ANY WAY YOU WANT IT: PUSHING THE LIMITS OF CHEMICAL PROTEIN SYNTHESIS

by

Michael Thomas Jacobsen

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biochemistry

The University of Utah

December 2015

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Michael Thomas Jacobsen**

has been approved by the following supervisory committee members:

Michael S. Kay	, Chair	**06/22/15** Date Approved
Wesley I. Sundquist	, Member	**06/22/15** Date Approved
Christopher Peter Hill	, Member	**06/22/15** Date Approved
David W. Grainger	, Member	**06/22/15** Date Approved
Adam Frost	, Member	**06/22/15** Date Approved

and by **Wesley I. Sundquist** , Chair/Dean of

the Department/College/School of **Biochemistry**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Chemical protein synthesis, via solid-phase peptide synthesis and chemoselective ligation of peptides, is a powerful approach for preparing peptides and proteins. These techniques enable complete atomic control over protein composition with both mechanistic and practical applications for biochemistry.

The foundation of this dissertation is formed by two ambitious chemical protein synthesis projects: DapA (Chapter 2) and Dpo4 (Chapter 5). DapA is a 312-residue protein whose folding depends on the well-studied chaperone GroEL/ES. The successful synthesis of DapA (in both L- and D- chirality) was used to demonstrate cross-chiral folding by GroEL/ES—a fundamental biological insight and potential tool for future mirror-image synthetic biology research.

However, the record-breaking synthesis of DapA was an arduous process requiring tremendous human and technical resources. The lessons from this project were then applied to the synthesis of our next target, Dpo4 (352 residues). Dpo4 is one of the shortest DNA/RNA polymerases, providing an accessible synthetic tool to amplify DNA/RNA for future synthetic biology studies. A new concept termed DOPPEL (Diversity-based Optimization of Peptide Properties to Enhance Ligations) was used to simplify this synthesis. Furthermore, various synthesis strategies and general advice for completing

mega-synthesis projects in the future are detailed in this chapter.

In both the DapA and Dpo4 projects, one of the most prominent challenges was the handling of poorly soluble peptide segments. These poorly soluble peptides can lead to dramatic yield losses and additional complexities.

In the third major synthesis project of this thesis, GroES, we overcame an even greater insolubility challenge. This 97-residue protein could not be synthesized due to extreme solubility challenges with its C-terminal half. In response to this challenge, a new chemical tool was developed to link a solubilizing peptide (“Helping Hand”) to the C-terminal half of the protein. Key to this approach is a new synthetic building block, Fmoc-Ddae-OH, which is easy to synthesize, incorporate, and cleanly remove once the solubilizing function is complete.

Overall, this dissertation pushes the limits of chemical peptide and protein synthesis, and provides exciting directions for the next wave of biochemists looking to use chemical protein synthesis to study interesting problems and engineering challenges.

TABLE OF CONTENTS

ABSTRACT iii

LIST OF TABLES vii

LIST OF FIGURES viii

ACKNOWLEDGMENTS xii

Chapters

1. INTRODUCTION TO CHEMICAL PROTEIN SYNTHESIS 1

    1.1 The Advantages of Chemical Protein Synthesis 1
    1.2 Native Chemical Ligation: The Critical Advance in CPS 4
    1.3 Dependence of NCL on an N-terminal Terminal Cysteine 5
    1.4 Facile Preparation of Peptide Thioesters by Fmoc-SPPS 7
    1.5 Dissertation Overview 8
    1.6 References 11

2. SYNTHESIS AND FOLDING OF A MIRROR-IMAGE ENZYME REVEALS AMBIDEXTROUS CHAPERONE ACTIVITY 15

    2.1 Abstract 16
    2.2 Significance 16
    2.3 Introduction 16
    2.4 Results 17
    2.5 Discussion 19
    2.6 Methods and Materials 19
    2.7 References 20
    2.8 Supporting Information 22

3. DESIGN AND CHARACTERIZATION OF EBOLAVIRUS GP PREHAIRPIN INTERMEDIATE MIMICS AS DRUG TARGETS 29

    3.1 Abstract 30
    3.2 Introduction 31
    3.3 Results and Discussion 33

3.4 Conclusion 40
3.5 Materials and Methods 40
3.6 References 45
3.7 Supporting Information 48

4. “HELPING HANDS” FOR SIMPLIFYING CHEMICAL PROTEIN SYNTHESIS: PREPARATION OF THE CHALLENGING GROES 56

4.1 Abstract 57
4.2 Introduction 58
4.3 Results 63
4.4 Discussion and Future Plans 74
4.5 Acknowledgments 75
4.6 Materials and Methods 75
4.7 References 85

5. TOWARDS THE TOTAL CHEMICAL SYNTHESIS OF A 352-RESIDUE DNA POLYMERASE 90

5.1 Abstract 91
5.2 Introduction 92
5.3 Results 98
5.4 Future Plans 121
5.5 Acknowledgments 122
5.6 Materials and Methods 122
5.7 References 155

LIST OF TABLES

2.S.1 Absorbance values (A562) at specified time points in refolding assays……..28

3.1 Biophysical analyses of N-trimer mimics via CD and AUC…….....34

3.S.1 eboIZN21 crystallographic data and refinement statistics……... 55

5.1 Mass spec summary of the initial nine peptides…….....128

LIST OF FIGURES

1.1 Mechanism of native chemical ligation……..... 10

2.1 Total chemical synthesis of 312-residue DapA……..... 17

2.2 Validation of the DapA A77C mutation…… 18

2.3 Analysis of synthetic unfolded L- and D-DapA……..... 18

2.4 Structural and functional characterization of synthetic folded L- and D-DapA…… 19

2.5 GroEL/ES-mediated refolding of synthetic L- and D-DapA……..... 19

2.S.1 Initial synthetic strategy for total chemical synthesis of 312-residue DapA…….. 24

2.S.2 Hydrolysis of DapA 2 thioester…… 25

2.S.3 Yields obtained during the total chemical synthesis of 312-residue L-DapA……..... 26

2.S.4 Yields obtained during the total chemical synthesis of 312-residue D-DapA……..27

2.S.5 Arginine-assisted refolding of DapA……27

2.S.6 SEC purification of DapA following arginine-assisted folding……..... 28

3.1 Model for membrane fusion mediated by enveloped virus surface glycoproteins……....31

3.2 Conservation of the ebolavirus GP N-trimer and design of peptide N-trimer mimics……....32

3.3 Biophysical analyses of ebolavirus N-trimer mimics……..33

3.4 Binding of the ebolavirus C-peptide to the N-trimer mimic……....34

3.5 Crystal structure of eboIZN21……..... 35

3.6 Validation of eboIZN39IQ as a phage display target…… 36

3.7 Validation of eboIZN21 as a phage display target…… 37

3.8 Comparing the two ebolavirus N-trimer mimics as phage display targets…….....38

3.9 Inhibition of filovirus entry by eboIZN39IQ…….....39

3.S.1 Gel filtration analysis of 240 uM eboIZN21…… 48

3.S.2 Binding of the ebolavirus C-peptide to the N-trimer mimic…….. 49

3.S.3 Hydrophobic interactions between N21 residues in the unliganded eboIZN21structure……50

3.S.4 Comparison of hydrophobic pockets in ebolavirus and HIV N-trimers…….. 51

3.S.5 Mirror-image phage display…….....52

3.S.6 Synthesis of D-eboIZN39IQ…….....53

3.S.7 Biophysical characterization of the D-versions of the Ebola N-trimer mimics……...54

4.1 Semi-permanent solubilizing methods…… 77

4.2 The challenging GroES synthesis……78

4.3 HMBA-based method to prepare GroES…… 79

4.4 Permanent poly-His method to prepare GroES…… 80

4.5 Fmoc-Lys(N3-Dtpp) approaches to helping hand..……..... 81

4.6 Fmoc-Ddae helping hand approach…… 82

4.7 Fmoc-Ddae method using C20…… 83

4.8 Synthesis of GroES using Fmoc-Ddae-OH approach……..... 84

5.1 Dpo4 crystal structure……..125

5.2 Initial breakdown of the 352-residue Dpo4……..... 126

5.3 HPLC of the initial nine peptides……... 127

5.4 Analysis of two additional scouting peptides……129

5.5 Final six-segment breakdown of the 352-residue Dpo4…… 130

5.6 Relationship between peptide segments and minimum total number of purifications……... 131

5.7 HPLC on the final six crude peptides…… 132

5.8 Analysis of pure [1-2] peptide…… 133

5.9 Analysis of pure [3] peptide…… 134

5.10 Analysis of pure [4] peptide…… 135

5.11 Analysis of pure [5-6] peptide…… 136

5.12 Analysis of pure [7-8] peptide…… 137

5.13 Analysis of pure [8-9] peptide…… 138

5.14 Optimization of [1-2] and [5-6] crude peptides…… 139

5.15 Optimization of [2] and convergent assembly…… 140

5.16 Identification of the His tag issue……..... 141

5.17 Additional PCR on recombinant constructs……..... 142

5.18 Assembly strategies for handling one cysteine……143

5.19 DOPPEL analysis at Cys31…… 144

5.20 Chemical assembly strategies for N-terminal half…….. 145

5.21 DOPPEL analysis at Ile101…… 146

5.22 Activity analysis of the double mutant (C31A/I101M) ……147

5.23 Chemical assembly strategies for C-terminal half……...148

5.24 Synthesis of [5-6] and [7-8] by Dawson and hydrazide

methods…….. 149

5.25 Final synthesis strategy for Dpo4……..... 150

5.26 MS on the N-terminal intermediates……..... 151

5.27 MS on the C-terminal intermediates……..... 152

5.28 MS on the key intermediates and final product……..... 153

5.29 Dpo4 future directions……..154

ACKNOWLEDGMENTS

The most important person I have to thank for this great experience is my advisor, Dr. Michael Kay. As everyone who has met him knows, he is very clever and scientifically innovative, but he is also a very caring and thoughtful individual. I especially thank him for introducing me to Ms. Elsie and Cynthia. I strongly thank Dr. Kay for his rescue mission to “return my scientific mojo”; I think he was successful in that respect.

The second person to acknowledge is Matthew Weinstock. Our partnership during the first half of my time in Dr. Kay’s lab was highly productive, both scientifically and personally. Matt was a great counterbalance to my persistent scientific personality. He was the “Professor Xavier” to my “Magneto.” I hope we can work together again someday.

I must acknowledge Vincent Aucagne (CNRS) for his active collaboration on the *coup de main*. Our continuing collaboration, and open sharing of ideas, has been very inspiring. I truly value our relationship.

I also wish to thank other members of Kay lab with whom I worked during this time. First, I acknowledge Mark Petersen for providing a spark of chemical expertise and rationale, as well as a fun personality. Next, I thank Debbie Eckert for providing wisdom and writing expertise, as well as advice on personal challenges. I have also enjoyed working directly with Sarah Apple, David Judd,

Tracy Clinton, Nick Francis, Nico Szabo, Maya Pandya, and Amanda Smith.

One very important element to success was the collegiality provided by the EEJMRB 3rd Floor. This was very helpful to me when I had to switch to "cloning mode" during the Dpo4 project. In particular, I found Dave Kemble, Frank Whitby, Katherine Ferrell, Vish Chandrasekaran, Steve Alam, Matthew Lalonde, John McCullough, and Jack Skalicky to be very helpful and insightful. Along these lines, I need to thank my Mass Spec friends: Rob Marquardt, Scott Antonetti (Mr. Scotch), James Cox (Dr. Bourbon), and Mario Gonzales.

I was also extremely impressed by the efficiency and friendliness of the Biochemistry staff. They are the best: Dawn Ashment, Rachel Bookman, Diana Sperry, Kay Willden, Jared Kirby, Linda VanOrden, and Jill Wilson. I think it is important to compliment the leadership of Wes Sundquist and Chris Hill. During my time in the Department of Biochemistry, I have been very impressed by the overall efficiency and camaraderie. I think that this attitude starts at the top, and I greatly respect their leadership in this regard.

I also thank Wes, Chris, and Adam Frost for their time on my committee. In addition to their general scientific advice and critiques, I also appreciate their respect for the less prominent topics that I discuss on Chemical Protein Synthesis, which they have (or is it get?) to review during our committee meetings together. I especially thank Dave Grainger for serving on my committee and for his personal mentoring. I wish we had more time together, as there is still so much that I can learn from him.

CHAPTER 1

INTRODUCTION TO CHEMICAL PROTEIN SYNTHESIS

1.1 The Advantages of Chemical Protein Synthesis

Chemical Protein Synthesis (CPS) can be simply defined as the total chemical synthesis of proteins without any recombinantly expressed components. This dissertation discusses the application and simplification of CPS toward studying interesting biological problems. There are three key advantages that CPS provides over classical recombinant expression:

(1) Direct accessibility to mirror-image D-peptide and proteins: Peptides and proteins composed of D-amino acids are attractive therapeutics due to their resistance to natural L-proteases[1]. D-peptides are the foremost research interest of the Kay lab, where they have developed a trimeric D-peptide inhibitor of HIV entry that possesses pM potency[2]. This group is also currently developing inhibitors to Ebola[3] and other viruses with similar mechanisms of cell entry[4]. The critical method that is needed to develop targeted, mirror-image D-peptide inhibitors is mirror-image phage display[5]. In this method, the target of interest is prepared in D-chirality and then screened against a natural (L-) phage library. The selected phage that bind the target are then sequenced. These peptide sequences are next synthesized in D-chirality.

Based on the law of symmetry, these D-peptides will bind to the natural L-target. A critical first step to using this method is the total chemical synthesis of the D-target. There is currently no other reliable method for producing D-proteins, although recombinant incorporation of individual D-residues into proteins has been achieved (review of methods for genetic code expansion[6]). Another notable application of D-proteins is the versatility that it can add to crystallization conditions. As has been reviewed[7], the availability of racemic protein mixtures provides access to an extended array of achiral space groups, in particular the P1 space group. Several examples from the Kent lab have highlighted this advantage for protein crystallization[8].

(2) Preparation of site-specifically modified proteins for biological studies: Although the preparation of full-length proteins (>300 aa) by total chemical synthesis is not yet routine (see discussion below), an interesting work-around to this challenge has been developed: expressed protein ligation (EPL)[9]. With this technology[10], semisynthetic proteins can be prepared by reaction of recombinantly expressed proteins with synthetic segments. Some recent demonstrations include fundamental studies with ubiquitin[11], alpha-synuclein[12], histones[13], and membrane proteins[14]. Although this intriguing method has been highly successful, this dissertation focuses on the challenges and applications of total chemical synthesis.

(3) Preparation of proteins with entirely new topologies: In addition to facilitating the incorporation of site-specific modifications, CPS provides an approach for creating new protein topologies. Specifically, protein domains can be stitched

together in new ways analogous to those used in polymer chemistry. For example, scaffolds or templates can be prepared that then anchor several protein domains via chemoselective and site-specific attachment[15].

The current standard approach for performing CPS combines two technologies: (1) solid-phase peptide synthesis (SPPS)[16] to produce peptide segments followed by (2) a chemoselective ligation strategy[17] to connect the peptide segments. In SPPS, peptides are prepared using couple-deprotection cycles for sequential introduction of individual protected amino acids (building on a solid support from the C- to N-terminus). Final deprotection and cleavage from resin is typically achieved using either a moderate or strong acid, depending on the particular protecting groups and resin. The repetitive nature of SPPS makes it amenable to automation, greatly accelerating the rate of peptide production. However, the accumulation of byproducts (e.g., terminations, deletions, and modifications) during synthesis generally limits SPPS to ~50 aa[18], although some prominent, highly-optimized examples have reported the successful synthesis (and purification) of 188[19] and 140 residue peptides[20].

Importantly, peptides produced by SPPS must be purified to a high degree and be well-characterized by analytical HPLC and MS. To routinely advance beyond the ~50 aa limit of SPPS, chemoselective methods[17, 21] are needed to selectively ligate these purified peptides.

1.2 Native Chemical Ligation: The Critical Advance in CPS

The original attempts to ligate peptides required protection of all internal reactive groups, so that selective amide formation could be performed at the peptide termini. However, this approach was significantly hindered by the poor solubility of protected peptide fragments[22]. The solubilities of protected peptides can be hard to predict; in some cases, protected peptides are totally insoluble[23]. Other challenges associated with ligating protected peptides included racemization at the activated ligation site[24] and the obvious difficulty of performing high-resolution purification on poorly soluble intermediates[21b]. A key discovery toward overcoming this challenge was the development of the "chemical ligation principle", where unprotected peptide segments could be chemoselectively ligated. Here, a peptide containing a C-terminal thioacid was selectively reacted with a peptide containing an N-terminal bromoacetyl group to form a chemically analogous thioester bond product. Although this approach provided an important advance to achieving soluble and purifiable ligation products, the reaction product contained a non-native thioester bond.

The key discovery in 1994 of Native Chemical Ligation (NCL) by Dawson and Kent solved these handling problems with protected peptides[25]. **Fig 1** shows the mechanism of NCL. Here, a peptide containing a C-terminal thioester can react with another thiol by reversible thiol-thioester exchange. In the presence of another peptide with an N-terminal thiol group, a reversible peptide-peptide thioester intermediate forms. This peptide-peptide intermediate can then undergo an irreversible S ->N acyl shift (likely via 5-membered ring intermediate) to form a

natural amide product.

Some key benefits of NCL include:

- The use of highly solubilizing and denaturing conditions, which (generally) allows for relatively high concentrations (several mM) of the starting peptides, making the reaction highly kinetically favorable,
- Straightforward HPLC and MS methods (generally) can be used to purify and analyze the stable ligation product,
- The two reacting peptides can be prepared, and then stored, in a relatively easy fashion. The peptide with an N-terminal Cys can be directly prepared by SPPS, while preparation of the C-terminal thioester peptide is also achievable directly by Boc-SPPS[26]. Preparation of the thioester peptides by Fmoc-SPPS is more challenging and will be discussed below in detail.

Thus, the development of NCL was critical to the field of peptide synthesis, as it provides an effective and clean procedure for producing synthetic proteins. However, as detailed in the next sections below, several challenges remain for the production of synthetic proteins to become routine.

1.3 Dependence of NCL on an N-terminal Terminal Cysteine

As introduced, NCL[25] possessed a significant limitation in its dependence on a Cys residue at the ligation site—as Cys is the least common residue in proteins. This dependence on Cys is a bottleneck, as many proteins do not contain Cys residues. Alternatively, some proteins that do contain Cys residues possess long stretches (>100 aa) that lack these residues. This problem has

been addressed in two general ways: the ligation-desulfurization strategy and a removable auxiliary approach. In the less common auxiliary approach[27], a C-terminal peptide is generated with an N-terminal auxiliary group (typically on a Gly residue) containing a thiol for reaction in NCL. Following the NCL reaction, the auxiliary group is then removed to generate a native peptide. The lower utility of this method is likely due to the slower kinetics observed in these ligations.

On the other hand, the ligation-desulfurization approach (pioneered by Dawson) has proven to be a reliable method for expanding NCL sites[28]. In the first iteration of this concept, NCL is still performed using Cys-based junctions; however, following ligation, the reaction site cysteine is desulfurized to an Ala residue using Ni metal. In other words, this method expands the suite of NCL sites to include Ala junctions. Danishefsky later refined this procedure by developing more gentle and robust metal-free reaction conditions by applying a free radical mechanism[29].

Addition of Ala to the NCL suite provided a huge benefit to the field, but even broader utility was provided in the next iteration of the ligation-desulfurization concept. As suggested in Dawson's original paper[28], even more ligation sites could be gained by introducing non-natural β- or γ-mercapto-thiol amino acids, which could then be desulfurized back to a native residue following ligation. This concept was first applied using a β-mercapto-phenylalanine[30]. It has since expanded[31] to several residues including Asp, Glu, Lys, Leu, Pro, Gln, Arg, Thr, Val, and Trp; however, the preparation of these building blocks is chemically demanding, with most requiring 7-16 synthetic steps. Unfortunately, only two of

these custom non-natural residues are commercially available: penicillamine (thio-Valine)[32] and thio-proline[33]. Thus, at this stage, significant progress has been made to expand past the dependence on Cys junctions, but much of this work is still relatively inaccessible.

1.4 Facile Preparation of Peptide Thioesters by Fmoc-SPPS

The preparation of thioesters by Boc-SPPS is relatively straightforward by simple incorporation of a mercaptopropionic acid building block on-resin followed by subsequent thioester formation[34]. This thioester is stable during the synthesis and cleavage steps. However, Boc-SPPS, compared to Fmoc-SPPS, is not ideal for routine laboratory use for the following reasons:

(1) Boc-SPPS requires trifluoroacetic acid (TFA) handing at every deprotection step of the synthesis (in contrast to just one time for cleavage in Fmoc-SPPS),

(2) Boc-SPPS requires highly corrosive and toxic hydrogen fluoride (HF) during the cleavage step[26],

(3) The harsh acidic conditions needed for Boc-SPPS (both for deprotection and cleavage) limits its use in the preparation of glycosylated and other custom amino acids[35].

Thus, we (and most other peptide labs) employ Fmoc-SPPS for routine peptide synthesis. However, at this time, thioesters cannot be routinely prepared using Fmoc-SPPS due to their instability to Fmoc deprotection conditions[36]. To address this problem, various crypto thioester/surrogate methods have been

developed (recently reviewed[37],[38]). However, most of these methods suffer from one of two problems: very slow kinetics to generate the thioester species or the requirement for specialized chemical syntheses to generate the thioester device. In contrast, the recently developed peptide hydrazide method[39] has provided a robust approach to preparing peptide hydrazides. In the peptide hydrazide method, thioesters are not directly assembled on-resin. Instead, they are later formed by activation of the hydrazide into an azide followed by in-situ thiolysis to generate a thioester in solution. This method has become the routine technique for all Kay lab synthesis projects. Additional details on the strengths and weaknesses of these methods are elaborated in great detail throughout the chapters.

1.5 Dissertation Overview

The primary goal of this dissertation has been to push the limits of chemical peptide and protein synthesis for studying interesting biological problems. However, along the way to pushing the limits, new tools and strategies were developed for simplifying the entire chemical synthesis process.

Indeed, Chapter 2 (“Synthesis and Folding of a Mirror-image Enzyme Reveals Ambidextrous Chaperone Activity”) starts with the description of our first successful protein chemical synthesis project (the 312-residue DapA protein). This published report describes how the natural chaperone GroEL/ES is ambidextrous, in that it is capable of folding both L- and D- chemically synthesized versions of DapA, a model chaperone client protein.

Next, Chapter 3 (“Design and Characterization of Ebolavirus GP Prehairpin Intermediate Mimics as Drug Targets”) provides a detour to shorter peptide-based approaches to D-peptide inhibitors of Ebolavirus cell entry.

Returning to CPS, Chapter 4 (“’Helping Hands’ for Simplifying Chemical Protein Synthesis: Preparation of the Challenging GroES”) introduces the synthetically challenging 97-residue GroES protein. The foundation of this challenge is the very poorly soluble C-terminal region of the protein, which could not be adequately purified in order to achieve the total synthesis. This problem was solved by the development of a straightforward method for temporarily introducing, and then cleaving, a solubilizing peptide (“Helping Hand”). This new tool for introducing helping hands, termed Fmoc-Ddae-OH, should be highly useful in the field.

Lastly, Chapter 5 (“Towards the Total Chemical Synthesis of a 352-residue DNA Polymerase”) details the sophisticated development of a six-piece synthesis strategy. Full-length synthetic protein was achieved using a methodical approach to breaking down this very challenging project.

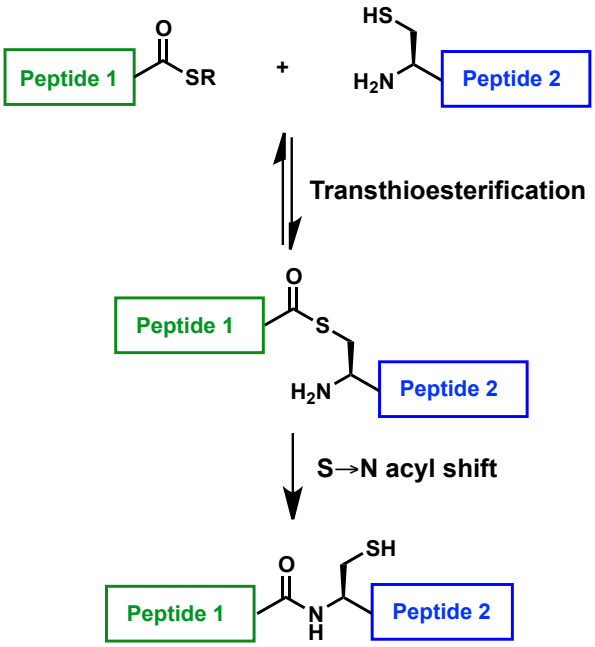


Figure 1: Mechanism of native chemical ligation

1.6 References

1. (a) Milton, R. C.; Milton, S. C.; Kent, S. B., Total chemical synthesis of a D-enzyme: the enantiomers of HIV-1 protease show reciprocal chiral substrate specificity. *Science* **1992,** *256* (5062), 1445-8; (b) Zawadzke, L. E. B., J.M., A racemic protein. *J Am Chem Soc* **1992,** *114* (10), 4002-4003.

2. Francis, J. N.; Redman, J. S.; Eckert, D. M.; Kay, M. S., Design of a modular tetrameric scaffold for the synthesis of membrane-localized d-peptide inhibitors of HIV-1 entry. *Bioconjug Chem* **2012**.

3. Clinton, T. R.; Weinstock, M. T.; Jacobsen, M. T.; Szabo-Fresnais, N.; Pandya, M. J.; Whitby, F. G.; Herbert, A. S.; Prugar, L. I.; McKinnon, R.; Hill, C. P.; Welch, B. D.; Dye, J. M.; Eckert, D. M.; Kay, M. S., Design and characterization of ebolavirus GP prehairpin intermediate mimics as drug targets. *Protein Science* **2015,** *24* (4), 446-463.

4. Eckert, D. M.; Kim, P. S., Mechanisms of viral membrane fusion and its inhibition. *Annu Rev Biochem* **2001,** *70*, 777-810.

5. Schumacher, T. N.; Mayr, L. M.; Minor, D. L., Jr.; Milhollen, M. A.; Burgess, M. W.; Kim, P. S., Identification of D-peptide ligands through mirror-image phage display. *Science* **1996,** *271* (5257), 1854-7.

6. Xie, J.; Schultz, P. G., An expanding genetic code. *Methods* **2005,** *36* (3), 227-38.

7. (a) Yeates, T. O.; Kent, S. B., Racemic protein crystallography. *Annu Rev Biophys* **2012,** *41*, 41-61; (b) Matthews, B. W., Racemic crystallography—easy crystals and easy structures: what's not to like? *Protein Science* **2009,** *18* (6), 1135-1138.

8. (a) Mandal, K.; Pentelute, B. L.; Tereshko, V.; Thammavongsa, V.; Schneewind, O.; Kossiakoff, A. A.; Kent, S. B., Racemic crystallography of synthetic protein enantiomers used to determine the X-ray structure of plectasin by direct methods. *Protein Sci* **2009,** *18* (6), 1146-54; (b) Pentelute, B. L.; Gates, Z. P.; Tereshko, V.; Dashnau, J. L.; Vanderkooi, J. M.; Kossiakoff, A. A.; Kent, S. B., X-ray structure of snow flea antifreeze protein determined by racemic crystallization of synthetic protein enantiomers. *J Am Chem Soc* **2008,** *130* (30), 9695-701; (c) Mandal, K.; Pentelute, B. L.; Bang, D.; Gates, Z. P.; Torbeev, V. Y.; Kent, S. B., Design, total chemical synthesis, and X-ray structure of a protein having a novel linear-loop polypeptide chain topology. *Angew Chem Int Ed Engl* **2012,** *51* (6), 1481-6; (d) Bunker, R. D.; Mandal, K.; Bashiri, G.; Chaston, J. J.; Pentelute, B. L.; Lott, J. S.; Kent, S. B.; Baker, E. N., A functional role of Rv1738 in Mycobacterium tuberculosis persistence suggested by racemic protein crystallography. *Proc Natl Acad Sci U S A* **2015,** *112* (14), 4310-5.

9. (a) Muir, T. W.; Sondhi, D.; Cole, P. A., Expressed protein ligation: a general method for protein engineering. *Proc Natl Acad Sci U S A* **1998,** *95* (12), 6705-10; (b) Muralidharan, V.; Muir, T. W., Protein ligation: an enabling technology for the biophysical analysis of proteins. *Nat Methods* **2006,** *3* (6), 429-38.

10. Muir, T. W., Semisynthesis of proteins by expressed protein ligation. *Annu Rev Biochem* **2003,** *72*, 249-89.

11. Hemantha, H. P.; Bavikar, S. N.; Herman-Bachinsky, Y.; Haj-Yahya, N.; Bondalapati, S.; Ciechanover, A.; Brik, A., Nonenzymatic polyubiquitination of expressed proteins. *J Am Chem Soc* **2014,** *136* (6), 2665-73.

12. Haj-Yahya, M.; Fauvet, B.; Herman-Bachinsky, Y.; Hejjaoui, M.; Bavikar, S. N.; Karthikeyan, S. V.; Ciechanover, A.; Lashuel, H. A.; Brik, A., Synthetic polyubiquitinated α-Synuclein reveals important insights into the roles of the ubiquitin chain in regulating its pathophysiology. *Proc Natl Acad Sci U S A* **2013,** *110* (44), 17726-17731.

13. Voigt, P.; Reinberg, D., Histone tails: ideal motifs for probing epigenetics through chemical biology approaches. *Chembiochem* **2011,** *12* (2), 236-52.

14. Focke, P. J.; Valiyaveetil, F. I., Studies of ion channels using expressed protein ligation. *Curr Opin Chem Biol* **2010,** *14* (6), 797-802.

15. van de Langemheen, H.; Quarles van Ufford, H. L.; Kruijtzer, J. A.; Liskamp, R. M., Efficient synthesis of protein mimics by sequential native chemical ligation. *Org Lett* **2014,** *16* (8), 2138-41.

16. (a) Merrifield, B., Solid phase synthesis. *Science* **1986,** *232* (4748), 341-347; (b) Merrifield, R. B., Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *J Am Chem Soc* **1963,** *85* (14), 2149-2154.

17. Hackenberger, C. P.; Schwarzer, D., Chemoselective ligation and modification strategies for peptides and proteins. *Angew Chem Int Ed Engl* **2008,** *47* (52), 10030-74.

18. Dawson, P. E.; Kent, S. B., Synthesis of native proteins by chemical ligation. *Annu Rev Biochem* **2000,** *69*, 923-60.

19. Li, C. H.; Yamashiro, D., The synthesis of a protein possessing growth-promoting and lactogenic activities. *J Am Chem Soc* **1970,** *92* (26), 7608-9.

20. Clark-Lewis, I.; Aebersold, R.; Ziltener, H.; Schrader, J. W.; Hood, L. E.; Kent, S. B., Automated chemical synthesis of a protein growth factor for hemopoietic cells, interleukin-3. *Science* **1986,** *231* (4734), 134-9.

21. (a) Verzele, D.; Madder, A., Patchwork protein chemistry: a practitioner's treatise on the advances in synthetic peptide stitchery. *Chembiochem* **2013,** *14* (9), 1032-48; (b) Kent, S. B., Total chemical synthesis of proteins. *Chem Soc Rev* **2009,** *38* (2), 338-51.

22. Kent, S. B., Chemical synthesis of peptides and proteins. *Annu Rev Biochem* **1988,** *57*, 957-89.

23. Sasaki, T.; Findeis, M. A.; Kaiser, E. T., Evaluation of the oxime resin based segment synthesis-condensation approach using RNase T1 as a model synthetic target. *J Org Chem* **1991,** *56* (9), 3159-3168.

24. Fridkin, M.; Patchornik, A., Peptide synthesis. *Annu Rev Biochem* **1974,** *43* (0), 419-43.

25. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B., Synthesis of proteins by native chemical ligation. *Science* **1994,** *266* (5186), 776-9.

26. Schnolzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B., In situ neutralization in boc-chemistry solid phase peptide synthesis. Rapid, high yield assembly of difficult sequences. *Int J Pept Protein Res* **1992,** *40* (3-4), 180-93.

27. Offer, J.; Boddy, C. N.; Dawson, P. E., Extending synthetic access to proteins with a removable acyl transfer auxiliary. *J Am Chem Soc* **2002,** *124* (17), 4642-6.

28. Yan, L. Z.; Dawson, P. E., Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurization. *J Am Chem Soc* **2001,** *123* (4), 526-33.

29. Wan, Q.; Danishefsky, S. J., Free-radical-based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides. *Angew Chem Int Ed Engl* **2007,** *46* (48), 9248-52.

30. Crich, D.; Banerjee, A., Native chemical ligation at phenylalanine. *J Am Chem Soc* **2007,** *129* (33), 10064-10065.

31. (a) He, Q. Q. F., G. M.; Liu, L., Design of thiol-containing amino acids for native chemical ligation at non-Cys sites. *Chin Chem Lett* **2013,** *24* (4), 265-269; (b) Malins, L. R.; Payne, R. J., Recent extensions to native chemical ligation for the chemical synthesis of peptides and proteins. *Curr Opin Chem Biol* **2014,** *22*, 70-8.

32. (a) Haase, C.; Rohde, H.; Seitz, O., Native chemical ligation at valine. *Angew Chem Int Ed Engl* **2008,** *47* (36), 6807-10; (b) Chen, J. W., Q.; Yuan, Y.; Zhu, J.; Danishefsky, S. J., Native chemical ligation at valine: a contribution to peptide and glycopeptide synthesis. *Angew Chem Int Ed Engl* **2008,** *47* (44), 8521-8524.

33. Shang, S.; Tan, Z.; Dong, S.; Danishefsky, S. J., An advance in proline ligation. *J Am Chem Soc* **2011,** *133* (28), 10784-6.

34. Hackeng, T. M.; Griffin, J. H.; Dawson, P. E., Protein synthesis by native chemical ligation: expanded scope by using straightforward methodology. *Proc Natl Acad Sci U S A* **1999,** *96* (18), 10068-73.

35. Okamoto, R.; Mandal, K.; Ling, M.; Luster, A. D.; Kajihara, Y.; Kent, S. B., Total chemical synthesis and biological activities of glycosylated and non-glycosylated forms of the chemokines CCL1 and Ser-CCL1. *Angew Chem Int Ed Engl* **2014,** *53* (20), 5188-93.

36. Li, X. K., T.; Aimoto, S., Direct preparation of peptide thioesters using an Fmoc solid-phase method. *Tetrahedron Lett* **1998,** *39* (47), 8669-8672.

37. Raibaut, L.; Ollivier, N.; Melnyk, O., Sequential native peptide ligation strategies for total chemical protein synthesis. *Chem Soc Rev* **2012,** *41* (21), 7001-7015.

38. Hojo, H., Recent progress in the chemical synthesis of proteins. *Curr Opin Struct Biol* **2014,** *26* (0), 16-23.

39. (a) Zheng, J. S.; Tang, S.; Qi, Y. K.; Wang, Z. P.; Liu, L., Chemical synthesis of proteins using peptide hydrazides as thioester surrogates. *Nat Protoc* **2013,** *8* (12), 2483-95; (b) Zheng, J. S.; Tang, S.; Guo, Y.; Chang, H. N.; Liu, L., Synthesis of cyclic peptides and cyclic proteins via ligation of peptide hydrazides. *Chembiochem* **2012,** *13* (4), 542-6; (c) Fang, G. M.; Wang, J. X.; Liu, L., Convergent chemical synthesis of proteins by ligation of peptide hydrazides. *Angew Chem Int Ed Engl* **2012,** *51* (41), 10347-50; (d) Fang, G. M.; Li, Y. M.; Shen, F.; Huang, Y. C.; Li, J. B.; Lin, Y.; Cui, H. K.; Liu, L., Protein chemical synthesis by ligation of peptide hydrazides. *Angew Chem Int Ed Engl* **2011,** *50* (33), 7645-9.

CHAPTER 2

SYNTHESIS AND FOLDING OF A MIRROR-IMAGE ENZYME REVEALS AMBIDEXTROUS CHAPERONE ACTIVITY

Matthew T. Weinstock, Michael T. Jacobsen, Michael S. Kay

Reprinted with permission from PNAS 111(32), 11679-11684

Synthesis and folding of a mirror-image enzyme reveals ambidextrous chaperone activity

Matthew T. Weinstock[1], Michael T. Jacobsen, and Michael S. Kay[2]

Department of Biochemistry, University of Utah School of Medicine, Salt Lake City, UT 84112-5650

Edited by Gregory A. Petsko, Weill Cornell Medical College, New York, NY, and approved July 9, 2014 (received for review June 11, 2014)

Mirror-image proteins (composed of D-amino acids) are promising therapeutic agents and drug discovery tools, but as synthesis of larger D-proteins becomes feasible, a major anticipated challenge is the folding of these proteins into their active conformations. In vivo, many large and/or complex proteins require chaperones like GroEL/ES to prevent misfolding and produce functional protein. The ability of chaperones to fold D-proteins is unknown. Here we examine the ability of GroEL/ES to fold a synthetic D-protein. We report the total chemical synthesis of a 312-residue GroEL/ES-dependent protein, DapA, in both L- and D-chiralities, the longest fully synthetic proteins yet reported. Impressively, GroEL/ES folds both L- and D-DapA. This work extends the limits of chemical protein synthesis, reveals ambidextrous GroEL/ES folding activity, and provides a valuable tool to fold D-proteins for drug development and mirror-image synthetic biology applications.

peptide synthesis | protein folding

All known living organisms use proteins composed of L-amino acids. Mirror-image proteins (composed of D-amino acids) are not found in nature and are promising therapeutic agents due to their resistance to degradation by natural proteases (1, 2). D-peptide inhibitors that target particular protein interfaces can be identified by mirror-image phage display (3, 4), in which a library of phage bearing L-peptides on their surface is screened against a mirror-image (D-) protein target. By symmetry, D-peptide versions of the identified sequences will bind to the natural L-target. Because D-protein targets must be chemically synthesized, this discovery method has thus far been limited to relatively small targets.

Through rigorous application of recent advances in chemical protein synthesis (reviewed in ref. 5), the production of larger synthetic D-proteins is becoming increasingly feasible [e.g., 204-residue D-VEGF dimer (6) and 84-residue D-MDM2/MDMX (7)]. However, many proteins are prone to misfolding, especially as their size and complexity increase (8). Molecular chaperones, such as the extensively studied GroEL/ES, mediate folding by preventing aggregation of many cellular proteins (9, 10). GroEL/ES is thought to interact with these diverse substrates via nonspecific hydrophobic interactions, but it is unknown whether it can fold mirror-image proteins. If natural chaperones cannot fold mirror-image proteins, then the folding of large/complex D-proteins into their active conformations will be a major challenge (in the absence of mirror-image chaperones, which are currently inaccessible).

The binding of substrates by GroEL is an intriguing instance of promiscuous molecular recognition. GroEL has been shown to interact transiently with ∼250 cytosolic proteins in *Escherichia coli* under normal growth conditions (8, 11). A subset of these proteins exhibit an absolute requirement for GroEL and its cochaperone GroES to avoid aggregation and fold into their native state (8, 12). Interestingly, sequence analysis of known GroEL/ES obligate substrates reveals no obvious consensus binding sequence (11), although structurally they are enriched in aggregation-prone folds (12).

Several lines of evidence suggest the predominant interactions between GroEL/ES and substrate proteins are hydrophobic. Protein substrates trapped in nonnative states have been shown to present hydrophobic surfaces that are otherwise buried in the core of the correctly folded protein, and a hydrophobic binding model is supported by the thermodynamics of binding of these nonnative states to GroEL (13). Additionally, the GroEL apical domain residues implicated in substrate binding are largely hydrophobic (14). Finally, previous studies on the basis of substrate interaction with GroEL using short model peptides have concluded that the most important determinant of substrate binding is the presentation of a cluster of hydrophobic residues (15–17).

The only evidence addressing the chiral specificity of GroEL/ES comes from a study that qualitatively demonstrated binding of a short D-peptide to GroEL (16). However, this NMR study required peptide concentrations that greatly exceed physiologic levels and did not localize the interaction to the substrate-binding region of GroEL. Only recently has it become feasible to directly test the stereospecificity of the GroEL/ES folding reaction by synthesizing the mirror-image version of a chaperone-dependent protein.

Due to great interest in mirror-image proteins as targets for drug discovery (6, 7, 18, 19) and mirror-image synthetic biology (20, 21), we were intrigued by the possibility that natural (L-) GroEL/ES could assist in the folding of D-proteins. Thus, we synthesized a D-version of a substrate protein and evaluated its folding by GroEL/ES. Furthermore, because most GroEL/ES substrate proteins are large (>250 residues), this project provided an excellent opportunity to demonstrate the power of chemical synthesis methodologies for producing previously inaccessible synthetic proteins.

Significance

This paper addresses a fundamental question: Can natural chaperones fold mirror-image proteins? Mirror-image proteins (composed of D-amino acids) are only accessible by chemical synthesis, but are protease resistant and therefore have tremendous potential as long-lived drugs. Many large/complex proteins depend on chaperones for efficient folding. Here we describe the total chemical synthesis of a 312-residue chaperone-dependent protein (DapA) in natural (L-) and mirror-image (D-) forms, the longest fully synthetic proteins yet reported. Using these proteins we show that the natural bacterial GroEL/ES chaperone is "ambidextrous"—i.e., it can fold both natural and mirror-image proteins via nonspecific hydrophobic interactions. Our study also provides proof-of-concept for the use of natural GroEL/ES to fold D-proteins for mirror-image drug discovery and synthetic biology applications.

Author contributions: M.T.W., M.T.J., and M.S.K. designed research; M.T.W. and M.T.J. performed research; M.T.W., M.T.J., and M.S.K. analyzed data; and M.T.W., M.T.J., and M.S.K. wrote the paper.

Conflict of interest statement: M.S.K. is a Scientific Director, consultant, and equity holder of the D-Peptide Research Division of Navigen, which is commercializing D-peptide inhibitors of viral entry.

This article is a PNAS Direct Submission.

[1]Present address: Synthetic Genomics, La Jolla, CA 92037.

[2]To whom correspondence should be addressed. Email: kay@biochem.utah.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1410900111/-/DCSupplemental.

Results

Selection of DapA as Model Protein. We began our investigation by searching for the smallest model protein that requires GroEL/ES for folding under physiologic conditions and has a robust activity assay that does not depend on complex chiral reagents (e.g., cofactors or other enzymes that would also require mirror-image synthesis). The *E. coli* DapA protein [4-hydroxy-tetrahydrodipicolinate synthase (EC 4.3.3.7)] (Fig. 1) meets these criteria. DapA is a 31-kDa protein that forms a homotetramer (22) and catalyzes the stereoselective (23) condensation of L-aspartate-β-semialdehyde and pyruvate to (4S)-4-hydroxy-2,3,4,5-tetrahydro-(2S)-dipicolinic acid (24–26), a key step in the biosynthesis of lysine and diaminopimelic acid, a cell-wall precursor.

DapA is highly enriched in GroEL/ES complexes under normal growth conditions (8) and is aggregated or degraded in GroEL-depleted cells (8, 12). In *E. coli* GroEL-depletion strains, cell death occurs via lysis due to loss of DapA activity, further demonstrating the dependence of DapA on GroEL/ES to adopt its native structure (27). Indeed, in vitro, DapA is absolutely dependent on GroEL/ES for proper folding in physiologic buffer at 37 °C (8). An added benefit of DapA as a model protein is that although it depends on chaperones for folding under physiological conditions, a chemical-folding procedure has been reported using 0.5 M arginine (8), providing an important positive control for enzyme activity independent of chaperone-mediated folding. Although a chemical refolding protocol for DapA was reported, many complex proteins cannot be folded by known chemical means, e.g., *E. coli* METK (28) and METF (8, 29).

Synthesis of L- and D-DapA. Because D-proteins can only be accessed through chemical synthesis, a synthetic route to DapA was devised. Synthetic peptides are routinely made using solid-phase peptide synthesis (SPPS), but a project of this magnitude (312 residues) is well outside the capability of current SPPS technology (generally ~50 residues). To access larger synthetic assemblies, chemoselective ligation techniques, especially native chemical ligation (30), are used to assemble peptide segments into larger constructs (reviewed in refs. 5, 31, and 32). Recent noteworthy synthetic proteins include tetraubiquitin [alone (33) and as part of a semisynthetic tetraubiquitinated α-synuclein (34)], covalent HIV protease dimer (35), L-/D-snow flea antifreeze protein (36), glycosylated EPO (37, 38), and the γ-subunit of F-ATPase (39).

For synthesizing DapA, we used a recently developed method to join peptide segments via native peptide bonds formed between a peptide with a C-terminal hydrazide (for selective conversion to a thioester) and a peptide with an N-terminal Cys (40–43). We selected this chemistry because of the convenient route to peptide hydrazides via Fmoc SPPS (less hazardous and more compatible with acid-sensitive modifications than Boc chemistry), the robustness of the native chemical ligation reaction (30, 44), and the ease of carrying out convergent protein assembly (vs. linear C- to N-assembly).

Our retrosynthetic analysis began by locating all Cys residues (potential ligation junctions) in DapA (Fig. 1*A*), all of which are located at acceptable ligation junctions (see refs. 40 and 45 for discussions of unacceptable junctions). This information allowed us to break the protein into six segments, leaving two segments >50 residues. To expand the range of potential ligation junctions, we used a free radical-based desulfurization reaction that enables selective conversion of unprotected Cys to Ala (46, 47). This technique allows one to substitute a Cys for a native Ala residue during peptide synthesis (for use in ligation) and then convert the Cys back to the native Ala following assembly. Using

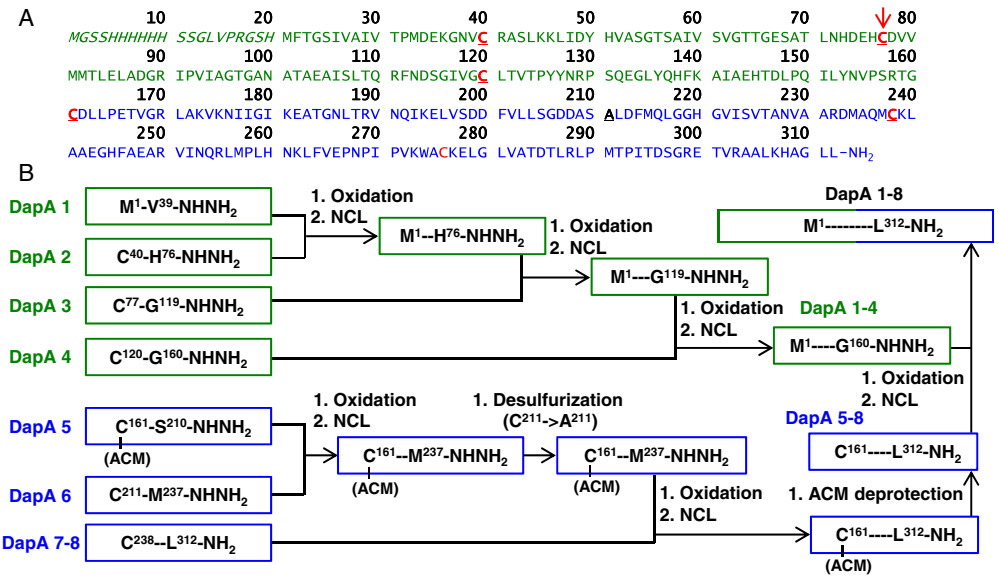


Fig. 1. Total chemical synthesis of 312-residue DapA. (*A*) Target amino acid sequence, including N-terminal half (DapA 1–4, green), C-terminal half (DapA 5–8, blue), N-terminal His tag and thrombin cleavage site (italics), cysteines (red), ligation sites (bold and underline), and A77C mutation (red arrow). (*B*) Final synthetic strategy, including peptide segments (DapA 1–4, green and DapA 5–8, blue) with ligation hydrazide and cysteine residues indicated. Oxidation (hydrazide to azide), native chemical ligation (NCL), and acetamidomethyl (ACM) cysteine deprotection steps are also indicated.

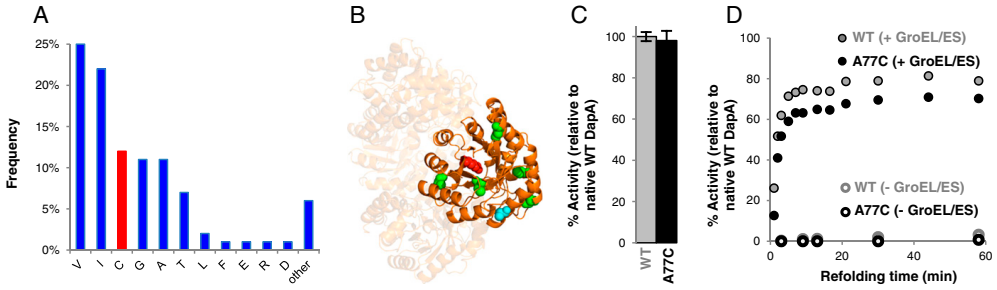


Fig. 2. Validation of the DapA A77C mutation. (*A*) Natural sequence diversity at position 77 from Protein BLAST analysis. (*B*) Structure of DapA tetramer (PDB ID code 1DHP) showing, on one subunit, the surface-exposed alanine at position 77 (cyan), natural cysteine residues (green), and catalytic lysine at position 181 in the active site (red). (*C*) Enzyme activity of recombinant native WT and A77C DapA. Error bars indicate SD of at least three measurements. (*D*) GroEL/ES-mediated refolding of recombinant WT and A77C DapA.

this method, we introduced two additional junction sites at A77 and A211, resulting in eight segments overall, ranging in size from 27 to 50 residues (DapA 1–8; Fig. S1). Using optimized SPPS reaction conditions and RP-HPLC column selection (*Materials and Methods*), we synthesized and purified all eight peptides.

Our initial strategy for the assembly of these eight segments required 12 steps (seven ligations, two desulfurizations, and three Acm removals; Fig. S1) and their associated purifications. Acm was used as an orthogonal Cys protecting group that prevents cyclization/polymerization of peptides containing both an activated C-terminal hydrazide and an N-terminal Cys, and also prevents Cys desulfurization. Following this scheme, we assembled the C-terminal segments (DapA 5–8), but were unable to assemble the N-terminal segments (DapA 1–4). A significant complication was the His thioester on DapA 2 (H76), which was highly susceptible to hydrolysis, leading to low reaction yields during the DapA 2/3 ligation step (Fig. S2). This difficulty, coupled with the large number of manipulations (and concomitant sample losses), resulted in a failure to assemble DapA 1–4 in usable yield.

We reasoned that we could simplify the assembly if we eliminated the desulfurization step necessary to convert the Cys to native Ala at the DapA 2–3 junction. Toward this end, we determined locations in our protein that would likely tolerate permanent mutation to Cys. BLAST analysis of the *E. coli* DapA identified the 1,000 most-similar homologs (>69% conservation, >49% identity), which were aligned to determine positions where Cys residues naturally occur. Fortuitously, 12% of the aligned sequences contained Cys at position 77, site of the DapA 2–3 junction (Fig. 2*A*). Next, we analyzed the DapA crystal structure to determine the likelihood of the A77C mutation to disrupt protein structure/function. The side chain of residue 77 is surface-exposed and not in close proximity to the active site or any native Cys residues (>12 Å to the nearest Cys; Fig. 2*B*). This analysis suggested that introduction of the A77C mutation would likely be well tolerated. Indeed, this mutation affected neither recombinant protein activity (Fig. 2*C*) nor its dependence on GroEL/ES for folding under physiological conditions (Fig. 2*D* and Table S1). We ultimately selected a final assembly strategy that incorporated both the A77C mutation and a unified DapA 7–8 segment (we were not able to produce high-quality DapA 1–2, 3–4, or 5–6 unified peptides). This final strategy yielded a seven-segment assembly scheme (Fig. 1*B*) that removed four synthetic steps (and associated purifications) from the initial scheme.

Following this simplified strategy, we successfully assembled the 312-residue synthetic DapA A77C (hereafter referred to as "DapA") in both L- and D- chiralities (Fig. 3 and *SI Text*), the longest synthetic peptides reported to date. The peptides were synthesized at milligram scale (1.1 and 1.7 mg of L- and D-DapA, respectively; Figs. S3 and S4). The synthetic L- and D-peptides behave identically to recombinant DapA on a C4 RP-HPLC column (Fig. 3*A*), and the major products possess the correct mass (Fig. 3 *B* and *C*).

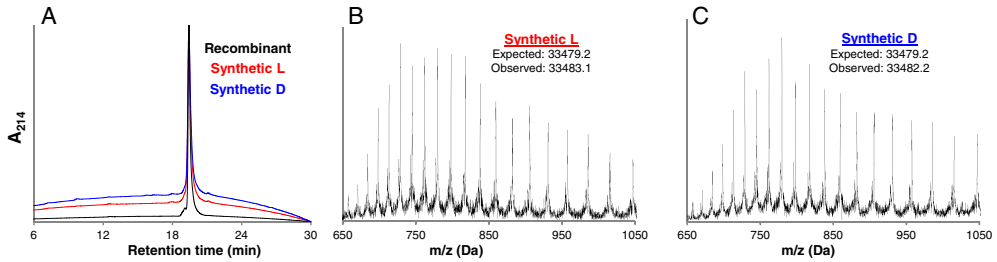


Fig. 3. Analysis of synthetic unfolded L- and D-DapA. (*A*) Analytical RP-HPLC of recombinant (black), synthetic L- (red), and synthetic D-DapA (blue) on C4 column (linear gradient 5–100% buffer B over 30 min; buffer A, 0.1% TFA in water; buffer B, 0.1% TFA in 10% water/90% acetonitrile). (*B* and *C*) LC-MS analysis of the synthetic L- and D-DapA, respectively. Observed masses were calculated using the Bayesian Protein Reconstruct tool in Analyst 1.5.1 software (AB Sciex) over the charge states covering 650–1,050 Da. See *SI Appendix* for larger, detailed mass spectra of the final synthetic products and HPLC and LC-MS characterization of all synthetic intermediates.

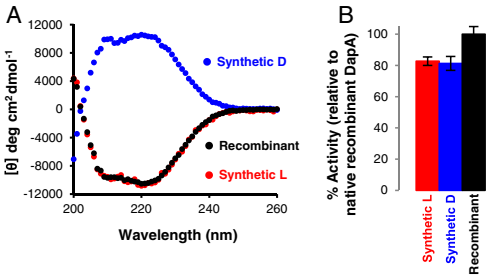


Fig. 4. Structural and functional characterization of synthetic folded L- and D-DapA. (*A*) Circular dichroism spectra of Arg-folded and SEC-purified recombinant (black), synthetic L- (red), and synthetic D-DapA (blue). (*B*) Enzyme activity of Arg-folded and SEC-purified synthetic L- and synthetic D-DapA compared with native recombinant DapA. Error bars indicate SD of at least three assays.

Our initial efforts to fold these synthetic peptides using GroEL/ES resulted in measurable enzymatic activity, albeit at relatively low levels (~20–40%; Table S1), likely due to microheterogeneity in the synthetic peptides (*SI Appendix*). Because active DapA assembles as a tetramer, we reasoned that we could enrich for "foldable" protein by using a chemical refolding procedure followed by size-exclusion chromatography (SEC).

Chemical-Mediated Folding of DapA. Chaperone-independent folding of DapA has been described using 0.5 M L-arginine (8), a protein refolding additive (48). This method was validated at 13 °C using recombinant DapA and works equally well with D-arginine (Fig. S5). Thus, L-arginine can be used to fold both L- and D-DapA. Importantly, this procedure also provides a chaperone-independent means to evaluate the activity of our synthetic constructs.

