THE UNIVERSITY OF CHICAGO

NEW CHEMICAL METHODS FOR THE SYNTHESIS OF PROTEINS AND THEIR APPLICATION TO THE ELUCIDATION OF PROTEIN STRUCTURE BY RACEMIC PROTEIN CRYSTALLOGRAPHY

VOLUME ONE

A DISSERTATION SUBMITTED TO THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES IN CANDIDACY FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMISTRY

BY
BRAD L. PENTELUTE

CHICAGO, ILLINOIS
MARCH 2008
TABLE OF CONTENTS

VOLUME ONE

page

LIST OF PUBLICATIONS .................................................. iv
LIST OF FIGURES ............................................................ vi
LIST OF SCHEMES ........................................................... x
LIST OF TABLES ............................................................. xi
ACKNOWLEDGEMENTS ..................................................... xii

CHAPTER

1. Introduction .............................................................. 1

2. Design and synthesis of a novel topological analogue of
crambin ........................................................................... 17

3. Kinetically-controlled ligation for the convergent synthesis of
proteins ........................................................................... 45

4. Selective desulfurization of Cys in the presence of Cys(Acm)
in polypeptides obtained by native chemical
ligation ............................................................................ 60

5. Total chemical synthesis by selective desulfurization of the mirror
image form of the hypothetical protein Rv1738 from
*Mycobacterium tuberculosis* .................................................. 79

6. Total chemical synthesis of snow flea antifreeze protein (sfAFP)
using modern chemical ligation methods: preparation of the
protein enantiomers L-sfAFP and D-sfAFP ........................................... 88

7. Total chemical synthesis and racemic crystallography used
to determine the novel Xray structure of a snow flea antifreeze
protein (sfAFP) ................................................................. 111

8. Significance and Future Perspectives ....................................... 133
**VOLUME TWO**

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Synthesis of peptides and peptide-thioester building blocks for use in convergent chemical ligation</td>
<td>140</td>
</tr>
<tr>
<td>B</td>
<td>Total synthesis of proteins by convergent chemical ligation of unprotected peptides</td>
<td>174</td>
</tr>
<tr>
<td>C</td>
<td>Design and implementation of undergraduate Fmoc-SPPS laboratories</td>
<td>219</td>
</tr>
<tr>
<td>D</td>
<td>Direct on-resin synthesis of peptide-(alpha)thiophenylesters for use in native chemical ligation</td>
<td>233</td>
</tr>
<tr>
<td>E</td>
<td>Solving the X-ray structure of the scorpion toxin BmBKTx1 by racemic crystallography and direct methods</td>
<td>246</td>
</tr>
<tr>
<td>F</td>
<td>Modular total chemical synthesis and racemic crystallization of Kaliotoxin</td>
<td>260</td>
</tr>
<tr>
<td>G</td>
<td>Comparative properties of Insulin-like Growth Factor 1 (IGF-1) and [Gly7D-Ala]IGF-1</td>
<td>272</td>
</tr>
<tr>
<td>H</td>
<td>Insights from atomic-resolution Xray structures of chemically-synthesized HIV-1 protease in complex with inhibitors</td>
<td>288</td>
</tr>
<tr>
<td>I</td>
<td>Total Chemical Synthesis of a Cysteine-Free 10FN3 monobody protein that targets the integrin α,β3</td>
<td>308</td>
</tr>
<tr>
<td>J</td>
<td>Design, synthesis, and NMR of site-specifically isotope labeled L-sfAFP</td>
<td>323</td>
</tr>
</tbody>
</table>
LIST OF PUBLICATIONS

Chapter 2
To be submitted. Pentelute, B.L., Gilbreth, R., Bang, D., Kurutz, J., Koide, S., Kent, S.B.H.

Chapter 3

Chapter 4

Chapter 6
To be submitted. Pentelute, B.L., Gates P., Vanderkooi, J., Kent, S.B.H.

Chapter 7

Appendix A

Appendix B

Appendix D
Appendix E
To be submitted. Mandal, K., Pentelute, B.L., Tereshko, V.T., Kossiakoff, A.A., Kent, S.B.H.

