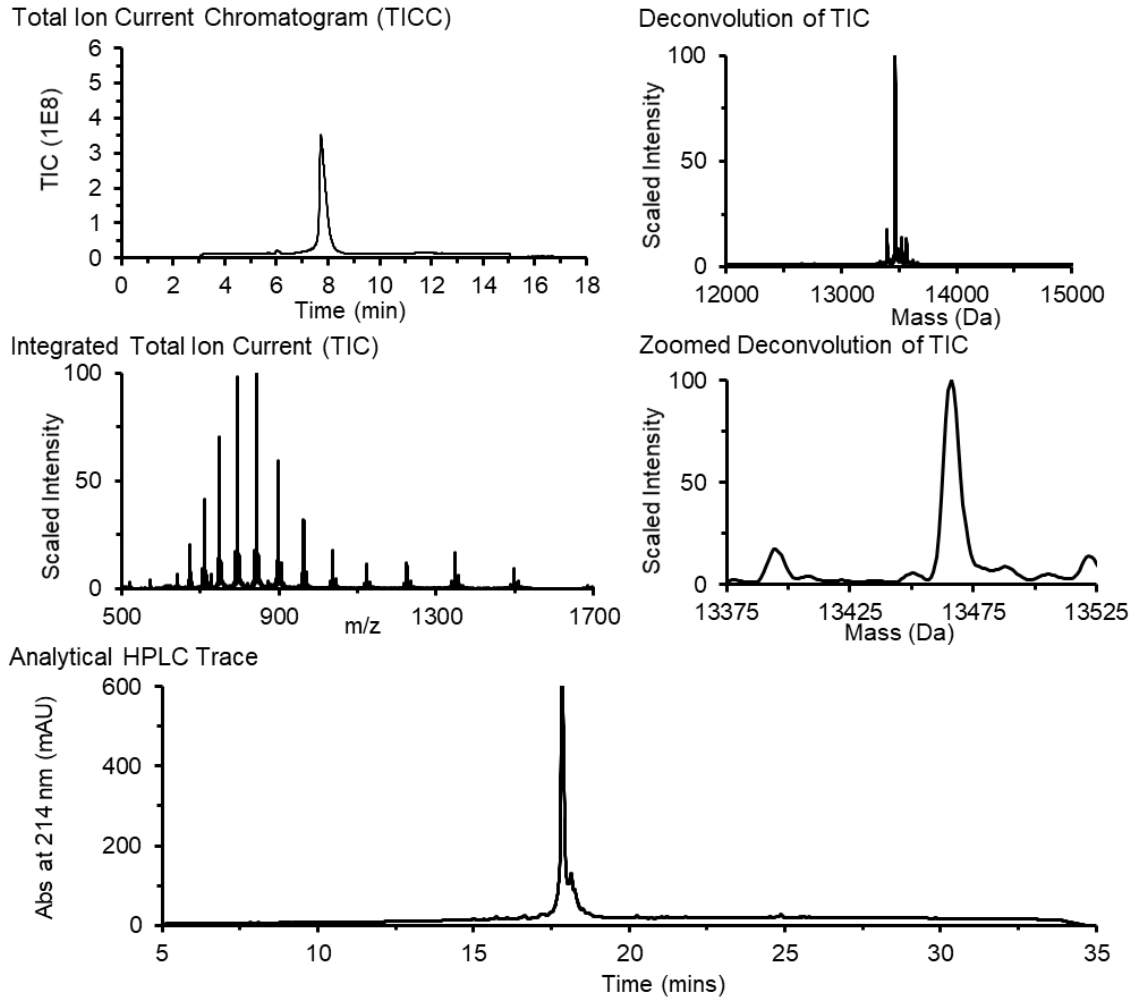
Percentage Purified: 50% (of the folded material)

Isolated Amount: 17.7 nmol (1.5 mL, 11.8 µM)

Isolated Yield: 0.25% (relative to resin loading)

Extrapolated Amount: 35 nmol (0.5 mg)

Purified Analytical Characterization of STIP1-K8-Ac



Purified RP-HPLC Method: Method 2

Purified LC-MS Method: Method 5

Calculated Mass: 13467 Da

Observed Mass: 13466 Da

Synthesis Scale: 14 µmol (70 mg, 0.2 mmol/g resin)

Percentage Cleaved: 100% (of the synthesis resin)

Percentage Folded: 100% (of the cleaved material)

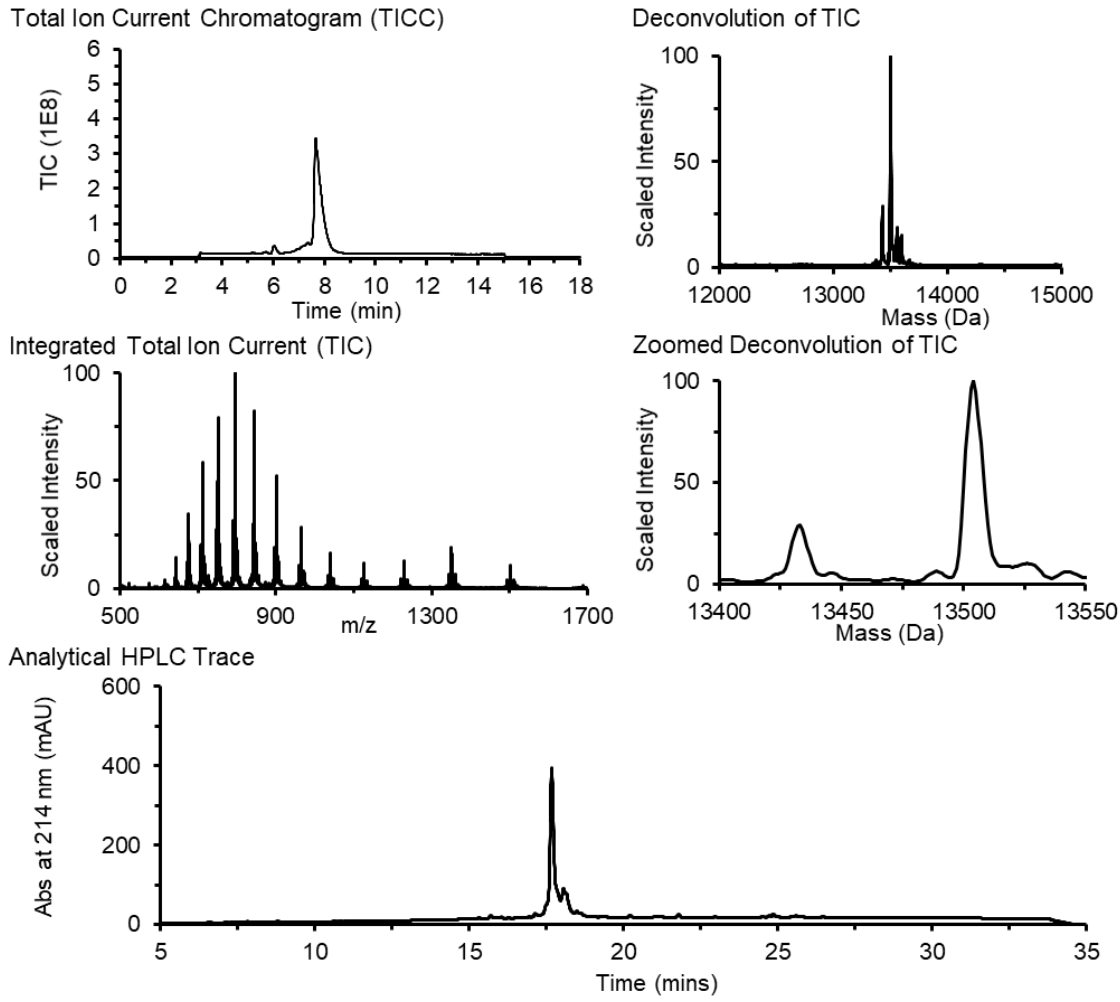
Percentage Purified: 50% (of the folded material)

Isolated Amount: 37 nmol (1.5 mL, 25 µM)

Isolated Yield: 0.5% (relative to resin loading)

Extrapolated Amount: 74.9 nmol (1 mg)

Purified Analytical Characterization of STIP1-S16-Phos



Purified RP-HPLC Method: Method 2

Purified LC-MS Method: Method 5

Calculated Mass: 13504 Da

Observed Mass: 13504 Da

Synthesis Scale: 14 µmol (70 mg, 0.2 mmol/g resin)

Percentage Cleaved: 100% (of the synthesis resin)

Percentage Folded: 100% (of the cleaved material)

Percentage Purified: 50% (of the folded material)

Isolated Amount: 19.5 nmol (1.5 mL, 13 µM)

Isolated Yield: 0.3% (relative to resin loading)

Extrapolated Amount: 39 nmol (0.5 mg)

3.5. Acknowledgements

Financial support for this work was provided by FOG Pharmaceuticals, NIH, and MIT discretionary funds (to B.L.P.). The NSF provided support for J.J.R.

3.6.Author Contributions

A.J.C. and B.L.P. conceptualized the research. A.J.C. and A.R. carried out the protein purifications, and A.J.C., L.L.S., and S.G. carried out the protein synthesis. A.J.C. and J.J.R. generated figures. A.J.C., A.L., and B.L.P. wrote the manuscript with input from all authors.

3.7.References

1. M. Zahn-Zabal, P.-A. Michel, A. Gateau, F. Nikitin, M. Schaeffer, E. Audot, P. Gaudet, P. D. Duek, D. Teixeira, V. Rech de Laval, K. Samarasinghe, A. Bairoch, L. Lane, The neXtProt knowledgebase in 2020: data, tools and usability improvements. *Nucleic Acids Research*. **48**, D328–D334 (2020).

2. M. Uhlen, C. Zhang, S. Lee, E. Sjöstedt, L. Fagerberg, G. Bidkhori, R. Benfeitas, M. Arif, Z. Liu, F. Edfors, K. Sanli, K. von Feilitzen, P. Oksvold, E. Lundberg, S. Hober, P. Nilsson, J. Mattsson, J. M. Schwenk, H. Brunnström, B. Glimelius, T. Sjöblom, P.-H. Edqvist, D. Djureinovic, P. Micke, C. Lindskog, A. Mardinoglu, F. Ponten, A pathology atlas of the human cancer transcriptome. *Science*. **357**, eaan2507 (2017).

3. E. Sjöstedt, W. Zhong, L. Fagerberg, M. Karlsson, N. Mitsios, C. Adori, P. Oksvold, F. Edfors, A. Limiszewska, F. Hikmet, J. Huang, Y. Du, L. Lin, Z. Dong, L. Yang, X. Liu, H. Jiang, X. Xu, J. Wang, H. Yang, L. Bolund, A. Mardinoglu, C. Zhang, K. von Feilitzen, C. Lindskog, F. Pontén, Y. Luo, T. Hökfelt, M. Uhlén, J. Mulder, An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science*. **367**, eaay5947 (2020).

4. R. Chen, Bacterial expression systems for recombinant protein production: E. coli and beyond. *Biotechnology Advances*. **30**, 1102–1107 (2012).

5. S. L. Irons, A. C. Chambers, O. Lissina, L. A. King, R. D. Possee, *Current Protocols in Protein Science*, in press, doi:10.1002/cpps.45.

6. J. Zhu, D. Hatton, "New Mammalian Expression Systems" in *New Bioprocessing Strategies: Development and Manufacturing of Recombinant Antibodies and Proteins*, B. Kiss, U. Gottschalk, M. Pohlscheidt, Eds. (Springer International Publishing, Cham, 2018; https://doi.org/10.1007/10_2016_55), pp. 9–50.

7. M. Hunter, P. Yuan, D. Vavilala, M. Fox, Optimization of Protein Expression in Mammalian Cells. *Current Protocols in Protein Science*. **95**, e77 (2019).

8. S. Gräslund, P. Nordlund, J. Weigelt, B. M. Hallberg, J. Bray, O. Gileadi, S. Knapp, U. Oppermann, C. Arrowsmith, R. Hui, J. Ming, S. dhe-Paganon, H. Park, A. Savchenko, A. Yee, A. Edwards, R. Vincentelli, C. Cambillau, R. Kim, S.-H. Kim, Z. Rao, Y. Shi, T. C. Terwilliger, C.-Y. Kim, L.-W. Hung, G. S. Waldo, Y. Peleg, S. Albeck, T. Unger, O. Dym, J. Prilusky, J. L. Sussman, R. C. Stevens, S. A. Lesley, I. A. Wilson, A. Joachimiak, F.

Collart, I. Dementieva, M. I. Donnelly, W. H. Eschenfeldt, Y. Kim, L. Stols, R. Wu, M. Zhou, S. K. Burley, J. S. Emtage, J. M. Sauder, D. Thompson, K. Bain, J. Luz, T. Gheyi, F. Zhang, S. Atwell, S. C. Almo, J. B. Bonanno, A. Fiser, S. Swaminathan, F. W. Studier, M. R. Chance, A. Sali, T. B. Acton, R. Xiao, L. Zhao, L. C. Ma, J. F. Hunt, L. Tong, K. Cunningham, M. Inouye, S. Anderson, H. Janjua, R. Shastry, C. K. Ho, D. Wang, H. Wang, M. Jiang, G. T. Montelione, D. I. Stuart, R. J. Owens, S. Daenke, A. Schütz, U. Heinemann, S. Yokoyama, K. Büssow, K. C. Gunsalus, Structural Genomics Consortium, Architecture et Fonction des Macromolécules Biologiques, Berkeley Structural Genomics Center, China Structural Genomics Consortium, Integrated Center for Structure and Function Innovation, Israel Structural Proteomics Center, Joint Center for Structural Genomics, Midwest Center for Structural Genomics, New York Structural GenomiX Research Center for Structural Genomics, Northeast Structural Genomics Consortium, Oxford Protein Production Facility, M. D. C. for M. M. Protein Sample Production Facility, RIKEN Structural Genomics/Proteomics Initiative, SPINE2-Complexes, Protein production and purification. *Nature Methods*. **5**, 135–146 (2008).

9. N. Hartrampf, A. Saebi, M. Poskus, Z. P. Gates, A. J. Callahan, A. E. Cowfer, S. Hanna, S. Antilla, C. K. Schissel, A. J. Quartararo, X. Ye, A. J. Mijalis, M. D. Simon, A. Loas, S. Liu, C. Jessen, T. E. Nielsen, B. L. Pentelute, Synthesis of proteins by automated flow chemistry. *Science*. **368**, 980 (2020).

10. M. Jbara, S. Pomplun, C. K. Schissel, S. W. Hawken, A. Boija, I. Klein, J. Rodriguez, S. L. Buchwald, B. L. Pentelute, Engineering Bioactive Dimeric Transcription Factor Analogs via Palladium Rebound Reagents. *J. Am. Chem. Soc.* **143**, 11788–11798 (2021).

11. S. B. H. Kent, CHEMICAL SYNTHESIS OF PEPTIDES AND PROTEINS. *Annu. Rev. Biochem.* **57**, 957–989 (1988).

12. P. Wang, S. Dong, J.-H. Shieh, E. Peguero, R. Hendrickson, M. A. S. Moore, S. J. Danishefsky, Erythropoietin Derived by Chemical Synthesis. *Science*. **342** (2013), pp. 1357–1360.

13. H. Sun, A. Brik, The Journey for the Total Chemical Synthesis of a 53 kDa Protein. *Acc. Chem. Res.* **52**, 3361–3371 (2019).

14. 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. *Nature Chemical Biology*. **13** (2017), p. 464.

15. A. Saebi, J. S. Brown, V. M. Marando, N. Hartrampf, N. M. Chumbler, S. Hanna, M. Poskus, A. Loas, L. L. Kiessling, D. T. Hung, B. L. Pentelute, Rapid Single-Shot Synthesis of the 214 Amino Acid-Long N-Terminal Domain of Pyocin S2. *ACS Chem. Biol.* **18**, 518–527 (2023).

16. R. K. Scopes, *Protein purification: principles and practice* (Springer Science & Business Media, 1993).

17. D. Walls, R. McGrath, S. T. Loughran, "A Digest of Protein Purification" in *Protein Chromatography: Methods and Protocols*, D. Walls, S. T. Loughran, Eds. (Humana Press, Totowa, NJ, 2011; https://doi.org/10.1007/978-1-60761-913-0_1), pp. 3–23.

18. F. Barany, Two-codon insertion mutagenesis of plasmid genes by using single-stranded hexameric oligonucleotides. *Proceedings of the National Academy of Sciences*. **82**, 4202–4206 (1985).

19. T. Marti, H. Otto, S. J. Rösselet, M. P. Heyn, H. G. Khorana, Consequences of amino acid insertions and/or deletions in transmembrane helix C of bacteriorhodopsin. *Proceedings of the National Academy of Sciences*. **89**, 1219–1223 (1992).

20. K. T. O'Neil, A. C. I. Bach, W. F. DeGrado, Structural consequences of an amino acid deletion in the B1 domain of protein G. *Proteins: Struct., Funct., Genet.* **41**, 323–333 (2000).

21. D. D. Jones, *Nucleic Acids Res.*, in press, doi:10.1093/nar/gni077.

22. M. D. Simon, Y. Maki, A. A. Vinogradov, C. Zhang, H. Yu, Y.-S. Lin, Y. Kajihara, B. L. Pentelute, d-Amino Acid Scan of Two Small Proteins. *J. Am. Chem. Soc.* **138**, 12099–12111 (2016).

23. Y. Yang, *Side Reactions in Peptide Synthesis* (Tsinghua University Press Limited; Academic Press, 2016; 10.1016/C2013-0-13700-2).

24. A. J. Callahan, S. Gandhesiri, T. L. Travaline, L. Lozano Salazar, S. Hanna, Y.-C. Lee, K. Li, O. S. Tokareva, J.-M. Swiecicki, A. Loas, G. L. Verdine, J. H. McGee, B. L. Pentelute, Single-Shot Flow Synthesis of D-Proteins for Mirror-Image Phage Display. *ChemRxiv.*

25. P. T. Wingfield, *Curr Protoc Protein Sci*, in press, doi:10.1002/0471140864.psa03fs13.

26. H. P. Jennissen, "Hydrophobic Interaction Chromatography" in *eLS* (2016; https://doi.org/10.1002/9780470015902.a0002678.pub4), pp. 1–12.

27. B. H. J. Hofstee, N. Frank Otillio, Modifying factors in hydrophobic protein binding by substituted agaroses. *Journal of Chromatography A*. **161**, 153–163 (1978).

28. P. M. Cummins, B. F. O'Connor, "Hydrophobic Interaction Chromatography" in *Protein Chromatography: Methods and Protocols*, D. Walls, S. T. Loughran, Eds. (Humana Press, Totowa, NJ, 2011; https://doi.org/10.1007/978-1-60761-913-0_24), pp. 431–437.

29. B. R. Kelemen, T. A. Klink, M. A. Behike, S. R. Eubanks, P. A. Leland, R. T. Raines, Hypersensitive substrate for ribonucleases. *Nucleic Acids Research*. **27**, 3696–3701 (1999).

30. 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).

31. A. A. Vinogradov, E. D. Evans, B. L. Pentelute, Total synthesis and biochemical characterization of mirror image barnase. *Chem. Sci.* **6**, 2997–3002 (2015).

32. A. Holmgren, Thioredoxin catalyzes the reduction of insulin disulfides by dithiothreitol and dihydrolipoamide. *Journal of Biological Chemistry*. **254**, 9627–9632 (1979).

33. T. C. Laurent, E. C. Moore, P. Reichard, Enzymatic Synthesis of Deoxyribonucleotides: IV. ISOLATION AND CHARACTERIZATION OF THIOREDOXIN, THE HYDROGEN DONOR FROM ESCHERICHIA COLI B. *Journal of Biological Chemistry*. **239**, 3436–3444 (1964).

34. A. Holmgren, P. Reichard, Thioredoxin 2: Cleavage with Cyanogen Bromide. *European Journal of Biochemistry*. **2**, 187–196 (1967).

35. S. Laps, H. Sun, G. Kamnesky, A. Brik, Palladium-Mediated Direct Disulfide Bond Formation in Proteins Containing S-Acetamidomethyl-cysteine under Aqueous Conditions. *Angewandte Chemie International Edition*. **58**, 5729–5733 (2019).

36. G. B. Vamisetti, G. Satish, P. Sulkshane, G. Mann, M. H. Glickman, A. Brik, On-Demand Detachment of Succinimides on Cysteine to Facilitate (Semi)Synthesis of Challenging Proteins. *J. Am. Chem. Soc.* **142**, 19558–19569 (2020).

37. Y. Yang, Z. Xu, C. He, B. Zhang, Y. Shi, F. Li, Structural insights into the recognition of γ-globin gene promoter by BCL11A. *Cell Research*. **29**, 960–963 (2019).

38. K. N. Swatek, D. Komander, Ubiquitin modifications. *Cell Research*. **26**, 399–422 (2016).

39. G. Satish, G. B. Vamisetti, A. Brik, "Chemical Synthesis of Ubiquitinated Proteins for Biochemical Studies" in *Total Chemical Synthesis of Proteins* (2021; https://doi.org/10.1002/9783527823567.ch14), pp. 383–410.

40. R. M. M. Brito, F. B. Rudolph, P. R. Rosevear, Conformation of NADP+ bound to a type II dihydrofolate reductase. *Biochemistry*. **30**, 1461–1469 (1991).

41. R. Y.-R. Wang, C. M. Noddings, E. Kirschke, A. G. Myasnikov, J. L. Johnson, D. A. Agard, Structure of Hsp90–Hsp70–Hop–GR reveals the Hsp90 client-loading mechanism. *Nature*. **601**, 460–464 (2022).

42. Nitika, C. M. Porter, A. W. Truman, M. C. Truttmann, Post-translational modifications of Hsp70 family proteins: Expanding the chaperone code. *Journal of Biological Chemistry*. **295**, 10689–10708 (2020).

43. G. L. Rosano, E. S. Morales, E. A. Ceccarelli, New tools for recombinant protein production in Escherichia coli: A 5-year update. *Protein Science*. **28**, 1412–1422 (2019).

44. C. Choudhary, C. Kumar, F. Gnad, M. L. Nielsen, M. Rehman, T. C. Walther, J. V. Olsen, M. Mann, Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. *Science*. **325**, 834–840 (2009).

45. N. Dephoure, C. Zhou, J. Villén, S. A. Beausoleil, C. E. Bakalarski, S. J. Elledge, S. P. Gygi, A quantitative atlas of mitotic phosphorylation. *Proceedings of the National Academy of Sciences*. **105**, 10762–10767 (2008).

4. Automated Flow Synthesis of Peptide-PNA Conjugates

The work presented in this chapter has been reproduced from the following publication with the permission of Springer Nature:

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 Central Science* 2022, 8, 2, 205-213

4.1. Introduction

Antisense oligonucleotide (ASO)-based therapeutic development is gathering momentum following the recent FDA approval of the drugs Eteplirsen (*1*), Golodirsen (*2*), Casimersen (*3*), Viltepso (*4*) (based on phosphorodiamidate morpholino oligomer, PMO) and Spinraza (*5*) (based on 2′-*O*-methoxyethyl-phosphorothioate). Next-generation technology for enhanced ASO tissue penetration using cell-penetrating peptides conjugated to PMO (PPMO) is under development at Sarepta Therapeutics, with promising clinical outcomes (*6*). As a charge-neutral ASO class, peptide nucleic acids (PNAs) are also under development (*7-8*).

PNAs are artificial single-stranded DNA-like molecules in which the sugar-phosphodiester moiety is replaced with an uncharged *N*-(2-aminoethyl) glycine unit and the nucleobases are attached via a methyl carbonyl linker (*7, 8*). Due to the attenuated electrostatic repulsion in PNA-DNA/RNA duplexes, PNAs can hybridize complementary DNA and RNA with higher affinity and specificity than DNA-DNA/RNA duplexes (*9*). Additionally, the amide-based PNA backbone offers unique physiochemical properties including superior chemical, thermal, and enzymatic stability (*10*).

PNAs have found a wide range of chemical and biological applications (*11-13*). PNAs are currently being developed as antisense agents for diseases including cancer (*14*), monogenic blood disorders (*15*), novel antibiotics (*13, 16*), and antivirals (*17, 18*), but to date have yet to become a drug. PNAs can be limited by low solubility, poor intracellular delivery, and renal clearance (*19*). Despite these limitations, some PNAs display promising biological activity (*11-13*). Specifically, longer PNA sequences (> 15-mer) can be endowed with higher affinity and lower off-target toxicities (*20*).

The covalent attachment of a cell-penetrating peptide (CPP) to a PNA can help to overcome some of the development challenges (*19*). CPP conjugation can improve the PNA solubility, but also reduces its physiological clearance rate by promoting cellular internalization (*19, 21*). Despite the utility of CPP-PNA conjugates, they can be toxic to cells and animals (*22*) and often require systematic structure-function studies that optimize delivery while minimizing toxicity (*23*).

Existing standard batch protocols afford efficient access to PNA sequences with fewer than 15 bases while longer sequences remain challenging to synthesize (*20*) (Figure 4.1a). PNA synthesis can be limited by on-resin aggregation and side-reactions including deletion, rearrangement, isomerization, and nucleobase additions, resulting in low yield of the desired product (*20*). Efforts to improve the synthetic efficiency include capping and double-couplings, yet these approaches still fail to produce long PNA sequences robustly.

As the causative agent of coronavirus disease 2019 (COVID-19), the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (*24*) is a pathogen of immense importance to global public health. Developing innovative direct-acting antiviral agents is needed to eliminate this virus. PNAs have been investigated and demonstrated efficacy against SARS-CoV *in vitro* (*17*). Given the high sequence similarity between SARS-CoV-2 and SARS-CoV (*15*), antisense PNAs provide a promising avenue to achieve SARS-CoV-2 inhibition through a mechanism of reducing the targeted RNA expression via sequence-specific steric blocking (Figure 4.1b).

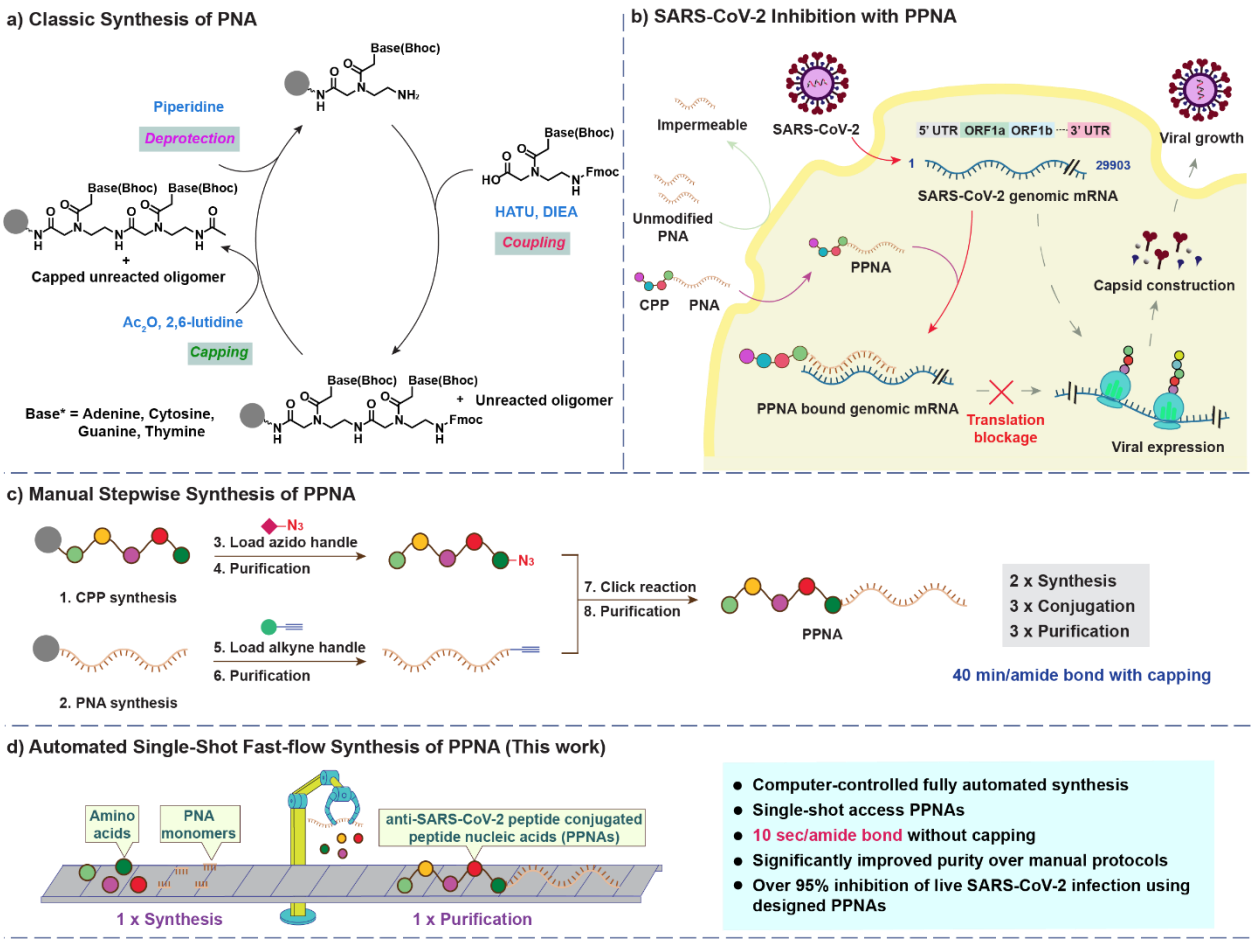


Figure 4.1 Automated single-shot technology can rapidly produce on-demand customized PPNA sequences that inhibit SARS-CoV-2.