After arginine-assisted folding of synthetic L- and D-DapA, we isolated tetrameric protein using SEC (Fig. S6). Following SEC, both the L- and D-DapA synthetic proteins have the expected CD spectra (Fig. 4*A*) and are enzymatically active (Fig. 4*B*), demonstrating that both L- and D-synthetic proteins are correctly folded and functional. As hoped, the SEC purification generated synthetic proteins with high specific activity (~80% compared with recombinant protein). However, the Arg-assisted refolding/SEC purification resulted in a substantial (>10-fold) yield loss, largely due to precipitation during dialysis and concentration steps.

Chaperone-Mediated Folding of DapA. With folded and equally active synthetic L- and D-DapA in hand, we were poised to perform the definitive experiment comparing the refolding of our synthetic L- and D-DapA by GroEL/ES. This experiment answers the question of whether GroEL/ES is ambidextrous (i.e., Can it fold a mirror-image protein?). The SEC-purified proteins were denatured for 1 h in denaturation buffer (containing 6 M GuHCl) and then diluted 100-fold into refolding buffer with or without GroEL/ES at 37 °C to initiate refolding. At specific time points, refolding was quenched by Mg chelation [1,2-diaminocyclohexanetetraacetic acid (CDTA)] followed by measurement of enzyme activity using a colorimetric assay (8). Interestingly, GroEL/ES refolded both synthetic L- and D-DapA, as demonstrated by the recovery of significant enzymatic activity (Fig. 5 and Table S1).

Discussion

The results presented here demonstrate that GroEL/ES is able to fold a D-protein and therefore does not manifest strict stereospecificity in folding its substrates. This result supports a substrate binding mechanism via nonspecific hydrophobic interactions followed by sequestration in the GroEL/ES cage (9, 10). Our study also provides proof-of-concept for the use of natural (L-) GroEL/ES to fold D-proteins for mirror-image drug discovery and synthetic biology applications.

To determine if the ability of GroEL/ES to fold D-proteins is universal, the most definitive approach would be the total chemical synthesis of D-GroEL (548 residues) and D-GroES (97 residues), followed by screening of a suite of well-characterized recombinant L-substrates in refolding assays. Though we observed no difference in the activity of chemically refolded synthetic L- vs. D-DapA, there was a noticeable difference in their chaperone-mediated refolding. More detailed folding studies (49) requiring additional material will be needed to determine whether this difference reflects a general chiral preference in the recognition and/or extent/rate of folding.

Although the synthetic proteins show high specific activity (~80% of recombinant protein; Fig. 4*B*), it will be important to improve their quality and yield to expand application of this work to even larger synthetic proteins. We speculate that subtle synthetic defects in our proteins include single-residue deletions, racemization (50), and aspartamide formation (51, 52).

Ultimately, the ability to chemically synthesize proteins of interest not only serves to advance mirror-image drug discovery efforts by making larger targets available, but also provides alluring possibilities for mirror-image synthetic biology (20) and complements efforts to synthesize other large biomolecules (e.g., synthetic genomes) (53). An intriguing prospect is the assembly of a mirror-image in vitro translation apparatus (including mirror-image ribosomal proteins in combination with mirror-image rRNAs; all but one of the 70S subunits are <300 residues), an effort that we have dubbed the "*D. coli*" project (18). Such a tool would not only provide a facile route to the production of mirror-image biomolecules for drug discovery, but would also facilitate the structural/biochemical study of highly toxic agents in (nontoxic) mirror-image form.

Materials and Methods

Peptide Synthesis and Ligation. Peptides were synthesized via Fmoc–SPPS on a commercial peptide synthesizer (Prelude; Protein Technologies, Inc). Peptide hydrazides were prepared on 2-hydrazine chlorotrityl resin (ChemPep). Peptide hydrazides were activated in 6 M GuHCl, 100 mM sodium phosphate,

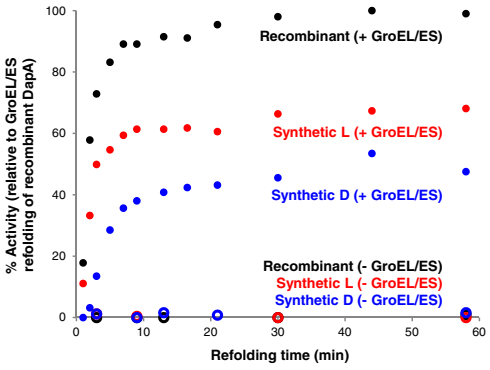


Fig. 5. GroEL/ES-mediated refolding of synthetic L- and D-DapA. Refolding of recombinant (black), synthetic L- (red), and synthetic D-DapA (blue) (250 nM) in the presence (closed circles) or absence (open circles) of 7 μM GroEL/ES. Data are normalized to the maximum point in the GroEL/ES refolding of recombinant DapA.

pH 3.0 (5–20 mM $NaNO_2$) at −20 °C for 20 min. Peptides were then ligated in 6 M GuHCl, 200 mM 4-mercaptophenylacetic acid (MPAA), 200 mM sodium phosphate, pH adjusted to 7.0–7.2, at 25 °C for 5–20 h. Ligation reactions were quenched by addition of freshly prepared tris(2-carboxyethyl)phosphine to ∼130 mM and incubated for >10 min.

Peptide Purification and Characterization. Analytical reverse-phase HPLC was performed using Phenomenex Jupiter 4-μm Proteo C12 90 Å (150 × 4.6 mm) or Phenomenex Jupiter 5-μm C4 300 Å (150 × 4.6 mm) columns. Preparative reverse-phase HPLC of crude peptides was performed on either Phenomenex Jupiter 4-μm Proteo C12 90 Å (250 × 21.2 mm) or Phenomenex Jupiter 10-μm C4 300 Å (250 × 21.2 mm) column. Semipreparative reverse-phase HPLC of ligation products was performed on a Phenomenex Jupiter 10-μm C4 300 Å (250 × 10 mm) column. Purified peptides were analyzed by LC/MS on a Phenomenex Aeris WIDEPORE 3.6-μm C4 (50 × 2.1 mm) column on an AB Sciex API 3000 LC/MS/MS system. The major observed deconvoluted masses from mass spectrometry were calculated using Bayesian Peptide and Protein Reconstruct Tools in Analyst 1.5.1 Software (AB Sciex). See *SI Appendix* for full characterization of all peptides.

Enzyme Activity Assay. Ten-microliter samples of DapA (250 nM) were added to 240 μL of DapA assay buffer [200 mM imidazole (pH 7.4), 35 mM Na pyruvate, 4 mM DL-aspartate-β-semialdehyde, 0.5 mg/mL *o*-aminobenzaldehyde, 12.5 mM CDTA] to initiate the enzyme activity assay (10 nM final enzyme concentration). The assay is quenched after 15 min of agitation at room temperature on a microplate shaker (800 rpm) by the addition of 50 μL 2 M HCl, developed by continuing the agitation for 1 h at room temperature, followed by measuring absorbance at 562 nm. Under these conditions, this assay demonstrates good linearity (A_{562} < 0.4 for WT recombinant DapA; saturation occurs at A_{562} >1.5).

Arginine-Assisted Folding. DapA constructs (both recombinant and synthetic) were dissolved in denaturation buffer [6 M GuHCl, 20 mM MOPS (pH 7.4), 100 mM KCl, 10 mM $MgCl_2$, 10 mM DTT] with 0.5 M arginine and diluted to final concentration of ∼37 μM. Samples were incubated at room temperature for 40 min, 13 °C for 20 min, and then dialyzed [Slide-A-Lyzer minidialysis cassettes 3500 molecular weight cutoff (MWCO)] against 100× volume of refolding buffer [20 mM MOPS (pH 7.4), 100 mM KCl, 10 mM $MgCl_2$, 10 mM sodium pyruvate, 1 mM DTT] with 0.5 M arginine for 2.5 h at 13 °C. Samples were then further dialyzed against 100× volume 100 mM ammonium bicarbonate (pH 8) for 1 h. The dialyzed sample was used directly in functional assays (post-Arg and pre-SEC) or concentrated by Vivaspin 500 10,000 MWCO centrifugal concentrators and further purified by SEC (Superdex 200 10/30; GE Healthcare) in 100 mM ammonium bicarbonate (pH 8) running buffer with a flow rate of 0.75 mL/min (post-Arg and post-SEC). Following SEC, samples were again concentrated and prepared for structural (CD spectroscopy) and functional assays (direct activity and GroEL/ES refolding).

Chaperone Refolding Assay. The DapA refolding assay (to evaluate GroEL/ES chaperone activity) was adapted from ref. 8. Twenty-five micromolar stocks of DapA were prepared from lyophilized powder (pre-SEC) or buffer exchanged (post-SEC) into denaturation buffer [20 mM MOPS (pH 7.4), 100 mM KCl, 10 mM $MgCl_2$, 10 mM DTT, 6 M GuHCl] and allowed to denature for 1 h at 25 °C. Refolding was initiated by diluting 100× into 37 °C refolding buffer [20 mM MOPS (pH 7.4), 100 mM KCl, 10 mM $MgCl_2$, 10 mM sodium pyruvate, 5 mM ATP] with or without 7 μM GroEL monomer and 7 μM GroES monomer. Final DapA concentrations used in refolding assays were 250 nM. At specific time points, 10-μL aliquots of the refolding reaction were added to 240 μL of DapA assay buffer, which simultaneously quenches chaperone-mediated refolding and initiates assay of enzyme activity (measured as described above).

ACKNOWLEDGMENTS. We thank Costa Georgopoulos and Debbie Ang for their helpful discussions, provision of protocols, gift of our initial stock of GroEL, and critical reading of the manuscript; Rob Marquardt for his mass spectrometry advice and training; the Gary Keck laboratory for use of its ozone generator for preparing DL-aspartate-β-semialdehyde; Janet Iwasa for advice on figure design; and Debra Eckert for critical review of the manuscript. Funding for this work was provided by National Institutes of Health (NIH) Predoctoral Training Grant F31CA171677 (to M.T.J.) and NIH Grants AI076168 and AI102347 (to M.S.K.).

1. Zawadzke LE, Berg JM (1992) A racemic protein. *J Am Chem Soc* 114(10):4002–4003.
2. Milton RC, Milton SC, Kent SB (1992) Total chemical synthesis of a D-enzyme: The enantiomers of HIV-1 protease show reciprocal chiral substrate specificity. *Science* 256(5062):1445–1448.
3. Schumacher TN, et al. (1996) Identification of D-peptide ligands through mirror-image phage display. *Science* 271(5257):1854–1857.
4. Welch BD, VanDemark AP, Heroux A, Hill CP, Kay MS (2007) Potent D-peptide inhibitors of HIV-1 entry. *Proc Natl Acad Sci USA* 104(43):16828–16833.
5. Verzele D, Madder A (2013) Patchwork protein chemistry: A practitioner's treatise on the advances in synthetic peptide stitchery. *ChemBioChem* 14(9):1032–1048.
6. Mandal K, et al. (2012) Chemical synthesis and X-ray structure of a heterochiral D-protein antagonist plus vascular endothelial growth factor protein complex by racemic crystallography. *Proc Natl Acad Sci USA* 109(37):14779–14784.
7. Liu M, et al. (2010) A left-handed solution to peptide inhibition of the p53-MDM2 interaction. *Angew Chem Int Ed Engl* 49(21):3649–3652.
8. Kerner MJ, et al. (2005) Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. *Cell* 122(2):209–220.
9. Horwich AL, Fenton WA (2009) Chaperonin-mediated protein folding: using a central cavity to kinetically assist polypeptide chain folding. *Q Rev Biophys* 42(2):83–116.
10. Kim YE, Hipp MS, Bracher A, Hayer-Hartl M, Hartl FU (2013) Molecular chaperone functions in protein folding and proteostasis. *Annu Rev Biochem* 82:323–355.
11. Houry WA, Frishman D, Eckerskorn C, Lottspeich F, Hartl FU (1999) Identification of in vivo substrates of the chaperonin GroEL. *Nature* 402(6758):147–154.
12. Fujiwara K, Ishihama Y, Nakahigashi K, Soga T, Taguchi H (2010) A systematic survey of in vivo obligate chaperonin-dependent substrates. *EMBO J* 29(9):1552–1564.
13. Lin Z, Schwartz FP, Eisenstein E (1995) The hydrophobic nature of GroEL-substrate binding. *J Biol Chem* 270(3):1011–1014.
14. Fenton WA, Kashi Y, Furtak K, Horwich AL (1994) Residues in chaperonin GroEL required for polypeptide binding and release. *Nature* 371(6498):614–619.
15. Chen L, Sigler PB (1999) The crystal structure of a GroEL/peptide complex: Plasticity as a basis for substrate diversity. *Cell* 99(7):757–768.
16. Wang Z, Feng Hp, Landry SJ, Maxwell J, Gierasch LM (1999) Basis of substrate binding by the chaperonin GroEL. *Biochemistry* 38(39):12537–12546.
17. Coyle JE, Jaeger J, Gross M, Robinson CV, Radford SE (1997) Structural and mechanistic consequences of polypeptide binding by GroEL. *Fold Des* 2(6):R93–R104.
18. Weinstock MT, Francis JN, Redman JS, Kay MS (2012) Protease-resistant peptide design-empowering nature's fragile warriors against HIV. *Biopolymers* 98(5):431–442.
19. Welch BD, et al. (2010) Design of a potent D-peptide HIV-1 entry inhibitor with a strong barrier to resistance. *J Virol* 84(21):11235–11244.
20. Forster AC, Church GM (2007) Synthetic biology projects in vitro. *Genome Res* 17(1):1–6.
21. Church GM, Regis E (2012) *Regenesis* (Basic, New York).
22. Mirwaldt C, Korndörfer I, Huber R (1995) The crystal structure of dihydrodipicolinate synthase from *Escherichia coli* at 2.5 A resolution. *J Mol Biol* 246(1):227–239.
23. Dobson RC, Gerrard JA, Pearce FG (2004) Dihydrodipicolinate synthase is not inhibited by its substrate, (S)-aspartate beta-semialdehyde. *Biochem J* 377(Pt 3):757–762.
24. Blickling S, et al. (1997) Reaction mechanism of *Escherichia coli* dihydrodipicolinate synthase investigated by X-ray crystallography and NMR spectroscopy. *Biochemistry* 36(1):24–33.
25. Devenish SR, Blunt JW, Gerrard JA (2010) NMR studies uncover alternate substrates for dihydrodipicolinate synthase and suggest that dihydrodipicolinate reductase is also a dehydratase. *J Med Chem* 53(12):4808–4812.
26. Yugari Y, Gilvarg C (1965) The condensation step in diaminopimelate synthesis. *J Biol Chem* 240(12):4710–4716.
27. McLennan N, Masters M (1998) GroE is vital for cell-wall synthesis. *Nature* 392(6672):139.
28. Ying BW, Taguchi H, Kondo M, Ueda T (2005) Co-translational involvement of the chaperonin GroEL in the folding of newly translated polypeptides. *J Biol Chem* 280(12):12035–12040.
29. Tang YC, et al. (2006) Structural features of the GroEL-GroES nano-cage required for rapid folding of encapsulated protein. *Cell* 125(5):903–914.
30. Dawson PE, Muir TW, Clark-Lewis I, Kent SB (1994) Synthesis of proteins by native chemical ligation. *Science* 266(5186):776–779.
31. Kent SB (2009) Total chemical synthesis of proteins. *Chem Soc Rev* 38(2):338–351.
32. Hackenberger CP, Schwarzer D (2008) Chemoselective ligation and modification strategies for peptides and proteins. *Angew Chem Int Ed Engl* 47(52):10030–10074.
33. Kumar KS, et al. (2011) Total chemical synthesis of a 304 amino acid K48-linked tetraubiquitin protein. *Angew Chem Int Ed Engl* 50(27):6137–6141.
34. Haj-Yahya M, et al. (2013) Synthetic polyubiquitinated α-Synuclein reveals important insights into the roles of the ubiquitin chain in regulating its pathophysiology. *Proc Natl Acad Sci USA* 110(44):17726–17731.
35. Torbeev VY, Kent SB (2007) Convergent chemical synthesis and crystal structure of a 203 amino acid "covalent dimer" HIV-1 protease enzyme molecule. *Angew Chem Int Ed Engl* 46(10):1667–1670.
36. Pentelute BL, Gates ZP, Dashnau JL, Vanderkooi JM, Kent SBH (2008) Mirror image forms of snow flea antifreeze protein prepared by total chemical synthesis have identical antifreeze activities. *J Am Chem Soc* 130(30):9702–9707.
37. Kochendoerfer GG, et al. (2003) Design and chemical synthesis of a homogeneous polymer-modified erythropoiesis protein. *Science* 299(5608):884–887.
38. Wang P, et al. (2013) Erythropoietin derived by chemical synthesis. *Science* 342(6164):1357–1360.

39. Wintermann F, Engelbrecht S (2013) Reconstitution of the catalytic core of F-ATPase $(\alpha\beta)3\gamma$ from *Escherichia coli* using chemically synthesized subunit γ. *Angew Chem Int Ed Engl* 52(4):1309–1313.
40. Fang GM, et al. (2011) Protein chemical synthesis by ligation of peptide hydrazides. *Angew Chem Int Ed Engl* 50(33):7645–7649.
41. Zheng JS, Tang S, Qi YK, Wang ZP, Liu L (2013) Chemical synthesis of proteins using peptide hydrazides as thioester surrogates. *Nat Protoc* 8(12):2483–2495.
42. Fang GM, Wang JX, Liu L (2012) Convergent chemical synthesis of proteins by ligation of peptide hydrazides. *Angew Chem Int Ed Engl* 51(41):10347–10350.
43. Zheng JS, Tang S, Guo Y, Chang HN, Liu L (2012) Synthesis of cyclic peptides and cyclic proteins via ligation of peptide hydrazides. *ChemBioChem* 13(4):542–546.
44. Johnson ECB, Kent SBH (2006) Insights into the mechanism and catalysis of the native chemical ligation reaction. *J Am Chem Soc* 128(20):6640–6646.
45. Hackeng TM, Griffin JH, Dawson PE (1999) Protein synthesis by native chemical ligation: Expanded scope by using straightforward methodology. *Proc Natl Acad Sci USA* 96(18): 10068–10073.
46. Wan Q, Danishefsky SJ (2007) Free-radical-based, specific desulfurization of cysteine: A powerful advance in the synthesis of polypeptides and glycopolypeptides. *Angew Chem Int Ed Engl* 46(48):9248–9252.
47. Yan LZ, Dawson PE (2001) Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurization. *J Am Chem Soc* 123(4):526–533.
48. Tsumoto K, et al. (2004) Role of arginine in protein refolding, solubilization, and purification. *Biotechnol Prog* 20(5):1301–1308.
49. Georgescauld F, et al. (2014) GroEL/ES chaperonin modulates the mechanism and accelerates the rate of TIM-barrel domain folding. *Cell* 157(4):922–934.
50. Benoiton NL (2006) *Chemistry of Peptide Synthesis* (CRC, Boca Raton, FL), pp 93–124.
51. Mergler M, Dick F, Sax B, Stähelin C, Vorherr T (2003) The aspartimide problem in Fmoc-based SPPS. Part II. *J Pept Sci* 9(8):518–526.
52. Mergler M, Dick F, Sax B, Weiler P, Vorherr T (2003) The aspartimide problem in Fmoc-based SPPS. Part I. *J Pept Sci* 9(1):36–46.
53. Gibson DG, et al. (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 329(5987):52–56.

Supporting Information

Weinstock et al. 10.1073/pnas.1410900111

SI Materials and Methods

Recombinant DapA Proteins. The coding region of DapA was PCR amplified from *Escherichia coli* BL21 (DE3) (Novagen) genomic DNA using two primers:

- DapA_fwd: TGACCATATGTTCACGGGAAGTATTGTCG
- DapA_rev: TGACGGATCCTCCCTAAACTTTACAGCAAACC

Amplified product was then inserted into the pET14b vector (Novagen) between NdeI and BamHI restriction sites, yielding a construct with an N-terminal hexahistidine tag and thrombin cleavage site (pET14b_DapA). The A77C mutant of DapA, pET14b_DapA_A77C, was generated by mutating position 77 from GCT to TGC using the Q5 Site-Directed Mutagenesis Kit (NEB) with pET14b_DapA as the template, using two primers:

- DapAmut_fwd: TGACGAACATTGCGATGTGGTGATGATG
- DapAmut_rev: TGATTTAAGGTAGCGGAC

Mutation was confirmed by sequencing of the entire gene.

Proteins were expressed in BL21 (DE3) cells (Novagen). Cultures were grown in autoinduction media (1) in shake flasks at 37 °C to an OD_{600} of 0.7–1 and then allowed to grow for an additional 14–18 h at 19 °C. Cell pellets were suspended in buffer A [50 mM Tris (pH 7.3), 300 mM NaCl, 5 mM imidazole, 5% glycerol] and disrupted by sonication. Samples were clarified by centrifugation, and protein was isolated from the supernatant by applying to His-Select Ni-affinity resin (Sigma), thoroughly washing with buffer A, and eluting in buffer A containing 250 mM imidazole. Purified fractions were pooled, dialyzed against buffer A without imidazole, spin concentrated to ∼8 mg/mL, aliquoted, flash-frozen, and stored at −80 °C. A portion of the material was further purified via RP-HPLC and lyophilized for use in refolding studies. The amino acid sequence of the recombinantly expressed DapA used in the paper is identical to the synthetic DapA constructs described below.

Recombinant GroEL and GroES. GroEL and GroES were expressed in *E. coli* DH5α from the pBRE-groESL+ plasmid [gift from C. Georgopoulos (University of Utah, Salt Lake City)], which contains the *E. coli groE* operon. Cultures were grown at 37 °C in shake flasks to an OD_{600} of 0.6. An equal volume of 55 °C media was added and the cultures were shifted to 43 °C for 3 h to induce expression of *groE*. Cell pellets were resuspended in buffer [50 mM Tris (pH 8), 1 mM EDTA, 1 mM DTT] containing 0.17 mg/mL PMSF and lysed by incubation with lysozyme, followed by sonication and centrifugation. GroEL and GroES were purified as described in ref. 2 with a 5-mL HiTrap Q HP column (GE Healthcare) used instead of a DE-52 column. Eluted fractions containing GroES were dialyzed at 4 °C into 25 mM Tris (pH 8 at 4 °C), 0.5 mM EDTA, 1 mM DTT. One-molar NaOAc (pH 4.6) was added to 75 mM and the sample was stirred on ice for 15 min, filtered through a 0.2-μm pore-size membrane, and then purified on a 1-mL HiTrap SP XL column (GE Healthcare) with a NaCl gradient (0–500 mM over 15 column volumes) in 50 mM NaOAc, 0.5 mM EDTA, 2 mM DTT (pH 4.6). Fractions containing GroES as judged by SDS/PAGE and liquid chromatography (LC)/MS analysis were pooled, dialyzed against 25 mM Tris (pH 7.5; pH 8 at 4 °C), 0.5 mM EDTA, 1 mM DTT, and concentrated in an Amicon Ultra-15 centrifugal concentrator. Concentrations were determined by absorbance at 280 mm (GroES) or Bradford method (GroEL). Glycerol was added to 10%, and aliquots were flash-frozen and stored at −80 °C.

Peptide Synthesis. Peptides were synthesized via Fmoc–SPPS on a commercial peptide synthesizer (Prelude instrument; Protein Technologies, Inc). L-amino acids were from AAPPTec; D-amino acids were from Peptides International and CBL Patras; and both L- and D-Val-Ser($\psi^{Me,Me}$pro)-OH pseudoproline dipeptides (incorporated in peptide segments DapA 2 and 5) were obtained from AAPPTec. Multiple batches of each peptide were synthesized at ∼15-μmol scale. Peptides containing the C-terminal region of DapA, i.e., DapA 8 and DapA 7–8, were synthesized on TentaGel R RAM resin (0.19 mmol/g, RAPP Polymere) to generate C-terminal peptide amides. All other peptides were synthesized on 2-hydrazine chlorotrityl resin (ChemPep) to generate C-terminal peptide hydrazides. For each hydrazide peptide, the first residue was manually attached to the resin to a desired density between 0.05 and 0.1 mmol/g by dissolving 0.075 mmol amino acid and 0.075 mmol Oxyma Pure (Novabiochem) in 1 mL 1:1 dimethylformamide (DMF):DCM, followed by activation with 0.08 mmol *N*,*N*′-diisopropylcarbodiimide for 10 min before addition to 180 mg of resin for 2 h. The resin was then washed with DMF, and unreacted sites were capped by treatment with 1/1 acetic anhydride/0.6 M *N*-methylmorpholine (NMM) in DMF for 15 min. After capping, the density of the first amino acid (desired 0.05–0.1 mmol/g) was confirmed by measuring Fmoc release, after 30-min treatment with 20% piperidine in DMF (assuming Fmoc ε_{301} = 7,800 $M^{-1}\cdot cm^{-1}$), using a spectrophotometer.

Automated peptide synthesis was performed with the following parameters. For Fmoc deprotection: (*i*) 2 mL of 20% piperidine in DMF; (*ii*) mix 3 × 3 min (new solvent delivered for each mixing cycle). For amino acid coupling: (*i*) 0.66 mL of 200 mM Fmoc-protected amino acid in NMP; (*ii*) 0.66 mL of 198 mM HATU in DMF; (*iii*) 0.5 mL of 600 mM NMM in DMF; and (*iv*) mix 1 × 25 min. For DMF washing (performed between deprotection and coupling steps): (*i*) 2 mL of DMF; (*ii*) mix 6 × 30 s (new solvent delivered for each mixing cycle). Upon completion of the peptide chain, resins were washed with DCM and dried (using vacuum) for >10 min and then cleaved for 2.5 h using a TFA mixture specific to the peptide sequence. Peptides containing Cys: 3.7 mL TFA, 100 μL water, 100 μL TIS, 100 μL EDT; peptides containing Met (±Cys): 3.7 mL TFA, 100 μL water, 100 μL TIS, 100 μL EDT, 60 mg NH_4I.

Following cleavage, the mixture was isolated from the resin via filtration, precipitated (and washed 3×) with cold diethyl ether, and peptide was collected by centrifugation. Pelleted peptides were dried overnight in a vacuum desiccator before HPLC purification.

HPLC and LC/MS. Analytical reverse-phase HPLC was performed on crude and purified peptides using Phenomenex Jupiter 4-μm Proteo C12 90 Å (150 × 4.6 mm) and Phenomenex Jupiter 5-μm C4 300 Å (150 × 4.6 mm) columns at 1 mL/min with a water/acetonitrile gradient in 0.1% TFA. The initial crude peptide analytical HPLC traces were used to scout the column (either C12 or C4) and gradient method to be used for the preparative-scale purification. Analytical HPLC runs were also used to present the purity of the peptide segments and ligation products (see *SI Appendix*).

Preparative reverse-phase HPLC of crude peptides was performed on either Phenomenex Jupiter 4-μm Proteo C12 90 Å (250 × 21.2 mm) or Phenomenex Jupiter 10-μm C4 300 Å (250 × 21.2 mm) at 10 mL/min with a water/acetonitrile gradient in 0.1% TFA. Semipreparative reverse-phase HPLC of ligation products was performed on a Phenomenex Jupiter 10-μm C4

300 Å (250 × 10 mm) at 5 mL/min with a water/acetonitrile gradient in 0.1% TFA.

Fractions collected from preparative and semipreparative runs (30-s intervals) were analyzed by LC/MS on a Phenomenex Aeris WIDEPORE 3.6-μm C4 (50 × 2.1 mm) column at 0.4 mL/min with a water/acetonitrile gradient in 0.1% formic acid on an AB Sciex API 3000 LC/MS/MS system. Fractions containing correct product (based on LC/MS) were pooled and lyophilized.

Analytical reverse-phase HPLC traces of purified peptides were normalized to the maximum intensity point of the trace. The major observed deconvoluted masses from mass spectrometry were calculated using Bayesian Peptide and Protein Reconstruct Tools in Analyst 1.5.1 Software (AB Sciex). The analytical gradients used for each peptide are listed in the *SI Appendix*.

Peptide Ligation. For a typical ligation reaction, the N-terminal peptide segment (containing C-terminal hydrazide for in situ conversion to azide and then thioester) was dissolved (0.8–3.0 mM) in activation buffer [6 M GuHCl, 100 mM sodium phosphate (pH 3.0)] and prechilled to −20 °C for 10 min. Next, hydrazide activation (conversion of hydrazide to azide) was performed by adding sodium nitrite to a final concentration of 5–20 mM (lower concentrations were found to diminish hydrolysis of DapA 2 thioester) from a 200-mM stock in water (pH adjusted to 3.8–4.2), mixing, and incubating at −20 °C for 20 min. The molar ratio of $NaNO_2$ to peptide hydrazide should be greater than 10 (3) to achieve sufficient activation. During the activation step, the C-terminal peptide segment (containing N-terminal cysteine) was dissolved in ligation buffer [6 M GuHCl, 200 mM 4-mercaptophenylacetic acid (MPAA) (4), 200 mM sodium phosphate, pH adjusted to 7.0–7.2]. Typically, a 2× molar ratio of the C-terminal peptide segment (cysteine peptide) over the N-terminal peptide segment (hydrazide peptide) was used in the ligations, except in the final ligation where an excess of DapA 1–4 over DapA 5–8 was used. In ligations involving DapA 2, 3, and 6 peptides, it was found that increasing the initial GuHCl concentration to 8 M improved solubility (final GuHCl concentration was ∼7 M). Following activation of the N-terminal segment, the solutions containing N- and C-terminal segments were combined, pH adjusted to 6.8–7.0 with 2 M or 6 M NaOH (using rapid mixing), and allowed to react at 25 °C for 5–20 h. This final pH adjustment should be done carefully to prevent thioester hydrolysis at pH values >7.0 (5). Accordingly, it was found that this pH adjustment step could be conveniently simplified by using activation and ligation buffers of different sodium phosphate concentrations. Here, the pH 3 activation buffer was 100 mM sodium phosphate, and the pH 7 buffer was 200 mM sodium phosphate. This approach ultimately reduced the amount of NaOH needed to adjust the final ligation pH to 6.8–7.0 after mixing the two buffers together.

Following ligation, reactions were quenched by addition of freshly prepared tris(2-carboxyethyl)phosphine (TCEP) to ∼130 mM and incubated for >10 min. Addition of TCEP is essential to reduce all MPAA in the solution; inadequate reduction with TCEP led to dramatic precipitation problems during workup. Further, the TCEP reduction should be performed at pH >5 to prevent MPAA precipitation (5). Following reduction, the volume was increased to 2.5 mL with 6 M GuHCl in 5% AcOH, and then brought to a total volume of 5 mL with 5% AcOH (final pH ∼2.0–3.0). Sample was then vortexed and sonicated thoroughly and clarified via centrifugation, and peptides were purified by HPLC.

Desulfurization. Desulfurization was based on ref. 6 with the modification of replacing *t*-butylthiol with glutathione (7). Desulfurization buffer [6 M GuHCl, 100 mM phosphate (pH 6.5)] was sparged with He and used to prepare a 550-mM stock of TCEP (pH readjusted with NaOH to 6.5 after the addition of TCEP) and a 53-mM stock of reduced glutathione. A 240-mM stock of VA-044 was prepared in sparged water. Peptide was dissolved in 230 μL of desulfurization buffer at a concentration of ∼3 mM. Reagents were added to the peptide solution in the following order: 204 μL glutathione stock, 204 μL VA-044 stock, and 720 μL TCEP stock. The reaction was layered with argon and incubated on a tube rotisserie overnight at 37 °C. Upon completion, the reaction volume was increased to 2.5 mL with 6 M GuHCl in 5% AcOH, and then brought to a total volume of 5 mL with 5% AcOH. The sample was sonicated and clarified via centrifugation, and peptides were purified by HPLC.

Acm Deprotection. Acm deprotection was performed as described in ref. 8 by adding AgOAc to HPLC fractions containing Acm-protected peptide to a final concentration of 20 mM, covering with argon, and incubating overnight with gentle agitation at 25 °C. DTT was added to 24 mM to quench the reaction, which resulted in an immediate metal–DTT precipitate, followed by centrifugation to clarify the solution, dilution with 0.1% TFA in water (to reduce acetonitrile concentration), and HPLC purification.

Enzyme Activity Assay. See main text.

Arginine-Assisted Folding. See main text.

Chaperone Refolding Assay. See main text.

DL-Aspartate-β-semialdehyde. DL-aspartate-β-semialdehyde (D/L–ASA) for use in refolding assays was prepared essentially as described by Black and Wright (9) from DL-allylglycine except that the purification on Dowex resin was omitted and the compound was aliquoted and stored in 1 M HCl at −80 °C. Stocks were neutralized with NaOH immediately before preparation of assay buffer. Racemic ASA was selected, because it has been demonstrated that the recognition of ASA by DapA is stereoselective; the opposite chirality substrate is neither a substrate nor an inhibitor of enzyme activity (10).

Circular Dichroism. All CD spectra were recorded on an AVIV Model 410 spectrophotometer (AVIV) in 100 mM ammonium bicarbonate buffer in a 1 mm QS quartz cuvette (Starna) at 25 °C. Wavelength scans were performed at 1-nm resolution with 1-s averaging time. Data from triplicate scans were averaged, blank subtracted, and normalized to mean residue ellipticity by the following equation: $[\theta] = 100 \times \theta / C \times l \times (n - 1)$, where C is concentration of protein in mM, l is path length in centimeters, and n is the number of residues in the protein. The concentrations of the protein samples used for CD experiments were 15.56 μM for recombinant DapA, 8.27 μM for synthetic D-DapA, and 5.13 μM for synthetic L-DapA.

Sequence Analysis. BLAST analysis was performed using the BLASTp algorithm with default parameters on the DapA protein sequence from *E. coli* BL21(DE3) (GenBank accession no. ACT44191.1). The sequences were aligned with the COBALT Constraint-Based Multiple Protein Alignment Tool (11) and analyzed with Jalview (12).

Structural Analysis. The crystal structure of DapA (PDB ID code 1DHP) was analyzed using the PyMol Molecular Graphics System (Schrödinger, LLC).

1. Studier FW (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 41(1):207–234.

2. Kamireddi M, Eisenstein E, Reddy P (1997) Stable expression and rapid purification of *Escherichia coli* GroEL and GroES chaperonins. *Protein Expr Purif* 11(1):47–52.

3. Zheng JS, Tang S, Guo Y, Chang HN, Liu L (2012) Synthesis of cyclic peptides and cyclic proteins via ligation of peptide hydrazides. *ChemBioChem* 13(4):542–546.
4. Johnson ECB, Kent SBH (2006) Insights into the mechanism and catalysis of the native chemical ligation reaction. *J Am Chem Soc* 128(20):6640–6646.
5. Zheng JS, Tang S, Qi YK, Wang ZP, Liu L (2013) Chemical synthesis of proteins using peptide hydrazides as thioester surrogates. *Nat Protoc* 8(12):2483–2495.
6. Wan Q, Danishefsky SJ (2007) Free-radical-based, specific desulfurization of cysteine: A powerful advance in the synthesis of polypeptides and glycopolypeptides. *Angew Chem Int Ed Engl* 46(48):9248–9252.
7. Haase C, Rohde H, Seitz O (2008) Native chemical ligation at valine. *Angew Chem Int Ed Engl* 47(36):6807–6810.
8. Bang D, Chopra N, Kent SBH (2004) Total chemical synthesis of crambin. *J Am Chem Soc* 126(5):1377–1383.
9. Black S, Wright NG (1955) Aspartic beta-semialdehyde dehydrogenase and aspartic beta-semialdehyde. *J Biol Chem* 213(1):39–50.
10. Dobson RC, Gerrard JA, Pearce FG (2004) Dihydrodipicolinate synthase is not inhibited by its substrate, (S)-aspartate beta-semialdehyde. *Biochem J* 377(Pt 3):757–762.
11. Papadopoulos JS, Agarwala R (2007) COBALT: Constraint-based alignment tool for multiple protein sequences. *Bioinformatics* 23(9):1073–1079.
12. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25(9):1189–1191.

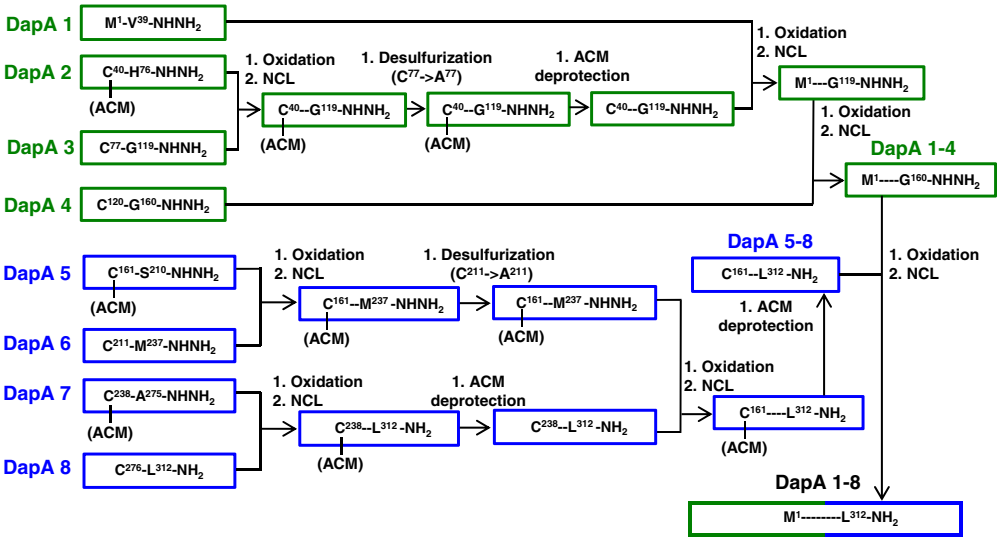


Fig. S1. Initial synthetic strategy for total chemical synthesis of 312-residue DapA, including peptide segments (DapA 1–4, green and DapA 5–8, blue) with ligation hydrazide and cysteine residues indicated. Oxidation (hydrazide to azide), native chemical ligation (NCL), and acetamidomethyl (ACM) cysteine deprotection steps are also indicated.

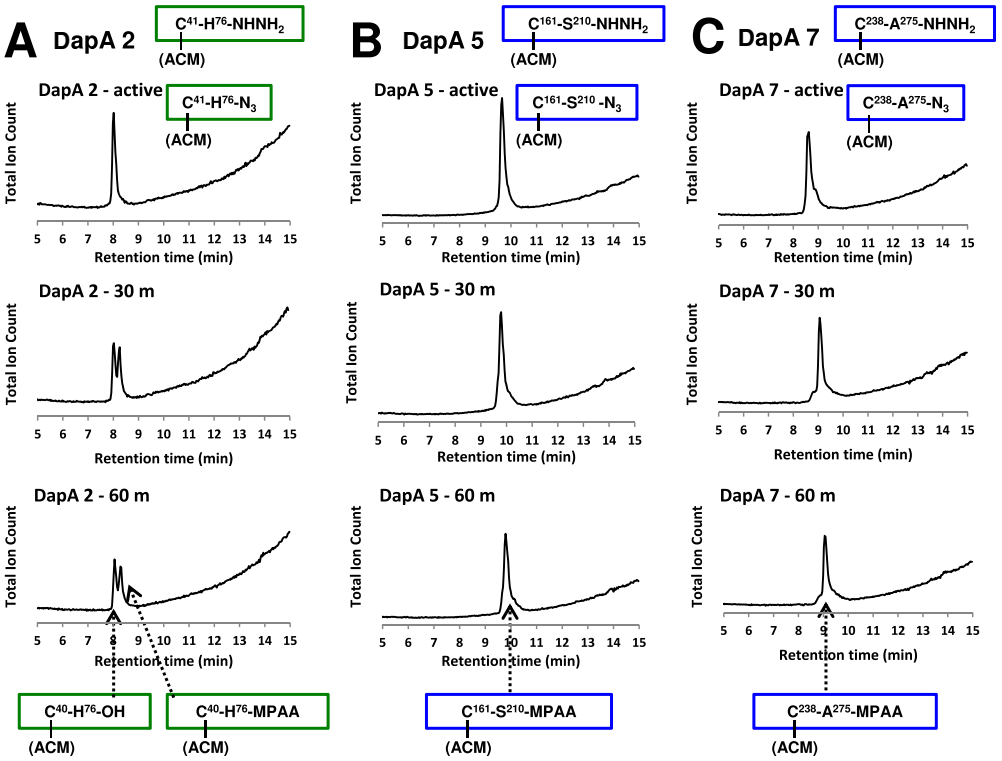


Fig. S2. Hydrolysis of DapA 2 thioester. Cys(ACM)-protected peptide hydrazides DapA 2 (*A*), 5 (*B*), and 7 (*C*) were dissolved in activation buffer at ~0.4 mM [6 M GuHCl, 100 mM sodium phosphate (pH 3.0)] and activated with $NaNO_2$ (10 mM) at −20 °C for 20 min to generate active C-terminal azides (*Top*). Azide groups were then converted to MPAA thioesters by equal-volume addition of ligation buffer (6 M GuHCl, 200 mM 4-mercaptophenylacetic acid, 200 mM sodium phosphate, pH adjusted to 7.0–7.2) and pH readjusted to 6.8–7.0. Postmixing time-points at 30 and 60 min (*Middle* and *Bottom*) were collected, quenched with 100 mM TCEP in 1% TFA, and analyzed by LC/MS. Significant hydrolysis of DapA 2 (histidine) thioester was observed in less than 60 min, whereas DapA 5 and 7 thioesters were stable for several hours.

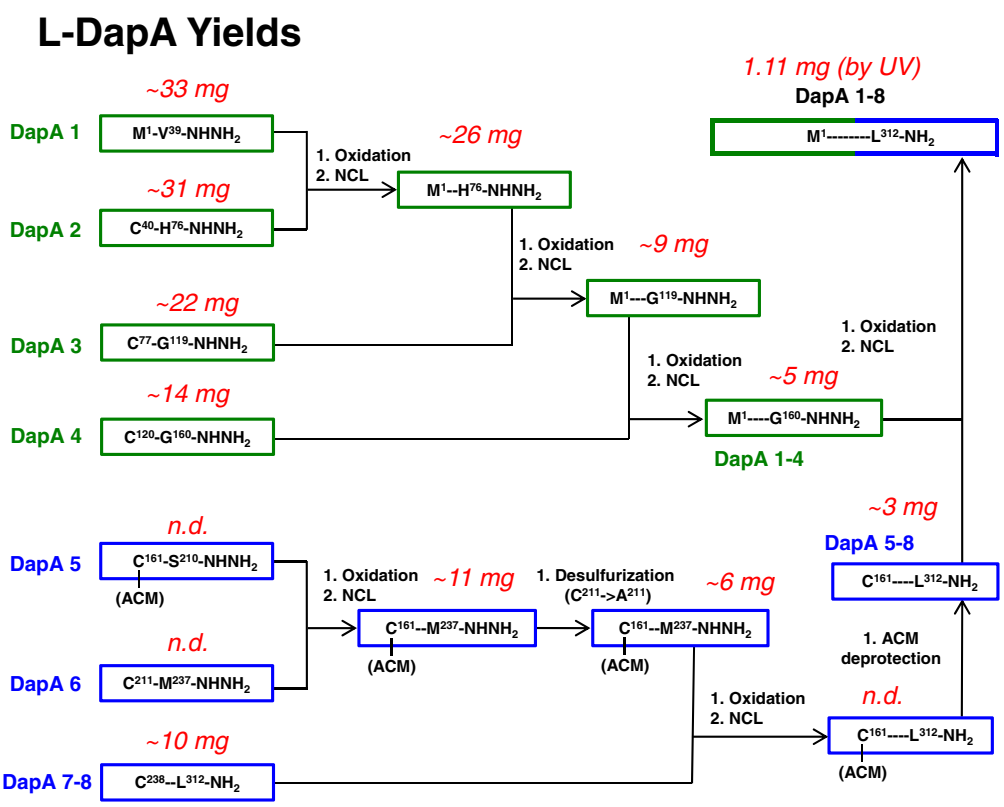


Fig. S3. Yields obtained during the total chemical synthesis of 312 residue L-DapA, including peptide segments (DapA 1–4, green and DapA 5–8, blue) with ligation hydrazide and cysteine residues. Oxidation (hydrazide to azide), NCL, and ACM cysteine deprotection steps are also indicated.

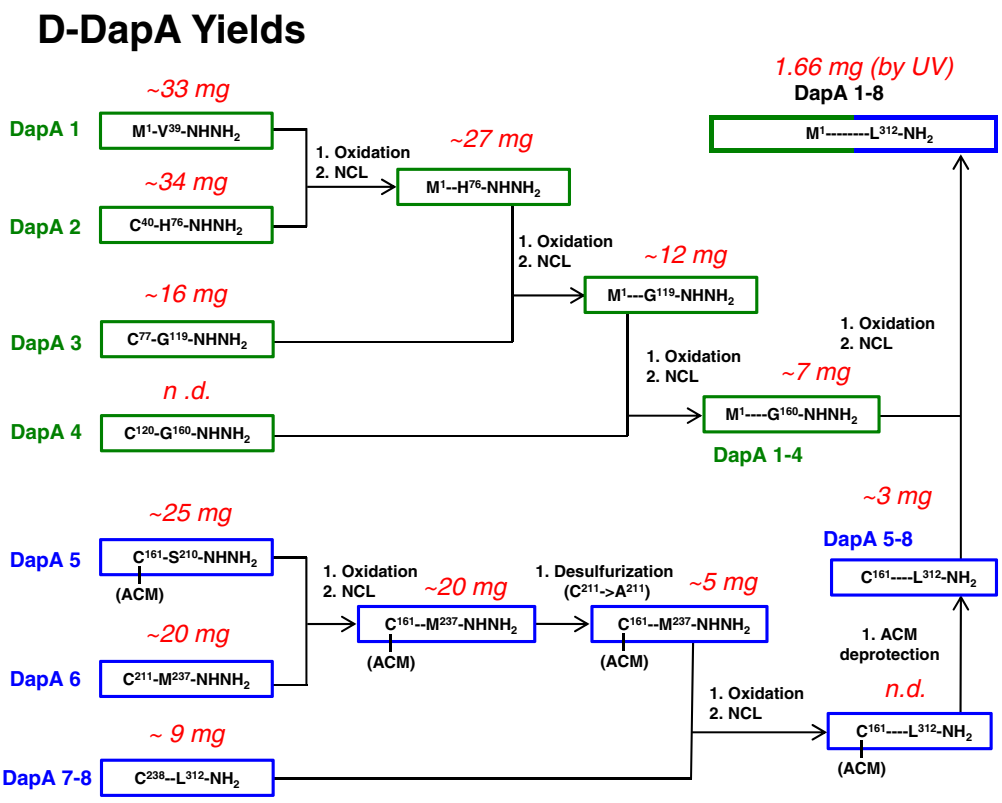


Fig. S4. Yields obtained during the total chemical synthesis of 312-residue D-DapA, including peptide segments (DapA 1–4, green and DapA 5–8, blue) with ligation hydrazide and cysteine residues. Oxidation (hydrazide to azide), NCL, and ACM cysteine deprotection steps are also indicated.

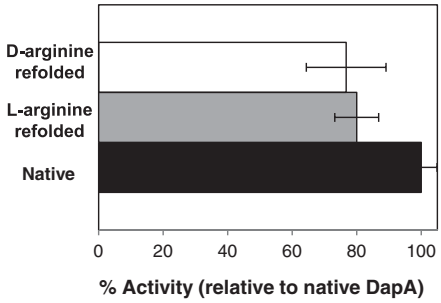


Fig. S5. Arginine-assisted refolding of DapA. Enzyme activity of recombinant DapA refolded with D-arginine or L-arginine (pre-SEC purification), normalized to native DapA. Error bars indicate SD of at least three assays.

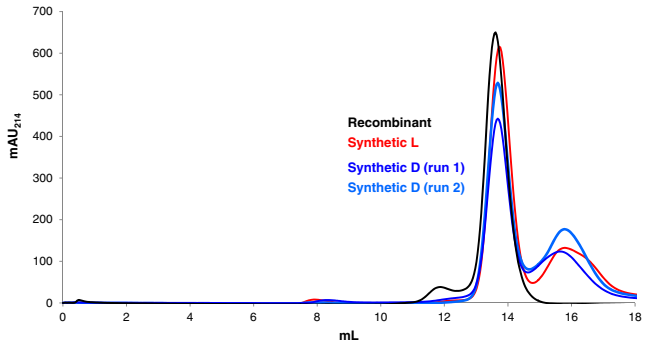


Fig. S6. SEC purification of DapA following arginine-assisted folding. Chromatograph from purification of DapA tetramers after arginine-assisted folding: synthetic L-DapA (one run) and synthetic D-DapA (two runs). Native recombinant DapA tetramer (normalized) is shown for comparison.

Table S1. Absorbance values (A_{562}) at specified time points in refolding assays

	Recombinant DapA*				Synthetic DapA A77C, pre-SEC				Synthetic DapA A77C, post-SEC†			
	WT		A77C		L-		D-		L-		D-	
Time, min	+GroEL/ES	−GroEL/ES	+GroEL/ES	−GroEL/ES	+GroEL/ES	−GroEL/ES	+GroEL/ES	−GroEL/ES	+GroEL/ES	−GroEL/ES	+GroEL/ES	−GroEL/ES
1	0.097		0.045		0.012		0.006		0.028		0.000	
2	0.192		0.146		0.045		0.007		0.084		0.008	
3	0.230	0.001	0.184	0.000	0.063	0.000	0.012	0.000	0.126	0.003	0.034	0.003
5	0.265		0.210		0.078		0.025		0.138		0.072	
7	0.272		0.225		0.085		0.033		0.150		0.090	
9	0.277	0.004	0.225	0.000	0.088	0.002	0.039	0.002	0.155	0.001	0.096	0.000
13	0.275	0.005	0.231	0.000	0.094	0.003	0.041	0.008	0.155		0.103	0.004
16.5	0.274		0.230		0.097		0.046		0.156		0.107	
21	0.292		0.241		0.100		0.046		0.153		0.109	0.002
30	0.293	0.008	0.248	0.000	0.110	0.005	0.047	0.003	0.168	0.000	0.115	
44	0.302		0.253		0.106		0.053		0.170		0.135	
58	0.293	0.012	0.250	0.003	0.108	0.008	0.053	0.011	0.172	0.000	0.120	0.004

Chaperone-mediated refolding of DapA recombinant WT vs. A77C, synthetic L- vs. D-A77C, pre-SEC purification, and synthetic L- vs. D-A77C, post-SEC purification.

*In Fig. 2*D*, normalized to native recombinant DapA (WT): 0.363.

†In Fig. 5, normalized to highest point in GroEL/ES refolding of recombinant DapA (A77C): 0.253.

Other Supporting Information Files

SI Appendix (PDF)

CHAPTER 3

DESIGN AND CHARACTERIZATION OF EBOLAVIRUS GP PREHAIRPIN INTERMEDIATE MIMICS AS DRUG TARGETS

Tracy R. Clinton, Matthew T. Weinstock, Michael T. Jacobsen, Nicolas Szabo-Fresnais, Maya J. Pandya, Frank G. Whitby, Andrew S. Herbert, Laura I. Prugar, Rena McKinnon, Christopher P. Hill, Brett D. Welch, John M. Dye, Debra M. Eckert, Michael S. Kay

Reprinted with permission from Protein Science 24(4), 446-463

THE PROTEIN SOCIETY

Design and characterization of ebolavirus GP prehairpin intermediate mimics as drug targets

Tracy R. Clinton,[1] Matthew T. Weinstock,[1] Michael T. Jacobsen,[1] Nicolas Szabo-Fresnais,[1,2] Maya J. Pandya,[1] Frank G. Whitby,[1] Andrew S. Herbert,[3] Laura I. Prugar,[3] Rena McKinnon,[4] Christopher P. Hill,[1] Brett D. Welch,[4] John M. Dye,[3] Debra M. Eckert,[1]* and Michael S. Kay[1]*

[1]Department of Biochemistry, University of Utah School of Medicine, Salt Lake City, Utah 84112-5650
[2]Cardiology Section, Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, Utah 84148
[3]U.S. Army Medical Research Institute of Infectious Diseases, Fort Detrick, Frederick, Maryland 21702-5011
[4]D-Peptide Research Division, Navigen, Inc., Salt Lake City, Utah 84108

Received 21 August 2014; Accepted 1 October 2014
DOI: 10.1002/pro.2578
Published online 6 October 2014 proteinscience.org

Abstract: Ebolaviruses are highly lethal filoviruses that cause hemorrhagic fever in humans and non-human primates. With no approved treatments or preventatives, the development of an anti-ebolavirus therapy to protect against natural infections and potential weaponization is an urgent global health need. Here, we describe the design, biophysical characterization, and validation of peptide mimics of the ebolavirus N-trimer, a highly conserved region of the GP2 fusion protein, to be used as targets to develop broad-spectrum inhibitors of ebolavirus entry. The N-trimer region of GP2 is 90% identical across all ebolavirus species and forms a critical part of the prehairpin intermediate that is exposed during viral entry. Specifically, we fused designed coiled coils to the N-trimer to present it as a soluble trimeric coiled coil as it appears during membrane fusion. Circular dichroism, sedimentation equilibrium, and X-ray crystallography analyses reveal the helical, trimeric structure of the designed N-trimer mimic targets. Surface plasmon resonance studies validate that the N-trimer mimic binds its native ligand, the C-peptide region of GP2. The longest N-trimer mimic also inhibits virus entry, thereby confirming binding of the C-peptide region during viral entry and the presence of a vulnerable prehairpin intermediate. Using phage display as a model system, we validate the suitability of the N-trimer mimics as drug screening targets. Finally, we describe the foundational work to use the N-trimer mimics as targets in mirror-image phage display, which will be used to identify D-peptide inhibitors of ebolavirus entry.

Keywords: ebolavirus; filovirus entry; ebolavirus GP2; prehairpin intermediate; designed coiled coil; N-trimer; phage display; mirror-image phage display

Additional Supporting Information may be found in the online version of this article.

Disclosure: D.M.E., B.D.W. and M.S.K. are Scientific Directors and equity holders of the D-Peptide Research Division of Navigen, which is commercializing D-peptide inhibitors of viral entry. D.M.E. and M.S.K. are also Navigen consultants.

Matthew T. Weinstock current address is Synthetic Genomics, Inc., La Jolla, California 92037

*Correspondence to: Debra M. Eckert; Department of Biochemistry, University of Utah School of Medicine, 15 N. Medical Drive East, Rm 4100, Salt Lake City, UT 84112. E-mail: deckert@biochem.utah.edu or Michael S. Kay; Department of Biochemistry, University of Utah School of Medicine, 15 N. Medical Drive East, Rm 4100, Salt Lake City, UT 84112. E-mail: kay@biochem.utah.edu

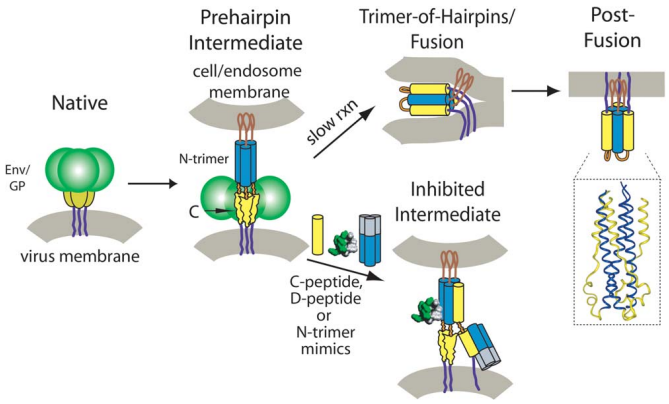


Figure 1. Model for membrane fusion mediated by enveloped virus surface glycoproteins. The HIV-1 and ebolavirus entry events are predicted to be similar. First, the surface glycoprotein (Env for HIV-1, GP for ebolavirus) facilitates viral attachment to the cell and, for ebolavirus, the virus is endocytosed and then cleaved by endosomal proteases. Engagement of the virus receptors (CD4 and a chemokine receptor for HIV-1, NPC1 for ebolavirus) is followed by a conformational change in Env/GP, and insertion of the fusion peptide/loop (brown) into the host cell membrane. At this stage, the virus is in a transient state that bridges both membranes, termed the "prehairpin intermediate," which is vulnerable to inhibition. In the absence of an inhibitor, the Env/GP structure slowly resolves into the highly stable trimer-of-hairpins structure, juxtaposing the two membranes, and leading to membrane fusion. The inset shows the high resolution structure of the ebolavirus trimer-of-hairpins (PDB: 2EBO).[27] In HIV, it has been shown that inhibitors that bind to either the N-trimer (blue) or C-peptide (yellow) regions are capable of inhibiting entry [as reviewed in Ref. (17)].