Appendix G

Appendix H
LIST OF FIGURES

Figure 1.1 Swollen peptide resin in DMF....................................................4
Figure 1.2 Native chemical ligation............................................................6
Figure 1.3 LCMS analysis of crude products from the stepwise Boc chemistry SPPS.............................................................8
Figure 2.1 Novel topological analogue of crambin......................................18
Figure 2.2 Retrosynthetic strategy #1 for the total synthesis of our novel topological analogue of crambin...........................................20
Figure 2.3 Retrosynthetic strategy #2 for the total synthesis of our novel topological analogue of crambin...........................................23
Figure 2.4 Analytical LCMS traces for the selective native chemical ligation of [Thz\(^{16}\)-Gly\(^{31}\)]\(^{-}\)alkylthioester and [Cys\(^{31}\)-Thr\(^{39}\)]\(^{-}\)alkylthioester.............25
Figure 2.5 Analytical LCMS traces for the ligation between [Thr\(^{1}\]-Lys\(^{10}\) (Asn\(^{46}\)-Thz\(^{10}\)-Ala\(^{15}\)] \(^{-}\)alkylthioester and [Cys\(^{16}\]-Thr\(^{39}\)] \(^{-}\)alkylthioester...........27
Figure 2.6 Retrosynthetic strategy #3 for the total synthesis of our novel topological analogue of crambin.............................................29
Figure 2.7 Analytical LCMS data for the synthesis of cyclic crambin[Thr\(^{1}\]-Lys\(^{10}\) (Asn\(^{46}\)-Cys\(^{40}\))]-Thr\(^{39}\)].........................................................31
Figure 2.8 Analytical HPLC data for the folding of fully unprotected cyclic crambin [Thr\(^{1}\]-Lys\(^{10}\) (Asn\(^{46}\)-Cys\(^{40}\))]-Thr\(^{39}\) (SH)\(_{6}\)........................................33
Figure 2.9 Analytical LCMS data for [Thr\(^{1}\]-Lys\(^{10}\) (Asn\(^{46}\)-Thz\(^{40}\)-Ala\(^{15}\)] \(^{-}\)alkylthioester.........................................................37
Figure 2.10 Analytical LCMS data for [Thr\(^{1}\]-Lys\(^{10}\) (Asn\(^{46}\)-Thz\(^{40}\)-Ala\(^{15}\)] \(^{-}\)ary/thioester.........................................................38
Figure 2.11 Analytical LCMS data for [Cys\(^{16}\]-Thr\(^{39}\)] \(^{-}\)alkylthioester...........40
Figure 2.12 Two-dimensional TOCSY \(^{1}\)H NMR spectra of the crambin analogue..............................................................42
Figure 3.1 Kinetically-controlled ligation of [Thz4-Ala15]- \(\alpha\)-thiophenylester and [Cys16-Leu25]- \(\alpha\)-thioalkylester.........................................................49
Figure 3.2 The final step of the fully convergent synthesis of crambin...........51
Figure 3.3 The crude ligation product from the reaction of [Thr1-Cys3]- \(\alpha\)-thiophenylester and [Cys4-Leu25]- \(\alpha\)-thioalkylester.........................56
Figure 3.4 Crude products from the reaction of [Thz26-Gly31]- \(\alpha\)-thiophenylester and [Cys32-Thr39]- \(\alpha\)-thioalkylester.................................57
Figure 4.1 Model study for the selective desulfurization of Cys to Ala in the presence of Cys(Acm) .................................................63
Figure 4.2 Analytical HPLC data for the synthesis of [A24P,S27P,S28P] amylin.................................................................65
Figure 4.3 Analytical HPLC data for the synthesis of [Met7Nle]EETI-II[1-28].......68
Figure 4.4 Analytical HPLC of crude [A24P,S27P,S28P] amylin after removal of Acm groups and oxidation.................................72
Figure 4.5 Analytical HPLC data for folding EETI-II.................................73
Figure 4.6  Analytical HPLC for the selective desulfurization of Cys in the presence of Met.................................................................74
Figure 4.7  Analytical HPLC of the selective desulfurization of Cys in the presence of both Cys(Acm) and Met.............................................75
Figure 5.1  The 94-residue amino acid sequence for Mtb Rv1738 and the synthetic strategy.................................................................80
Figure 5.2  Analytical LCMS traces for synthesis of D-Rv1738.....................83
Figure 5.3  Purified synthetic product of L-Mtb Rv1738................................89
Figure 6.1  The 81-residue glycine-rich amino acid sequence for sfAFP and the synthetic strategy.................................................................92
Figure 6.2  Analytical LCMS traces for synthetic folded L- and D-sfAFP........94
Figure 6.3  The CD spectra of the protein enantiomers D-sfAFP and L-sfAFP....95
Figure 6.4  Ice recrystallization inhibition assay........................................96
Figure 6.5  LCMS analysis of L-[Thz1-Gly12]-thioester................................101
Figure 6.6  LCMS analysis of L-[Thz13-Gly27]-thioester................................102
Figure 6.7  LCMS analysis of L-[Thz28-Gly42]-thioester...............................103
Figure 6.8  LCMS analysis of L-[Cys43-Pro81]-COOH.................................104
Figure 6.9  Analytical HPLC traces of aliquots of the ligation of [Thz28-Gly42] thioester and [Cys43-Pro81]-COOH...........................................105
Figure 6.10 Analytical HPLC traces of aliquots from the ligation of [Thz13-Gly27] thioester and [Cys28-Pro81]-COOH.................................106
Figure 6.11 Analytical HPLC traces of aliquots taken from the ligation of [Thz1 Gly12]-thioester and [Cys13-Pro81]-COOH........................................107
Figure 6.12 Folding/Disulfide formation for Synthetic sfAFP(Cys1-Pro81)........108
Figure 6.13 Ice recrystallization inhibition assay series................................109
Figure 7.1  Racemic crystallization of [D-sfAFP + L-sfAFP].............................115
Figure 7.2  The pseudo-racemate crystal form of sfAFP...............................117
Figure 7.3  The xray structure of snow flea antifreeze protein (sfAFP) ..........118
Figure 7.4  SigmaA-weighted 2Fo-Fc electron density map of the proposed ice binding surface for P2; L-sfAFP and P1<bar> {D-sfAFP + LsfAFP}......................................................................................120
Figure 7.5  Analytical LCMS traces for synthetic folded L-Se-sfAFP................122
Figure 8.1  Sequential segment condensation and fully convergent synthetic strategies ..............................................................................134
Figure 8.2  Amino acids needed for the extension of native chemical ligation to Xxx-Yyy ligation sites.........................................................136
Figure A.1  Schematic guide to the protocols ..............................................178
Figure A.2  Preparation of aminomethyl-copoly(styrene-divinylbenzene) resin.........................................................................................179
Figure A.3  Boc solid phase peptide synthesis .............................................179
Figure A.4  HF cleavage apparatus setup ....................................................180
Figure A.5  Boc SPPS of peptide- thioesters and transthioesterification with MPAA HPLC profiles ..........................................................181
Figure B.1  Ligation of unprotected peptide segments by means of native chemical ligation ........................................213
Figure B.2  Chemical protein synthesis overview ..........................................................214
Figure B.3  Sequential versus convergent assembly of proteins ..................................215
Figure B.4  Native chemical ligation of Cram[Thz\textsuperscript{26}-Gly\textsuperscript{31}]-"thioester and Cram[Cys\textsuperscript{32}-Asn\textsuperscript{46}] .................................................................216
Figure B.5  Kinetically controlled ligation of Cram[Thr\textsuperscript{1}-Ala\textsuperscript{15}]-"thio(4-carboxymethyl)phenyl ester and Cram[Cys\textsuperscript{16}-Leu\textsuperscript{25}]-"thioalkylester ......................................................217
Figure B.6  Final assembly of Crambin ...........................................................................218
Figure E.1  The LC-MS profiles of the folded and purified synthetic BmBKTx1 enantiomers ..............................................................248
Figure E.2  X-ray structure of BmBKTx1 .................................................................251
Figure F.1  The amino acid sequence and synthetic strategy used to prepare Kaliotoxin ..................................................................................261
Figure F.2  Analytical traces for the characterization of synthetic D-KTX and L-KTX ..............................................................262
Figure F.3  Pictures of Kaliotoxin crystals ....................................................................264
Figure F.4  Ligation of L-KTX [Thz\textsuperscript{14}-Lys\textsuperscript{27}]-"thioester and [Cys\textsuperscript{28}-Lys\textsuperscript{38}]-COOH .................................................................268
Figure F.5  Ligation of L-KTX [Gly\textsuperscript{1}-Gln\textsuperscript{13}]-"thioester and [Cys\textsuperscript{14}-Lys\textsuperscript{38}]-COOH and folding of [Gly\textsuperscript{1}-Lys\textsuperscript{38}]-COOH in one-pot .................................................................269
Figure I.1  10FN3 ........................................................................................................309
Figure I.2  The synthetic strategy used for the preparation of 10FN3 .......................311
Figure I.3  Analytical LCMS traces for the one-pot synthesis of 10FN3 ............313
Figure I.4  Desulfurization [A39C,A76C]10FN3(1-96) .......................................................315
Figure I.5  1-D NMR of 10FN3 ................................................................................320
Figure J.1  Sequence and LC-MS analysis of folded, purified site specifically isotope-labeled L-sfAFP ..................................................325
Figure J.2  \textsuperscript{15}N-H HSQC spectrum of isotope-labeled sfAFP analogue #2 ........327
Figure J.3  \textsuperscript{15}N-H HSQC spectrum of isotope-labeled sfAFP analogue #2 ..........328
Figure J.4  \textsuperscript{15}N-\textsuperscript{1}H HSQC of isotope-labeled sfAFP ........................................330
Figure J.5  \textsuperscript{13}C-\textsuperscript{1}H HSQC spectrum of isotope-labeled sfAFP analogue #2 ......332
Figure J.6  LCMS analysis of the unlabeled peptide Thz\textsuperscript{1}-Lys-Gly-Ala-Asp-Gly-Ala-His-Gly-Val-Asn-Gly\textsuperscript{12}-CO-S-CH\textsubscript{2}-CH\textsubscript{2}-CO-Leu-Pro-COOH ........................................336
Figure J.7  LCMS analysis of the unlabeled peptide Thz\textsuperscript{13}-Pro-Gly-Thr-Ala-Gly-Ala-Gly-Ser-Val-Gly-Gly-Pro-Gly\textsuperscript{27}-CO-S-CH\textsubscript{2}-CH\textsubscript{2}-CO-Ile-COOH ........................................337
Figure J.8  LCMS analysis of the isotope labeled peptide Thz\textsuperscript{28}-Asp-Gly*-Gly*-His-Gly*-Gly*-Asn-Gly*-Gly*-Asn-Gly*-Asn-Pro-Gly\textsuperscript{42}-CO-S-CH\textsubscript{2}-CH\textsubscript{2}-CO-Leu-Pro-COOH ........................................338
Figure J.9  LCMS analysis of the isotope labeled peptide Cys$^{43}$-Ala-Gly-Gly-Val-Gly-Gly-Ala-Ser-Gly-Thr-Gly-Val-Gly*-Gly*-Arg-Gly*-Gly*-Lys-Gly*-Gly*-Ser-Gly*-Thr-Pro-Lys-Gly-Ala-Asp-Gly-Ala-Pro-Gly*-Ala-Pro$^{81}$-COOH .......................................................... 339

Figure J.10  Analytical HPLC traces of aliquots of the ligation of labeled-(2)-[Thz$^{28}$-Gly$^{42}$]-"thioester .................................................. 340

Figure J.11  Analytical HPLC traces of the ligation of [Thz$^{13}$-Gly$^{27}$]"thioester and labeled-(2)-[Cys$^{28}$-Pro$^{81}$]-COOH .......... 341

Figure J.12  Analytical HPLC traces of aliquots taken from the ligation of [Thz$^{1}$-Gly$^{12}$]"thioester labeled-(2)-[Cys$^{13}$-Pro$^{81}$]-COOH. .................................................. 342

Figure J.13  Folding/disulfide formation for synthetic labeled-(analogue #2)-(Cys$^{1}$-Pro$^{81}$) sfAFP ....................................................... 343

Figure J.14  LCMS analysis of the peptide Thz$^{1}$-Lys-Gly-Ala-Asp-Gly-Ala-His-Gly-Val-Asn-Gly$^{12}$-CO-S-CH$_2$-CH$_2$-CO-Ile-Pro-COOH after HF cleavage .......................................................... 345

Figure J.15  LCMS analysis of the crude peptide Thz$^{13}$-Pro-Gly*-Thr-Ala-Gly-Ala-Ala-Gly-Ser-Val-Gly-Gly-Pro-Gly$^{27}$-CO-S-CH$_2$-CH$_2$-CO-Ile-Pro-COOH .................................................. 346

Figure J.16  LCMS analysis of the crude peptide Thz$^{28}$-Asp-Gly-Gly-His-Gly-Gly-Asn-Gly-Gly-Asn-Gly-Ala-Asp-Gly-Ala-Pro-Gly$^{42}$-CO-S-CH$_2$-CH$_2$-CO-Ile-COOH ..................................................... 347


Figure J.18  Analytical HPLC traces of aliquots of the ligation of [Thz$^{28}$-Gly$^{42}$]-"thioester and labeled-(1)-[Cys$^{43}$-Pro$^{81}$]-COOH ................. 349

Figure J.19  Analytical HPLC traces of aliquots of the ligation of labeled-(1)-[Thz$^{13}$-Gly$^{27}$]-"thioester and labeled-(1)-[Cys$^{28}$-Pro$^{81}$]-COOH. .......................................................... 350

Figure J.20  Analytical HPLC traces of aliquots taken from the ligation of labeled-(1)-[Thz$^{1}$-Gly$^{12}$]"thioester and labeled-(1)-[Cys$^{13}$-Pro$^{81}$]- .......................................................... 351

Figure J.21  Folding/Disulfide formation for Synthetic sfAFP labeled-(analogue #1)-(Cys$^{1}$-Pro$^{81}$) .......................................................... 352
**LIST OF SCHEMES**