(a) Typical coupling, deprotection, and capping procedures during PNA assembly in the solid phase. (b) An antisense PPNA binding to the 5' UTR of the SARS-CoV-2 genomic transcript can prevent the expression of viral genes and subsequently inhibit viral growth. (c) Workflow for the manual stepwise synthesis of PPNAs with click chemistry. (d) An automated fast-flow synthesizer allows for rapid manufacture of PPNAs in a single-shot. Fmoc: 9-fluorenylmethoxycarbonyl; HATU: O-(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate. DIEA: N,N-diisopropylethylamine. Ac2O: acetic anhydride. Fmoc: 9-fluorenylmethoxycarbonyl. Bhoc: benzhydryloxycarbonyl.

In order to find the most effective sequence for a given indication, investigation of multiple PNA candidates in a high-throughput manner is required. To remove the obstacles for the rapid production of these target-specific libraries of PNA strands, we developed a highly efficient technology in which chemistry is matched with an automated fast-flow instrument. This fully automated platform increases the PNA synthesis rate by nearly an order of magnitude relative to commercial state-of-the-art instruments with

improved crude purity. The new flow synthesis protocol described here only requires 10 seconds for each amide bond formation between PNA monomers, a significant improvement over commercial microwave peptide synthesizers26 (10 min/amide bond at 45 °C) or the DNA synthesizer Expedite 890927 (32 min/coupling cycle with capping at room temperature). Further, with flow-based synthesis, no capping or double couplings are needed. A variable temperature design increases coupling efficiency, while reducing on-resin aggregation and other side-reactions.

Instead of using common stepwise synthesis (*24*) via click chemistry (*29*) (Figure 4.1c), the high coupling efficiency achieved with our flow technology allows direct manufacture of long (>15-mer) PNA conjugates with cell-penetrating peptides in a single-shot (Figure 4.1d). This production strategy is convenient for simultaneous investigation of the bioactivity, toxicity, and cell uptake of multiple PNAs in view of establishing structure-function relationships.

To demonstrate the reliability and applicability of our method, we synthesized an 18-mer PPNA which hybridizes to the β-thalassemia gene sequence, IVS2-654 (*30*), for splice-correction that was covalently attached to a 12-mer CPP. The enhanced green fluorescence protein (EGFP) assay (*30*) showed that the single-shot synthesized PPNA presented 3-fold activity compared to transfected PNA. The utility of this automated technology was further demonstrated by the synthesis of eight PPNAs targeting SARS-CoV-2 genomic RNA in a single day. One of the designed PPNAs targeting the 5' untranslated region (5'UTR) reduced the viral titers by over 95% in a live infection assay.

4.2. Results

4.2.1. Automated Microscale Flow Synthesizer Design

The automated flow PNA synthesizer consists of seven modules including a central control computer, solution storage system, three HPLC pumps, three multi-position valves, heating elements, reaction zone, and a UV-visible detector. A modular script in the Mechwolf programming environment (*31*) controls the instrument (Figure 4.2a). During a coupling reaction, three HPLC pumps draw reagents stored under nitrogen atmosphere from the storage module and the desired PNA monomer, activator,

and base solutions are merged using a valve. The mixture flows through a module electrically heated at 70 ˚C, forming an activated ester. The activated PNA monomer flows next through the reaction zone, a packed bed of resin maintained at 70 ˚C, where amide bond formation is completed within 10 s. During the deprotection step, the piperidine solution flows through a room temperature loop controlled by a multi-position valve (Figure 4.2a) and meets with the 70 ˚C reactor to generate an ~40 ˚C environment, which enables rapid and efficient deprotection with minimized nucleobase adducts. An in-line UV-visible detector was used to monitor the composition of the spent reagent solution (Figure 4.2b). The deprotection efficiency and the mass transfer rate through the resin can be inferred in through the Fmoc-removal absorbance chromatogram (*32*). This modular and in-line detection design allows for chemistry optimization not only for PNAs but also other biopolymer syntheses.

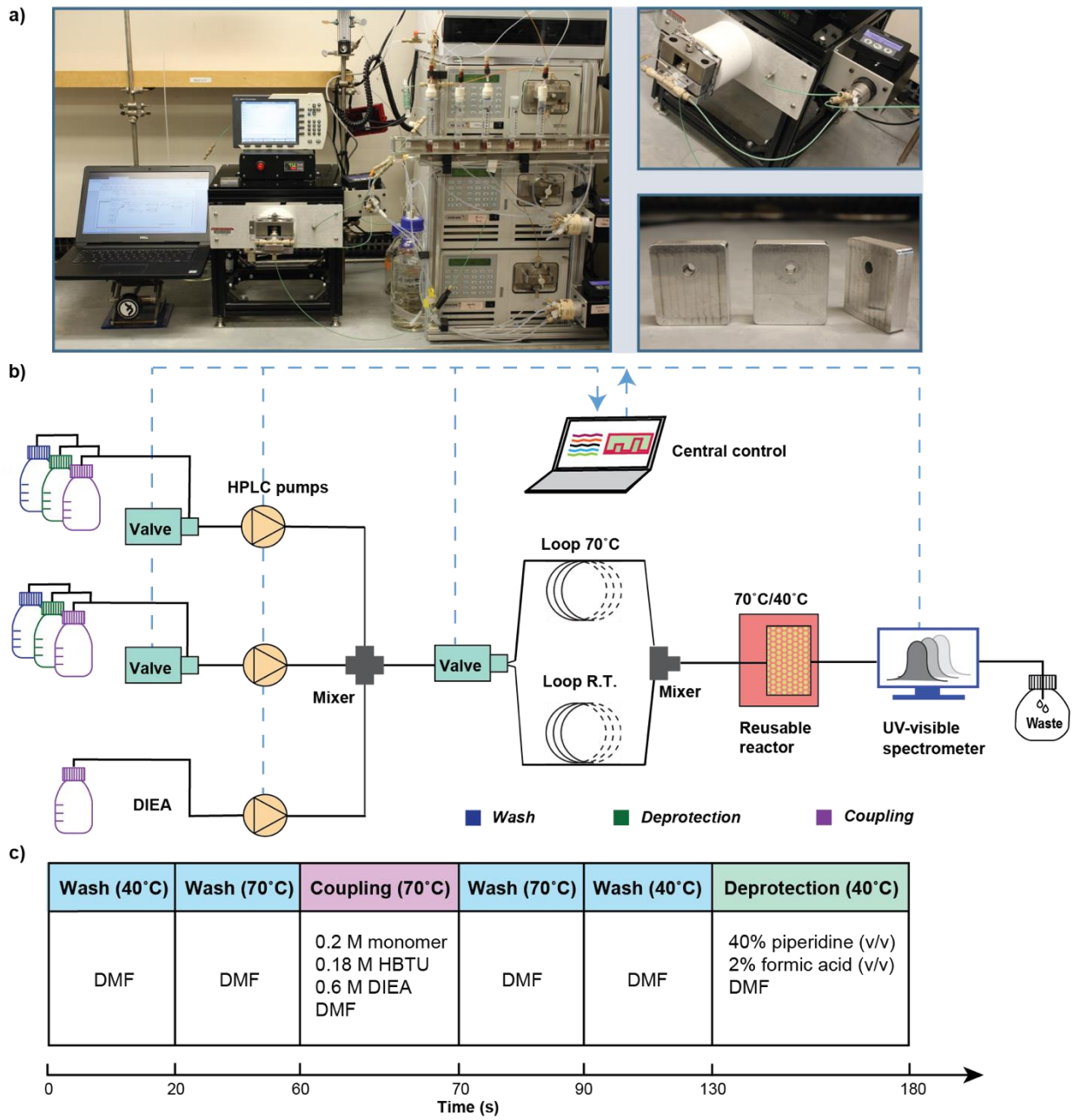


Figure 4.2 Automated flow PNA synthesis enables 10-second amide bond formation and complete solid-phase synthesis cycles in three minutes.

(a) Photographs of the automated flow solid-phase synthesizer: full image, variant temperature valve, and reusable reactors. (b) Process flow diagram. Amino acids, activating agents, and DIEA are mixed by three HPLC pumps. Two multi-position valves control the selection of the amino acids and activating agents. The third valve controls variant temperature flow paths. Amino acids are activated in the 70 ˚C loop, then flowed over the resin bed housed in a reusable reactor. The effluent is passed through a UV-visible spectrometer to waste. (c) Cycle diagram showing the duration of each step, the solution composition during each step after mixing, and the time used at each step. HBTU:

O-(benzotriazol-1-yl)-N,N,N,N'-tetramethyluronium hexafluorophosphate; DMF: N,N-dimethylformamide.

The reusable reactor body (Figure 4.2a) is designed for a 7.5 µmol scale synthesis and is typically loaded with 15 mg of 0.5 mmol/g Rink Amide resin. While reducing expensive monomer consumption and overall cost, this design can deliver milligrams of pure product, which is often sufficient for subsequent biological characterization. A detailed workflow and timeline for a synthesis process is listed in Figure 4.2c, which presents the amide bond formation step and the overall solid-phase PNA synthesis cycle within three minutes. Additional details can be found in Sections 4.4.5 to 4.4.7.

Entry	Activator	Couple temp. (˚C)	Depro. sol.	Depro. temp. (˚C)	Couple time	Purity[b] (%)	Isomer (%)	Deletion[c] (%)	Base add. (%)	Unknown (%)
1	PyAOP	RT	20% piperidine	RT	30 min	***57***	15	4	2	22
2	PyAOP	70	20% piperidine	70	10 sec	***70***	9	10	7	4
3	PyAOP	70	20% piperazine	70	10 sec	***70***	7	9	4	10
4	PyAOP	70	20% morpholine	70	10 sec	***61***	12	8	8	11
5	PyAOP	70	20% piperidine, 1% Formic acid	70	10 sec	***68***	11	10	3	8
6	PyAOP	70	20% piperidine, 2% Formic acid	70	10 sec	***65***	11	11	2	11
7	PyAOP	70	20% piperidine, 1% Formic acid	40	10 sec	***76***	5	9	<1	9
8	PyAOP	80	20% piperidine, 1% Formic acid	45	10 sec	***65***	8	9	2	16
9	PyAOP	90	20% piperidine, 1% Formic acid	50	10 sec	***47***	12	10	3	28
10	HATU	70	20% piperidine, 1% Formic acid	40	10 sec	***81***	5	9	<1	4
11	HBTU	70	20% piperidine, 1% Formic acid	40	10 sec	***90***	<1	6	<1	3

Table 18 Evaluation of reaction conditions for the automated flow PNA synthesis[a]

[a]The PNA sequence: ACTG-Gly-$CONH_2$. Conditions: manual synthesis: 100 mg Rink Amide resin (0.18 mmol/g), 6 eq PNA monomer, 5.8 eq PyAOP, 6 eq DIEA, and 6 eq 2,5-lutidine in DMF. Automated flow synthesis: 15 mg Rink Amide resin (0.5 mmol/g), 10 eq PNA monomer, 9.6 eq activator, 30 eq DIEA in DMF, flow rate: 2.5 mL/min. See SI for details. [b]The crude purity was determined with HPLC UV absorbance at 280 nm. [c]Total deletions of A, C, T, G, and Gly were summed. PyAOP: (7-Azabenzotriazol-1-yloxy)trispyrrolidinophosphonium hexafluorophosphate.

4.2.2. Optimization of Automated PNA Synthesis

In-line UV-visible monitoring combined with liquid chromatography-mass spectrometry (LC-MS) and high-performance liquid chromatography (HPLC) product characterization allows for rapid optimization of PNA synthesis conditions. We began the optimization with a manually synthesized 4-mer PNA in ~4 h (*20, 33*) with 57% crude purity (Table 18, entry 1). The off-target products contained ~15% isomers, ~4% deletions, and ~2% nucleobase adducts (see Section 4.4.8). Next, the same PNA sequence was synthesized on the automated flow synthesizer in 15 min at 70 ˚C (Table 18, entry 2). High temperature accelerated both on-target (70%) and off-target reactions, especially nucleobase adducts (7%) and deletions (10%). Further base screening, including piperazine and morpholine, revealed that piperidine was optimal for the Fmoc-deprotection at 70 ˚C (Table 18, entry 3-4).

A major focus of the high temperature optimization was the prevention of piperidine adduct formation during the deprotection steps. Addition of formic acid to the deprotection solutions to prevent base-mediated aspartimide formation is a common practice (*34*), but we found the lower basicity is also useful in preventing nucleobase adduct formation. Formic acid (1% solution in 20% piperidine-DMF, v/v) was therefore used as an additive for deprotection and successfully decreased the base-adduct ratio to 3% (Table 18, entry 5). Higher formic acid concentration can further inhibit the formation of base adducts but results in additional impurities (Table 18, entry 6). The most effective strategy, however, was to decrease the deprotection temperature to 40 ˚C, which was controlled by a multi-position valve and a “T” connector (Figure 4.2a). Under this condition, only negligible nucleobase adducts (<1%) were generated (Table 18, entry 7).

Higher coupling temperatures (80 ˚C and 90 ˚C) for synthesis were tested and increased amounts of impurities (16% and 28%, respectively) were found, indicating the synthesis temperature of 70 ˚C on this flow instrument is optimal (Table 18, entry 8-9). A significant increase in purity (81%) was achieved by using HATU as an activator, but 5% isomerization was still observed (Table 18, entry 10). Finally, after replacing HATU with HBTU, we found that the condition of entry 11 in Table 18 yielded 90% pure PNA with <1% isomers and <1% nucleobase adducts. The entry 11 conditions therefore were used in all subsequent PNA syntheses.

4.2.3. Automated Single-Shot Synthesis of PPNA

The optimized flow recipe provides PNA sequences with significantly improved purity to manual protocols in a fraction of the time. As depicted in Figure 4.3, using batch protocols, production of a 4-mer PNA was completed in 4 h. In comparison, our flow protocol provided the same sequence in 15 min with superior purity (90% vs 57% for batch synthesis), as demonstrated by both HPLC and LC-MS chromatograms (Figure 4.3a-d). To further test the synthetic efficiency of this methodology, an 18-mer PNA with a three-lysine (K_3) linker (*35*), which hybridizes to the β-thalassemia gene sequence (IVS2-654), was prepared in about one hour with 60% crude purity. We purified 4 mg of pure PNA with >95% purity for downstream biological characterizations (Figure 4.3e, 4.3f).

This efficient protocol allows for single-shot synthesis of PPNAs. Conjugating cell-penetrating peptides (CPPs) with PNAs through click chemistry is a strategy for PNA intracellular delivery (*36, 37*). However, this method can be time consuming (Figure 4.1c) and might require multiple purifications thereby reducing yields. To streamline PPNA preparation, we established a single-shot strategy to produce PPNAs using our automated platform. A 12-mer CPP, RXRRBRRXRRBR-CONH2 (Bpep, X = 6-aminohexanoic acid; B = β-alanine) (*38*) was pre-synthesized onto the resin on a peptide synthesizer developed previously in our lab (*39*), and the Bpep-loaded resin was transferred directly on our microscale instrument for PPNA synthesis. In this way, an anti-IVS2-654 PPNA was synthesized in 1.7 h with 52% crude purity and 1.2 mg of pure material (>95% purity) were obtained after purification (Figure 4.3g, 4.3h). This automated protocol not only achieved rapid production but also enabled an efficient synthesis of PPNA sequences that might be difficult to synthesize with batch.

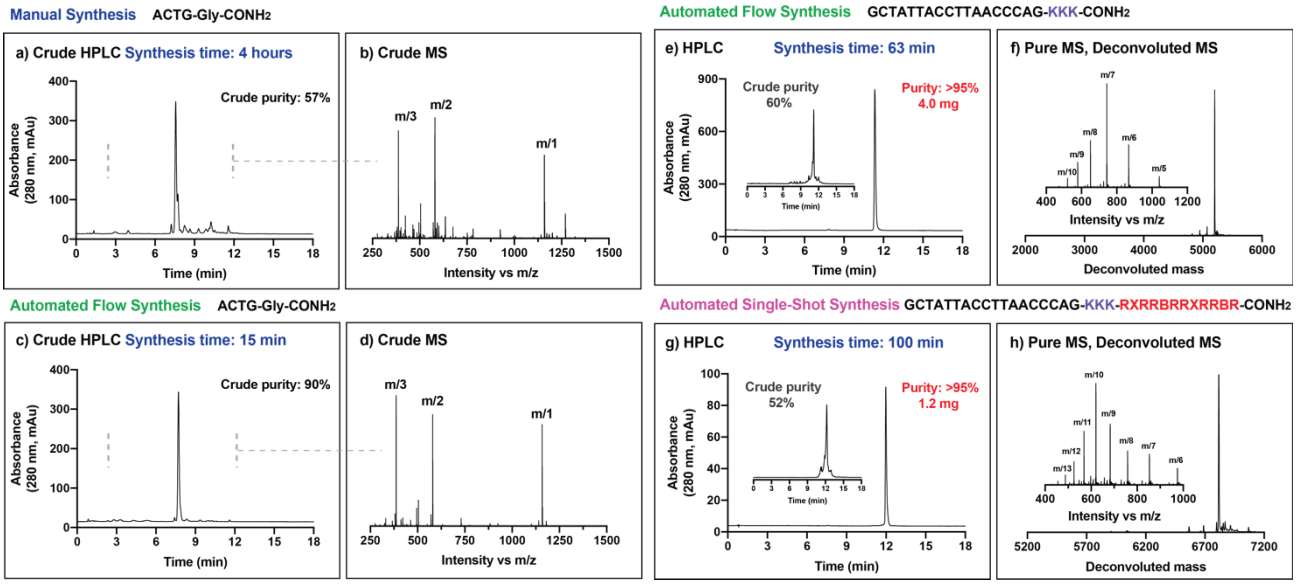


Figure 4.3 Automated fast-flow synthesis of PNAs and PPNAs with enhanced purity over manual protocols.

(a-d) Crude HPLC/LC-MS traces of manually- and flow-synthesized 4-mer PNA. (e-f) Crude/pure HPLC, pure LC-MS traces of flow synthesized anti-IVS2-654 PNA. (g-h) Crude/pure HPLC, pure LC-MS traces of a single-shot synthesized anti-IVS2-654 PPNA.

4.2.4. Synthetic anti-IVS2-654 PPNA Shows Enhanced Activity in an EGFP Assay

The bioactivity of synthesized PPNAs was demonstrated by an enhanced green fluorescent protein (EGFP) assay in HeLa 654 cells (*30*). In this assay, the HeLa cells are stably engineered with an EGFP-coding sequence interrupted by an intron from the human β-globin gene (IVS2-654). The intron contains a cryptic splice site that leads to retention of a β-globin fragment in the EGFP mRNA sequence, resulting in the translation of a non-fluorescent protein. The anti-IVS2-654 PNA hybridizes to the aberrant β-globin 5' splice site, forcing the splicing machinery to use the normal splice sites and producing a functional, fluorescent EGFP. The PNA activity in the nucleus is therefore correlated with EGFP fluorescence and can be analyzed using flow cytometry, reported here as mean fluorescence intensity (MFI) relative to cells treated with phosphate-buffered saline (PBS) as vehicle. We used the synthesis platform for rapid production of an 18-mer anti-IVS2-654 PNA, a scramble 18-mer PNA (negative control), and an anti-IVS2-654 PPNA at milligram scale with >95% purity.

a)

Entry	Name	Sequence (5'-3')	Quantity	Purity
1	Commercial PNA	GCTATTACCTTAACCCAG-KKK-$CONH_2$	0.3 mg	>95%
2	Synthesized PNA	GCTATTACCTTAACCCAG-KKK-$CONH_2$	4.0 mg	>95%
3	Scramble PNA	CCTCATACATACGTACGT-KKK-$CONH_2$	3.3 mg	>95%
4	PPNA (Bpep)	GCTATTACCTTAACCCAG-KKK-Bpep-$CONH_2$	1.2 mg	>95%

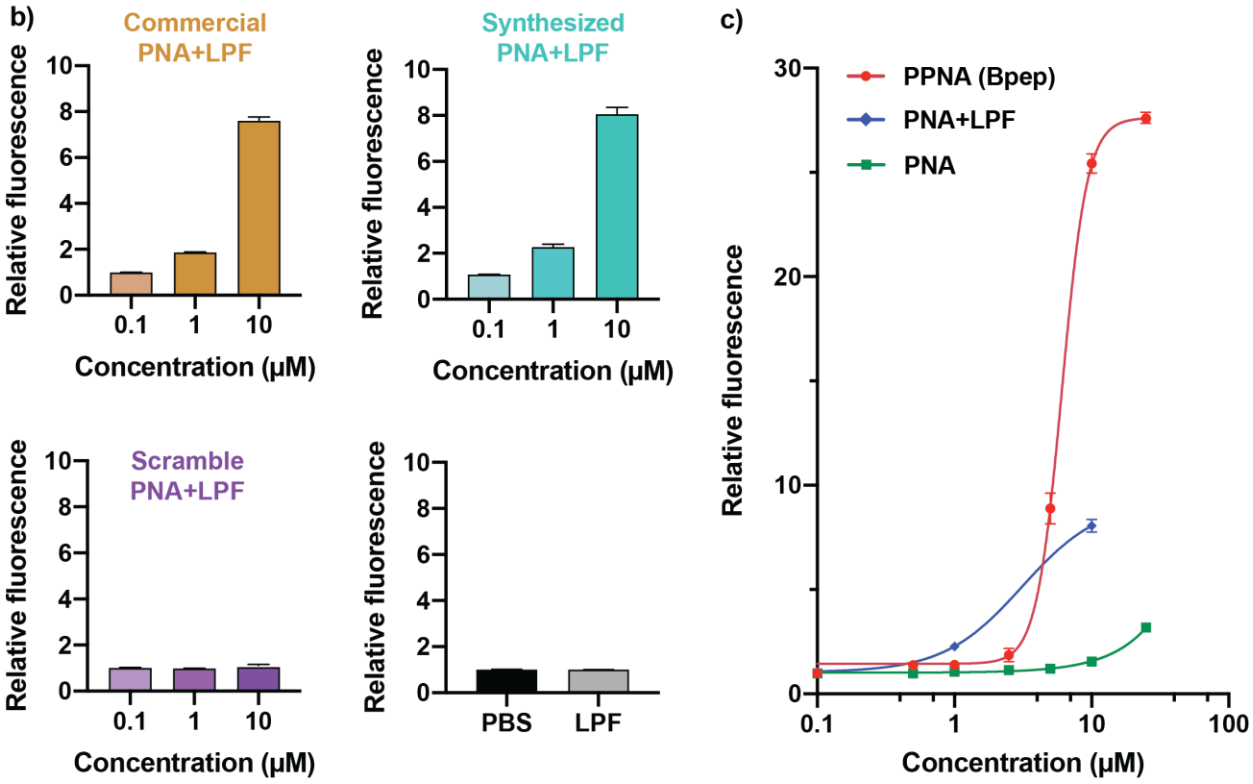


Figure 4.4 Automated flow-synthesized PNAs are active in cell assays.

(a) Names, sequences, quantities, and purities of EGFP PNAs and PPNA. (b) The relative fluorescence (to PBS vehicle-treated cells) of flow-synthesized anti-IVS2-654 PNA and relevant scramble PNA are compared to commercial anti-IVS2-654 PNA, as determined by an EGFP assay in HeLa-654 cells. PBS: Phosphate-Buffered Saline; LPF: Lipofectamine. (c) Dose-response curves corresponding to activity in the EGFP assay for synthesized unmodified PNA, PPNA, and PNA with lipofectamine. Activity is shown as fluorescence intensity relative to PBS vehicle-treated HeLa-654 cells.

A commercial anti-IVS2-654 PNA was obtained from PNA Bio® as a positive control (Figure 4.4a). As depicted in Figure 4.4b, the synthesized PNA and commercial PNA, with delivery mediated by lipofectamine, presented similar activity in the EGFP

assay, validating the function of the flow-produced material. The fluorescence, however, did not increase when cells were treated with the scramble PNA, indicating sequence-dependent activity (Figure 4.4b). Importantly, the single-shot synthesized PPNA improved the activity compared to both unmodified PNA (16-fold) and transfected PNA (3-fold) at 10 μM (Figure 4.4c). This result highlights the importance of cell-penetrating peptides but also demonstrates the potential to quickly build an active PPNA library for further applications.

4.2.5. Synthetic anti-SARS-CoV-2 PPNA Shows Over 95% Viral Inhibition in a Live Infection Assay

The 5’ UTR of the coronavirus genome is responsible for important biological functions, such as viral replication, transcription (*40*) and packaging (*41*). In previous studies, synthetic antisense agents such as PNAs and phosphorodiamidate morpholino oligomers (PMOs) targeting various sites in the 5’ UTR of mouse hepatitis virus (*42, 43*) and SARS-CoV (*17, 44*) demonstrated viral inhibition. Recently, PPMOs were shown to effectively inhibit SARS-CoV-2 replication (*45*). To our knowledge, inhibition of SARS-CoV-2 with PPNAs has not been reported, however.

To further demonstrate the utility of our automated single-shot technology, within one day we synthesized eight long PPNAs (>20mer) targeting the 5’UTR, the transcription regulatory site (TRS), and the polyprotein 1a/b translation start site (AUG) of SARS-CoV-2. After purification, each PPNA was obtained in milligram amounts at >90% purity (Figure 4.5a). These PPNAs were evaluated for inhibition of SARS-CoV-2 replication in Vero-E6 cells. Anti-IVS2-654 PNA, which has no intracellular target present in this assay, was selected as a negative control, while a known inhibitor of SARS-CoV-2 viral growth, EK1 (*46, 47*), was used as a positive control. A multiplicity of infection (M.O.I.) of 0.1 was used for virus inoculation for 2 h. PPNA treatments were given at the same time as viral infection and then continued after virus removal. Subsequently, the viral RNA levels were measured 48 h after initial infection. We observed dose-dependent reduction of the viral RNA replication with increasing concentrations of PPNAs (Figure 4.5b). Sequences designed to target the 5’UTR and AUG region were efficacious. The 5’UTR-3 sequence reduced titers by 75% in the live SARS-CoV-2 assay, which was at the same level as

EK1. Nearly complete inhibition was achieved with sequence 5'UTR-1 at 10 μM (IC50 = 0.8 μM, Figure 4.5c).

a)

Entry	PPNA Name	Sequence (5'→3')	Location	Quantity	Purity
1	anti-IVS2-654	GCTATTACCTTAACCCAG-KKK-Bpep-CONH2	--	1.2 mg	>95%
2	5'UTR-1	TTGGTTGGTTTGTTACCT-KKK-Bpep-CONH2	21-39	1.5 mg	>95%
3	5'UTR-2	AGTTACTCGTGTCCTGTC-KKK-Bpep-CONH2	155-172	1.4 mg	>90%
4	5'UTR-3	CTTACCTTTCGGTCACAC-KKK-Bpep-CONH2	248-265	1.3 mg	>90%
5	TRS-1	AAGTTCGTTTAGAGAACA-KKK-Bpep-CONH2	60-77	0.9 mg	>95%
6	TRS-2	TTTTAAAGTTCGTTTAGA-KKK-Bpep-CONH2	65-82	1.0 mg	>95%
7	TRS-3	AGATTTTAAAGTTCGTTT-KKK-Bpep-CONH2	68-85	1.0 mg	>95%
8	AUG	CTCTCCATCTTACCTTTC-KKK-Bpep-CONH2	254-271	1.5 mg	>90%

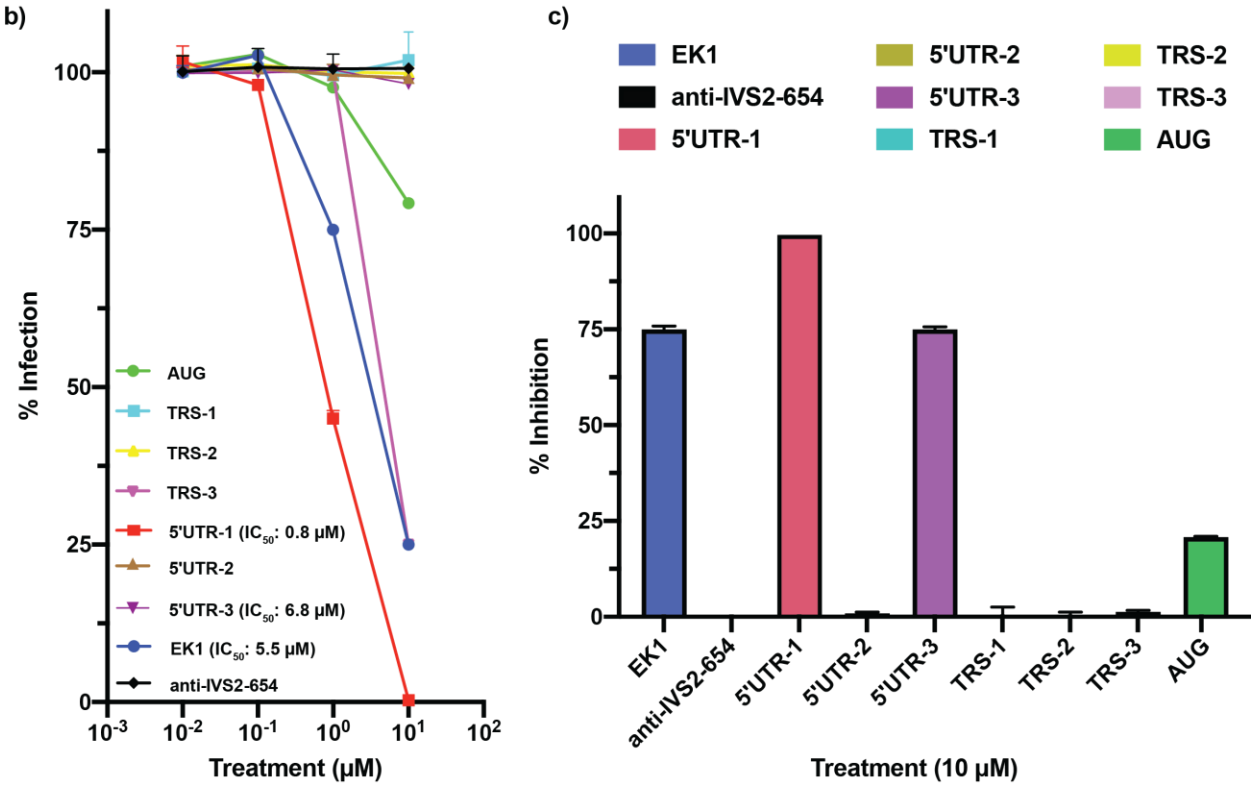


Figure 4.5 Dose-dependent inhibition of live SARS-CoV-2 replication by a synthetic PPNA library

(a) Names, sequences, target location, quantities, and purities of synthetic PPNAs. Locations on the SARS genomic RNA are based on GenBank NC_045512. Bases targeting the SARS-CoV-2 leader-TRS (nt 70-75) are in green; Bases targeting the AUG translation start site are in blue. (b) Increasing concentrations of PPNAs result in inhibition

of viral replication as measured by total viral RNA present. EK1 as a positive control, and anti-IVS2-654 as a negative control. (c) Viral inhibition with PPNAs at 10 μM.