Introduction

Ebolaviruses are enveloped, negative-strand RNA viruses that cause severe hemorrhagic fever.[1] Since its identification in 1976, there have been over 20 reported natural ebolavirus outbreaks, the majority since 2000, and several accidental laboratory exposures with an overall mortality rate >60%.[2] Alarmingly, in 2014 the largest known outbreak is occurring in western Africa[3] and has crossed international borders. Currently, no vaccines or therapeutics are FDA approved. Because of ease of transmission, high mortality, and potential for a severe impact on public health, the CDC places ebolaviruses in its highest category of potential agents of bioterrorism.[4] There is a vital need for preventatives and/or therapeutics to protect against future natural, accidental, or deliberate outbreaks.

Ebolavirus entry into host cells, a critical step to infection, is mediated by the viral surface glycoprotein (GP), a class I fusion protein.[5] GP comprises two disulfide-linked subunits, one surface exposed (GP1) and one embedded in the viral membrane (GP2).[6,7] Following binding to host cells via cell surface attachment factors, the virus is endocytosed. Endosomal cysteine proteases, cathepsins B and L, cleave off much of GP1, exposing the binding site for the receptor, endosomal NPC1.[8–12] At this point, the fusion mechanism is thought to mimic that of other well characterized viral class I fusion proteins, such as HIV-1 and influenza[13–15] (Fig. 1). GP2 forms a transient conformation ("prehairpin intermediate") embedded in both the virus (via the transmembrane domain) and host cell (via the fusion loop) membranes. This prehairpin intermediate exposes a trimeric coiled coil, formed by the N-terminal region (N-trimer), and the C-terminal region (C-peptide). Slow collapse of the intermediate into a very stable trimer-of-hairpins structure, with the C-peptide binding into the grooves on the N-trimer, juxtaposes the virus and cell membranes, leading to membrane fusion. In ebolavirus entry, the low pH of the endosome contributes to the stability of the trimer-of-hairpins.[16]

In the case of HIV-1, the prehairpin intermediate has been exploited to develop highly potent viral entry inhibitors (Fig. 1). Peptides and proteins that bind with high affinity to either the N-trimer or C-peptide regions prevent formation of the trimer-of-hairpins, thereby halting viral entry [as reviewed in Ref. (17)]. The most potent of the HIV entry inhibitors, chol-PIE12-trimer, binds to the conserved hydrophobic pockets of the HIV N-trimer and inhibits HIV entry with low picomolar potency.[18] Since filoviruses share a similar mechanism of entry as HIV-1, they are likely vulnerable to inhibitors that similarly target the prehairpin intermediate.

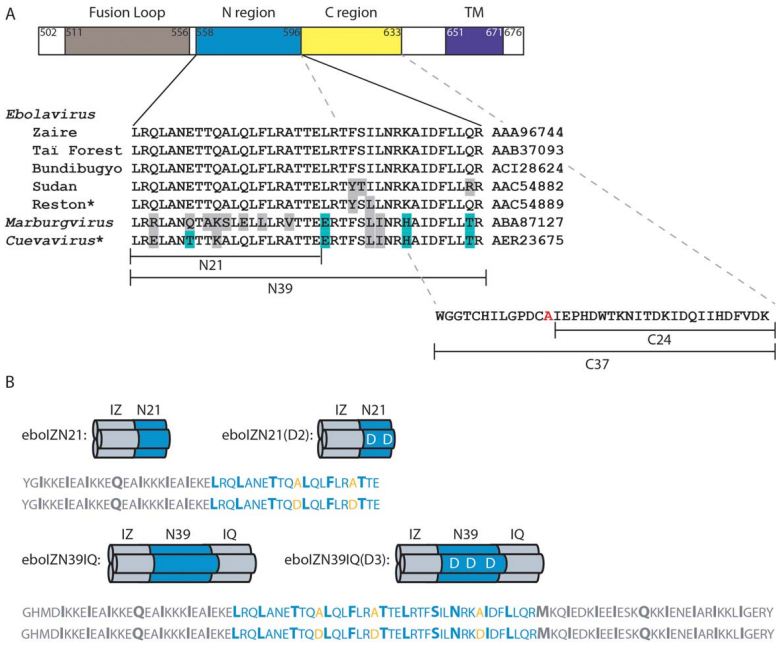


Figure 2. Conservation of the ebolavirus GP N-trimer and design of peptide N-trimer mimics. (A) Schematic of the primary structure of ebolavirus GP2, indicating the fusion loop (brown), N-trimer (blue), C-peptide (yellow), and transmembrane domain (TM, purple), all shown approximately to scale. The sequences of the N-trimer region (residues 558–596) and the C-peptide region (residues 597–633) (*Zaire ebolavirus* species, representative Mayinga strain isolated in Zaire in 1976[60]) contained in the peptides described in this study are indicated. In the C-peptide, cysteine 609, which is proposed to disulfide bond with GP1,[76] is mutated to alanine in our constructs (red). Below the Zaire N-trimer is an alignment of the sequences from the 4 additional ebolavirus species plus Marburgvirus and Cuevavirus filoviruses. Genbank accession codes are indicated (right). Conserved changes (score of 0 or higher in BLOSUM62 matrix[77]) are highlighted in gray, nonconserved in cyan. Notably, 3/5 and 5/5 ebolavirus species are 100% identical in the N39 and N21 regions, respectively. The 2014 epidemic is caused by the *Zaire ebolavirus* species and is 100% identical in this region.[3] *Reston and likely Cuevavirus (Lloviu virus) are not pathogenic to humans. (B) Schematics and sequences of the N-trimer mimics and their corresponding binding site mutants [eboIZN21 and eboIZN21(D2); eboIZN39IQ and eboIZN39IQ(D3)]. The designed coiled coils, IZ_m and IQ, are shown in gray, while the ebolavirus N-trimer is shown in blue. The a and d positions of the coiled-coil heptad repeats are indicated by a larger bolded font, including a stutter at the N-terminal end of the ebolavirus N-trimer as seen in the crystal structures,[27,28] where the coil is underwound, leading to an atypical 3-4-4-3 pattern (instead of the standard 3-4, or a-g, periodicity of a heptad repeat). The alanine residues along the C-peptide binding groove that are mutated to aspartate in the binding site mutants are shown (orange).

There are five known species of ebolavirus, four of which are pathogenic to humans. The vast majority of promising preventative and therapeutic candidates with efficacy against ebolavirus in animal models, such as vaccines, antibodies, and antisense compounds (e.g.,[19–23]), are species-specific, resulting in limited breadth and difficulty in combating emerging species. The N-trimer of the prehairpin intermediate provides a highly conserved target for potential broad-spectrum inhibitors. Indeed, although the overall sequence identity of GP across all known ebolavirus species is only 42%, the N-peptide region is 90% identical, and all changes are conservative [Fig. 2(A)].

Here, we describe the development of ebolavirus N-trimer mimics that will be useful in a variety of drug discovery platforms to screen small molecule, antibody, and peptide libraries for entry inhibitors that target this conserved region. Specifically, we have designed and characterized peptide mimics of ebolavirus N-trimers, validated their use as drug discovery tools, and explored conditions that can be applied directly to phage display drug discovery endeavors. In addition, using one of our N-trimers as an inhibitor itself (to

target the C-peptide region of the prehairpin intermediate), we have confirmed the vulnerability of the ebolavirus GP prehairpin intermediate to entry inhibition.

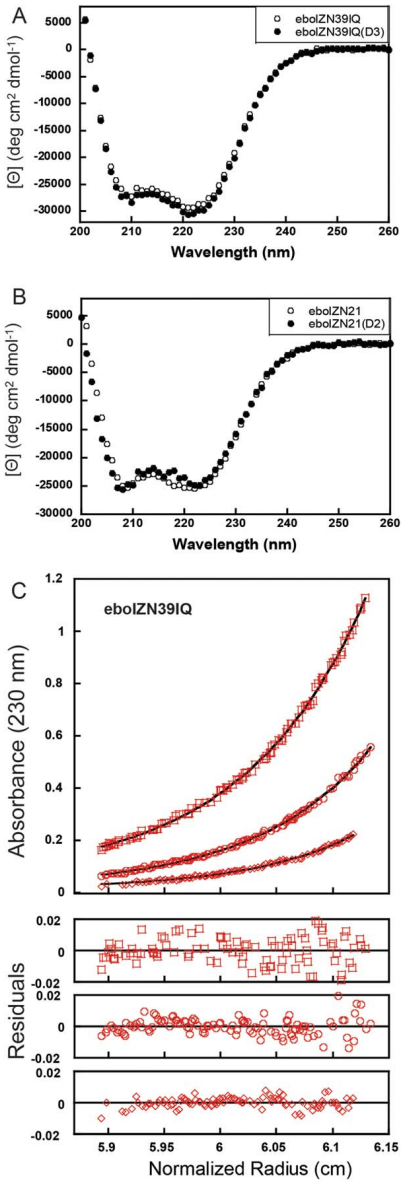


Figure 3. Biophysical analyses of ebolavirus N-trimer mimics. (A) CD spectra of 11.4 μ*M* eboIZN39IQ and 11.1 μ*M* eboIZN39IQ(D3) at 25°C. Both spectra indicate a highly helical conformation with 81 and 83% helicity, respectively. (B) CD spectra of 18.0 μ*M* eboIZN21 and 25.3 μ*M* eboIZN21(D2) at 25°C, also indicate a highly helical conformation with 73 and 71% helicity, respectively. (C) Analytical ultracentrifugation (AUC) sedimentation equilibrium analysis of eboIZN39IQ, shown as representative AUC data. 10, 5, and 2.5 μ*M* peptide solutions were centrifuged at 18,000, 21,000, and 24,000 rpm at 4°C on a Beckman XLA. All data were globally fit to a single ideal species, and an observed molecular weight of 39,762 Da was determined for an M_{obs}/M_{calc} of 3.31. The data (open symbols) and fit (solid lines) are shown for the lowest speed.

Results and Discussion

N-trimer mimic design

Based on our previous HIV-1 work,[24,25] we set out to design soluble peptide mimics of the N-trimer region of the ebolavirus GP prehairpin intermediate by fusing stable, soluble, designed trimeric coiled coils to the N-trimer sequence [Fig. 2(B)]. As with HIV-1, the ebolavirus N-trimer aggregates when produced in isolation. We were interested in presenting the entire N-trimer groove as well as a smaller, more conserved region of the N-trimer to provide flexibility in drug screening. Our initial designs, in which we fused the coiled coil IZ_m(24) to the N-terminus of N-trimer segments of 29 and 39 amino acids, were aggregated as determined by analytical ultracentrifugation (AUC) sedimentation equilibrium experiments (data not shown). To overcome this problem, we fused an additional trimeric coiled coil, GCN4-pI_QI' (IQ)[26] to the C-terminus of the ebolavirus N-trimer segment. The resulting peptide, eboIZN39IQ presents the full ebolavirus N-trimer (determined from available trimer-of-hairpins crystal structures[27,28]) as a trimeric coiled coil, as shown by circular dichroism (CD) [Fig. 3(A)] and AUC [Fig. 3(C) and Table I]. eboIZN39IQ is very stable, as indicated by similar CD spectra at 25, 37, and 50°C (Table I). The ultimate goal for an ebolavirus N-trimer mimic is to use it as a target in drug screening to identify inhibitors of ebolavirus entry. Since these inhibitors will bind to the virus in the endosome, all biophysical analyses were performed at pH 5.8 to mimic endosomal pH.

To produce a smaller target that presents a 100% identical region of the N-trimer (across all ebolavirus species), IZ_m was fused to the N-terminal 21 amino acids of the N-trimer, resulting in eboIZN21 (Fig. 2). Circular dichroism indicates that eboIZN21 is highly helical [Fig. 3(B)], and AUC and gel filtration studies show that it is largely trimeric with a slight tendency to form higher order aggregates (Table I and Supporting Information Fig. S1). X-ray crystallography studies confirm the trimeric

Table I. *Biophysical Analyses of N-Trimer Mimics Via CD and AUC*

Peptide	$[\theta_{222\ nm}]$ (deg cm^2 dmol^{-1}) 25°C	$[\theta_{222\ nm}]$ (deg cm^2 dmol^{-1}) 37°C	$[\theta_{222\ nm}]$ (deg cm^2 dmol^{-1}) 50°C	M_{obs}/M_{calc} 4°C
eboIZN39IQ	−29,400	−27,900	−27,100	3.24
eboIZN39IQ(D3)	−30,400	−29,300	−28,400	3.22
eboIZN21	−25,500	−24,000	−22,900	3.54
eboIZN21(D2)	−24,800	−22,800	−21,800	3.15

CD scans were performed on the same samples of 11.4 μ*M* eboIZN39IQ, 11.1 μ*M* eboIZN39IQ(D3), 18.0 μ*M* eboIZN21 and 25.3 μ*M* eboIZN21(D2) in 50 m*M* sodium phosphate, pH 5.8, 150 m*M* NaCl at 25, 37, and 50°C. The peptides were allowed to equilibrate at each temperature for 10 min, after which no change in signal was seen over time. Sedimentation equilibrium analysis was performed on each peptide at three concentrations each (a starting concentration and two twofold dilutions, with typical starting concentrations between 10 and 30 μ*M*) and a minimum of two speeds, but typically three speeds (18,000, 21,000, and 24,000 rpm). Each data set was globally fit to a single ideal species. Each sedimentation equilibrium analysis was performed 2–4 times and averaged for the above table.

coiled-coil structure of eboIZN21 (below). As seen with eboIZN39IQ, eboIZN21 is very stable, showing similar CD spectra at 25, 37, and 50°C (Table I). We also attempted to produce mimics presenting the C-terminal portion of the N-trimer, but they were not soluble (data not shown) and were not studied further.

As negative controls for binding studies and drug discovery efforts, we produced mutant N-trimer mimics aimed at abolishing the C-peptide binding site. Specifically, alanines found along the C-peptide binding groove were mutated to aspartate, introducing binding-disruptive charges along the groove (Fig. 2). The resulting peptides are termed eboIZN39IQ(D3) and eboIZN21(D2). Using CD and AUC, we confirmed that these mutants maintained the stable coiled-coil structure and trimeric nature of their wild-type counterparts (Fig. 3 and Table I).

C-peptide binding characterization

To validate that eboIZN39IQ presents the native conformation of the N-trimer found in the prehairpin intermediate, we characterized binding to its native C-peptide ligand (Fig. 2), which binds along the entire groove of the N-trimer in the postfusion trimer-of-hairpins conformation. Surface plasmon resonance (SPR) analysis (ProteOn XPR36, Bio-Rad) of the interaction of the full-length C-peptide, eboC37, with eboIZN39IQ showed a dissociation constant of 14 n*M* (Fig. 4), with no binding to the D3 negative control. This tight binding affinity is of the same magnitude as the HIV-1 N-trimer/C-peptide interaction[29] and indicates that eboIZN39IQ presents a native N-trimer. A shortened C-peptide (eboC24), missing the 13 N-terminal residues of eboC37, bound to eboIZN39IQ with a dissociation constant of ~300 n*M* and did not bind to the D3 negative control (Supporting Information Fig. S2). eboIZN21 was less well behaved on an SPR surface, and we were unable to obtain reproducible data using this target.

Crystal structure of eboIZN21

To visualize how the N-trimer is presented in the absence of the native C-peptide ligand, as in our

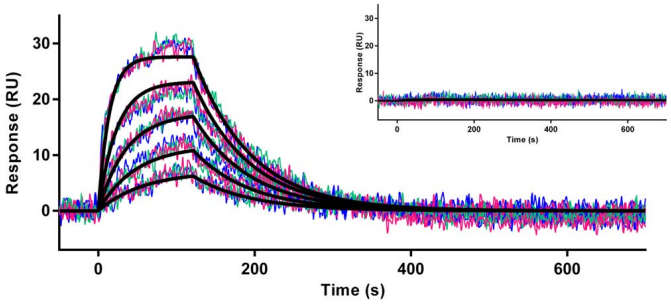


Figure 4. Binding of the ebolavirus C-peptide to the N-trimer mimic. Sensorgram of eboC37 flowed over eboIZN39IQ in a triplicate twofold dilution series starting at 60 n*M*, plotted with second-order second-neighbor-smoothing with a Savitzky–Golay filter (Prism 6, GraphPad Software). Each replicate dilution series is shown as a distinct color. The kinetic fit of the raw data is shown and yields $k_a = 9.6 \times 10^5\ M/s^1$, $k_d = 0.014/s^1$, and a K_D of 14 n*M*. Inset: The same eboC37 dilutions flowed over an eboIZN39IQ(D3) surface. No binding was observed.

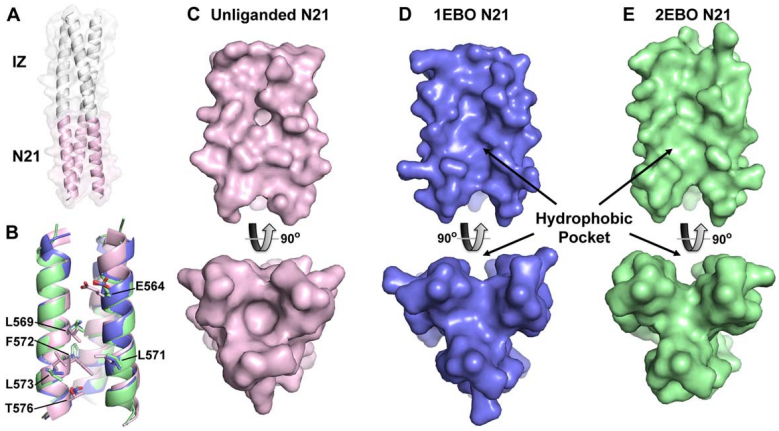


Figure 5. Crystal structure of eboIZN21. (A) Cartoon rendering with a semitransparent surface of the unliganded eboIZN21 structure. The IZ trimerization domain (white) and N21 region (pink) are indicated. The N21 region of available ebolavirus N-trimer structures is shown in isolation in panels B–E. (B) Overlay of the N21 region of the unliganded structure with the N21 region of the two previously solved ebolavirus GP2 core structures containing C-peptide (PDB IDs: 1EBO and 2EBO shown as blue and green, respectively. This color scheme is maintained in panels C–E). Residues that line the N21 groove and have significantly different rotamer conformations in the unliganded structure are shown as sticks and labeled. These residues occupy some of the equivalent space occupied by C-peptide (not shown) in the liganded structures resulting in a less prominent hydrophobic pocket when viewed (in subsequent panels) as a surface. (C) Surface representation of the unliganded N21 region. The bottom panel is the view of N21 from the bottom along its threefold axis and is rotated approximately 90° as compared to the top panel. (D, E) Similar views to (C) of the N21 region from the structures containing C-peptide. The prominent hydrophobic pocket in the 1EBO and 2EBO structures appears to be induced by ligand binding since the pocket is nearly absent in the unliganded structure.

mimics, we determined the X-ray crystal structure of eboIZN21 to 2.15 Å. eboIZN21 crystallized as a symmetrical trimer in space group P321 with one monomer in the asymmetric unit. The structure reveals that eboIZN21 is a continuous trimeric coiled coil, as designed [Fig. 5(A)]. Comparing our structure with the two previously reported structures of the ebolavirus 6-helix bundle (PDB IDs: 1EBO[28] and 2EBO[27]) revealed good overall agreement between the N21 residues of our unliganded structure and the C-peptide-bound structures, as indicated by root mean square deviations (rmsd) of 1.2 Å (across 63 atoms, 1EBO) and 1.4 Å (across 61 atoms, 2EBO) when trimers are aligned on Cα residues [Fig. 5(B)]. However, surface renderings show that a hydrophobic pocket in the N21 region of the 6-helix bundle structures, which accommodates residues 619–626 of the bound C-peptide, is collapsed in the isolated eboIZN21 structure [Fig. 5(C–E)].

The collapse of this pocket in the unliganded structure results from the side-chain conformations of several residues that fill the pocket. Specifically, in the absence of C-peptide, residues L569, L571, F572, L573, and T576 adopt alternate rotamers to pack together via hydrophobic interactions and thus alter the surface contours of the ligand binding pocket (Supporting Information Fig. S3). The side chain of E564 also adopts an alternate conformation to occupy a distinct portion of the pocket (toward the top of the pocket in Fig. 5). Therefore, as seen with the analogous hydrophobic ligand-binding pocket in the HIV gp41 N-trimer [comparing structures in Refs. (29–32), e.g., Supporting Information Fig. S4], our eboIZN21 unliganded structure indicates that the ebolavirus GP N21 pocket is induced by ligand binding and can likely adopt various conformations depending on the specific ligand.

We used the MONSTER protein interaction server[33] to calculate the solvent accessible surface area (SASA) buried at the interface of the ebolavirus and HIV hydrophobic pockets with their C-peptides. The crystal structures of the ebolavirus[27,28] and HIV[30,31] 6-helix bundles reveal that in each case the pocket interacts with 8 C-peptide residues (ITDKIDQI for ebolavirus and WMEWDREI for HIV) [Supporting Information Fig. S4(A,C)]. The buried SASA at the N21 pocket/8-mer C-peptide interface is similar in the 1EBO and 2EBO structures at 393/348 and 387/325 Å^2, respectively. These values are comparable to SASA buried at the HIV gp41 pocket/8-mer

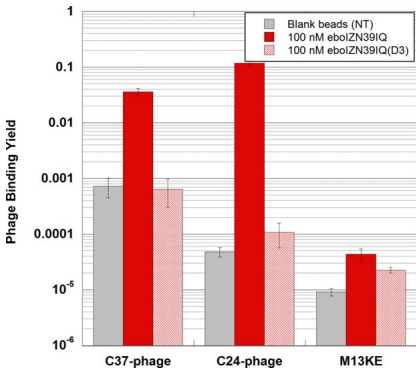


Figure 6. Validation of eboIZN39IQ as a phage display target. Clonal phage expressing ebolavirus C-peptides (eboC37 or eboC24) were incubated with biotinylated eboIZN39IQ in solution followed by capture via magnetic streptavidin beads. Negative target controls include the binding site mutant, eboIZN39IQ(D3), and magnetic beads with no target (NT). Binding of M13KE (phage with no peptide clone) to all targets was also assayed. The fraction of phage bound is reported. Error bars represent standard error across triplicate experiments.

C-peptide interface (349/310 Å^2).[30] Finally, a similar analysis between the HIV pocket and our D-peptide entry inhibitor, PIE12 [Supporting Information Fig. S4(B)], reveals that 416/391 A^2 of SASA is buried at that interface.[32] Given the comparable size of the pocket/C-peptide interface in the ebolavirus and HIV 6-helix bundle structures, combined with the high anti-HIV potency of PIE12-trimer (a trimerized version of PIE12, designed to bind all three pockets of the prehairpin intermediate),[32] it is reasonable to expect that potent pocket-specific peptide inhibitors of ebolavirus can also be discovered.

Phage display target validation

Phage display is a powerful screening technology that is used to screen billions of peptides or antibodies against a target of interest to identify specific inhibitors of protein/protein interactions. Indeed, HIV N-trimer mimics were successfully used in phage display screens of both scFv antibodies and D-peptides (see "Mirror-Image Phage Display" below) to identify potent, broadly neutralizing HIV entry inhibitors.[25,29,32,34,35] To validate the ebolavirus N-trimer mimics as discovery targets in the context of phage display, we produced phage clones expressing the native binding partners eboC37 and eboC24 [Fig. 2(A)] and assayed their ability to bind to our N-trimer mimics in phage clone binding assays. These experiments were designed to verify ligand binding and to define the best conditions for future phage display discovery efforts.

Phage display selections can be conducted in two formats: solid- and solution-phase. In solid-phase selections, the target is bound to a solid support (here, biotinylated ebolavirus N-trimer mimic is attached to streptavidin-coated magnetic beads), and then the phage are incubated with the immobilized target. Since common phage display libraries are multivalent (multiple copies of the library molecule are expressed on the surface of the phage, due to fusion to multicopy coat proteins), avidity effects improve the apparent binding constant of the library clones. This avidity-induced affinity boost is beneficial when screening naïve phage libraries, where initial binders typically have low target affinities. In solution phase, where both target and phage are incubated in solution, avidity effects are reduced. Following incubation, the bound complexes are captured through a brief interaction with a solid support (again, in this case, through a biotinylated target and streptavidin beads). At equivalent target concentrations, solution-phase selection is more stringent than solid-phase selection. The higher stringency of solution-phase is useful when screening second-generation libraries for affinity maturation (e.g., peptide binding consensus libraries or antibody variable loop mutagenesis libraries), where tight binders must be distinguished from a background of moderate binders.

Both eboC37 and eboC24 clonal phage bound to eboIZN39IQ target significantly over background (both empty beads and negative control eboIZN39IQ(D3) beads) using solution-phase clonal phage binding assays carried out at pH 5.8 to mimic the endosomal environment (Fig. 6). Also, binding of M13KE empty phage to both eboIZN39IQ and eboIZN39IQ(D3) was minimal. These data validate eboIZN39IQ as a phage display target. In addition, these data demonstrate that eboIZN39IQ(D3) serves as an effective negative control, as its clonal C-peptide phage binding is comparable to that of blank beads. In this format, both C-peptide clones bound at similar levels to eboIZN39IQ, although eboC37 had greater background binding to both negative controls.

Specific binding (over two orders of magnitude over background) was seen when using eboIZN21 as a target in a solid-phase eboC24 clonal phage binding assay, also validating eboIZN21 as a phage display target (Fig. 7). With the low level of eboC24 phage binding to eboIZN21(D2) (similar to eboC24 binding to blank beads), the binding site mutant is also verified as a negative control. In addition, only a very low level of M13KE empty phage binding to eboIZN21 and eboIZN21(D2) was observed.

Low phage background binding to targets is required in order to discern specific binding during

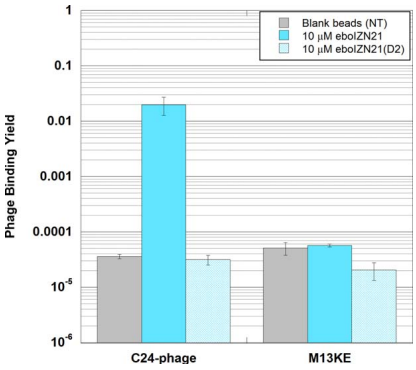


Figure 7. Validation of eboIZN21 as a phage display target. Clonal phage expressing an ebolavirus C-peptide (eboC24) were incubated with biotinylated eboIZN21 bound to streptavidin magnetic beads (solid-phase conditions). Negative target controls include the binding site mutant, eboIZN21(D2), and magnetic beads with no target (NT). Binding of M13KE (phage with no peptide clone) to all targets was also assayed. The fraction of phage bound is reported. Error bars represent standard error for triplicate experiments.

phage panning rounds. To evaluate this property for our two N-trimer mimics, we analyzed empty M13KE phage binding to both targets under varying conditions (Fig. 8). In both solid- and solution-phase formats, M13KE phage showed significantly higher binding to eboIZN39IQ beads than to blank beads. For the eboIZN21 target, phage background binding was drastically reduced in comparison to eboIZN39IQ, and binding of M13KE phage was similar to both target and blank beads in solid and solution phase. Under stringent conditions (solution phase, 100 n*M* target) the M13KE background binding to eboIZN39IQ was minimized, and a large affinity difference for eboC24 binding to eboIZN39IQ versus eboIZN21 could be seen. This affinity difference is likely to be biologically relevant because the trimer-of-hairpins structures[27,28] show that the binding site of eboC24 extends past the C-terminus of N21. These data indicate the importance of choosing proper stringency conditions in performing phage display selections.

The first step of a phage display discovery process is to screen a naïve phage display library for binding to the desired target. In such a first selection, where the library diversity only partially samples the large potential sequence space (e.g., a naïve peptide 12-mer library has 20^{12} ($>10^{15}$) possible sequences, whereas the typical diversity of a phage display library is $<10^{10}$), the best binders identified are usually modest, with low- to mid-micromolar affinities. Therefore, the selection pressure applied during phage panning must also be modest. Standard naïve phage display starting conditions are 10 μ*M* target presented on solid-phase (i.e., 30 μL of 10 μ*M* target immobilized onto magnetic beads).[25] As illustrated in Figure 8, M13KE binding to eboIZN39IQ is nearly saturated at this condition, and therefore it would not be possible to identify binding over background. 10% phage binding is considered saturating, as binding yields of even strong binders do not generally exceed this level (likely due to proteolysis of displayed peptides). Under the same conditions, the eboIZN21 background binding is >600-fold lower and similar to blank bead binding, ideal starting conditions for naïve phage display. Therefore, eboIZN21 is an optimized target for phage display discovery efforts. Additionally, the eboC24-phage can serve as an important positive control during naïve phage display to validate the conditions used to capture weak, but specific binders. Notably, in addition to having ideal behavior in phage display, the N21 region is also identical across all ebolavirus species and highly conserved among filoviruses (95% conserved) [Fig. 2(A)]. eboIZN39IQ is an ideal target for higher stringency solution-phase phage display and could be used to screen secondary libraries for affinity optimization of ligands identified from the naïve library. This could be especially useful for extending the binding interface of the ligands further along the N-trimer groove.

Mirror-image phage display

Mirror-image phage display is an innovative adaptation of standard peptide phage display and is used to identify D-peptides that bind to a target of interest[25,36] (Supporting Information Fig. S5). D-peptides are composed of D-amino acids and are the mirror-image of naturally occurring L-peptides. D-peptides have several important potential advantages as drug candidates (as reviewed[37]). As peptides, they are capable of blocking large protein/protein interactions, which is generally not possible for small molecules. In addition, because they are resistant to protease degradation,[38] D-peptides should possess a longer *in vivo* half life and reduced immunogenicity.[39] In our HIV-1 entry inhibitor discovery program,[18,25,29,32] we used mirror-image phage display and protein design to develop the highly potent and broadly acting D-peptide entry inhibitor, chol-PIE12-trimer, which is now in advanced preclinical studies. Their resistance to endosomal proteases makes D-peptides especially attractive as ebolavirus entry inhibitors. Our future studies will therefore use the N-trimer mimics to discover D-peptide inhibitors of ebolavirus entry by mirror-image phage display.

In traditional peptide phage display, a library of phage, each with a unique peptide displayed on its surface, is screened against a target.[40] In mirror-image phage display, the target is chemically

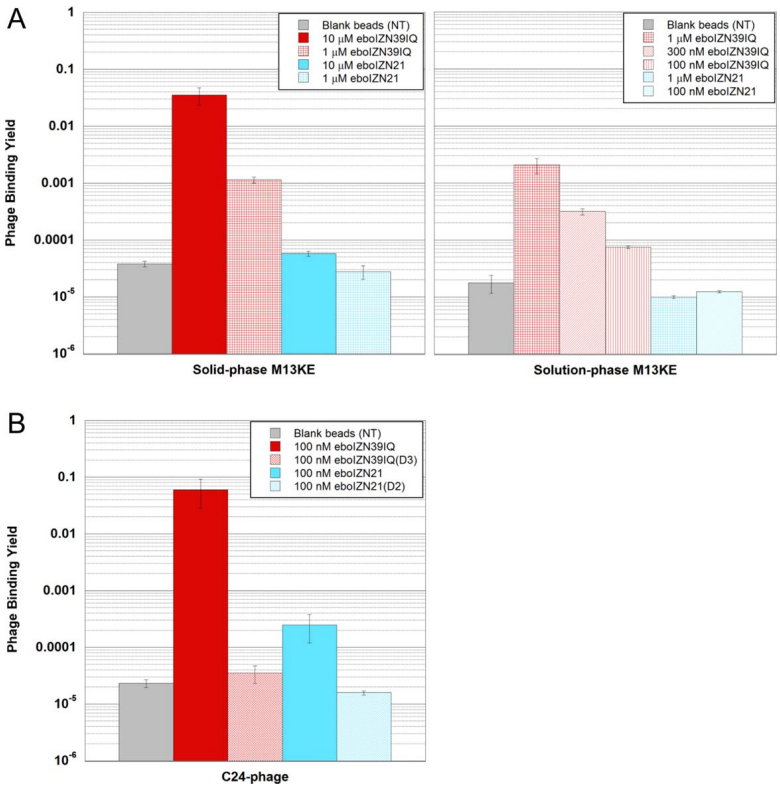


Figure 8. Comparing the two ebolavirus N-trimer mimics as phage display targets. (A) Phage background binding is greater to eboIZN39IQ than to eboIZN21. Phage binding assay showing M13KE control phage binding to biotinylated eboIZN39IQ and eboIZN21 under both solid-phase (left) and solution-phase (right) conditions. Magnetic beads with no target (NT) were used as a negative control. The fraction of phage bound is reported. Error bars represent the range for duplicate experiments (solid phase) and standard error for four or more replicates (solution phase). (B) High stringency solution-phase binding shows an affinity difference for the specific binding of eboC24 to the two N-trimer mimics. Clonal phage expressing eboC24 were incubated with biotinylated N-trimer in solution followed by capture via magnetic streptavidin beads. Negative target controls include the binding site mutants and magnetic beads with no target (NT). For NT, error bars represent standard error across triplicate experiments. The remaining error bars represent the range for duplicate experiments.

synthesized from D-amino acids and therefore forms the mirror-image structure of the natural L-target (Supporting Information Fig. S5). Phage display using the D-target is performed, and identified L-peptides that bind the D-target are then chemically synthesized using D-amino acids. By the law of symmetry, these D-peptides bind the natural L-target. Unlike with traditional phage display, mirror-image phage display targets are limited in size to those that can be chemically synthesized (although this size limit is continually expanding with modern chemical protein synthesis advances[41].

In order to prepare ebolavirus N-trimers as mirror-image phage display targets, we synthesized them as D-peptides. At 48 amino acids each, D-eboIZN21 and D-eboIZN21(D2) were synthesized through standard solid-phase peptide synthesis (SPPS) techniques. Importantly, even though the 101-residue length of eboIZN39IQ is beyond the scope of standard SPPS, modern chemoselective ligation techniques[42] allow for its assembly from multiple peptide segments. D-eboIZN39IQ, was assembled using native chemical ligation[43] and metal-free desulfurization,[44] in which cysteine residues are

introduced at native alanine sites to facilitate ligation and then converted back to alanine through desulfurization [Supporting Information Fig. S6(A)]. D-eboIZN39IQ was assembled from three synthetic segments of 27–41 residues. The production of D-eboIZN39IQ(D3) required a different assembly strategy, as one of the alanines used in D-eboIZN39IQ as a ligation junction is mutated to aspartate. Therefore, D-eboIZN39IQ(D3) was assembled from two synthetic segments of 33 and 68 amino acids. The final peptide products were confirmed by LC/MS [e.g., Supporting Information Fig. S6(B,C)].

The law of symmetry dictates that D-peptides will adopt the mirror-image structure of their L-counterparts. CD analysis of D-eboIZN21 confirms it possesses mirror-image helical structure compared to its L-peptide counterpart [Supporting Information Fig. S7(A)]. SPR analysis of D-eboC37 binding to D-eboIZN39IQ shows a similar binding affinity (6 n*M*) to the L-peptide interaction and validates the functionality of D-eboIZN39IQ [Supporting Information Fig. S7(B)]. Preliminary phage display experiments with these D-targets demonstrate the same M13KE binding properties as the L-versions (data not shown), verifying the strategy of screening naïve libraries with the D-eboIZN21 target and employing D-eboIZN39IQ for subsequent affinity optimization efforts when higher stringency is appropriate.

Vulnerability of the Ebolavirus prehairpin intermediate to a high potency inhibitor

A prerequisite for the success of drug discovery efforts targeting the ebolavirus N-trimer mimics is the exposure of a vulnerable prehairpin intermediate during viral entry. Exogenous C-peptides derived from the transmembrane subunit of the envelope glycoprotein have been used to validate this vulnerable prehairpin intermediate in a variety of viruses (e.g., HIV, SARS, and many paramyxoviruses[45–47]). For ebolavirus, an early report showed C-peptide inhibition activity at mM concentrations,[48] and more recent reports describe improved inhibitory activity (mid μM) of C-peptides with an endosomal localization tag.[49,50] Our ebolavirus N-trimer, eboIZN39IQ, provides an additional tool with which to explore the vulnerability of the prehairpin intermediate. In support of this strategy, peptide mimics of the HIV-1 N-trimer inhibit HIV entry at mid nM concentrations by binding to the C-peptide region of the exposed intermediate.[24]

Indeed, eboIZN39IQ inhibited entry in our pseudovirus system in which ebolavirus GP (representative species, *Zaire ebolavirus*) is expressed on the surface of an HIV particle [Fig. 9(A)], with an average IC_{50} of 320 n*M*. Importantly, the anti-ebolavirus activity of our negative control, eboIZN39IQ(D3), is ~30-fold diminished, with an IC_{50} of 11 μ*M*. It is difficult to determine the exact nature of the modest

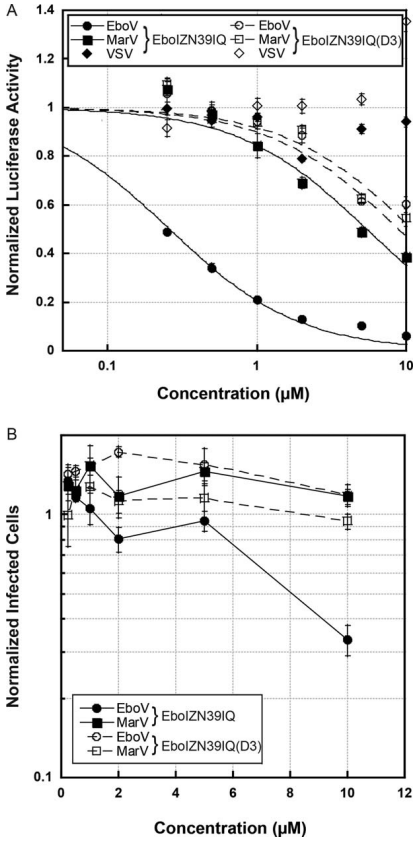


Figure 9. Inhibition of filovirus entry by eboIZN39IQ. (A) A representative pseudovirion assay looking at the inhibitory activity of eboIZN39IQ and the negative control, eboIZ-N39IQ(D3) against ebolavirus (EboV), marburgvirus (MarV), and VSV retroviral pseudotypes. Each point represents the average of quadruplicate measurements normalized to uninhibited control. Error bars represent normalized standard errors. For this particular assay, eboIZN39IQ IC_{50}s are 260 n*M* against ebolavirus and 5.4 μ*M* against marburgvirus. The eboIZN39IQ(D3) IC_{50}s are 8.9 μ*M* against ebolavirus and 11 μ*M* against marburgvirus. (B) Data for the authentic filovirus immunofluorescence inhibition assay. Each point represents the average of quadruplicate measurements normalized to vehicle control. Strong inhibition of ebolavirus is seen at 10 μ*M* eboIZN39IQ, with an average 33% (±4%) of infected cells compared to vehicle control.

eboIZN39IQ(D3) activity, as it is not seen against a vesicular stomatitis virus glycoprotein pseudotype (VSV), and no morphological changes (indicative of

toxicity) were observed. It is possible the modest eboIZN39IQ(D3) activity could be due to residual prehairpin intermediate binding activity. eboIZN39IQ demonstrated modest activity against marburgvirus pseudovirions (another member of the filovirus family), at an IC_{50} of 5.7 μM, although this was only ~2-fold better than the eboIZN39IQ(D3) anti-marburgvirus activity.

The ability of eboIZN39IQ to inhibit the entry of wild-type ebolavirus and marburgvirus was also assessed using a filovirus immunofluorescence assay under BSL4 conditions [Fig. 9(B)]. Although eboIZN39IQ was significantly less potent in this assay, there is 67% inhibition of entry at the highest concentration tested (10 μM) and no inhibition by our negative D3 control. Also, no activity was seen against marburgvirus. Potency differentials between pseudovirus systems and authentic filoviruses have been seen for other fusion inhibitors [for example in Ref. (51)]. Taken together, these data validate the presence of a vulnerable prehairpin intermediate during the ebolavirus entry process.

Unlike HIV-1, ebolavirus enters cells via endocytosis and initiates membrane fusion late in the endosomal pathway. Therefore, ebolavirus entry inhibitors will have to enter into and be active in endosomes. Although eboIZN39IQ does not possess a specific tag to localize it to endosomes, it is highly charged on its surface (with both positive and negatively charged side chains), and, interestingly, the inhibitory activity we observed in both the pseudovirus and authentic ebolavirus systems was dependent on the presence of the standard viral assay additive DEAE-dextran. It seems likely that the highly charged N-trimer mimic associates with the anionic cell membrane, especially in the presence of the cationic DEAE-dextran that would reduce electrostatic repulsion between the negative charges of eboIZN39IQ and the membrane, allowing it to access the endosome more efficiently than C-peptides.[50] As a structured peptide, eboIZN39IQ would also likely resist proteolysis longer than unstructured C-peptides.

Conclusion

In summary, we have designed and characterized two mimics of the highly conserved ebolavirus GP N-trimer region as it appears in the prehairpin intermediate during viral entry. In addition, through our clonal phage display experiments, we have functionally validated eboIZN39IQ and eboIZN21 as drug discovery targets, especially for phage display screens. Finally, with the characterization of the inhibitory activity of eboIZN39IQ, we have further validated the vulnerability of the ebolavirus prehairpin intermediate by demonstrating potent inhibition. These N-trimer mimics should be valuable for the discovery of small molecules, antibodies and/or peptides that inhibit ebolavirus entry. Specifically, our group is interested in the discovery of D-peptide inhibitors of ebolavirus using mirror-image phage display, and with the two targets and their binding site mutants synthesized in the D configuration, we are now poised for those selections.

It is noteworthy that in addition to the remarkable conservation of the N-trimer region across all ebolavirus species, it is also highly conserved across the filovirus family [see Fig. 2(A)]. Therefore, discovery efforts will likely identify inhibitors with broad filovirus activity. If such activity is suboptimal, it should be possible to design analogous mimics of the marburgvirus N-trimer region and use them in concert with the ebolavirus targets. For example in phage display, rounds of panning could alternate between the ebolavirus and marburgvirus targets, selecting specifically for an inhibitor of both viruses. Although the vast majority of natural filovirus outbreaks have been caused by ebolavirus, marburgvirus still poses a risk both as a natural pathogen (with three outbreaks in the last 10 years) and as a bioterror agent,[52] making the discovery of a broad-spectrum inhibitor desirable.

In addition to serving as drug targets, the ebolavirus N-trimer mimics should be useful as cell biological tools. For example, fluorescently labeled N-trimers could be used in cell culture experiments to track the appearance of the prehairpin intermediate during the viral entry event. Such studies would advance insight into filovirus entry dynamics.

Materials and Methods

Reagents

Plasmids and cells were obtained from the indicated sources: pEBB-HXB2 (gift from B. Chen)[53], SV-ZeboGPΔmuc and SVMarVGP (gift from M. Farzan),[54] BLR(DE3)pLysS *E. coli* (EMD Millipore, Billerica, MA), BL21-Gold(DE3)pLysS *E. coli* and XL-1 Blue *E. coli* (Agilent Technologies, Santa Clara, CA). pNL4–3.Luc.R-E- (N. Landau)[55,56] and HOS-CD4-fusin (N. Landau)[57,58] were obtained from the NIH AIDS Research and Reference Program. Mammalian cells were propagated in standard tissue culture medium, Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal calf serum and L-glutamate (Life Technologies, Grand Island, NY).

Recombinant peptide production and purification

The DNA encoding eboIZN39IQ and eboIZN39IQ(D3) was produced via PCR gene synthesis. The IZ_m and IQ fragments were PCR amplified from plasmids encoding HIV-1 N-trimer mimics [e.g. in Ref. (59)]. An *Nde*I site was included in the 5′ PCR primer for IZ_m, and a *Bam*HI site was included in

the 3′ PCR primer for IQ. The ebolavirus N39 sequence from the species *Zaire ebolavirus*[60] was synthesized in two overlapping oligos with optimized codons and companion primers. All internal primers contained complementary sequences so the three separate components, IZ_m, N39, and IQ could be annealed and amplified together. The resulting DNA fragment was cloned into the *Nde*I/*Bam*HI cloning sites of pKA8, validated by sequencing and expressed in BLR(DE3)pLysS cells using an autoinduction protocol. Specifically, cultures were inoculated from a single colony and grown overnight at 37°C in autoinduction media.[61] The resultant peptide has an N-terminal His tag (His_8) followed by a TEV cleavage site (ENLYFQG). A single tyrosine was placed at the end of the sequence to facilitate concentration determination via absorbance at 280 nm. The peptides were resuspended from inclusion bodies using Ni-binding buffer (20 n*M* sodium phosphate pH 8.0, 300 m*M* NaCl, 10 m*M* imidazole) + 6*M* GuHCl, and purified via gravity flow Ni affinity chromatography (HIS-Select Nickel Affinity Gel, Sigma Aldrich, St. Louis, MO). The purified peptides were dialyzed into 5% acetic acid and further purified by reverse phase HPLC on a C18 column (Vydac, Grace, Columbia, MD) and lyophilized. Peptide powder was resuspended in water and diluted to 0.2 mg/mL in 50 m*M* sodium phosphate pH 6.5, 0.5 m*M* EDTA, 1 m*M* DTT and digested with a solubility-enhanced tobacco etch virus NIa protease (TEVse, based on published modifications[62,63]) overnight at 30°C. The digested peptide was dialyzed into 5% acetic acid and then HPLC purified and lyophilized. The final peptide sequences are: GHMD**IKKEIEAIKKEQEAIKKKIEAIEKE***LRQLAN ETTQ(A/D)LQLFLR(A/D)TTELRTFSILNRK(A/D)ID FLLQR***MKQIEDKIEEIESKQKKIENEIARIKKLIG ERY**, with IZ_m and IQ shown in bold, the ebolavirus N-trimer in italics, and the three alanine positions that are changed to aspartate in the D3 mutant in parentheses.

Biotinylated eboIZN39IQ and eboIZN39IQ(D3) for SPR analysis and phage display were expressed from plasmids that are modified from those described above. Using PCR, a CGG sequence was added N-terminal to IZ (GHMCGGDIKK. . .). Expression and purification were as described above with additional reduction steps included to keep the cysteine reduced during purification (100 m*M* DTT treatment after Ni++ affinity chromatography and 50 m*M* TCEP treatment after TEV digestion). The purified protein was biotinylated with EZ-link Maleimide–PEG2–biotin (Thermo Scientific, Waltham, MA). The purified lyophilized powder was resuspended at 1 m*M* in freshly prepared reaction buffer (6 *M* GuHCl, 150 m*M* NaCl, 100 m*M* Na_2HPO_4, 5 m*M* TCEP) and the biotinylation reagent was added at 5 m*M* and allowed to react for 4 h at RT. The biotinylated peptides were purified by reverse phase HPLC on a C18 column (Waters) and lyophilized. The mass of the peptide was confirmed by LC/MS (AB Sciex API 3000 LC/MS/MS system, Framingham, MA).

Peptide synthesis

eboIZN21, eboIZN21(D2), eboC37 and eboC24 were chemically synthesized using solid-phase peptide synthesis (SPPS) with Fmoc-amino acids (AAPPTec, Louisville, KY, and CBL Biopharma, Boulder, CO) on a Prelude peptide synthesizer [Protein Technologies (PTI), Tucson, AZ]. A single tyrosine was placed at the N-terminus of both eboIZN21 and eboIZN21(D2) to facilitate concentration determination via absorbance at 280 nm. The peptides were synthesized on TentaGel R RAM resin (Rapp Polymere, Germany) to yield C-terminal amide peptides. Standard synthesis scales were 25–32 μmol per peptide. Standard amino acid coupling was as follows: 3 × 3 min deprotection with 20% piperidine in DMF followed by 25 min couplings with 72.2 m*M* amino acid (200 m*M* stocks in NMP), 71.5 m*M* HATU (198 m*M* stock in DMF), and 166.7 m*M* NMM (600 m*M* stock in DMF). Biotinylation was achieved with N-Biotinyl-NH-$(PEG)_2$-COOH DIPEA (Novabiochem, EMD Millipore) coupling for 2 h. N-terminal capping was accomplished in 30 min with 2 mL acetic anhydride and 2 mL 0.6*M* NMM. Peptide cleavage from resin was accomplished offline with 92.5% TFA, 2.5% EDT, 2.5% TIS, 2.5% H_2O when the peptide contained Met or Cys residue(s) or with 95% TFA, 2.5% TIS, 2.5% H_2O in the absence of any Met/Cys residues followed by precipitation/washing with diethylether. All peptides were purified by reverse-phase HPLC on a Waters (Milford, MA) BEH X-Bridge C18 column (10 μm, 300 Å, 19 × 250 mm) with a water/ACN gradient in 0.1% TFA. All peptides were lyophilized and their molecular weight verified by LC/MS.

D-eboIZN39IQ was assembled from three synthetic peptide segments via native chemical ligation/metal-free desulfurization (Supporting Information Fig. S6). D-peptides were synthesized via Fmoc-SPPS on a PTI PS3 peptide synthesizer at 100 μmol scale. The C-terminal peptide was synthesized on Rink Amide AM resin LL (Novabiochem) and the other two segments were synthesized on Dawson Dbz AM resin (Novabiochem). The C-terminal segment contained an N-terminal cysteine residue in the place of a native alanine for use in native chemical ligation (CIDFLLQRMKQIEDKIEEIESKQKKIENEIARIKKLIGERY). For the same reason, the middle segment contained an N-terminal Boc-L-thiazolidine-4-carboxylic acid (Boc-Thz-OH, Bachem, Torrance, CA) as its N-terminal residue in place of the native alanine at that position((Thz)-NETTQALQLFLRATTELRTFSILNRK). The N-terminus of

the N-terminal peptide (GHMDIKKEIEAIKKEQEAIKKKIEAIEKELRQL) was biotinylated with N-Biotinyl-NH-(PEG)2-COOH DIPEA (Novabiochem). For peptides synthesized on Dawson Dbz AM resin, the C-terminal linker was converted to the resin bound *N*-acyl-benzimidazolinone (Nbz) according to manufacturer instructions. Cleavage of all peptides was performed according to standard procedures. Peptides were purified by reverse-phase HPLC on a Waters BEH X-Bridge C18 column (10 μm, 300 Å, 19 × 250 mm) with a water/acetonitrile gradient in 0.1% TFA. Ligations were performed according to Ref. (64) with peptide concentrations ~2 m*M*. Following ligation between the C-terminal and middle segments, the N-terminal Thz was converted to cysteine by dissolving the purified ligation product in 6*M* GuHCl, 200 m*M* sodium phosphate, 200 m*M* methoxyamine HCl, pH 4. After Thz to Cys conversion was achieved, the buffer was brought to 200 m*M* MPAA and 20 m*M* TCEP, the pH was adjusted to 7, and the N-terminal peptide was added to the solution for the final ligation. Following purification of the ligation product by reverse-phase HPLC, the cysteine residues at the ligation junctions were converted to the native alanine residues via a metal-free, radical-mediated desulfurization strategy essentially as described in Ref. (44) except that *t*-butylthiol was replaced with glutathione, and desulfurizations were performed at 37°C. eboIZN39IQ(D3) was synthesized in an analogous (though simplified) manner using two peptide segments.

Preparation of peptide samples for biophysical analysis

For biophysical analyses, peptide stocks were prepared in water from lyophilized peptide at concentrations of 400 μ*M* or greater for a minimum absorbance at 280 nm of 0.1 in a 1 cm pathlength cuvette. Stocks were centrifuged at 18,000*g* for 10 min to remove aggregates. Absorbance at 280 nm (using ϵ_{280} of 1408 M/cm[1] for tyrosine) was used to determine stock concentrations.[65] For eboIZN39IQ and eboIZN39IQ(D3), both recombinant and synthetic, UV absorbance consistently overestimated the concentration of the stocks (as evidenced by an unusually high 260/280 ratio as well as CD traces whose shape depicted ideal coiled coils but whose signal had a lower than expected absolute value). This overestimation is likely due to peptide bond absorbance contributing to the signal at 280 nm (101 amino acids with only one near-UV absorbing residue, a tyrosine). Therefore, the concentrations of these stocks were determined via quantitative amino acid analysis, which was performed using a Hitachi L-8800 Amino Acid Analyzer (Hitachi High-Technologies Corporation, Tokyo, Japan). Peptides were hydrolyzed in 5.7*N* HCl overnight at 105°C in sealed ampoules, and then analyzed via ion-exchange chromatography and postcolumn derivatization with ninhydrin using the Hitachi Analyzer. The peptides were then diluted to the desired concentration in 50 m*M* sodium phosphate pH 5.8, 150 m*M* NaCl. For eboIZN21 and eboIZN21(D2), all experiments described in this paper were performed with biotinylated peptide. For eboIZN39IQ and eboIZN39IQ(D3), CD, AUC and viral infectivity were performed with nonbiotinylated material, whereas SPR and phage display used biotinylated material.

Circular dichrosim

Circular dichroism (CD) data were obtained using an AVIV Model 410 spectrophotometer (AVIV, Lakewood, NJ). Samples were analyzed in a 1 mm pathlength quartz cuvette at 25, 37, and 50°C. Prior to CD analysis, prepared samples (in 50 m*M* sodium phosphate pH 5.8, 150 m*M* NaCl) were centrifuged at 18,000*g* for 10 min to remove aggregates. CD data were scanned in triplicate and buffer subtracted. Final CD data were presented according to mean residue ellipticity equation $[\theta] = 100 \times \theta/[(n - 1) \times (\ell) \times (c)]$, where θ is the observed ellipticity, $n - 1$ is the number of peptide bonds, ℓ is the pathlength in cm, and c is the peptide concentration in mM. Percent helicity was calculated from the ellipticity at 222 nm according to the Lifson–Roig-based helix-coil model which defines the dependence on chain length and temperature as described.[66] Due to aggregation observed with eboIZN39IQ and eboIZN39IQ(D3) upon initial dilution in the CD buffer, their final concentrations were corrected from the original amino acid analysis values based on the ratio of ellipticity at 222 nm post- and precentrifugation ($\theta_{222\text{-postspun}}/\theta_{222\text{-prespun}}$).