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scheme 1.1</strong></td>
<td>Modern chemical protein synthesis roadmap</td>
<td>2</td>
</tr>
<tr>
<td><strong>Scheme 1.2</strong></td>
<td>Boc ‘in situ neutralization’ SPPS synthetic cycle</td>
<td>5</td>
</tr>
<tr>
<td><strong>Scheme 3.1</strong></td>
<td>Key steps for the realization of convergent synthesis using native chemical ligation</td>
<td>47</td>
</tr>
<tr>
<td><strong>Scheme 3.2</strong></td>
<td>Fully convergent synthesis of the model protein crambin from six peptide segments</td>
<td>48</td>
</tr>
<tr>
<td><strong>Scheme 4.1</strong></td>
<td>General synthetic strategy for Cys-containing targets using native chemical ligation combined with selective desulfurization</td>
<td>58</td>
</tr>
<tr>
<td><strong>Scheme 4.2</strong></td>
<td>Synthetic strategy used for the synthesis of [A24P,S27P, S28P]amylin</td>
<td>64</td>
</tr>
<tr>
<td><strong>Scheme 4.3</strong></td>
<td>Synthetic Strategy used for the total synthesis of EETI-II[1-28]</td>
<td>66</td>
</tr>
<tr>
<td><strong>Scheme J.1</strong></td>
<td>Strategy for site-specific isotope labeling of the sfAFP</td>
<td>324</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 5.1  The D-peptides prepared for the synthesis of the D-enantiomorph of Rv1738........................................................................................................86
Table 7.1  The X-ray data collection and refinement statistics for the quasi-racemate sfAFP........................................................................................................126
Table 7.2  Backbone RMSD (Å) deviations between molecules in the quasi-racemate........................................................................................................127
Table 7.3  The X-ray data collection and refinement statistics for L-sfAFP........................................................................................................................127
Table 7.4  The X-ray data collection and refinement statistics for racemic sfAFP................................................................................................................128
Table B.1  The X Troubleshooting section for NCL, KCL ........................................202
Table E.1  The X-ray data collection and refinement statistics for the crystal structure of BmBKTx1..............................................................256
Table F.1  The X-ray data statistics for the D L KTX..................................................265
Table J.1  Heteronuclear NOEs for the ~30 $^{15}$N-$^1$H HSQC cross peaks the site-specifically isotope-labeled sfAFP analogue #2........329
ACKNOWLEDGEMENTS

I would like to thank my supportive family for helping me accomplish the arduous task of becoming a scientist. Michelle you are great, thanks my love. I would like to thank Steve for being a spectacular advisor and it was a pleasure to work with you! I enjoyed the insightful and continual scientific conversations with my friends and colleagues who helped me through the tough times Duhee Bang and Zak Gates. I have also enjoyed working with Yohei Soma, Kalyan Mandel, Thomas Durek, Erik Johnson, and other members in the Kent Lab. I am very grateful for the time and help Valentina Tereshko provided during our crystallographic investigations. I would also like to thank Ryan Gilbreth for helping with the NMR investigation of the crambin. Other University of Chicago members that I am grateful to for their time and education are Tony Kossiakoff, Shohei Koide, Rustem Ismagilov, and Josh Kurutz. I would also like to thank my committee members Steve Kent, Chuan He, and Bob Keenan.
Chapter 1. Introduction

Our goal is to use the tools of organic chemistry to tailor the protein in almost unlimited ways in order to study proteins as natural products, i.e. chemical compounds with interesting properties, that occur in the natural world.

Proteins are essential components of the molecular machinery of life. A detailed understanding of protein molecules, a comprehension at the atomic level, is required for the complete formulation of life. Understanding the functioning of protein molecules is currently a major area of research. Protein research spans many aspects of science, ranging from single molecule studies to the description of proteomes. The function of a protein is determined by the precise folded (‘tertiary’) structure of a protein molecule. One proven approach to the study of proteins centers on the production of homogenous protein of defined composition, which is then used to determine the structure and function of the protein molecule. Once an understanding of the structure and function of individual protein molecules has been achieved, system wide studies can be carried out to investigate how proteins interact with one another in the context of the proteome.

Typically, the very basics of protein science start with recombinant DNA expression, isolation, and purification of a target protein. The structure of the target protein is then systematically varied by modern molecular biology techniques, such as site-directed mutagenesis[1], and the effects of these structural changes are evaluated by modern biophysical methods aimed to gather new information such as binding affinities, phosphorylation sites, or catalytic mechanisms. For example, Ala scanning mutagenesis[2] is used to identify what residues are necessary for protein function. In this approach, protein libraries are built by replacing the native residue with Ala, with the aim of making a causal inference as to the specific residues necessary for function. Site-directed mutagenesis and related techniques are powerful and of great utility, but because this approach relies on recombinant expression in living cells, research on proteins must be designed to answer questions with the 20 natural L-amino acids. This restriction limits protein science!
To circumvent this problem, the field has developed an arsenal of tactics. One such approach, that is still being developed, is nonsense-codon suppression with misloaded tRNA molecules for the incorporation of a non-natural amino acid into a protein molecule.[3] This allows for the incorporation into a protein molecule by recombinant expression of new chemical functionalities such alkynes and photocrosslinking agents. However, considerable work is still needed before this approach becomes common practice.

A more versatile alternative to recombinant expression of proteins would be total chemical synthesis. The synthesis of proteins to give chemical control over every atom in a protein molecule is no longer an unrealistic goal. The desire to undertake the total chemical synthesis of proteins originated from the work of the great chemist Emil Fischer, who in the first decade of the twentieth century aspired to the total synthesis of enzyme molecules.[4] Current, state-of-the-art protein science has shown that with chemical access we can systematically dissect the function of a protein molecule and then use that knowledge to tailor its properties in an unprecedented manner.[5]

The modern approach to the total synthesis of proteins is shown in Scheme 1.1.

Scheme 1.1. Modern chemical protein synthesis roadmap
The first step to achieve the total synthesis of a protein is to devise a synthetic strategy. The synthetic strategy must use established chemical tactics that produce the full-length polypeptide in reasonable amounts, such that cost in time and material is not prohibitive. The synthetic strategy is commonly guided by the use of the modern methods for the chemical ligation of unprotected peptide segments[6], developed in the Kent laboratory. Once such a strategy is devised, stepwise solid phase peptide synthesis (SPPS)[7] is used to prepare the necessary peptide building blocks. The peptides are purified by RP-HPLC to homogeneity as judged by mass spectrometry. The building blocks are then ligated to produce the target full-length linear polypeptide that is then folded to form the unique tertiary (quaternary) structure of the functional protein molecule. The 3D structure is then determined. The steps involved in total chemical protein synthesis are iterative and are guided by empirical results; hence, when problems are encountered a new approach can be devised to surmount the challenge(s).

Why does the above chemical protein synthesis roadmap provide a robust first step? The core of protein synthesis is successful because peptides can be prepared by SPPS and later ligated by native chemical ligation (NCL)[6] to prepare the full-length polypeptide. The chemical nature of the products can be confirmed in a fast and efficient manner using analytical high performance liquid chromatography coupled to an electrospray mass spectrometer (LCMS)[8]. Here I will briefly summarize why these techniques—SPPS, NCL, and LCMS—are core for the chemical synthesis of proteins.

**Stepwise solid phase peptide synthesis (SPPS) is a rapid and chemically efficient technique for the assembly of ~50 residue or less polypeptides.** SPPS has been highly optimized such that each coupling is >99.5% efficient.[7] During chain elongation the growing peptide is effectively in solution thereby displaying solution phase coupling kinetics.[9] Protected peptides are prepared at high concentrations because of the resin’s favorable solubilizing effect. The growing peptide resin swells as a function of chain length (Figure 1.1).[9]
Two different sets of protecting group tactics are commonly used for SPPS, based on either the tertiarybutyloxycarbonyl (Boc) N-alpha protecting group, or the fluorenlymethylloxycarbonyl (Fmoc) N-alpha protecting group. Boc chemistry SPPS uses graduated acidolysis and side chain benzyl-based protection, while Fmoc chemistry SPPS uses different chemical mechanisms for removal of the Fmoc N-alpha protecting group (base-catalyzed elimination) and the side chain trityl- and butyl-based protection (acidolysis).[9]
Boc chemistry SPPS is the preferred method for the preparation of peptide thioesters, because the thioester moiety is labile to the conditions used to remove the Fmoc N-alpha protecting group.[10] The Boc chemistry SPPS synthetic cycle used in this work is shown in **Scheme 1.2**.

**Scheme 1.2.** Cycle used for the synthesis of peptides by manual Boc chemistry ‘in situ neutralization’ SPPS.

This synthetic cycle is based on the manual ‘in situ neutralization’ technique, developed as general approach to prepare virtually any peptide in a highly efficient manner.[7] The total time required for each cycle of amino acid addition is 15-20 minutes. This enables the manual synthesis of a 30 residue peptide in a single (long) day. In situ neutralization is the preferred chemistry for Boc-SPPS because the phenomenon of aggregating ‘difficult’ sequences is minimized.
Native chemical ligation (NCL) is a practical and robust reaction for the covalent joining of two unprotected peptides in aqueous solution. For simple, practical reasons, the stitching together (‘ligation’) of unprotected peptide segments must be used in order to overcome peptide solubility problems, and to insure effective purification and characterization of synthetic intermediates. The need for unambiguous covalent joining of unprotected peptide segments led the Kent laboratory to develop ‘chemical ligation’ (i.e. chemoselective reaction) approaches. Native chemical ligation – the thioester-mediated reaction of a peptide–thioester with an N-terminal cysteine-peptide to give a native amide (‘peptide’) bond at the ligation site (Figure 1.2) – is the most robust and useful amide-forming ligation chemistry developed to date.[6, 11] Reactions are carried out in aqueous 6 M guanidine.HCl at neutral pH, are typically complete within a few hours and give near quantitative yields of the desired product.