4.3. Discussion

We developed an automated flow technology that provides a route to produce long (>15-mer) PNA-peptide conjugates with high purine content, whereas traditional methods may require several rounds of optimization. Furthermore, this automated synthesis enables access to PNA sequences on an accelerated timescale, more rapid than traditional automated protocols that use commercial microwave peptide synthesizers or the DNA synthesizer Expedite 8909. No additional capping or double couplings are needed even for the assembly of long PNA chains with our approach.

We took advantage of this streamlined flow synthesis to produce eight anti-SARS-CoV-2 peptide-conjugated PNAs in a single day. These novel synthetic PPNAs were tested in a SARS-CoV-2 cell entry assay. A PPNA sequence targeting the 5'UTR of SARS-CoV-2 was found to reduce the viral titer by >95%. Given the favorable toxicological profile of PNA compounds (*12*), further investigations on the therapeutic window of the PPNAs reported here are warranted.

Although PNAs face challenges in translation to the clinic, the introduction of the efficient and robust synthesis technology described here represents a step forward by enabling on-demand rapid production of candidate antisense oligonucleotides, not only for COVID-19 but also for other diseases and emerging pathogens.

4.4. Materials and Methods

4.4.1. Materials

Fluorenylmethyloxycarbonyl (Fmoc) protected PNA monomers, moA, moC, moG, and moT (Figure 4.6) were purchased from PNA bio. Commercial anti-IVS2-654 PNA sequence was purchased from PNA bio. Fmoc-protected amino acids Fmoc-Gly-OH; Fmoc-Lys(Boc)-OH; Fmoc-Arg(Pbf)-OH; 6-aminohexanoic acid and β-alanine were purchased from the Novabiochem-line from Sigma Millipore; *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%), and (7-azabenzotriazol-1-yloxy)tripyrrolidinophospho-nium hexa-fluorophosphate (PyAOP, ≥97.0%) were purchased from P3 Biosystems. Biosynthesis OmniSolv® grade *N*, *N*-dimethylformamide (DMF) was purchased from EMD Millipore (DX1732-1). AldraAmine trapping agents (for 1000 – 4000 mL DMF, catalog number Z511706), Diisopropylethylamine (DIEA; 99.5%, biotech grade, catalog number 387649), piperidine (≥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. H-Rink Amide (0.50 mmol/g loading or 0.18 mmol/g) resin (Figure 4.6) were purchased from PCAS Biomatrix. Dicholormethane (DCM) was obtained anhydrous from Sigma-Aldrich (St. Louis, MO). HPLC-grade Acetonitrile was purchased from VWR International (Philadelphia, PA) and LC-MS grade acetonitrile was purchased from Sigma-Aldrich (St. Louis, MO). Water for HPLC was purified to 18.2 MΩ·cm resistivity on a Millipore Milli-Q system. CytoTox 96® Non-Radioactive Cytotoxicity Assay was purchased from Promega. MEM, fetal bovine serum, and penicillin-streptomycin were purchased from Thermo Fisher Scientific. All other reagents and solvents were purchased from Sigma-Aldrich (St. Louis, MO) as the purest anhydrous grades available, and used without further purification. Unless specified otherwise, all solvents used were kept over activated 3 Å molecular sieves.

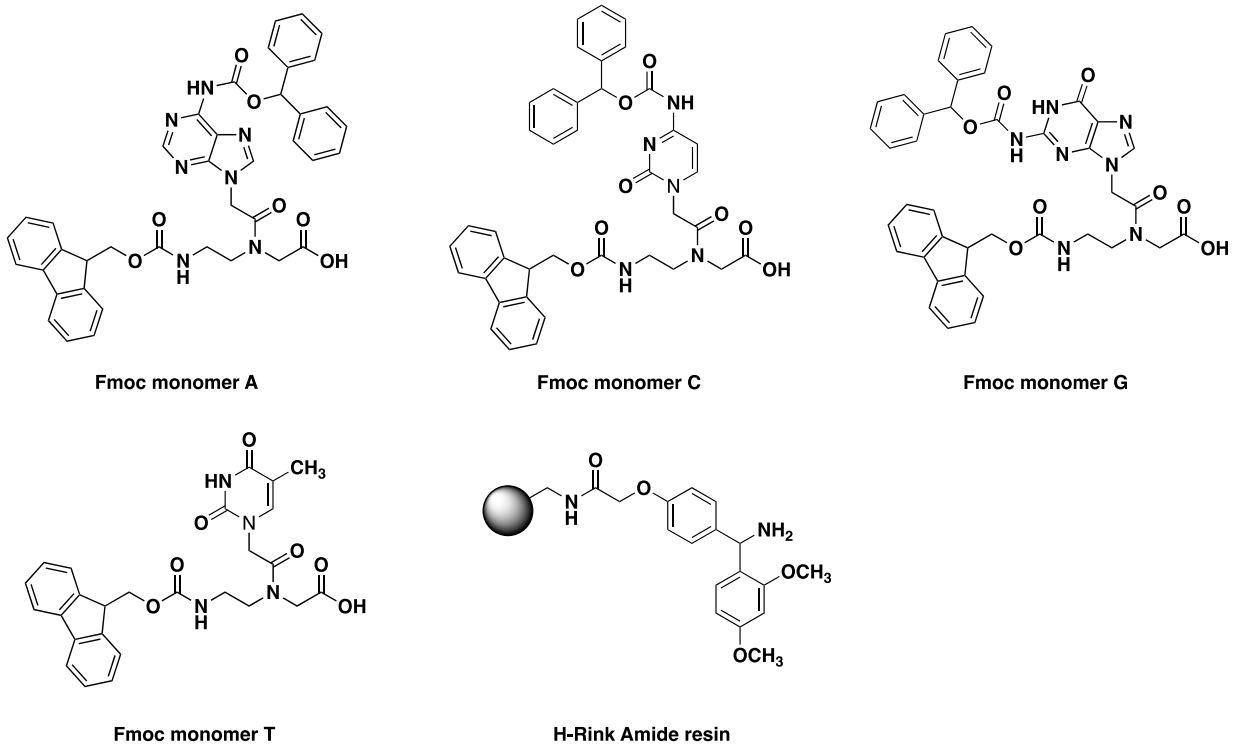


Figure 4.6 Structures of four PNA monomers and the H-Rink Amide resin.

4.4.2. Liquid chromatography–mass spectrometry (LCMS) analysis

All PNA solutions were filtered and then diluted to approximately 0.1 mg/mL. LC-MS chromatograms and associated high-resolution mass spectra were acquired using an Agilent 6550 iFunnel Q-TOF LCMS system (abbreviated as 6550) or an Agilent 6545 Accurate-Mass Q-TOF LCMS system (abbreviated as 6545). Solvent compositions used in the LC-MS are 0.1% formic acid in H_2O (solvent A) and 0.1% formic acid in acetonitrile (solvent B). The following LC-MS methods were used:

Condition 1: 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 (2GHz), and low mass range (m/z in range 100 to 1700). The solvent mixtures used were as above. Column: Phenomenex Luna C18 (3 um, 1 x 150 mm, 100 Å silica); Flow Rate: 0.05 mL/min; Gradient: 1% B 0-2 min, linearly ramp from 1% B to 61% B 2 to 14 min, hold 90% B from 14.1 to 18 min. Post time is 1% B for 5 min. MS data was acquired from 4 to 14 min.

Condition 2: Agilent 6545 (1290 Infinity HPLC system with Dual Jet Stream ESI ionization followed by Accurate-Mass Q-TOF MS. MS is run in positive ionization mode, extended dynamic range (2GHz), and standard mass range, m/z in range 100 to 3000). The solvent mixtures used were as above. Column: Agilent Zorbax 300SB-C18 (5 um, 2.1 x 150 mm, 100 Å silica); Flow Rate: 0.6 mL/min; Gradient: 1% B 0-2 min, linearly ramp from 1% B to 61% B 2 to 11 min, linearly ramp from 61% B to 95% B 11 to 12 min. Post time is 1% B for 3 min. MS data was acquired from 4 to 11 min.

Data were processed using Agilent MassHunter Workstation Qualitative Analysis Version B.06.00 with BioConfirm software, and plotted with the Prism 8.0 software.

4.4.3. High-performance liquid chromatography (HPLC) analysis

Analytical HPLC was carried out on an Agilent 1200 series system with UV detection at 280 nm. Column: Phenomenex Kinetex, (100 x 2.1 mm, 2.6 um, 100 Å silica); flow rate 0.375 mL/minute; Solvent System: A linear gradient of acetonitrile with 0.08% TFA additive (solvent B) in water with 0.1% TFA additive (solvent A) was used. 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 method lasted 25 min in total. Crude HPLC purities were determined by manual integration of all signals in the area of 2 to 13 min.

4.4.4. Procedure for manual PNA synthesis

1) To a 20-mL disposable reaction vessel (Torviq), add 100 mg of Rink Amide resin (0.18 mmol/g) and swell the resin in amine free DMF for 1 hour1.
2) Activate Glycine amino acid (3 eq) to be added by adding (7-Azabenzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate (PyAOP, 2.9 eq), DMF, and N,N-diisopropylethylamine (DIEA, 3 eq).
3) Transfer the activated monomer solution to the reaction vessel.
4) The coupling steps were carried out at room temperature for 30 min.
5) Vacuum off coupling solution from the resin.
6) Wash resin six times with 10 mL of 50:50 DMF/DCM.
7) Close and bleed vacuum line.

8) Add 6 mL of 20% piperidine solution in DMF to begin the deprotection. Stir for 5 min.

9) Vacuum off coupling solution from the resin.

10) Wash resin three times with 10 mL of 50:50 DMF/DCM.

11) Close and bleed vacuum line.

12) Repeat steps 8 to 11 for one more time.

13) Activate Fmoc-protected PNA monomer (6 eq) to be added by adding PyAOP (5.8 eq), DMF, DIEA (6 eq), and 2,5-lutidine (6 eq).

14) The coupling steps were carried out at room temperature for 30 min.

15) Vacuum off coupling solution from the resin.

16) Wash resin six times with 10 mL of 50:50 DMF/DCM.

17) Close and bleed vacuum line.

18) Add 6 mL of 20% piperidine solution in DMF to begin the deprotection. Stir for 5 min.

19) Vacuum off coupling solution from the resin.

20) Wash resin three times with 10 mL of 50:50 DMF/DCM.

21) Close and bleed vacuum line.

22) Repeat steps 18 to 21 for one more time.

23) Repeat steps 13 to 22 for the remaining PNA monomer additions.

24) Wash six times with 1 mL methanol.

25) Lyophilize overnight.

4.4.5. Procedure for automated 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. The synthesis order is characterized in Figure 4.7. For each chain elongation cycle, steps 1-6 were repeated.

For single-shot cell-penetrating peptides (CPP) conjugated PNAs (PPNAs) synthesis, CPP was pre-synthesized on the H-Rink Amide resin (0.5 mmol/g) with a peptide synthesizer developed in our lab (*4*). The CPP bound resin was used directly for automated PPNA synthesis on Tiny Tides. Both $(Lys)_3$ linker and PNA sequences were synthesized directly on Tiny Tides.

Step	Description	Monomer Pump	Reagent Pump	DIEA Pump	Temp.	Time	Flow Rate
1	Wash	DMF	DMF	--	40 ˚C	20 sec	2.5 mL/min x 2 pumps
2	Wash	DMF	DMF	--	70 ˚C	40 sec	2.5 mL/min x 2 pumps
3	Coupling	0.2 M monomer (10 strokes, 10 eq) in DMF	0.18 M HBTU (10 strokes, 9.6 eq) in DMF	10% v/v DIEA (10 strokes, 30 eq) in DMF	70 ˚C	10 sec	2.5 mL/min x 3 pumps

4	Wash	DMF	DMF	--	70 ˚C	20 sec	2.5 mL/min x 2 pumps
5	Wash	DMF	DMF	--	40 ˚C	40 sec	2.5 mL/min x 2 pumps
6	Deprotection	DMF	40% Piperidine, 2% Formic acid in DMF	--	40 ˚C	50 sec	2.5 mL/min x 2 pumps

Table 19 Summary of coupling and deprotection steps performed during automated flow PNA synthesis.

4.4.6. PNA 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 filtrated with a Nylon 0.22 µm syringe filter, frozen with liquid nitrogen, and lyophilized to dried powder. Finally, the crude PNA was weighed.

4.4.7. PNA purification protocols

Method 1: HPLC purification was carried out on an Agilent 1200 series system with UV detection at 260 nm. Column: Phenomenex Aeris PEPTIDE XB C18, (250 x 2.1 mm, 2.6 um, 100 Å silica); Flow rate 1.0 mL/minute; Solvent system: 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. 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.

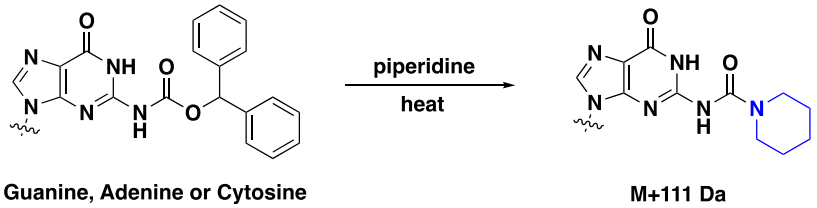
Method 2: HPLC purification was carried out on an Agilent 1200 series system with UV detection at 260 nm. Column: Phenomenex Kinetex C18, (100 x 2.1 mm, 2.6 um, 100 Å silica); Flow rate 0.375 mL/minute; Solvent system: A linear gradient of acetonitrile with a 0.08% TFA additive (solvent B) in water with a 0.1% TFA additive (solvent A) was used. 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 total duration of the method was 125 min. Fractions were collected every minute.

Note: **Method 1** was used for peptide nucleic acid (PNA) purification. Most of the cell-penetrating peptide conjugated PNAs (PPNAs) suffer from the challenging separation of target product and single residue C deletion side-product after purification with **method 1**. A second purification with **method 2** can help remove most of the C-deletion side-product and improve the product purity.

4.4.8. Side reactions during Fmoc PNA synthesis

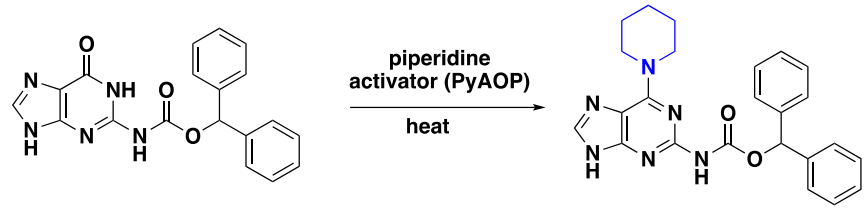
The common side reactions during high-temperature PNA synthesis are listed in Figure S3. The acid-labile benzhydryloxycarbonyl (Bhoc) protecting group is commonly used to protect nucleobase exocyclic amino groups for Fmoc PNA synthesis. At elevated temperature, it can be cleaved by piperidine to yield a byproduct with an apparent +111 Da. mass (Figure 4.7a top). If trace activator is left on the resin bed before deprotection, the amination of nucleobases could be accomplished via the presence of activator and piperidine (*48*), and yield a byproduct with an apparent +67 Da. mass (Figure 4.7a bottom). We have found that a longer wash after coupling can help solve this problem. Backbiting of the newly deprotected amino group can lead to formation of a ketopiperazine that gets washed away, causing deletion of the N-terminal PNA unit (Figure 4.7b). Base and heat can promote transamidation of the nucleobase to the N-terminus, resulting in an isomeric, secondary amine-containing product (Figure 4.7c).

a) Base adducts



Guanine, Adenine or Cytosine

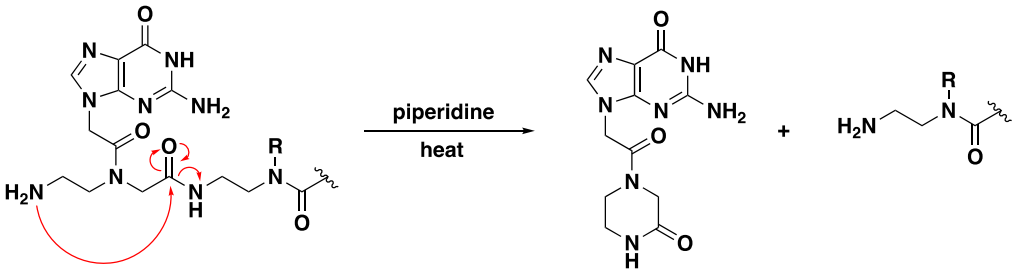
M+111 Da



Guanine or Thymine

M+67 Da

b) N-terminus deletion



c) Transamidation

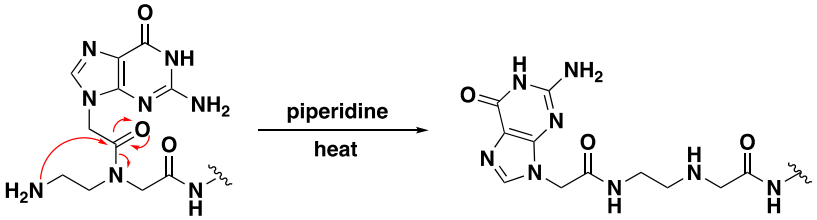


Figure 4.7 Side reactions during Fmoc PNA synthesis at high temperature.

4.4.9. Design of the fully automated PNA synthesizer

The overall design of the automated synthesizer Tiny Tides (*4*) is shown below in Figure 4.8.

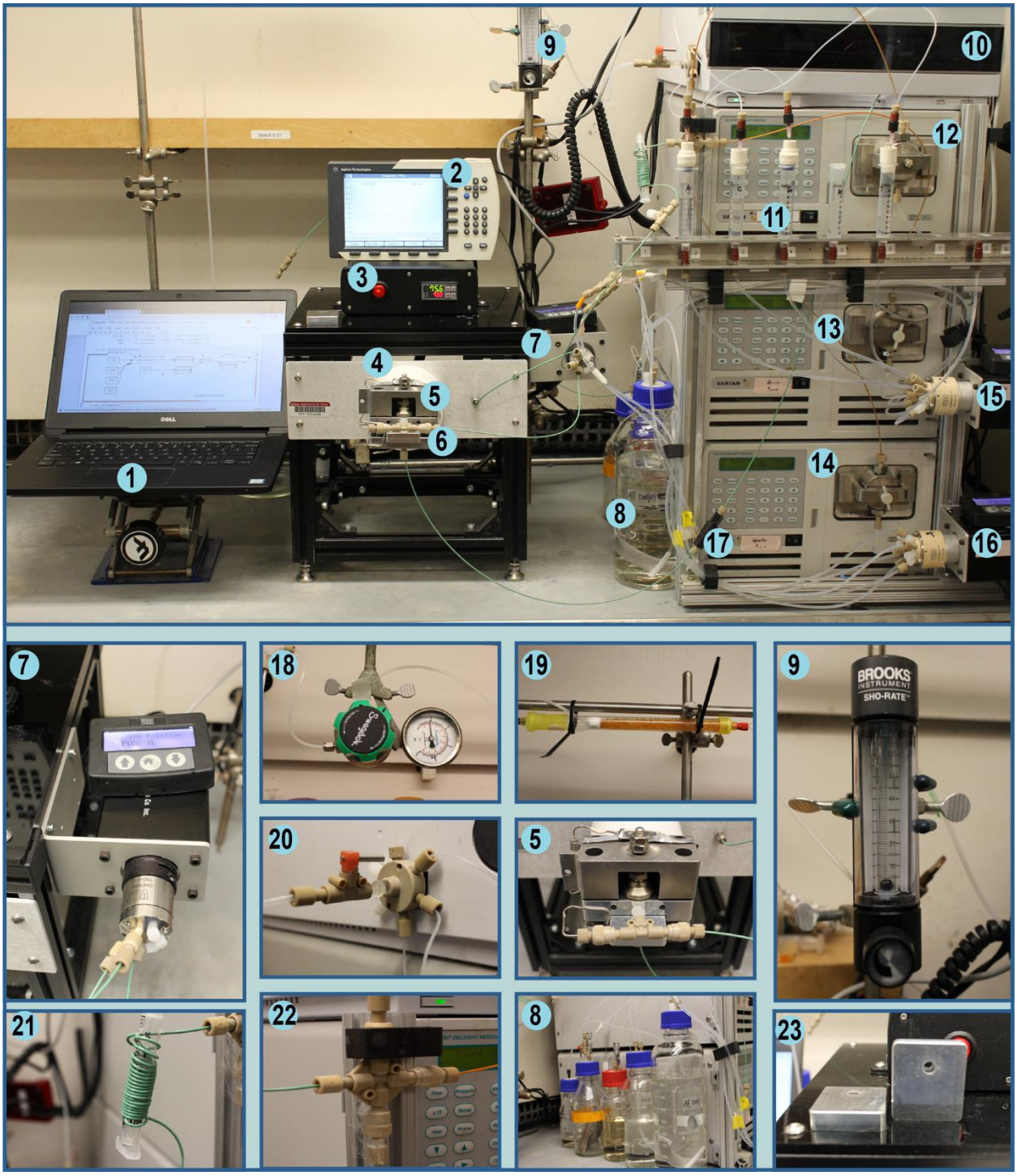


Figure 4.8 Overview and some major components of the automated PNA synthesizer.

Individual components are labeled with numbers.

(1) Computer: The instrument is controlled with an in-house Python script[4]. The programming code is available on an open-source GitHub repository.

(2) UV Detector: Agilent 1260 UV detector (Part #: 1260DAD) fitted with a micro flow cell 6 mm path length (Part #:01200-90130 Rev. B). Effluent was monitored at 280 nm.

(3) Temperature Controller: A Misumi temperature controller (Part #: MTMNRD) was used to drive power for an OMEGA 120V 300W heater cartridge (Part#: HDC00059).

(4) Heating Coil: 3 feet of 316L stainless steel tubing (0.064" OD x 0.020" ID) wrapped around a heated aluminum core.

(5) Reactor: Custom aluminum and stainless-steel reactor head. A screw-handle on top holds a cartridge containing resin in the solvent flow-path[4].

(6) T-Connector: A T-connector (IDEX Part #: P713) connected with two check valves was used to deliver solutions from different temperature loops.

(7) 6-Position valves: VICI brand EUHA body fitted with a 6-position Chem-Inert model 150-0376H valve face.

(8) Solvent and reagent reservoirs: Deprotection, neutralization, coupling base solutions and DMF was stored in Schott brand media bottles. The bottles were fitted with machined polypropylene caps to supply nitrogen and maintain positive nitrogen head pressure. Activated 3 Å molecular sieves were added into the DMF bottle.

(9) Flow meter: A nitrogen flow meter (BROOKS Model #1250AR6071MRSVV) is placed before the monomer solutions but not the reagent lines to prevent excess nitrogen leakage from the monomer containers.

(10) UV monitor: The Agilent 1200 infinity instant pilot controller gives complete control, system monitoring, and signal plotting.

(11) Monomer reservoirs: Monomer solutions were stored in 5 mL polypropylene syringes under positive pressure of dry nitrogen supplied through custom polypropylene caps.

(12) DIEA HPLC Pump: Agilent 210 series HPLC pump fitted with a 5 mL SS pump head. One stroke delivers 40 µL of solution.

(13) Monomer HPLC Pump: Agilent 210 series HPLC pump fitted with a 5 mL SS pump head. One stroke delivers 40 µL of solution.

(14) Activator HPLC Pump: Agilent 210 series HPLC pump fitted with a 5 mL SS pump head. One stroke delivers 40 µL of solution.

(15) 10-Position valve-1 : VICI brand EUHA body fitted with a 10-position Chem-Inert model 16P-0713L valve face. This valve was used to switch DMF and monomer solutions.

(16) 10-Position valve-2: VICI brand EUHA body fitted with a 10-position Chem-Inert model 16P-0713L valve face. This valve was used to switch DMF, deprotection solution, and activator solution.

(17) Back pressure regulator: 20 psi back pressure regulator (IDEX, P-791).

(18) Pressure regulator: An in-line pressure regulator (SWAGELOK, 15 psi) allows for precise back pressure control over the entire system.

(19) Silica gel desiccant column: Nitrogen from the hood manifold is plumbed through an activated silica gel desiccant column.

(20) Low-pressure manifold assembly: A 6-port low-pressure manifold assembly (IDEX P152) was used to transfer nitrogen to different monomer reservoirs.

(21) PEEK loop mixer: A spiral peek loop was used to help further mix the coupling reagent.

(22) Cross mixer: The streams from the three pumps meet from three directions in a PEEK cross-mixer (IDEX P-723-01).

(23) Reusable reactor: The reusable reactor was designed to accommodate a 7.5 micromole synthesis (~15 mg 0.5 mmol/g resin in ~600 µL). This smaller reactor saved expensive reagents, while still producing enough PNA for initial biological studies.

Tubing: Tubing from the pump to the mixer T is PEEK 1/16” OD x 0.02” ID tubing (McMaster Part#: 51085K45). Tubing from the Reagents to the valves is PFA Nat 1/16” OD x 0.030” ID tubing (IDEX Part #: 1514L). All other tubing is PEEK 1/16” OD x 0.030” ID tubing (McMaster Part#: 51085K48).

Fittings: Unless otherwise specified, all fittings were IDEX brand Super Flangeless™ PEEK fittings.

4.4.10. EGFP assay and cytotoxicity studies

4.4.10.1. Flow cytometry assay for PNA and PPNA constructs

HeLa-654 cells were maintained in MEM supplemented with 10% (v/v) fetal bovine serum and 1% (v/v) Pen Strep at 37 °C and 5% CO_2. Twenty-four hours before treatment,

HeLa cells were plated at a density of 5,000 cells per well in a 96-well plate. 1mM stocks of the compounds for treatment were made in cation-free PBS and concentration was quantified by absorbance at 260 nm using a UV/Vis spectrophotometer. The extinction coefficient of $\varepsilon_{260} = 175{,}000\ \mathrm{M^{-1}cm^{-1}}$ was provided by the manufacturers of the commercial PNA sample.

For the PPNA conjugate and the corresponding PNA control, each compound was diluted to the target concentration (0.1, 0,5 1, 2.5, 5, 10 and 25 µM) in MEM supplemented with 10% (v/v) fetal bovine serum and 1% (v/v) Pen/Strep. To treat the cells, the overnight growth media was aspirated from each well and 100 µL of a given diluted compound was applied to each well.

For the transfected PNA constructs, transfections of negative controls were prepared with Lipofectamine 3000 reagent according to the manufacturer instructions. Briefly, PNA stocks were diluted into 30 µL un-supplemented MEM and combined with 0.6 µL of the P3000™ Reagent. After a 5-minute incubation at room temperature, 0.75 µL of the Lipofectamine® 3000 Reagent was added, and the mixture was subsequently diluted to the target PNA concentrations (0.1, 1, 10 µM) in MEM supplemented with 10% (v/v) fetal bovine serum and 1% (v/v) Pen Strep. The un-transfected controls that corresponded to these experiments were prepared in the same process, without the addition of lipofectamine reagents (Figure 4.9).