Analytical ultracentrifugation sedimentation equilibrium

Using an Optima XL-A Analytical Ultracentrifuge (Beckman Coulter, Brea, CA), sedimentation equilibrium analysis was performed on each peptide at three concentrations (a starting concentration and two 2-fold dilutions, with typical starting concentrations between 10 and 30 μ*M*). Dilutions were prepared in matching buffer (50 m*M* sodium phosphate, 150 m*M* NaCl, pH 5.8), and the same buffer was used for blanks. Each sample was spun until equilibrium, typically ~24 h, at a minimum of two speeds, but typically three speeds (18,000, 21,000 and 24,000 rpm). Each data set was globally fit to a floating molecular weight single ideal species with a nonlinear least squares algorithm as implemented in HETEROANALYSIS.[67] Fits are reported as the observed (i.e., fit) molecular weight divided by the calculated molecular weight of a monomer (M_{obs}/M_{calc}). Buffer densities and protein partial specific volumes were calculated with SEDNTERP (version

1.09).[68] For the biotinylated peptides, partial specific volumes were adjusted based on reported values for PEG.[69]

Surface plasmon resonance

SPR analysis was conducted on the Bio-Rad (Hercules, CA) ProteOn XPR36 instrument in PBS* running buffer (50 m*M* sodium phosphate, 150 m*M* NaCl, pH 5.8) + 0.1 mg/mL BSA and 0.01% Tween 20. Approximately 600 RUs of biotin-eboIZN39IQ (in the presence of eboC37) and biotin-eboIZN39IQ(D3) targets (200 n*M* stocks ultracentrifuged for 30 min at 45,000 rpm) were loaded at 67 n*M* on to the NLC neutravidin-coated chip (Bio-Rad), followed by blocking with 450 μ*M* biotin. Using the one-shot kinetics method, a twofold dilution series was performed in triplicate at RT starting at 60 n*M* for eboC37 and a threefold dilution series in triplicate at RT starting at 800 n*M* for eboC24. 10 min dissociation time for eboC37 and 5 min dissociation time for eboC24 were used to ensure the response fully recovered to baseline prior to the next injection. Data were corrected by subtracting blank surface and blank buffer reference injections, and the kinetics (for eboC37) and equilibrium (for eboC24) were globally fit to the Langmuir model for 1:1 binding[70] using ProteOn Manager software (Bio-Rad).

Crystallization

eboIZN21 was dissolved in ddH20 to a concentration of ~10 mg/mL and centrifuged at 18,000*g* for 10 min. Sitting-drop vapor-diffusion crystal trials were set up using a Phoenix crystallization robot (Art Robbins Instruments, Sunnyvale, CA). Crystals grew at 4°C in drops containing a 2:1 ratio of peptide to well solution, which consisted of 30% (v/v) 1,2-propanediol, 100 m*M* HEPES pH 7.5, 20% (v/v) PEG-400. The crystals were flash frozen in liquid nitrogen without the need for additional cryoprotection and determined to be in space group P321 with unit cell dimensions $a = b = 38.51$ Å, $c = 72.59$ Å.

Data collection, structure determination, and refinement

A native dataset was collected at beam line 7-1 of the Stanford Synchrotron Radiation Lightsource. Data were integrated and scaled to 2.15 Å resolution using HKL2000.[71] In order to rule out the possibility of twinning, data were initially scaled in space group P3 and analyzed with the program Xtriage,[71] which indicated that the data are untwinned and that the correct space group is P321. A model that consisted of a canonical helix appropriate to the size and sequence of the IZ domain and the N21 region of ebolavirus GP (PDB ID: 1EBO) was used for molecular replacement using the program Phaser.[72] A single eboIZN21 monomer was found in the asymmetric unit with the trimeric structure generated by the crystallographic threefold. Subsequent model building, structure refinement, and validation were performed with Coot,[73] PHENIX Refine,[74] and MolProbity[75] software, respectively. The final model was refined to crystallographic R/R_{free} values of 0.272/0.294 with good geometry (Supporting Information Table S1). Additional refinements were carried out in space group P3 allowing all possible twin laws, to further verify (by monitoring Rfree) that the correct space group is P321 with a monomer in the asymmetric unit. A composite omit map agreed well with the final model, indicating good sidechain density throughout. The atomic coordinates and structure factors have been deposited in the Protein Data Bank, http://www.pdb.org (PDB ID: 4R0R).

Clonal phage production

Forward and reverse sandwich oligonucleotides encoding the C-peptide clones were designed based on the primary sequence of each clone. The forward and reverse eboC37 oligonucleotides were: ATGCGG TACCTTTCTATTCTCATTCTTGGGGCGGCACCTGC CATATTCTGGGCCCGGATTGCGCGATTGAACCGC ATGATTGGACCAAAA and CCTTTTCGGCCGAACC CCCACCTTTATCCACAAAATCATGAATAATCTGAT CAATTTTATCGGTAATGTTTTTGGTCCAATCATGC GGTT. The forward and reverse eboC24 oligonucleotides were: ATGCGGTACCTTTCTATTCTCATTCT ATTGAACCGCATGATTGGACCAAAAACATTACCG and CCTTTTCGGCCGAACCCCCACCTTTATCCAC AAAATCATGAATAATCTGATCAATTTTATCGGTAA TGTTTTTGGTCCAATCATGCGGTT. The oligonucleotide sandwich was annealed with 5 μg of each primer in 50 μL total volume in ddH_2O by heating to 95°C and slow cooling and then extended with Klenow Fragment [New England Biolabs (NEB), Ipswich, MA] according to the manufacturer's protocol. The inserts and M13KE cloning vector backbone (NEB) were digested with Acc65I and EagI-HF. The insert DNA was EtOH precipitated, gel purified from a 6% TBE acrylamide gel, extracted from the gel by incubating gel slices in a minimal volume of extraction buffer (100 m*M* NaOAc pH 4.5, 1 m*M* EDTA, 0.1% SDS) for 16 h at 37°C, and ethanol precipitated. The inserts and plasmid backbone were ligated and transformed into SS320 electrocompetent cells and plated on LB/IPTG/X-gal plates (LB agar, 25 μg/mL tetracycline, 1 m*M* IPTG, 0.1 mg/mL X-gal). The DNA from specific phage plaques was PCR amplified and Sanger sequenced (Eton Bioscience, Durham, NC), and those containing the correct DNA were subsequently amplified from a single plaque.

Phage amplification

A single plaque was added to XL-1 Blue cells (OD_{600} 0.5–1), diluted to 40 mL of OD_{600} 0.05 in LB +25 μg/mL tetracycline, and shaken at 220 rpm at 37°C for

4.5–5 h. Cells were pelleted by centrifugation, and the phage supernatant was sterile filtered. Phage were precipitated by adding 1/6th volume of PEG-NaCl [20% w/v polyethylene glycol-8000 (Fisher Scientific, Pittsburgh, PA), 2.5*M* NaCl] and incubating overnight at 4°C. Precipitated phage were then pelleted via centrifugation and resuspended in TBS (50 m*M* Tris–HCl, 150 m*M* NaCl, pH 7.4). They were PEG-precipitated again (~1 h on ice), centrifuged and resuspended in 200 μL TBS. Aliquots were flash frozen and stored at −20°C with a working stock left at 4°C if imminent experiments were planned.

Clonal phage binding assay

For each phage binding reaction, 30 μL streptavidin-coated magnetic beads (Life Technologies, Dynabeads MyOne, Streptavidin T1) at 10 mg/mL were magnetically pelleted and washed with 3.3× bead volume TBS. The beads were then blocked in 3.3× bead volume 100% SB (Thermo Scientific, SuperBlock Blocking Buffer in TBS, pH 7.4) for 10 min at RT and rinsed with equal volume of 100% SB* (SB adjusted to pH 5.8 with HCl). Solution-phase beads were then resuspended in 3.3× bead volume of 100% SB* and stored at 4°C for up to 24 h. Solid-phase beads were resuspended in 3.3× bead volume PBS* + 10% SB*. To load target on to solid-phase beads, 1× bead volume of an appropriate target concentration (e.g., 10 μL of 10 μ*M* target for 10 μL beads) was added and incubated for 10 min followed by addition of 3.3× bead volume 5 m*M* D-Biotin (in PBS* + 10% SB*) and incubation for an additional 5 min. For blank (no target) beads, 3.3× bead volume of 5 m*M* D-Biotin was added and incubated for 5 min. All beads were then magnetically pelleted, washed in PBS*, and resuspended in 1× bead volume PBS* + 10% SB*.

Solid-phase binding reactions were incubated in 96-well format (Costar, sterile polystyrene, V-bottom, nontreated, Corning, Corning, NY) with shaking at 700 rpm for 2 h at RT either as 30 or 100 μL reactions in 1× PBST* (PBS* + 0.01% Tween 20) + 10% SB* and 10^{10} plaque-forming units (pfu) of the phage clone. All washes and elution were done on the KingFisher Duo magnetic particle processor (Thermo Scientific). The binding reaction was mixed on the KingFisher for 1 min at medium speed and the beads collected by 5 s dips of the magnet through the sample, repeated five times (5 × 5 s). All washes were done with PBST* (wash 1: 700 μL; wash 2: 800 μL; wash 3: 900 μL; washes 4–7: 1000 μL), mixed at slow speed for 1.5 min, and beads collected 3 × 3 s. Bound phage were eluted with 50 μL EB (0.2 *M* glycine, pH 2.2) for 10 min, beads collected 5 × 5 s, and neutralized with 7.5 μL NB (1 *M* Tris, pH 9.1). Dilutions of eluted phage were used to infect XL-1 Blue cells and then plated in top agar (40% LB agar/60% LB) on LB/IPTG/X-gal plates. Blue plaques were then counted to determine phage titers.

Solution-phase binding reactions were performed similarly to solid-phase 30 μL reactions. Instead of adding target-loaded beads to the binding reaction, an appropriate volume of 10× soluble target was added to the reaction (final 1× soluble target) just before phage were added. Additionally, on the Kingfisher Duo, target and bound phage were pulled down in a rapid 1 min magnetic pelleting step (1 min slow mixing, 5 × 5 s bead collect). All washes were done with PBST* except wash 1 which contained 5 m*M* D-Biotin to block unoccupied streptavidin sites (wash 1: 150 μL; wash 2: 700 μL; wash 3: 800 μL; wash 4: 900 μL; wash 5: 1000 μL), mixed at slow speed for 25 s, and beads collected 3 × 3 s.

Pseudovirus infectivity assays

Single-cycle pseudovirions were produced with a pNL4-3 HIV-1 genome (with firefly luciferase inserted into the nef gene and frameshift mutations in both Env and Vpr) and filovirus GP on their surface (or VSV for a specificity control). These pseudovirions were produced by co-transfecting 293T human embryonic kidney cells with the described HIV-1 genome (pNL4-3.Luc.R-E-) and a plasmid encoding the desired virus glycoprotein (SV-ZeboGPΔmuc for *Zaire ebolavirus* GP lacking the mucin domain, SV-MarVGP for marburgvirus (Musoke strain) GP and pMDG VSV-G for VSV) in the presence of polyethylenimine (PEI) transfection reagent (Polysciences, Inc., Warrington, PA). Pseudovirus-containing supernatant was collected and filtered through a 0.45 μ*M* filter 38–43 h post-transfection. For ebolavirus and marburgvirus, pseudovirions were concentrated by centrifuging through a 20% sucrose/TNE (10 m*M* Tris pH 7.6, 100 m*M* NaCl, 1 m*M* EDTA) cushion (26,000 rpm, 2 h), and the pellet resuspended in TNE, aliquoted and stored at −80°C.

To measure inhibition of infectivity, 90 μL of each inhibitor dilution and 8.9 μg/mL DEAE–dextran were added to HOS–CD4–fusion cells in a 96-well format. For each assay, a total of six inhibitor dilutions were tested, each in quadruplicate. The plates were transferred to BSL3, and 10 μL of pseudovirus diluted in media was added to each well 30–60 min after the inhibitor addition (final DEAE–dextran concentration of 8 μg/mL). Virus was diluted in order to yield a robust luciferase signal. 24 h later, all wells were inspected under a light microscope to check for gross morphological changes. Virus and inhibitor were removed via aspiration, and fresh media was replenished. 20–24 h later, the cells were lysed, and the luciferase activity was measured (Bright-Glo luciferase assay system, Promega, Madison, WI). To determine IC_{50} values, the data from each inhibitor concentration series were normalized

to the noninhibitor control signal and fit to a Langmuir equation: $y = 1/(1+[\text{inhibitor}]/IC_{50})$ (Kaleidagraph, Synergy Software, Reading, PA). The curve fit was weighted by the normalized standard error of each concentration point (with a minimum error allowed of 1%). Provided IC_{50}s are averages of two to four replicate experiments.

Filovirus immunofluorescence assays

Vero cells were seeded in 96 well black plates. Peptides and vehicle control were diluted to 1.1X final concentration in culture media [MEM, 5% FBS, gentamicin (10 μg/mL)] and 90 μL/well were incubated on the plate for 1 h at 37°C. In the BSL4 10 μL of virus diluted in media and DEAE–dextran was added to each well (final 8 μg/mL DEAE–dextran). Ebola virus (Kikwit, a representative strain of the *Zaire ebolavirus* species) and Marburg virus (Ci67) were diluted in culture media to yield a robust signal in the assay (~20% infected cells). After 1 h at 37°C, the virus and peptides were removed, the wells were washed with PBS and media with peptide was used to replenish the wells. At 24 h postinfection, each well was visualized via light microscope to look for any gross morphological abnormalities. At 48 h postinfection, the wells were washed with PBS and then the cells fixed with 10% formalin. After blocking, the fixed cells were incubated with GP-specific mAb (9G4 for Marburg virus, KZ52 for Ebola virus) followed by incubation with FITC-labeled secondary antibody (goat antimouse or antihuman, respectively). Nuclei were stained with Hoechst solution. Cells were imaged using an Operetta high content device (PerkinElmer, Waltham, MA) and images were analyzed using Harmony software to determine percent of infected cells in a given well. Data were plotted normalized to the vehicle control.

ACKNOWLEDGMENTS

We thank Hyung Kim and Dennis Winge for amino acid analysis. We also thank Yu-Chan Chen, Andrew Steiner, Ruei-Lin Hsu, and Niladri Sinha for help with molecular biology, protein purification, and preliminary characterization of the N-trimer mimics and C-peptides. For help with peptide synthesis and purification, we thank Maritza Quintero and Dasha Pruss, and we thank Matthew Movsesian for assistance and advice. This research was funded by a University of Utah Funding Incentive Seed grant to D.M.E. and M.S.K., NIH grant AI102347 to B.D.W. and M.S.K, and NIH grant GM82545 to D.M.E. and C.P.H.. We thank the U.S. Air Force for support of T.R.C.. Portions of this research were carried out at the Stanford Synchrotron Radiation Light Source, (SSRL), which is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the NIH, NIGMS. Opinions, conclusions, interpretations, and recommendations are those of the authors and are not necessarily endorsed by the U.S. Army, U.S. Air Force, the NIH or NIGMS. The mention of trade names or commercial products does not constitute endorsement or recommendation for use by the Department of the Army, the Department of Defense, or the U.S. Air Force.

REFERENCES

1. Sanchez A, Kahn AS, Zaki SR, Nabel GJ, Ksiazek TG, Peters CJ (2001) Fields virology. Philadelphia: Lippincott, Williams and Wilkins.
2. Centers for Disease Control and Prevention (2014) Chronology of Ebola hemorrhagic fever outbreaks. http://www.cdc.gov/vhf/ebola/resources/outbreak-table.html.
3. Baize S, Pannetier D, Oestereich L, Rieger T, Koivogui L, Magassouba N, Soropogui B, Sow MS, Keita S, De Clerck H, Tiffany A, Dominguez G, Loua M, Traore A, Kolie M, Malano ER, Heleze E, Bocquin A, Mely S, Raoul H, Caro V, Cadar D, Gabriel M, Pahlmann M, Tappe D, Schmidt-Chanasit J, Impouma B, Diallo AK, Formenty P, Van Herp M, Gunther S (2014) Emergence of Zaire Ebola virus disease in Guinea—Preliminary report. New Engl J Med 371:1418–1425.
4. Bossi P, Garin D, Guihot A, Gay F, Crance JM, Debord T, Autran B, Bricaire F (2006) Bioterrorism: Management of major biological agents. Cell Mol Life Sci 63: 2196–2212.
5. White JM, Delos SE, Brecher M, Schornberg K (2008) Structures and mechanisms of viral membrane fusion proteins: Multiple variations on a common theme. Crit Rev Biochem Mol Biol 43:189–219.
6. Volchkov VE, Feldmann H, Volchkova VA, Klenk HD (1998) Processing of the Ebola virus glycoprotein by the proprotein convertase furin. Proc Natl Acad Sci USA 95:5762–5767.
7. Volchkov VE, Volchkova VA, Stroher U, Becker S, Dolnik O, Cieplik M, Garten W, Klenk HD, Feldmann H (2000) Proteolytic processing of Marburg virus glycoprotein. Virology 268:1–6.
8. Carette JE, Raaben M, Wong AC, Herbert AS, Obernosterer G, Mulherkar N, Kuehne AI, Kranzusch PJ, Griffin AM, Ruthel G, Dal Cin P, Dye JM, Whelan SP, Chandran K, Brummelkamp TR (2011) Ebola virus entry requires the cholesterol transporter Niemann–Pick C1. Nature 477:340–343.
9. Chandran K, Sullivan NJ, Felbor U, Whelan SP, Cunningham JM (2005) Endosomal proteolysis of the Ebola virus glycoprotein is necessary for infection. Science 308:1643–1645.
10. Cote M, Misasi J, Ren T, Bruchez A, Lee K, Filone CM, Hensley L, Li Q, Ory D, Chandran K, Cunningham J (2011) Small molecule inhibitors reveal Niemann–Pick C1 is essential for Ebola virus infection. Nature 477: 344–348.
11. Misasi J, Chandran K, Yang JY, Considine B, Filone CM, Cote M, Sullivan N, Fabozzi G, Hensley L, Cunningham J (2012) Filoviruses require endosomal cysteine proteases for entry but exhibit distinct protease preferences. J Virol 86:3284–3292.
12. Schornberg K, Matsuyama S, Kabsch K, Delos S, Bouton A, White J (2006) Role of endosomal cathepsins

in entry mediated by the Ebola virus glycoprotein. J Virol 80:4174–4178.
13. Eckert DM, Kim PS (2001) Mechanisms of viral membrane fusion and its inhibition. Annu Rev Biochem 70: 777–810.
14. Lee JE, Saphire EO (2009) Ebolavirus glycoprotein structure and mechanism of entry. Future Virol 4:621–635.
15. White JM, Schornberg KL (2012) A new player in the puzzle of filovirus entry. Nat Rev Microbiol 10:317–322.
16. Harrison JS, Higgins CD, Chandran K, Lai JR (2011) Designed protein mimics of the Ebola virus glycoprotein GP2 alpha-helical bundle: Stability and pH effects. Protein Sci 20:1587–1596.
17. Root MJ, Steger HK (2004) HIV-1 gp41 as a target for viral entry inhibition. Curr Pharm Des 10:1805–1825.
18. Francis JN, Redman JS, Eckert DM, Kay MS (2012) Design of a modular tetrameric scaffold for the synthesis of membrane-localized D-peptide inhibitors of HIV-1 entry. Bioconjug Chem 23:1252–1258.
19. Dye JM, Herbert AS, Kuehne AI, Barth JF, Muhammad MA, Zak SE, Ortiz RA, Prugar LI, Pratt WD (2012) Postexposure antibody prophylaxis protects nonhuman primates from filovirus disease. Proc Natl Acad Sci USA 109:5034–5039.
20. Geisbert TW, Lee AC, Robbins M, Geisbert JB, Honko AN, Sood V, Johnson JC, de Jong S, Tavakoli I, Judge A, Hensley LE, Maclachlan I (2010) Postexposure protection of non-human primates against a lethal Ebola virus challenge with RNA interference: A proof-of-concept study. Lancet 375:1896–1905.
21. Marzi A, Feldmann H, Geisbert TW, Falzarano D (2011) Vesicular stomatitis virus-based vaccines for prophylaxis and treatment of filovirus infections. J Bioterror Biodef S1 (4).
22. Qiu X, Audet J, Wong G, Pillet S, Bello A, Cabral T, Strong JE, Plummer F, Corbett CR, Alimonti JB, Kobinger GP (2012) Successful treatment of ebola virus-infected cynomolgus macaques with monoclonal antibodies. Sci Transl Med 4:138ra181.
23. Warren TK, Warfield KL, Wells J, Swenson DL, Donner KS, Van Tongeren SA, Garza NL, Dong L, Mourich DV, Crumley S, Nichols DK, Iversen PL, Bavari S (2010) Advanced antisense therapies for postexposure protection against lethal filovirus infections. Nat Med 16:991–994.
24. Eckert DM, Kim PS (2001) Design of potent inhibitors of HIV-1 entry from the gp41 N-peptide region. Proc Natl Acad Sci USA 98:11187–11192.
25. Eckert DM, Malashkevich VN, Hong LH, Carr PA, Kim PS (1999) Inhibiting HIV-1 entry: discovery of D-peptide inhibitors that target the gp41 coiled-coil pocket. Cell 99:103–115.
26. Eckert DM, Malashkevich VN, Kim PS (1998) Crystal structure of GCN4-pIQI, a trimeric coiled coil with buried polar residues. J Mol Biol 284:859–865.
27. Malashkevich VN, Schneider BJ, McNally ML, Milhollen MA, Pang JX, Kim PS (1999) Core structure of the envelope glycoprotein GP2 from Ebola virus at 1.9-A resolution. Proc Natl Acad Sci USA 96:2662–2667.
28. Weissenhorn W, Carfi A, Lee KH, Skehel JJ, Wiley DC (1998) Crystal structure of the Ebola virus membrane fusion subunit, GP2, from the envelope glycoprotein ectodomain. Mol Cell 2:605–616.
29. Welch BD, VanDemark AP, Heroux A, Hill CP, Kay MS (2007) Potent D-Peptide Inhibitors of HIV-1 Entry. Proc Natl Acad Sci USA 104:16828–16833.
30. Chan DC, Fass D, Berger JM, Kim PS (1997) Core structure of gp41 from the HIV envelope glycoprotein. Cell 89:263–273.
31. Weissenhorn W, Dessen A, Harrison SC, Skehel JJ, Wiley DC (1997) Atomic structure of the ectodomain from HIV-1 gp41. Nature 387:426–430.
32. Welch BD, Francis JN, Redman JS, Paul S, Weinstock MT, Reeves JD, Lie YS, Whitby FG, Eckert DM, Hill CP, Root MJ, Kay MS (2010) Design of a potent D-peptide HIV-1 entry inhibitor with a strong barrier to resistance. J Virol 84:11235–11244.
33. Salerno WJ, Seaver SM, Armstrong BR, Radhakrishnan I (2004) MONSTER: Inferring noncovalent interactions in macromolecular structures from atomic coordinate data. Nucleic Acids Res 32: W566–W568.
34. Miller MD, Geleziunas R, Bianchi E, Lennard S, Hrin R, Zhang H, Lu M, An Z, Ingallinella P, Finotto M, Mattu M, Finnefrock AC, Bramhill D, Cook J, Eckert DM, Hampton R, Patel M, Jarantow S, Joyce J, Ciliberto G, Cortese R, Lu P, Strohl W, Schleif W, McElhaugh M, Lane S, Lloyd C, Lowe D, Osbourn J, Vaughan T, Emini E, Barbato G, Kim PS, Hazuda DJ, Shiver JW, Pessi A (2005) A human monoclonal antibody neutralizes diverse HIV-1 isolates by binding a critical gp41 epitope. Proc Natl Acad Sci USA 102: 14759–14764.
35. Montgomery DL, Wang YJ, Hrin R, Luftig M, Su B, Miller MD, Wang F, Haytko P, Huang L, Vitelli S, Condra J, Liu X, Hampton R, Carfi A, Pessi A, Bianchi E, Joyce J, Lloyd C, Geleziunas R, Bramhill D, King VM, Finnefrock AC, Strohl W, An Z (2009) Affinity maturation and characterization of a human monoclonal antibody against HIV-1 gp41. MAbs 1:462–474.
36. Schumacher TN, Mayr LM, Minor DL, Jr., Milhollen MA, Burgess MW, Kim PS (1996) Identification of D-peptide ligands through mirror-image phage display. Science 271:1854–1857.
37. Weinstock MT, Francis JN, Redman JS, Kay MS (2012) Protease-resistant peptide design-empowering nature's fragile warriors against HIV. Biopolymers 98:431–442.
38. Zawadzke LE, Berg JM (1992) A racemic protein. J Am Chem Soc 114:4002–4003.
39. Dintzis HM, Symer DE, Dintzis RZ, Zawadzke LE, Berg JM (1993) A comparison of the immunogenicity of a pair of enantiomeric proteins. Proteins 16:306–308.
40. Noren KA, Noren CJ (2001) Construction of high-complexity combinatorial phage display peptide libraries. Methods 23:169–178.
41. Weinstock MT, Jacobsen MT, Kay MS (2014) Synthesis and folding of a mirror-image enzyme reveals ambidextrous chaperone activity. Proc Natl Acad Sci USA 111: 11679–11684.
42. Hackenberger CP, Schwarzer D (2008) Chemoselective ligation and modification strategies for peptides and proteins. Angew Chem Int Ed Engl 47:10030–10074.
43. Blanco-Canosa JB, Dawson PE (2008) An efficient Fmoc-SPPS approach for the generation of thioester peptide precursors for use in native chemical ligation. Angew Chem Int Edit 47:6851–6855.
44. Wan Q, Danishefsky SJ (2007) Free-radical-based, specific desulfurization of cysteine: A powerful advance in the synthesis of polypeptides and glycopolypeptides. Angew Chem Int Ed Engl 46:9248–9252.
45. Joshi SB, Dutch RE, Lamb RA (1998) A core trimer of the paramyxovirus fusion protein: parallels to influenza virus hemagglutinin and HIV-1 gp41. Virology 248:20–34.
46. Liu IJ, Kao CL, Hsieh SC, Wey MT, Kan LS, Wang WK (2009) Identification of a minimal peptide derived from heptad repeat (HR) 2 of spike protein of SARS-

CoV and combination of HR1-derived peptides as fusion inhibitors. Antiviral Res 81:82–87.
47. Wild C, Greenwell T, Matthews T (1993) A synthetic peptide from HIV-1 gp41 is a potent inhibitor of virus-mediated cell–cell fusion. AIDS Res Hum Retroviruses 9:1051–1053.
48. Watanabe S, Takada A, Watanabe T, Ito H, Kida H, Kawaoka Y (2000) Functional importance of the coiled-coil of the Ebola virus glycoprotein. J Virol 74:10194–10201.
49. Higgins CD, Koellhoffer JF, Chandran K, Lai JR (2013) C-peptide inhibitors of Ebola virus glycoprotein-mediated cell entry: Effects of conjugation to cholesterol and side chain-side chain crosslinking. Bioorgan Med Chem Lett 23:5356–5360.
50. Miller EH, Harrison JS, Radoshitzky SR, Higgins CD, Chi X, Dong L, Kuhn JH, Bavari S, Lai JR, Chandran K (2011) Inhibition of Ebola virus entry by a C-peptide targeted to endosomes. J Biol Chem 286: 15854–15861.
51. Basu A, Li B, Mills DM, Panchal RG, Cardinale SC, Butler MM, Peet NP, Majgier-Baranowska H, Williams JD, Patel I, Moir DT, Bavari S, Ray R, Farzan MR, Rong L, Bowlin TL (2011) Identification of a small-molecule entry inhibitor for filoviruses. J Virol 85: 3106–3119.
52. Roffey R, Tegnell A, Elgh F (2002) Biological warfare in a historical perspective. Clin Microbiol Infect 8:450–454.
53. Chen BK, Saksela K, Andino R, Baltimore D (1994) Distinct modes of human immunodeficiency virus type 1 proviral latency revealed by superinfection of nonproductively infected cell lines with recombinant luciferase-encoding viruses. J Virol 68:654–660.
54. Kuhn JH, Radoshitzky SR, Guth AC, Warfield KL, Li W, Vincent MJ, Towner JS, Nichol ST, Bavari S, Choe H, Aman MJ, Farzan M (2006) Conserved receptor-binding domains of Lake Victoria marburgvirus and Zaire ebolavirus bind a common receptor. J Biol Chem 281:15951–15958.
55. Connor RI, Chen BK, Choe S, Landau NR (1995) Vpr is required for efficient replication of human immunodeficiency virus type-1 in mononuclear phagocytes. Virology 206:935–944.
56. He J, Choe S, Walker R, Di Marzio P, Morgan DO, Landau NR (1995) Human immunodeficiency virus type 1 viral protein R (Vpr) arrests cells in the G2 phase of the cell cycle by inhibiting p34cdc2 activity. J Virol 69:6705–6711.
57. Deng H, Liu R, Ellmeier W, Choe S, Unutmaz D, Burkhart M, Di Marzio P, Marmon S, Sutton RE, Hill CM, Davis CB, Peiper SC, Schall TJ, Littman DR, Landau NR (1996) Identification of a major co-receptor for primary isolates of HIV-1. Nature 381:661–666.
58. Landau NR, Littman DR (1992) Packaging system for rapid production of murine leukemia virus vectors with variable tropism. J Virol 66:5110–5113.
59. Hamburger AE, Kim S, Welch BD, Kay MS (2005) Steric accessibility of the HIV-1 gp41 N-trimer region. J Biol Chem 280:12567–12572.
60. Sanchez A, Trappier SG, Mahy BW, Peters CJ, Nichol ST (1996) The virion glycoproteins of Ebola viruses are encoded in two reading frames and are expressed through transcriptional editing. Proc Natl Acad Sci USA 93:3602–3607.
61. Studier FW (2005) Protein production by auto-induction in high density shaking cultures. Protein Expr Purif 41:207–234.
62. Blommel PG, Fox BG (2007) A combined approach to improving large-scale production of tobacco etch virus protease. Protein Expr Purif 55:53–68.
63. van den Berg S, Lofdahl PA, Hard T, Berglund H (2006) Improved solubility of TEV protease by directed evolution. J Biotechnol 121:291–298.
64. Blanco-Canosa JB, Dawson PE (2008) An efficient Fmoc-SPPS approach for the generation of thioester peptide precursors for use in native chemical ligation. Angew Chem Int Ed Engl 47:6851–6855.
65. Edelhoch H (1967) Spectroscopic determination of tryptophan and tyrosine in proteins. Biochemistry 6:1948–1954.
66. Rohl CA, Baldwin RL (1997) Comparison of NH exchange and circular dichroism as techniques for measuring the parameters of the helix-coil transition in peptides. Biochemistry 36:8435–8442.
67. Cole JL (2004) Analysis of heterogeneous interactions. Methods Enzymol 384:212–232.
68. Laue T, Shah B, Ridgeway T, Pelletier S. Computer-aided interpretation of analytical sedimentation data for proteins. (1992) Analytical ultrancentrifugation in biochemistry and polymer science. Roy Soc Chem pp 90–125.
69. Wohlfarth C. Partial specific volume of poly(ethylene glycol). In: Lechner MD, Arndt KF, Eds. (2010) Polymer solutions. Berlin: Springer.
70. O'Shannessy DJ, Brigham-Burke M, Soneson KK, Hensley P, Brooks I (1993) Determination of rate and equilibrium binding constants for macromolecular interactions using surface plasmon resonance: Use of nonlinear least squares analysis methods. Anal Biochem 212:457–468.
71. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. In: Carter J, CW, Sweet RM, Eds. (1997) Methods in enzymology. San Diego: Academic Press, pp 307–326.
72. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. J Appl Crystallogr 40:658–674.
73. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. Acta Crystallogr D 60:2126–2132.
74. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: A comprehensive python-based system for macromolecular structure solution. Acta Crystallogr D 66:213–221.
75. Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: All-atom structure validation for macromolecular crystallography. Acta Crystallogr D 66:12–21.
76. Jeffers SA, Sanders DA, Sanchez A (2002) Covalent modifications of the ebola virus glycoprotein. J Virol 76:12463–12472.
77. Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci USA 89:10915–10919.

Supplementary Information

Figure S1

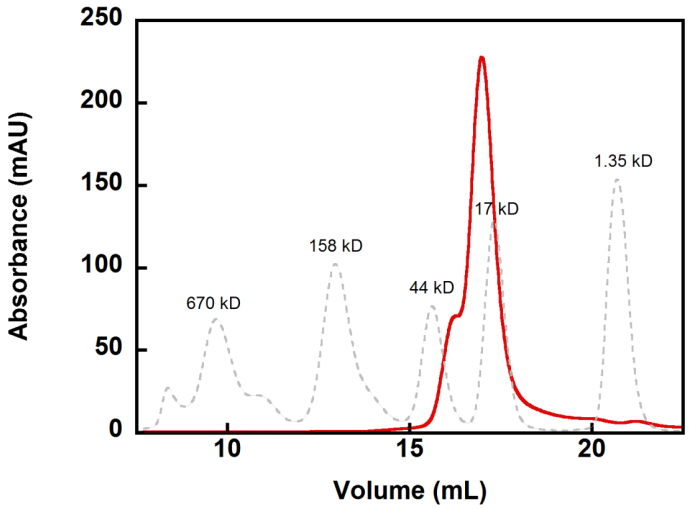


Figure S1: Gel filtration analysis of 240 µM eboIZN21 (red, recorded at 215 nm) in 50 mM sodium phosphate pH 5.8, 100 mM NaCl using a Superdex 200 column on an ÄKTApurifier (GE Healthcare Life Sciences) at room temperature with a 0.5 mL/min flowrate. Gel filtration standards (Bio-Rad) and their molecular weights are overlaid (grey, recorded at 280 nm). The predominant peak is consistent with a trimer, with a left shoulder containing higher order assemblies.

Figure S2

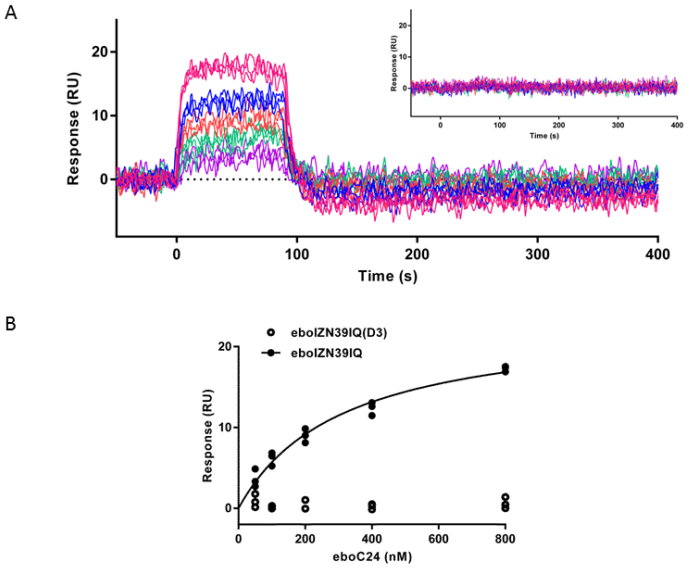


Figure S2: Binding of the ebolavirus C-peptide to the N-trimer mimic. Sensorgram of eboC24 flowed over eboIZN39IQ in a triplicate 2-fold dilution series starting at 800 nM (ProteOn XPR36, Bio-Rad), plotted with 2nd order 2-neighbor-smoothing with a Savitzky-Golay filter (Prism 6, GraphPad Software). Each dilution is shown as a distinct color. Equilibrium response data were averaged over one minute and fitted using non-linear least-squares analysis with Prism 6. The fit indicates a K_D of 310 nM. Inset: The same eboC24 dilutions flowed over an eboIZN39IQ(D3) surface. No binding was observed.

Figure S3

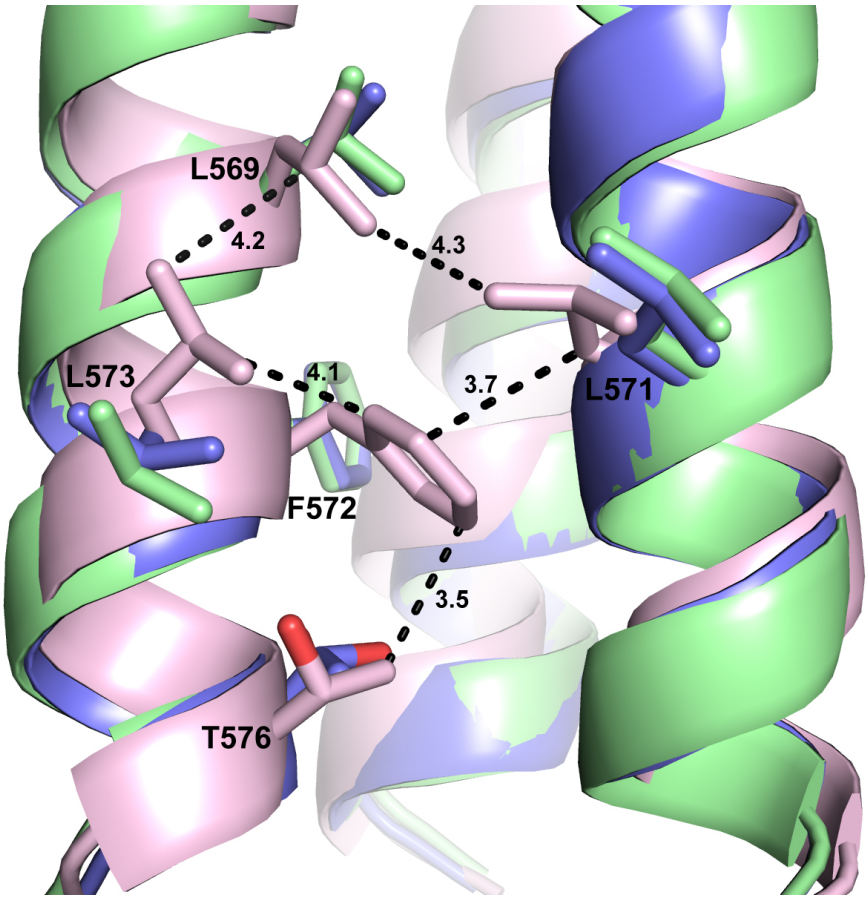


Figure S3: Hydrophobic interactions between N21 residues in the unliganded eboIZN21 structure. Shown is an overlay of the N21 region of available ebolavirus N-trimer structures in a similar view to that in Fig. 5B. Residues L569, L571, F572, L573, and T576 adopt alternate conformations in the unliganded state compared to the structures containing C-peptide and form hydrophobic interactions among themselves in the absence of ligand (dotted black lines; hydrophobic interactions for the C-peptide containing structures are not shown).

Figure S4

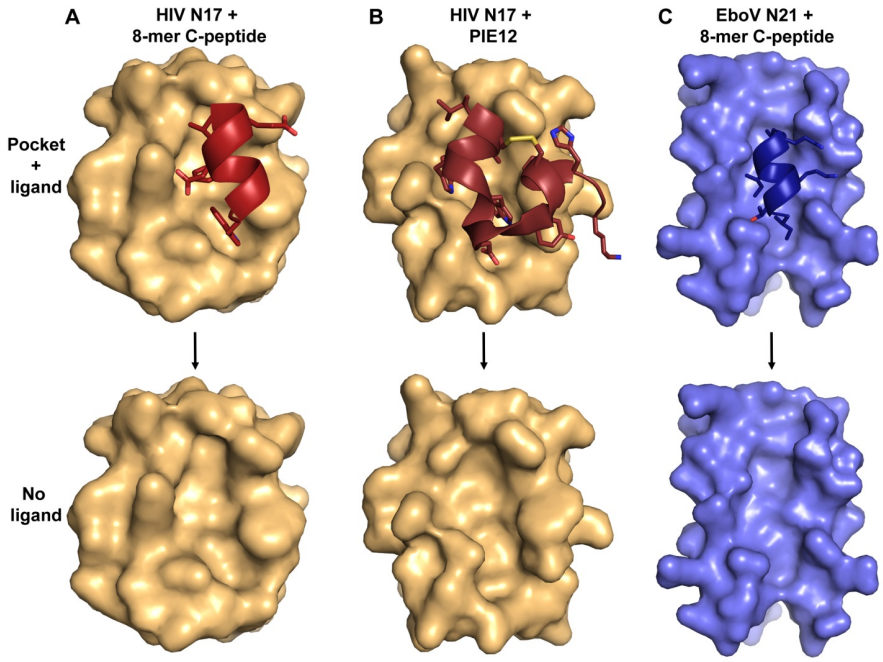


Figure S4: Comparison of hydrophobic pockets in ebolavirus and HIV N-trimers. A) Surface representation of the N17 region comprising a hydrophobic pocket in the HIV gp41 N-trimer (orange) including a cartoon representation of the eight residues (8-mer, red) of the HIV C-peptide that interact with the pocket. C-peptide residues that specifically contact pocket-forming residues are shown as sticks. The bottom panel is the same view as the top panel but without the ligand shown. B) A similar view of the HIV gp41 pocket but with the D-peptide ligand, PIE12 (dark red), bound. A comparison of the C-peptide-bound and PIE12-bound pockets indicates the shape of the pocket is ligand induced. C) A similar view of the ebolavirus N21 region (blue) from the 1EBO crystal structure showing the hydrophobic pocket with and without the 8-mer region of the ebolavirus C-peptide (dark blue) that interacts with the pocket.

Figure S5

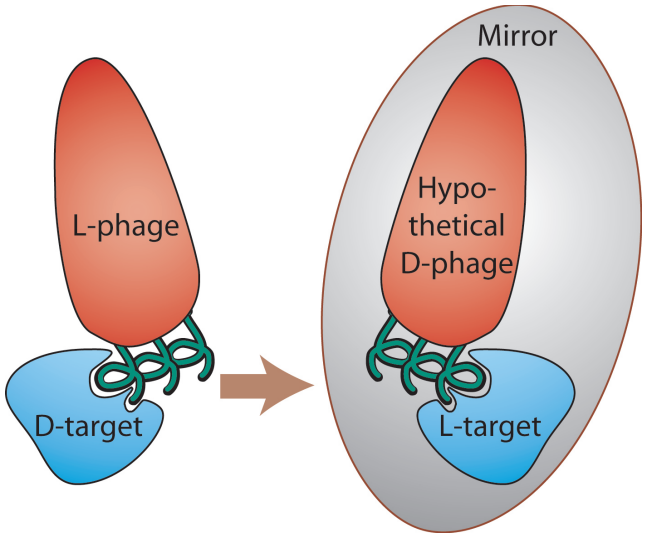


Figure S5: Mirror-Image Phage Display. In mirror-image phage display, the peptide/protein target is synthesized with D-amino acids (D-target) and forms the mirror-image of the natural L-target. Phage expressing a library of natural L-peptides (L-phage) are screened for binding to the D-target. The peptides from the specific phage clone binders are then synthesized with D-amino acids (mimicking a D-phage), and by the law of symmetry, the D-peptides will bind the natural L-target (adapted from (36)).

Figure S6

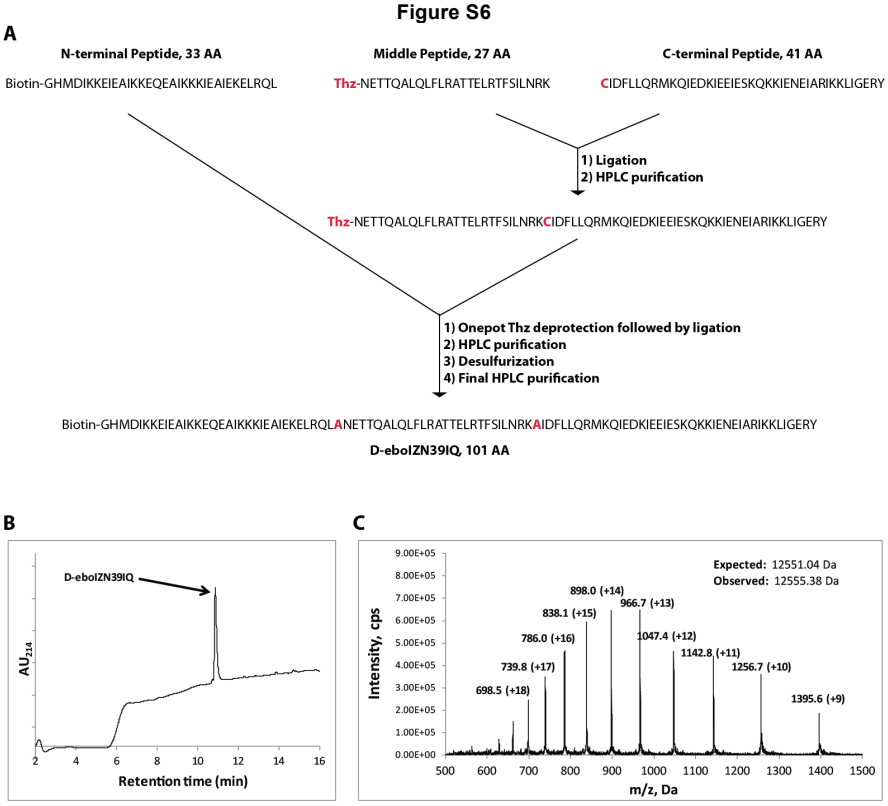


Figure S6: Synthesis of D-eboIZN39IQ. A) D-eboIZN39IQ (101 AA, with N-terminal biotin) was assembled using native chemical ligation and metal-free desulfurization. The native alanines and the residues used to replace them for native chemical ligation are indicated (red). B) HPLC analysis of final purified product D-eboIZN39IQ, using XBridge BEH130 C18 column, 2.1 x 50 mm, 5 to 90% acetonitrile gradient over 14 min, (c) MS validation showing final product with correct mass.

Figure S7

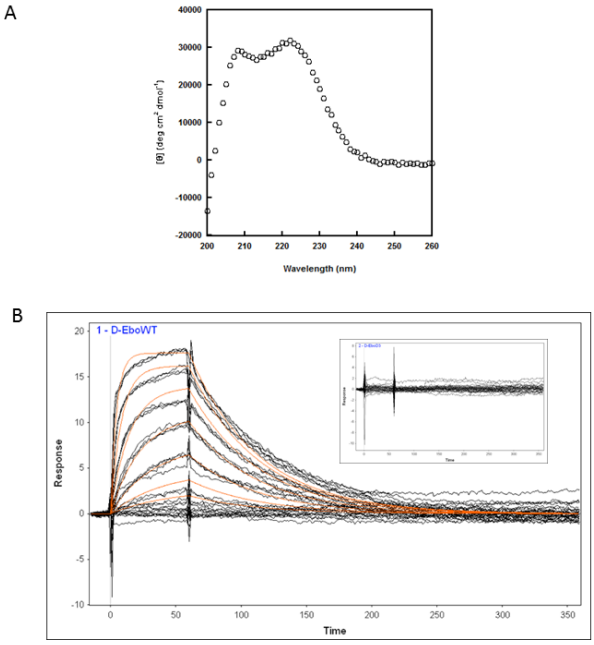


Figure S7: Biophysical characterization of the D-versions of the Ebola N-trimer mimics. A) CD spectrum of 10 μM D-eboIZN21 at 4 °C in 50 mM sodium phosphate pH 5.8, 150 mM NaCl is indicative of a highly helical conformation with 80% helicity. The positive values are as expected for this mirror-image helix. B) Analysis of binding of D-eboC37 to D-eboIZN39IQ via SPR (Biacore 3000). D-eboC37 was flowed over D-eboIZN39IQ (and D-eboIZN39IQ(D3); inset) in a 7-member 2-fold dilution series starting at 60 nM in duplicate. The fit indicates a K_D of 5.8 nM. No binding was observed to D-eboIZN39IQ(D3).

SPR methods: SPR analysis was conducted on a CM5 sensor chip (GE Healthcare) loaded with ~10,000 RU streptavidin followed by capturing ~400 RU biotin-D-eboIZN39IQ (at 40 nM in PBST* running buffer). Using Kinject, a 2-fold dilution series of D-eboC37 was flowed over the chip in duplicate at RT starting at 60 nM. A five minute dissociation time was used to ensure the response fully recovered to baseline prior to the next injection.

Table S1: eboIZN21 crystallographic data and refinement statistics

Data	
Space Group (a, b, c)	P321 (38.51, 38.51, 72.59)
Resolution (Å)	40.0 – 2.15
Resolution (Å) (high-resolution shell)	(2.23 – 2.15)
# Reflections measured	94,206
# Unique reflections	3,680
Redundancy	25.6
Completeness (%)	99.9 (99.7)
<I/σI>	18 (1.9)
Mosaicity (°)	0.68
Rpim[a]	0.010 (0.234)
Refinement	
Resolution (Å)	20.0 – 2.15
Resolution (Å) – (high-resolution shell)	(2.46 – 2.15)
# Reflections used for refinement	3,281
# Reflections in Rfree set	355
Rcryst[b]	0.272 (0.316)
Rfree[c]	0.294 (0.432)
RMSD: bonds (Å) / angles (°)	0.002 / 0.503
<B> (Å^2): protein atoms / # non-hydrogen atoms	63.0 / 434
<B> (Å^2): water molecules / # water molecules	59.7 / 6
ϕ/ψ most favored (%) / additionally allowed (%)	96 / 4

Values in parenthesis refer to data in the high resolution shell.

[a] Rpim = SQRT(1/N-1)*Σ|I-<I>|/ΣI where I is the intensity of an individual measurement and <I> is the corresponding mean value.

[b] Rcryst = Σ||*Fo*|-|*Fc*||/Σ|*Fo*|, where |*Fo*| is the observed and |*Fc*| the calculated structure factor amplitude.

[c] Rfree is the same as Rcryst calculated with a randomly selected test set of reflections that were never used in refinement calculations.

CHAPTER 4

"HELPING HANDS" FOR SIMPLIFYING CHEMICAL PROTEIN SYNTHESIS: PREPARATION OF THE CHALLENGING GROES

Michael T. Jacobsen[1], Mathieu Galibert[2], Michael S. Kay[1,3], Vincent Aucagne[2,3]

1- Department of Biochemistry, University of Utah School of Medicine, 15 N Medical Drive East Rm 4100, Salt Lake City, UT 84112-5650 (USA)

2- Centre de Biophysique Moleculaire, CNRS UPR 4301, Rue Charles Sadron, 45071 Orleans Cedex 2 (France)

3- Correspondence: kay@biochem.utah.edu (MSK), aucagne@cnrs-orleans.fr (VA)

4.1 Abstract

One of the most persistent, challenging problems in chemical peptide and protein synthesis is the handling of insoluble peptides. Although Native Chemical Ligation (NCL) has revolutionized peptide synthesis by providing a means to stitch together unprotected peptide segments, the handling of certain insoluble peptides remains a bottleneck to routinely preparing large proteins. Here, we introduce two new Solid Phase Peptide Synthesis (SPPS) building blocks, Fmoc-Ddae-OH and Fmoc-Lys(N_3-Dtpp)-OH, for incorporating and then selectively removing highly-solubilizing peptide sequences ("Helping Hands") into difficult peptides. Two key design elements in this new strategy are (1) simple, on-resin attachment of the helping hand sequence at Lys residues, and (2) selective, in-solution removal of the helping hand by gentle hydrazine treatment to regenerate the native Lys side chain. Introduction and selective removal of a helping hand sequence (Lys_6) via the Fmoc-Ddae approach is demonstrated using a model peptide, C20. This method is then applied to the chemical synthesis of the 97-aa co-chaperonin GroES, assembled via NCL. The highly insoluble peptide comprising the C-terminal region of GroES is rescued by this solubilizing helping hand approach. Furthermore, we show that removal of the helping hand, with hydrazine, can be conveniently performed in a one-pot reaction with other synthetic steps. This robust and easy-to-use new tool will be broadly useful for the synthesis of complex, poorly soluble peptides and proteins.

4.2 Introduction

4.2.1 Native Chemical Ligation and Protein Synthesis

The development of Native Chemical Ligation (NCL)[1] by Kent's group was a critical discovery toward simplifying the total chemical synthesis of peptides and proteins. Using NCL, individual peptides containing N-terminal Cys and C-terminal thioesters can be chemoselectively ligated to generate larger synthetic products. These starting peptide segments are usually prepared by solid-phase peptide synthesis (SPPS), using either Fmoc-[2] or Boc-SPPS[3] methods, and then purified by reverse-phase HPLC.

Although this two-part strategy (SPPS of peptide segments followed by NCL) has been incredibly successful—including the total synthesis of the 312-residue DapA protein[4] and 166-residue, 50,825 Da polymer-modified EPO[5]—very difficult challenges remain with handling poorly soluble and aggregation-prone sequences. Based on these synthetic challenges, we reasoned that a very useful tool for the field would be an easy-to-use, reversible linker for introducing semi-permanent solubilizing groups ("helping hands"). Below, we document several previous efforts in the literature for addressing this solubility challenge, followed by an introduction to our new tool.

4.2.2 Published Semi-permanent Solubilizing Methods

One of the earliest demonstrations of attaching cationic residues (Lys and Arg) to improve peptide solubility and handling comes from research on solubilizing hydrophobic transmembrane peptides/proteins. Led by Deber's

group, covalently installed poly-Lys and poly-Arg sequences on the termini of transmembrane helices greatly improved handling properties of these insoluble proteins[6]. These covalent strategies are not directly applicable in this work because the cationic sequences cannot be removed, but the key concept of using cationic groups to solubilize difficult/hydrophobic peptides has been established. Later, this same group employed a temporary solubilizing strategy via disulfide-bond attachment (**Fig 4.1A**); in this case, they developed a "PEG-a-Cys" reagent for modifying Cys side-chains with polydisperse PEG[7]. Although this disulfide strategy was very helpful in improving peptide solubility, it would be unstable under the highly reducing conditions present in NCL.

One of the most prominent examples of a temporary solubilizing tag is the thioester Arg_n tag, first demonstrated by the Kent[8] and Aimoto[9] groups (**Fig 4.1B**). In this method, the peptide of interest is directly prepared (on-resin) with a highly solubilizing thioester leaving group (Arg_n). This modification can be directly introduced by Boc peptide synthesis; incorporation via Fmoc-SPPS is more complicated due to the instability of thioesters to Fmoc deprotection[10]. However, a conceptually similar approach was recently demonstrated via Fmoc-SPPS using an ortho-aminoanilide cryptothioester method (**Fig 4.1C**)[11]. The thioester method has been used in several cases to endow peptides with greatly improved solubility and handling properties (see examples[12]). Unfortunately, the solubilizing tag is eliminated during the ligation reaction itself: it is either displaced by a more reactive thiol present in the reaction or eliminated during formation of the NCL product. Thus, the solubility enhancement cannot be

retained over the course of the synthesis.

Another route for solubilizing tags (**Fig 4.1D, E**) employs C-terminal base-labile linkers. In these cases, either glycolic acid (**Fig 4.1D**) or 4.hydroxymethylbenzoic acid (HMBA) (**Fig 4.1E**) are used to create base-sensitive ester linkages that are compatible with Boc[13] or Fmoc-SPPS[14], respectively. On the other end of the spectrum, an acid-labile (HF) linker (PAM, phenylacetamido) (**Fig 4.1F**) was recently used[15] for solubilizing a difficult peptide produced by Fmoc-SPPS. Overall, these base and acid-sensitive linkers have been limited to peptide termini.

Danifshefsky's group developed an Allyl-based protection approach for introducing solubilizing groups[16]. Here, custom Fmoc-Glu and Fmoc-Lys building blocks were prepared with allylic ester and allylic carbamate linkers, respectively, containing solubilizing guanidine groups (**Fig 4.1G**). Deprotection of the allyl groups was achieved with tetrakis(triphenylphosphine) palladium(0) in DMSO in the presence of excess triphenylsilane to generate native peptide structure. Although these reagents were shown to be highly useful in solubilizing EPO peptide segments, their five-step syntheses were relatively complex, and only one Arg group was incorporated per building block.

Lei Liu's group recently developed a photosensitive linker for incorporating solubilizing sequences (**Fig 4.1H**)[17]. Here, an Fmoc-Gln building block was prepared that contains an Alloc group for Pd deprotection followed by on-resin incorporation of the solubilizing sequence. Later selective removal of the solubilizing sequence and restoration of a native Gln residue is achieved by

irradiation with UV light (proceeding through the ortho nitrobenzyl linker).

Another very promising demonstration of solubilizing tags also comes from Liu's group: Gly^{RBM}, where RBM stands for **r**emovable **b**ackbone **m**odification (**Fig 4.1l**)[18]. Here, an Fmoc-Gly building block was prepared that again uses Alloc protection for incorporation of the solubilizing sequence. An elegant approach was then developed for selective removal of the solubilizing sequence—inspired by Johnson's N-(2-hydroxy-4.methoxy benzyl) (Hmb) group[19]. Protection of the 2-OH with an autocleavable linker (N-methyl-N-[2-(methylamino)ethyl]carbamoyl group)[20] provides a two-step deprotection mechanism: auto-elimination at pH 7 during the ligation reaction, followed by TFA-sensitive elimination of the Hmb group to restore a native Gly residue within the peptide.