Figure 1.2. Native chemical ligation is a highly chemo-and regio-selective reaction, that enables the precise covalent joining of two polypeptides in the presence of all the unprotected functional groups typically found in ribosomally-translated polypeptide chains. Key functionalities are highlighted. Red: -“thioester; blue: N-terminal cysteine
Analytical LCMS is a rapid and effective tool to confirm the purity and precise mass of synthetic peptides and proteins. Every synthetic step ranging from the confirmation of SPPS products to protein folding can be effectively monitored by LCMS.[12] LCMS is simple and rapid.[8] It functions by eluting the synthetic peptide from an analytical reverse phase HPLC column by a 0.1% trifluoroacetic acid-acetonitrile-water eluant containing a gradient of increasing organic composition. As compounds elute from the column, they are electrosprayed into an ion trap MS thereby providing a realtime mass-to-charge readout. The final result is an LCMS chromatogram that details product homogeneity and corresponding mass-to-charge ratios over the LC peak observed at 214 nm. This technique provides a very complete analytical picture as to the nature of the chemical products at all stage of a synthesis. The MS instrument settings are general for smaller peptides, and need slight modification for larger (> 100 residue) polypeptide and for folded protein constructs. Figure 1.3 illustrates the power of the LCMS technique.
**Figure 1.3.** (Upper) LCMS analysis of crude products from the stepwise Boc chemistry SPPS of the peptide Thz-RLRWREKELVGVLARLNPADRNVEIGDELSVAR-thioester. The vertical axis is OD(214nm). (Inset) The electrospray MS of the principal component eluting at 29.8min had a mass of 4287.2 ± 0.5 Daltons, and corresponds to the desired product (calculated mass 4287.7 Daltons (average isotope composition). (Lower) Expanded view of the HPLC data in the region of interest. Identities and mass differences of some of the minor components are as labeled.

The main component monitored at 214nm is the desired product and appears as a symmetrical peak with side products eluting before and after. The LCMS readout shows +3H and +4H charge states corresponding to a mass of 4287.2 ± 0.5 Daltons, and immediately confirms that the desired 4287.7 ± 0.5 Da (average isotopes) peptide was synthesized. Closer examination details the side products that are present in the final product as shown in the bottom panel of the **Figure 1.3.** This high-resolution analysis provides a precise, high-resolution analytical tool that gives essential information for synthetic optimization.

Despite the SPPS, chemical ligation, and LCMS technologies described above, at the time of starting this thesis research the chemical protein synthesis toolkit is still subject to important limitations. Novel chemistries were needed to expand the size and
scope of proteins that can be prepared by total chemical synthesis. The average size of a protein is 300 amino acids and the tools necessary to construct a protein of this size had yet to be developed. And, to expand the utility of native chemical ligation as the main engine for chemical protein synthesis, methods need to be developed that allow for ligation at residues other than cysteine. In addition, even more practical methods for the synthesis of peptide thioesters are needed for Fmoc-SPPS. It is difficult to prepare peptide thioesters by Fmoc chemistry SPPS and this limits the use of chemical protein synthesis because 95% of all peptides are made by the Fmoc chemistry SPPS.

**Racemic Crystallography:**

Once synthetic access to a chemical target has been established, the experimental determination of protein structure is a key step in understanding the molecular basis of biological function. High-resolution protein structures are most often determined by Xray crystallography.[13] But if a target protein will not form well-ordered crystals, it is not possible to determine the molecular structure by Xray diffraction methods. Crystal growth is the bottleneck for the modern determination of a protein structure.[14]

A potential solution to this problem is provided by the prediction that a racemic mixture made up of a native L-protein molecule and its enantiomer (i.e. the mirror image form of the protein, D-protein) would crystallize more readily than the native L-protein.[15] Racemic protein solutions are hypothesized to favor crystal growth because the achiral space group P<1bar> is accessible and presents the highest probability of crystal formation for proteins.

The hypothesis of facile crystal growth from racemic mixtures has yet to be tested. In order to explore the benefits of racemic crystallization the *mirror image protein* needs to be prepared and this is only possible by the use of chemistry.[16] Indeed, all proteins found in nature are chiral: that is, a protein molecule cannot be superimposed on its own mirror image molecular form (‘enantiomer’). Linear polypeptide chains are constructed in living cells from the 19 L-alpha amino acids and the achiral amino acid glycine. The newly-formed polypeptide chain folds to give the
three-dimensional structure that is the defining characteristic of a protein molecule; it is this tertiary structure that gives rise to the functional properties of the protein. The fold is inherently chiral, even at the level of the polypeptide backbone itself: for example, alpha helices found in natural proteins made from L-amino acids (and glycine) are right-handed. It has been shown that if all the chiral centers in the 19 L-amino acids are inverted in chemically synthesized polypeptides, the resulting folded D-protein molecules are the mirror image of the natural L-proteins, contain left-handed alpha helices, and have reciprocal chiral functional properties.[16]

Prior racemic protein crystallography investigations, lead by Berg et al.[17] and later by Kim et al.[18], focused to show proof-of-concept experiments by the structure determination of racemic rubredoxin and monellin, respectively. Work on rubredoxin[19] served to show that racemic mixtures could crystallize in P<1bar>. However, the original idea was to determine the structure by de novo methods because centrosymmetry lends to quantized phases thereby simplifying the computation. The de novo determination of rubredoxin failed and the structure had to be determined by molecular replacement. Berg et al. also pointed out that centrosymmetric crystals should lend to excellent electron density and low errors in the final model due to the restriction of phases.[17] Later studies involved the structural determination of racemic monellin.[18] The monellin worked focused on the differences between D and L molecules and refinement in P1 or P<1bar>. Kim et al. concluded that the racemic monellin crystal was not truly centrosymmetric because of small differences between the mirror image forms of monellin in the crystal.[18]

Despite the initial scientific investigations put-forth, little is known about racemic protein crystallography. On theoretical grounds, racemic protein mixtures are argued to allow for more favorable crystallization[15], de novo structure determination[20], superior electron density[21], and structural models with low error[17]. None of these advantages have been experimentally worked out. As pointed earlier, crystal growth is sometimes the rate-limiting step for structural determination, consequently new methods are needed to favor crystallization. The de novo determination of a Xray structure eliminates the need for additional experiments like
MAD and MIR to determine protein phases[22], thereby saving time and money. Superior electron density will aid the structure determination process by allowing for rapid model building and structural models with low error help to provide correct structural details. These questions concerning racemic crystallography can be fully investigated when combined with total chemical protein synthesis.

**Scope of thesis:**

My thesis research aimed to address: (1) the expansion of chemical tactics available for the rapid and efficient preparation of larger and more complex proteins; and, (2) the use of racemic mixtures to crystallize and determine the Xray structure of proteins.

**Chapters 2-5** present new tactics for the synthesis of proteins. The new tactics allow for the convergent synthesis and selective desulfurization of proteins. Convergent synthesis of proteins increases the yields and size of proteins targets accessible to total chemical synthesis. The same method developed for the convergent synthesis of proteins allows for the preparation of cyclic proteins from more than one polypeptide. Selective desulfurization is a chemical tactic that allows for ligations at Ala residues, in the presence of protected Cys residues, thereby increasing synthetic versatility.

Some of the new methods described in the first part of the thesis were used in **Chapters 5-6** to prepare the mirror image forms of proteins needed for racemic protein crystallography. The mirror image form of a protein can only be prepared by chemical synthesis. Moreover, **Chapter 7 and Appendices E-F** documents the exploration of racemic protein mixtures as an avenue to determine novel Xray structures that are difficult to crystallize. We also show, for the first time, that direct methods and racemic crystallography can be used in a simple fashion to determine the Xray structure.

**Chapter 2** describes the convergent chemical synthesis of a novel topological analogue derived from the protein crambin. In particular, in order to make this knotted molecule a new chemical method was developed—kinetically controlled ligation. Structural studies by NMR and Xray are currently under way to elucidate the precise 3D structure of this novel protein.
Chapter 3 represents a detailed presentation of kinetically controlled ligation exemplified by the convergent synthesis of crambin from six unprotected polypeptide segments. Kinetically controlled ligation when combined with native chemical ligation allows for the facile preparation of proteins in high yields and has proved to be key to expanding the size of proteins accessible by total chemical synthesis.

Chapter 4 describes the selective desulfurization of polypeptides obtained from native chemical ligation. Raney nickel can be used to convert a Cys to an Ala residue in a selective fashion in the presence of protected Cys(Acm). To demonstrate the utility of this reaction EETI-II and amylin were synthesized.

Chapter 5 extends selective desulfurization to prepare a larger 94-residue protein from Mycobacterium tuberculosis of unknown tertiary structure. Both the L form of the protein and the enantiomeric D form were prepared to facilitate crystallization and Xray structure determination (see later chapters).

Chapter 6 presents an optimized chemical synthesis of the novel antifreeze protein isolated from Canadian snow fleas. Both mirror image forms of this protein were prepared, and were equally effective in inhibiting ice recrystallization.

Chapter 7 presents a new method for the facile determination of protein crystal structures by the racemic approach. We used the sfAFP protein as a test bed to show crystal formation is more likely to occur from a racemic mixture than a solution of L alone. A novel L-Se-sfAFP derivative was prepared and used to determine the Xray structure by (pseudo)racemic protein crystallography. This approach allowed the Xray structure determination from a quasi racemic solution where MAD phasing was used.

Finally, Chapter 8, summarizes the findings presented here and points toward future research that is needed to further our ability to synthesize and use chemistry to understand proteins at the molecular level.
**Appendix A** is a detailed Kent lab procedure for the preparation of unprotected peptides thioesters by Boc-SPPS. These peptides are later used in native chemical ligation.

**Appendix B** provides exact procedures for the convergent chemical synthesis of proteins by kinetically controlled ligation. To showcase the chemistry involved crambin was synthesized from four unprotected peptides.