Cells were incubated for 22 hours at 37 °C and 5% CO_2 with the peptide treatment, and then the treatment media was removed and transferred to a separate 96-well plate for the LDH cytotoxicity assay. 30 µL Trypsin EDTA 0.25 % (150 µL) was added to the cells and incubated for 10 min at 37 °C and 5% CO_2. To quench the trypsin, 70 µL of MEM supplemented with 10% (v/v) fetal bovine serum and 1% (v/v) Pen Strep was added to each well. The dissociated cells in media were transferred to 96-well V-bottom plate and spun at 500 rcf for 3 min. The supernatant was removed and the pellets were each washed with 100 µL of phosphate-buffered saline (PBS), and the plate was spun again. The supernatant was again removed and the pellets were resuspended in 300 µL PBS with 2% FBS (v/v) and 2 µg/mL propidium iodide in water. Flow cytometry analysis was carried out on a BD LSRII flow cytometer. Gates were applied to the data to ensure that only data from healthy, living cells were taken into the account. Cells that were highly

positive for propidium iodide or had forward/side scatter readings that were sufficiently different from the main cell population were excluded. Each histogram contains at least 5,000 gated events.

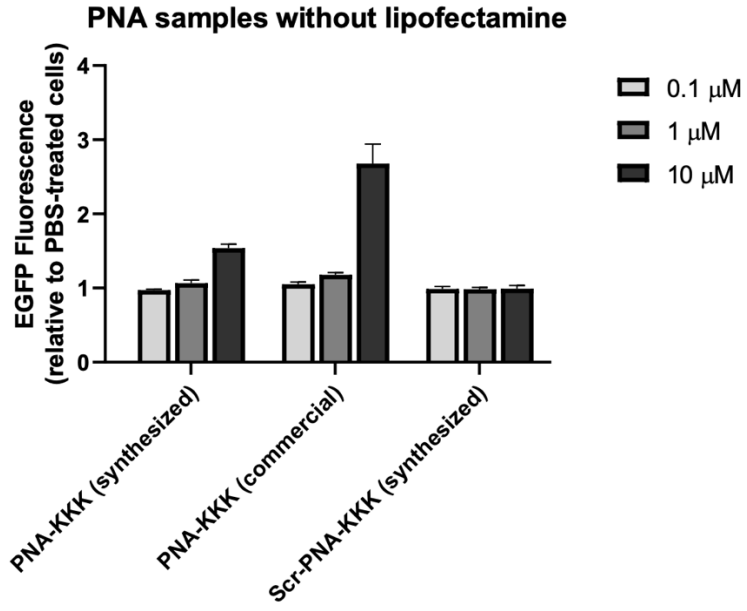
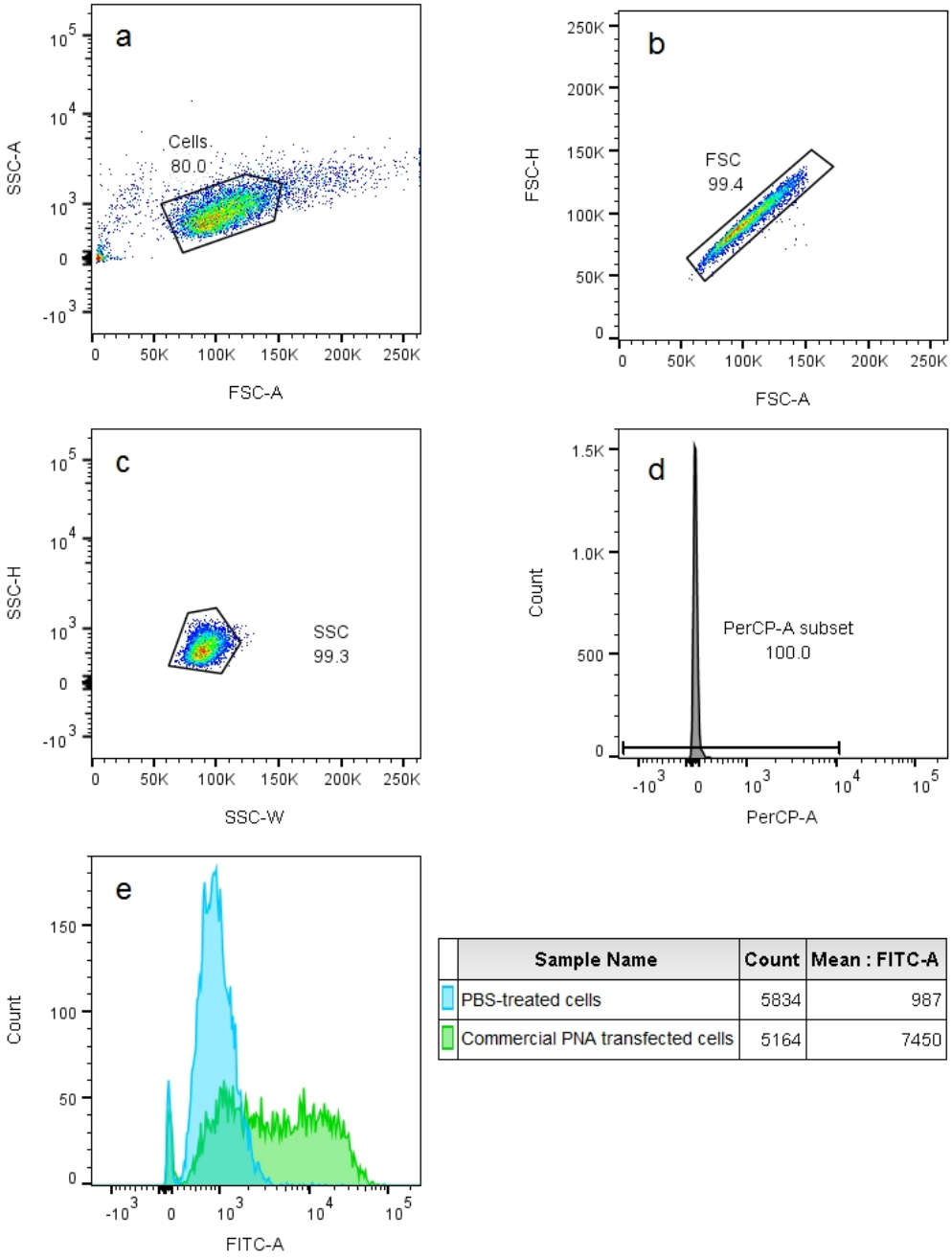


Figure 4.9 Cellular uptake of PNA compounds without transfection

The automated synthesized anti-IVS2-654 PNA and relevant scramble PNA are compared to commercial PNA for their cell uptake and exon-skipping activity without Lipofectamine transfection. Cell entry is determined by relative fluorescence (to PBS-treated cells) in the EGFP assay in HeLa-654 cells, and error bars represent three technical replicates. Without transfection, the anti-IVS2-654 PNA compounds show significant ($p<0.0001$) cell entry at 10 μM. The commercial PNA sample does show higher cell entry than the synthesized anti-IVS2-654 PNA ($p<0.05$). However, given the equivalent exon-skipping activity shown in the assays with transfected commercial and synthesized anti-IVS2-654 (main text. Figure 4.4b), the difference in activity shown here likely represents increased cell penetration of the commercial sample rather than a difference in DNA hybridization or exon-skipping activity. The increased LDH release from the commercial sample (significantly greater than the synthesized anti-IVS2-654 PNA, $p<0.05$. Figure 4.6a) suggests that increased membrane permeability correlates with this difference in uptake.

4.4.10.2.Flow cytometry gating



	Sample Name	Count	Mean : FITC-A
	PBS-treated cells	5834	987
	Commercial PNA transfected cells	5164	7450

Figure 4.10 Example of the gating strategy for flow cytometry.

HeLa 654 cells were gated using a (a) forward versus side scatter log area plot to remove debris, followed by a (b) forward linear height by forward linear area to remove doublets and a (c) side linear height by side linear width. Cells positive for propidum iodide were excluded from the final population (d). Representative histograms (e) of the eGFP signal in PBS-treated cells (blue) and cells transfected with 10 μM of the commercial PNA sample (green) are shown.

4.4.10.3. LDH cytotoxicity assay

HeLa-654 cells were maintained in MEM supplemented with 10% (v/v) fetal bovine serum and 1% (v/v) penicillin-streptomycin at 37 °C and 5% CO_2. The cells were treated with PNA or PPNA compounds as previously described in their preparation for Flow Cytometry. Before cells were trypsinized for Flow Cytometry analysis, 50 µL of the supernatant treatment media was transferred to another clear-bottom 96-well plate for the LDH assay. The assay was performed using the CytoTox 96® Non-Radioactive Cytotoxicity Assay (Promega) according to the included technical bulletin and using the provided solutions. The absorbance was measured on a BioTek Epoch Microplate Spectrophotometer at 490 nm. The positive control is a cell lysis with the supplied lysis buffer and the negative control is cells treated with a vehicle of 2% PBS in MEM supplemented with 10% (v/v) fetal bovine serum and 1% (v/v) penicillin-streptomycin. The data are worked up by subtracting the absorbance of untreated cells from all of the treatment conditions, including the cell lysis, and then dividing by the corrected lysis value. A value of 0 reflects no additional LDH release over cell baseline and a value of 100% reflects total LDH release from the cytosol into the treatment media (Figure S11 and S12).

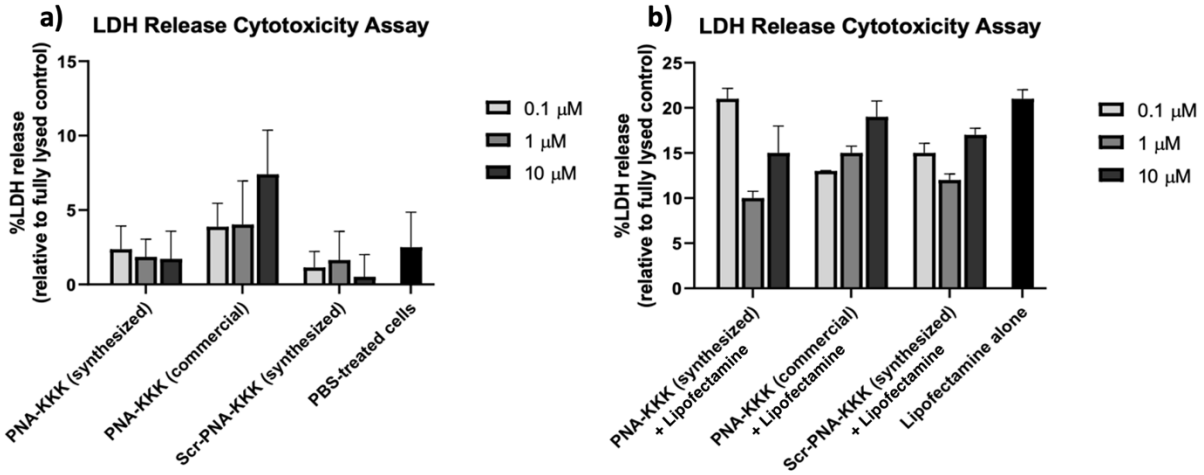


Figure 4.11 Cytotoxicity of PNA Compounds

HeLa-654 cells were treated with each construct at 0.1, 1, or 10 µM for twenty-two hours, with (a) or without (b) the addition of Lipofectamine® 3000 Transfection Reagent. The y-axis is the ratio of LDH release compared with the cell lysis control, and error bars represent three technical replicates. Without transfection, no compound elicited LDH release from the membrane significantly more than vehicle-treated cells. The transfected

PNA compounds did not show significant LDH release above the cells treated with only Lipofectamine® 3000 Transfection Reagent.

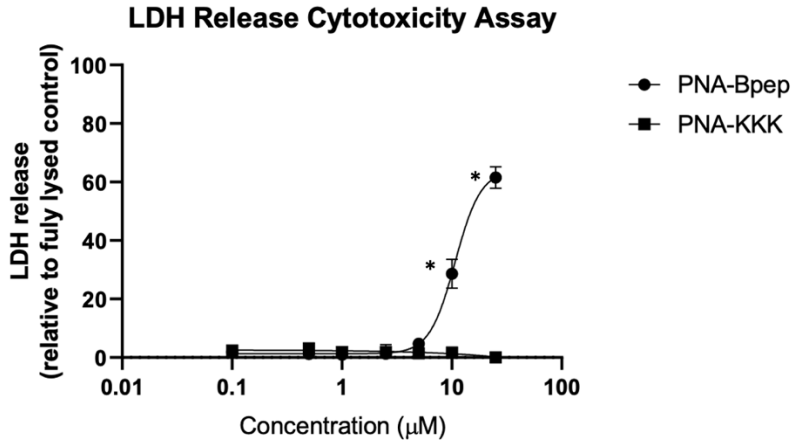


Figure 4.12 Cytotoxicity of PPNA compared to synthesized anti-IVS2-654 PNA

HeLa-654 cells were treated with each construct at 0.1, 0.5, 1, 2.5, 5, 10, or 25 μM for twenty-two hours. The y-axis is the ratio of LDH release compared with the cell lysis control. At 10 and 25 μM, the PPNA construct demonstrated significant LDH release ($p < 0.001$) compared to the PNA alone, suggesting cytotoxicity at these higher doses. Although the LDH leakage suggests membrane disruption at higher PPNA doses, this is unlikely to be responsible for the compound's cell penetration. The PPNA conjugate begins to show cell penetration and exon-skipping activity at 5μM (Figure 4.4c), before significant LDH leakage occurs.

4.4.11. SARS-CoV-2 inhibition assays

All live SARS-COV-2 viral work was performed at the Iowa State University College of Veterinary Medicine's BSL-3 facility under approval by the institutional biosafety committee.

SARS-CoV-2 (Seattle Strain) was obtained from Beiresources (ATCC) and expanded in Vero E6 cells (ATCC), quantified by qRT-PCR (IDT) and frozen at -80 ˚C in 1 mL aliquots. Vero cells were then plated for 90% confluency in 96 well flat bottom plates a day in DMEM 10% FBS 1% Pen/Strep (Gibco). PNAs and EK1 (*46, 47*) used in this study were diluted in final varying concentrations as indicated (0-10 μM) in triplicate. Virus was added to PNAs or EK1 at multiplicity of infection (MOI) 0.1 and incubated for 2 hours on the cells. Cells were then rinsed and media replaced with fresh DMEM 10% FBS 1% Pen/Strep containing similar PNA concentrations. Supernatants were harvested 2 days

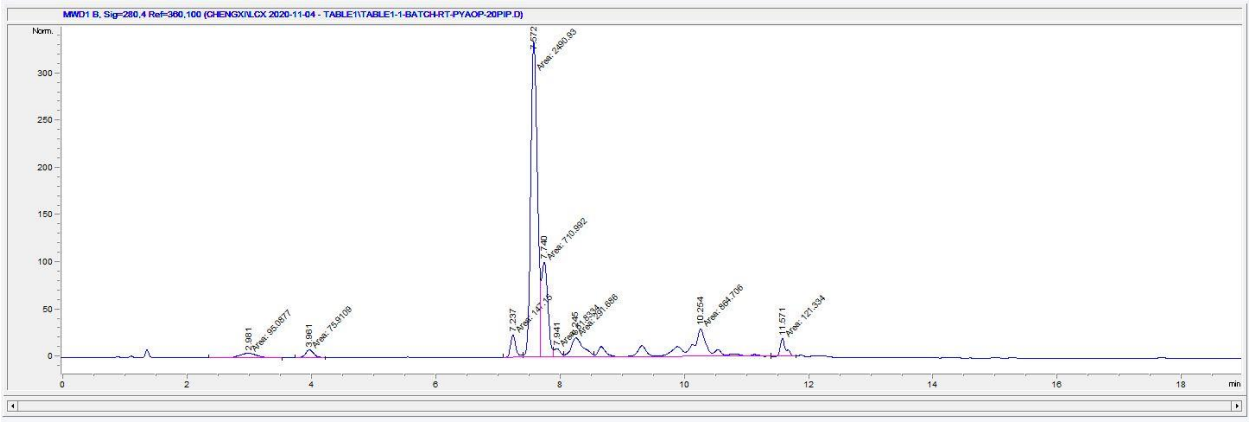
post-infection, mixed 1:1 with Trizol reagent (Invitrogen), and the RNA was isolated according to the manufacturer protocols. Viral RNA was amplified on a Quant Studio 3 (Applied Biosystems) using a Probe based LUNA qRT-PCR detection system (NEB Bio) and commercial SARS-CoV-2 N detection primer (IDT) according to manufacturer directions followed by verification of neutralization using a second set of SAR-CoV-2 F detection primers.

4.4.12. HPLC traces of 4-mer PNAs for condition optimization

A 4-mer PNA "ACTG-Gly-$CONH_2$" was synthesized manually in batch with the conditions in Section 4.5.4 as a control. Ten 4-mer PNAs with same sequences were produced to optimize PNA synthesis in flow. After synthesis, five milligrams of each resin-bound PNA was washed with DCM and dried under vacuum. After cleavage, each crude sample was analyzed with HPLC using the methods in Section 4.5.7.

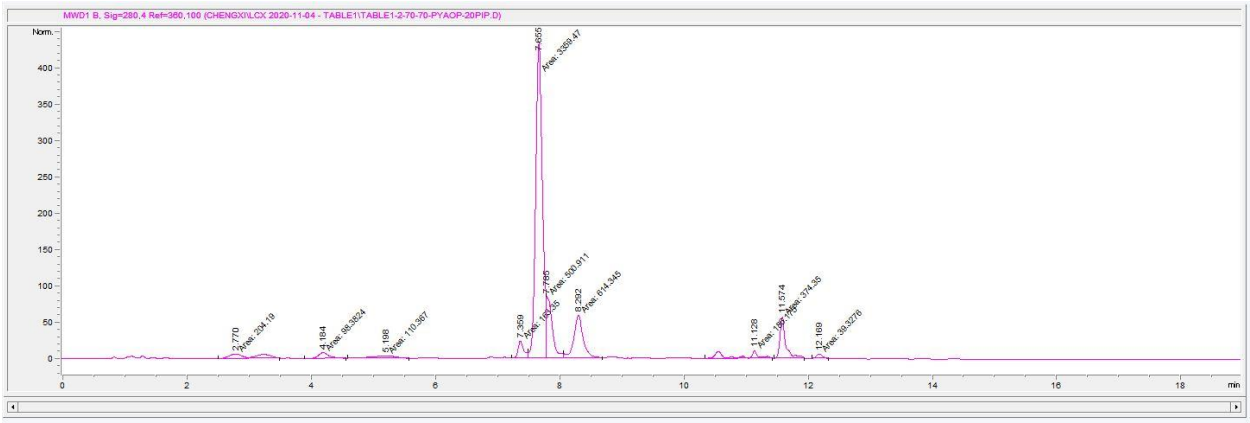
Crude purity of each PNA sample was determined through integration of the chromatogram from 2 minutes to 14 minutes, which covers the desired product and structurally related side-products. The attribution of the peaks can be summarized as follows: 2-6 min: A, C, T, G deletions; 7.6 min: target product; 7.7 min: isomer; 8.2 min: an analogue which was generated during analysis (amide group at C-terminus changed to carboxylic group, R-$CONH_2$ to R-COOH, +1 Da. mass), which was summed together with the product. 8.5-11 min: peaks in this area were summered as unknown side products. 11.1 min: base adduct (Figure 4.7a bottom), 11.5: base adducts (Figure 4.7a top).

Table 18, Entry 1. Sequence: ACTG-Gly-$CONH_2$. Manual synthesis.



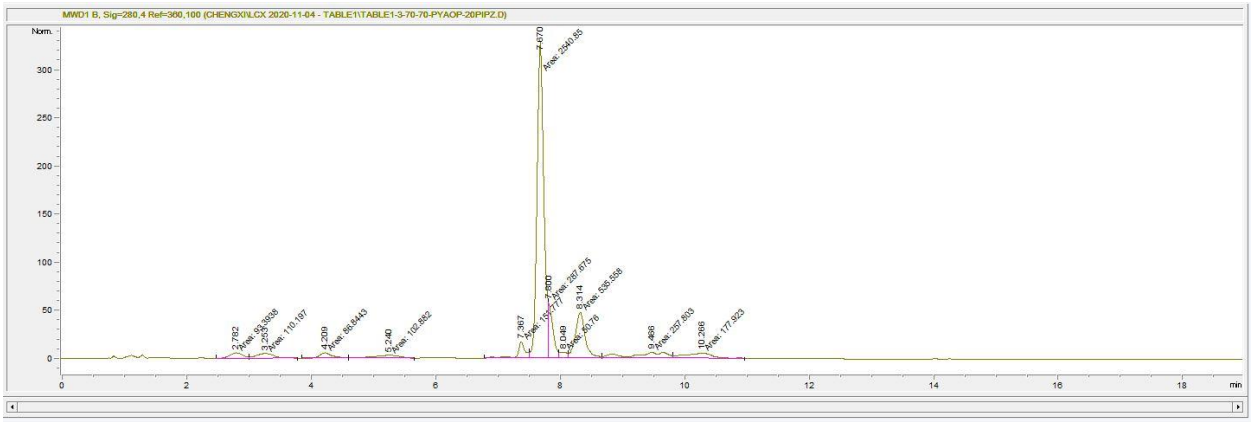
#	Time	Area	Height	Width	Area%	Symmetry
1	2.981	95.1	4.7	0.3398	1.957	1.198
2	3.961	75.9	8.2	0.1543	1.562	0.908
3	7.237	147.1	24	0.1022	3.028	0
4	7.572	2490.9	335.1	0.1239	51.258	0.857
5	7.74	711	100.8	0.1176	14.631	0.712
6	7.941	61.8	9.3	0.1105	1.272	0.554
7	8.245	291.7	20.6	0.2366	6.002	0.587
8	10.254	864.7	29.4	0.4899	17.794	2.023
9	11.571	121.3	18.9	0.1071	2.497	0.579

Table 18, Entry 2. Sequence: ACTG-Gly-$CONH_2$. Automated flow synthesis.



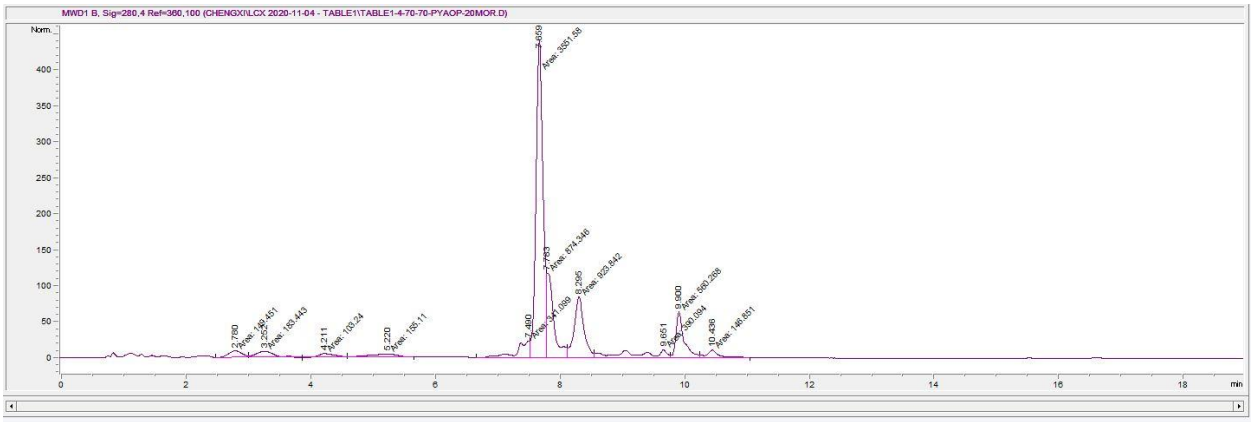
#	Time	Area	Height	Width	Area%	Symmetry
1	2.77	204.2	6.7	0.5063	3.613	0.327
2	4.184	98.4	8	0.2048	1.741	0.771
3	5.198	110.4	3.5	0.5258	1.953	1.517
4	7.359	163.4	23.9	0.1139	2.891	0
5	7.655	3359.5	433.6	0.1291	59.451	0.827
6	7.785	500.9	84	0.0716	8.864	0
7	8.292	614.3	58.7	0.1744	10.872	0.908
8	11.128	186.2	10.6	0.2938	3.295	2.679
9	11.574	374.4	57.1	0.1093	6.625	0.599
10	12.169	39.3	6.2	0.1061	0.696	0.798

Table 18, Entry 3. Sequence: ACTG-Gly-$CONH_2$. Automated flow synthesis.



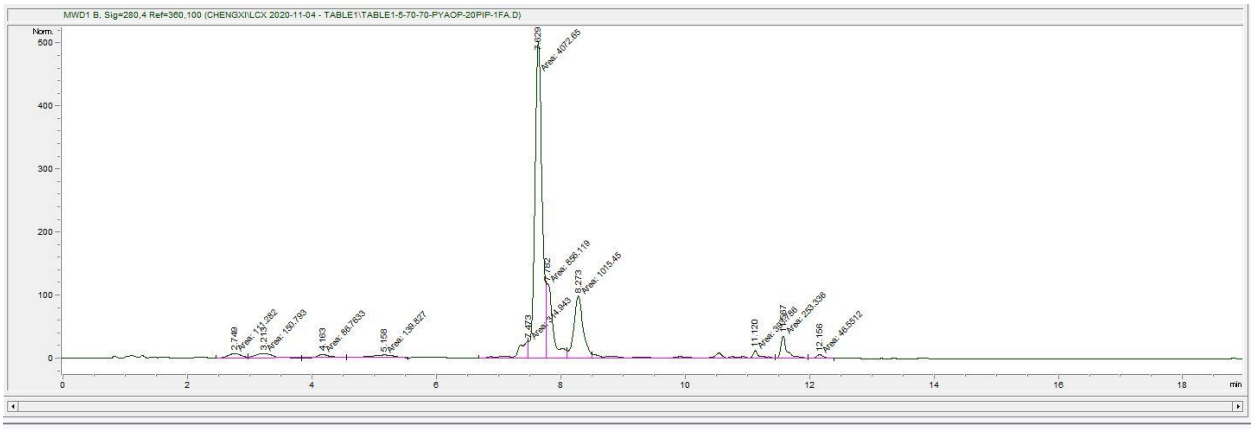
#	Time	Area	Height	Width	Area%	Symmetry
1	2.782	93.4	6	0.2579	2.125	0.948
2	3.253	110.2	5.2	0.3508	2.507	0.923
3	4.209	86.8	5.8	0.2485	1.976	0
4	5.24	102.9	3.1	0.5461	2.341	1.584
5	7.367	151.8	17.2	0.1468	3.453	0
6	7.67	2540.9	328.9	0.1288	57.804	0.857
7	7.8	287.7	58.6	0.0592	6.545	0
8	8.049	50.8	6.3	0.1347	1.155	0.928
9	8.314	535.6	47.9	0.1863	12.184	0.939
10	9.466	257.8	6.4	0.668	5.865	1.581
11	10.266	177.9	5.6	0.5295	4.048	1.434

Table 18, Entry 4. Sequence: ACTG-Gly-$CONH_2$. Automated flow synthesis.



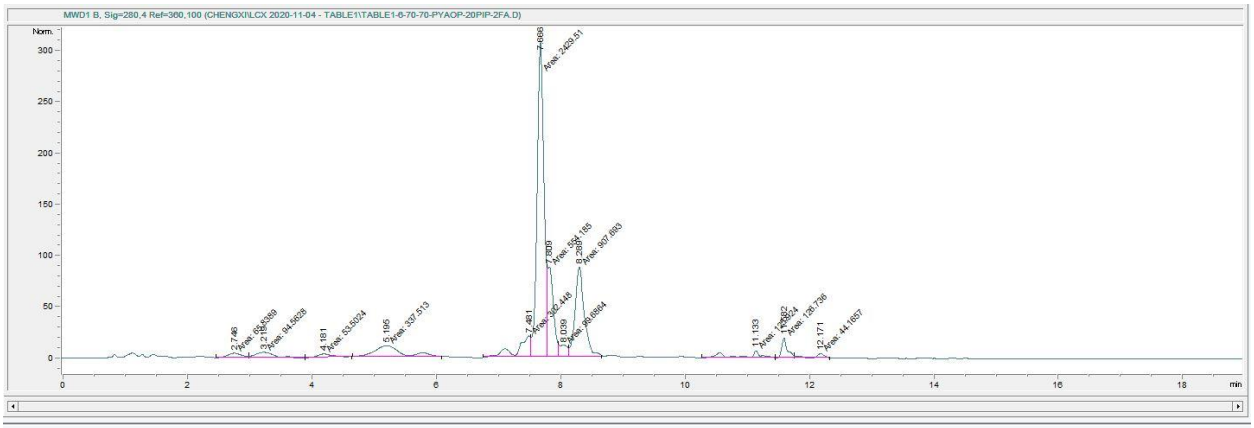
#	Time	Area	Height	Width	Area%	Symmetry
1	2.78	149.5	9.6	0.2591	2.025	0.949
2	3.252	183.4	8.7	0.35	2.486	0
3	4.211	103.2	5.6	0.3098	1.399	0.542
4	5.22	155.1	4.6	0.5611	2.102	1.895
5	7.49	341.1	23.2	0.2452	4.622	8.207
6	7.659	3551.6	441	0.1342	48.129	0.82
7	7.783	874.3	117.4	0.1084	11.849	0
8	8.295	923.8	84.4	0.1823	12.519	0.887
9	9.651	390.1	11.4	0.5723	5.286	7.007
10	9.9	560.3	63.7	0.1465	7.592	0.547
11	10.436	146.9	10.8	0.2277	1.990	0.706

Table 18, Entry 5. Sequence: ACTG-Gly-$CONH_2$. Automated flow synthesis.