Lastly, two cases involving enzyme-cleavable linkers (**Fig 4.1J, K**) have been published for introducing solubilizing groups, using HIV-protease (**Fig 4.1J**)[12b] and Carboxypeptidase (**Fig 4.1K**) [21].

4.2.3 A New Helping Hand Approach

These methods described above all address solubility problems with peptide segments; however in most cases, they are not sufficiently robust for routine use (e.g., due to their complex reagent synthesis, complex or uncontrolled cleavage conditions, limited placement within peptide sequences, or lability during NCL. Based on these challenges, we desire a new chemical tool (SPPS building block) with six important properties:

(1) Simple to prepare, low-cost, and usable in both L- and D-peptide syntheses,

(2) Fully compatible and easy to incorporate in peptides using Fmoc-SPPS,

(3) Compatible with NCL, desulfurization, hydrazide activation, and other methods used in assembly of peptide segments,

(4) Provides a unique functional group to attach the desired solubilizing peptide sequence for improving solubility and handling properties,

(5) Can be incorporated at diverse sites in nearly all peptide segments of the desired protein synthesis project (i.e., not limited to N- or C-termini),

(6) After the solubilizing application is complete, the introduced sequence can be simply and selectively removed to generate the native peptide sequence.

In response to this challenge, we describe the synthesis and application of Fmoc-Lys(N_3-Dtpp)-OH (**Fig 4.1L**) and Fmoc-Ddae-OH (**Fig 4.1M**) for improving the solubility of difficult peptide sequences. Using these reagents (Fmoc-Ddae-OH in particular), we can simply attach and remove, via gentle hydrazine treatment, highly solubilizing peptide sequences that we have termed helping hands (Fr. *coup de main*). We then applied a Lys_6 helping hand to facilitate the synthesis and handling of a difficult region of the GroES chaperonin protein. Because these new tools possess the six properties listed above, we believe they will enjoy wide utility in diverse chemical peptide/protein synthesis projects.

4.3 Results

4.3.1 Failed Synthesis of GroES – a "Difficult Protein" that Needs a Helping Hand

The Kay lab has a keen interest[22] in developing mirror-image D-peptides as therapeutics[23], as D-peptides have great therapeutic potential due to their resistance to natural (L-) proteases[24]. D-peptide inhibitors have been successfully developed using mirror-image phage display (MIPD)[25]; however, a key prerequisite for developing these inhibitors via MIPD is a D-protein target, which requires total chemical synthesis. Furthermore, we recognized an additional challenge in the preparation of large, complex mirror-image proteins: their precise folding in the absence of mirror-image chaperones. Accordingly, we showed that the *E. coli* GroEL/ES chaperone is ambidextrous (capable of folding both natural and mirror-image proteins)[4]. In order to generalize this observation, we are pursuing the total chemical synthesis of the *E. coli* GroEL/ES (538/97 residues) chaperone, so that we can then use D-GroEL/ES to evaluate the cross-chiral folding of a variety of recombinant L-client proteins.

On the path to synthesizing GroEL/ES, we unexpectedly found the synthesis of the short GroES co-chaperonin to be highly challenging due to purification and solubility difficulties localized to its C-terminal region. In these exploratory studies, we pursued the chemical synthesis of GroES using three different synthesis strategies (via Fmoc-SPPS). First (**Fig 4.2A**), we attempted to synthesize the full 97-aa protein as a single peptide following a highly optimized synthesis SPPS strategy (several well-spaced pseudoproline dipeptides and low-

density TentaGel R RAM resin). Although we were able to detect some correct full-length product, we were unable to cleanly isolate it by HPLC purification (**Fig 4.2B**). At this point, it should be mentioned that one literature report has described the successful chemical synthesis of GroES using Boc-SPPS in one segment[26], though neither HPLC nor MS data were provided to critically evaluate the quality of this material.

We next pursued GroES using NCL strategies, by dividing the protein into two (**Fig 4.2C, D**) or three segments (**Fig 4.2E, F**). In this case, the thioester-forming segments were prepared as C-terminal hydrazides for subsequent activation and in-situ thiolysis for NCL reaction[27]. However, these two strategies also proved very challenging—primarily due to difficulties with the C-terminal region of GroES. In the two-segment strategy, we obtained high-quality N-peptide but found the C-peptide to be poorly soluble and practically unpurifiable as tested on three different HPLC columns of different chain lengths: C4, C12, and C18 (**Fig 4.2D**). With the three-segment strategy (**Fig 4.2F**), we found that the N- and middle peptides were of good quality, while the C-terminal peptide could not be dissolved for HPLC analysis and purification.

These surprising challenges in our exploratory GroES syntheses prompted us to determine the nature of the problem and devise a chemical strategy to solve it. We speculate that the difficulty with the C-terminal region of GroES is due to its high density of acidic and hydrophobic residues. In fact, the N-terminal peptide possesses a pI of ~10.5, while the C-terminal peptide has a much lower pI of ~4.0. Based on this analysis, we decided to synthesize an

artificial version of the GroES-C protein with a covalent C-terminal Lys_6 sequence (**Fig 4.2G**) in order to more appropriately balance its charge profile. Pleasantly, we found the crude HPLC profile of GroES-C-K_6 (**Fig 4.2H**) to be much better resolved on three different columns, and the peptide solubility and handling to be greatly improved compared to unmodified control GroES-C.

4.3.2 Attempt to Synthesize GroES Using Semi-permanent HMBA Linker

Based on the promising results obtained with covalent attachment of poly-Lys to GroES-C, we tested an already published method for semi-permanently modifying this peptide—via HMBA linker[14]—in order to synthesize the full-length protein. **Fig 4.3A, B** outline this two-segment approach. The N-peptide was prepared on hydrazine resin, while the C-peptide was prepared on TentaGel R RAM resin. For the difficult GroES-C peptide, a C-terminal Lys_6 sequence was first installed on-resin, followed by an HMBA linker, a Cys(ACM) for attachment of fluorescent probes in future mechanistic studies, and then the remaining peptide sequence. The two starting peptides were purified by HPLC (**Fig 4.3C**), and then NCL was performed to generate a full-length construct (**Fig 4.3D**). Next, desulfurization of the ligation junction Cys^{42} was achieved using a free radical method[28], generating the correct product (**Fig 4.3D**).

After generating this full-length material (more soluble than the individual C-terminal segment), we tested removal of the Lys_6 using the published base-mediated method. Unfortunately, we encountered challenges with maintaining

the solubility of our full-length construct in denaturing conditions at the high pH required to completely remove the solubilizing group. Further, overnight incubation at pH ~8.5 (highest pH that we could maintain soluble sample) showed incomplete removal of the Lys_6 sequence. These complications with cleaving the HMBA linker under denaturing conditions encouraged us to pursue another route to synthetic GroES.

4.3.3 Synthesis of GroES Using Permanent C-terminal Poly-His

Our ultimate goal in the synthesis of GroES is to prepare the native protein sequence for future mechanistic studies. Thus, we were initially hesitant to prepare a version modified with a permanent C-terminal Lys_6 sequence due to the potential impact on chaperone assembly and activity. However, from literature, we found a C-terminal His_6 version of GroES was active. In this SPR study[29], a Ni-NTA SPR chip was used to immobilize GroES-His_6, which was then used to demonstrate an ATP-dependent association with GroEL. Based on these data, we pursued the synthesis of GroES using our two-segment strategy (**Fig 4.4A, B**), but retaining a permanent C-terminal His_6 in the final construct.

Although we were able to purify the GroES-C-H_6 peptide (**Fig 4.4C**), we found that its solubility and handling properties were worse than our original C-terminal Lys_6 version. Specifically, the dissolution of GroES-C-H_6 was more challenging compared to GroES-C-K_6; with the poly-Lys version, this peptide (at 30 μmol scale) can be dissolved relatively straightforwardly in ~12 mL of HPLC buffer, whereas 30 mL of HPLC buffer was inadequate for dissolving the poly-His

version. Additional rounds of dissolution were needed using 6 M GuHCl to fully dissolve the batch of GroES-C-H_6. Nevertheless, we ligated purified GroES-C-H_6 with GroES-N to generate full-length product (**Fig 4.4D**). Similar to the previous approach, this product was then desulfurized at Cys^{42} to generate GroES with an additional C-terminal His_6 (**Fig 4.4E**).

The lyophilized synthetic GroES-His_6 was then dissolved in 6 M GuHCl, 25 mM Tris, 50 mM NaCl, pH 7.5 and refolded by overnight dialysis at 4 °C (MWCO 3500, two exchanges) into this same buffer without GuHCl. In collaboration with George Lorimer (University of Maryland), the dialyzed material was then tested in an activity assay. Specifically, the ability of synthetic GroES-H_6 to inhibit steady-state ATPase activity of a single-ring mutant of GroEL (termed SR1) was evaluated (**Fig 4.4F**)[30]. Here, GroES-H_6 demonstrated SR1 binding and inhibition of ATPase activity similar to recombinant controls: GroES and H_6-GroES. However, the profile of the concentration dependence was slightly different in the synthetic sample compared to recombinant controls. This difference could be due to an altered monomer/heptamer equilibrium in GroES-H_6 compared to WT and H_6-GroES, as has been suggested in the literature[31]. However, this discrepancy might also be due to defects in the synthetic sample. Based on these somewhat vague activity data and the undesirable solubility properties of the GroES-C-H_6 peptide, we decided to attempt synthesis of native GroES using a scar-free helping hand approach.

4.3.4 Development of Helping Hand Using the C20 Model Peptide

A simple model peptide, C20 (Ac-DWTKNITDKIDQIIHDFVDK-NH_2), was used to develop the Helping Hand solubilizing strategy. C20 is derived from the C-terminal heptad repeat region (CHR) of the Ebola GP2 protein[32], which serves a critical role in viral entry[33]. A Tat-modified form of the extended Ebola C24 peptide (YGRKKRRQRRR-GSG--IEPHDWTKNITDKIDQIIHDFVDK) was recently shown to inhibit viral entry[34]. The C20 peptide has several properties that make it an appealing model peptide for developing our new Helping Hand strategy:

(1) It possesses good solubility in HPLC buffer (0.1% TFA in water/acetonitrile), and 6 M GuHCl, in order to cleanly develop our new approach;

(2) The purity of the crude synthetic product is high, and it resolves well by RP-HPLC on several different columns, simplifying data interpretation;

(3) The peptide sequence contains several important residues: one Trp for accurate concentration determination, three Lys as potential insertion sites for helping hands, and good sequence diversity (ten different residues).

Accordingly, we prepared crude C20 peptides that incorporated either Fmoc-Lys(Dde) or Fmoc-Lys(N_3-Dtpp) as reference points to develop our helping hand strategy (see data below).

4.3.5 Implementation of the Fmoc-Lys(N_3-Dtpp) Approach

The inspiration behind the two new building blocks, Fmoc-Lys(N_3-Dtpp) and Fmoc-Ddae-OH, described in this work and synthesized in the Aucagne lab comes from the reaction of nitrogen-containing nucleophiles (primary amines) with 2-acylcyclohexane-1,3-diones (see detailed review[35]) (**Fig 4.5A**). The formed enamine bonds can then be selectively cleaved by treatment with hydrazine, regenerating the original primary amine; importantly, these enamine bonds are stable to Fmoc and Boc deprotection conditions. This strategy has been used by Bycroft[36] to develop the well-established amine-protecting group Dde, (4,4.**d**imethyl-2,6-**d**ioxocyclohexylidene)**e**thyl (Dde), which is used in Fmoc-Lys(Dde) and Dde-Lys(Fmoc) building blocks offered by several commercial suppliers.

Our first attempt to install helping hands involved the new building block Fmoc-Lys(N_3-Dtpp) (1-azido-5-[1,3-**d**imethyl-2,4,6(1H,3H,5H)-**t**rioxo**p**yrimidine-5-ylidene]**p**entyl) (**Fig 4.1L**). The N_3-Dtpp linker was recently shown[37] to be stable to a variety of conditions used in chemical protein synthesis projects (including HPLC and NCL). In this published work, the linker was cleaved with 1 M hydrazine, pH 9.5 or more mild conditions with 1 M hydroxylamine, pH 8.5. Next, N_3-Dtpp-Val-OH was used to assemble, on solid-phase beads, a human mucin (MUC1) glycoprotein of 160 residues by applying a series of copper-catalyzed azide-alkyne “click” reactions, resulting in ~15 kDa (non-glycosylated) and ~20 kDa (glycosylated) proteins.

We prepared C20 and C20(N_3-Dtpp) peptides and observed clean

incorporation by SPPS **(Fig 4.5B, C)**. We also performed preliminary copper-catalyzed azide-alkyne cycloaddition with 4.pentynoic acid to demonstrate successful on-resin reaction with the azide group **(Fig 4.5B, C)**. Having established that the reagent could be incorporated and participate in a click reaction, we considered various strategies for introducing and removing solubilizing sequences (**Fig 4.5D-G**), with the goal to identify the most efficient and useful method. In our first strategy (**Fig 4.5D**), we attempted to use a click reaction to attach an unprotected poly-Lys peptide (Lys_6) to the on-resin C20(N_3-Dtpp) peptide. Here, poly-Lys was prepared on an amide resin, and the N-terminus was modified with 4.pentynoic acid to generate an N-terminal alkyne. Cleavage with TFA then produced an unprotected peptide with C-terminal amide. Unfortunately, we were unable to achieve observable poly-Lys click products, which we attributed to a phase separation issue between the hydrophobic protected on-resin peptide and the highly charged unprotected poly-Lys in solution.

In our second strategy (**Fig 4.5E**), we looked to overcome the phase separation issue by using protected poly-Lys (in contrast to unprotected peptide used above). In this case, the poly-Lys peptide (Lys_6) was prepared on 2-chlorotrityl-chloride resin and cleaved using two different conditions: 1% TFA in DCM and 20% TFA in DCM, to prepare protected forms. Although these protected peptides were easily synthesized, their poor handling properties made this approach non-ideal due to the unpredictable nature of working up, storing, and solubilizing protected peptide fragments. Although this approach may be

feasible in the future, we did not pursue it further as it steers away from our goal of making a helping hand approach that is easy to incorporate.

In our third strategy (**Fig 4.5F**), we recognized that the unique azide group could be directly reduced to an amine[38], thus providing a primary amine for simple introduction of Lys_n via Fmoc-SPPS. Unfortunately, reduction of the azide led to elimination of the linker (likely by formation of a six-membered ring intermediate), ultimately generating the native Lys side chain before any additional modifications could be performed. This reduction strategy could likely be tried again by synthesizing a new reagent with a longer side chain, in order to minimize cycle formation and subsequent self-elimination.

In our fourth strategy (**Fig 4.5G**), we designed an "adaptor approach" via copper-catalyzed azide-alkyne "click" reaction. In this design, the azide group was reacted with Fmoc-propargylamine to generate a distant primary amine for subsequent Fmoc-SPPS introduction of Lys_n. Although we were able to successfully generate C20(Fmoc-triazole-Dtpp) on-resin, we eventually decided against this approach because some on-resin sequences were found to be highly sensitive to oxygen damage when performing copper-click reaction. In particular, we observed minor histidine oxidation byproducts (+16 Da) in cases when strict, inert procedures were not carefully followed[39]. For our new helping hand approach to be generally useful to the community, it should be clean and simple-to-perform, even in nonspecialist labs.

4.3.6 Implementation of the Fmoc-Ddae-OH Approach

Based on our experience with Fmoc-Lys(N_3-Dtpp), we wished to design a new, more user-friendly approach for installing helping hands sequences. This approach (**Fig 4.6**) employs our second new building block: Fmoc-Ddae-OH, (N-Fmoc-1-(4,4.**d**imethyl-2,6-**d**ioxocyclohexylidene)-2-[2-(2-**a**minoethoxy)ethoxy]-**e**than-1-ol). Here, we were again inspired by the Dde-work from Bycroft[36a]. In this case, we designed an approach that involves the already commercially available (in both L- and D-) reagent Fmoc-Lys(Dde). We first synthesized C20 model peptide with a central Lys(Dde) (**Fig 4.7A**), creating C20(Dde). Next, we deprotected the Dde group with 3% hydrazine, producing C20 peptide (**Fig 4.7B**). Fmoc-Ddae-OH was then added to the peptide resin to react with the newly revealed unique primary amine, producing C20(Fmoc-Ddae) (**Fig 4.7C**), which was right-shifted on RP-HPLC due to the hydrophobic Fmoc and (4,4.dimethyl-2,6-dioxocyclohexylidene) groups. Time course studies showed this reaction to be nearly complete after 2 h. On-resin Fmoc-SPPS was then performed to prepare C20(K_6-Ddae) (**Fig 4.7D**), which showed a dramatic left-shift on HPLC, as expected. The full batch of C20(K_6-Ddae) peptide was then cleaved from the resin using standard TFA cleavage conditions.

This new strategy for producing C20 with a (K_6-Ddae) helping hand was very simple to employ using standard coupling and deprotection conditions. We next evaluated the selective removal of the (K_6-Ddae) helping hand via hydrazine treatment. Here, we conducted a time-course study (**Fig 4.7D**) to measure the conversion of C20(K_6-Ddae) into native C20. C20(K_6-Ddae) (~5 mM) was

dissolved in 6 M GuHCl, 100 mM phosphate, pH 7.5. Cleavage of the (K_6-Ddae) group was achieved by addition of 1 M hydrazine to the solution (final pH adjusted to 7.5). Clean elimination of the (K_6-Ddae) group was achieved with a $t_{1/2}$ of ~1 h.

These results with C20 validated our new tool for introducing helping hands. This tool possesses the six properties listed in the introduction: simple-to-prepare, easy-to-incorporate, stable to assembly conditions, provides a unique functional group, diverse sequence placement (at common Lys residues), and simple, selective removal of the solubilizing sequence.

4.3.7 Successful Synthesis of GroES Using Fmoc-Ddae Approach

The best metric for the general utility of our new Fmoc-Ddae Helping Hand approach would be to synthesize and rescue a difficult peptide. In this case, we returned to the challenging GroES synthesis, especially the intractable GroES-C peptide. Accordingly, we designed a revised GroES synthesis strategy that employed the new Fmoc-Ddae approach (**Fig 4.8A,B**). GroES-C(K_6-Ddae) peptide was synthesized and purified (**Fig 4.8C**), and then ligated to purified GroES-N via NCL to generate full-length GroES(K_6-Ddae) (**Fig 4.8D**). We next desulfurized the ligation site Cys^{42} (**Fig 4.8E**) and removed the helping hand sequence after 2 h treatment with 1 M hydrazine (**Fig 4.8F**) to create native GroES. Interestingly, the desulfurization and hydrazine steps were done in a one-pot reaction. Specifically, desulfurization of the peptide was first performed using

the free radical method (6 M GuCl, pH 6.5, 20 mM reduced glutathione, 30 mM VA-044, 300 mM TCEP) for 4 h. Second, in the same reaction (without an intervening purification), 1 M hydrazine was added to the solution (pH adjusted to 7.5) to remove the helping hand. This one-pot capability for helping hand removal is another appealing attribute of this method.

In conclusion, we were finally able to achieve synthetic GroES protein using this new, simple helping hand strategy.

4.4 Discussion and Future Plans

The new Fmoc-Ddae-OH approach for introducing and removing helping hands was demonstrated in a simple C20 model peptide, as well as the highly challenging GroES protein. The introduced helping hand sequence (K_6) was sufficient to assist in solubilizing the difficult GroES-C terminal peptide, enabling synthesis of the full-length product. Importantly, the entire procedure for introducing Fmoc-Ddae-OH and the helping hand sequence, as well as the later removal with 1 M hydrazine, was very simple to perform using standard peptide synthesis procedures. This simple method will be amenable to automation for future synthesis projects.

The possibility of using an azide-alkyne click approach with the Fmoc-Lys(N3-Dtpp) reagent is also still feasible for future research projects. The four options described above with this reagent, although not yet generalizable for broad utility, have potential for alternative/complementary helping hand approaches.

The immediate next steps with this helping hand approach are to repeat the procedure on a large scale (including yields at all synthetic steps) and generate sufficient material for activity testing in collaboration with the Lorimer lab. Interesting D-/L- studies involving GroES can now be pursued using this new technique.

4.5 Acknowledgments

Xiang Ye and George Lorimer (University of Maryland) for performing GroES activity testing; M.S.K.(University of Utah Research Foundation Seed Grant and AI076168).

4.6 Materials and Methods

4.6.1 Peptide Characterization

All peptides were prepared on a Prelude commercial peptide synthesizer (Protein Technologies, Inc.) Peptides with C-terminal amides were prepared on TentaGel R RAM resin (0.19 mmol/g, RAPP Polymere). Peptides with C-terminal hydrazides were prepared on in-house modified hydrazine-chlorotrityl resin based on published protocol[27c].

Peptides were analyzed on the following columns (from Phenomenex):

- Jupiter 5u C4 300A, 4.6 x 150 mm
- Jupiter 4u C12 Proteo 90A, 4.6 x 150 mm
- Aeris WIDEPORE C4 3.6u, 2.1 x 50 mm

Peptides were purified on the following column (from Phenomenex):

- Jupiter 5u C4 300A, 10 x 250 mm

Mass spectrometry characterization of peptides was performed using an AB Sciex 3000 LC/MS/MS system.

4.6.2 Native Chemical Ligation and Desulfurization

NCL reactions were performed according to standard protocols in the field[40], with specific adjustments for peptide hydrazides[27, 41]. Desulfurization reactions were performed according to the established free-radical protocol[28], using reduced glutathione as the thiol scavenger.

4.6.3 Cleavage of Helping Hands with 1 M Hydrazine

Cleavage of helping hands was performed in 6 M GdnHCl in 100 mM phosphate, 50 mM DTT, pH 7.5, with addition of 1 M hydrazine. Careful attention should be made to adjust the solution pH after adding hydrazine.

4.6.4 Copper Catalyzed Azide-Alkyne Cycloaddition Reactions

Copper-catalyzed azide-alkyne was performed under inert conditions using three-way valve set-up. Several cycles of vacuum-argon degassing is performed on all buffers before use. Copper bromide dimethyl sulfide complex (Sigma) is used as the catalyst. Typical reaction conditions (small-scale testing) used ~2 μmol of peptide resin, ~5 μmol of alkyne, ~10 μmol copper, and ~20 μmol DIPEA, in 500 μL of degassed NMP. Reactions were quenched with AcOH.

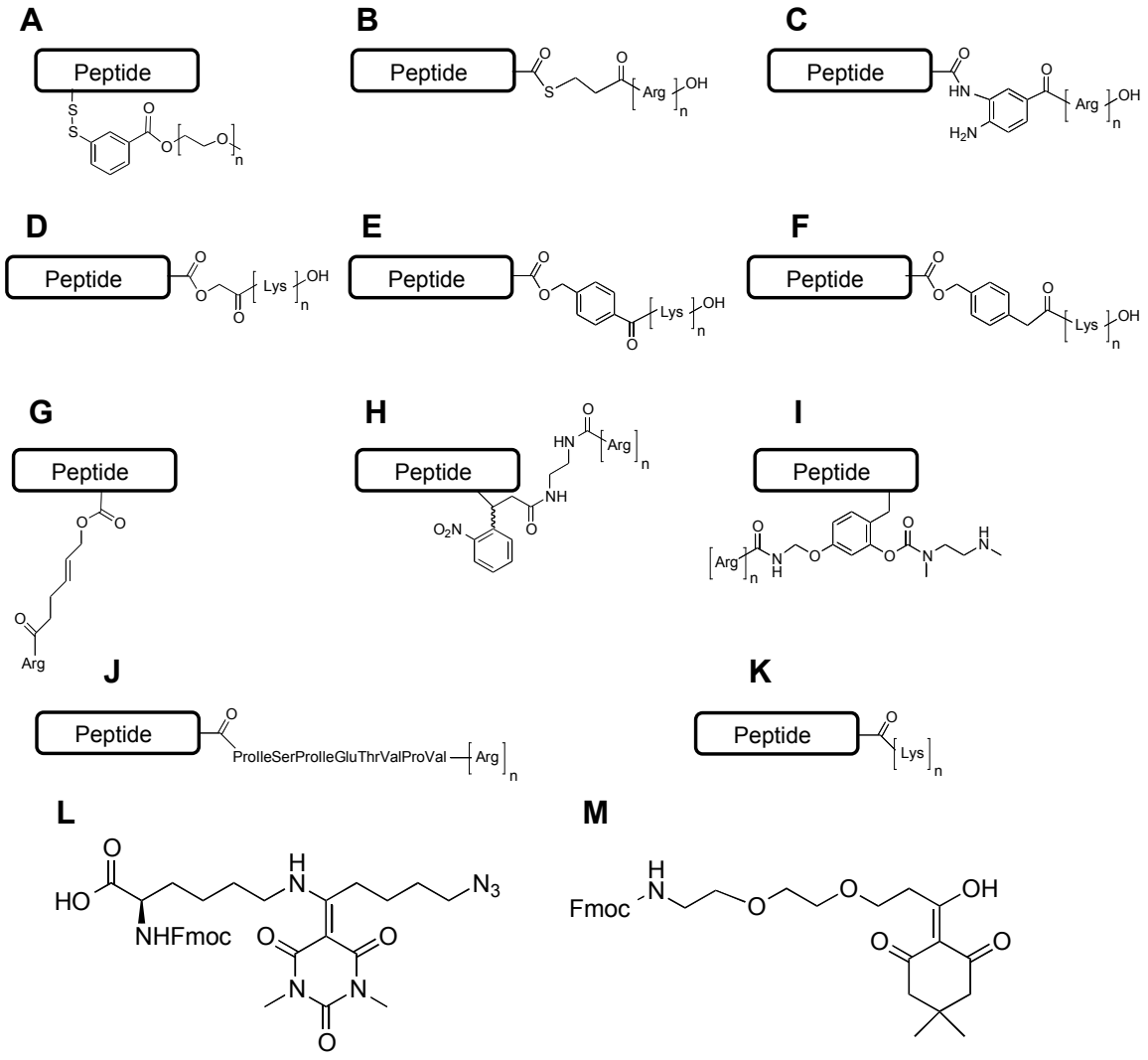


Figure 4.1: Semi-permanent solubilizing methods. (**A**) PEG-a-Cys; (**B**) Thioester tag; (**C**) *o*-aminoanilide cryptothioester tag; (**D**) Base-sensitive glycolamide-linked tag; (**E**) Base-sensitive HMBA-linked tag; (**F**) Acid-sensitive PAM-linked tag; (**G**) Allyl-linked Arg tags; (**H**) Photocleavable tag; (**I**) pH-sensitive autocleavable tag; (**J**) HIV protease-sensitive tag; (**K**) Carboxypeptidase-sensitive tag; (**L**) Fmoc-Lys(N_3-Dtpp)-OH; (**M**) Fmoc-Ddae-OH.

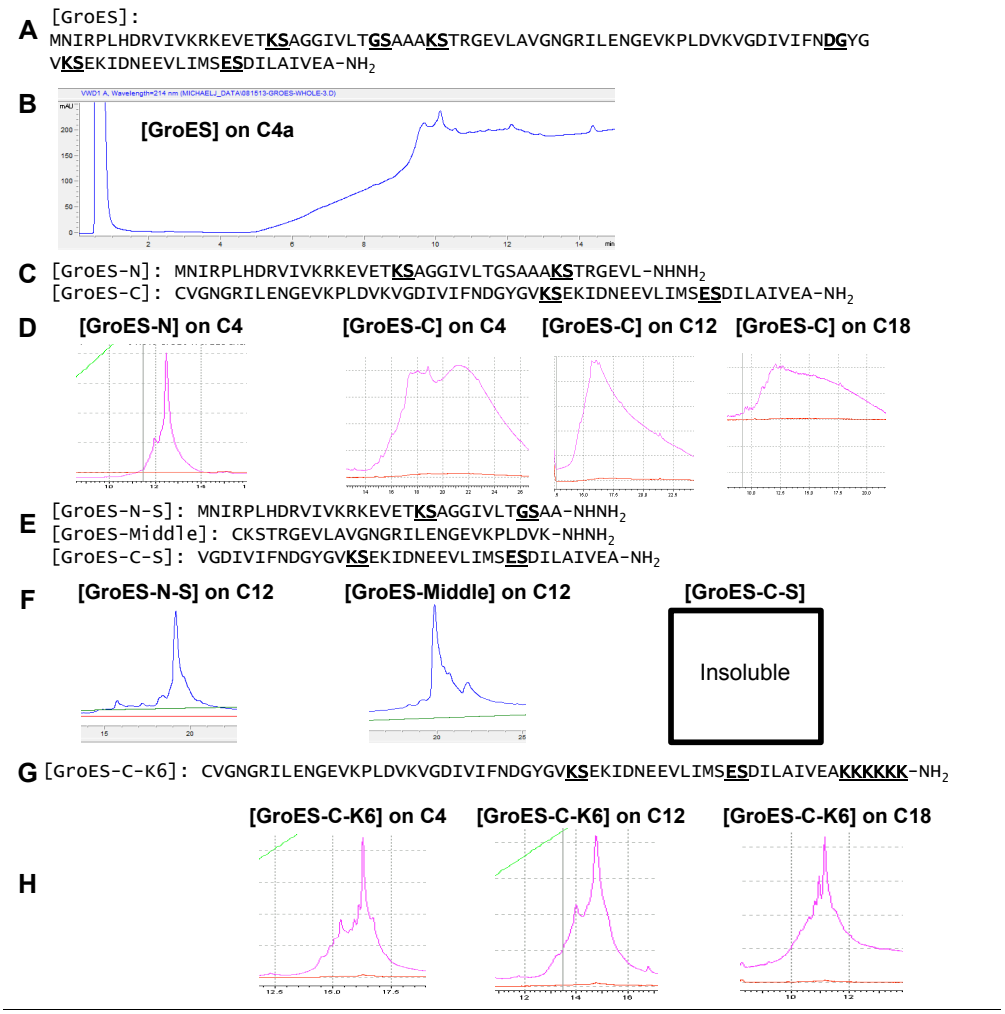


Figure 4.2: The challenging GroES synthesis. (**A**) Full-length GroES sequence; (**B**) HPLC analysis of crude GroES on C4a column; (**C**) GroES-N and GroES-C peptide sequences; (**D**) HPLC analysis of crude GroES-N and GroES-C peptides on C4, C12, and C18 columns. (**E**) GroES-N-s, GroES-Middle, and GroES-C-s peptide sequences; (**F**) GroES-N-s, GroES-Middle, and GroES-C-s peptides on C12 column; (**G**) GroES-C-K6 peptide sequence; (**H**) GroES-C-K6 on C4, C12, and C18 columns. C4 column is Phenomenex Jupiter (4.6 x 150 mm); C4a is Phenomenex Aeris (2.1 x 50 mm); C12 is Phenomenex Proteo (4.6 x 150 mm); and C18 is Waters (4.6 x 50 mm). The incorporations of pseudoproline dipeptides are indicated in bold and underline within the peptide sequences.

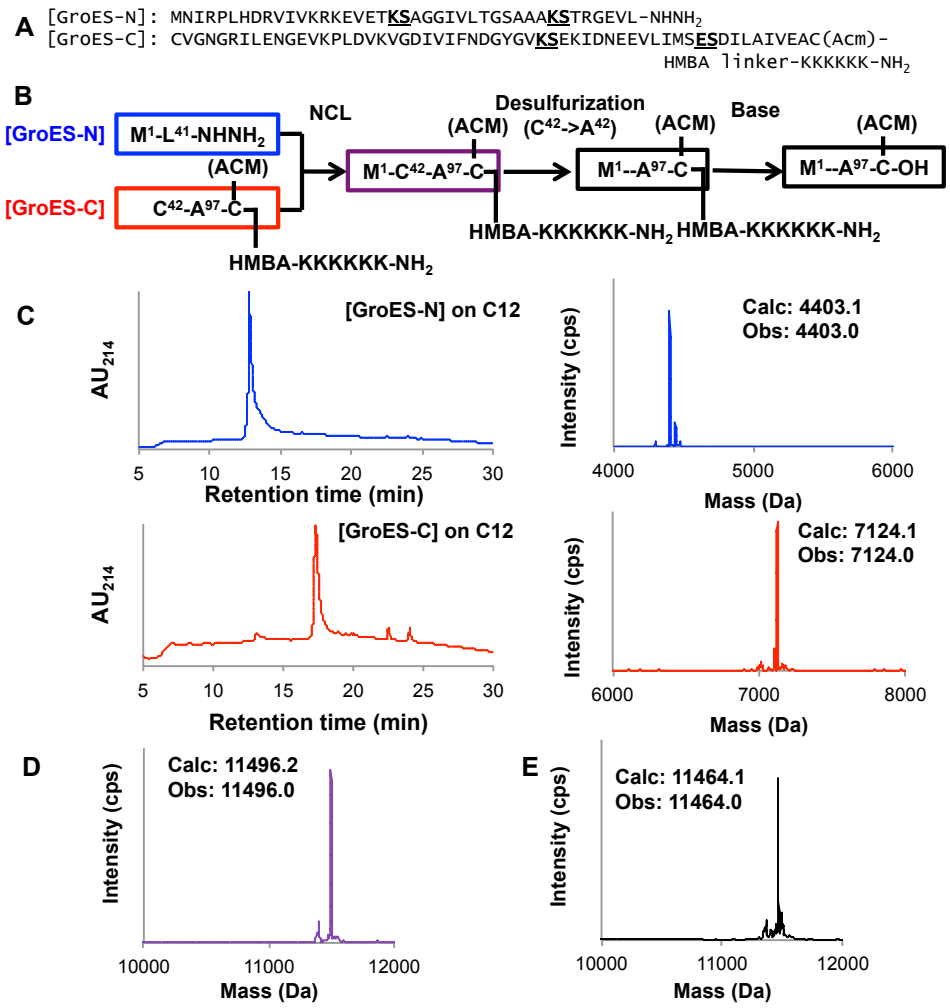


Figure 4.3: HMBA-based method to prepare GroES. (**A**) GroES-N and GroES-C peptide sequences; (**B**) Assembly strategy; (**C**) C12 (Phenomenex Proteo, 4.6 x 150 mm) HPLC and MS data on the purified GroES-N and GroES-C peptides (**D**) MS data on the purified ligation product; (**E**) MS data on the purified desulfurization product. The incorporations of pseudoproline dipeptides are indicated in bold and underline within the peptide sequences.

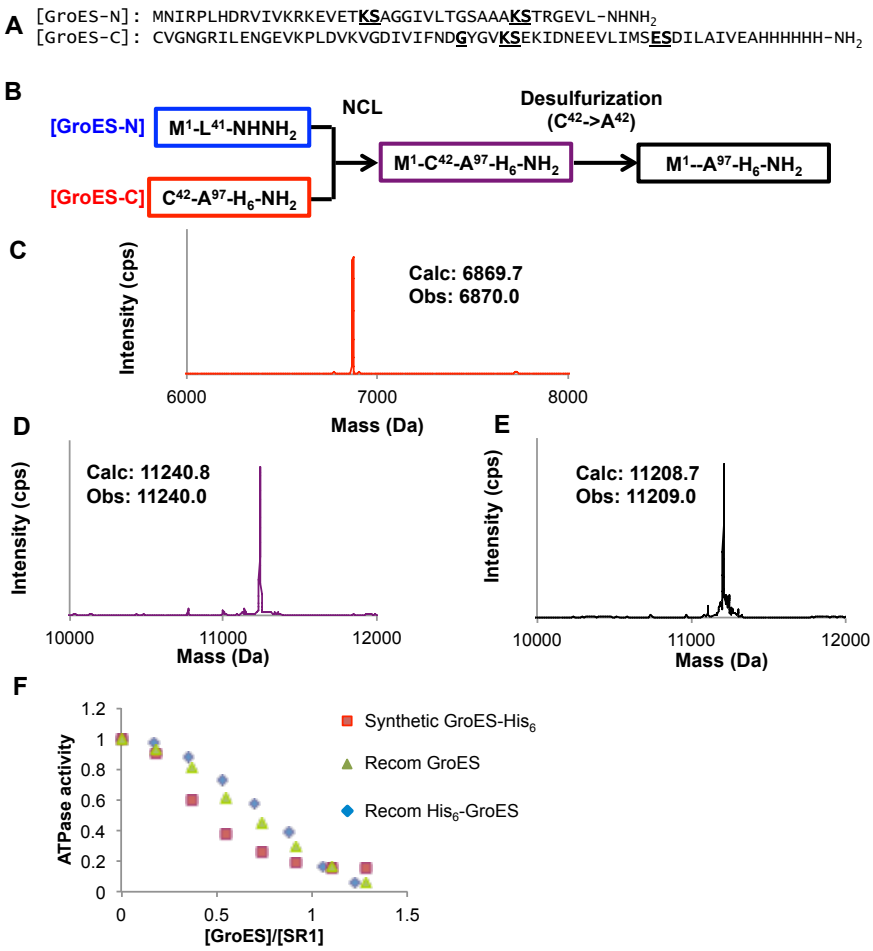


Figure 4.4: Permanent poly-His method to prepare GroES. (**A**) GroES-N and GroES-C peptide sequences; (**B**) Assembly strategy; (**C**) MS data on the purified GroES-C peptide; (**D**) Mass spec data on the purified ligation product; (**E**) Mass spec data on the purified desulfurization product; (**F**) ATPase activity assay for GroES constructs. Pseudo-proline dipeptides and N-substituted Gly residues, used to improve synthesis quality, are indicated in bold and underline within the peptide sequences.

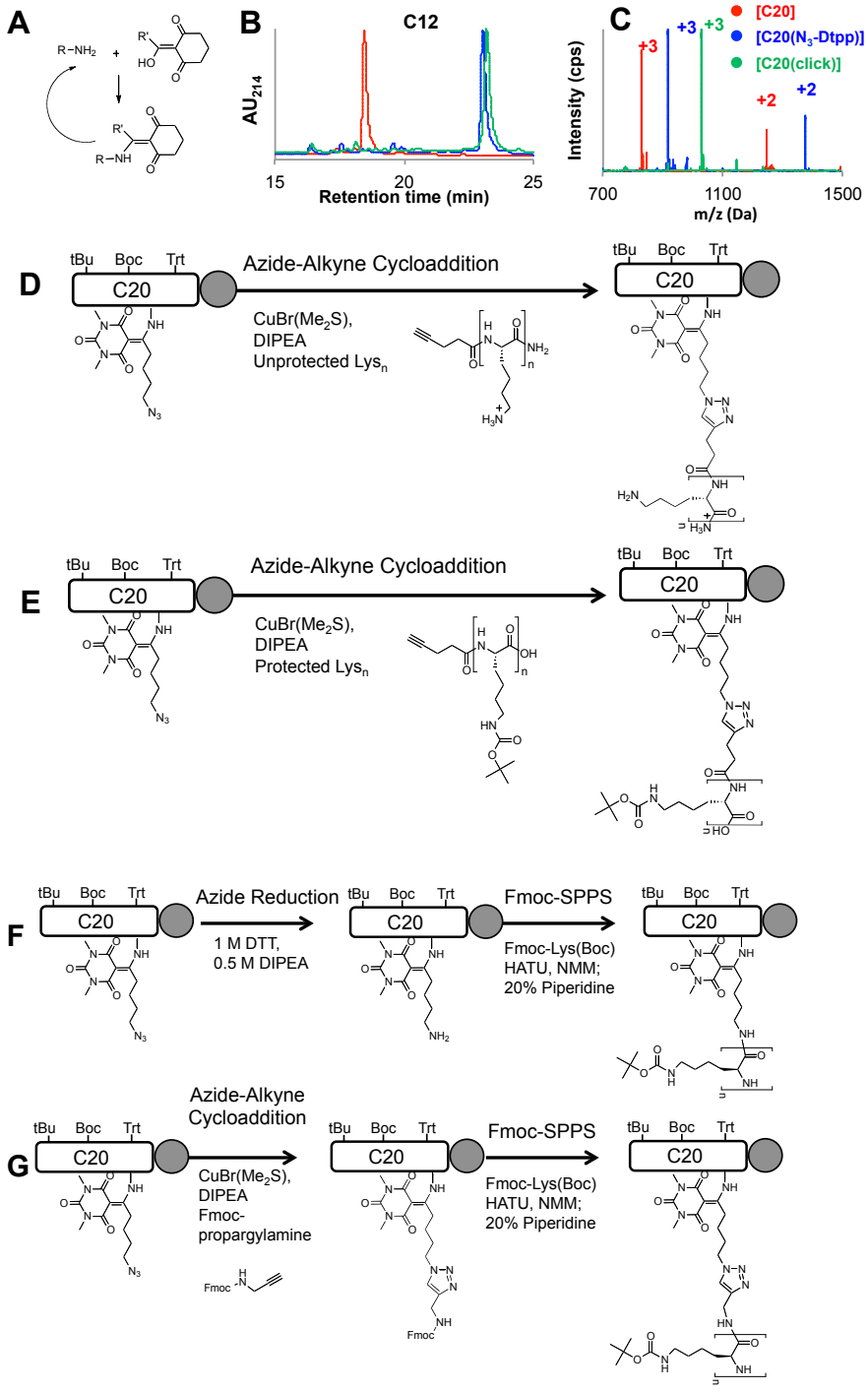


Figure 4.5: Fmoc-Lys(N_3-Dtpp) approaches to helping hand. (**A**) Reaction of 2-acylcyclohexane-1,3-diones with primary amine; (**B**) Crude [C20], [C20(N3-Dtpp)], and [C20(click)] on C12 column (4.6 x 150 mm); (**C**) Mass spec data on [C20], [C20(N3-Dtpp)], and [C20(click)]; (**D-G**) Four helping hand strategies.

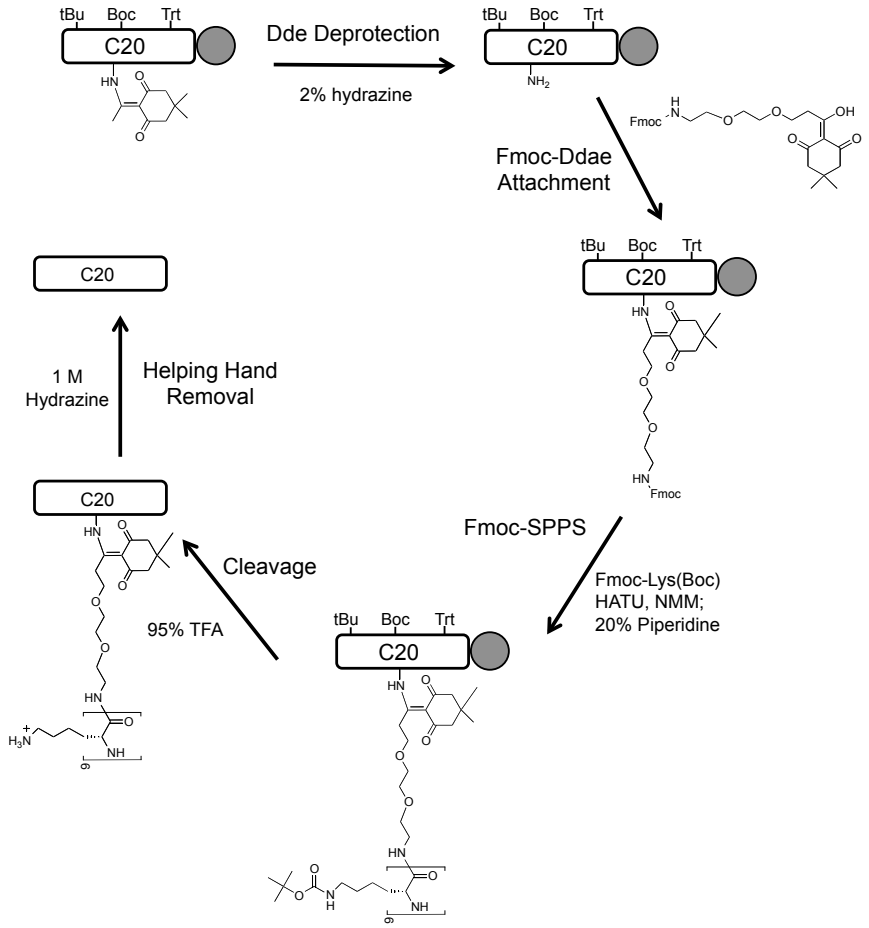


Figure 4.6: Fmoc-Ddae helping hand approach. Fmoc-Ddae-OH method for introducing helping hands proceeds using an Fmoc-Lys(Dde) intermediate, which is then deprotected with hydrazine. The single primary amine is reacted with Fmoc-Ddae-OH to generate a unique anchor for introducing helping hand sequences by Fmoc-SPPS. Cleavage of the helping hand is achieved with 1 M hydrazine to generate a native peptide.

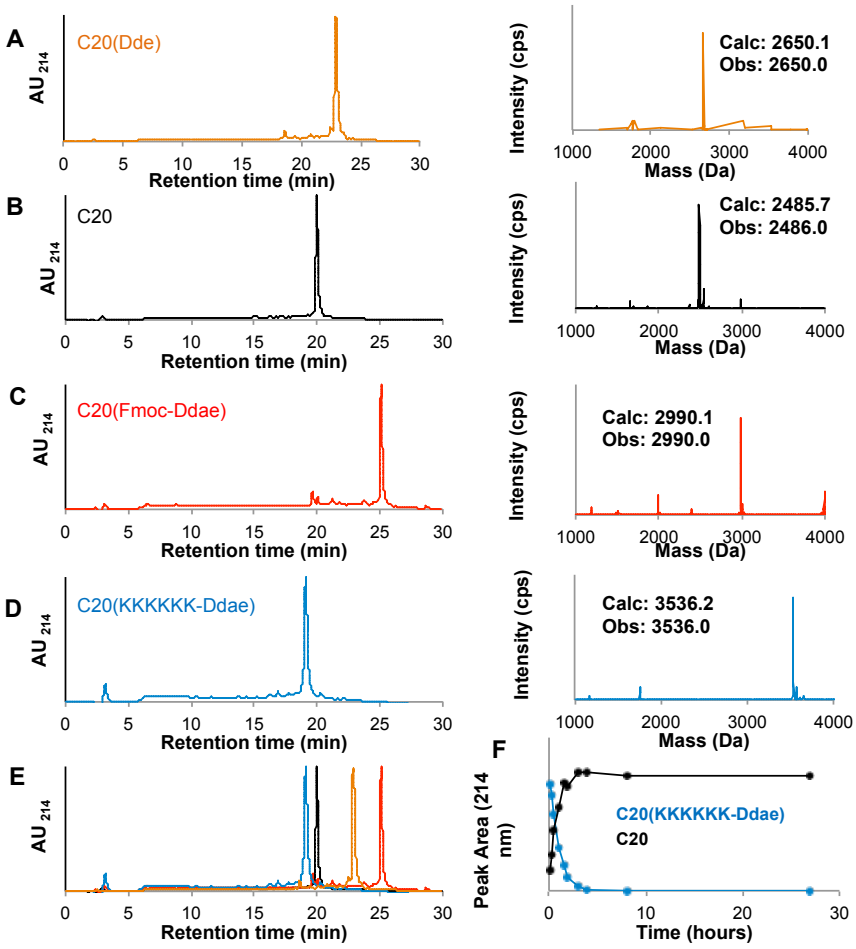


Figure 4.7: Fmoc-Ddae-OH method using C20. (**A**) HPLC and MS analysis of C20(Dde); (**B**) HPLC and MS analysis of C20; (**C**) HPLC and MS analysis of C20(Fmoc-Ddae); (**D**) HPLC and MS analysis of C20(KKKKKK-Ddae); (**E**) Overlay of HPLC traces; (**F**) Time-course study on helping hand cleavage. HPLC data on crude peptides were acquired on C12-Phenomenex (4.6 x 150 mm).

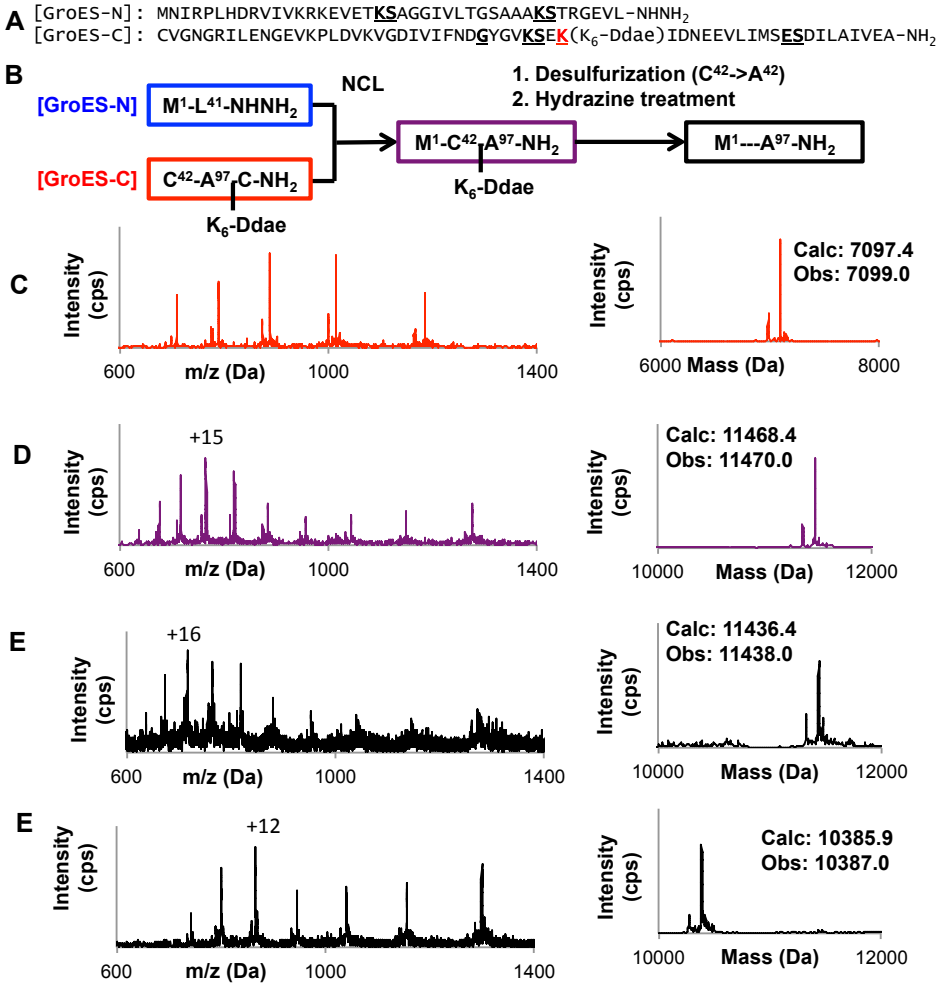


Figure 4.8: Synthesis of GroES using Fmoc-Ddae-OH approach. (**A**) GroES-N and GroES-C peptide sequences; (**B**) Assembly strategy; (**C**) MS data on the purified GroES-C peptide; (**D**) MS data on the purified ligation product; (**E**) MS data on the purified desulfurization and hydrazine-treated product. The incorporation of pseudoproline dipeptides and N-substituted Gly residues, used to improve synthesis quality, are indicated in bold and underline within the peptide sequences. Note: an Ile/Leu deletion byproduct is evident in this pilot study, which can be addressed in future syntheses.

4.7 References

1. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B., Synthesis of proteins by native chemical ligation. *Science* **1994,** *266* (5186), 776-9.

2. (a) Carpino, L. A. H., G. Y., The 9-fluorenylmethoxycarbonyl function, a new base-sensitive amino-protecting group. *J Am Chem Soc* **1970,** *92* (19), 5748-5749; (b) Carpino, L. A. H., G. Y., The 9-fluorenylmethoxycarbonyl amino-protecting group. *J Org Chem* **1972,** *37* (22), 3404-3409.

3. Schnolzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B., In situ neutralization in boc-chemistry solid phase peptide synthesis. Rapid, high yield assembly of difficult sequences. *Int J Pept Protein Res* **1992,** *40* (3-4), 180-93.

4. Weinstock, M. T.; Jacobsen, M. T.; Kay, M. S., Synthesis and folding of a mirror-image enzyme reveals ambidextrous chaperone activity. *Proc Natl Acad Sci U S A* **2014,** *111* (32), 11679-84.

5. Kochendoerfer, G. G.; Chen, S. Y.; Mao, F.; Cressman, S.; Traviglia, S.; Shao, H.; Hunter, C. L.; Low, D. W.; Cagle, E. N.; Carnevali, M.; Gueriguian, V.; Keogh, P. J.; Porter, H.; Stratton, S. M.; Wiedeke, M. C.; Wilken, J.; Tang, J.; Levy, J. J.; Miranda, L. P.; Crnogorac, M. M.; Kalbag, S.; Botti, P.; Schindler-Horvat, J.; Savatski, L.; Adamson, J. W.; Kung, A.; Kent, S. B.; Bradburne, J. A., Design and chemical synthesis of a homogeneous polymer-modified erythropoiesis protein. *Science* **2003,** *299* (5608), 884-7.

6. (a) Melnyk, R. A.; Partridge, A. W.; Yip, J.; Wu, Y.; Goto, N. K.; Deber, C. M., Polar residue tagging of transmembrane peptides. *Biopolymers* **2003,** *71* (6), 675-85; (b) Cunningham, F.; Deber, C. M., Optimizing synthesis and expression of transmembrane peptides and proteins. *Methods* **2007,** *41* (4), 370-80.

7. Pomroy, N. C.; Deber, C. M., Conjugation of polyethylene glycol via a disulfide bond confers water solubility upon a peptide model of a protein transmembrane segment. *Anal Biochem* **1999,** *275* (2), 224-30.

8. Johnson, E. C.; Kent, S. B., Towards the total chemical synthesis of integral membrane proteins: a general method for the synthesis of hydrophobic peptide-thioester building blocks. *Tetrahedron Lett* **2007,** *48* (10), 1795-1799.

9. Sato, T.; Saito, Y.; Aimoto, S., Synthesis of the C-terminal region of opioid receptor like 1 in an SDS micelle by the native chemical ligation: effect of thiol additive and SDS concentration on ligation efficiency. *J Pept Sci* **2005,** *11* (7), 410-6.

10. Li, X. K., T.; Aimoto, S., Direct preparation of peptide thioesters using an Fmoc solid-phase method. *Tetrahedron Lett* **1998,** *39* (47), 8669-8672.

11. Wang, J. X.; Fang, G. M.; He, Y.; Qu, D. L.; Yu, M.; Hong, Z. Y.; Liu, L., Peptide o-aminoanilides as crypto-thioesters for protein chemical synthesis. *Angew Chem Int Ed Engl* **2015,** *54* (7), 2194-8.

12. (a) Boerema, D. J.; Tereshko, V. A.; Kent, S. B., Total synthesis by modern chemical ligation methods and high resolution (1.1 A) X-ray structure of ribonuclease A. *Biopolymers* **2008,** *90* (3), 278-86; (b) Johnson, E. C.; Malito, E.; Shen, Y.; Rich, D.; Tang, W. J.; Kent, S. B., Modular total chemical synthesis of a human immunodeficiency virus type 1 protease. *J Am Chem Soc* **2007,** *129* (37), 11480-90; (c) Lahiri, S.; Brehs, M.; Olschewski, D.; Becker, C. F., Total chemical synthesis of an integral membrane enzyme: diacylglycerol kinase from Escherichia coli. *Angew Chem Int Ed Engl* **2011,** *50* (17), 3988-92.

13. Englebretsen, D. R. A., P. F., Boc SPPS of two hydrophobic peptides using a "solubilising tail" strategy: dodecaalanine and chemotactic protein 1042-55. *Tetrahedron Lett* **1996,** *37* (46), 8431-8434.