**Appendix C** is a series a laboratory exercises developed for undergraduate organic chemistry aimed to teach Fmoc-SPPS. The lab exercises comprised of experiments for the preparation of amino methyl resin that was then used to synthesize YGGFL.

**Appendix D** describes a new procedure for the Boc-SPPS on-resin preparation of peptide arylthioesters for use in native chemical ligation.

**Appendix E** explores the use of racemic crystallization to determine the Xray structure of the toxin BmbKTX1. This protein target was previously reported to be recalcitrant to crystallization even at concentrations of 100mg/mL. Once crystals and diffraction data was obtained the structure was determined by direct methods.

**Appendix F** points toward the Xray structure determination of Kaliotoxin by the racemic method. In particular, preliminary data has been collected from a twinned crystal and attempts to optimize crystal growth are in-progress. The D and L forms of the protein were prepared from a modular three-piece one-pot approach.

**Appendix G** illustrates that non-natural amino acids can be incorporated into proteins by total chemical synthesis and then be used to answer biological questions. Insulin-life growth factor 1 (IGF-1) was prepared in which Gly7 was substituted with D-Ala and the effects were investigated.

**Appendix H** is a modern interpretation of the HIV-1 protease in complex with inhibitors. The study was undertaken by resolving the Xray structures at atomic-resolution.
Appendix I documents the synthesis of a monobody derived from the 10th type fibronectin domain (10FN3). This protein was prepared from three polypeptides using a one-pot approach and desulfurization.

Appendix J is a first attempt to study L-sfAFP by NMR by the incorporation of site-specifically isotope labeled Gly residues. The design, synthesis, and NMR characterization is reported.
References:


Chapter 2. Design and synthesis of a novel topological analogue of crambin

Crambin contains a critical salt bridge between the side chain $\delta$guanidinium moiety of Arg$^{10}$ and the $\alpha$carboxylate of C-terminal Asn$^{46}$. We have shown that this salt bridge guides the formation of correct disulfide bonds and contributes to the tightly folded overall structure of crambin.[1] The structural importance of the salt bridge led us to ask: can we use chemistry to replace the salt bridge with a covalent bond; and, will current state-of-the-art chemical methods allow the synthesis of this novel topological analogue? Moreover, will the reduced full-length cyclic topology fold to form the native disulfides? If the analogue designed here folds and shares a similar structure to native crambin, does this suggest a covalent bond can serve as protein salt bridge surrogate?

We envisioned that with an amide bond we could replace the salt bridge between side chain of Arg$^{10}$ and the C-terminal $\alpha$carboxylate of Asn$^{46}$ by mutating Arg$^{10}$ to Lys$^{10}$. The resulting amide, between the $\epsilon$amine group of Lys$^{10}$ and C-terminus as shown in Figure 2.1, would lend the desired complex structure of crambin. The replacement of a salt bridge with a covalent bond has never been done for a protein and is only possible by total chemical synthesis. This chemical transformation turns crambin into a complex interpenetrating polypeptide topology that is fascinating (Figure 2.1). First, the novel crambin molecule will have a start (N-terminus) and no end (i.e. no C-terminus). Second, the structure will have an extraordinary ‘knot’ topology in which two disulfide bonds lock the N-terminal nine residues, so that this linear polypeptide penetrates through the covalent polyamide ring. Our challenge was to design a practical chemical synthesis for this molecule and to determine the structure of this unprecedented protein.
**Figure 2.1.** A) A cartoon representation of crambin (PDB 1EJG) and highlighted in red is the salt bridge between Arg\(^{10}\) and the C-terminal "carboxylate of Asn\(^{46}.\) B) A graphic showing the mutation of Arg\(^{10}\) to Lys\(^{10}\) and possibility of forming an amide bond. C) A expanded view of (B). D) A freehand sketch of the polypeptide backbone (black line) in relation to the disulfide bonds (yellow lines). By forming an amide bond between the C-terminus and the epsilon amino group of the side chain of Lys\(^{10}\) (thick red line) the N-terminus threads through a polyamide ring and is held in position by disulfide bonds.

This novel protein topology is reminiscent of, but not identical to, the backbone N-to-C covalently cyclized plant proteins known as ‘cyclotides’.[2] Cyclotides are a newly discovered class of ribosomally-translated proteins that contain a cyclic polypeptide
backbone with the N- and C- termini joined by an amide bond.[3] In addition to the cyclic backbone, cyclotides have a so-called ‘cysteine knot’ topology, in which two disulfide bonds and two strands of polypeptide backbone form an embedded ring that is penetrated by a third disulfide bond. Cyclotide proteins have exceptional stability and have interesting biological activities that include anti-HIV and anti-microbial activities thereby making them an important class of targets to study.[4]

**Synthetic design.** Our first retrosynthetic strategy is shown in Figure 2.2 and is the simplest chemical route to prepare the molecule. This strategy relies on building a branched 46-residue Cys-polypeptide-thioester by stepwise solid phase peptide synthesis (SPPS)[5] and then cyclizing the protein by an intramolecular native chemical ligation (NCL)[6] of [Thr1-Lys10(Asn46-Cys16)-Ala15]-alkylthioester. In order to do this, we need to synthesize the [1-15]-thioester on the solid-phase resin, and then build residues (from Asn46 to Cys16) from the side chain of Lys10. The ε-amine group of Lys10 must be protected with a group stable to the conditions of Boc chemistry ‘in situ naturalization’ SPPS for the preparation of [1-15]-thioester, and that can then be selectively removed under conditions that leave the protected peptide resin intact. For these reasons, we chose the base labile Fmoc[7] group to protect the ε-amine group of Lys10. After the removal of the Fmoc-group, SPPS would be continued on the side chain of Lys10, starting from residue Asn46 and ending with Cys16.
Figure 2.2. Retrosynthetic strategy #1 for the total synthesis of our novel topological analogue of crambin. This route, although potentially the simplest, was considered unlikely to work because stepwise SPPS of the branched 46-residue crambin polypeptide seemed likely to be not feasible.[8] For this reason, it was replaced by an improved synthetic design (shown in Figure 2.3).
Synthetic Strategy #2. Not convinced that the initial synthetic design shown in Figure 2.2 was feasible, we designed a new approach that relied on the development of a new chemical tactic called ‘kinetically controlled ligation’ [9], which allows for the convergent synthesis of proteins. This new chemistry was developed because the native chemical ligation reaction cannot be used to condense two Cys-peptides both of which contain thioester moieties: cyclization and/or oligimerization would result in a complex mixture of products. Such a convergent synthetic approach requires independent control of the reactivity of both the cysteine and “thioester functionalities. While fully reversible N-terminal cysteine protection as a 1,3-thiazolidine-4-R-carboxylic acid (Thz) turned out to be straightforward[10], a similarly practical means for controlling the reactivity of the C-terminal “thioester had not been realized.

Figure 2.3 represents a second, convergent, retrosynthetic strategy that uses kinetically controlled ligation and native chemical ligation to transform crambin into a protein molecule of complex topology. This approach is referred to as strategy #2. The devised synthesis relies on two critical steps. [Thr$^{1}$-Lys$^{10}$ (Asn$^{46}$-Thz$^{40}$)-Ala$^{15}$]$^{a}$alky/thioester needs to be prepared by stepwise SPPS using Fmoc side-chain protected Lys$^{10}$ as described earlier. Second, two highly selective ligation reactions need to be carried out. The ligation reaction between [Thz$^{16}$-Gly$^{31}$]$^{a}$alky/thioester and [Cys$^{32}$-Thr$^{39}$]$^{a}$alky/thioester needs to be carried out to give [Cys$^{16}$-Thr$^{39}$]$^{a}$alky/thioester as the main product. The selective aspect of this reaction was based on the anticipated fact that a peptidyl-Gly-thioester would react much faster than a peptidyl-Thr-thioester under native chemical ligation conditions. Thus, the desired product will form and not the side products from cyclization or oligomerization of [Cys$^{16}$-Thr$^{39}$]$^{a}$alky/thioester. The second ligation reaction between [Thr$^{1}$-Lys$^{10}$ (Asn$^{46}$-Thz$^{40}$)-Ala$^{15}$]$^{a}$alky/thioester and [Cys$^{16}$-Thr$^{39}$]$^{a}$alky/thioester also needs to be carried out such that the main product is [Thr$^{1}$-Lys$^{10}$ (Asn$^{46}$-Cys$^{16}$)-Ala$^{15}$]$^{a}$alky/thioester. Again, this selective reaction needs to meet the same requirements of the first. If the desired full length branched polypeptide-thioester is obtained, known
chemistry would then be used to convert the Thz-moiet to a Cys- and to cyclize the polypeptide by native chemical ligation, followed by formation of the disulfide bonds of the target folded protein.
Figure 2.3. Retrosynthetic strategy #2 for the total synthesis of our novel protein, which was replaced by our improved design (shown in Figure 2.6). Note that some of
(Figure 2.3 cont.) the peptide segments are shown in the opposite of the conventional orientation, i.e. with the N-terminal at the right and the C-terminal at the left. This route had to be abandoned because of low reaction yields for the attempted selective ligation reactions A and B.

Data for this second synthetic route are given below (Figures 2.4 & 2.5). Strategy #2 had to be abandoned because there was insufficient selectivity of the native chemical ligation reactions based on differences in reaction rates for peptidyl–Gly-thioester and peptidyl–Ala-thioester with peptidyl–Thr thioesters. Thus, product mixtures were excessively complex and synthetic yields were low.