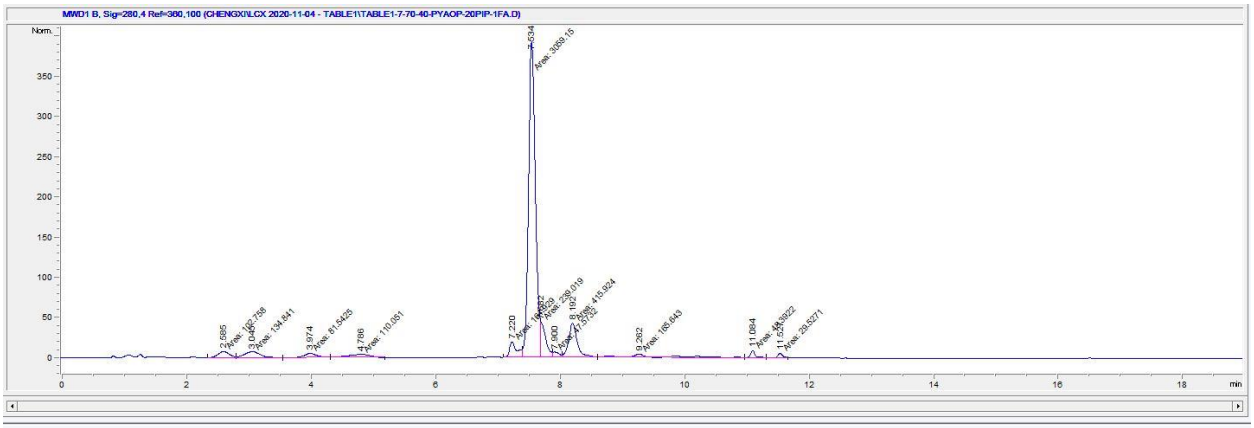
#	Time	Area	Height	Width	Area%	Symmetry
1	2.749	111.3	7.4	0.2508	1.495	0.955
2	3.213	150.8	7.5	0.3355	2.026	0
3	4.163	86.8	5.6	0.2599	1.166	0
4	5.158	139.8	4.6	0.5059	1.879	1.664
5	7.473	314.9	26.9	0.1952	4.232	0
6	7.629	4072.6	501.5	0.1353	54.729	0.812
7	7.782	856.1	118.2	0.1207	11.505	0.165
8	8.273	1015.5	98.9	0.1712	13.646	0.837
9	11.12	393.8	12.2	0.5376	5.292	5.065
10	11.567	253.3	35.3	0.1196	3.404	0
11	12.156	46.6	5.8	0.1337	0.626	0.864

Table 18, Entry 6. Sequence: ACTG-Gly-$CONH_2$. Automated flow synthesis.



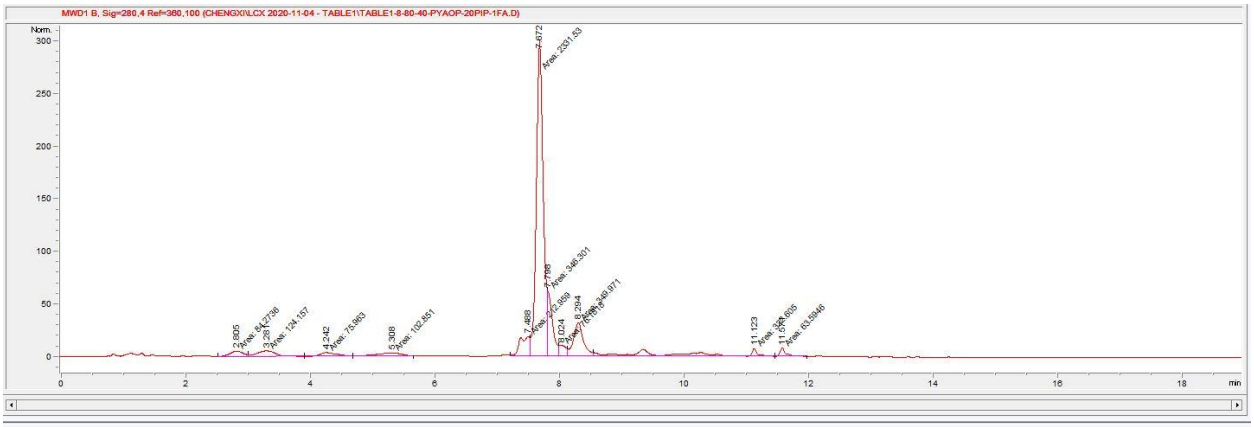
#	Time	Area	Height	Width	Area%	Symmetry
1	2.746	65.8	4.3	0.258	1.280	0
2	3.219	94.6	5	0.3146	1.839	0
3	4.181	53.5	3.2	0.2766	1.041	0
4	5.195	337.5	10.9	0.5181	6.564	0.799
5	7.481	302.4	20.4	0.2468	5.882	5.472
6	7.666	2429.5	307	0.1319	47.251	0.874
7	7.809	554.2	87.7	0.1053	10.778	0.302
8	8.039	99.7	11.3	0.1258	1.939	0.9
9	8.289	907.7	87.3	0.1734	17.653	0.773
10	11.133	125.9	6.9	0.3032	2.449	2.502
11	11.582	126.7	19.3	0.1093	2.465	0.607
12	12.171	44.2	4.3	0.1728	0.859	1.569

Table 18, Entry 7. Sequence: ACTG-Gly-$CONH_2$. Automated flow synthesis.



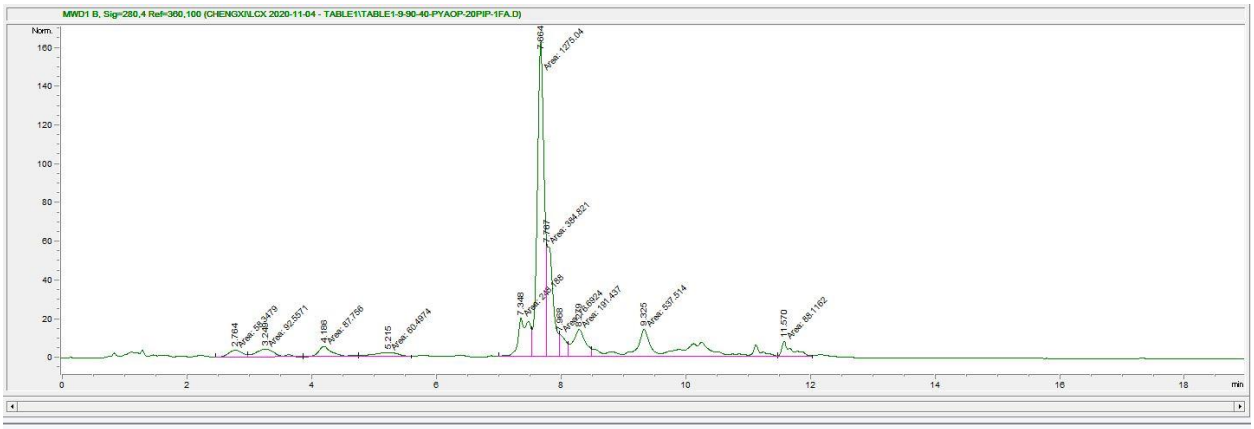
#	Time	Area	Height	Width	Area%	Symmetry
1	2.585	102.8	7.8	0.219	2.234	0.971
2	3.045	134.8	7.5	0.3006	2.932	0.953
3	3.974	81.5	5.4	0.252	1.773	0
4	4.786	110.1	4.2	0.4409	2.393	1.416
5	7.22	164.9	19.4	0.1414	3.586	0
6	7.534	3059.1	392.8	0.1298	66.513	0.841
7	7.682	239	44.4	0.0653	5.197	0
8	7.9	47.6	6.7	0.1188	1.034	0.257
9	8.192	415.9	42.7	0.1622	9.043	0
10	9.262	165.6	4.4	0.6294	3.601	0.549
11	11.084	48.4	8.8	0.0918	1.052	0.694
12	11.527	29.5	5.5	0.0899	0.642	0.88

Table 18, Entry 8. Sequence: ACTG-Gly-$CONH_2$. Automated flow synthesis.



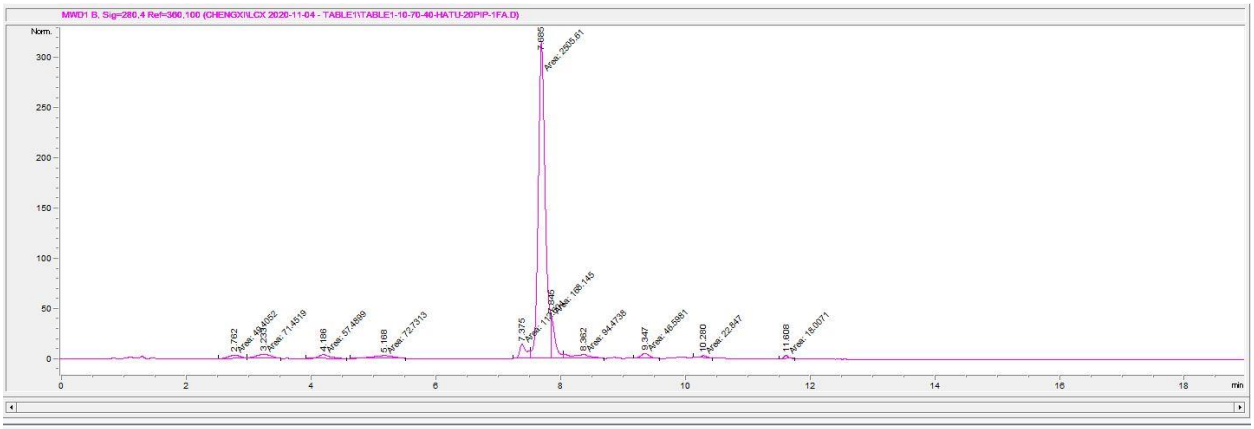
#	Time	Area	Height	Width	Area%	Symmetry
1	2.805	84.3	5.3	0.2668	2.055	1.045
2	3.281	124.2	5.5	0.376	3.028	0
3	4.242	76	3.8	0.337	1.853	0.536
4	5.308	102.9	3.4	0.5104	2.508	1.753
5	7.488	213	18.5	0.1915	5.194	3.758
6	7.672	2331.5	299.8	0.1296	56.861	0.864
7	7.798	346.3	62.5	0.0663	8.446	0
8	8.024	76.2	9.9	0.1276	1.858	0.358
9	8.294	350	31.7	0.1841	8.535	0.852
10	11.123	332.6	7.5	0.7435	8.112	8.761
11	11.571	63.6	8	0.1318	1.551	0.478

Table 18, Entry 9. Sequence: ACTG-Gly-$CONH_2$. Automated flow synthesis.



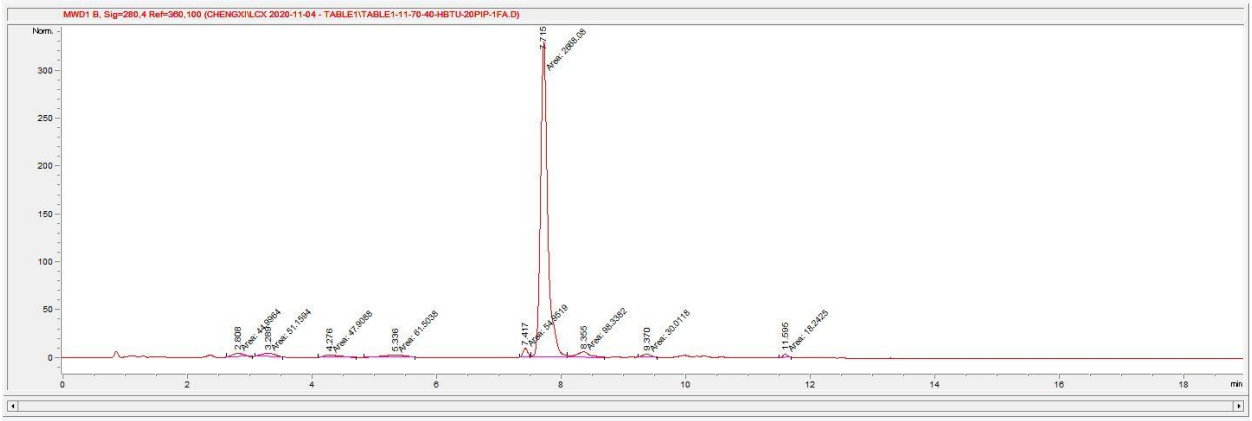
#	Time	Area	Height	Width	Area%	Symmetry
1	2.764	58.3	3.7	0.2639	1.882	0.953
2	3.249	92.6	4.2	0.3717	2.985	0
3	4.186	87.8	5.4	0.2723	2.830	0.595
4	5.215	60.5	2.1	0.4775	1.951	1.475
5	7.348	248.2	19.6	0.2106	8.004	0.407
6	7.664	1275	162.5	0.1308	41.117	0.913
7	7.767	384.8	57	0.0987	12.410	0
8	7.968	76.7	11.2	0.114	2.473	0
9	8.279	191.4	13.8	0.2312	6.173	0.869
10	9.325	537.5	13.9	0.6434	17.334	0
11	11.57	88.1	7.8	0.1879	2.842	0.287

Table 18, Entry 10. Sequence: ACTG-Gly-$CONH_2$. Automated flow synthesis.



#	Time	Area	Height	Width	Area%	Symmetry
1	2.762	49.4	3.5	0.237	1.535	0.99
2	3.233	71.5	4	0.2943	2.219	1.156
3	4.186	57.5	4.1	0.235	1.786	0.641
4	5.168	72.7	2.9	0.4204	2.259	1.376
5	7.375	112.6	14.2	0.1317	3.498	0
6	7.685	2505.6	314.8	0.1326	77.829	0.823
7	7.845	168.1	42.6	0.0658	5.223	0
8	8.362	94.5	4.2	0.3732	2.935	1.728
9	9.347	46.6	4.9	0.1573	1.447	1.013
10	10.28	22.8	2.4	0.1596	0.710	1.282
11	11.608	18	3.5	0.087	0.559	0.864

Table 18, Entry 11. Sequence: ACTG-Gly-$CONH_2$. Automated flow synthesis.



#	Time	Area	Height	Width	Area%	Symmetry
1	2.808	45	3.6	0.209	1.463	0.985
2	3.289	51.2	3.3	0.2568	1.664	0.873
3	4.276	47.9	2.8	0.2839	1.558	0.488
4	5.336	61.5	2.4	0.4273	2.000	1.489
5	7.417	55	9.8	0.0936	1.787	0.845
6	7.715	2668.1	329.5	0.135	86.761	0.766
7	8.355	98.3	6	0.2725	3.198	1.076
8	9.37	30	3.3	0.151	0.976	1.07
9	11.595	18.2	3.8	0.0803	0.593	0.955

4.4.13. HPLC and LCMS data of PNAs and PPNAs for EGFP assay

An 18-mer anti-IVS2-654 PNA with sequence “GCTATTACCTTAACCCAG-KKK-$CONH_2$” was ordered from PNA Bio with 3-week lead time as a positive control. The purity of this commercial 18-mer PNA was analyzed with HPLC and LCMS before using.

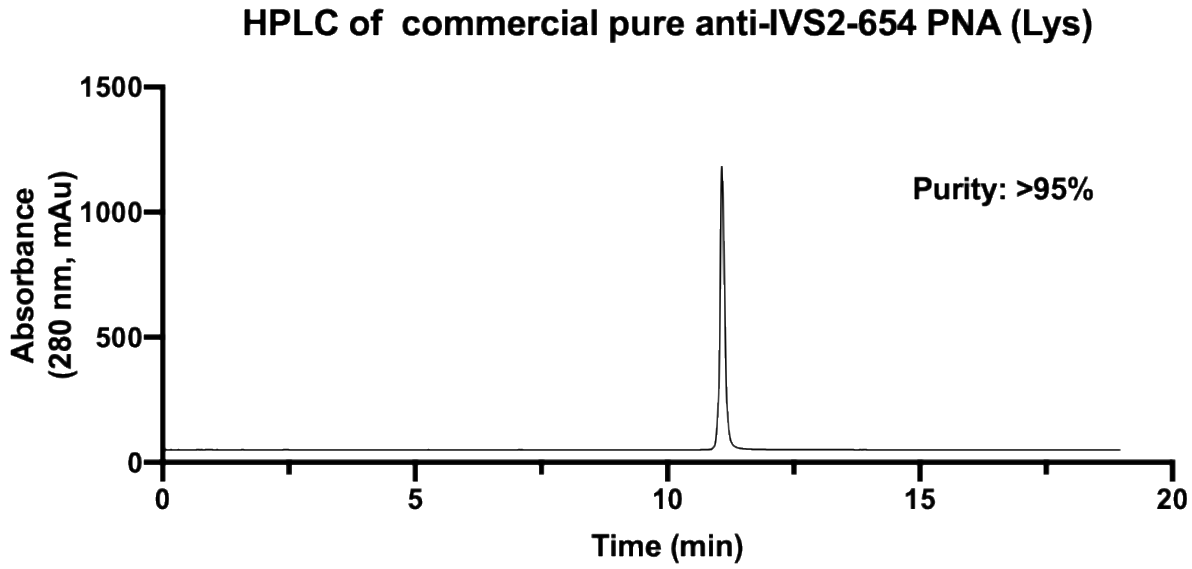
Figure 4.4, Entry 1. Sample: Commercial PNA

Sequence: GCTATTACCTTAACCCAG-KKK-$CONH_2$.

Synthesis method: Ordered from PNA Bio, 3-week lead time.

Purity: >95%

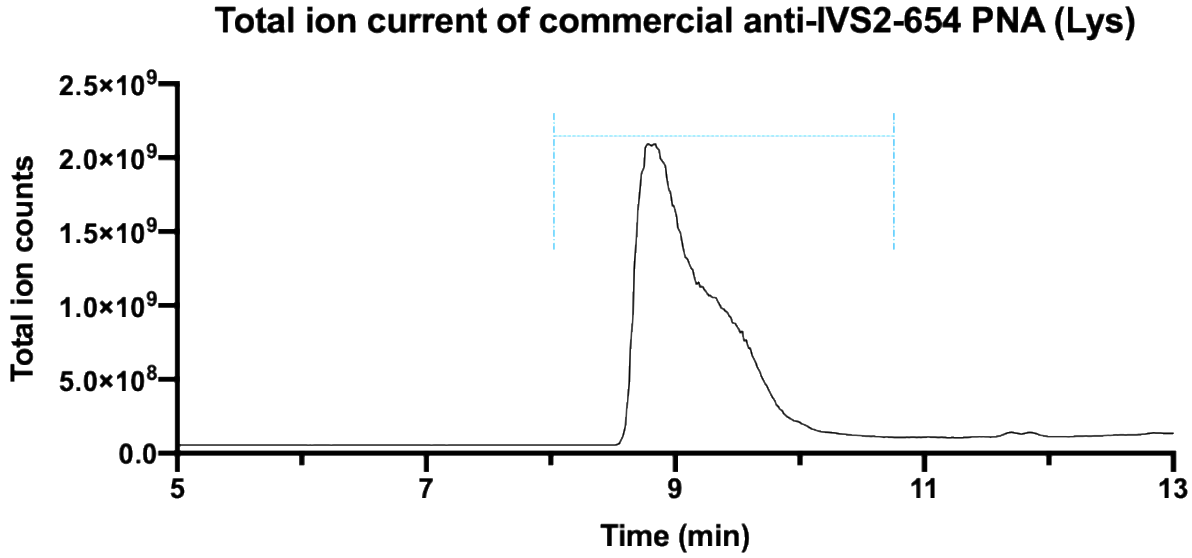
HPLC method: Section 4.4.3



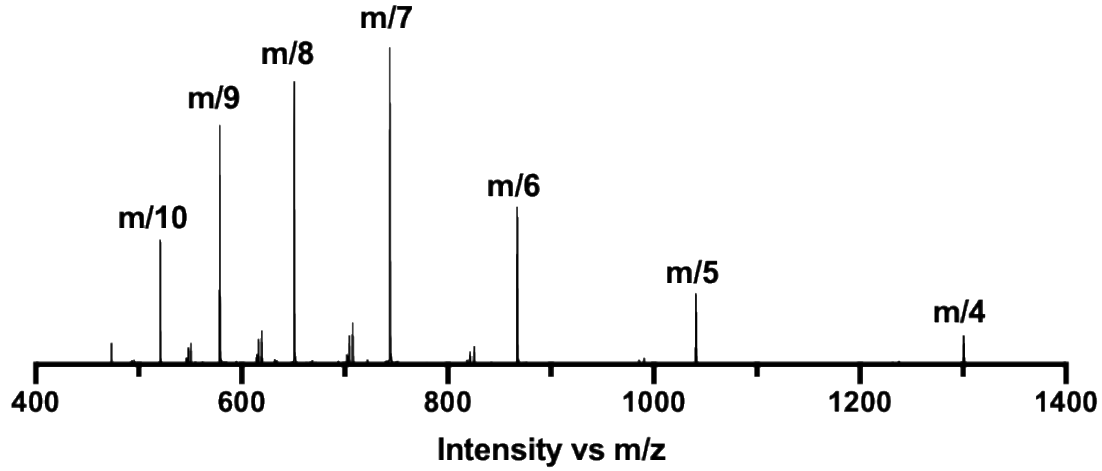
LCMS method: Condition 1

Calculated: 5196.21 Da

Observed: 5196.27 Da



MS data of commercial anti-IVS2-654 PNA (Lys)



Deconvoluted commercial anti-IVS2-654 PNA (Lys)

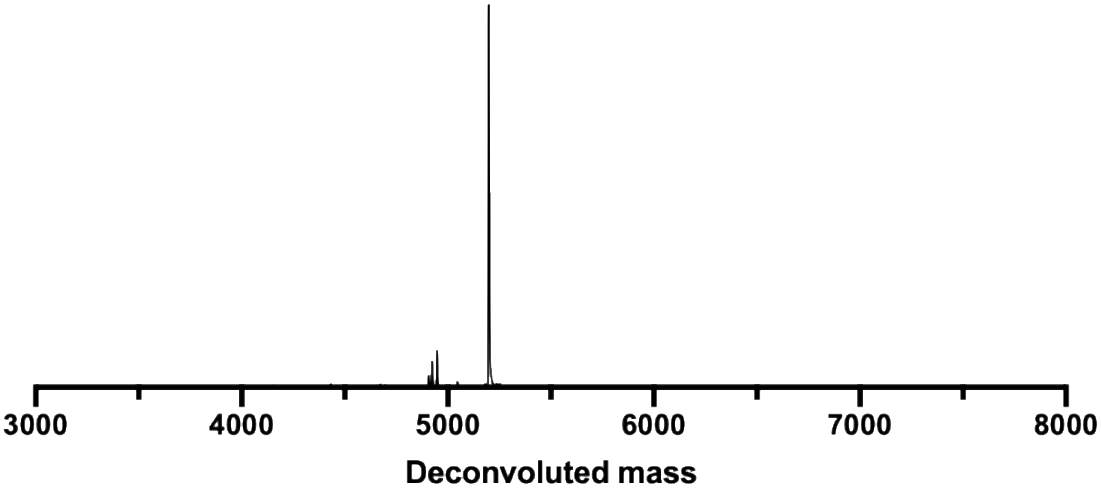


Figure 4.4, Entry 2. Sample: Synthesized PNA

Sequence: GCTATTACCTTAACCCAG-KKK-$CONH_2$.

Synthesis method: Automated flow synthesis, 63 minutes.

Resin: 15 mg Rink Amide resin (0.5 mmol/g, 7.5 µmmol).

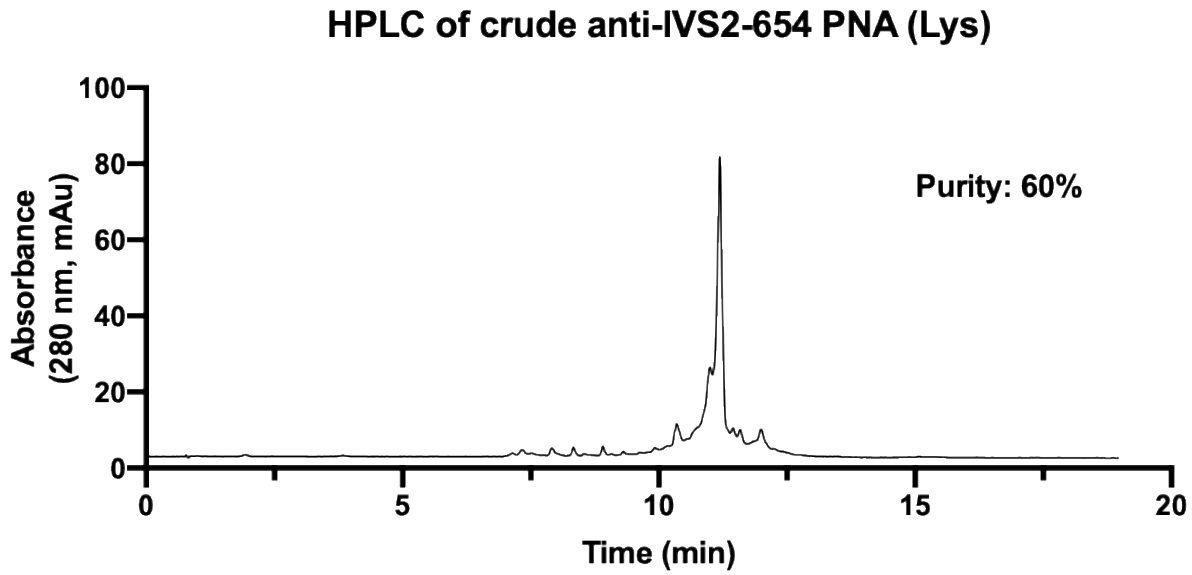
Crude purity: 60%

Pure purity: >95%

Obtained sample: 4 mg, 10% yield.

Purification method: method 1.

HPLC method: Section 4.4.3



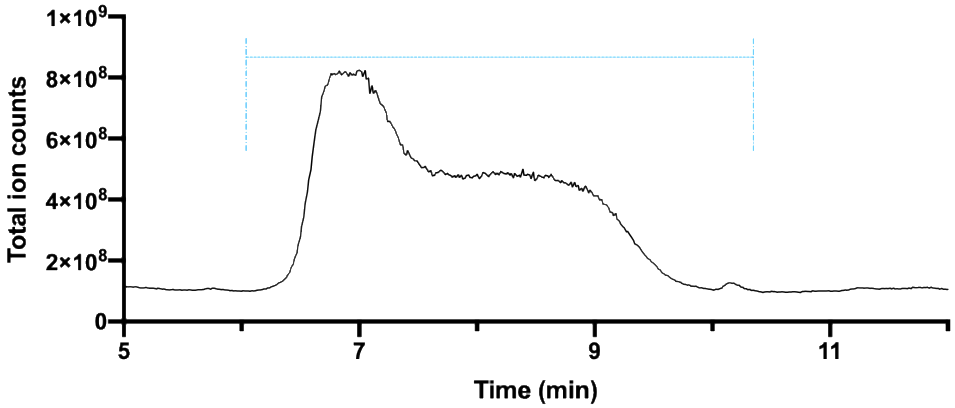
Crude LCMS trace was shown here to demonstrate the synthesis efficiency.