14. (a) Hossain, M. A.; Belgi, A.; Lin, F.; Zhang, S.; Shabanpoor, F.; Chan, L.; Belyea, C.; Truong, H. T.; Blair, A. R.; Andrikopoulos, S.; Tregear, G. W.; Wade, J. D., Use of a temporary "solubilizing" peptide tag for the Fmoc solid-phase synthesis of human insulin glargine via use of regioselective disulfide bond formation. *Bioconjug Chem* **2009,** *20* (7), 1390-6; (b) Yang, S. H.; Wojnar, J. M.; Harris, P. W.; DeVries, A. L.; Evans, C. W.; Brimble, M. A., Chemical synthesis of a masked analogue of the fish antifreeze potentiating protein (AFPP). *Org Biomol Chem* **2013,** *11* (30), 4935-42; (c) Harris, P. W.; Brimble, M. A., Toward the total chemical synthesis of the cancer protein NY-ESO-1. *Biopolymers* **2010,** *94* (4), 542-50.

15. Peigneur, S.; Paolini-Bertrand, M.; Gaertner, H.; Biass, D.; Violette, A.; Stocklin, R.; Favreau, P.; Tytgat, J.; Hartley, O., delta-Conotoxins synthesized using an acid-cleavable solubility tag approach reveal key structural determinants for NaV subtype selectivity. *J Biol Chem* **2014,** *289* (51), 35341-50.

16. Tan, Z.; Shang, S.; Danishefsky, S. J., Rational development of a strategy for modifying the aggregatibility of proteins. *Proc Natl Acad Sci U S A* **2011,** *108* (11), 4297-302.

17. Huang, Y. C.; Li, Y. M.; Chen, Y.; Pan, M.; Li, Y. T.; Yu, L.; Guo, Q. X.; Liu, L., Synthesis of autophagosomal marker protein LC3-II under detergent-free conditions. *Angew Chem Int Ed Engl* **2013,** *52* (18), 4858-62.

18. Zheng, J. S.; Yu, M.; Qi, Y. K.; Tang, S.; Shen, F.; Wang, Z. P.; Xiao, L.; Zhang, L.; Tian, C. L.; Liu, L., Expedient total synthesis of small to medium-sized membrane proteins via fmoc chemistry. *J Am Chem Soc* **2014,** *136* (9), 3695-704.

19. Johnson, T. Q., M.; Owen, D.; Sheppard, R. C., A reversible protecting group for the amide bond in peptides. Use in the synthesis of 'difficult sequences'. *J. Chem. Soc., Chem. Commun.* **1993,** (4), 369-372.

20. (a) Wahlström, K., Planstedt, O., Undén, A. , Synthesis and purification of aggregation-prone hydrophobic peptides by the incorporation of an Fmoc dipeptide with the peptide bond protected with a modified 2-hydroxy-4-methoxybenzyl (Hmb) group. *Tetrahedron Lett* **2008,** *49* (24), 3921-3924; (b) Wahlström, K., Planstedt, O., Undén, A., A carbamoyl-protective group for tyrosine that facilitates purification of hydrophobic synthetic peptides. *Tetrahedron Lett* **2008,** *49* (23), 3779-3781.

21. Chemuru, S.; Kodali, R.; Wetzel, R., Improved chemical synthesis of hydrophobic Abeta peptides using addition of C-terminal lysines later removed by carboxypeptidase B. *Biopolymers* **2014,** *102* (2), 206-21.

22. (a) Francis, J. N.; Redman, J. S.; Eckert, D. M.; Kay, M. S., Design of a modular tetrameric scaffold for the synthesis of membrane-localized d-peptide inhibitors of HIV-1 entry. *Bioconjug Chem* **2012**; (b) Welch, B. D.; Francis, J. N.; Redman, J. S.; Paul, S.; Weinstock, M. T.; Reeves, J. D.; Lie, Y. S.; Whitby, F. G.; Eckert, D. M.; Hill, C. P.; Root, M. J.; Kay, M. S., Design of a potent D-peptide HIV-1 entry inhibitor with a strong barrier to resistance. *J Virol* **2010,** *84* (21), 11235-44; (c) Welch, B. D.; VanDemark, A. P.; Heroux, A.; Hill, C. P.; Kay, M. S., Potent D-peptide inhibitors of HIV-1 entry. *Proc Natl Acad Sci U S A* **2007,** *104* (43), 16828-33.

23. Weinstock, M. T.; Francis, J. N.; Redman, J. S.; Kay, M. S., Protease-resistant peptide design-empowering nature's fragile warriors against HIV. *Biopolymers* **2012,** *98* (5), 431-42.

24. (a) Milton, R. C.; Milton, S. C.; Kent, S. B., Total chemical synthesis of a D-enzyme: the enantiomers of HIV-1 protease show reciprocal chiral substrate specificity. *Science* **1992,** *256* (5062), 1445-8; (b) Zawadzke, L. E. B., J.M., A racemic protein. *J Am Chem Soc* **1992,** *114* (10), 4002-4003.

25. (a) Schumacher, T. N.; Mayr, L. M.; Minor, D. L., Jr.; Milhollen, M. A.; Burgess, M. W.; Kim, P. S., Identification of D-peptide ligands through mirror-image phage display. *Science* **1996,** *271* (5257), 1854-7; (b) Mandal, K.; Uppalapati, M.; Ault-Riche, D.; Kenney, J.; Lowitz, J.; Sidhu, S. S.; Kent, S. B., Chemical synthesis and X-ray structure of a heterochiral {D-protein antagonist plus vascular endothelial growth factor} protein complex by racemic crystallography. *Proc Natl Acad Sci U S A* **2012,** *109* (37), 14779-84; (c) Liu, M.; Pazgier, M.; Li, C.; Yuan, W.; Li, C.; Lu, W., A left-handed solution to peptide inhibition of the p53-MDM2 interaction. *Angew Chem Int Ed Engl* **2010,** *49* (21), 3649-52.

26. Mascagni, P. T., M.; Ball, H.; Lim, M.; Ellis, R.J.; Coates, A., Chemical synthesis of 10 kDa chaperonin: biological activity suggests chaperonins do not require other molecular chaperones. *FEBS* **1991,** *286* (1-2).

27. (a) Fang, G. M.; Li, Y. M.; Shen, F.; Huang, Y. C.; Li, J. B.; Lin, Y.; Cui, H. K.; Liu, L., Protein chemical synthesis by ligation of peptide hydrazides. *Angew Chem Int Ed Engl* **2011,** *50* (33), 7645-9; (b) Fang, G. M.; Wang, J. X.; Liu, L., Convergent chemical synthesis of proteins by ligation of peptide hydrazides. *Angew Chem Int Ed Engl* **2012,** *51* (41), 10347-50; (c) Zheng, J. S.; Tang, S.; Qi, Y. K.; Wang, Z. P.; Liu, L., Chemical synthesis of proteins using peptide hydrazides as thioester surrogates. *Nat Protoc* **2013,** *8* (12), 2483-95.

28. Wan, Q.; Danishefsky, S. J., Free-radical-based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides. *Angew Chem Int Ed Engl* **2007,** *46* (48), 9248-52.

29. Nieba, L. N.-A., S. E. Persson, A.; Hämäläinen, M.; Edebratt, F.; Hansson, A.; Lidholm, J.; Magnusson, K.; Karlsson, A. F.; Plückthun, A., BIACORE analysis of histidine-tagged proteins using a chelating NTA sensor chip. *Anal Biochem* **1997,** *252* (2), 217-228.

30. (a) Horwich, A. L.; Burston, S. G.; Rye, H. S.; Weissman, J. S.; Fenton, W. A., Construction of single-ring and two-ring hybrid versions of bacterial chaperonin GroEL. *Methods Enzymol* **1998,** *290*, 141-6; (b) Grason, J. P.; Gresham, J. S.; Widjaja, L.; Wehri, S. C.; Lorimer, G. H., Setting the chaperonin timer: the effects of K+ and substrate protein on ATP hydrolysis. *Proc Natl Acad Sci U S A* **2008,** *105* (45), 17334-8; (c) Grason, J. P.; Gresham, J. S.; Lorimer, G. H., Setting the chaperonin timer: a two-stroke, two-speed, protein machine. *Proc Natl Acad Sci U S A* **2008,** *105* (45), 17339-44.

31. Zondlo, J.; Fisher, K. E.; Lin, Z.; Ducote, K. R.; Eisenstein, E., Monomer-heptamer equilibrium of the Escherichia coli chaperonin GroES. *Biochemistry* **1995,** *34* (33), 10334-9.

32. (a) Weissenhorn, W.; Carfi, A.; Lee, K. H.; Skehel, J. J.; Wiley, D. C., Crystal structure of the Ebola virus membrane fusion subunit, GP2, from the envelope glycoprotein ectodomain. *Mol Cell* **1998,** *2* (5), 605-16; (b) Malashkevich, V. N.; Schneider, B. J.; McNally, M. L.; Milhollen, M. A.; Pang, J. X.; Kim, P. S., Core structure of the envelope glycoprotein GP2 from Ebola virus at 1.9-A resolution. *Proc Natl Acad Sci U S A* **1999,** *96* (6), 2662-7.

33. (a) Eckert, D. M.; Kim, P. S., Mechanisms of viral membrane fusion and its inhibition. *Annu Rev Biochem* **2001,** *70*, 777-810; (b) Harrison, S. C., Viral membrane fusion. *Nat Struct Mol Biol* **2008,** *15* (7), 690-8.

34. Miller, E. H.; Harrison, J. S.; Radoshitzky, S. R.; Higgins, C. D.; Chi, X.; Dong, L.; Kuhn, J. H.; Bavari, S.; Lai, J. R.; Chandran, K., Inhibition of Ebola

virus entry by a C-peptide targeted to endosomes. *J Biol Chem* **2011,** *286* (18), 15854-61.

35. Rubinov, D. B.; Rubinova, I. L.; Akhrem, A. A., Chemistry of 2-Acylcycloalkane-1,3-diones. *Chem Rev* **1999,** *99* (4), 1047-1066.

36. (a) Bycroft, B. W. C., W. C.; Chhabra, S. R.; Hone, N. D. , A novel lysine-protecting procedure for continuous flow solid phase synthesis of branched peptides *J. Chem. Soc., Chem. Commun.* **1993,** (9), 778-779; (b) Chhabra, S. R. K., A. N.; Bycroft, B. W., Versatile Dde-based primary amine linkers for solid phase synthesis. *Tetrahedron Lett* **1998,** *39* (21), 3585-3588; (c) Chhabra, S. R. H., B.; Evans, D. J.; White, P. D.; Bycroft, B. W.; Chan, W. C., An appraisal of new variants of Dde amine protecting group for solid phase peptide synthesis *Tetrahedron Lett* **1998,** *39* (12), 1603-1606.

37. Galibert, M. P., V.; Piller, F.; Aucagne, V.; Delmas, A. F. , Combining triazole ligation and enzymatic glycosylation on solid phase simplifies the synthesis of very long glycoprotein analogues. *Chem Sci* **2015,** *Accepted Manuscript*.

38. Tornoe, C. W. D., P.; Porreca, F.; Meldal, M., α-Azido acids for direct use in solid-phase peptide synthesis. *J Pept Sci* **2000,** *6* (12), 594–602.

39. Hong, V.; Presolski, S. I.; Ma, C.; Finn, M. G., Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation. *Angew Chem Int Ed Engl* **2009,** *48* (52), 9879-83.

40. Johnson, E. C.; Kent, S. B., Insights into the mechanism and catalysis of the native chemical ligation reaction. *J Am Chem Soc* **2006,** *128* (20), 6640-6.

41. Zheng, J. S.; Tang, S.; Guo, Y.; Chang, H. N.; Liu, L., Synthesis of cyclic peptides and cyclic proteins via ligation of peptide hydrazides. *Chembiochem* **2012,** *13* (4), 542-6.

CHAPTER 5

TOWARDS THE TOTAL CHEMICAL SYNTHESIS OF A 352-RESIDUE DNA POLYMERASE

Michael T. Jacobsen[1], Mark E. Petersen[1], David P. Judd[1], Matthew T. Weinstock[3], Ben Stranges[2], George M. Church[2], Michael S. Kay[1,4]

1- Department of Biochemistry, University of Utah School of Medicine, 15 N Medical Drive East Rm 4100, Salt Lake City, UT 84112-5650

2- Department of Genetics, Harvard Medical School, Boston, MA

3- Synthetic Genomics, La Jolla, CA 92037.

4- Correspondence: kay@biochem.utah.edu (MSK)

5.1 Abstract

This chapter describes our efforts to synthesize a 352-residue DNA polymerase: Dpo4 (DNA polymerase IV) from *Sulfolobus solfataricus*. The total chemical synthesis of a DNA polymerase has long been an interest of the synthetic biology and chemical protein synthesis fields. An entirely synthetic polymerase would provide unmatched access to specific modifications for mechanistic studies and an important first step toward generating mirror-image life, as a D-polymerase would be an excellent tool for amplifying mirror-image L-DNA. In this study, we selected Dpo4 as our synthetic target because 1) it is one of the shortest thermostable DNA polymerases found in Nature; 2) it has been well-characterized both structurally and functionally; 3) it possesses lesion bypass ability; and 4) a mutant form (Y12A) can extend RNA in addition to DNA.

In this chapter, we detail the rational design of a six-segment strategy for synthesizing this enzyme. Several design optimizations were validated using a PCR assay for screening different mutants in recombinant Dpo4. In addition to the various synthetic tools and optimizations described here, we formally introduce the DOPPEL (Diversity-based Optimization of Peptide Properties to Enhance Ligations) concept. DOPPEL is the introduction of subtle sequence changes that can be made to greatly simplify chemical synthesis. With this approach, we identified a Dpo4 “DOPPELganger” (C31A/I101M) that has robust activity, but is much more synthetically tractable. We also introduce a TEV-cleavable Lys tag to improve solubility and the overall synthesis. Currently, we have prepared micro quantities of the full-length construct, and are scaling up.

5.2 Introduction

5.2.1 Selection of Dpo4 Polymerase

Dpo4 (DNA polymerase IV), 352 residues, was identified in 2001 using a homology search of the published *Sulfolobus solfataricus* genome[2]. *S. solfataricus* is a thermophilic archaeal bacterium that was discovered in the Solfatara volcano crater near Naples, Italy[3]. Dpo4 is a thermostable DNA polymerase that is capable of performing PCR[4]. In fact, a blend of Dpo4 and Taq was capable of amplifying UV-irradiated DNA that could not be amplified by Taq alone[5].

Similar to all known DNA polymerases, the structure of Dpo4 has been described as a right hand composed of “thumb”, “palm”, and “finger” domains[1, 6] that wraps around the DNA. The palm domain is generally well-conserved across polymerases and provides three carboxylic acid groups that are key to catalysis (in Dpo4, these are Asp^{7}, Asp^{105}, and Glu^{106}, which are conserved throughout the Y-family)[1]. The thumb and finger domains are more structurally variable and are generally associated with other important functions: translocation, DNA positioning, and processivity[6]. However, unlike most other DNA polymerases, Dpo4 also possesses a fourth domain termed the “little finger” or wrist/pad (**Fig 5.1**). The suggested function of the little finger is primarily to facilitate DNA association upstream of the active site via electrostatic interactions[7]. Additionally, domain-swapping studies between different Y-family polymerases suggest that the little finger influences processivity and the response to different types of DNA lesions[8]. A key structural feature of Dpo4 (and all Y-family members) is its

solvent-accessible active site that provides the ability to bypass DNA lesions.

Specifically, in Y-family members, the thumb and finger domains are significantly smaller compared to the other polymerases[7]. In all other families, the active site is much tighter[9], as a protein conformational change occurs upon binding of the correct complementary nucleotide. A variation of an “induced fit” mechanism has been used to describe this concept[10]. In fact, the error rate of Y-family DNA polymerases[11] is several orders of magnitude higher compared to polymerases in all other families[12]. However, it should be emphasized that this large difference in error rate is also influenced by other proofreading and mismatch repair activities—not just due to this active site accessibility[13]. On the other hand, the active site accessibility and subsequent translesion synthesis ability of Y-family DNA polymerases provide a critical role in Nature, as these enzymes prevent stalling of higher fidelity enzymes that perform most cellular DNA replication[13].

Another interesting feature of Dpo4 is its potential to perform not just DNA polymerase activity, but also RNA polymerase activity. In this case, a mutant of Dpo4 (Y12A) showed significant RNA polymerase activity, in contrast to WT[14]. However, the Y12A mutant showed less efficient overall DNA/RNA polymerase activity compared to WT. This result has been explained by a steric gate hypothesis[15], where the Y12A mutation leads to decreased discrimination at the 2’-OH group—as was shown in two crystal structures of this mutant[16]. Interestingly, nonspecific ribonucleotide incorporation by DNA polymerases, as a general phenomenon to all polymerase families, has recently attracted great

interest[17].

In conclusion, the relatively short Dpo4 is a very appealing synthetic target based not only on its DNA polymerase activity, but also on its other interesting capabilities: translesion synthesis and potential for RNA polymerase activity.

5.2.2 Dpo4 (352 aa) from the Perspective of Chemical Protein Synthesis

Total chemical protein synthesis (CPS) is an exciting field of chemistry that is dependent on the optimized application of two established procedures: solid-phase peptide synthesis (SPPS)[18] and native chemical ligation (NCL)[19]. In most cases, peptide segments are prepared by SPPS, using either Fmoc[20] or Boc[21] methods, and then these segments are chemoselectively ligated together using NCL.

Our recent synthesis of 312-residue DapA in both L- and D- chirality was an impressive achievement that extended the length record for CPS, as well as providing an interesting biological insight into chaperone-mediated protein folding[22]. Size limits in protein chemical synthesis have recently been reviewed[23]. Prior to DapA, the longest reported synthetic protein was the iterative synthesis of isopeptide-linked tetraubiquitin of 304-aa[24]. Most major CPS projects are in the range of 100 – 200 aa[25], with only these two published examples >300 aa. The routine production of proteins >300 aa is still a significant challenge.

The proposed synthesis of 352-aa Dpo4 would push this limit even further. Based on our experience with DapA, the two primary limitations to synthesizing

long proteins are (1) cumulative sample losses over sequential purification steps, and (2) handling of poorly soluble synthetic intermediates. Thus, the total chemical synthesis of Dpo4 will require a highly optimized strategy and introduction of new tools to address these limitations, as detailed throughout the chapter.

5.2.3 Synthesis Strategy and DOPPEL

Our previous synthesis of 312-residue DapA was a laborious, difficult process that required a strong intellectual partnership with a high degree of persistence (over 2 years). Although we were able to achieve this mega-synthesis twice (in L- and D- chirality), there were some aspects of the synthesis that were suboptimal. Ultimately, these challenges must be addressed for the chemical synthesis of large proteins to become routine. Some of these challenges with DapA included:

- Several inefficient and ultimately failed synthesis strategies were pursued before the final optimized route was selected. For many of these strategies, several 30 μmole batches of purified peptide were required, and many rounds of ligation (NCL) and HPLC purification were needed to determine the viability of a particular strategy.
- The quality of the synthetic L- and D- proteins (pre- and post-folded) was not ideal. Although the biological question in the DapA paper could be answered with these materials, the synthetic quality was lacking at two stages. First, the final synthetic proteins, prior to an orthogonal SEC

purification, were poorly active (~15.30% compared to recombinant protein). Second, even after an SEC purification of this material that isolated foldable tetrameric material, it was only ~80% active compared to recombinant protein. The quality of our synthetic material must be improved for future mega-synthesis projects.

- Significant yield losses accumulated over the course of the synthesis process that would be problematic in longer and more complex synthesis projects. Many of the peptides in the chemical synthesis of DapA required several batches (~100 μmol) of starting material. Thus, projects that involve more handling steps would require even more starting material (purified peptides), ultimately becoming resource-prohibitive.

Based on these challenges from the DapA project, we wished to pursue the synthesis of the longer Dpo4 protein using a more methodical strategy. We identified two important design criteria before pushing forward on this total chemical synthesis:

(1) We need to scout and then develop a highly optimized synthesis strategy to obtain high yield and quality, as well as highly functional material;

(2) We wish to introduce a new tool that can be used to routinely simplify and improve the overall synthesis process.

Concerning the first criterion (an optimized synthesis strategy), we made extensive use of pseudoproline dipeptides and 2,4-dimethoxybenzyl (Dmb) backbone amide protecting groups, low-density resins, diverse HPLC column chemistries, and strategic optimizations at all steps of the SPPS and NCL

processes to produce the highest quality peptides and assembly strategy. These exact optimizations are detailed in the sections below.

Concerning the second criterion (a new tool), in addition to using several state-of-the-art protecting groups and assembly strategies that helped enhance the overall synthesis, we desire a tool that can generally assist in all major synthesis projects. Here, we formally introduce a new tool termed DOPPEL: Diversity-based Optimization of Peptide Properties to Enhance Ligations. The key concept of DOPPEL is the introduction of synthesis-friendly mutations ("quiet mutations") that do not significantly perturb the structure and function of the desired synthetic target. In other words, we look to make well-informed, subtle sequence changes that can significantly improve the chemical synthesizability (quality and/or yield) of the target protein. These changes should not have a major impact on the function of the target protein and are first validated using recombinant protein, as will be shown below.

The DOPPEL concept, like many of the ideas in this chapter, originated from our previous challenges with the DapA synthesis. To achieve DapA, we made a single sequence change, A77C, which aided the synthesis by eliminating two synthetic handling steps: desulfurization and ACM-removal. In fact, without this change, we were unable to assemble the N-terminal half of DapA. The rationale for making the A77C version of DapA included analysis of a BLAST sequence alignment and inspection of the published crystal structure. Ultimately, the mutant was validated by recombinant protein activity testing that showed negligible effects on either DapA activity or refolding by GroEL/ES.

In this Dpo4 synthesis, we formalize this idea into the more comprehensive DOPPEL concept by evaluating several different mutants or DOPPEL versions of Dpo4 (“DOPPELgangers”). The justifications behind these mutations are described below, including both rationale for why the mutation would assist in the synthesis and why it is predicted to have minimal impact on structure/function. We then evaluated the effects of these mutations on enzyme activity via recombinant expression methods.

5.3 Results

5.3.1 Exploratory Synthesis and Peptide Scouting of Dpo4

As the 352-residue Dpo4 construct is 40 residues longer than our previous highly challenging synthesis of 312-residue DapA, we reasoned that a very efficient synthesis strategy would be needed. Therefore, before attempting the total synthesis of the entire Dpo4 protein, we decided to first scout the properties (synthesis quality and solubility) of the individual regions within the protein—to inform the final synthesis strategy. In other words, the peptides that make up the Dpo4 protein sequence can be individually synthesized and analyzed to identify any highly challenging sequences. Accordingly, the Dpo4 protein sequence was then divided at potential NCL junction sites into nine contiguous peptides ranging from 26 – 49 residues (**Figure 5.2 A, B**, peptides [s1] to [s9]). Peptides [s1] to [s8] were prepared as C-terminal hydrazides for exploratory testing of ligation reactivity, while peptide [s9] was prepared as C-terminal amide. Peptide N-termini were prepared with either Cys or Pen (Penicillamine, a thiolated version

of Val[26]) to explore ligation reactivity. All peptides were synthesized by standard Fmoc-SPPS (30 μmol scale).

HPLC and MS were then used to characterize the crude peptides (**Figure 5.3 A-I, Table 5.1**). Preliminary scouting of peptide solubility was also performed in the buffers used for HPLC purification (0.1% TFA in acetonitrile/water) and NCL reactions (6 M GuHCl, pH 7).

Fortunately, all nine peptides showed a major product with the correct mass (**Figure 5.3 A-I, Table 5.1**). Across all peptides, the most common byproducts were: Ile/Leu/Val/Pen deletions (-113/-113/-99/-130 Da), aspartimide formation (-18 Da), and various adducts typically observed to a small degree in crude peptides: TFA (+97 Da), OtBu/tBu (+56 Da), and an unresolved +40 Da modification.

Overall, we found that peptides comprising the C-terminal half of Dpo4 ([s5], [s6], [s7], [s8], and [s9]) showed slightly higher synthesis quality compared to the N-terminal half. These C-terminal peptides are rich in synthesis-friendly ("kinky") Pro residues and potential pseudoproline sites[27] (Xaa-Ser and Xaa-Thr), which encouraged us to pursue longer combined segments in our final synthesis strategy. In contrast, we found peptide [s3] to be highly challenging due to a significant degree of aspartimide formation (-18 Da aspartimide and +67 Da piperidine adduct byproducts), as well as Ile/Leu (-113 Da) deletions.

The C-terminal region of peptide [s3] contains four potential aspartimide sites (Asp-Glu, Asp-Ile, Asp-Lys, and Asp-Tyr)[28]. Indeed, synthesis of a short peptide ([ASIDE], **Figure 5.4 A, B**) containing just this region already showed

measurable aspartimide formation[29]. To address this challenge, we considered troubleshooting a series of different aspartimide-preventing additives in our Fmoc deprotection solution (e.g., HOBt, Oxyma-based additives[30], or formic acid[31]); however, we reasoned that we could reduce this problem by simply revising our final synthesis strategy. In this case, the difficult region (-**D**EAYL**D**IS**D**KVR**D**-) should be repositioned from the C- to the N-terminus of a peptide segment so as to reduce the cumulative piperidine (base) exposure time and subsequent aspartimide formation. This repositioning of the ASIDE sequence greatly reduced the aspartimide problem, as observed in the final peptides.

During our initial peptide scouting, we also found peptide [s1] to be poorly soluble under ligation conditions (<1 mM in 6 M GuHCl). This could be problematic during the subsequent assembly of our protein, as it would be difficult to drive any reactions containing this peptide by increasing the concentration of [s1]. Furthermore, any post-ligation purifications involving this segment are likely to be complicated by this solubility challenge. In response, we prepared a His_6-tagged version of this peptide ([His-s1]), reasoning that the His_6 sequence would provide a permanent solubilizing function (**Figure 5.4 C, D**). Model peptide testing showed that the His_6 addition greatly improved peptide solubility (>1 mM) (data not shown). Installation of an N-terminal His_6-tag would also facilitate the expression of our recombinant Dpo4 constructs used below. Fortunately, the quality of the [His-s1] peptide was not significantly impacted by the His_6-tag addition. Furthermore, the C-terminal peptide [s9] was adequately soluble (>1 mM) in 6 M GuHCl, providing additional support to place a His_6

sequence at the N-terminus.

In addition to scouting the synthesis quality and solubility of these peptides, we also used them to perform exploratory testing on their ligation kinetics. This testing is highly informative, as it will assist in identifying the most reactive and reliable ligation junctions for connecting individual peptide segments. In the course of this testing, we observed slow ligations (>8 h to obtain a major product peak) for segments associated with the C-terminal half: Leu thioester between [s5] + [s6], Leu thioester between [s6] + [s7], and Thr thioester between [s8] and [s9]. These observations are consistent with Dawson's published thioester reactivities[32]. We also unfortunately observed prohibitively slow NCL kinetics in all reactions involving N-terminal Penicillamine (Pen) groups. Pen (β,β-dimethylcysteine) is a β-thiol derivative of Val that is used to ligate at Val residues[33] by a ligation-desulfurization strategy[34]. We found that even reactions involving Ala thioesters (e.g., reaction of [s1] with [s2]) were very slow, with negligible product after 24 h. This result was surprising, as Pen was reported to ligate with sluggish[32] Val thioesters (i.e., forming a Val-Val junction)[33]. However, the general utility of Pen in NCL has since been questioned[35]. Ultimately, we decided to avoid Val junctions, using only Ala-based junctions (via Cys junctions). It should be mentioned that many custom cysteine surrogates have been developed for ligating at other residues, as recently reviewed[36]. However, most of these cysteine surrogates are commercially unavailable and are still synthetically challenging to produce (>7 steps), although Payne's recent work is highly encouraging on this front[37].

5.3.2 Six-Segment, Convergent Assembly Strategy

Based on the lessons from our peptide scouting, we selected our six peptide segments, ranging from 45 – 72 residues, for assembling Dpo4 (**Fig 5.5 A, B**). However, a brief note on the peptide nomenclature is needed before further discussion. The final six peptides are combinations of the initial nine scouting peptides.

Although these new peptides are relatively long (thus more difficult to synthesize and purify compared to shorter peptides), we rationalized that more difficult initial HPLC purifications are preferred over the losses and complexities associated with additional HPLC purifications of synthetic intermediates. Based on our experience, each additional HPLC purification adds at least three potential challenges:

(1) An additional complexity associated with maintaining solubilized peptide during the transition from NCL reaction buffer (6 M GuHCl, 200 mM phosphate, plus other additives) to purification conditions (0.1% TFA in water/acetonitrile), and

(2) The development of a high-resolution HPLC purification method that can sometimes be very difficult, due to closely eluting reactants and products.

(3) A decrease in yield (typically >20%) simply due to HPLC contact.

Considering this issue of balancing peptide length with the number of intermediate purifications more quantitatively, we recognized a simple relationship between number of peptide segments and minimum total number of purifications: $p = (n*2) – 1$, where p is the minimum total number of purifications

and n is the number of peptide segments (**Fig 5.6**). **Fig 5.6a** shows the minimum number of purifications required for a particular number of peptide segments based on this relationship.

To simplify the calculations, one-pot strategies[38] are omitted, as well as desulfurizations and various deprotection steps that add complexity as the number of segments increases. Thus, by decreasing the number of starting peptide segments from nine (**Fig 5.6b**) to six (**Fig 5.6c**), we have decreased the overall complexity by reducing the minimal number of purifications from 17 to 11.

Importantly, this design concept—using longer peptides with fewer purifications versus shorter peptides with more purifications—required us to employ optimized Fmoc-SPPS and HPLC purifications to generate high-quality peptides. HPLC traces of the six crude peptides are provided in **Fig 5.7**. HPLC and MS data demonstrate that very high-quality purifications of these peptides could be achieved (**Fig 5.8, 5.9, 5.10, 5.11, 5.12,** and **5.13**)

We made extensive use of pseudoproline dipeptides in all six peptides (**Fig 5.5**) to improve overall synthesis quality[27, 39]. Asp-Ser and Asp-Thr pseudoproline dipeptides were especially important, as they prevent aspartimide formation. In general, as many pseudoprolines as possible were used in the assembly, with two exceptions:

(1) In Pro and pseudoproline-rich peptides, over-incorporation was avoided (e.g., Arg-Thr and Glu-Thr pseudoprolines were not used in the already kinky [8-9] peptide),

(2) Ile-based pseudoprolines were avoided as D-Ile (ultimately needed to

synthesize D-Ile-Ser and D-Ile-Thr pseudoprolines) is unusually expensive, and we plan to synthesize D-Dpo4 for mirror-image polymerase studies using the same synthetic route employed here.

In addition to the extensive incorporation of pseudoproline dipeptides, we strategically introduced 2,4-dimethoxybenzyl (Dmb) N-substituted Gly^{40} to improve synthesis quality in peptides [1-2] and [5-6]. In the case of [1-2], we prepared a truncated version of this peptide and showed that insertion of (Dmb)-Gly greatly reduced His deletion byproducts associated with the N-terminal His_6 sequence (**Fig 5.14 A**). Based on this result, we then incorporated (Dmb)-Gly in the full [1-2] sequence.

In the case of [5-6], we found the crude peptide to be relatively low quality (**Fig 5.7 D and 5.14 G**), although a clearly isolated correct peak was observed. The relatively low quality of [5-6] is not entirely surprising due to its low density of Pro and pseudoproline residues compared to the other peptides. To address this problem, we simplified it by focusing on a truncated version: peptide [6] (**Fig 5.14 B**). This peptide was then prepared with Asp-Thr pseudoproline dipeptide (**Fig 5.14 C**). Impressively, early insertion of a single (Dmb)Gly greatly improved quality (**Fig 5.14 D**).

Alternatively, we investigated the effect of alternative Lys side-chain protecting groups to improve synthesis quality. In this case, we substituted three Lys(Boc) in the sequence with one of two alternatives: Lys(Mtt) or Lys (Dde) (**Fig 5.14 E, F**). We hypothesized that the more bulky protecting groups Mtt and Dde could reduce on-resin aggregation compared to the more compact Boc protecting

group. It has been reported in the literature that these protecting groups can improve synthesis quality in short peptides. However, as can be seen in **Fig 5.14 E & F**, these substitutions did not improve synthesis quality, and thus Boc protection was retained at all Lys positions

The dramatic synthesis improvement seen in [6] with one (Dmb)Gly was then expanded to the more difficult context of full-length [5-6]. Again, the overall quality of the longer full-length crude peptide was significantly improved (compare **Fig 5.14 G & H**).

However, outside these two cases in [1-2] and [5-6], we avoided further incorporation of (Dmb)Gly throughout the project due to difficulty in coupling the next (N-terminal) residue[40]. The other Gly residues are located in positions not ideal for incorporating (Dmb)Gly protection for a variety of reasons:

(1) They precede a difficult-to-couple hindered residue,

(2) They are located next to an already introduced pseudoproline dipeptide, which makes them likely to be unnecessary,

(3) They are too late (too N-terminal) in the peptide to have a measurable effect on overall synthesis quality.

Nevertheless, we note one particular appeal of the commercially available (Dmb)Gly in improving synthesis quality: it can be directly used in both L- and D-synthesis projects, as Gly is an achiral residue. In contrast, the established pseudoproline dipeptides (Xaa-Ser and Xaa-Thr, detailed above) are commercially limited to L-chirality, so these must be custom-prepared in mirror image for future D- synthesis projects (an effort already underway with Mark

Petersen).

Lastly, we wish to emphasize the importance of performing Fmoc-SPPS at very low densities (~0.2 mmol/g) for improving synthesis quality. **Fig 5.15** shows the crude synthesis quality of an exploratory version of peptide [2] when prepared at two different densities: ~0.5 (**Fig 5.15 A**) and ~0.2 mmol/g (**Fig 5.15 B**). Based on this result, all peptides were subsequently synthesized at densities <0.25 mmol/g. We could then outline the assembly strategy, applying two rules:

(1) Employ a convergent assembly to maximize overall efficiency[41],

(2) Plan the order of assembly so that the last ligation reaction utilizes a highly reactive and reliable thioester.

Concerning the order of assembly, there are five thioesters to consider: Lys, Ile, Ala, Lys, and Lys (**Fig 5.5 B, Fig 5.15 C**). The highly reactive Ala junction[32] between peptides [4] and [5-6] was selected for the final ligation reaction. Our logic is that because the last ligation uses the most precious synthetic material, this reaction should be performed under the most favorable conditions, i.e., using the most optimal available thioester. Further, although Lys thioesters are highly reactive, they have recently been found to form unreactive thiolactone species[42]—another reason to avoid Lys as the final thioester.

Thus, our assembly strategy was split into two halves centered on the Ala junction (**Fig 5.15 C**): N-terminal half with peptides [1-2], [3], and [4] and C-terminal half with peptides [5-6], [7-8], and [8-9]. Starting with this framework, we describe the challenges that we encountered in the total synthesis of these two halves, followed by the necessary DOPPEL changes. These five sections

described below include:

(1) N-terminal Half: How to Handle the His tag?

(2) N-terminal Half: The C31 Conundrum

(3) N-terminal Half: I101-Inspired Precipitation

(4) C-terminal Half: Desulfurization Dilemmas

(5) C-terminal Half: Taking a Step Back in Ligations.

5.3.3 (1) N-terminal Half: How to Handle the His tag?

An important observation during the peptide scouting stage was the poor solubility of the [s1] peptide, which was remedied by installation of an N-terminal His_6 tag. Thus, our recombinant Dpo4 construct was produced with an N-terminal His_6 tag (WT construct, **Fig 5.16 A**), which was also used to facilitate the purification via Ni affinity resin.

Our first efforts to perform PCR using this WT construct were relatively unsuccessful. **Fig 5.16 B** shows the results of a PCR assay using M13KE plasmid template, and two short 20 nt primers with the following PCR conditions: 15 seconds denaturation at 86, 89, 90.6, or 92 °C, 30 seconds annealing at 56°, and 1 minute extension at 60°, repeated over 25 cycles (generating a product ~200 nt). The positive control ("NEB") is a commercially available Dpo4 protein from a similar thermophile, *Sulfolobus islandicus*. Here, the NEB sample was active up to 92 °C, while the WT sample was only active at 86 °C. To determine this discrepancy in activity between our *S. solfataricus* and the NEB *S. islandicus* samples, we first identified the sequence variations. Based on the published

molecular weight of the NEB reagent, we were able to deduce the likely *S. islandicus* strain from a BLAST search of the reference *S. solfataricus*. Compared to our *S. solfataricus* sequence (WP_009993137.1), the *S. islandicus* strain (WP_014513455.1) contained 24 sequence changes (out of 352 residues) although most of these changes are relatively conservative: Lys/Arg, Ile/Val (**Fig 5.16 C)**. However, we decided against engineering our vector to mimic the *S. islandicus* sequence due to significant peptide scouting that had already been performed on our native sequence.

In addition to the protein sequence differences (**Fig 5.16 C)**, our construct contained an additional N-terminal His_6 tag. To determine if the His-tag interfered with activity, we generated a new construct containing a TEV-cleavable site between the His_6 tag and the protein sequence (TEV construct, **Fig 5.16 A**). This construct allowed us to purify the protein using a Ni affinity column and then remove the N-terminal His_6 tag by TEV protease digestion (courtesy of the Hill lab). Unfortunately, we were unable to digest this particular construct with TEV, likely due to limited accessibility of the TEV site. To overcome this issue, we generated a new Dpo4 construct containing an extended GG insert between the His_6 tag and the protein sequence (TEVX construct, **Fig 5.16 A**). Using very high concentrations of TEV (~0.1 mg/mL per ~1 mg/mL Dpo4) for an extended digestion period (>12 h), we were able to achieve nearly complete TEV cleavage to generate the His-tag free protein (Naked WT, NWT construct, **Fig 5.16 A)**. Pleasantly, we found that the His-tag-free Dpo4 construct showed PCR ability similar to the NEB sample (**Fig 5.16 B**).

Unfortunately, chemical synthesis of His-tag-free Dpo4 would be more difficult without the solubilizing N-terminal His_6 tag, as described in the peptide scouting sections above. Therefore, we then wondered if the inhibitory effect of the His_6 tag is strictly dependent on (a) the simple presence of an N-terminal His_6 sequence, or (b) the location of our N-terminal His_6 being directly abutted against the N-terminus of Dpo4. To address this question, we evaluated the PCR ability of our TEVX construct (**Fig 5.17 A**) prior to TEV cleavage. To our surprise, the TEVX construct is nearly as active as the NWT Dpo4 protein (**Fig 5.17 A**). This important result means that we can maintain the N-terminal His_6 in our synthesis strategy by simply distancing it from the Dpo4 protein via TEV linker. Alternatively, the tag could still be cleaved by TEV digestion following total synthesis (though this would involve a handling step that would diminish yield).

During the course of our investigation of WT Dpo4 polymerase activity, we also developed more gentle conditions that allowed for robust activity of the original WT construct (**Fig 5.17 B**): 5 sec denaturation at 84, 84.9, 86, or 87.2 °C, 30 sec annealing at 56°, and 1 min extension at 60°, for 25 cycles.

These conditions employ a shorter denaturation time and lower denaturation temperatures compared to our original method. These gentle conditions were used to evaluate our DOPPEL changes in the next sections.

One last design optimization was developed during this stage of testing. Based on the data showing that the TEVX and NWT constructs were significantly more active than the original WT construct, we recognized that this could be exploited to further assist the chemical synthesis. Instead of incorporating a His_6

sequence, we reasoned that any solubilizing sequence could be introduced in the synthetic construct N-terminal to the TEV site for two reasons:

(1) The His6 sequence is not important for purification of synthetic constructs, so any solubilizing sequence could be introduced, and

(2) The final synthetic construct can be cut with TEV to generate NWT sample, so the temporary solubilizing group could be removed before activity testing.

Thus, we modified our synthesis plan to incorporate a Lys_7-TEV site at the N-terminus instead of the original His_6 sequence. Based on solubility data from GroES in Chapter 3, poly-Lys is more solubilizing than the corresponding poly-His.

5.3.4 (2) N-terminal Half: The C31 Conundrum

Dpo4 (**Fig 5.2, 5.5**) possesses only one native Cys residue: C31. Unfortunately, this one single Cys adds significant complexity to the overall assembly. In order to effectively explain this added complexity, it is best to first consider the assembly strategy as if there were no native Cys residues, as shown in **Fig 5.18 A**. Here (with no Cys constraints), the six segments can be assembled in the most optimal manner (based on solubility, preferred protecting groups, and thioester method), and then one single global desulfurization reaction is performed at the last step to eliminate the five junction cysteines.

However, the presence of one native Cys constrains the overall assembly strategy. Specifically, there are two different approaches that can be used to

retain Cys31. In the first approach (**Fig 5.18 B**), the C31 position is internal.

After the whole protein has been assembled and desulfurized using the same approach as **Fig 5.18 A**, an extra deprotection step would then be required. However, obtaining a high-resolution purification between the protected Cys protein and deprotected Cys protein in the final full-length product would likely be extremely challenging within the context of a 352-aa protein (i.e., ~100 Da change within a ~42,000 Da protein or ~0.2% difference). Moreover, this deprotection step adds an extra handling step at the end of a major synthesis project—an undesirable prospect. In the second approach (**Fig 5.18 C**), the C31 position is incorporated as a terminal orthogonally-protected Cys. The protein would then be assembled maintaining this orthogonally-protected Cys C31. After desulfurizing all Cys that are C-terminal to C31, a separate handling step to deprotect C31 would be required. This step would then be followed by ligation of a short 30-aa peptide that is N-terminal to C31. In addition to adding two extra handling steps, this approach would require a high-resolution purification between 362-aa and 332-aa proteins, which would also likely be challenging. In summary, both “one Cys” approaches are limiting for three reasons:

(1) An additional Cys deprotection step, associated purification, and loss in yield would be required,

(2) Very challenging, hard-to-resolve purifications would be introduced, and

(3) These extra steps decrease the overall convergency of the approach.

Thus, we reasoned that mutation of C31 to any other residue would greatly simplify the overall synthesis strategy. To consider this possibility, we

employ DOPPEL.

To evaluate potential changes at position C31, we first performed a BLAST search of the top 500 sequences in the NCBI “Reference Proteins (refseq)” database. Intriguingly, selection of the Reference Proteins Database greatly reduced redundancy compared to the default database entry: “Non-redundant protein sequences (nr)”. The top 500 hits were aligned using the NIH COBALT tool[43] and then exported to Jalview[44] for further analysis. Within Jalview, a redundancy filter was applied at 95%, which reduced the number of hits to 312. These data were then imported into the Weblogo program[45] for presenting sequence alignment data. Based on these data, Gly was surprisingly favored at position 31 (**Fig 5.19 A**). Investigation of the crystal structure produced two interesting results: first, C31 is found within a beta-sheet structure and does not possess any direct polar contacts with neighboring residues (**Fig 5.19 B**), and second, it is relatively buried within neighboring hydrophobic residues. Finally, a preliminary investigation via Rosetta simulation ranked Phe as the least disruptive mutation at this position. Based on these data, we evaluated three different single mutants: C31F (hydrophobic change supported by crystal structure and Rosetta), C31S (classic isosteric mutation), and C31A (another isosteric mutation). The C31G mutant was not prepared due to its likely disruptive effect within a beta-sheet structure.

During preliminary studies using DNA extension assays, we found that the C31F mutant was the most disruptive mutant (data not shown), and it was not further studied. The other two mutants, C31S and C31A, were then evaluated

using gentle PCR conditions (5 seconds denaturation at 84, 84.9, 86, or 87.2 °C, 30 seconds annealing at 56°, and 1 minute extension at 60°, for 25 cycles). **Fig 5.19 C, D** shows the less disruptive effect of C31A compared to C31S. Thus, C31A was selected as the favored mutation at this position.

5.3.5 (3) N-terminal Half: I101-Inspired Precipitation

The assembly of the N-terminal half of Dpo4 can be performed in an N->C (**Fig 5.20 A**) or C->N (**Fig 5.20 B**) direction. Our initial attempt proceeded via the N->C pathway. Here, peptides [1-2], [3], and [4] were prepared as C-terminal peptide hydrazides, while peptides [3] and [4] were equipped with N-terminal Cys. During this initial trial, we found that the [1-2] plus [3] ligation proceeded in a straightforward manner, with minimal difficulties associated with either the ligation kinetics or the purification step. However, we surprisingly encountered a severe aggregation problem when attempting to ligate [1-3] to [4].

This difficult ligation involves a poorly reactive Ile thioester[32], and the reaction kinetics we observed were therefore very slow, requiring overnight reactions (>12 h) to observe significant product (>30%). However, during the reaction, significant and irreversible precipitation occurred despite the denaturing conditions (6 M GuHCl). Unfortunately, this precipitation could not be rescued: DMF, 8 M GuHCl, water/acetonitrile, nor elevated temperature (~50 °C) had a significant effect. Furthermore, based on the reaction input, we found that the precipitation contained both ligation product [1-4] and reactant [4], whereas most of the unreacted [1-3] remained in solution. This difficulty forced us to reconsider

our assembly strategy.

In response to the N->C challenge, we tested a C->N direction assembly strategy (**Fig 5.20 B**). Here, some issues with protecting group compatibilities must first be introduced. Although the peptide hydrazide method is a significant advance for accessing thioesters via Fmoc-SPPS[46], it is incompatible with the most common N-terminal protecting group, Thz (Thiazolidine)[47]. Unfortunately, an easy-to-use, high yield, and commercially viable alternative to Thz, which is compatible with hydrazides has not yet been established, although many alternatives are under development[38c, 48]. The incompatibility of Thz with hydrazides originates at the activation step (**Fig 5.20 C**); here, the $NaNO_2$-mediated activation converts a hydrazide ($-NHNH_2$) into a reactive acyl azide ($-N_3$) intermediate that can then be displaced by a thiol to generate the active thioester for NCL. However, this activation procedure disables the Thz protecting group by converting it into a Cys and other unidentified byproducts (data not shown). Thus, an alternative protection scheme must be used for N-terminal protection in combination with peptide hydrazides for N->C ligations.

However, it is important to note that peptide hydrazides are compatible (**Fig 5.20 D)** with Thz-ring opening conditions (conversion of Thz -> Cys), via methoxyamine hydrochloride treatment ($MeONH_2$). Thus, peptide hydrazides can be used for C->N ligations when used in combination with other thioester methods.

Therefore, we used Dawson's Dbz method[49] (**Fig 5.20 E**) for preparing peptide [3] as a thioester. Specifically, peptide [3] was assembled on Dawson

Dbz NovaSyn TGF resin (Novabiochem, 0.22 mmol/g) to prepare a C-terminal Dbz (diaminobenzoate) on-resin. Next, the Dbz was activated by addition of p-nitrophenyl chloroformate and base to generate an Nbz product (N-acyl-benzimidazolinone). Following peptide cleavage, the [3]-Nbz peptide was purified and used for subsequent conversion to a thioester. With purified peptides [3]-Nbz and [4] in hand, we again attempted to prepare the N-terminal half. Unfortunately, this reaction was again very slow (incomplete after >18 h at 37 °C using 2-3 mM peptide concentrations), and precipitation became visually obvious under ligation conditions in 6 M GuHCl. Due to the slow reaction kinetics and poor peptide solubility, we were unable to isolate a usable amount of product.

Based on the results using both N->C and C->N assembly strategies, the synthetic difficulty of the N-terminal half is likely due to two compounding factors:

(1) Slow ligation kinetics at the Ile junction between peptides [3] and [4],

(2) The aggregation propensity of peptide [4] under ligation conditions.

We reasoned that this challenge offers another case for DOPPEL. The aggregation problem could be solved by simply changing the ligation junction residue (I101) to a more reactive thioester. The application of a more reactive thioester will accelerate NCL kinetics and minimize aggregation simply by reducing the reaction time.

To evaluate potential changes at position I101, we utilized the sequence alignment data and crystal structure data from the C31 analysis (above). For I101, we found via alignment that Pro was the most statistically favored residue (**Fig 5.21 A**). This substitution would not be helpful, as Pro is the least reactive

thioester[32]. After further evaluating the crystal structure, we found that I101 was located within a hydrophobic pocket along a beta-sheet (**Fig 5.21 B, C**). Based on this information, we decided to make two conservative mutations: I101M and I101F, both of which would significantly increase thioester reactivity.

Preliminary DNA extension data led us to remove I101F from consideration, as it was very disruptive. In contrast, I101M was found to be minimally disruptive compared to WT (**Fig 5.21 D)** as demonstrated in a PCR experiment (5 sec denaturation at 84, 84.9, 86, or 87.2 C, 30 sec annealing at 56, and 1 min extension at 60, for 25 cycles). Thus, I101M was selected at this favored mutation at this position. Preliminary testing of with a [3]-Met thioester peptide confirmed accelerated reactivity and significantly less precipitation compared to [3]-Ile (data not shown).

At this stage, two mutations have been identified, using DOPPEL, to enhance the Dpo4 synthesis: C31A/I101M. **Fig 5.22** shows the activity testing data on this new double mutant, confirming minimal effect on activity under gentle PCR conditions.

5.3.6 (4) C-terminal Half: Desulfurization Dilemmas

The success of our Dpo4 design hinges on our ability to perform global desulfurization at the last step of the synthesis. In this case, we will be converting five Cys to Ala within a >300 residue protein. However, Brik's group reported[24] significant challenges in desulfurizing seven Cys in their final synthetic 304-aa tetraubiquitin. Ultimately, they had to redesign their synthesis strategy to include

intermediate desulfurization steps. For our Dpo4 synthesis, we would like to avoid this problem. After expending considerable effort preparing the full-length construct, it would be extremely disappointing not to be able to desulfurize the final ligation junctions. Thus, in order to investigate this potential problem before consuming precious synthetic material, we recombinantly engineered artificial constructs of Dpo4 with varying numbers of Cys residues to predict the degree of desulfurization difficulty. Here, three artificial Dpo4 constructs were prepared, with the Cys locations indicated:

- Multi-Cys construct: 31, 57, 102, 121, 155, 222, 283;
- N-Cys construct: 31, 57, 102, 121; and
- C-Cys construct: 31, 155, 222, 283.

Multi-Cys was generated to create the most challenging desulfurization construct, while N-Cys and C-Cys were prepared to localize any challenging Cys to either the N- or C-terminal halves of the full-length protein. From our preliminary testing, we found the desulfurization of seven Cys to be highly challenging, whereas desulfurization of either three or four Cys residues was feasible. Although we have concluded that two desulfurizations should be employed to favor success in our total Dpo4 synthesis, further studies are underway. More broadly (outside of the current Dpo4 project), these constructs will be used to inform the most ideal conditions for performing desulfurizations, as previous cases in the Kay lab have been relatively simple: one or two Cys per short peptide (<150 aa).

5.3.7 (5) C-terminal Half: Taking a Step Back in Ligations

Similar to the N-terminal half, the C-terminal half can be assembled in two different ways: N->C (**Fig 5.23 A**) or C->N (**Fig 5.23 B**). Although this choice may seem inconsequential, it ultimately has a significant effect on the overall yield of the C-terminal half—and consequently the entire protein

Using peptide hydrazides, the N->C assembly of this half must be performed using a Cys(ACM) protection at the N-terminus (**Fig 5.23 A**). Unfortunately, this approach would require at least three synthetic steps, with the last step being silver-mediated ACM deprotection, which we have found to be highly challenging (and low yielding) when handling long poorly-soluble synthetic intermediates. In contrast: note that in the synthesis of the N-terminal half described above (**Fig 5.20**), no protection is needed for the unreactive (non-Cys) N-terminus.

On the other hand, the C-terminus could also (theoretically) be assembled using a method that does not include peptide hydrazides, but rather exploits the benefits of a traditional NCL approach using Thz-protection. Here, Thz-peptide thioesters could be used to prepare [5-6] and [7-8]. Then, peptide [8-9] can be prepared with N-terminal Cys and C-terminal amide/acid. Using this approach, the C-terminal half could be assembled in either one or two steps (**Fig 5.23 D, E**)—with both strategies avoiding ACM deprotection. This is an exciting prospect from two perspectives:

(1) The overall efficiency of the synthesis would be greatly improved,

(2) This elegant approach entails an appealing combination of classical and

modern thioester methods. The classical thioester assembly (C->N) would be used to assemble the C-terminal half, while the modern peptide hydrazide method (N->C) would be used to assemble the N-terminal half.

Based on this potentially improved synthesis strategy, we decided to prepare peptides [5-6] and [7-8] as Thz-peptide thioesters using the older Dawson Dbz method[49]. **Fig 5.24A** shows the crude HPLC and MS data on peptide [Thz-5-6-Nbz], while **Fig 5.24B** shows data for the peptide hydrazide [Cys(ACM)-5-6- $NHNH_2$].

Unfortunately, the quality of the Nbz-produced peptide is clearly worse than the peptide hydrazide. The [Cys(ACM)-5-6-$NHNH_2$] peptide shows a clearly isolated peak, whereas the [Thz-5-6-Nbz] peptide is a complex mixture that will be highly challenging to purify. **Fig 5.24C** and **5.24D** show the crude HPLC and MS data for the [Thz-7-8-Nbz] and [7-8-$NHNH_2$] peptides. In this case, the decreased quality with the Nbz peptide is even more pronounced, with a significant -1997 Da byproduct appearing as the main species. Most of the byproducts observed with the Dawson Dbz method were one of three types: (1) inadequate conversion of Dbz->Nbz; (2) nonspecific peptide modification with p-nitrophenyl chloroformation; or (3) impurities likely due to branching at the Dbz core. Based on lower quality of the Nbz-peptides, we postponed this approach Thus, our final approach will maintain the ACM-based strategy (**Fig 5.25,** data presented below).

5.3.8 Final Strategy and Pilot Synthesis of Dpo4

Based on all of the scouting data and DOPPEL work described above, we identified our synthesis strategy and DOPPEL changes:

- Six-piece, convergent assembly strategy (Fig 5.25)
- C31A and I101M DOPPEL changes
- Introduction of an N-terminal polyLys sequence, cleavable with TEV protease

We next pursued a pilot total chemical synthesis of Dpo4 using this strategy. Due to the convergency of this approach, we were able to independently assemble the two protein halves. First, peptide [1-2] was reacted [3] generating [1-3] (**Fig 5.26 A**), which was then reacted with [4] to generate the N-terminal half: [1-4] (**Fig 5.26 B**). Introduction of the I101M mutation was essential to assembling the N-terminal half, as the [1-3] plus [4] ligation was complete after only 2 h. The N-terminal half was then desulfurized at positions C57 and C102 to generate [1-4]-desulfurized (**Fig 5.26 C**).

For the C-terminal half of the protein, [ACM-5-6] was ligated to [7-8] to generate [ACM-5-8] (**Fig 5.27 A**), which was then reacted with [8-9] to generate the C-terminal half of the protein: [ACM-5-9] (**Fig 5.26 B**). Removal of the ACM protecting group produced the ligation-competent [5-9] (**Fig 5.26 C**).

With both halves purified (see **Fig 5.28 A, B** for closer analysis of the Mass Spec data), the final, full-length ligation could be performed. In this case, very limited material was available (<1 mg of each half); nevertheless, the final ligation showed traces of full-length product by MS (**Fig 5.28 C**). Unfortunately,

insufficient material was obtained for further testing due to limited starting material in this pilot synthesis.

5.4 Future Plans

Push Forward the Dpo4 Synthesis: The Dpo4 synthesis will be scaled-up in order to produce usable quantities of functional protein. Some of this work has been performed as batches of peptides in the C-terminal have already been synthesized. In parallel with the synthesis scale-up, optimized work-up conditions for the full-length recombinant product will be performed. Here, recombinant Dpo4 will be dissolved in ligation/desulfurization buffer (6 M GuHCl), and different conditions will be evaluated for improving the yields of purification via HPLC. The transition from 6 M GuHCl to HPLC conditions (0.1% TFA in water/acetonitrile) has proven to be a difficult, low-yield step of the assembly process. This can be optimized using more recombinant material.