**Results for synthetic strategy #2.** The branched peptide [Thr¹-Lys¹⁰(Asn⁴⁶-Thz⁴⁹)-Ala¹⁵]-¹alky/θthioester, [Thz¹⁶-Gly³¹]-¹alky/θthioester, and [Cys³¹-Thr³⁹]-¹alky/θthioester was prepared SPPS and purified by preparative RP-HPLC.

The first selective native chemical ligation between [Thz¹⁶-Gly³¹]-¹alky/θthioester and [Cys³¹-Thr³⁹]-¹alky/θthioester resulted in polypeptide [Thz¹⁶-Thr³⁹]-¹alky/θthioester (Figure 2.4). The reaction was carried out by dissolving 8 micromole of each peptide in 4 mL of pH = 6.8 buffer containing 6 M GuHCl, 0.2 M phosphate, and 0.1% thiophenol (v/v); reaction was substantially complete in 4 hours. The Thz¹⁶ was converted to Cys¹⁶ by 0.2 M methoxyamine at pH 4 and the product was purified by HPLC.
Figure 2.4. Analytical LCMS traces for the selective native chemical ligation of \([\text{Thz}^{16}\text{-Gly}^{31}]^\circ \text{alkylthioester and [Cys}^{31}\text{-Thr}^{39}]^\circ \text{alkylthioester: A) } t = 0 \text{ and B) } t = 4 \text{ hrs. In (B), the main peak is the desired product [Thz}^{16}\text{-Thr}^{39}]^\circ \text{alkylthioester (ob = 3462.6 } \pm 0.5 \text{ Da, ca = 3461.8 Da). The peak labeled (\&) is the ligation catalyst thiolphenol. Peaks (1-2) are unknown byproducts. The chromatographic separations were carried out on Vydac C_4 2.1 X 150 mm column using a linear gradient of 5-65 \% buffer B over 15 min (buffer A = 0.1\% TFA in H_2O; buffer B = 0.08\% TFA in acetonitrile). The inset is the ESMS. Calculated masses were based on average isotope composition.}

The results from the second attempted selective native chemical ligation between [Thr^{1}\text{-Lys}^{10}\text{(Asn}^{46}\text{-Thz}^{40}\text{-Ala}^{15}]^\circ \text{alkylthioester and [Cys}^{16}\text{-Thr}^{39}]^\circ \text{-alkylthioester.}
*alky/thioester are shown in Figure 2.5. The reaction was on a small scale in which 0.36 umol of each peptide was dissolved in 200 uL of pH = 6.8 buffer containing 6 M GuHCl, 0.2 M phosphate, 0.1 % (v/v) thiolphenol, and 30 mM TCEP-HCl. The reaction was followed for sometime and resulted in the formation of full-length polypeptide [Thr\(^1\)-Lys\(^{10}\) (Asn\(^{46}\)-Thz\(^{40}\))-Thr\(^{39}\)]-*alky/thioester. However, excessive levels of side reactions occurred as indicated by multiple HPLC peaks in Figure 2.5C. At this point, the approach was abandoned in favor of strategy #3.
Figure 2.5. Analytical LCMS traces for the ligation between [Thr\textsuperscript{1}-Lys\textsuperscript{10}(Asn\textsuperscript{46}-Thz\textsuperscript{40})-Ala\textsuperscript{15}]\textsuperscript{-alkylthioester} and [Cys\textsuperscript{16}-Thr\textsuperscript{39}]\textsuperscript{-alkylthioester}: A) t = 0 and B) t = 8 hrs. C) The cyclic reaction product from (B). In (C), a number of products were
(Figure 2.5 cont.) formed and continued to get worse. Consequently, the approach was abandoned. The peak labeled (&) is the ligation catalyst thiolphenol. The chromatographic separations were carried out on Vydc C4 2.1 X 150 mm column using a linear gradient of 5-65 % buffer B over 15 min (buffer A= 0.1% TFA in H2O; buffer B = 0.08% TFA in acetonitrile). The inset is the ESMS. Calculated masses were based on average isotope composition.

**Synthetic Strategy #3.** The synthetic design that finally proved to be effective is shown in Figure 2.6, and is referred to as strategy #3. We sought to increase the reaction rate differences in the attempted selective ligation reactions. To achieve this increase in rate differences, we exploited the large reaction rate differences between peptide-\(^{\text{aryl}}\)thioesters and peptide-\(^{\text{alky}}\)thioesters. We expected that, *in the absence of added thiol*, a peptidyl-\(^{\text{aryl}}\)thioester would react much faster than the corresponding peptidyl-\(^{\text{alky}}\)thioester with a Cys-peptide. The difference in reaction rates can be used to control the outcome of a reaction involving two peptide-\(^{\text{thioesters}}, \text{so that only a single peptide-thioester ligation product is formed. We called this technique ‘kinetically controlled ligation’. The kinetically controlled ligation concept allows more flexible synthetic strategies and has already proved to be useful for the synthesis of a variety of proteins in our laboratory.\cite{11, 12}*

Strategy #3 relies on two critical steps. [Thr\(^1\)-Lys\(^{10}\)(Asn\(^{46}\)-Thz\(^{40}\))-Ala\(^{15}\)]-\(^{\text{aryl}}\)thioester needs to be prepared by SPPS, using Fmoc side-chain protected Lys\(^{10}\) as described earlier. Second, a kinetically controlled ligation reaction between [Thr\(^1\)-Lys\(^{10}\)(Asn\(^{46}\)-Thz\(^{40}\))-Ala\(^{15}\)]-\(^{\text{aryl}}\)thioester and [Cys\(^{16}\)-Thr\(^{39}\)]-\(^{\text{alky}}\)thioester needs to be carried out to give the full-length branched 46 residue polypeptide. At this stage, known chemistry is used to convert the Thz\(^{40}\) moiety to a Cys\(^{40}\), to cyclize by native chemical ligation to form the target cyclic polypeptide, and to then form the disulfides of the folded topological analogue protein molecule. We chose to protect the side chain thiols of internal Cys residue not at sites of ligation as Cys(Acm), because earlier test reactions with unprotected Cys residues resulted in lower reaction yields arising from internal thiolactone formation.
Figure 2.6. Our improved retrosynthetic design, strategy #3, for the total synthesis of the novel topological analogue of crambin by a convergent route using optimized kinetically controlled ligation.
**Results for synthetic strategy #3.** The branched peptide [Thr\[^1\]-Lys\[^{10}\](Asn\[^{46}\]-Thz\[^{40}\])\(-Ala\[^{15}\])\(-^aalky/thioester and the linear peptide [Cys\[^{16}\] -Thr\[^{39}\])\(-^aalky/thioester were prepared by stepwise SPPS, and purified by preparative RP-HPLC. [Thr\[^1\]-Lys\[^{10}\](Asn\[^{46}\]-Thz\[^{40}\])\(-Ala\[^{15}\])\(-^aryl/thioester was generated by the transthioesterification of the initial [Thr\[^1\]-Lys\[^{10}\](Asn\[^{46}\]-Thz\[^{40}\])\(-Ala\[^{15}\])\(-^alky/thioester using mercaptophenylacetic acid (MPAA), and then purified.[13]

As our design had anticipated, the ligation reaction between [Thr\[^1\]-Lys\[^{10}\](Asn\[^{46}\]-Thz\[^{40}\])\(-Ala\[^{15}\])\(-^aryl/thioester and [Cys\[^{16}\] -Thr\[^{39}\])\(-^alky/thioester resulted in formation of the full-length polypeptide [Thr\[^1\]-Lys\[^{10}\](Asn\[^{46}\]-Thz\[^{40}\])\(-Thr\[^{39}\])\(-^alky/thioester cleanly and in good yield (Figure 2.7 A & B). The kinetically controlled ligation reaction was carried out by dissolving 10.5 micromole of each peptide in 5 mL of pH = 6.8 0.2 M phosphate buffer, containing 6 M GuHCl and 30 mM TCEP-HCl. After the ligation reaction was complete, the Thz\[^{40}\] moiety was converted to Cys\[^{40}\] by treatment with 0.2 M methoxyamine.HCl at pH 4. A total of 9.4 micromole (90 %) of product peptide [Thr\[^1\]-Lys\[^{10}\](Asn\[^{46}\]-Cys\[^{40}\])\(-Thr\[^{39}\])\(-^alky/thioester was isolated by solid phase extraction (SPE). Then, 6.8 umol of this product was dissolved in 10 mL of 6 M GuHCl, 0.2 M phosphate pH 6.8, containing 150 mM (4-carboxymethyl)thiophenol (mercaptophenylacetic acid, MPAA), and 30 mM TCEP.HCl and in 3 hours the desired cyclized full-length polypeptide was obtained by native chemical ligation, as shown in Figure 2.7 C. Typically, native chemical ligations that involve a peptidyl-Thr-thioester are slow and may need as much as 24 hrs to give reasonable reaction yields, even using a high concentration of the MPAA thiol catalyst. However, for the case reported here the reaction rate was accelerated because of the favorable nature of the intramolecular reaction.[14] After preparative reverse phase HPLC, purified fractions were pooled to give 5.9 micromole (87%) of the desired peptide. The purification of the cyclic peptide was complicated because it eluted over a wide range of gradient elution conditions, behavior similar to that of hydrophobic membrane spanning peptides.[15] This
required the pooling of most fractions that contained product, even if some impurities were also present.