LCMS method: condition 1

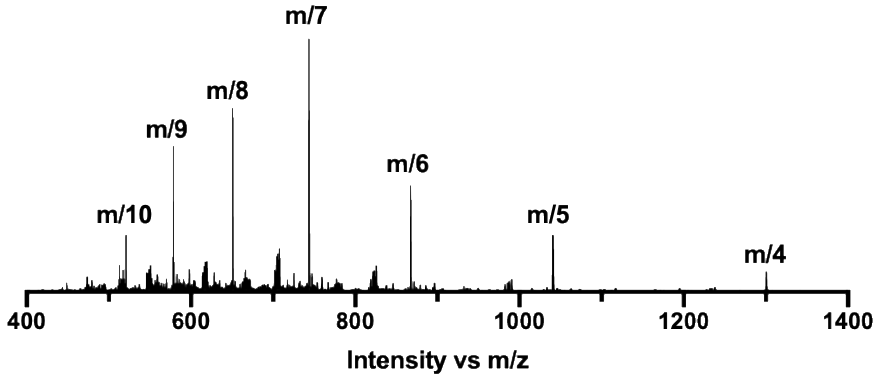
Calculated: 5196.21 Da

Observed: 5196.25 Da

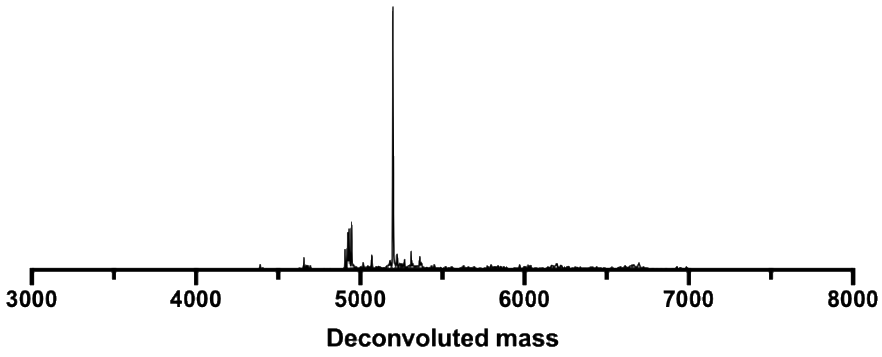
Total ion current of crude anti-IVS2-654 PNA (Lys)



MS data of crude anti-IVS2-654 PNA (Lys)

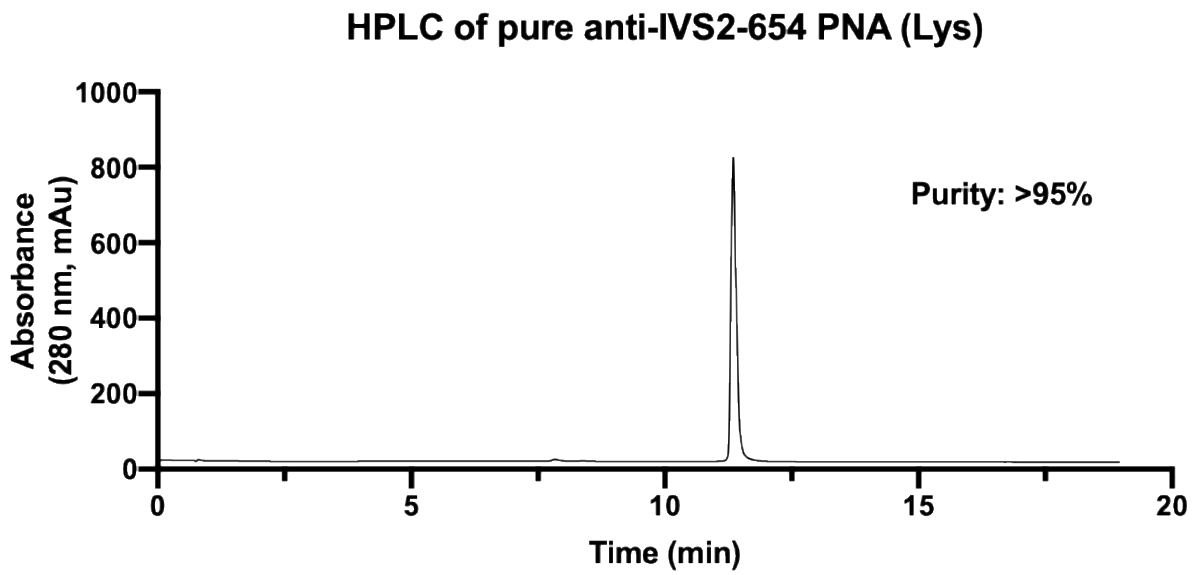


Deconvoluted crude anti-IVS2-654 PNA (Lys)



Pure sample after purification

HPLC method: Section 4.4.3



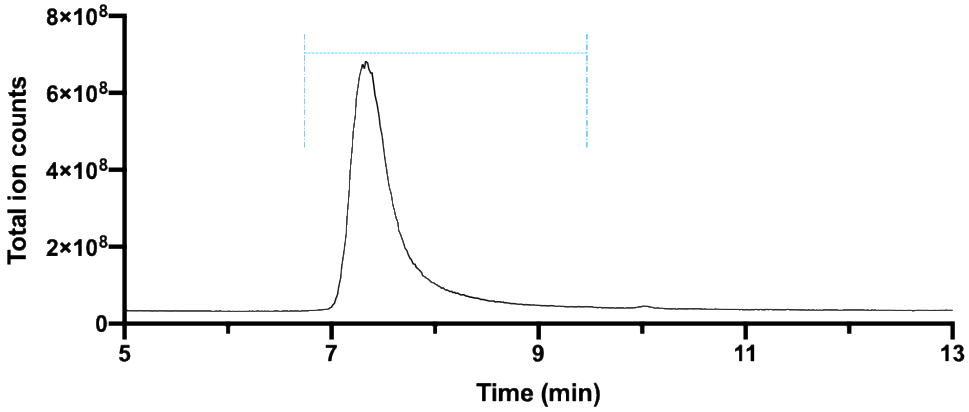
Pure sample after purification

LCMS method: condition 1

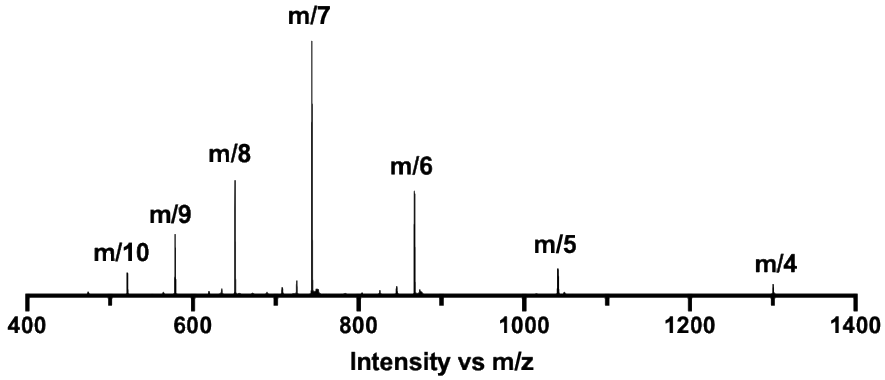
Calculated: 5196.21 Da

Observed: 5196.25 Da

Total ion current of pure anti-IVS2-654 PNA (Lys)



MS data of pure anti-IVS2-654 PNA (Lys)



Deconvoluted pure anti-IVS2-654 PNA (Lys)

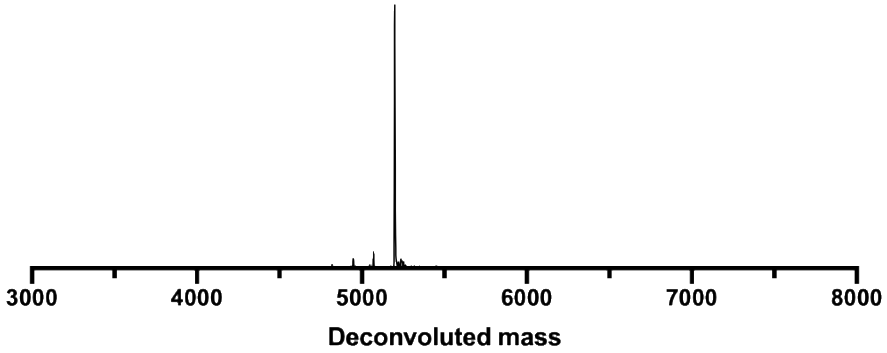


Figure 4.4, Entry 3. Sample: Scramble PNA

Sequence: CCTCATACATACGTACGT-KKK-$CONH_2$.

Synthesis method: Automated flow synthesis, 63 minutes.

Resin: 15 mg Rink Amide resin (0.5 mmol/g, 7.5 µmmol).

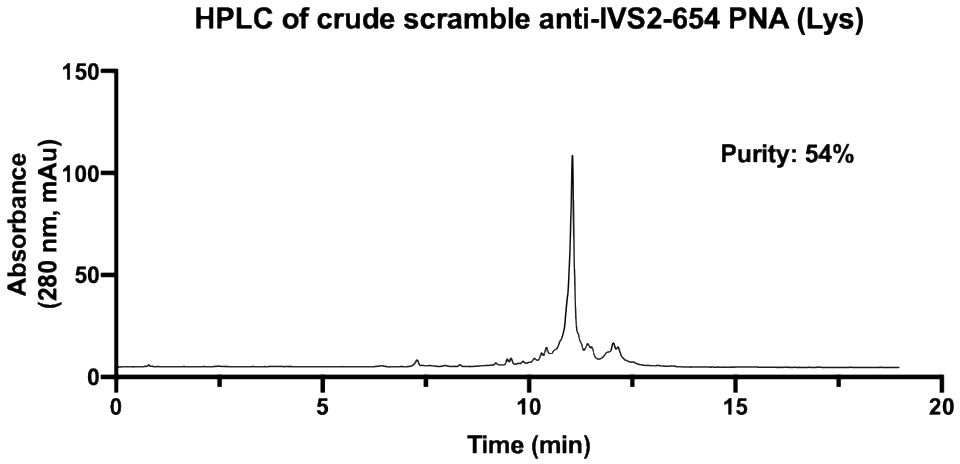
Crude purity: 54%

Pure purity: >95%

Obtained sample: 3.3 mg, 8.5% yield.

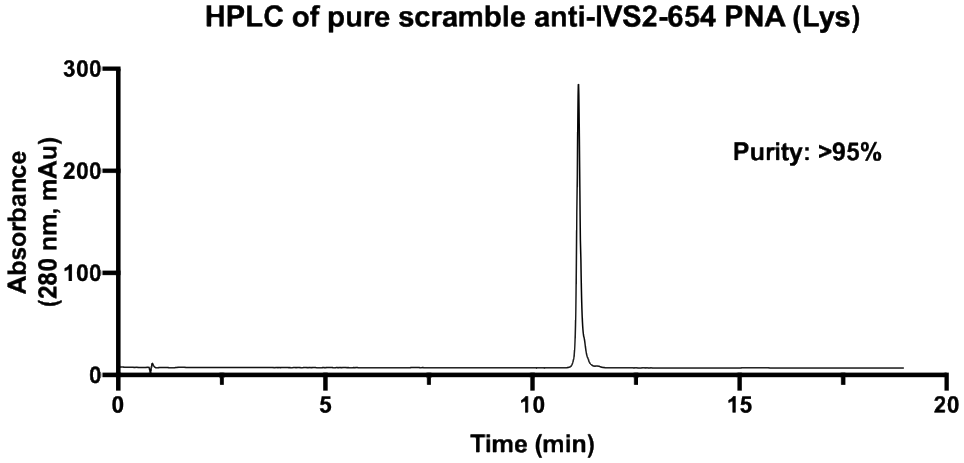
Purification method: method 1.

HPLC method: Section 4.4.3



Pure sample after purification

HPLC method: Section 4.4.3

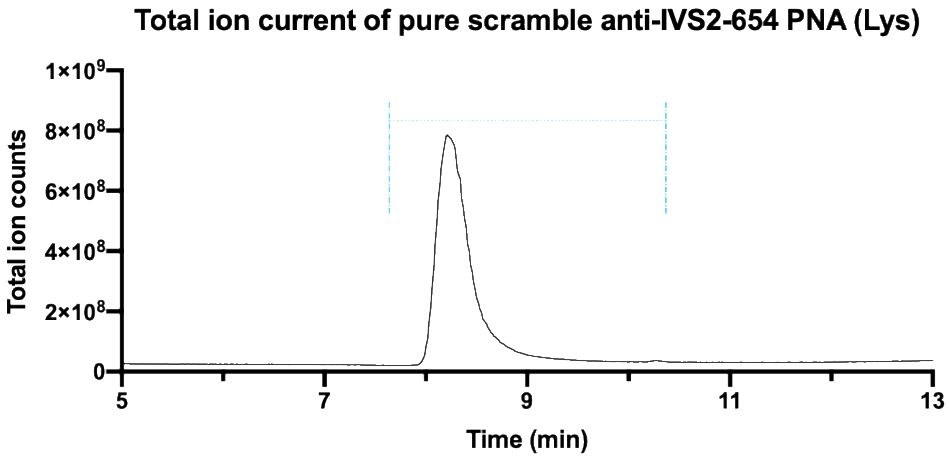


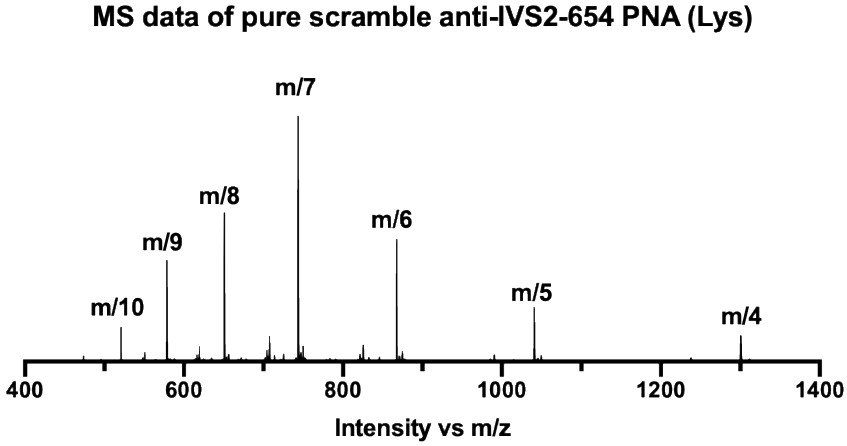
Pure sample after purification

LCMS method: condition 1

Calculated: 5196.21 Da

Observed: 5196.27 Da





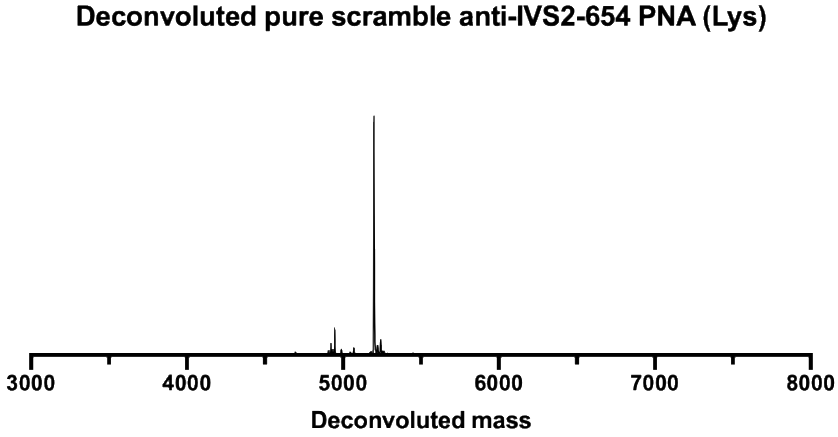


Figure 4.4, Entry 4. Sample: anti-IVS2-654 PPNA (Bpep)

Sequence: GCTATTACCTTAACCCAG-KKK-Bpep-$CONH_2$.

(Bpep = RXRRBRRXRRBR, X = 6-aminohexanoic acid; B = β-alanine)

Synthesis method: Automated flow synthesis, ~1.6 hours.

Resin: 20 mg Bpep loaded Rink Amide resin (3.5 μmmol).

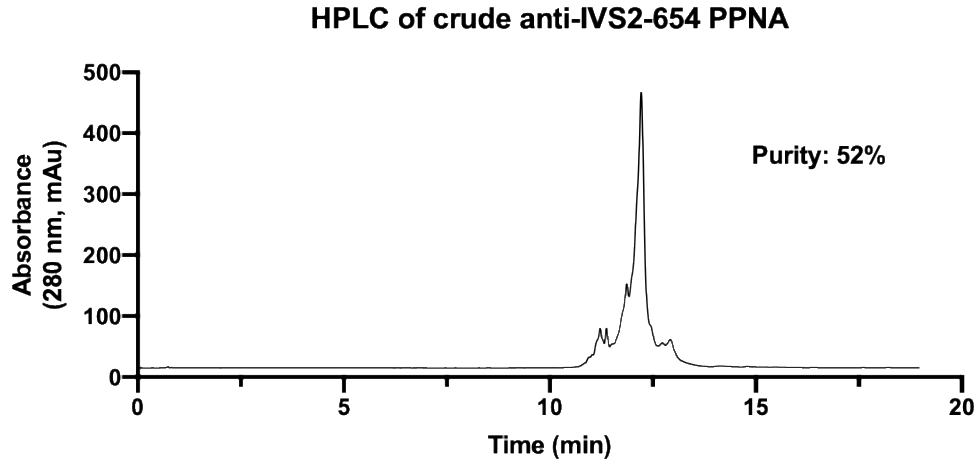
Crude purity: 52%

Pure purity: >95%

Obtained sample: 1.2 mg, 5% yield.

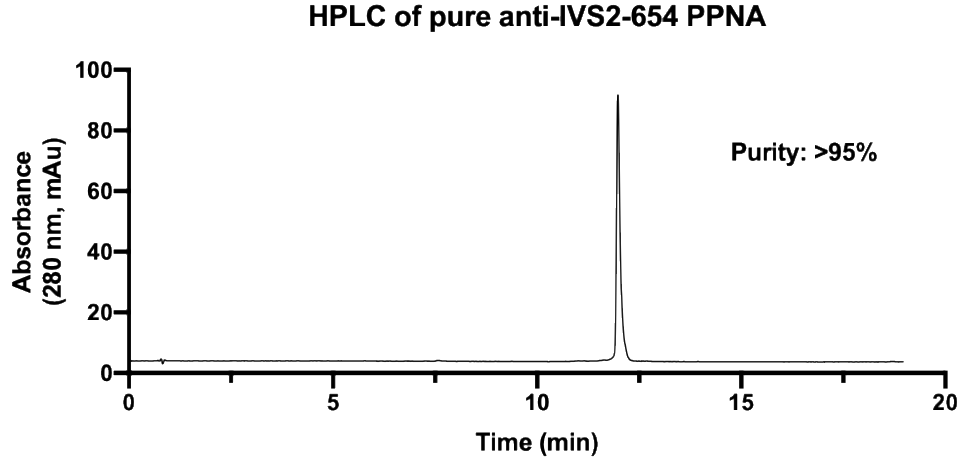
Purification method: method 1 and method 2.

HPLC method: Section 4.4.3



Pure sample after purification

HPLC method: Section 4.4.3

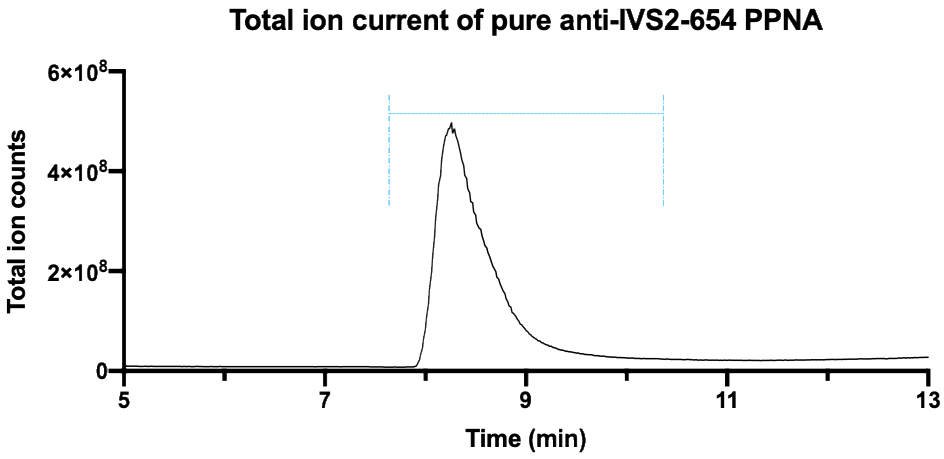


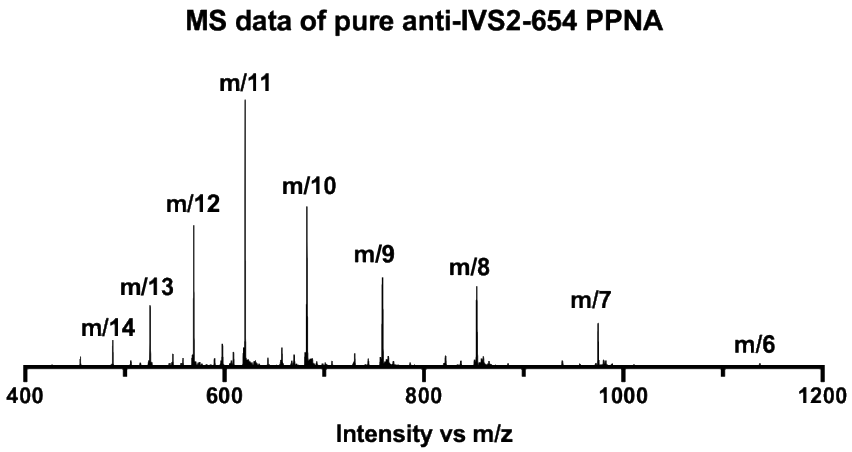
Pure sample after purification

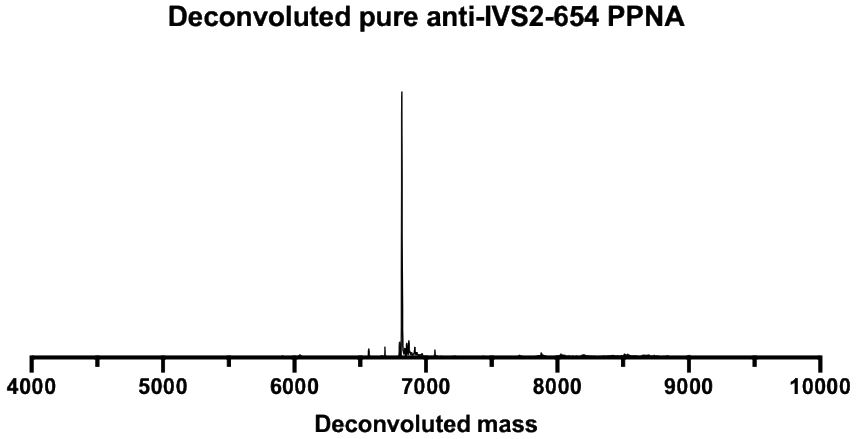
LCMS method: condition 1

Calculated: 6813.26 Da

Observed: 6813.43 Da







4.4.14. HPLC and LCMS data of PNAs and PPNAs for BL3 COVID-19 infection assay

The same anti-IVS2-654 PPNA sequence in Figure 4.4, Entry 4 was used as a negative control in BL3 COVID-19 infection assay (Figure 4.5, Entry 1).

Figure 4.5, Entry 2. Sample: 5'-UTR-1 PPNA (Bpep)

Sequence: TTGGTTGGTTTGTTACCT-KKK-Bpep-$CONH_2$.

(Bpep = RXRRBRRXRRBR, X = 6-aminohexanoic acid; B = β-alanine)

Synthesis method: Automated flow synthesis, ~1.6 hours.

Resin: 20 mg Bpep loaded Rink Amide resin (3.5 μmmol).

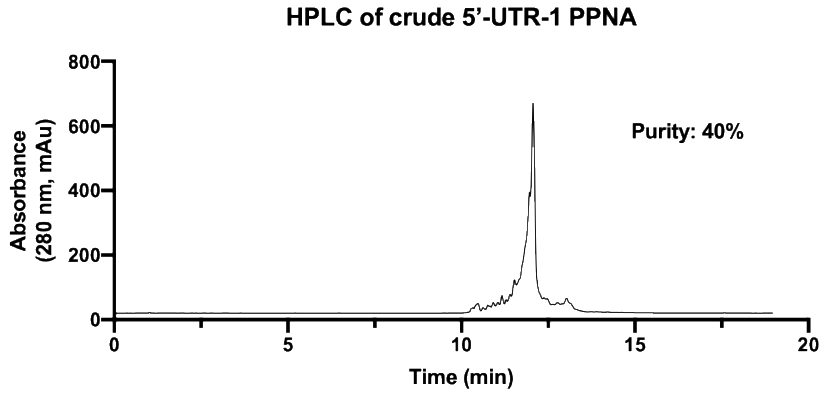
Crude purity: 40%

Pure purity: >95%

Obtained sample: 1.5 mg, 6% yield.

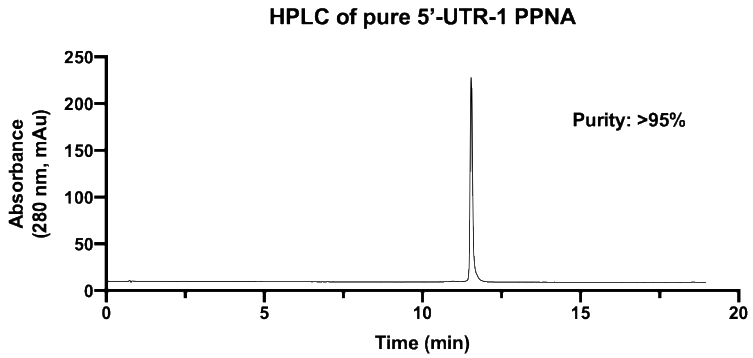
Purification method: method 1 and method 2.

HPLC method: Section 4.4.3



Pure sample after purification

HPLC method: Section 4.4.3

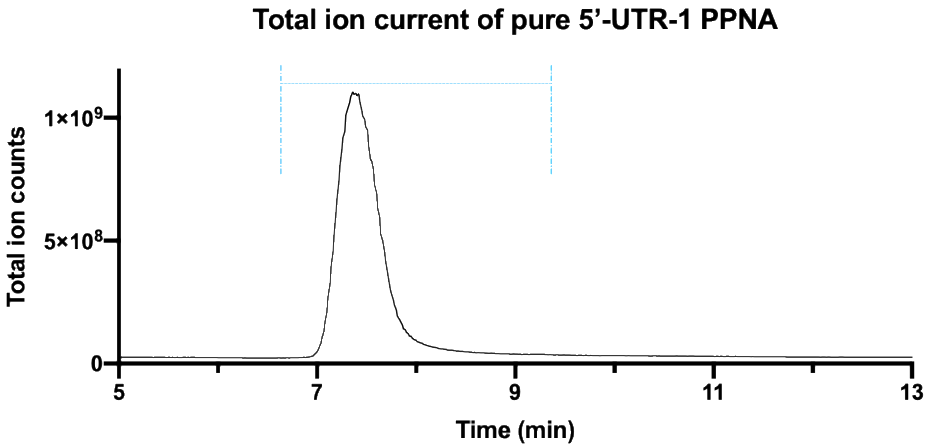


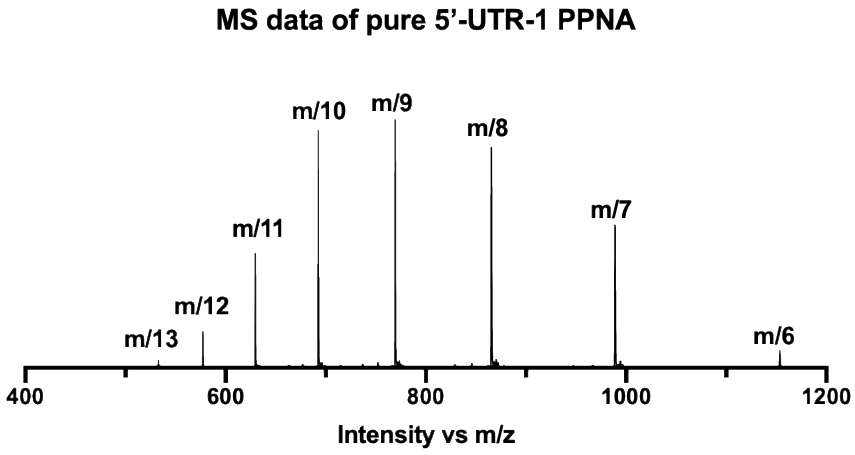
Pure sample after purification

LCMS method: condition 1

Calculated: 6912.23 Da

Observed: 6912.34 Da





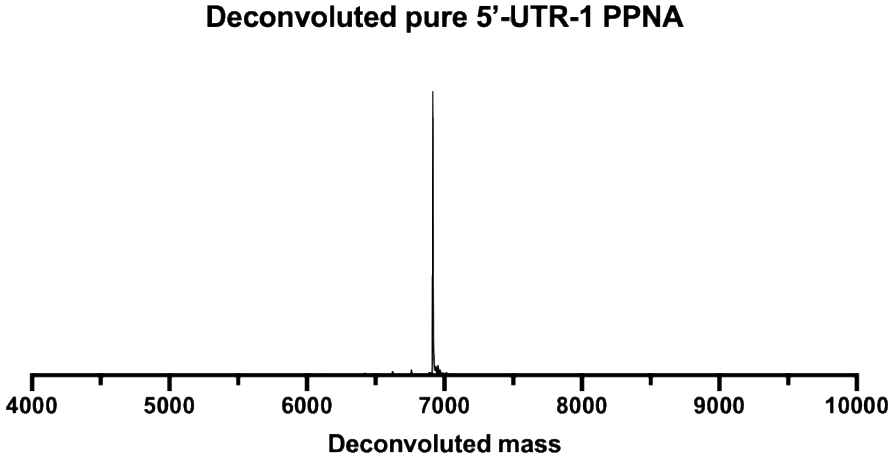


Figure 4.5, Entry 3. Sample: 5'-UTR-2 PPNA (Bpep)

Sequence: AGTTACTCGTGTCCTGTC-KKK-Bpep-$CONH_2$.

(Bpep = RXRRBRRXRRBR, X = 6-aminohexanoic acid; B = β-alanine)

Synthesis method: Automated flow synthesis, ~1.6 hours.

Resin: 20 mg Bpep loaded Rink Amide resin (3.5 μmmol).