Further Optimize the Dpo4 Synthesis: Two new Thz-based protecting groups are under development for simplifying the assembly of the C-terminal half ([5-9]). These groups were designed to be peptide hydrazide compatible. See **Fig 5.29** for details. Here, we found that Meoc-Thz-OH (a simpler version of the published Tbeoc-Thz-OH[47]) reagent was stable to peptide hydrazide activation. We also found that the conversion of Meoc-Thz into Thz was greatly accelerated by adjusting pH to >8, which overcomes a previous limitation of this reagent[50]. We have also begun preliminary investigations into using pNB(*p*-nitrobenzyl)-Thz-OH as an alternative to the Thz-protecting group; this group is theoretically

stable to hydrazide activation and can be selectively reduced to the acid-labile pAB(*p*-aminobenzyl)-Thz-OH (ultimately forming Thz-OH) using reducing conditions[51], such as Pd-based hydrogenation[52] or Zn reduction[53].

5.5 Acknowledgments

Dave Kemble, Frank Whitby, Katherine Ferrell, Vish Chandrasekaran (University of Utah) for advice and help with cloning and protein expression. Katherine Ferrell for providing TEV stocks.

5.6 Materials and Methods

5.6.1 Peptide Characterization

All peptides were prepared on a Prelude commercial peptide synthesizer (Protein Technologies, Inc.) Peptides with C-terminal amides were prepared on TentaGel R RAM resin (0.19 mmol/g, RAPP Polymere). Peptides with C-terminal hydrazides were prepared on in-house modified hydrazine-chlorotrityl resin based on published protocol[46b]. Peptides with C-terminal Dbz/Nbz on Dawson Dbz NovaSyn TGF resin (0.22 mmol/g, Novabiochem).

Peptides were analyzed on the following columns (from Phenomenex):

- Jupiter 5u C4 300A, 4.6 x 150 mm
- Jupiter 4u C12 Proteo 90A, 4.6 x 150 mm
- Luna 5u C18(2) 100 A, 4.6 x 250 mm
- Aeris WIDEPORE C4 3.6u, 2.1 x 50 mm

Peptides (following ligations, desulfurizations, and ACM-deprotections) were

purified on the following semiprep-scale columns (from Phenomenex):

- Jupiter 5u C4 300A, 10 x 250 mm
- Jupiter 10u C4 300A, 10 x 250 mm
- Jupiter 4u C12 Proteo 90A, 10 x 250 mm

Crude peptides were purified on the following prep-scale columns (from Phenomenex):

- Jupiter 10u C4 300A, 21.2 x 250 mm
- Jupiter 4u C12 Proteo 90A, 21.2 x 250 mm

Mass spectrometry characterization of peptides was performed using an AB Sciex 3000 LC/MS/MS system.

5.6.2 Native Chemical Ligation and Desulfurization

NCL reactions were performed according to standard protocols in the field[54], with specific adjustments for peptide hydrazides[46-47, 55]. Desulfurization reactions were performed according to the established free-radical protocol[56], using reduced glutathione as the thiol scavenger.

5.6.3 Cloning and Protein Expression

Dpo4 *S. solfataricus* sequence (WP_009993137.1) was cloned into pEXP5.CT vector. Mutant versions of Dpo4 were produced either by Quickchange strategy or ‘round-the-horn’ PCR methodology. All mutants were confirmed by sequencing using T7 primers. Dpo4 was expressed using either BL21(DE3) pLysS (all constructs except for TEV-containing constructs) or BL21-

CodonPlus (DE3)-RIPL cells (for TEV-containing constructs). Proteins were expressed by autoinduction method. After overnight expression, cells were pelleted and then lysed with lysosome and sonication. The lysed material (soluble fraction) was collected by centrifugation and applied to Ni-NTA beads for purification, followed by overnight dialysis (4 °C) into 50 mM Tris pH 8, 50 mM NaCl, 1 mM DTT. After dialysis, samples were purified by HPLC on a Phenomenex C4 column and then lyophilized.

5.6.4 Refolding Protocol

Lyophilized Dpo4 powder was suspended in denaturation buffer (6 M GuHCl, 50 mM Tris pH 8, 50 mM NaCl) to ~25 μM and then dialyzed overnight at 4 °C, using Slide-a-Lyzer dialysis cassettes 3,500 MWCO, into the same buffer without GuHCl. Sampels were then spun to remove any precipitation and protein concentration was measured using an extinction coefficient of 19,200 cm^{-1} M^{-1}.

5.6.5 PCR Protocol

PCR activity assays (for comparing activity of our Dpo4 recombinant constructs) was performed on a Biorad C1000 Touch Thermal Cycler. For each 25 μl reaction, 140 ng of M13KE plasmid template DNA and 1 μM concentrations of two primers:

- M13 1470: cgcaactatcggtatcaagc
- -96 gIII: ccctcatagttagcgtaacg.

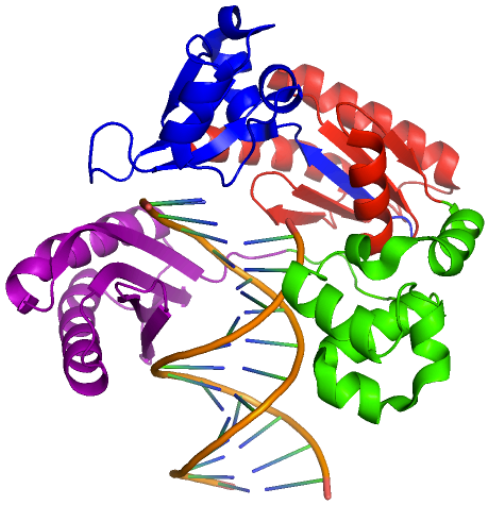


Figure 5.1: Dpo4 crystal structure. Crystal structure shows the four key polymerase domains: thumb (green), palm (red), finger (blue), and little finger (purple). Image generated using PyMol (1JXL.pdb[1]).

A

```
        10         20         30         40         50         60         70
MIVLFVDFDY FYAQVEEVLN PSLKGKPVVV AVFSGRFEDS GAVATANYEA RKFGVKAGIP IVEAKKILPN
        80         90        100        110        120        130        140
AVYLPMRKEV YQQVSSRIMN LLREYSEKIE IASIDEAYLD ISDKVRDYRE AYNLGLEIKN KILEKEKITV
       150        160        170        180        190        200        210
TVGISKNKVF AKIAADMAKP NGIKVIDDEE VKRLIRELDI ADVPGIGNIT AEKLKKLGIN KLVDTLSIEF
       220        230        240        250        260        270        280
DKLKGMIGEA KAKYLISLAR DEYNEPIRTR VRKSIGRIVT MKRNSRNLEE IKPYLFRAIE ESYYKLDKRI
       290        300        310        320        330        340        350
PKAIHVVAVT EDLDIVSRGR TFPHGISKET AYSESVKLLQ KILEEDERKI RRIGVRFSKF IEAIGLDKFF
352
 DT
```

B

Segment	Length	Sequence
[s1]	31	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVVA
[s2]	40	VFSGRFEDSGAVATANYEARKFGVKAGIPIVEAKKILPNA
[s3]	49	VYLPMRKEVYQQVSSRIMNLLREYSEKIEIASIDEAYLDISDKVRDYRE
[s4]	44	AYNLGLEIKNKILEKEKITVTVGISKNKVFAKIAADMAKPNGIK
[s5]	38	VIDDEEVKRLIRELDIADVPGIGNITAEKLKKLGINKL
[s6]	26	VDTLSIEFDKLKGMIGEAKAKYLISL
[s7]	39	ARDEYNEPIRTRVRKSIGRIVTMKRNSRNLEEIKPYLFR
[s8]	43	AIEESYYKLDKRIPKAIHVVAVTEDLDIVSRGRTFPHGISKET
[s9]	42	AYSESVKLLQKILEEDERKIRRIGVRFSKFIEAIGLDKFFDT

Figure 5.2: Initial breakdown of the 352-residue Dpo4. (**A**) Protein sequence of *Sulfolobus solfataricus* Dpo4, NCBI reference: NP_343798.1 (**B**) Breakdown of the protein sequence into nine peptides for exploratory synthesis and scouting of peptide properties, with segment name and peptide length indicated.

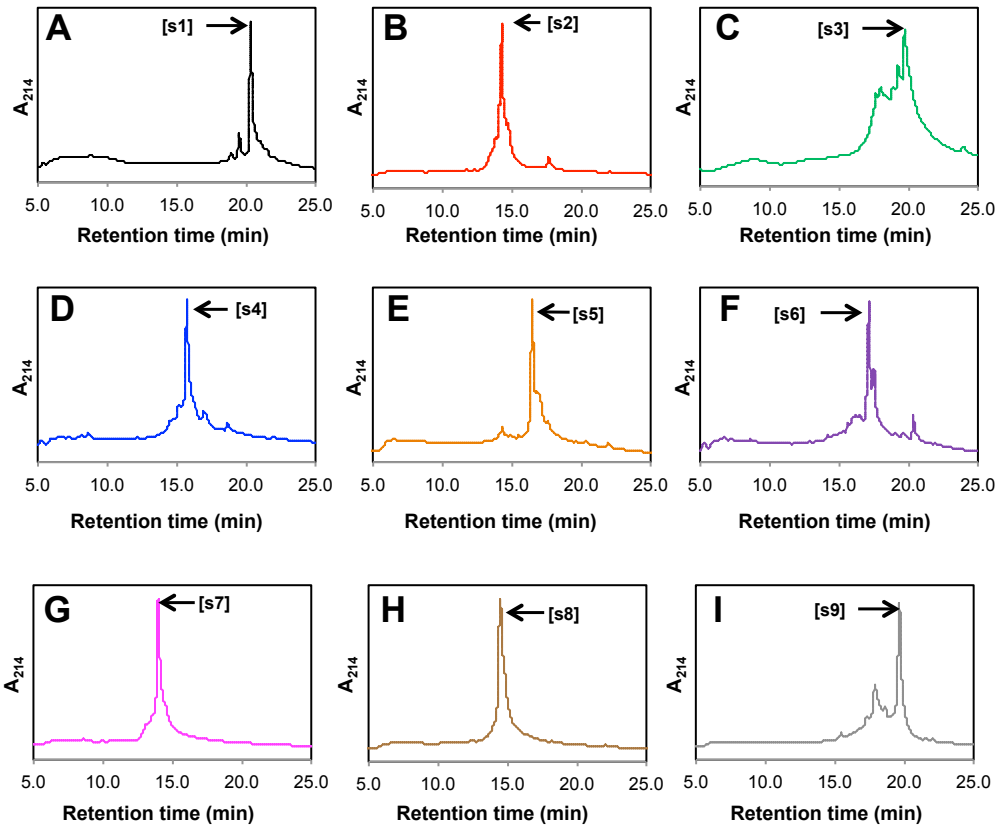
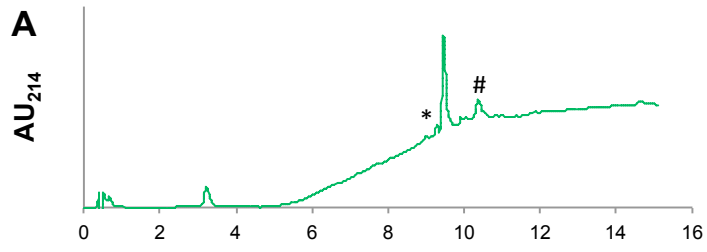


Figure 5.3: HPLC of the initial nine peptides. (**A**) – (**I**) Crude HPLC traces of the nine initial scouting peptides covering the Dpo4 sequence. Samples were analyzed on a C12-Phenomenex column (4.6 x 150 mm).

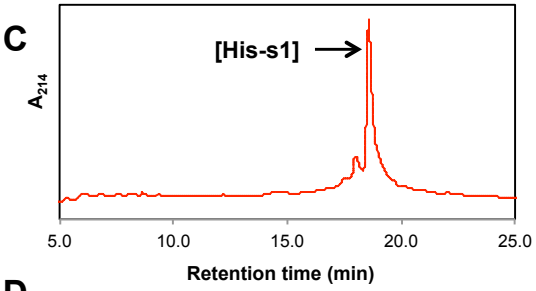
Table 5.1: Mass spec summary of the initial nine peptides. Mass Spec of the initial nine peptides, including the observed (deconvoluted) masses and major byproducts for each peptide. Sites of pseudoproline dipeptides are indicated in bold and underline.

Segment	Length	Full Peptide Sequence	Calc Mass (Da)	Obs Mass (Da)	Major Byproducts
[s1]	31	MIVLFVDFDYFYAQVEEVLNPSLKGKPVVV A-NHNH2	3545.1	3545.2	3431.5 (-Ile/Leu), 3413.8 (-Met), 3446.2 (-Val)
[s2]	40	PenFSGRFEDSGAV**AT**ANYEARKFGVKAGI PIVEAKKILPNA-NHNH2	4268.8	4267.9	4136.7 (-Pen), 4307.5 (+40), 4490.5 (+Fmoc)
[s3]	49	PenYLPMRKEVYQQ**VS**SRIMNLLREYSEKI EI**AS**IDEAYLDISDKVRDYRE-NHNH2	6001.8	6002.5	5889.3 (-Ile/Leu), 5986.1 (aspartimide), 5872.6 (-Ile/Leu + aspartimide)
[s4]	44	CYNLGLEIKNKILEKEKIT**VT**VGISKNKVFAKI AADMAKPNGIK-NHNH2	4846.8	4846.7	4733.1 (-Ile/Leu), 4718.4 (-Glu), 2456.5 (truncation)
[s5]	38	PenIDDEEVKRLIRELDIADVPGIGNITAEKL KKLGINKL-NHNH2	4275.9	4275.1	4315.1 (+40 Da), 4300.7 (+26 Da), 4260 (aspartimide)
[s6]	26	PenDTLSIEFDKLKGMIGEAKAKYLISL-NHNH2	2929.4	2928.8	2798.2 (-Pen), 2968.9 (+40 Da), 2839.9 (-90 Da)
[s7]	39	CRDEYNEPIRTRVR**KS**IGRI**VT**MKRNSRNL EEIKPYLFR-NHNH2	4852.6	4852.1	4948.1 (TFA adduct), 2889.9 (Truncation), 4909.6 (OtBu adduct)
[s8]	43	CIEESYYKLDKRIPKAIHVVA**VT**EDLDI**VS**RG RTFPHGISKET-NHNH2	4929.6	4928.9	4968.9 (+40 Da), 4852.2 (-77 Da), 4913.1 (aspartimide)
[s9]	42	CYSESVKLLQKILEEDERKIRRIGVRFSKFIE AIGLDKFFDT-NH2	5020.8	5020.7	4907.5 (-Ile/Leu), 4892.3 (-Gln/Lys), 4921 (-Val)



B

Segment	Length	Full Peptide Sequence	Calc Mass (Da)	Obs Mass (Da)	Major Byproducts
[ASIDE]	19	CSIDEAYLDISDKVRDYRE-NHNH2	2304.5	2303.8	2369.3 (Piperidine adduct), 2285.9 (aspartimide)



D

Segment	Length	Full Peptide Sequence	Calc Mass (Da)	Obs Mass (Da)	Major Byproducts
[His-s1]	37	HHHHHHMIVLFVDFDYFYAQVEEVLNPSLK GKPVVVA-NHNH2	4368.0	4367.7	4254.2 (-Ile/Leu/Asn), 4230.7 (-His)

Figure 5.4: Analysis of two additional scouting peptides. (**A**) HPLC analysis of crude [ASIDE] peptide; (**B**) Mass Spec summary of [ASIDE] peptide; (**C**) HPLC analysis of [His-s1]; (**D**) Mass Spec summary of [His-s1] peptide. * and # indicate -18 Da aspartimide and +67 Da piperidine adduct byproducts. [ASIDE] was analyzed on C4-Aeris column (2.1 x 50 mm) and [His-s1] was analyzed on C12-Phenomenex column (4.6 x 150 mm).

A

```
         2         12         22         32         42         52         62
MHHHHHHGGI VLFVDFDYFY AQVEEVLNPS LKGKPVVVCV FSGRFEDSGA VATANYEARK FGVKAGIPIV
        72         82         92        102        112        122        132
EAKKILPNAV YLPMRKEVYQ QVSSRIMNLL REYSEKIEIA SIDEAYLDIS DKVRDYREAY NLGLEIKNKI
       142        152        162        172        182        192        202
LEKEKITVTV GISKNKVFAK IAADMAKPNG IKVIDDEEVK RLIRELDIAD VPGIGNITAE KLKKLGINKL
       212        222        232        242        252        262        272
VDTLSIEFDK LKGMIGEAKA KYLISLARDE YNEPIRTRVR KSIGRIVTMK RNSRNLEEIK PYLFRAIEES
       282        292        302        312        322        332        342
YYKLDKRIPK AIHVVAVTED LDIVSRGRTF PHGISKETAY SESVKLLQKI LEEDERKIRR IGVRFSKFIE
       352
AIGLDKFFDT GG-NH2
```

B

Segment	Length	Sequence	Pseudoprolines & (Dmb)Gly	Thioester
[1-2]	64	MHHHHHHGGIVLFVDFDYFYAQVEE VLNPSLKGKPVVVCVFSGRFEDSGA VATANYEARKFGVK	(Dmb)G, FS, DS, AT	Lys
[3]	45	AGIPIVEAKKILPNAVYLPMRKEVY QQVSSRIMNLLREYSEKIEI	VS, YS	Ile
[4]	53	ASIDEAYLDISDKVRDYREAYNLGL EIKNKILEKEKITVTVGISKNKVFA KIA	VT	Ala
[5-6]	67	ADMAKPNGIKVIDDEEVKRLIRELD IADVPGIGNITAEKLKKLGINKLVD TLSIEFDKLKGMIGEAK	DT, (Dmb)G	Lys
[7-8]	61	AKYLISLARDEYNEPIRTRVRKSIG RIVTMKRNSRNLEEIKPYLFRAIEE SYYKLDKRIPK	VT, ES	Lys
[8-9]	72	AIHVVAVTEDLDIVSRGRTFPHGIS KETAYSESVKLLQKILEEDERKIRR IGVRFSKFIEAIGLDKFFDTGG	VT, VS, ES, FS, DT	n/a

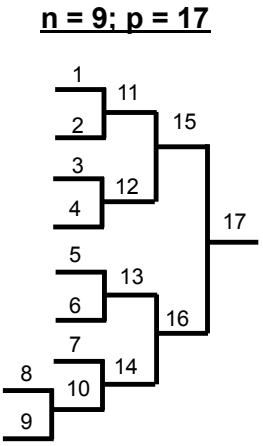
Figure 5.5: Final six-segment breakdown of the 352-residue Dpo4. (**A**) Protein sequence of *Sulfolobus solfataricus* Dpo4, NCBI reference: NP_343798.1; (**B**) Breakdown of the protein sequence into six peptides for the total synthesis. Sites of pseudoproline dipeptides and (Dmb)Gly are indicated in bold and underline.

A

n	p
1	1
2	3
3	5
4	7
5	9
6	11
7	13
8	15
9	17
10	19

n, peptides
p, purifications
p = (n*2) - 1

B



C

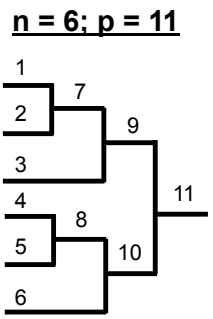


Figure 5.6: Relationship between peptide segments and minimum total number of purifications. (**A**) Table summarizing the relationship between peptides and purifications; (**B**) Optimal assembly strategy starting with nine peptides; (**C**) Optimal assembly strategy starting with six peptides.

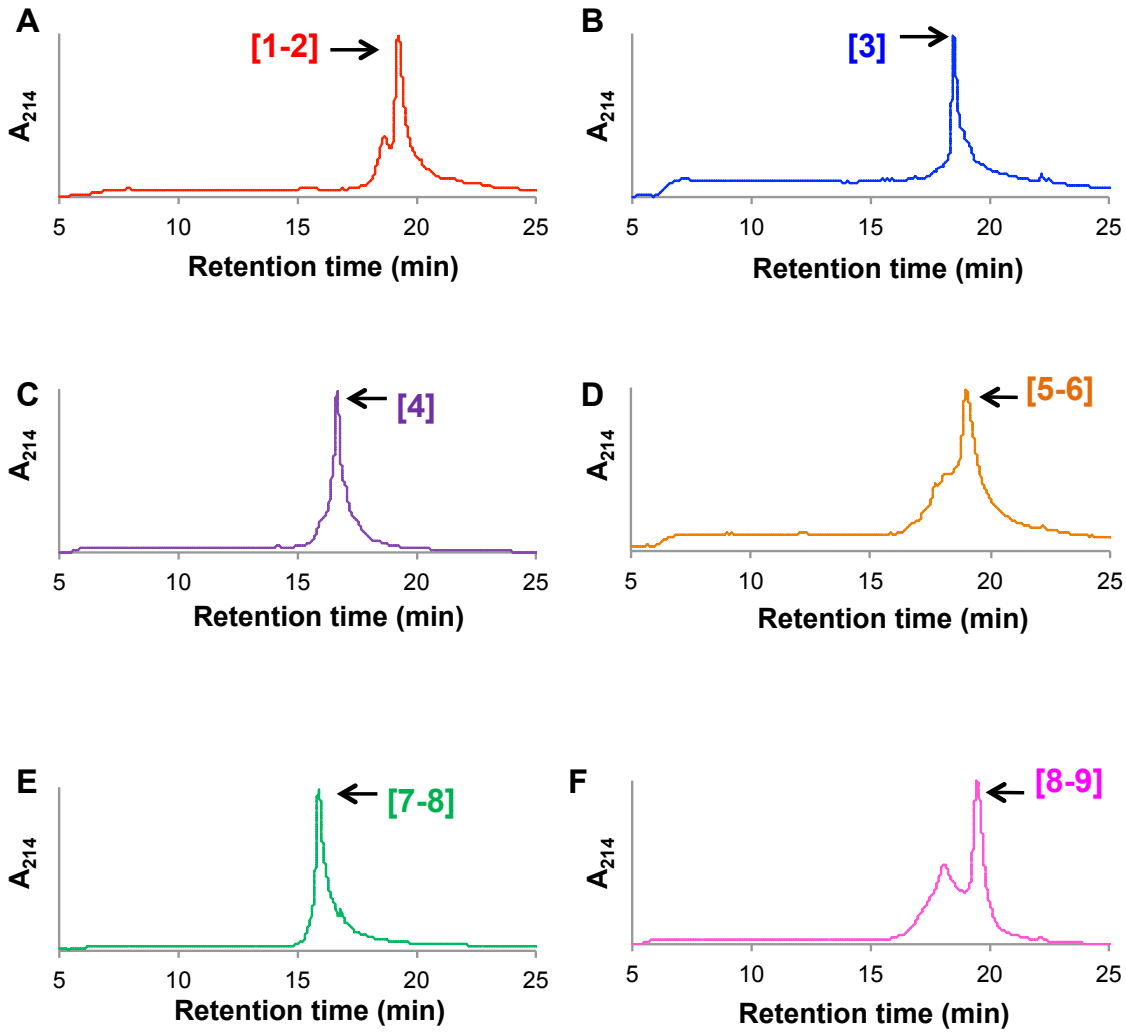


Figure 5.7: HPLC on the final six crude peptides. (**A**) Peptide [1-2]; (**B**) Peptide [3]; (**C**) Peptide [4]; (**D**) Peptide [5-6]; (**E**) Peptide [7-8]; (**F**) Peptide [8-9]. All samples were analyzed on C12-Phenomenex column (4.6 x 150 mm).

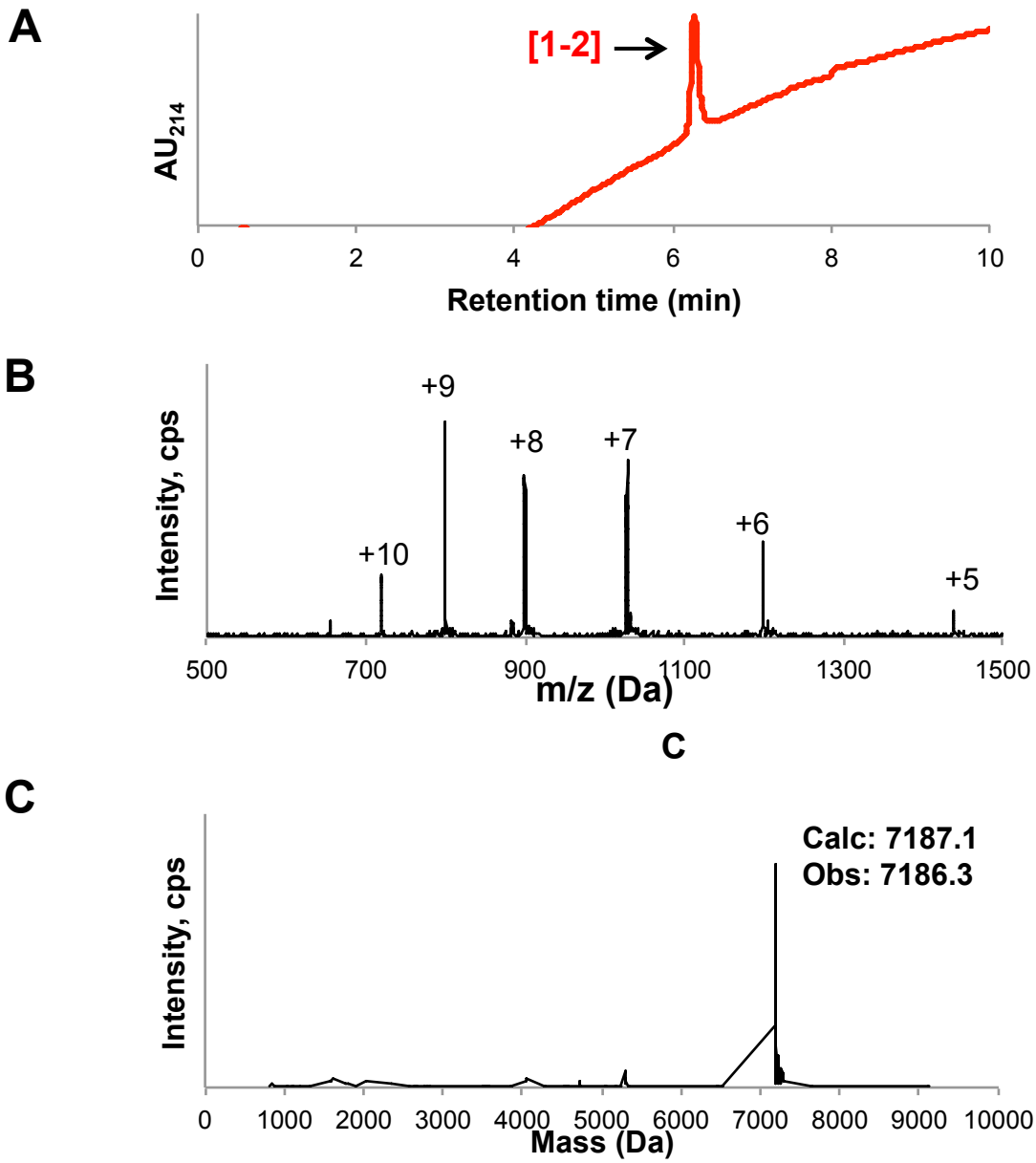


Figure 5.8: Analysis of pure [1-2] peptide. (**A**) HPLC on C4-Aeris column; (**B**) Observed ions (charge states); (**C**) Deconvoluted Spectra. Calculated theoretical mass is based on the peptide sequence, while observed mass is the mass average determined from BioAnalyst software (AB Sciex).

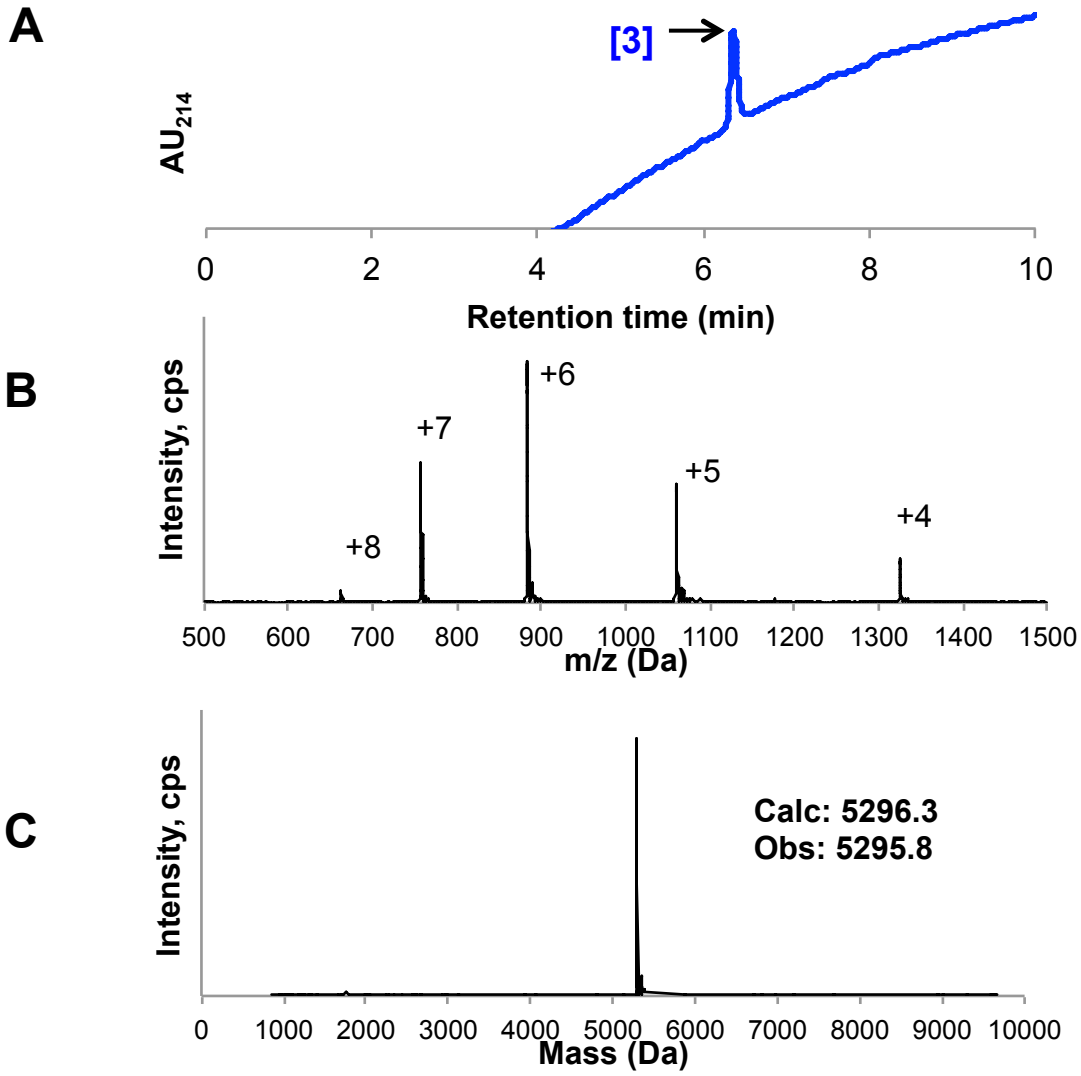


Figure 5.9: Analysis of pure [3] peptide. (**A**) HPLC on C4-Aeris column; (**B**) Observed ions (charge states); (**C**) Deconvoluted Spectra. Calculated theoretical mass is based on the peptide sequence, while observed mass is the mass average determined from BioAnalyst software (AB Sciex).

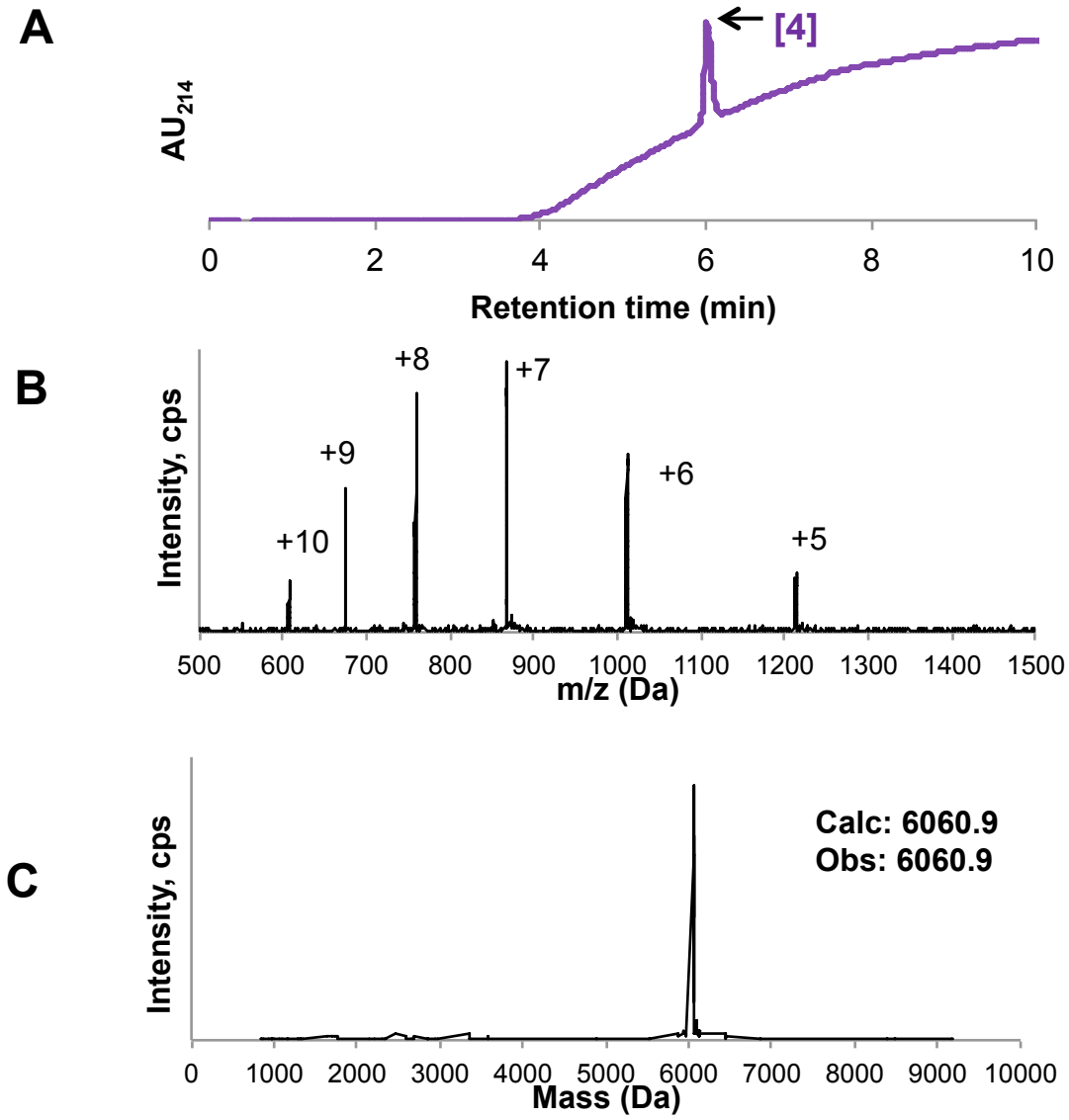


Figure 5.10: Analysis of pure [4] peptide. (**A**) HPLC on C4-Aeris column; (**B**) Observed ions (charge states); (**C**) Deconvoluted Spectra. Calculated theoretical mass is based on the peptide sequence, while observed mass is the mass average determined from BioAnalyst software (AB Sciex).

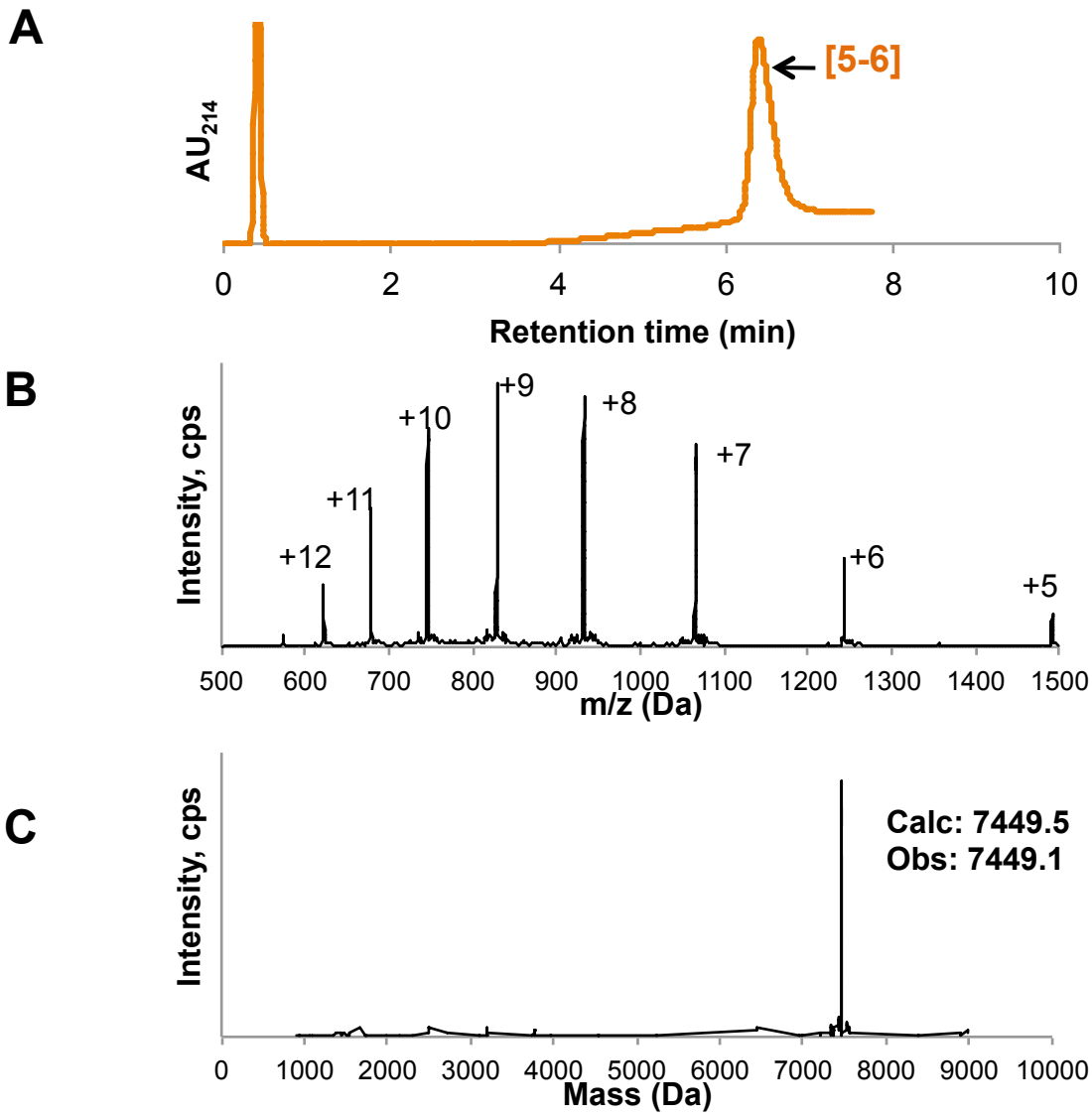


Figure 5.11: Analysis of pure [5-6] peptide. (**A**) HPLC on C4-Aeris column; (**B**) Observed ions (charge states); (**C**) Deconvoluted Spectra. Calculated theoretical mass is based on the peptide sequence, while observed mass is the mass average determined from BioAnalyst software (AB Sciex).

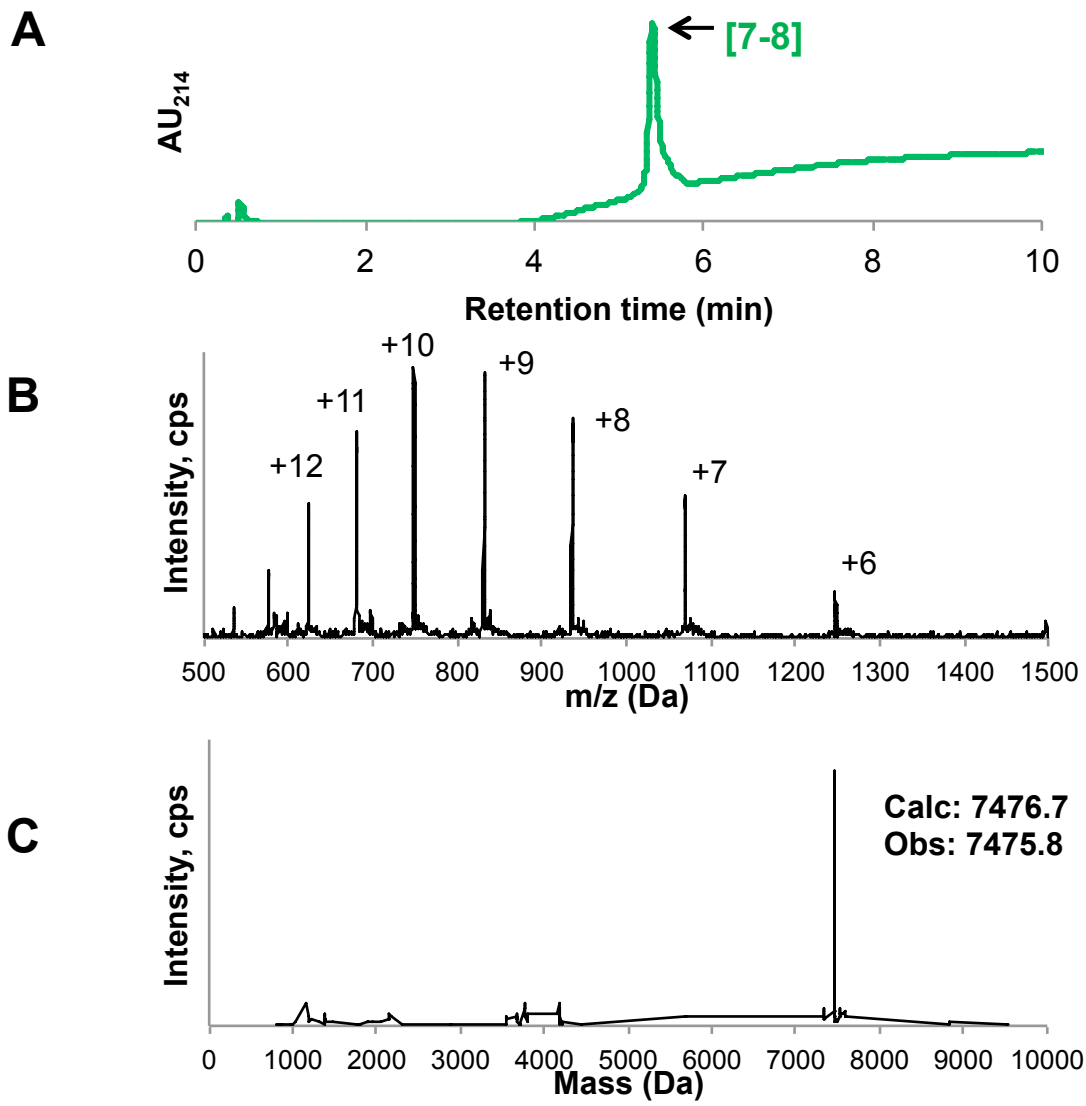


Figure 5.12: Analysis of pure [7-8] peptide. (**A**) HPLC on C4-Aeris column; (**B**) Observed ions (charge states); (**C**) Deconvoluted Spectra. Calculated theoretical mass is based on the peptide sequence, while observed mass is the mass average determined from BioAnalyst software (AB Sciex).

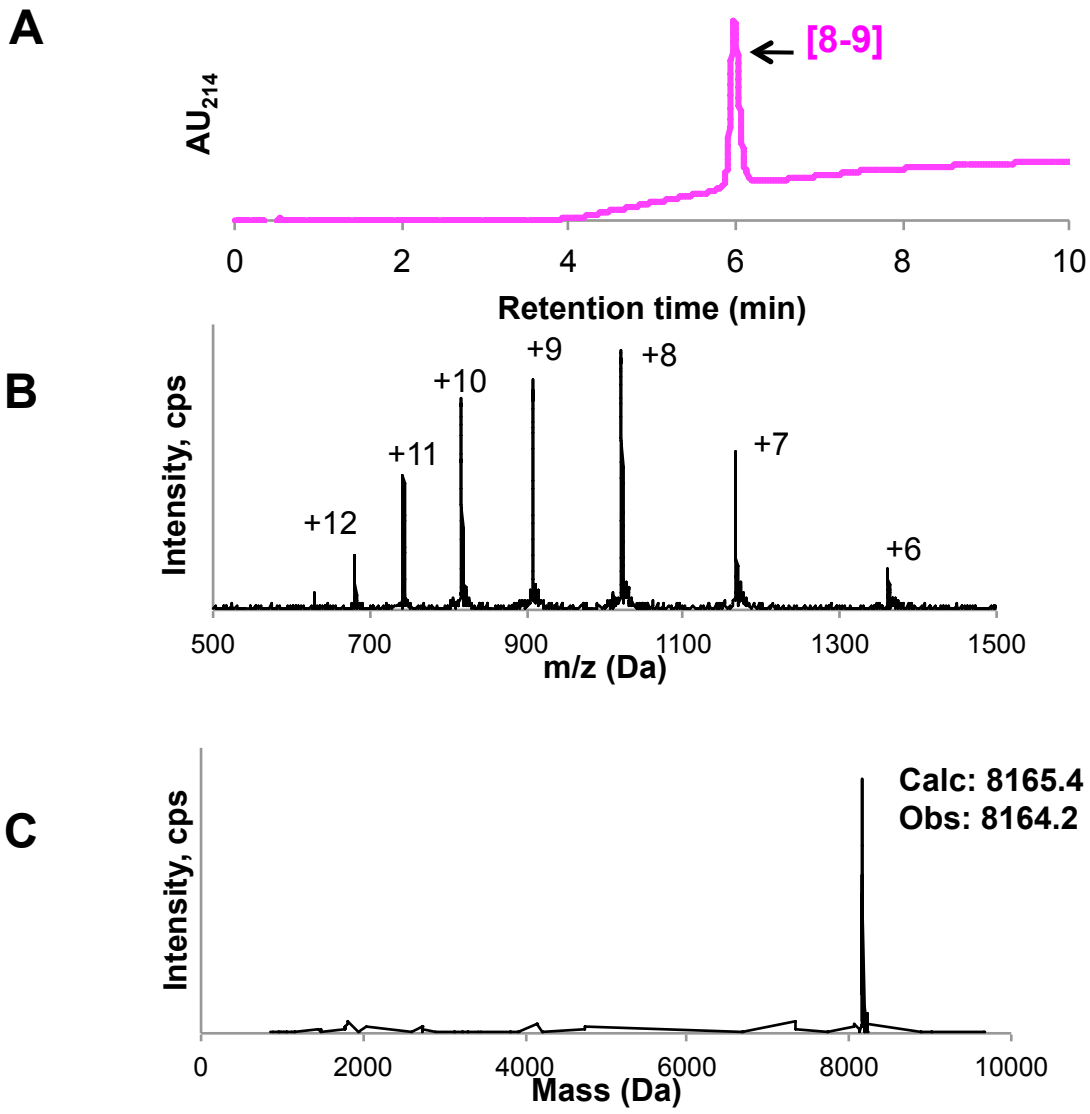


Figure 5.13: Analysis of pure [8-9] peptide. (**A**) HPLC on C4-Aeris column; (**B**) Observed ions (charge states); (**C**) Deconvoluted Spectra. Calculated theoretical mass is based on the peptide sequence, while observed mass is the mass average determined from BioAnalyst software (AB Sciex).

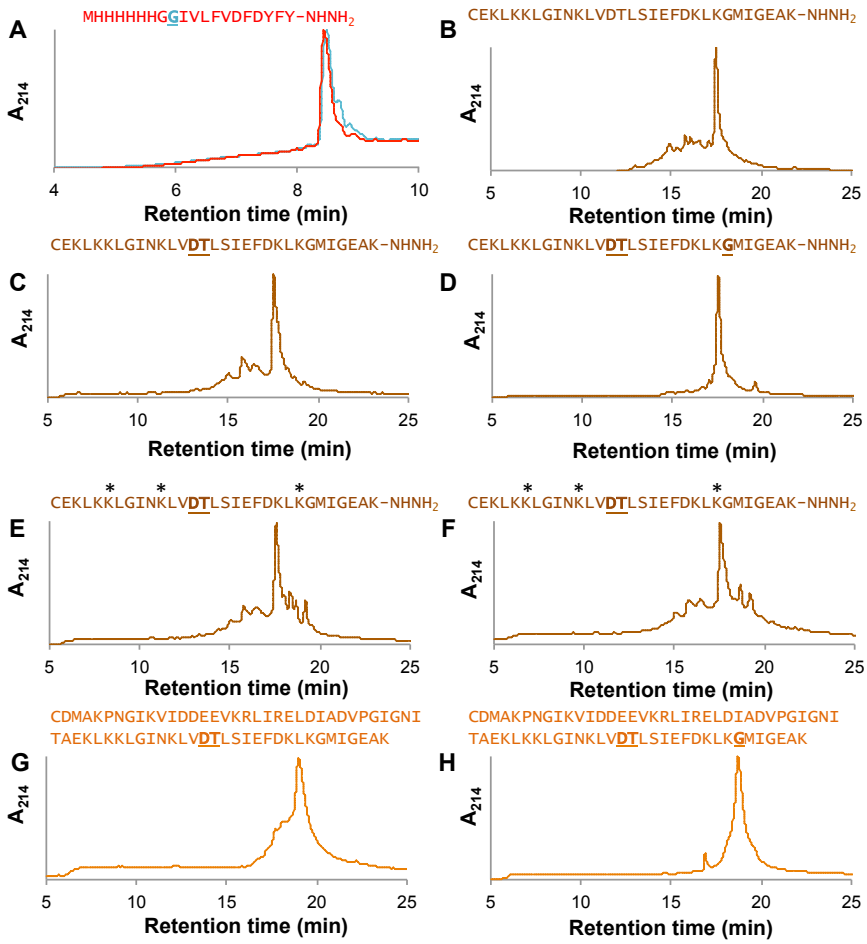


Figure 5.14: Optimization of [1-2] and [5-6] crude peptides. (**A**) [1-2], with (Dmb)Gly position indicated in blue; (**B**) [6]-only; (**C**) [6] with DT-psi; (**D**) [6] with DT-psi and (Dmb)Gly; (**E**) [6] with DT-psi and select Lys(Mtt) incorporation; (**F**) [6] with DT-psi and select Lys(Dde) incorporation; (**G**) [5-6] with DT-psi; (**D**) [5-6] with DT-psi and (Dmb)Gly. HPLC of [1-2] peptide analyzed on C4-Aeris column (2.1 x 50 mm); [6] and [5-6] peptides were analyzed on C12-column (4.6 x 150 mm). * indicates sites of alternative Lys protecting groups.

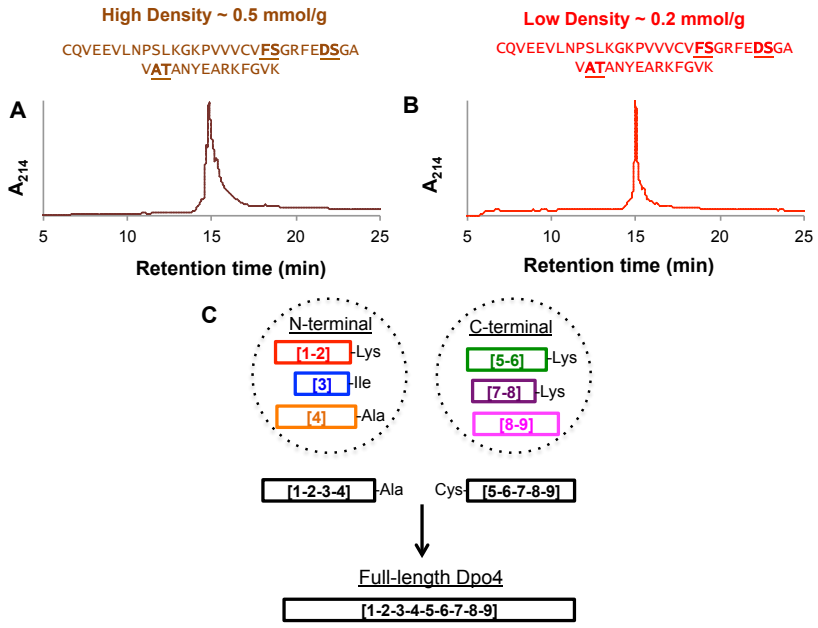


Figure 5.15: Optimization of [2] and convergent assembly. HPLC analysis of (**A**) [2] at 0.5 mmol/g density; and (**B**) [2] at 0.2 mmol/g. Peptides were analyzed on C12-column (4.6 x 150 mm). (**C**) Summary of the convergent assembly strategy based on the N- and C-terminal halves of Dpo4.

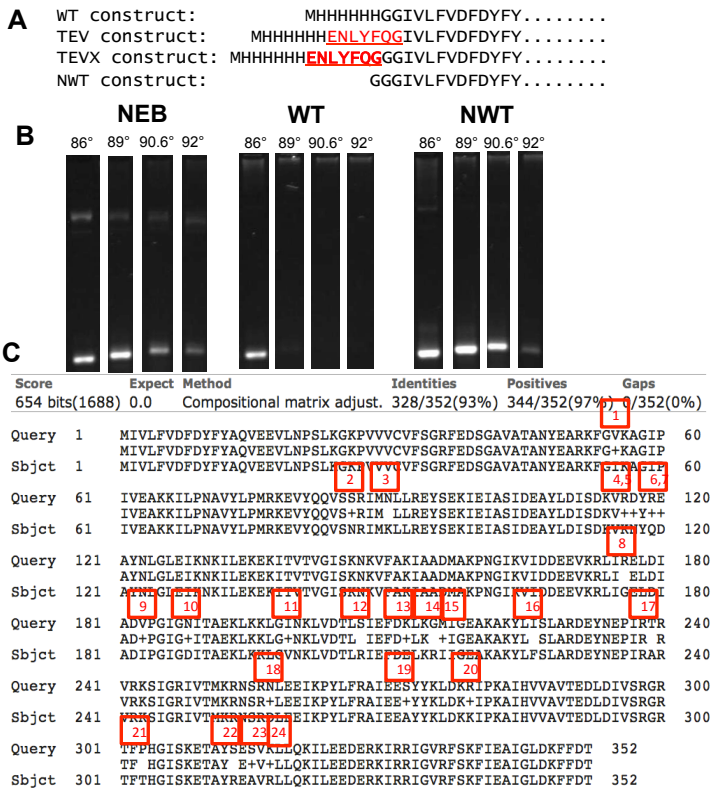


Figure 5.16: Identification of the His tag issue. (**A**) N-terminal sequences of various Dpo4 constructs; (**B**) 1% agarose gels on PCR activity assays; (**C**) Sequence comparison of the WT and NEB samples. Note that the PCR activity data were run on the same gel, but lanes have been repositioned to simplify the data presentation.