Figure 2.7. Analytical LCMS data for the synthesis of cyclic crambin[Thr^{1}-Lys^{10}(Asn^{46}-Cys^{40})-Thr^{30}] using synthetic strategy #3. The insets show electrospray MS data for the principal component. A) time = 0 h for the kinetically controlled ligation reaction, B) time = 4 h for the same reaction, which was complete. Prior to the analysis shown, the Thz^{40} - moiety had been converted to a Cys^{40} - by treatment of
(Figure 2.7 cont.) the crude reaction mixture with 0.2M MeONH$_2$.HCL ast pH 4. The observed and calculated masses for the principal component were: $ob = 5436.6 \pm 0.5$ Da, $ca = 5436.8$ Da. C) The cyclic reaction product from (B), after reverse phase purification. Masses: $ob = 4946.0 \pm 0.5$ Da, $ca = 4946.39$ Da. The chromatographic separations were performed on a self-packed Varian microsorb C4 2.1 X 50 mm column using a linear gradient of 5-65 % buffer B over 15 min at 40°C at a flow rate of 0.5 mL/min (buffer A = 0.1% TFA in H$_2$O; buffer B = 0.08% TFA in acetonitrile).

With cyclic crambin [Thr$^1$-Lys$^{10}$-(Asn$^{46}$-Cys$^{40}$)-Thr$^{39}$]$_2$(SH)$_2$(SAcm)$_4$ in hand the four Acm protecting groups were removed using AgOAc under standard conditions[11] to give the unprotected crambin [Thr$^1$-Lys$^{10}$-(Asn$^{46}$-Cys$^{40}$)-Thr$^{39}$]$_2$(SH)$_6$ (ob = 4662.0 $\pm 0.8$ Da, ca = 4662.4 Da). For a small-scale Acm removal reaction 0.56 micromoles (93%) was isolated starting from 0.6 micromoles and when the reaction was scaled up, slightly lower yields were obtained. Then 3.2 micromole (~15 milligrams) of this peptide was dissolved in 30 mL of pH = 7.9 folding buffer containing 8 mM cysteine, 1 mM cystine·2HCl, 100 mM Tris, 2 M GuHCl. Within 30 minutes, while stirring the reaction, a single product containing three disulfide bonds was formed, as confirmed by LCMS (Figure 2.8). This product had a molecular mass (calculated mass = 4656.4 Da, observed mass = 4656.6 $\pm 0.8$ Da) expected for the target protein topological analogue, and reflecting the loss of 5.4 $\pm 0.8$ Daltons consistent with the formation of the three disulfides. The product was purified by reverse phase HPLC to give 1.5 umol (47%) of desired product.

Discussion. We had anticipated that the folding of the full-length polypeptide might be difficult because the target molecule has a critical salt bridge replaced with a covalent bond. Surprisingly, folding proceeded smoothly. It is important to note that native crambin is an “inside out” globular protein in which the folded, disulfide-containing protein is more hydrophobic than the unfolded reduced polypeptide as indicated by relative elution times on reverse phase HPLC.[10] However, we found that the crambin [Thr$^1$-Lys$^{10}$-(Asn$^{46}$-Cys$^{40}$)-Thr$^{39}$]$_2$(SH)$_6$ and the final folded crambin topoligcal analogue with three disulfides have the same retention time on reverse phase HPLC. This suggests that the ‘inside out’ tertiary structure of the folded
crambin molecule had already formed prior to disulfide formation. This observation may allow for new experiments to understand the folding pathway of crambin. It may be possible to crystallize and determine the Xray and/or NMR structure of this folding intermediate.

**Figure 2.8.** Analytical HPLC data for the folding of fully unprotected cyclic crambin [Thr\(^1\)-Lys\(^{10}\) (Asn\(^{46}\)-Cys\(^{40}\))-Thr\(^{39}\)](SH)\(_6\). The insets show electrospray MS data for the principal component. A) Fully unprotected cyclic crambin (SH)\(_6\) (ob = 4662.0 ±0.8
(Figure 2.8 cont.) Da, ca = 4662.39 Da). B) Crude folded cyclic crambin (ob = 4657.2 ±0.7 Da, ca = 4656.39 Da). C) Folded cyclic crambin after reverse phase HPLC purification (ob = 4656.6 ±0.8 Da, ca = 4656.39 Da). The chromatographic separations were performed on a self-packed Varian microsorb C4 2.1 X 50 mm column using a linear gradient of 5-65 % buffer B over 15 min at 40 °C with a flow rate of 0.5 mL/min (buffer A= 0.1% TFA in H2O; buffer B = 0.08% TFA in acetonitrile).

Experiments are under way to determine the Xray or NMR structure of this novel synthetic crambin topological analogue. A preliminary two-dimensional TOCSY 1H NMR spectra showing the fingerprint region of the analogue is shown in the experimental section (Figure 2.12). Our results suggest that it is now possible by the use of chemistry to replace a critical salt bridge in a protein with a covalent bond as to give novel products with intriguing properties. This approach may also be used to engineer proteins that are super thermostable.
Experimental:

Analytical HPLC: Peptide compositions were confirmed by analytical reverse phase LC-MS using a gradient of acetonitrile versus 0.1% trifluoroacetic acid (TFA) in water. For the work reported in this paper unless otherwise noted, analytical HPLC was carried out as follows: Varian microsorb self-packed C4 2.1 X 50 mm column using a linear gradient of 5-65 % buffer B over 15 min at 40 °C with a flow rate of 0.5 mL/min (buffer A= 0.1% TFA in H2O; buffer B = 0.08% TFA in acetonitrile). The eluent was monitored at 214 nm, with on-line ion trap electrospray mass spectrometry (MS).

Peptide Segment Synthesis (peptide-α-carboxylate or peptide-α-thioester): Peptides were prepared manually by “in situ neutralization” Boc chemistry stepwise solid phase peptide synthesis,[5] on -OCH2-Pam-resins (free α-carboxyl peptides) or on HSCH2CH2CO-Leu-OCH2-Pam-resin (α-thioester peptides), or on 4-methylbenzhydrylamine-resin.[14] Side-chain protection for amino acids was as follows: Arg(Tos), Asn(Xan), Asp(OcHex), Cys(4-CH3Bzl), Cys(Acm), Glu(OcHex), His(Bom), Lys(2Cl-Z), Ser(Bzl), Thr(Bzl), Tyr(Br-Z). After completion of the chain assembly, the peptide-resin was N-alphadeprotected by treatment with TFA, and the peptide was cleaved from the resin support and the side chain protecting groups were simultaneously removed by treatment with anhydrous HF containing p-cresol (90:10, vol/vol) for 1 h at 0 °C. After complete evaporation of the HF under reduced pressure, crude peptide products were precipitated and triturated with chilled diethyl ether, and the peptide products were dissolved in 50% aqueous acetonitrile containing 0.1% TFA and lyophilized.

Boc-SPPS synthesis of [Thr1-Lys10(Asn46-Thz40)-Ala15]-α-alkylthioester: The sequence TTCys(Acm)Cys(Acm)PSIVAK(Thz-PDGYAN)SNFNA-COS-CH2-CH2-Arg-Arg-Leu-COOH was prepared to give a peptide with mass (observed mass (ob.)
2943.6 ± 0.3 Da, calculated average mass (ca.) 2944.09 Da. The analytical traces for the synthetic product are shown in Figure 2.9. The branched portion of the peptide was prepared after chain elongation of [Thr¹-Lys¹⁰(Fmoc)-Ala¹⁵]-acylthioester. The Thr¹ Boc group was removed and then the free amine was protected by reaction with N-2-chlorobenzoyloxycarbonyl succinimide (Z(2-Cl)-OSu). After confirming completion of the reaction by ninhydrin, the side chain Fmoc group of Lys¹⁰ was removed using 20 % piperidine in DMF at 0 C for 10 min. Then, the Asn⁴⁶ was coupled to the free amine and chain elongation was carried out as normal.
Figure 2.9. LC analysis of the crude peptide after HF cleavage (A). The LC-MS trace in (B) is the peptide after purification and the inset is the ESMS for the major peak. The retention times for (A) and (B) do not match because different gradients and columns were used. For the LC trace in (A) a linear gradient of 5-65 % buffer B over 60 min (C3 Agilent 4.6X150mm) and in (B) a linear gradient of 5-65 % buffer B over 15 min (C4 self-packed Varian microsorb 2.1X50mm).
Generation of [Thr\textsuperscript{1}-Lys\textsuperscript{10}(Asn\textsuperscript{46}-Thz\textsuperscript{40})-Ala\textsuperscript{15}]\textsuperscript{\text{*}alky/thioester} for kinetically controlled ligation: The crude peptide [Thr\textsuperscript{1}-Lys\textsuperscript{10}(Asn\textsuperscript{46}-Thz\textsuperscript{40})-Ala\textsuperscript{15}]\textsuperscript{\text{*}alky/thioester} was treated with 0.2 M mercaptothiophenylacetic acid (MPAA),[13] 6 M GuHCl, 0.2 M phosphate, 20 mM TCEP HCl, for 2-3 hrs at pH = 7 and then purified by preparative HPLC. The product was analyzed by LC-MS to give a mass of ob. 2580.6 ± 0.8 Da, ca. 2580.54 Da (Figure 2.10).

![Figure 2.10](image-url)
(Figure 2.10 cont.) material after 2 hrs. The peak label (&) is MPAA. The LC-MS trace in (C) is the product peptide after purification and the inset is the ESMS for the major peak. For LC analysis in (A) and (B) a C3 Agilent 4.6X150mm column was used with a linear gradient of 5-65 % buffer B over 20 min. For the LC-MS trace in (C), a linear gradient of 5-65 % buffer B over 15 min with column a C4 self-packed Varian microsorb 2.1X50mm column.