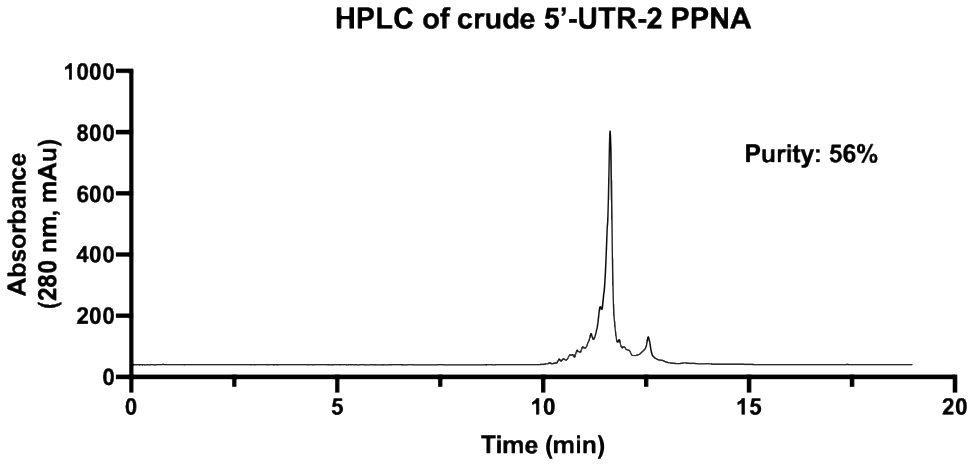
Crude purity: 56%

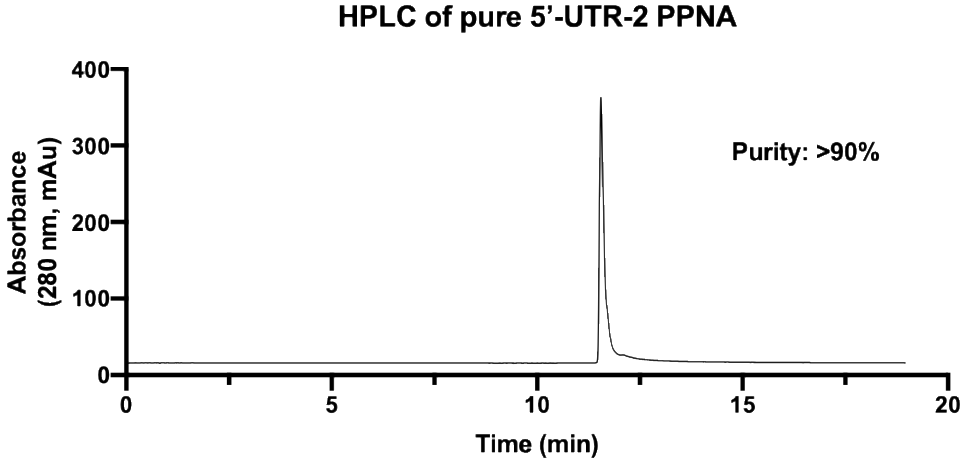
Pure purity: >90%

Obtained sample: 1.4 mg, 6% yield.

Purification method: method 1 and method 2.

HPLC method: Section 4.4.3





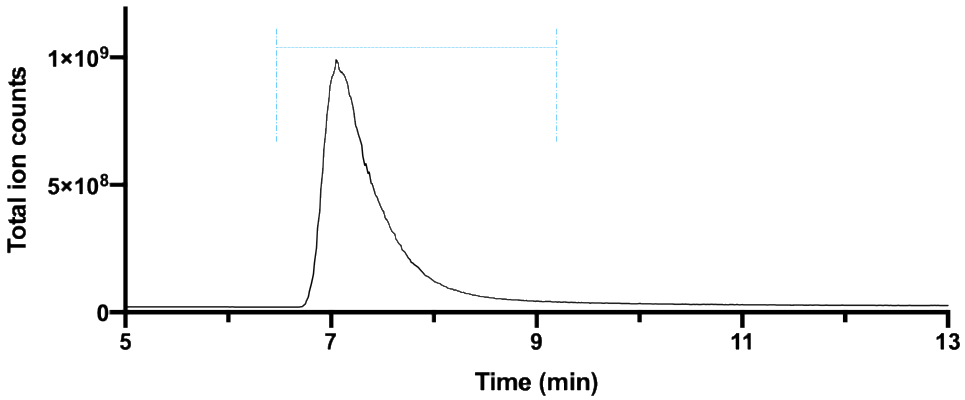
Pure sample after purification

LCMS method: condition 1

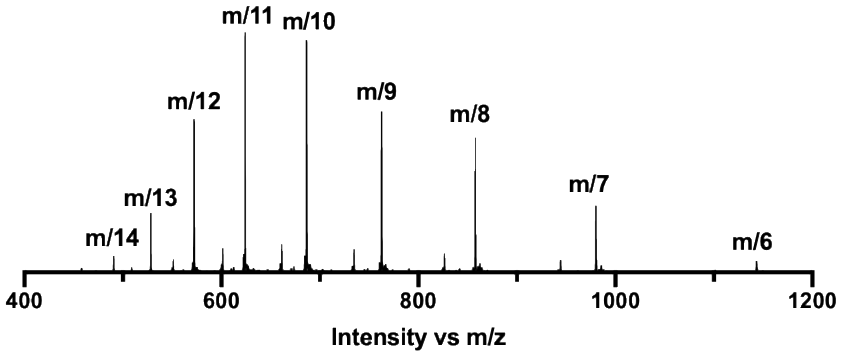
Calculated: 6851.24 Da

Observed: 6851.35 Da

Total ion current of pure 5’-UTR-2 PPNA



MS data of pure 5’-UTR-2 PPNA



Deconvoluted pure 5’-UTR-2 PPNA

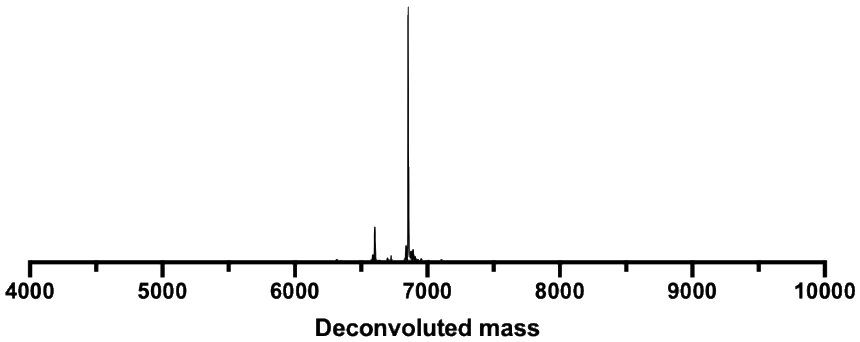


Figure 4.5, Entry 4. Sample: 5'-UTR-3 PPNA (Bpep)

Sequence: CTTACCTTTCGGTCACAC-KKK-Bpep-$CONH_2$.

(Bpep = RXRRBRRXRRBR, X = 6-aminohexanoic acid; B = β-alanine)

Synthesis method: Automated flow synthesis, ~1.6 hours.

Resin: 20 mg Bpep loaded Rink Amide resin (3.5 μmmol).

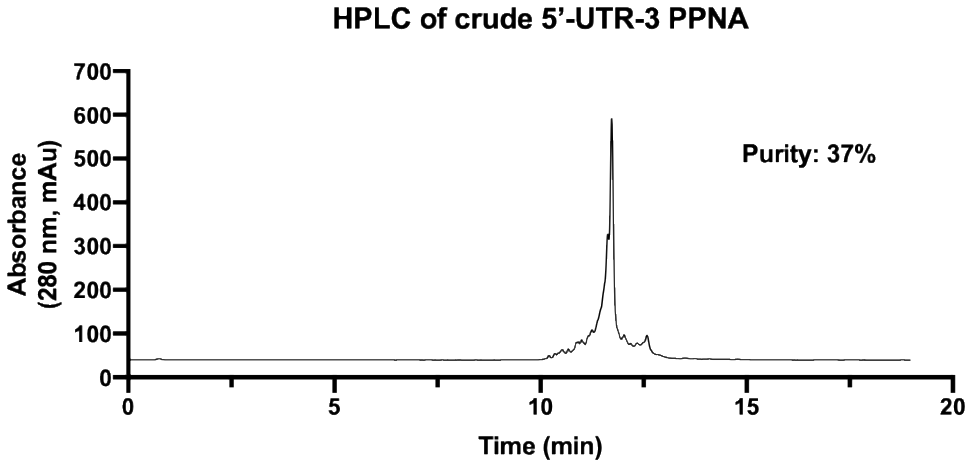
Crude purity: 37%

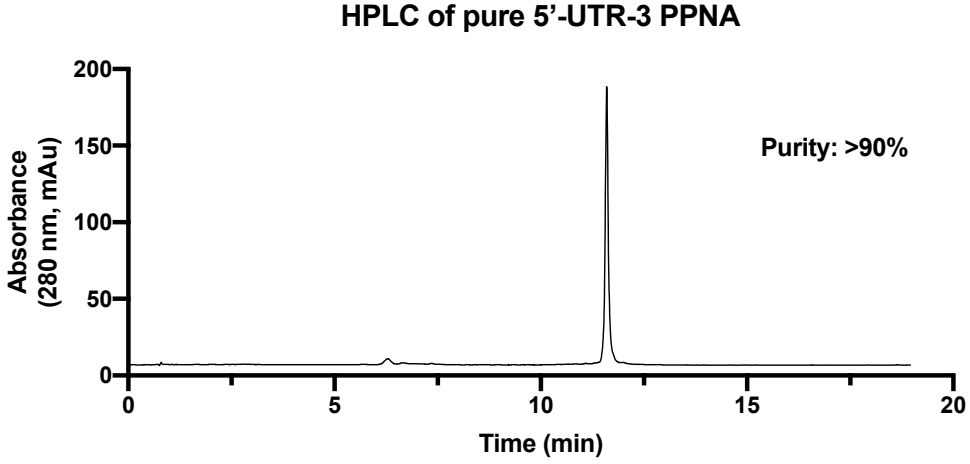
Pure purity: >90%

Obtained sample: 1.3 mg, 5% yield.

Purification method: method 1 and method 2.

HPLC method: Section 4.4.3



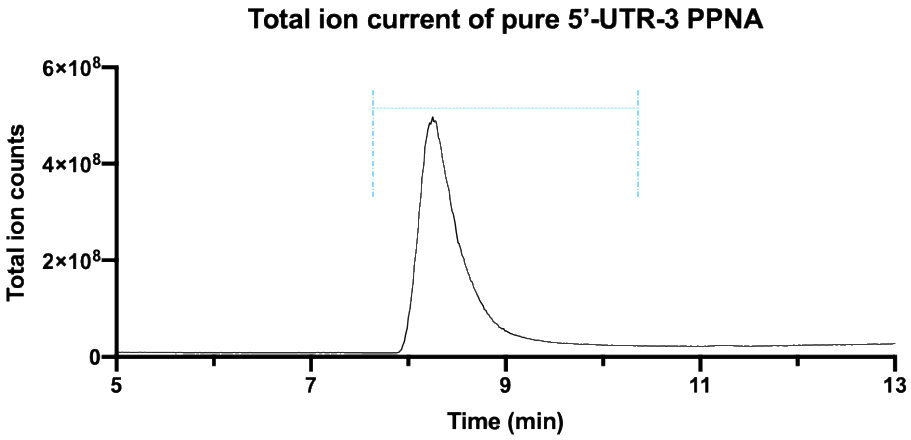


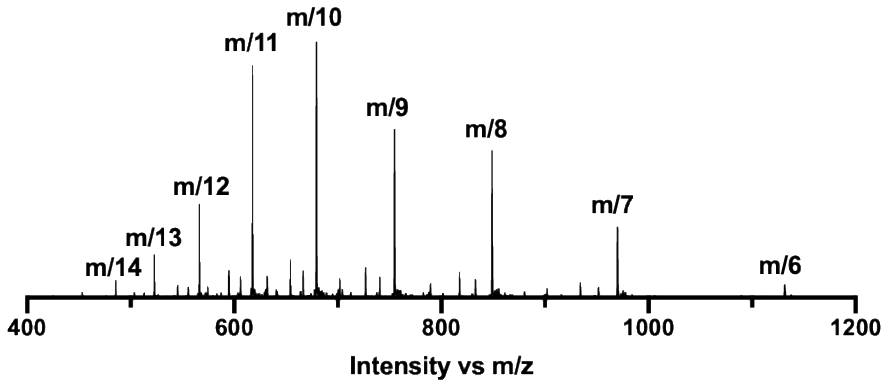
Pure sample after purification

LCMS method: condition 1

Calculated: 6780.24 Da

Observed: 6780.36 Da





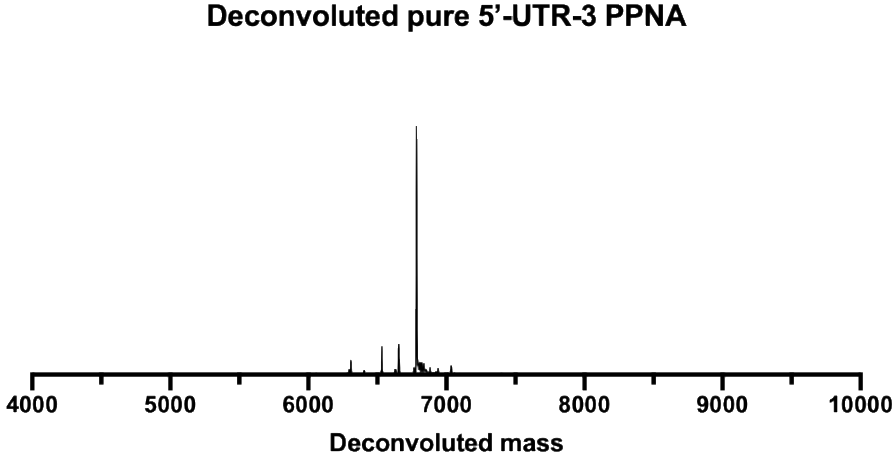


Figure 4.5, Entry 5. Sample: TRS-1 PPNA (Bpep)

Sequence: AAGTTCGTTTAGAGAACA-KKK-Bpep-$CONH_2$.

(Bpep = RXRRBRRXRRBR, X = 6-aminohexanoic acid; B = β-alanine)

Synthesis method: Automated flow synthesis, ~1.6 hours.

Resin: 20 mg Bpep loaded Rink Amide resin (3.5 μmmol).

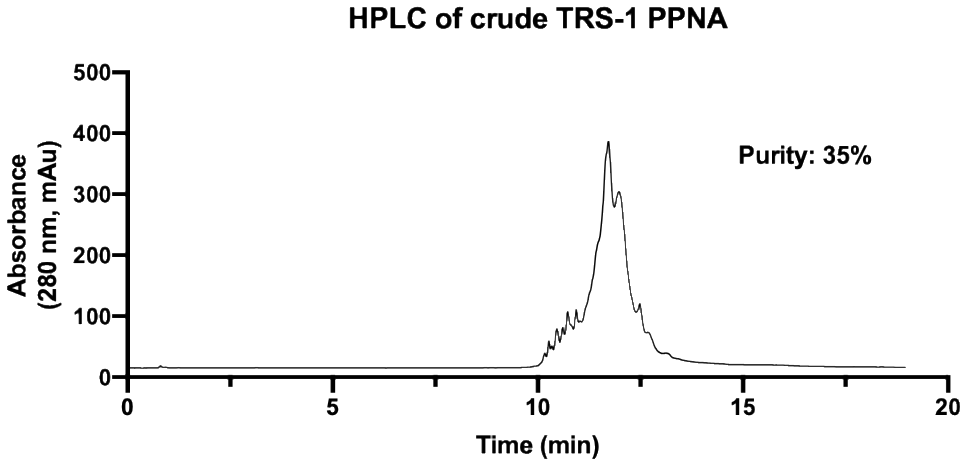
Crude purity: 35%

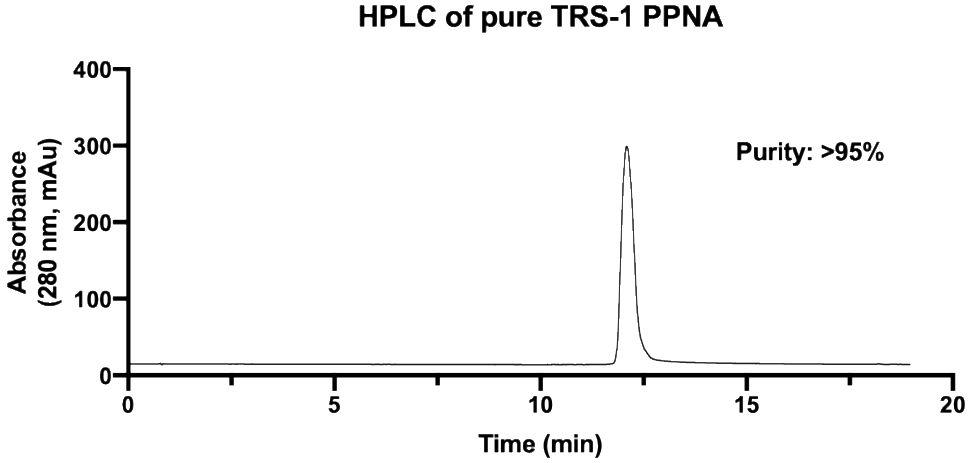
Pure purity: >95%

Obtained sample: 0.9 mg, 3.7% yield.

Purification method: method 1 and method 2.

HPLC method: Section 4.4.3



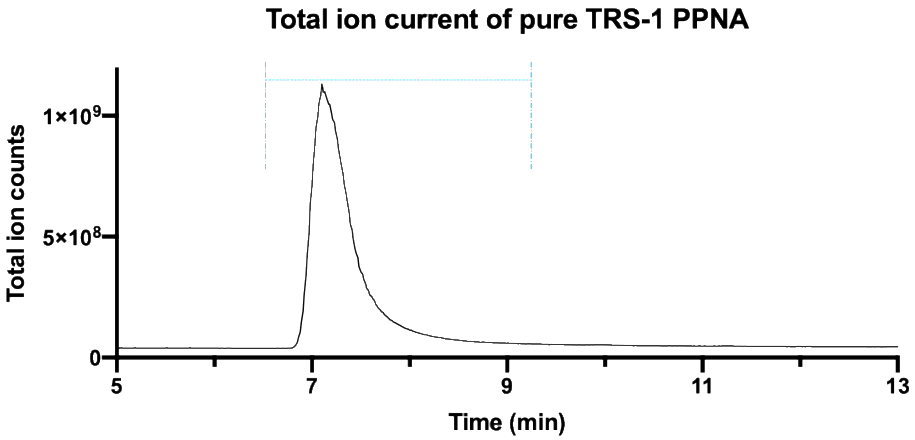


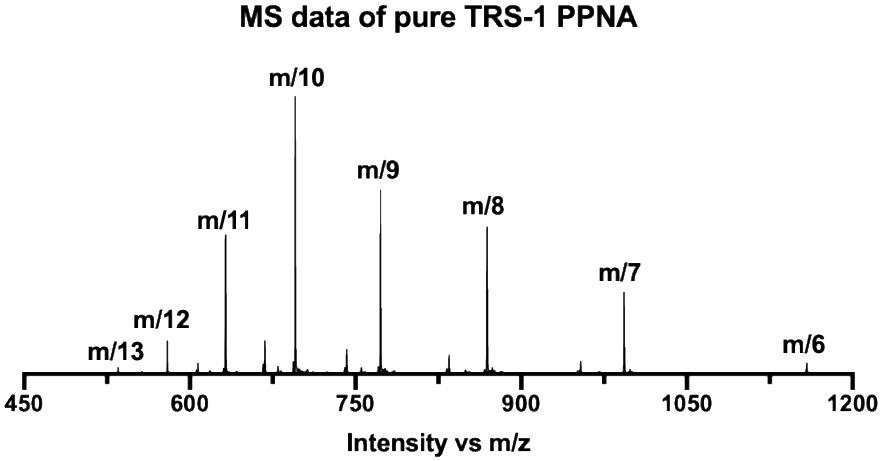
Pure sample after purification

LCMS method: condition 1

Calculated: 6941.30 Da

Observed: 6941.26 Da





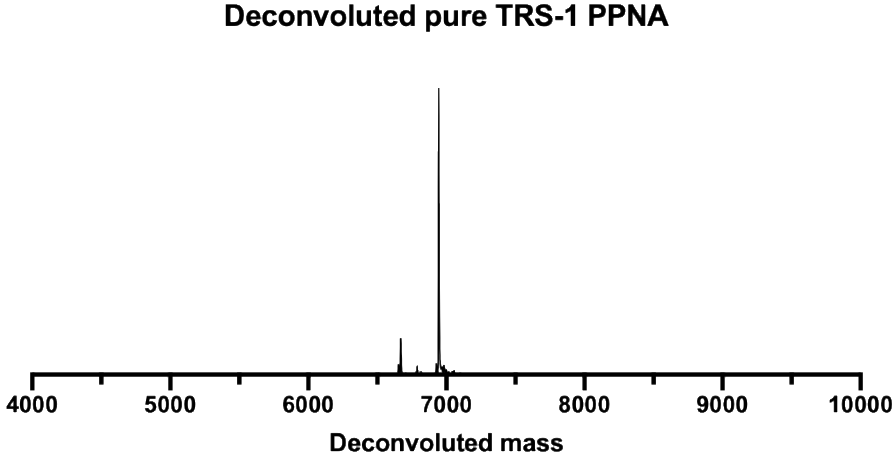


Figure 4.5, Entry 6. Sample: TRS-2 PPNA (Bpep)

Sequence: TTTTAAAGTTCGTTTAGA-KKK-Bpep-$CONH_2$.

(Bpep = RXRRBRRXRRBR, X = 6-aminohexanoic acid; B = β-alanine)

Synthesis method: Automated flow synthesis, ~1.6 hours.

Resin: 20 mg Bpep loaded Rink Amide resin (3.5 μmmol).

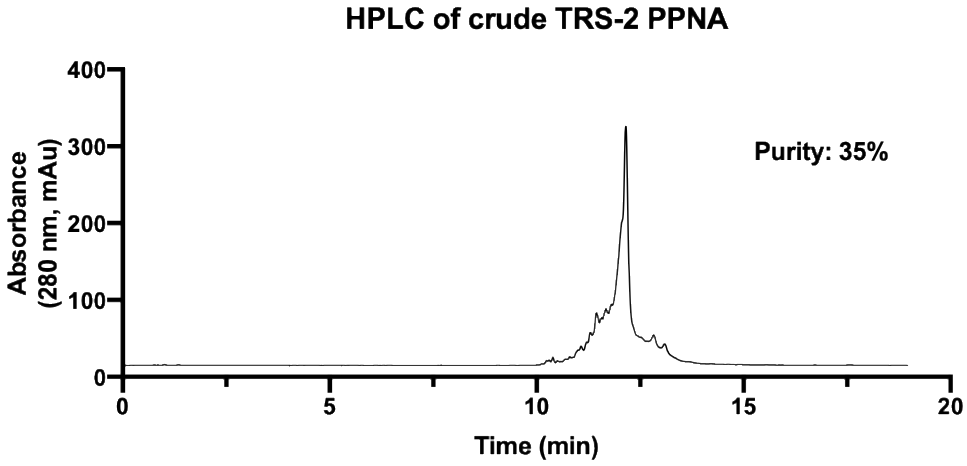
Crude purity: 35%

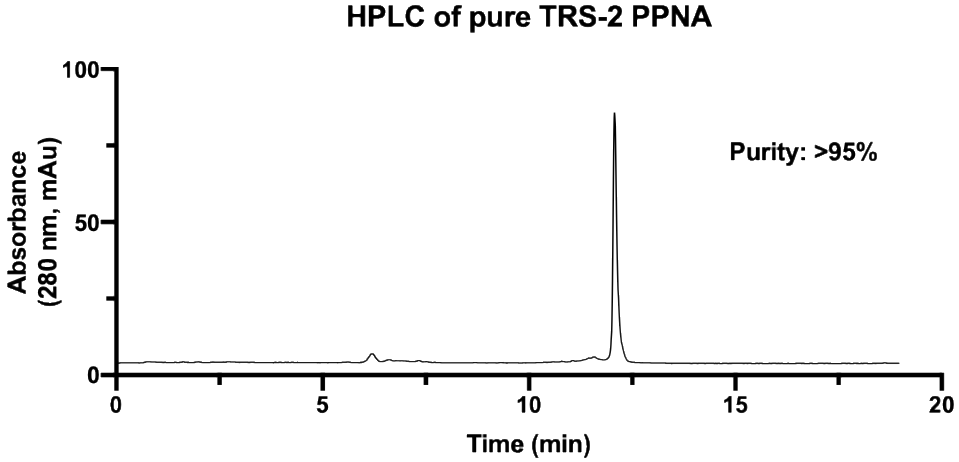
Pure purity: >95%

Obtained sample: 1.0 mg, 4% yield.

Purification method: method 1 and method 2.

HPLC method: Section 4.4.3



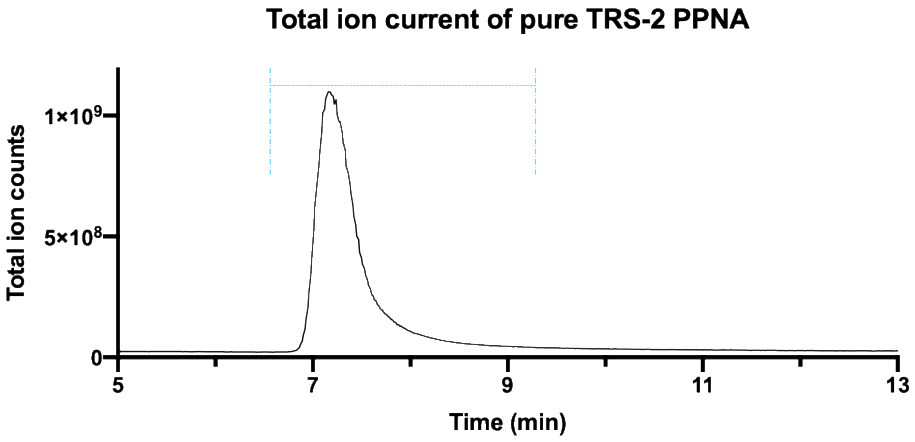


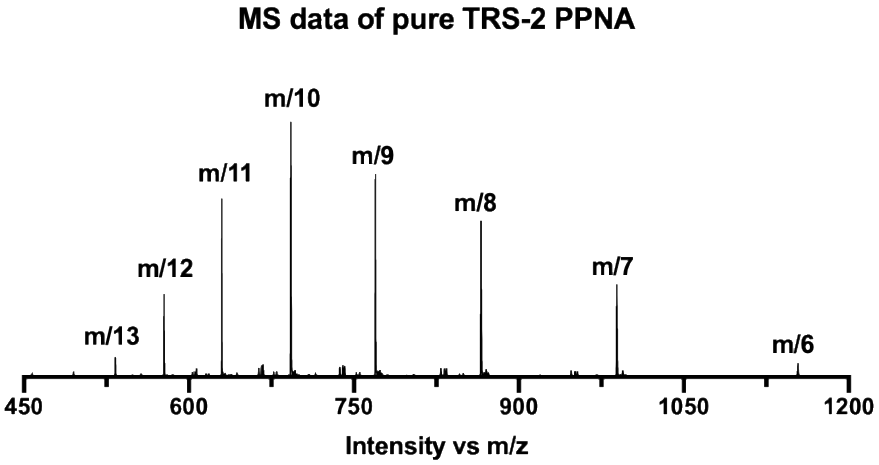
Pure sample after purification

LCMS method: condition 1

Calculated: 6913.27 Da

Observed: 6913.20 Da





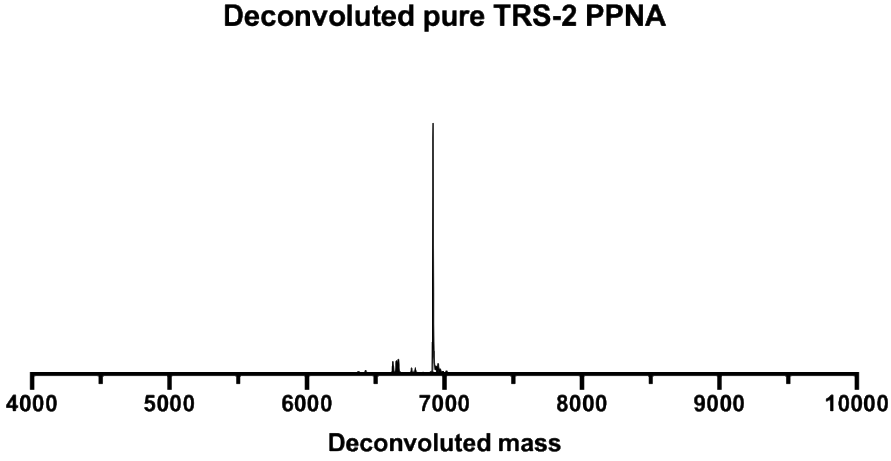


Figure 4.5, Entry 7. Sample: TRS-3 PPNA (Bpep)

Sequence: AGATTTTAAAGTTCGTTT-KKK-Bpep-$CONH_2$.

(Bpep = RXRRBRRXRRBR, X = 6-aminohexanoic acid; B = β-alanine)

Synthesis method: Automated flow synthesis, ~1.6 hours.

Resin: 20 mg Bpep loaded Rink Amide resin (3.5 μmmol).