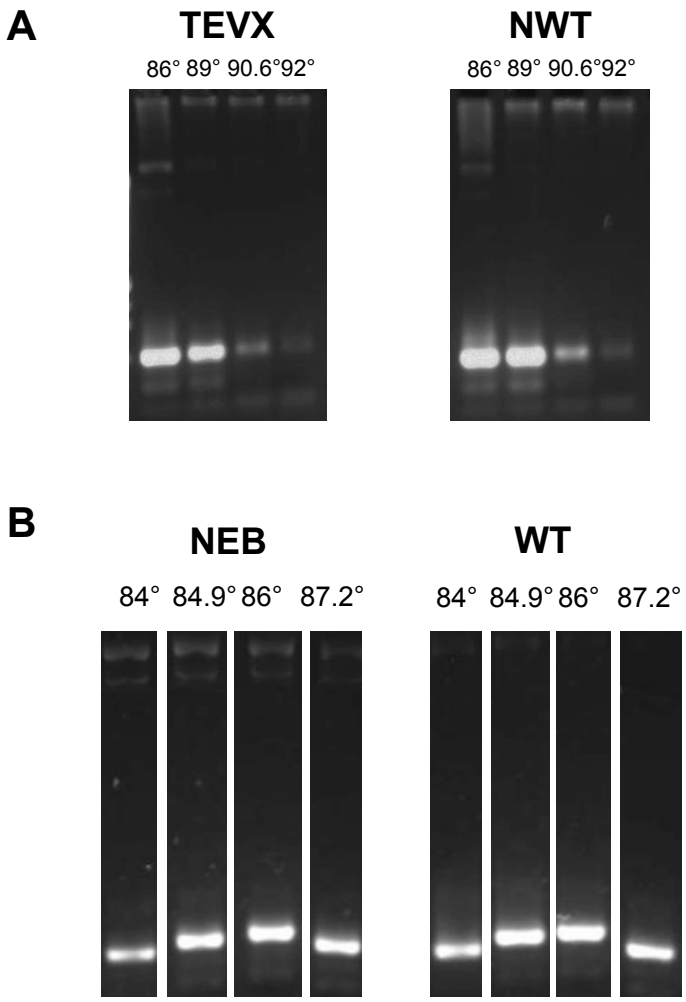


Figure 5.17: Additional PCR on recombinant constructs. (**A**) 1% agarose gel on PCR activity assays using original denaturation conditions; (**B**) 1% agarose gel on PCR activity assays using gentle denaturation conditions. Note that PCR activity data were run on the same gel, but lanes have been repositioned to simplify data presentation.

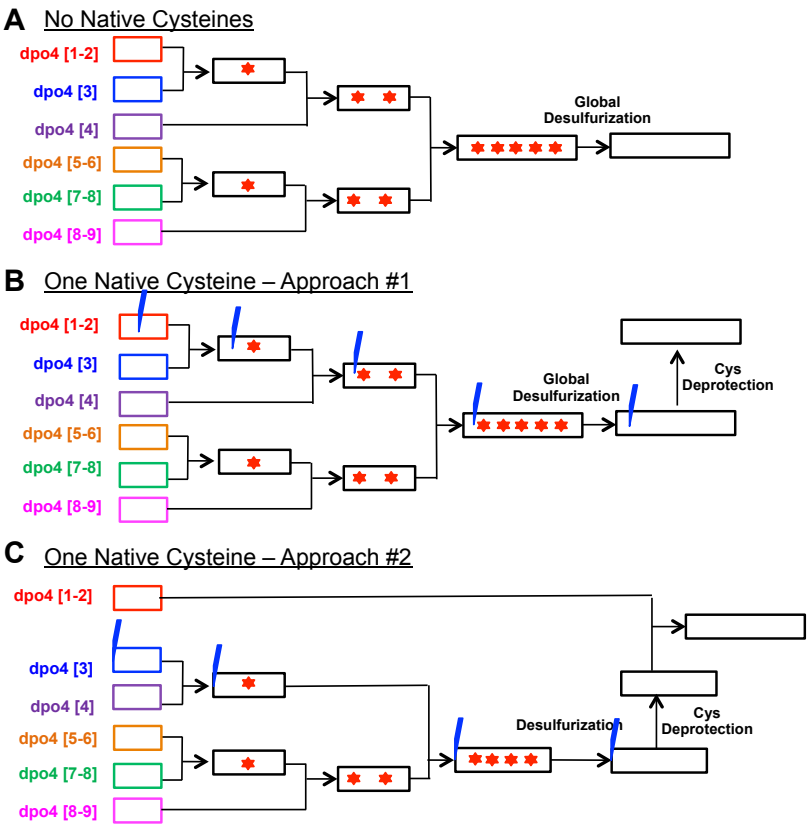


Figure 5.18: Assembly strategies for handling one cysteine. (**A**) No Cys assembly strategy; (**B**) One Cys assembly strategy – approach #1; (**C**) One Cys assembly strategy – approach #2. Red stars indicate junction site Cys that need to be desulfurized. Blue dagger indicates the one native Cys that needs to be protected from desulfurization.

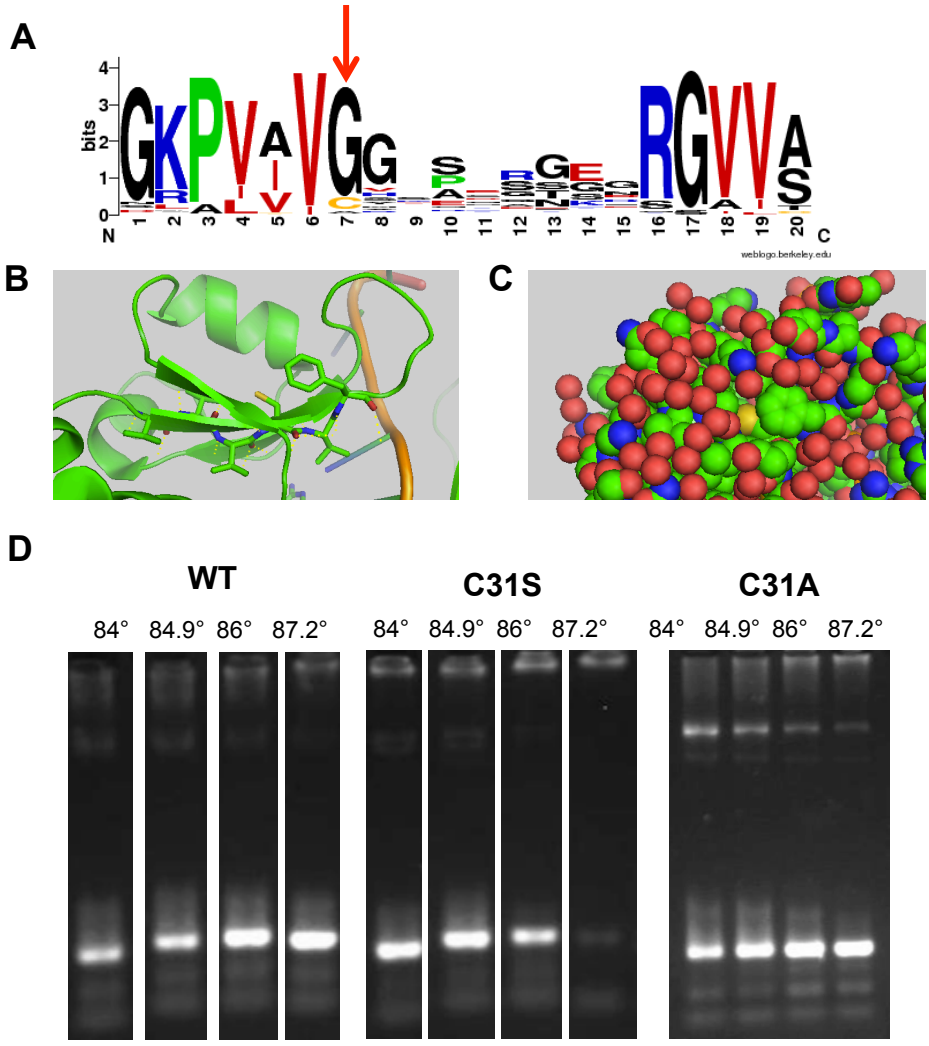


Figure 5.19: DOPPEL analysis at Cys31. (**A**) Sequence alignment data, with position 31 indicated with red arrow; **(B, C**) Analyses of the crystal structure at position 31 (1JXL.pdb) Sulfur residue at C31 is shown in orange. (**D**) 1% agarose gel on PCR activity assays using gentle denaturation conditions. Note that WT and C31S PCR activity data were run on the same gel, but lanes have been repositioned to simplify data presentation. C31A was analyzed on a different gel.

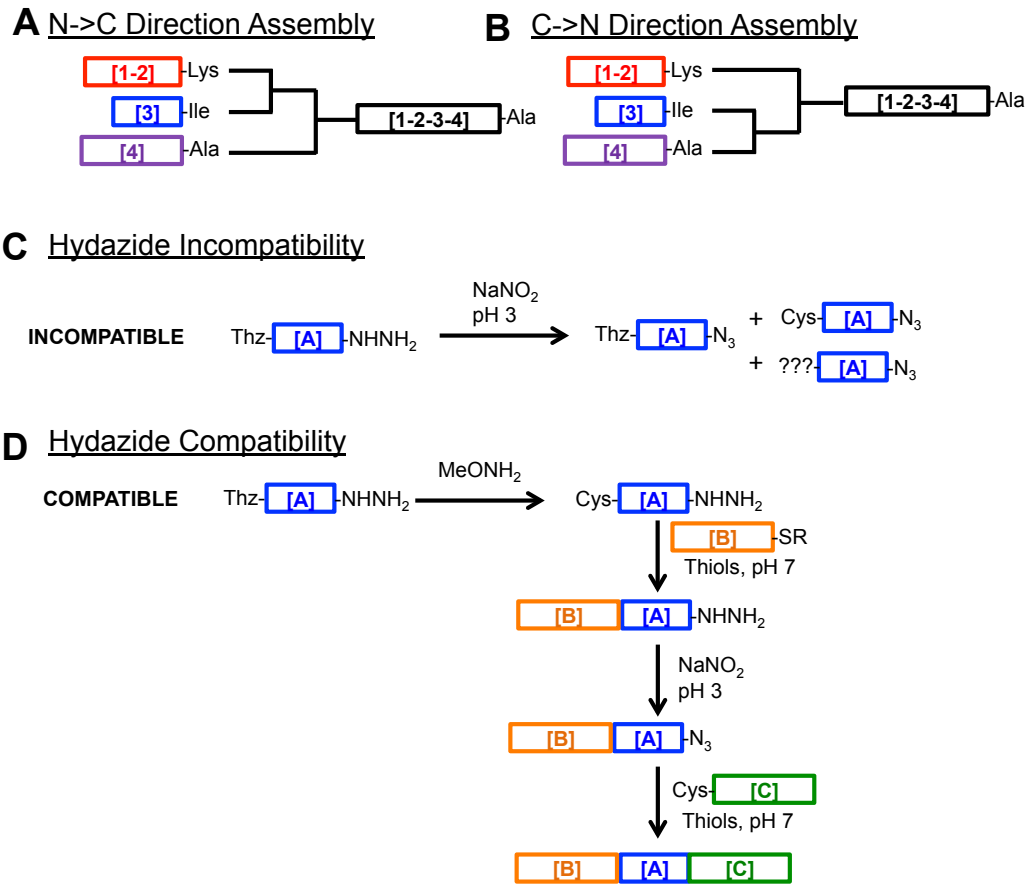


Figure 5.20: Chemical assembly strategies for N-terminal half. (**A**) N->C direction assembly; **(B)** C->N direction assembly; **(C)** Demonstration of the hydrazide-Thz incompabibility; (**D**) Demonstration of the hydrazide-Thz compatibility; (**E**) General activation method for Dawson Dbz/Nbz method to produce thioesters.

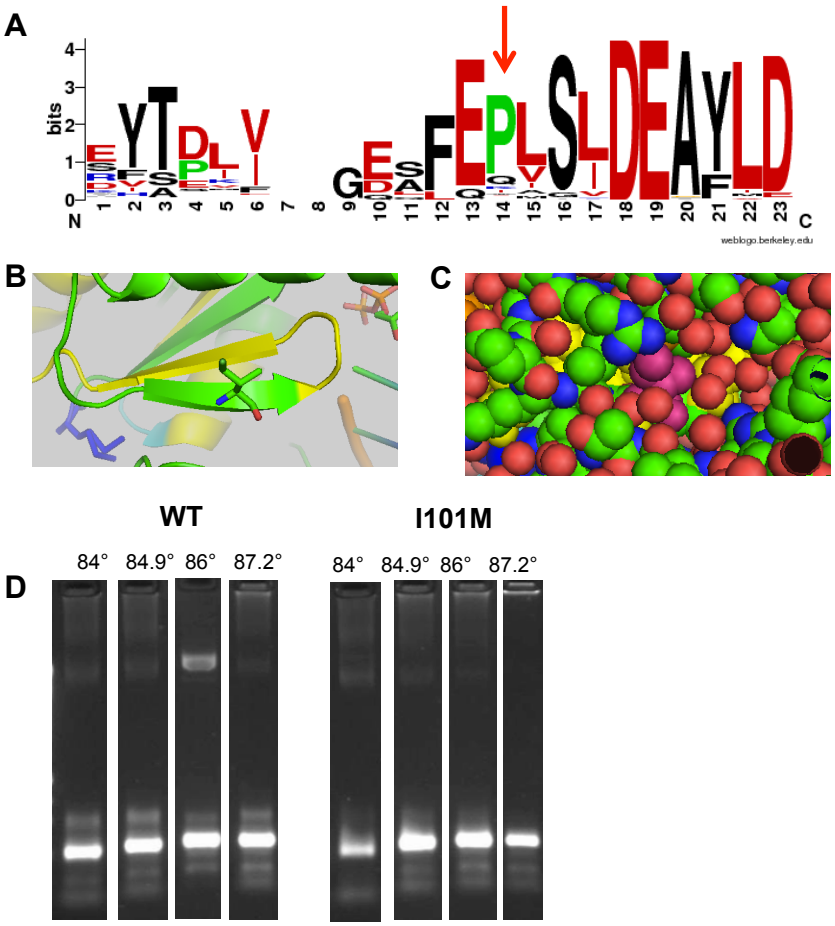


Figure 5.21: DOPPEL analysis at Ile101. (**A**) Sequence alignment data, with position 101 indicated with red arrow; **(B, C**) Analyses of the crystal structure at position 101 (1JXL.pdb) In right panel, residue at I101 is shown in pink. (**D**) 1% agarose gel on PCR activity assays using gentle denaturation conditions. Note that PCR activity data were run on the same gel, but lanes have been repositioned to simplify data presentation.

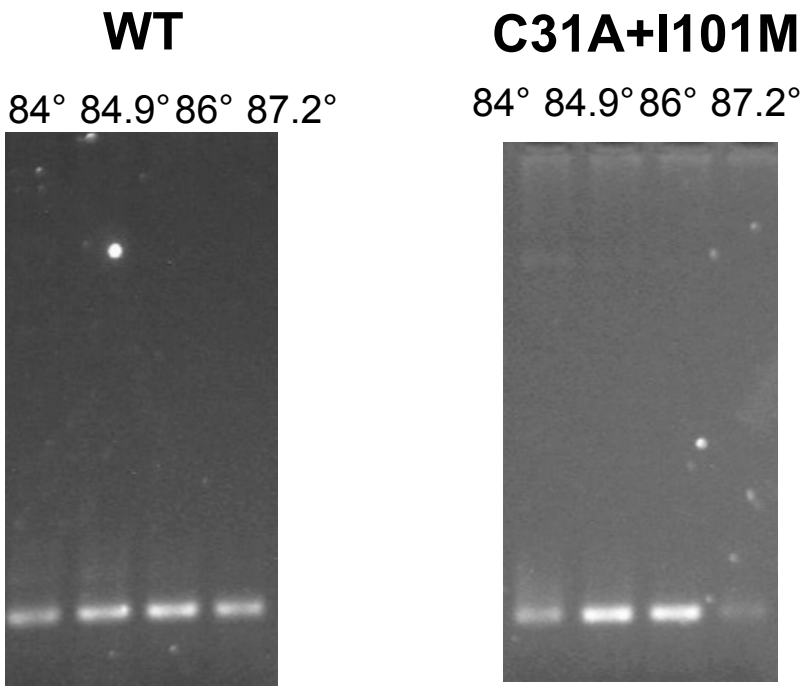


Figure 5.22: Activity analysis of the double mutant (C31A/I101M). 1% agarose gel on PCR activity assays using gentle denaturation conditions. Note that PCR activity data were run on the same gel.

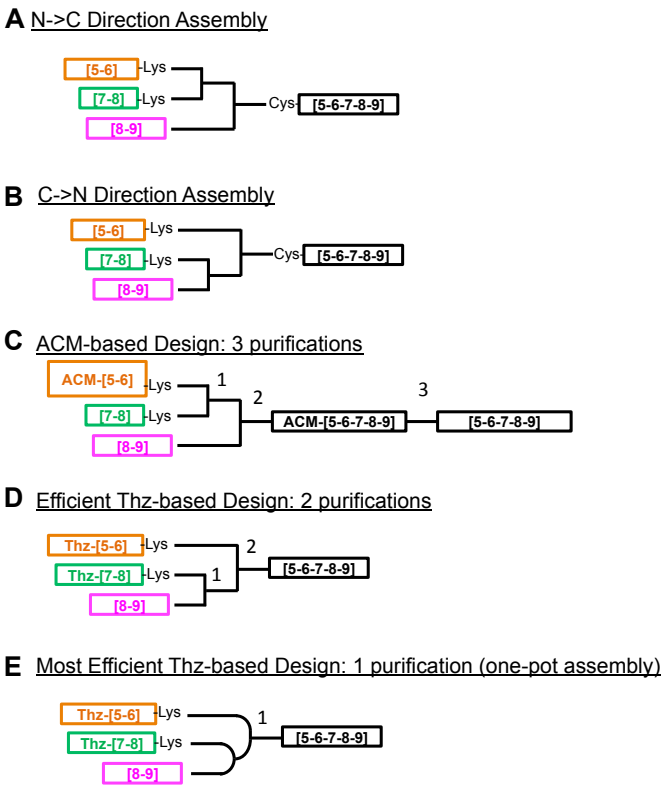


Figure 5.23: Chemical assembly strategies for C-terminal half. (**A**) N->C direction assembly; **(B)** C->N direction assembly; **(C)** ACM-based approach; (**D**) Two-step Thz-based approach; (**E**) One-step Thz-based approach.

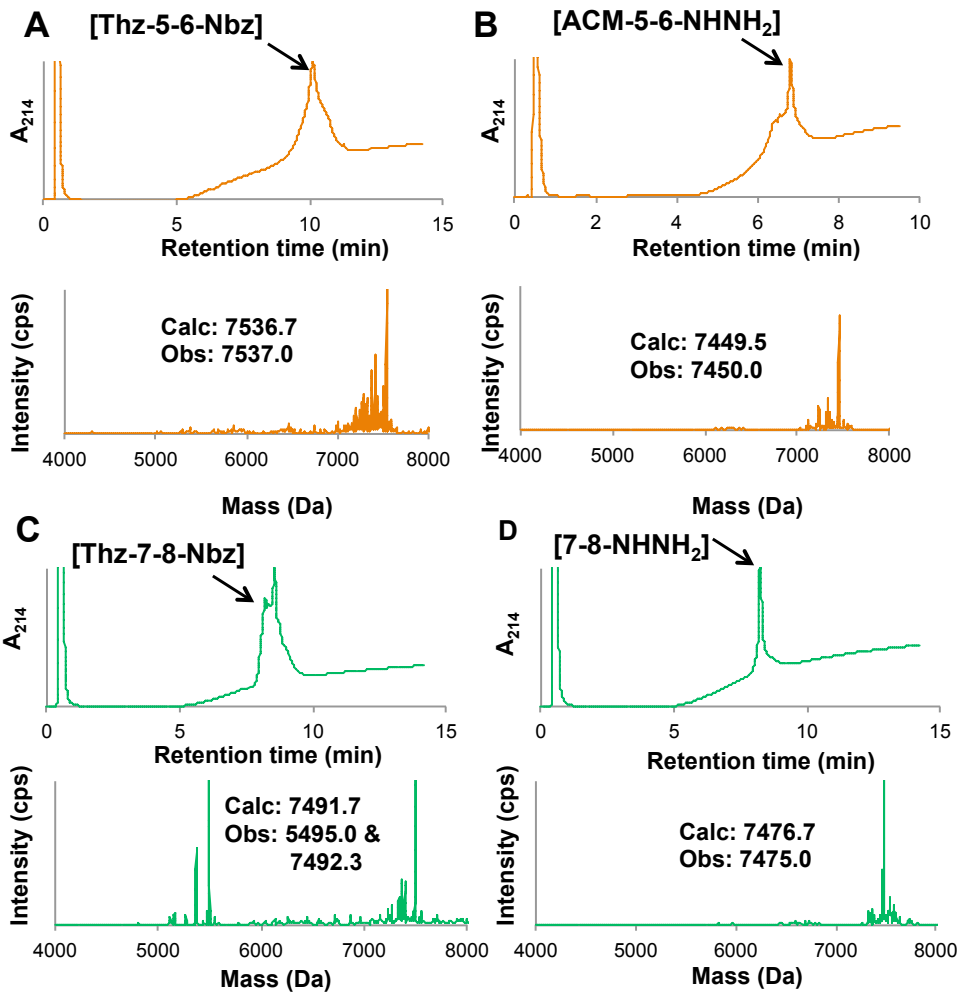


Figure 5.24: Synthesis of [5-6] and [7-8] by Dawson and hydrazide methods. (**A**) HPLC and MS of [Thz-5-6-Nbz]; (**B**) HPLC and MS of [ACM-5-6-NHNH2]; (**C**) HPLC and MS of [Thz-7-8-Nbz]; (**D**) HPLC and MS of [7-8-NHNH2]; Crude peptides were analyzed on C4-Aeris column (2.1 x 50 mm).

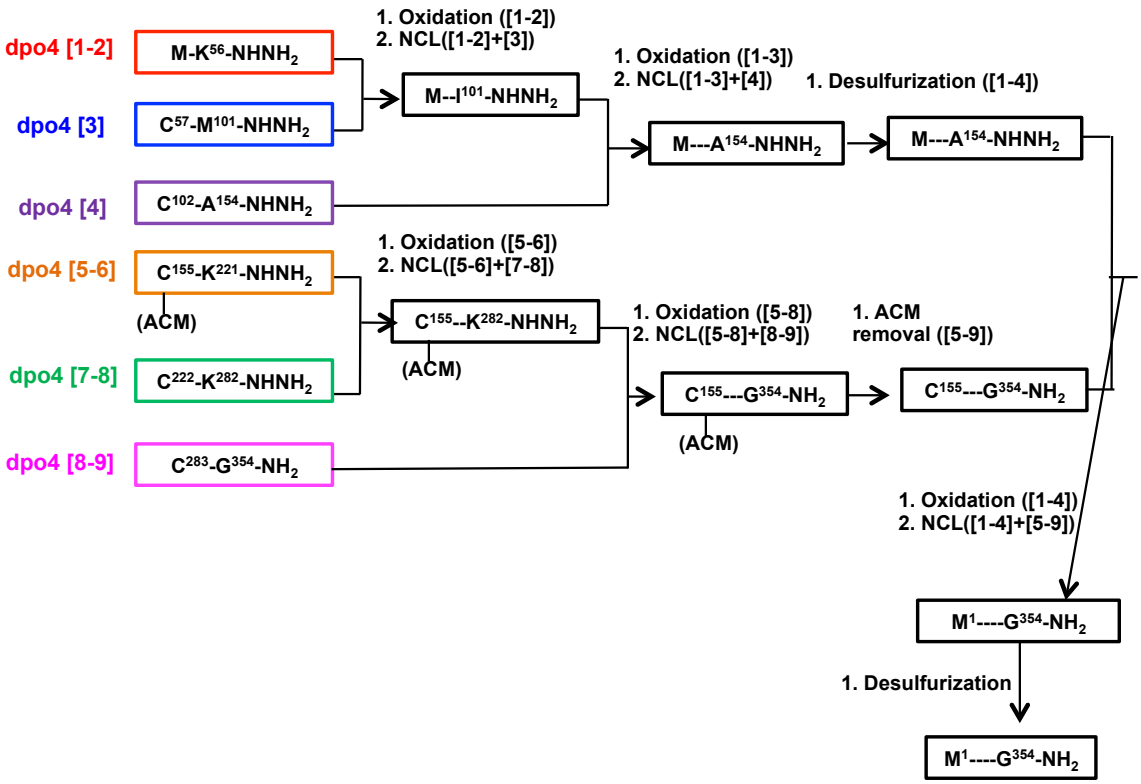


Figure 5.25: Final synthesis strategy for Dpo4.

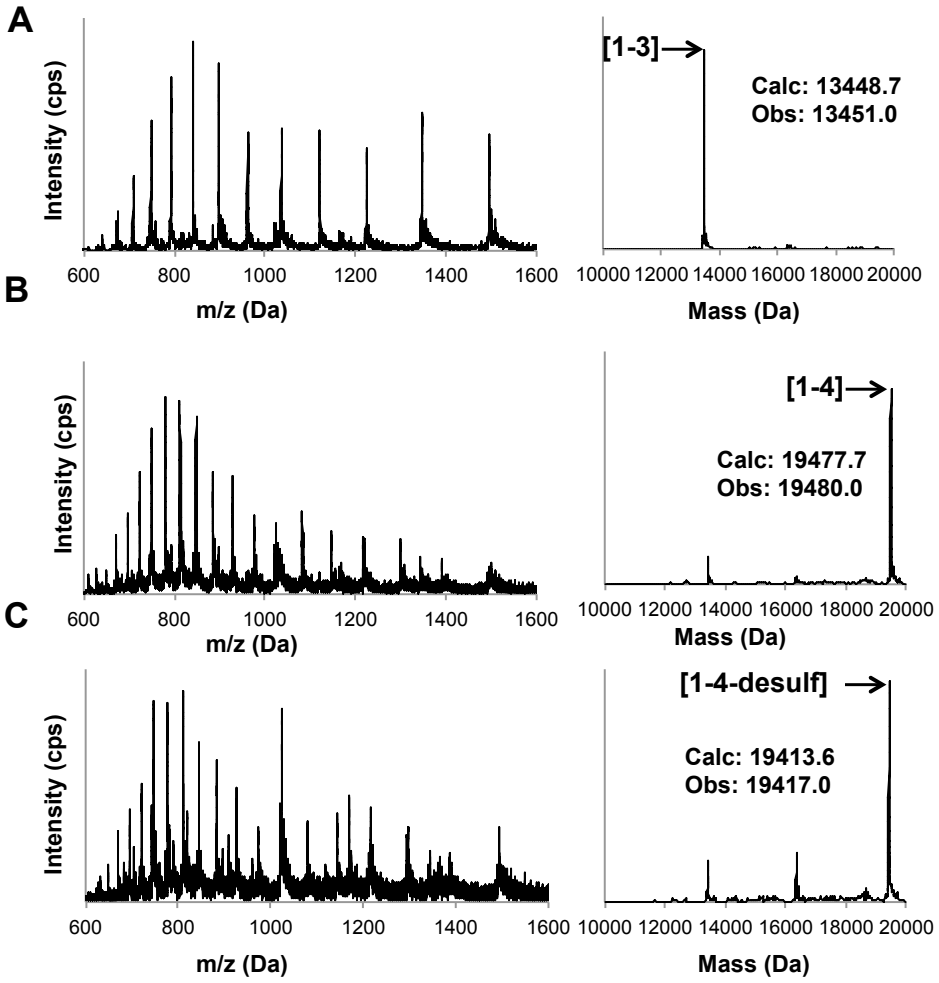


Figure 5.26: MS on the N-terminal intermediates. (**A**) Peptide [1-3]; (**B**) Peptide [1-4]; (**C**) Peptide [1-4-desulfurized];

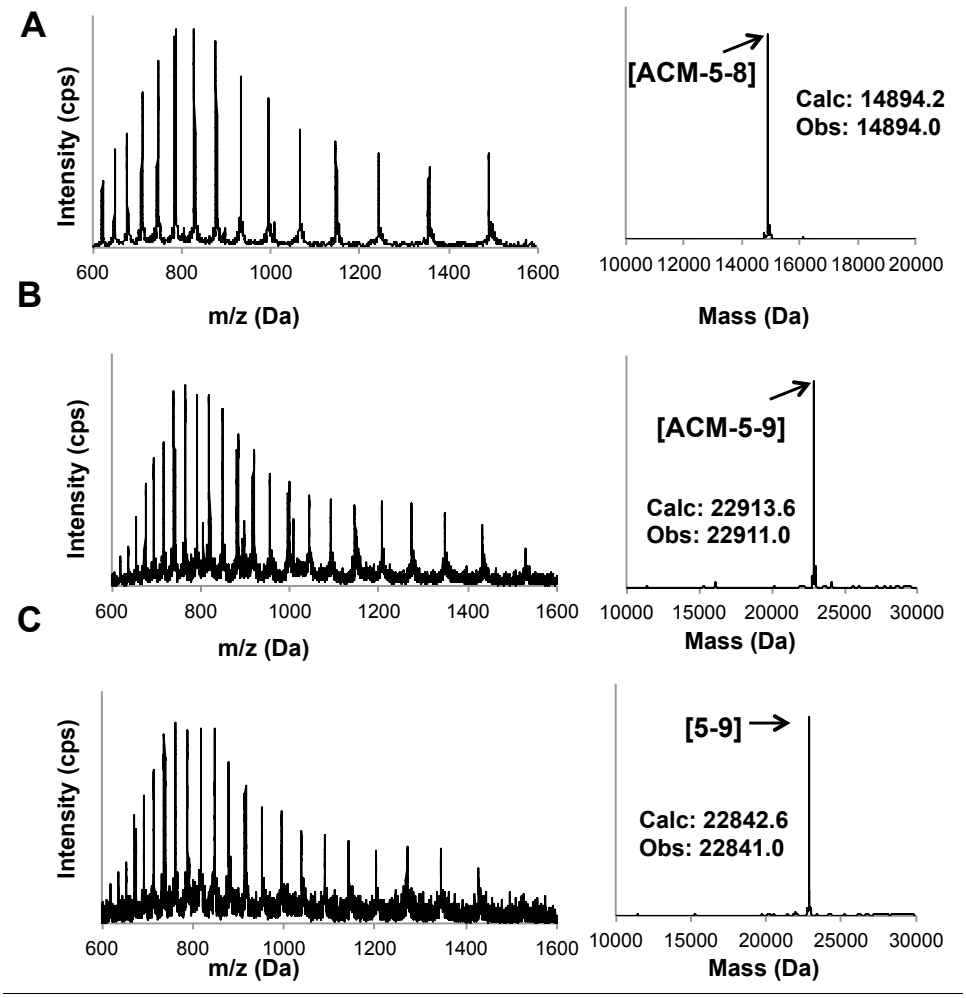


Figure 5.27: MS on the C-terminal intermediates. (**A**) Peptide [ACM-5.8]; (**B**) Peptide [ACM-5.9]; (**C**) Peptide [5.9];

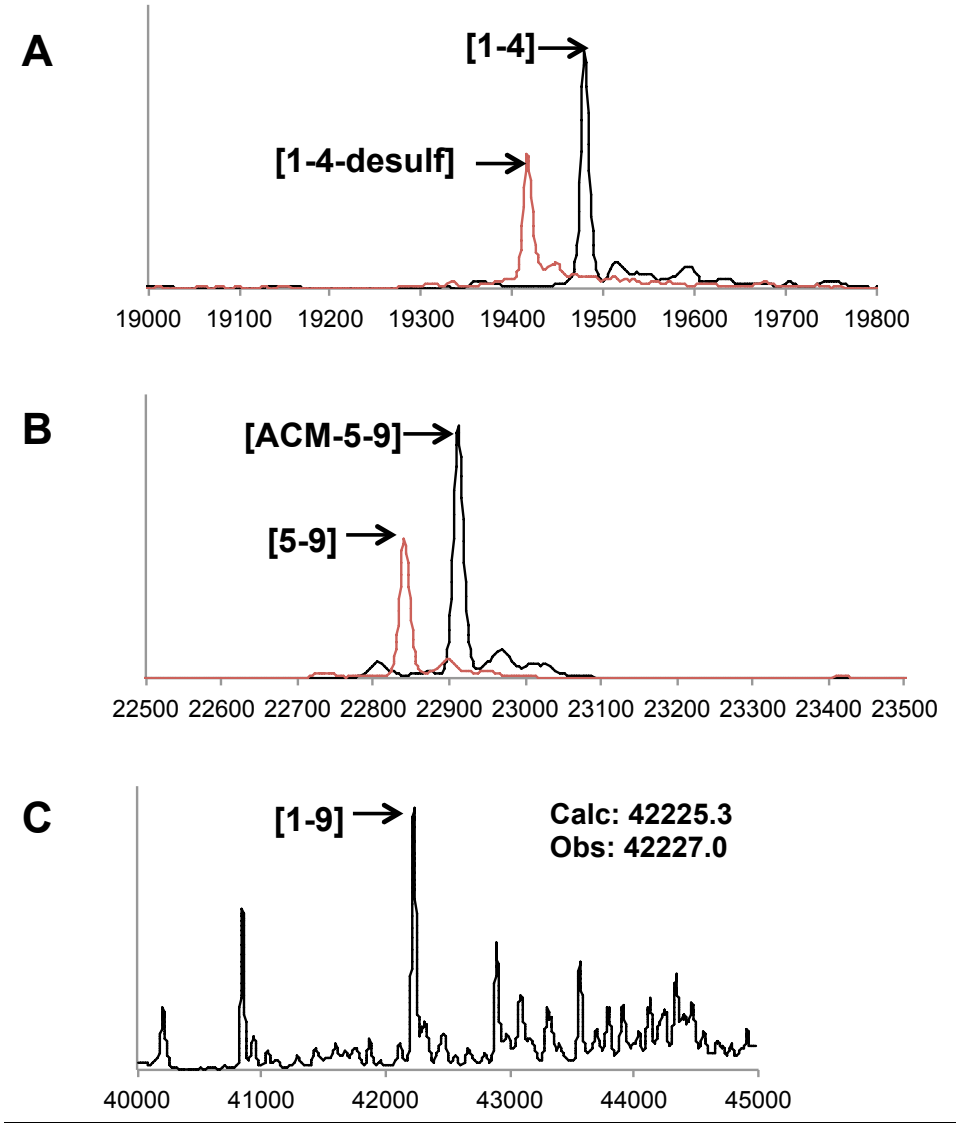


Figure 5.28: MS on the key intermediates and final product. (**A**) Magnified MS trace on peptides [1-4] and [1-4]-desulfurized; (**B**) Magnified MS trace on peptides [5.9] and [ACM-5.9]; (**C**) Magnified MS trace on full-length ligated product.

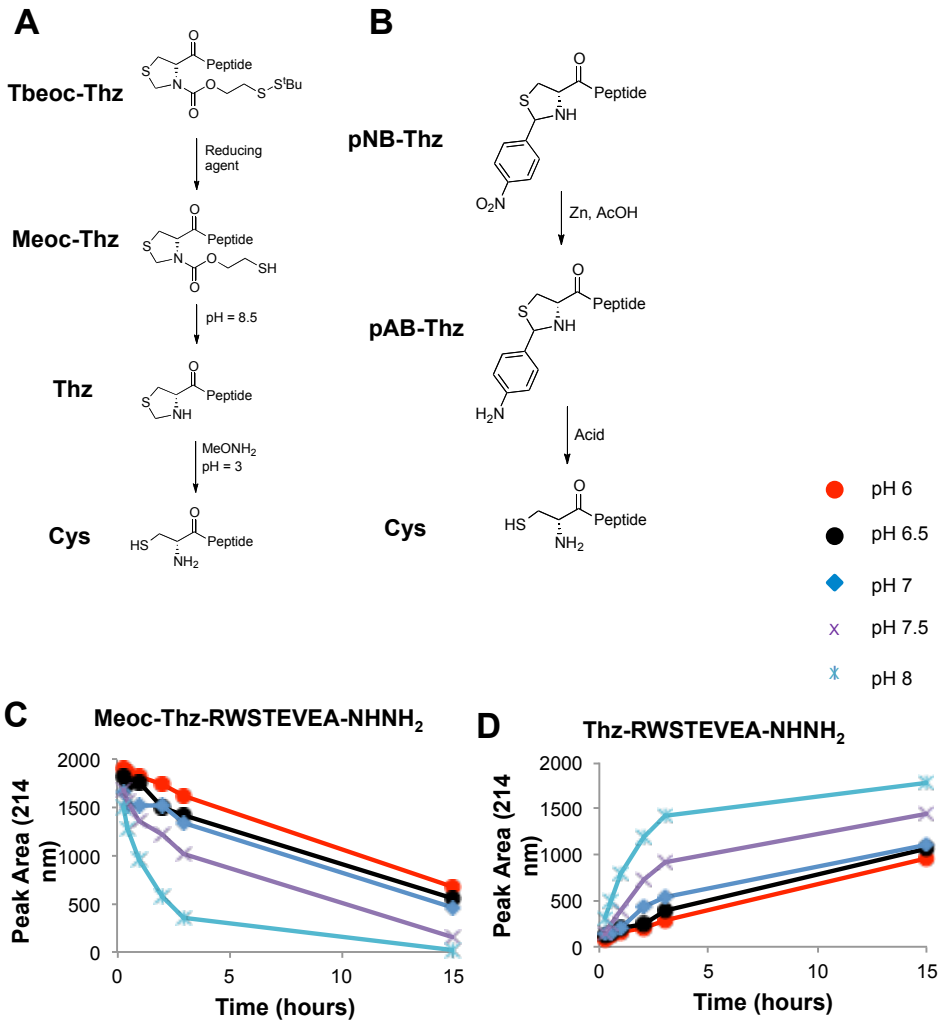


Figure 5.29: Dpo4 future directions. (**A**) Deprotection mechanism for Tbeoc-Thz; (**B**) Deprotection mechanism for pNB-Thz; (**C**) Time-course of Meoc-Thz-RWSTEVEA-NHNH2 peak area under different pH values; (**D**) Time-course of Thz-RWSTEVEA-NHNH2 peak area under different pH values.

5.7 References

1. Ling, H.; Boudsocq, F.; Woodgate, R.; Yang, W., Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. *Cell* **2001,** *107* (1), 91-102.

2. She, Q.; Singh, R. K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M. J.; Chan-Weiher, C. C.; Clausen, I. G.; Curtis, B. A.; De Moors, A.; Erauso, G.; Fletcher, C.; Gordon, P. M.; Heikamp-de Jong, I.; Jeffries, A. C.; Kozera, C. J.; Medina, N.; Peng, X.; Thi-Ngoc, H. P.; Redder, P.; Schenk, M. E.; Theriault, C.; Tolstrup, N.; Charlebois, R. L.; Doolittle, W. F.; Duguet, M.; Gaasterland, T.; Garrett, R. A.; Ragan, M. A.; Sensen, C. W.; Van der Oost, J., The complete genome of the crenarchaeon Sulfolobus solfataricus P2. *Proc Natl Acad Sci U S A* **2001,** *98* (14), 7835-40.

3. Zillig, W. S., K. O.; Wunderl, S.; Schulz, W.; Priess, H.; Scholz, I., The Sulfolobus-"Caldariella" group: Taxonomy on the basis of the structure of DNA-dependent RNA polymerases. *Arch Microbiol* **1980,** *125* (3), 259-269.

4. Boudsocq, F.; Iwai, S.; Hanaoka, F.; Woodgate, R., Sulfolobus solfataricus P2 DNA polymerase IV (Dpo4): an archaeal DinB-like DNA polymerase with lesion-bypass properties akin to eukaryotic poleta. *Nucleic Acids Res* **2001,** *29* (22), 4607-16.

5. McDonald, J. P.; Hall, A.; Gasparutto, D.; Cadet, J.; Ballantyne, J.; Woodgate, R., Novel thermostable Y-family polymerases: applications for the PCR amplification of damaged or ancient DNAs. *Nucleic Acids Res* **2006,** *34* (4), 1102-11.

6. Steitz, T. A., DNA polymerases: structural diversity and common mechanisms. *J Biol Chem* **1999,** *274* (25), 17395-8.

7. Yang, W., Damage repair DNA polymerases Y. *Curr Opin Struct Biol* **2003,** *13* (1), 23-30.

8. Boudsocq, F.; Kokoska, R. J.; Plosky, B. S.; Vaisman, A.; Ling, H.; Kunkel, T. A.; Yang, W.; Woodgate, R., Investigating the role of the little finger domain of Y-family DNA polymerases in low fidelity synthesis and translesion replication. *J Biol Chem* **2004,** *279* (31), 32932-40.

9. Kool, E. T., Active site tightness and substrate fit in DNA replication. *Annu Rev Biochem* **2002,** *71*, 191-219.

10. Doublie, S. S., M. R.; Ellenberger, T., An open and closed case for all polymerases. *Structure* **1999,** *7* (2), R31-R35.

11. Ohmori, H.; Friedberg, E. C.; Fuchs, R. P.; Goodman, M. F.; Hanaoka, F.; Hinkle, D.; Kunkel, T. A.; Lawrence, C. W.; Livneh, Z.; Nohmi, T.; Prakash, L.;

Prakash, S.; Todo, T.; Walker, G. C.; Wang, Z.; Woodgate, R., The Y-family of DNA polymerases. *Mol Cell* **2001,** *8* (1), 7-8.

12. (a) McCulloch, S. D.; Kunkel, T. A., The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res* **2008,** *18* (1), 148-61; (b) Bebenek, K.; Kunkel, T. A., Analyzing fidelity of DNA polymerases. *Methods Enzymol* **1995,** *262*, 217-32.

13. Vaisman, A.; Ling, H.; Woodgate, R.; Yang, W., Fidelity of Dpo4: effect of metal ions, nucleotide selection and pyrophosphorolysis. *EMBO J* **2005,** *24* (17), 2957-67.

14. Sherrer, S. M.; Beyer, D. C.; Xia, C. X.; Fowler, J. D.; Suo, Z., Kinetic basis of sugar selection by a Y-family DNA polymerase from Sulfolobus solfataricus P2. *Biochemistry* **2010,** *49* (47), 10179-86.

15. Joyce, C. M., Choosing the right sugar: how polymerases select a nucleotide substrate. *Proc Natl Acad Sci U S A* **1997,** *94* (5), 1619-22.

16. Kirouac, K. N.; Suo, Z.; Ling, H., Structural mechanism of ribonucleotide discrimination by a Y-family DNA polymerase. *J Mol Biol* **2011,** *407* (3), 382-90.

17. (a) Williams, J. S.; Kunkel, T. A., Ribonucleotides in DNA: origins, repair and consequences. *DNA Repair (Amst)* **2014,** *19*, 27-37; (b) Jinks-Robertson, S.; Klein, H. L., Ribonucleotides in DNA: hidden in plain sight. *Nat Struct Mol Biol* **2015,** *22* (3), 176-8.

18. Merrifield, R. B., Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *J Am Chem Soc* **1963,** *85* (14), 2149-2154.

19. (a) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B., Synthesis of proteins by native chemical ligation. *Science* **1994,** *266* (5186), 776-9; (b) Kent, S. B., Total chemical synthesis of proteins. *Chem Soc Rev* **2009,** *38* (2), 338-51.

20. (a) Carpino, L. A. H., G. Y., The 9-fluorenylmethoxycarbonyl function, a new base-sensitive amino-protecting group. *J Am Chem Soc* **1970,** *92* (19), 5748-5749; (b) Carpino, L. A. H., G. Y., The 9-fluorenylmethoxycarbonyl amino-protecting group. *J Org Chem* **1972,** *37* (22), 3404-3409.

21. Schnolzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B., In situ neutralization in boc-chemistry solid phase peptide synthesis. Rapid, high yield assembly of difficult sequences. *Int J Pept Protein Res* **1992,** *40* (3-4), 180-93.

22. Weinstock, M. T.; Jacobsen, M. T.; Kay, M. S., Synthesis and folding of a mirror-image enzyme reveals ambidextrous chaperone activity. *Proc Natl Acad Sci U S A* **2014,** *111* (32), 11679-84.

23. Verzele, D.; Madder, A., Patchwork protein chemistry: a practitioner's treatise on the advances in synthetic peptide stitchery. *Chembiochem* **2013,** *14* (9), 1032-48.

24. Kumar, K. S.; Bavikar, S. N.; Spasser, L.; Moyal, T.; Ohayon, S.; Brik, A., Total chemical synthesis of a 304 amino acid K48-linked tetraubiquitin protein. *Angew Chem Int Ed Engl* **2011,** *50* (27), 6137-41.

25. (a) Milton, R. C.; Milton, S. C.; Kent, S. B., Total chemical synthesis of a D-enzyme: the enantiomers of HIV-1 protease show reciprocal chiral substrate specificity. *Science* **1992,** *256* (5062), 1445-8; (b) Kochendoerfer, G. G.; Chen, S. Y.; Mao, F.; Cressman, S.; Traviglia, S.; Shao, H.; Hunter, C. L.; Low, D. W.; Cagle, E. N.; Carnevali, M.; Gueriguian, V.; Keogh, P. J.; Porter, H.; Stratton, S. M.; Wiedeke, M. C.; Wilken, J.; Tang, J.; Levy, J. J.; Miranda, L. P.; Crnogorac, M. M.; Kalbag, S.; Botti, P.; Schindler-Horvat, J.; Savatski, L.; Adamson, J. W.; Kung, A.; Kent, S. B.; Bradburne, J. A., Design and chemical synthesis of a homogeneous polymer-modified erythropoiesis protein. *Science* **2003,** *299* (5608), 884-7; (c) Durek, T.; Torbeev, V. Y.; Kent, S. B., Convergent chemical synthesis and high-resolution x-ray structure of human lysozyme. *Proc Natl Acad Sci U S A* **2007,** *104* (12), 4846-51; (d) Torbeev, V. Y.; Kent, S. B., Convergent chemical synthesis and crystal structure of a 203 amino acid "covalent dimer" HIV-1 protease enzyme molecule. *Angew Chem Int Ed Engl* **2007,** *46* (10), 1667-70; (e) Mandal, K.; Uppalapati, M.; Ault-Riche, D.; Kenney, J.; Lowitz, J.; Sidhu, S. S.; Kent, S. B., Chemical synthesis and X-ray structure of a heterochiral {D-protein antagonist plus vascular endothelial growth factor} protein complex by racemic crystallography. *Proc Natl Acad Sci U S A* **2012,** *109* (37), 14779-84; (f) Wang, P.; Dong, S.; Shieh, J. H.; Peguero, E.; Hendrickson, R.; Moore, M. A.; Danishefsky, S. J., Erythropoietin derived by chemical synthesis. *Science* **2013,** *342* (6164), 1357-60; (g) Wilkinson, B. L.; Stone, R. S.; Capicciotti, C. J.; Thaysen-Andersen, M.; Matthews, J. M.; Packer, N. H.; Ben, R. N.; Payne, R. J., Total synthesis of homogeneous antifreeze glycopeptides and glycoproteins. *Angew Chem Int Ed Engl* **2012,** *51* (15), 3606-10.

26. Haase, C.; Rohde, H.; Seitz, O., Native chemical ligation at valine. *Angew Chem Int Ed Engl* **2008,** *47* (36), 6807-10.

27. (a) Wohr, T. W., F.; Nefzi, A.; Rohwedder, B.; Sato, T.; Sun, X.; Mutter, M., Pseudo-prolines as a solubilizing, structure-disrupting protection technique in peptide synthesis *J Am Chem Soc* **1996,** *118* (39), 9218-9227; (b) Mutter, M. N., A.; Sato, T.; Sun, X.; Wahl, F.; Wohr, T., Pseudo-prolines (psi Pro) for accessing "inaccessible" peptides. *Pept Res* **1995,** *8* (3), 145-153.

28. Lauer, J. L. F., C. G.; Fields, G. B., Sequence dependence of aspartimide formation during 9-fluorenylmethoxycarbonyl solid-phase peptide synthesis. *Letters in Pep Sci* **1995,** *1* (4), 197-205.

29. Yang, Y. S., W. V.; Schneider, K.; Thornqvist, S.; Chait, B. T.; Tam, J. P., Aspartimide formation in base-driven 9-fluorenylmethoxycarbonyl chemistry. *Tetrahedron Lett* **1994,** *35* (52), 9689-9692.

30. Subiros-Funosas, R. E.-F., A.; Albericio, F., Use of Oxyma as pH modulatory agent to be used in the prevention of base-driven side reactions and its effect on 2-chlorotrityl chloride resin. *Biopolymers* **2012,** *98* (2), 89-97.

31. Michels, T.; Dolling, R.; Haberkorn, U.; Mier, W., Acid-mediated prevention of aspartimide formation in solid phase peptide synthesis. *Org Lett* **2012,** *14* (20), 5218-21.

32. Hackeng, T. M.; Griffin, J. H.; Dawson, P. E., Protein synthesis by native chemical ligation: expanded scope by using straightforward methodology. *Proc Natl Acad Sci U S A* **1999,** *96* (18), 10068-73.

33. Haase, C. R., H.; Seitz, O., Native chemical ligation at valine. *Angew Chem Int Ed Engl* **2008,** *47* (36), 6807-6810.

34. Yan, L. Z.; Dawson, P. E., Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurization. *J Am Chem Soc* **2001,** *123* (4), 526-33.

35. (a) Chen, J. W., Q.; Yuan, Y.; Zhu, J.; Danishefsky, S. J., Native chemical ligation at valine: a contribution to peptide and glycopeptide synthesis. *Angew Chem Int Ed Engl* **2008,** *47* (44), 8521-8524; (b) Yang, S. H.; Wojnar, J. M.; Harris, P. W.; DeVries, A. L.; Evans, C. W.; Brimble, M. A., Chemical synthesis of a masked analogue of the fish antifreeze potentiating protein (AFPP). *Org Biomol Chem* **2013,** *11* (30), 4935-42; (c) Medini, K.; Harris, P. W.; Hards, K.; Dingley, A. J.; Cook, G. M.; Brimble, M. A., Chemical synthesis of a pore-forming antimicrobial protein, caenopore-5, by using native chemical ligation at a glu-cys site. *Chembiochem* **2015,** *16* (2), 328-36.

36. (a) Malins, L. R.; Payne, R. J., Recent extensions to native chemical ligation for the chemical synthesis of peptides and proteins. *Curr Opin Chem Biol* **2014,** *22*, 70-8; (b) Wong, C. T. T., C. L.; Li, X., Synthetic cysteine surrogates used in native chemical ligation. *Molecular Biosystems* **2013,** *9* (5), 826-833; (c) He, Q. Q. F., G. M.; Liu, L., Design of thiol-containing amino acids for native chemical ligation at non-Cys sites. *Chin Chem Lett* **2013,** *24* (4), 265-269.

37. Malins, L. R. C., K. M.; Payne, R. J., Chemoselective sulfenylation and peptide ligation at tryptophan. *Chem Sci* **2014,** *5* (1), 260-266.

38. (a) Bang, D.; Kent, S. B., A one-pot total synthesis of crambin. *Angew Chem Int Ed Engl* **2004,** *43* (19), 2534-8; (b) Bang, D.; Pentelute, B. L.; Kent, S. B., Kinetically controlled ligation for the convergent chemical synthesis of proteins. *Angew Chem Int Ed Engl* **2006,** *45* (24), 3985-8; (c) Tang, S. S., Y. Y.;

Wang, Z. P.; Mei, K. R.; Chen, X.; Cheng, J. Y.; Zheng, J. S.; Liu, L., An Efficient One-Pot Four-Segment Condensation Method for Protein Chemical Synthesis. *Angew Chem Int Ed Engl* **2015,** *Epub ahead of print*.

39. Haack, T. M., M., Serine derived oxazolidines as secondary structure disrupting, solubilizing building blocks in peptide synthesis. *Tetrahedron Lett* **1992,** *33* (12), 1589-1892.

40. Johnson, T. Q., M.; Owen, D.; Sheppard, R. C., A reversible protecting group for the amide bond in peptides. Use in the synthesis of 'difficult sequences'. *J Chem Soc Chem Comm* **1993,** (4), 369-372.

41. Hendrickson, J. B., Systematic Synthesis Design. 6. Yield Analysis and Convergency. *J Am Chem Soc* **1977,** *99* (16), 5439-5450.

42. Siman, P. K., S. V.; Nikolov, M.; Fischle, W.; Brik, A., Convergent Chemical Synthesis of Histone H2B Protein for the Site-Specific Ubiquitination at Lys34. *Angew Chem Int Ed Engl* **2013,** *52* (31), 8059-8063.

43. Papadopoulos, J. S.; Agarwala, R., COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* **2007,** *23* (9), 1073-9.

44. Waterhouse, A. M.; Procter, J. B.; Martin, D. M.; Clamp, M.; Barton, G. J., Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009,** *25* (9), 1189-91.

45. Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E., WebLogo: a sequence logo generator. *Genome Res* **2004,** *14* (6), 1188-90.

46. (a) Fang, G. M.; Li, Y. M.; Shen, F.; Huang, Y. C.; Li, J. B.; Lin, Y.; Cui, H. K.; Liu, L., Protein chemical synthesis by ligation of peptide hydrazides. *Angew Chem Int Ed Engl* **2011,** *50* (33), 7645-9; (b) Zheng, J. S.; Tang, S.; Qi, Y. K.; Wang, Z. P.; Liu, L., Chemical synthesis of proteins using peptide hydrazides as thioester surrogates. *Nat Protoc* **2013,** *8* (12), 2483-95.

47. Fang, G. M.; Wang, J. X.; Liu, L., Convergent chemical synthesis of proteins by ligation of peptide hydrazides. *Angew Chem Int Ed Engl* **2012,** *51* (41), 10347-50.

48. (a) Pan, M.; He, Y.; Wen, M.; Wu, F.; Sun, D.; Li, S.; Zhang, L.; Li, Y.; Tian, C., One-pot hydrazide-based native chemical ligation for efficient chemical synthesis and structure determination of toxin Mambalgin-1. *Chem Commun (Camb)* **2014,** *50* (44), 5837-9; (b) Li, J.; Li, Y.; He, Q.; Li, Y.; Li, H.; Liu, L., One-pot native chemical ligation of peptide hydrazides enables total synthesis of modified histones. *Org Biomol Chem* **2014,** *12* (29), 5435-41; (c) Boll, E.; Ebran, J. P.; Drobecq, H.; El-Mahdi, O.; Raibaut, L.; Ollivier, N.; Melnyk, O., Access to

large cyclic peptides by a one-pot two-peptide segment ligation/cyclization process. *Org Lett* **2015,** *17* (1), 130-3.

49. Blanco-Canosa, J. B. D., P. E., An efficient Fmoc-SPPS approach for the generation of thioester peptide precursors for use in native chemical ligation. *Angew Chem Int Ed Engl* **2008,** *47* (36), 6851-8655.

50. Wang, J. X.; Fang, G. M.; He, Y.; Qu, D. L.; Yu, M.; Hong, Z. Y.; Liu, L., Peptide o-aminoanilides as crypto-thioesters for protein chemical synthesis. *Angew Chem Int Ed Engl* **2015,** *54* (7), 2194-8.

51. Muttenthaler, M.; Ramos, Y. G.; Feytens, D.; de Araujo, A. D.; Alewood, P. F., p-Nitrobenzyl protection for cysteine and selenocysteine: a more stable alternative to the acetamidomethyl group. *Biopolymers* **2010,** *94* (4), 423-32.

52. Schwarz, H.; Arakawa, K., The Use of p-Nitrobenzyl Esters in Peptide Synthesis1. *J Am Chem Soc* **1959,** *81* (21), 5691-5695.

53. Gowda, D. C.; Mahesh, B.; Gowda, S., Zinc-catalyzed ammonium formate reduc- tions : Rapid and selective reduction of ali- phatic and aromatic nitro compounds. *Indian J Chem, SEC B* **2001,** *40B*, 75-77.

54. Johnson, E. C.; Kent, S. B., Insights into the mechanism and catalysis of the native chemical ligation reaction. *J Am Chem Soc* **2006,** *128* (20), 6640-6.

55. Zheng, J. S.; Tang, S.; Guo, Y.; Chang, H. N.; Liu, L., Synthesis of cyclic peptides and cyclic proteins via ligation of peptide hydrazides. *Chembiochem* **2012,** *13* (4), 542-6.

56. Wan, Q.; Danishefsky, S. J., Free-radical-based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides. *Angew Chem Int Ed Engl* **2007,** *46* (48), 9248-52.