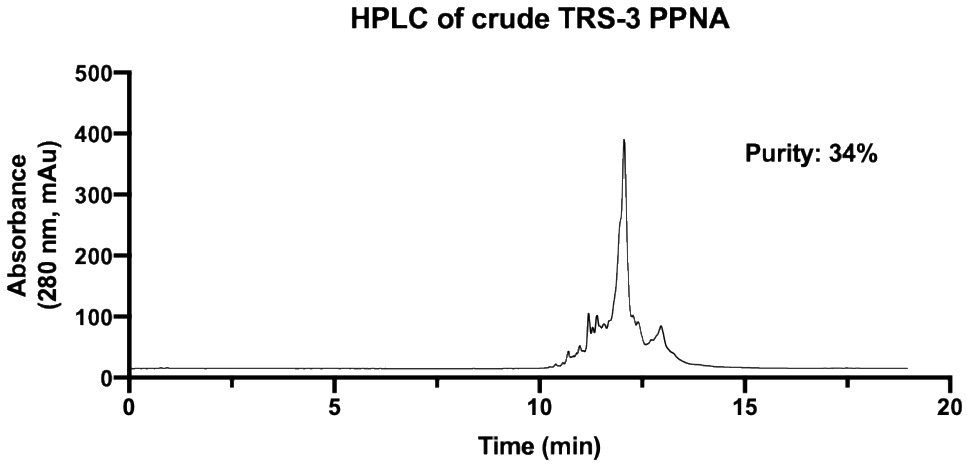
Crude purity: 34%

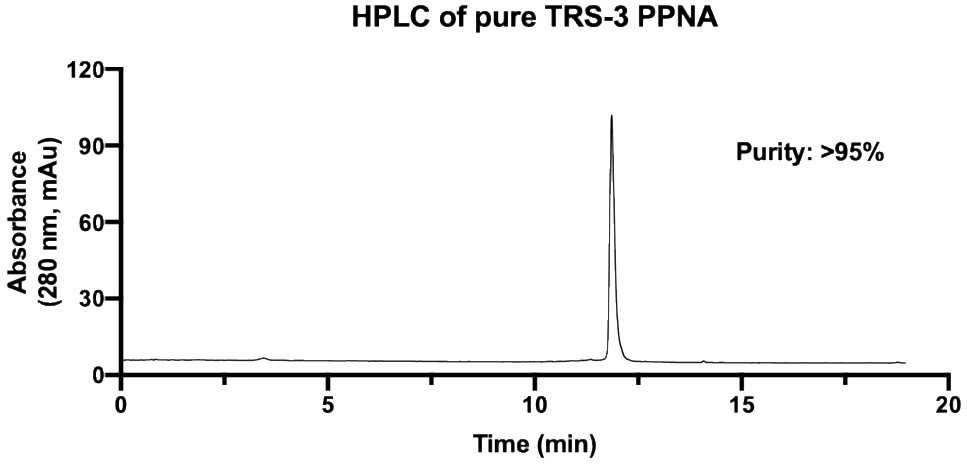
Pure purity: >95%

Obtained sample: 1.0 mg, 4% yield.

Purification method: method 1 and method 2.

HPLC method: Section 4.4.3



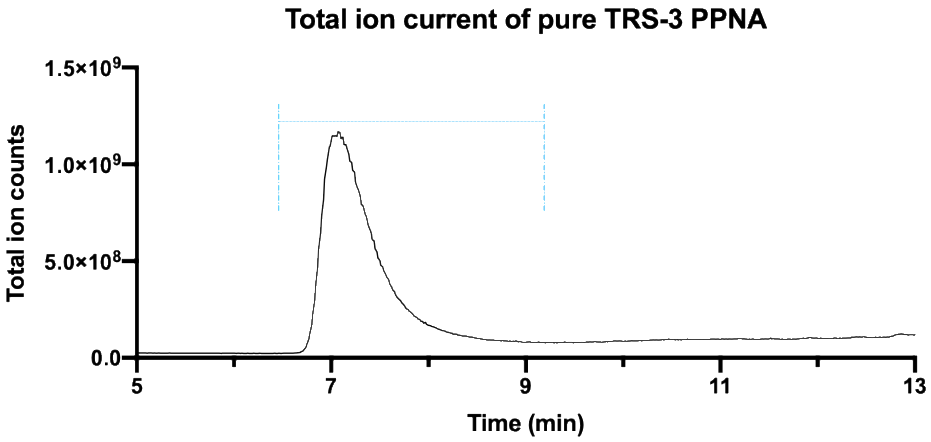


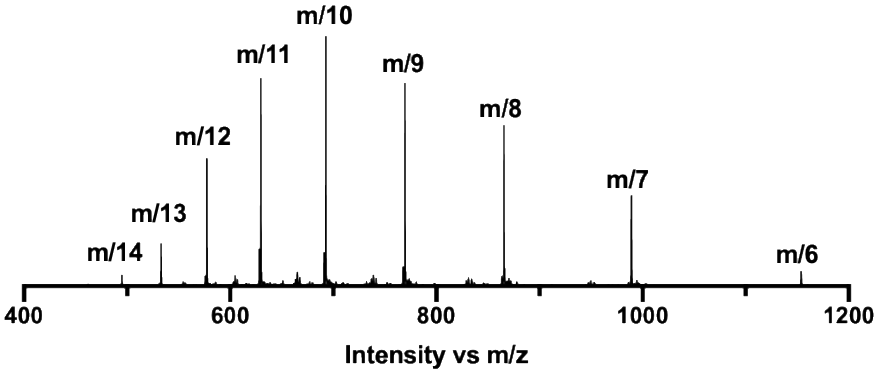
Pure sample after purification

LCMS method: condition 1

Calculated: 6913.27 Da

Observed: 6913.32 Da





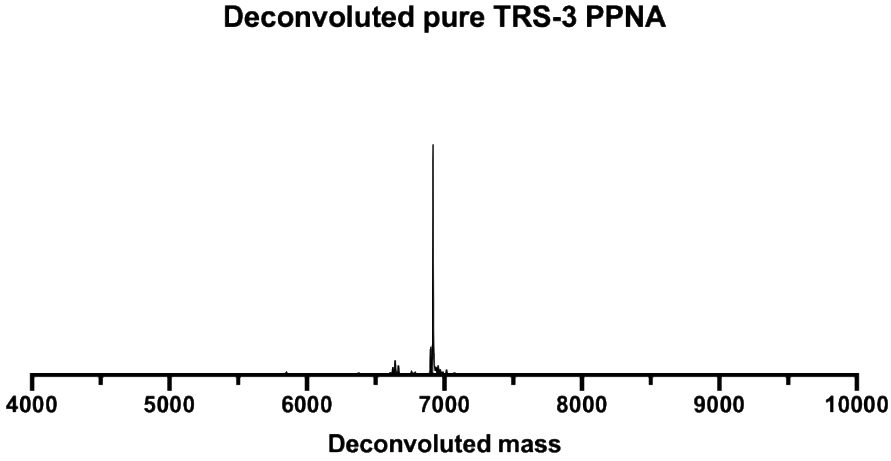


Figure 4.5, Entry 8. Sample: AUG PPNA (Bpep)

Sequence: CTCTCCATCTTACCTTTC-KKK-Bpep-$CONH_2$.

(Bpep = RXRRBRRXRRBR, X = 6-aminohexanoic acid; B = β-alanine)

Synthesis method: Automated flow synthesis, ~1.6 hours.

Resin: 20 mg Bpep loaded Rink Amide resin (3.5 μmmol).

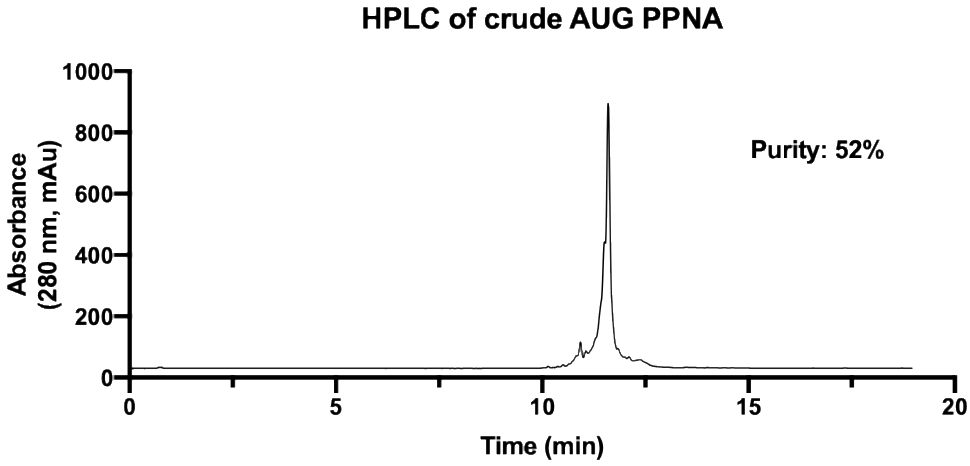
Crude purity: 52%

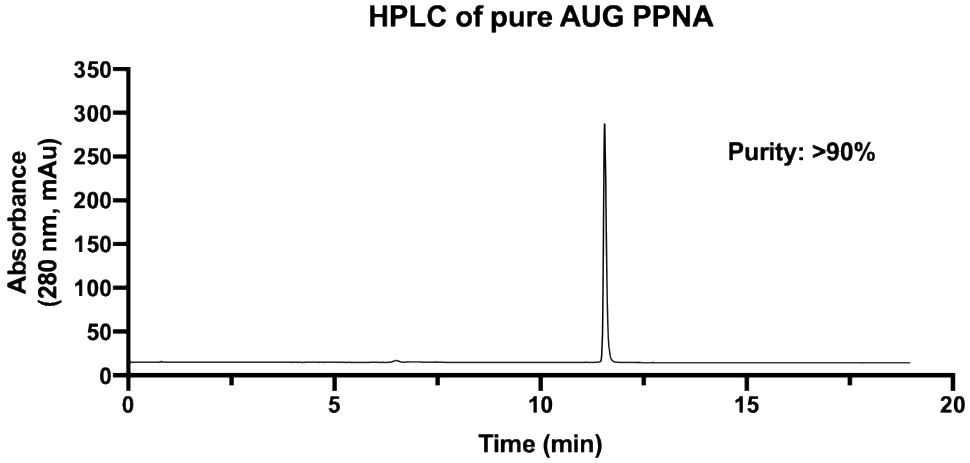
Pure purity: >90%

Obtained sample: 1.5 mg, 6.3% yield.

Purification method: method 1 and method 2.

HPLC method: Section 4.4.3





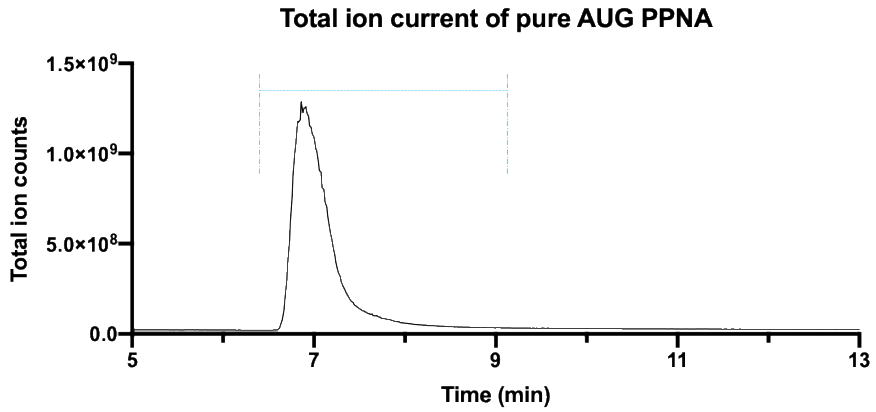
Pure sample after purification

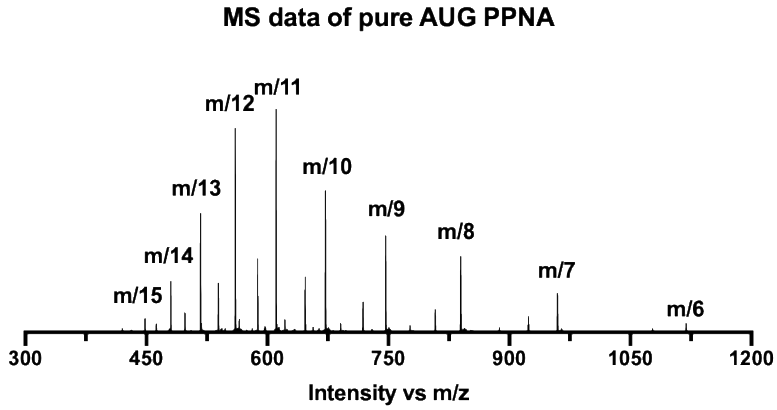
LCMS method: condition 1

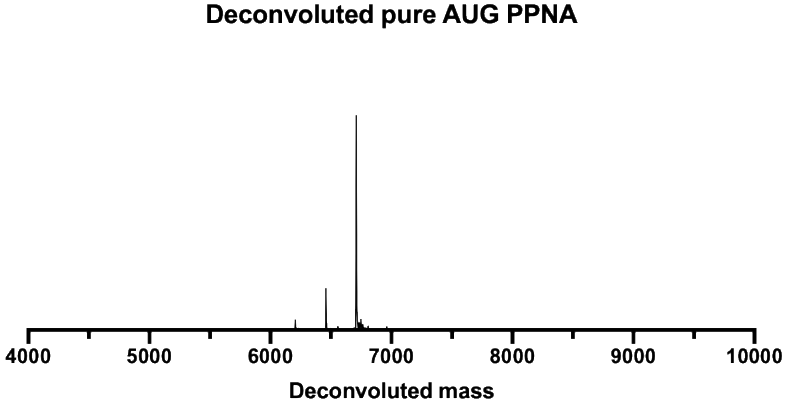
Calculated: 6706.22 Da

Observed: 6706.29 Da

Note: The major impurity is a single C-deletion side-product, which is hard to remove, even after two purifications.







4.4.15. LC-MS Data of purified EK1 peptide

A previously reported SARS-CoV-2 fusion inhibitor, EK1, was used as a positive control for the virus infection assay. Peptide was synthesized on an automated peptide synthesizer and purified with Biotage.

Figure 5, Positive control. Sample: EK1

Sequence: SLDQINVTFLDLEYEMKKLEEAIKKLEESYIDLKEL-$CONH_2$

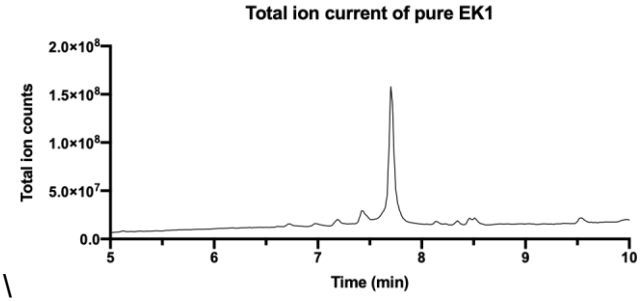
Synthesis method: automated peptide synthesis, 90 °C, HATU (or PyAOP), DIEA base, 20% Piperidine, 28 eq monomer, 1.5 hours.

Purification condition: Biotage Sfär C18 D column (Duo 100 Å 30 µm, 12 g), 5 to 60% acetonitrile over 12 column volume.

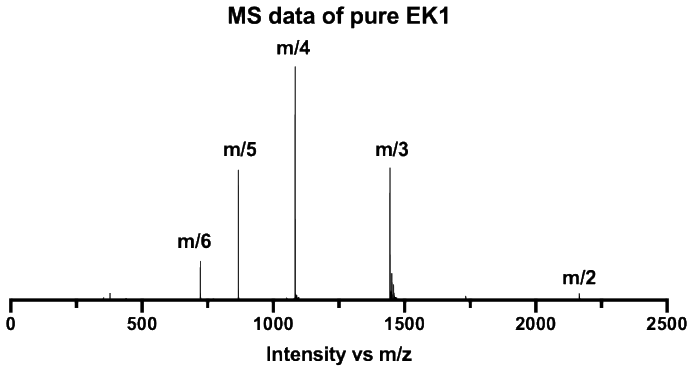
LCMS Method: Agilent Zorbax 300SB-C3 column, gradient: 1 to 91% acetonitrile over 15 mins.

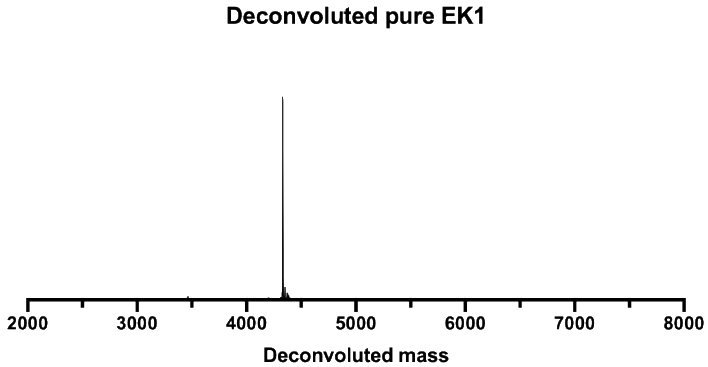
Calculated: 4328.26 Da

Observed: 4328.17 Da



\





4.5. Acknowledgments

Financial support for this work was provided by Novo Nordisk A/S (to B.L.P.). A.J.C. is supported by the Koch Institute MIT School of Science Fellowship in Cancer Research. C.E.F. is supported by the National Science Foundation (NSF) Graduate Research Fellowship program (4000057441). We thank Andrew Wilson at Detailed Dynamic for help with the design and manufacture of the automated instrument.

4.6.References

1. Syed, Y. Y., Eteplirsen: First Global Approval. Drugs 2016, 76 (17), 1699-1704.
2. Heo, Y.-A., Golodirsen: First Approval. Drugs 2020, 80 (3), 329-333.
3. Shirley, M., Casimersen: First Approval. Drugs 2021, 81 (7), 875-879.
4. Dhillon, S., Viltolarsen: First Approval. Drugs 2020, 80 (10), 1027-1031.
5. Prakash, V., Spinraza - A Rare Disease Success Story. Gene Ther. 2017, 24 (9), 497-497.
6. Sarepta Therapeutics Clinical Update: https://investorrelations.sarepta.com/node/20621/pdf.
7. Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O., Sequence-selective Recognition of DNA by Strand Displacement with a Thymine-Substituted Polyamide. Science 1991, 254 (5037), 1497-1500.
8. Siddiquee, S., A Review of Peptide Nucleic Acid. Adv. Tech. Bio. Med. 2015, 03 (02), 1-10.
9. Egholm, M.; Buchardt, O.; Christensen, L.; Behrens, C.; Freier, S. M.; Driver, D. A.; Berg, R. H.; Kim, S. K.; Norden, B.; Nielsen, P. E., PNA Hybridizes to Complementary Oligonucleotides Obeying the Watson–Crick Hydrogen-Bonding Rules. Nature 1993, 365 (6446), 566-568.
10. Demidov, V. V.; Potaman, V. N.; Frank-Kamenetskil, M. D.; Egholm, M.; Buchard, O.; Sönnichsen, S. H.; Nlelsen, P. E., Stability of Peptide Nucleic Acids in Human Serum and Cellular Extracts. Biochem. Pharmacol. 1994, 48 (6), 1310-1313.
11. Gupta, A.; Mishra, A.; Puri, N., Peptide Nucleic Acids: Advanced Tools for Biomedical Applications. J. Biotechnol. 2017, 259, 148-159.
12. Sharma, C.; Awasthi, S. K., Versatility of Peptide Nucleic Acids (PNAs): Role in Chemical Biology, Drug Discovery, and Origins of Life. Chem. Biol. Drug Des. 2017, 89 (1), 16-37.
13. Montazersaheb, S.; Hejazi, M. S.; Nozad Charoudeh, H., Potential of Peptide Nucleic Acids in Future Therapeutic Applications. Adv. Pharm. Bull. 2018, 8 (4), 551-563.
14. Alagpulinsa, D. A.; Yaccoby, S.; Ayyadevara, S.; Shmookler Reis, R. J., A Peptide Nucleic Acid Targeting Nuclear RAD51 Sensitizes Multiple Myeloma Cells to Melphalan Treatment. Cancer Biol. Ther. 2015, 16 (6), 976-986.
15. Hoban, M. D.; Orkin, S. H.; Bauer, D. E., Genetic Treatment of a Molecular Disorder: Gene Therapy Approaches to Sickle Cell Disease. Blood 2016, 127 (7), 839-848.
16. Maekawa, K.; Azuma, M.; Okuno, Y.; Tsukamoto, T.; Nishiguchi, K.; Setsukinai, K.-i.; Maki, H.; Numata, Y.; Takemoto, H.; Rokushima, M., Antisense Peptide Nucleic Acid - Peptide Conjugates for

Functional Analyses of Genes in Pseudomonas Aeruginosa. Bioorg. Med. Chem. 2015, 23 (22), 7234-7239.

17. Ahn, D.-G.; Lee, W.; Choi, J.-K.; Kim, S.-J.; Plant, E. P.; Almazán, F.; Taylor, D. R.; Enjuanes, L.; Oh, J.-W., Interference of Ribosomal Frameshifting by Antisense Peptide Nucleic Acids Suppresses SARS Coronavirus Replication. Antivir. Res. 2011, 91 (1), 1-10.

18. Zeng, Z.; Han, S.; Hong, W.; Lang, Y.; Li, F.; Liu, Y.; Li, Z.; Wu, Y.; Li, W.; Zhang, X.; Cao, Z., A Tat-conjugated Peptide Nucleic Acid Tat-PNA-DR Inhibits Hepatitis B Virus Replication In Vitro and In Vivo by Targeting LTR Direct Repeats of HBV RNA. Mol. Ther. Nucleic Acids 2016, 5, e295.

19. Quijano, E.; Bahal, R.; Ricciardi, A.; Saltzman, W. M.; Glazer, P. M., Therapeutic Peptide Nucleic Acids: Principles, Limitations, and Opportunities. Yale J. Biol. Med. 2017, 90 (4), 583-598.

20. Tailhades, J.; Takizawa, H.; Gait, M. J.; Wellings, D. A.; Wade, J. D.; Aoki, Y.; Shabanpoor, F., Solid-Phase Synthesis of Difficult Purine-Rich PNAs through Selective Hmb Incorporation: Application to the Total Synthesis of Cell Penetrating Peptide-PNAs. Front. Chem. 2017, 5 (81).

21. Rogers, F. A.; Lin, S. S.; Hegan, D. C.; Krause, D. S.; Glazer, P. M., Targeted Gene Modification of Hematopoietic Progenitor Cells in Mice Following Systemic Administration of a PNA-peptide Conjugate. Mol. Ther. 2012, 20 (1), 109-118.

22. Li, Q.; Xu, M.; Cui, Y.; Huang, C.; Sun, M., Arginine-rich Membrane-permeable Peptides are Seriously Toxic. Pharmacol. Res. Perspect. 2017, 5 (5), e00334.

23. Schissel, C. K.; Mohapatra, S.; Wolfe, J. M.; Fadzen, C. M.; Bellovoda, K.; Wu, C.-L.; Wood, J. A.; Malmberg, A. B.; Loas, A.; Gómez-Bombarelli, R.; Pentelute, B. L., Interpretable Deep Learning for De Novo Design of Cell-Penetrating Abiotic Polymers. bioRxiv 2020, DOI:10.1101/2020.04.10.036566.

24. Wu, F.; Zhao, S.; Yu, B.; Chen, Y.-M.; Wang, W.; Song, Z.-G.; Hu, Y.; Tao, Z.-W.; Tian, J.-H.; Pei, Y.-Y.; Yuan, M.-L.; Zhang, Y.-L.; Dai, F.-H.; Liu, Y.; Wang, Q.-M.; Zheng, J.-J.; Xu, L.; Holmes, E. C.; Zhang, Y.-Z., A New Coronavirus Associated with Human Respiratory Disease in China. Nature 2020, 579 (7798), 265-269.

25. Wang, H.; Li, X.; Li, T.; Zhang, S.; Wang, L.; Wu, X.; Liu, J., The Genetic Sequence, Origin, and Diagnosis of SARS-CoV-2. Eur. J. Med. Chem. 2020, 39 (9), 1629-1635.

26. Hansen, A. M.; Bonke, G.; Hogendorf, W. F. J.; Björkling, F.; Nielsen, J.; Kongstad, K. T.; Zabicka, D.; Tomczak, M.; Urbas, M.; Nielsen, P. E.; Franzyk, H., Microwave-assisted Solid-phase Synthesis of Antisense acpP Peptide Nucleic Acid-peptide Conjugates Active Against Colistin- and Tigecycline-resistant E. coli and K. pneumoniae. Eur. J. Med. Chem. 2019, 168, 134-145.

27. Mayfield, L. D.; Corey, D. R., Automated Synthesis of Peptide Nucleic Acids and Peptide Nucleic Acid–Peptide Conjugates. Anal. Biochem. 1999, 268 (2), 401-404.

28. Gogoi, K.; Mane, M. V.; Kunte, S. S.; Kumar, V. A., A Versatile Method for the Preparation of Conjugates of Peptides with DNA/PNA/Analog by Employing Chemo-selective Click Reaction in Water. Nucleic Acids Res. 2007, 35 (21), e139.

29. Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B., A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective “Ligation” of Azides and Terminal Alkynes. Angew. Chem. Int. Ed. 2002, 41 (14), 2596-2599.

30. Sazani, P.; Kang, S.-H.; Maier, M. A.; Wei, C.; Dillman, J.; Summerton, J.; Manoharan, M.; Kole, R., Nuclear Antisense Effects of Neutral, Anionic and Cationic Oligonucleotide Analogs. Nucleic Acids Res. 2001, 29 (19), 3965-3974.

31. 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.

32. 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.

33. Braasch, D. A.; Nulf, C. J.; Corey, D. R., Synthesis and Purification of Peptide Nucleic Acids. Curr. Protoc. Nucleic Acid Chem. 2002, 9 (1), 4.11.1-4.11.18.

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

35. Egholm, M. B., O.; Nielsen, P. E.; Berg, R. , Peptide Nucleic Acids (PNA). Oligonucleotide Analogues with an Achiral Peptide Backbone. J. Am. Chem. Soc. 1992, (144), 1895-1897.

36. Pooga, M.; Soomets, U.; Hällbrink, M.; Valkna, A.; Saar, K.; Rezaei, K.; Kahl, U.; Hao, J.-X.; Xu, X.-J.; Wiesenfeld-Hallin, Z.; Hökfelt, T.; Bartfai, T.; Langel, Ü., Cell Penetrating PNA Constructs Regulate Galanin Receptor Levels and Modify Pain Transmission in Vivo. Nat. Biotechnol. 1998, 16 (9), 857-861.

37. Gait, M. J.; Arzumanov, A. A.; McClorey, G.; Godfrey, C.; Betts, C.; Hammond, S.; Wood, M. J. A., Cell-Penetrating Peptide Conjugates of Steric Blocking Oligonucleotides as Therapeutics for Neuromuscular Diseases from a Historical Perspective to Current Prospects of Treatment. Nucleic Acid Ther. 2018, 29 (1), 1-12.

38. Fadzen, C. M.; Holden, R. L.; Wolfe, J. M.; Choo, Z. N.; Schissel, C. K.; Yao, M.; Hanson, G. J.; Pentelute, B. L., Chimeras of Cell-Penetrating Peptides Demonstrate Synergistic Improvement in Antisense Efficacy. Biochemistry 2019, 58 (38), 3980-3989.

39. 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.

40. Madhugiri, R.; Fricke, M.; Marz, M.; Ziebuhr, J., RNA Structure Analysis of Alphacoronavirus Terminal Genome Regions. Virus Res. 2014, 194, 76-89.

41. Masters, P. S., Coronavirus Genomic RNA Packaging. Virology 2019, 537, 198-207.

42. Burrer, R.; Neuman, B. W.; Ting, J. P. C.; Stein, D. A.; Moulton, H. M.; Iversen, P. L.; Kuhn, P.; Buchmeier, M. J., Antiviral Effects of Antisense Morpholino Oligomers in Murine Coronavirus Infection Models. J. Virol. 2007, 81 (11), 5637-5648.

43. Neuman, B. W.; Stein, D. A.; Kroeker, A. D.; Paulino, A. D.; Moulton, H. M.; Iversen, P. L.; Buchmeier, M. J., Antisense Morpholino-Oligomers Directed against the 5′ End of the Genome Inhibit Coronavirus Proliferation and Growth. J. Virol. 2004, 78 (11), 5891-5899.

44. Neuman, B. W.; Stein, D. A.; Kroeker, A. D.; Churchill, M. J.; Kim, A. M.; Kuhn, P.; Dawson, P.; Moulton, H. M.; Bestwick, R. K.; Iversen, P. L.; Buchmeier, M. J., Inhibition, Escape, and Attenuated Growth of Severe Acute Respiratory Syndrome Coronavirus Treated with Antisense Morpholino Oligomers. J. Virol. 2005, 79 (15), 9665-9676.

45. Rosenke, K.; Leventhal, S.; Moulton, H. M.; Hatlevig, S.; Hawman, D.; Feldmann, H.; Stein, D. A., Inhibition of SARS-CoV-2 in Vero Cell Cultures by Peptide-conjugated Morpholino Oligomers. J. Antimicrob. Chemother. 2021, 76 (2), 413-417.

46. Xia, S.; Yan, L.; Xu, W.; Agrawal, A. S.; Algaissi, A.; Tseng, C.-T. K.; Wang, Q.; Du, L.; Tan, W.; Wilson, I. A.; Jiang, S.; Yang, B.; Lu, L., A Pan-coronavirus Fusion Inhibitor Targeting the HR1 Domain of Human Coronavirus Spike. Sci. Adv. 2019, 5 (4), eaav4580.

47. Xia, S.; Liu, M.; Wang, C.; Xu, W.; Lan, Q.; Feng, S.; Qi, F.; Bao, L.; Du, L.; Liu, S.; Qin, C.; Sun, F.; Shi, Z.; Zhu, Y.; Jiang, S.; Lu, L., Inhibition of SARS-CoV-2 (previously 2019-nCoV) Infection by a Highly Potent Pan-coronavirus Fusion Inhibitor Targeting Its Spike Protein that Harbors a High Capacity to Mediate Membrane Fusion. Cell Res. 2020, 30 (4), 343-355.

48. Lakshman, M. K.; Frank, J. A Simple Method for C-6 Modification of Guanine Nucleosides. *Org. Biomol. Chem.* **2009,** *7*, 2933-2940.