Boc-SPPS of [Cys$^{16}$-Thr$^{39}$]-$^*_{-}$alkythioester: The sequence

Cys$^{16}$RLPGTPEALCys(Acm)ATYTGcys(Acm)IIIPGAT-CO-S-CH$_2$-CH$_2$-CO-Arg-Arg-Ala-COOH was prepared by manual Boc chemistry ‘in situ neutralization’ stepwise SPPS and cleaved/deprotected using HF under the conditions described above. The peptide was analyzed by LCMS (Figure 2.11) to give a mass of ob. 3035.4 ± 0.8Da, ca. 3035.32 Da.
Figure 2.11. LC analysis of the crude peptide after HF cleavage (A). The LC-MS trace in (B) is the peptide after purification and the inset is the ESMS for the major peak. The retention times for (A) and (B) do not match because different gradients and columns were used. For the LC trace in (A) a linear gradient of 5-65 % buffer B over 60 min (C3 Agilent 4.6X150mm) and in (B) a linear gradient of 5-65 % buffer B over 15 min (C4 self-packed Varian microsorb 2.1X50mm).
**Preparative HPLC:** Peptides were purified on C4, C8, or C18 silica with columns of dimension 22 X 250 mm, 10 X 250 mm, or 10 X 100 mm. The silica used was TP Vydac or self-packed Varian Microsorb or Agilent Zorbax. Crude peptides (50-300 milligrams) were dissolved in 5% acetonitrile/95% (0.1%TFA in water) to a concentration of ~ 20 mgs/mL and loaded onto the prep column by pumping at a flow rate of 5-10 mL/min. After the non-peptidic material had eluted, as judged by the re-establishment of the 214nm baseline, the peptidic components were eluted at a flow rate of 10 mL per minute using a shallow gradient (e.g. 20%B-40%B over 60 minutes) of increasing concentrations of solvent B (0.1% TFA in acetonitrile) in solvent A (0.1% TFA in water). The exact gradient used was determined by the elution behaviour of the desired peptide, as assessed by prior analytical HPLC and confirmed by preliminary runs at low loading on the preparative column being used. Fractions containing the pure target peptide were identified by analytical LC and MALDI MS, and were combined and lyophilized.
2D $^1$H NMR. 2D-TOCSY spectra were taken on a Varian Inova 600 spectrometer using previously published protocols.[1] 2.0 milligrams of the crambin analogue was dissolved in 500 microliters of 75% $d^6$ acetone in water (v/v).

Figure 2.12. Two-dimensional TOCSY $^1$H NMR spectra showing the fingerprint region of the analogue.
References:


**Chapter 3.** Kinetically-controlled ligation for convergent synthesis of proteins

In the past decade, total chemical synthesis has proved to be a robust and reproducible method for making proteins of defined tertiary structure and full biological activity.[1-5] This success has been made possible by the introduction of chemical ligation methods, based on the chemoselective reaction of unprotected peptides under mild conditions in aqueous solution.[6]

Here we report a set of chemical tactics based on a novel principle – ‘kinetically-controlled ligation’ – that enables a highly practical, fully convergent strategy for the synthesis of a target protein. Our objective is the fully convergent synthesis of proteins from four or more peptide segments, using modern native chemical ligation methods. Convergent synthesis is inherently parallel and is thus more efficient.[7] In a convergent approach, the two halves of the target sequence are prepared from multiple peptide segments and condensed in a final step to give the full length polypeptide chain. In the fully convergent strategy for the chemical synthesis of a target polypeptide chain reported here, novel synthetic routes are used to make the two halves of the target polypeptide chain. Thus, the left-hand half of the target polypeptide is made by sequential C-to-N ligations in the presence of a potentially reactive moiety at the C-terminal, while the right hand half of the target sequence is made by unprecedented sequential ligations in the N-to-C direction to give a segment with a potentially reactive moiety at the N-terminal.

Native chemical ligation, the thioester-mediated covalent joining of unprotected peptide segments at a cysteine residue,[8] is the most practical and most widely used chemoselective ligation method. Until now, most synthetic proteins have been prepared from just two peptide segment building blocks. Recent improvements have been focused on the preparation of proteins by sequential ligation of three or four peptide segments.[1-3, 5, 9] For syntheses involving multiple segments, existing tactics for the use of native chemical ligation are suitable only for sequential ligations towards the N-terminal from a C-terminal Cys-peptide segment. An N-terminal Cys residue can be reversibly protected to give a PG-Cys-peptide1-thioester, so that reaction with a Cys-peptide2 gives only the desired PG-Cys-peptide1-peptide2; removal of the protecting group to give the
Cys-peptide1-peptide2 enables the synthesis to be continued by sequential ligation towards the N-terminal of the target polypeptide chain.\[4\] No corresponding set of chemical tactics currently exists for protecting the thioester moiety of a peptide-\textalpha\thioester, preventing N-to-C sequential assembly of the peptide segments by native chemical ligation, and also preventing fully convergent protein synthesis using this chemistry.\[10\]

The major challenge was to control the intrinsic dual reactivity of a bifunctional Cys-peptide2-\textalpha\thioester under native chemical ligation reaction conditions, so that it will react with a peptide1-\textalpha\thioester to yield only a single product (see Scheme 3.1A).\[11, 12\] Like most researchers who use native chemical ligation, until now we have routinely prepared relatively unreactive peptide-\textalpha\COSCH2CH2CO-Leu (i.e. alkyl) thioesters.\[13\] Typically, native chemical ligation of such a peptide-\textalpha\thioalkylester and a Cys-peptide has been carried out in the presence of thiophenol as a catalyst; the reactive species is assumed to be the peptide-\textalpha\thiophenylester.\[14\] We conjectured that a preformed peptide-\textalpha\thiophenylester would react with a Cys-peptide so much faster than a standard peptide-\textalpha\thioalkylester that, in the same solution under competitive reaction conditions in the absence of exogenous thiophenol, this large rate difference would make the standard peptide-\textalpha\thioalkylester effectively unreactive. This turned out to be the case.
Scheme 3.1. Key steps for the realization of convergent synthesis using native chemical ligation. A. Kinetically-controlled reaction of a peptide-α-thiophenylester with a Cys-peptide-α-thioalkylester to give a single ligation product. B. The conversion of a Thz-peptide-α-thioalkylester to a Cys-peptide-α-thioalkylester, without damage to the thioalkylester moiety.

In order to present concrete examples of the chemical challenges involved in the fully convergent synthesis of a protein from multiple peptide segments and to illustrate how we were able to effectively address those challenges, we used a synthesis of the model protein crambin from six peptide segments.[3, 15-17] The sequence of the target molecule and the synthetic design are shown in Scheme 3.2.
**Scheme 3.2.** Fully convergent synthesis of the model protein crambin from six peptide segments. Ligations #1 and #2 are kinetically-controlled and assemble the N-terminal half of the target sequence in the C-to-N direction; ligation #3 is also kinetically controlled, and with native chemical ligation #4, assembles the C-terminal half of the molecule in the N-to-C direction. Ligation #5 is the final native chemical ligation of the two halves of the target molecule.

The efficacy of our kinetic-control tactics for determining the outcome of a ligation reaction is shown in ligation #1 in the crambin synthesis (see **Scheme 3.2**, Box A), the reaction of Thz-ProSerIleValAlaArgSer-AsnPheAsnAla-athiophenylester with Cys-ArgLeuProGlyThrPro-GluAlaLeu-athioalkylester in pH 6.6 aqueous buffer in the absence of added thiophenol. The results are shown in **Figure 3.1A&B**. The ligation reaction was complete in one hour and gave the desired product Thz-ProSerIleValAlaArgSerAsnPheAsnAlaCysArgLeuPro-GlyThrProGluAlaLeu-athioalkylester in near quantitative yield. Only trace amounts of byproducts from undesired reaction of the alkyl thioester were formed (see [Inset] **Figure 3.1B**): <1.5% cyclized byproduct, <<1% oligomer. Similar results were obtained for the two other kinetically-controlled ligations in the crambin synthesis (**Scheme 3.2**, ligations #2 and #3; results are in, **Figures 3.3 & 3.4**).
Figure 3.1. Kinetically-controlled ligation of [Thz4-Ala15]-α-thiophenylester and [Cys16-Leu25]-α-thioalkylester. (A) t = 0. (B) t = 1 hour. The ligation gave almost exclusively the [Thz4-Leu25]-α-thioester, and the reaction was complete. B[inset]: (i) arrow indicates cyclized byproduct (cyclic-[Cys16-Leu25]; (ii) arrow indicates byproduct [Thz4-Leu25]-[Cys16-Leu25] α-thioalkylester. (C), Conversion of [Thz4-Leu25]-α-thioalkylester to [Cys4-Leu25]-α-thioalkylester, using 0.2M methoxyamine·HCl added directly to the ligation reaction mixture. The reaction was completed in two hours.

The other key chemical transformation for the use of native chemical ligation in a fully convergent synthesis is to be able to convert a Thz-peptide-α-thioalkylester to give a Cys-peptide-α-thioalkylester (Scheme 3.1B). This was accomplished by treatment with 0.2M methoxyamine·HCl. The thioester moiety was unaffected during the quantitative conversion
from the Thz-peptide-ðthioalkylester to the Cys-peptide-ðthioalkylester: compare the products shown in Figure 3.1C with the starting mixture in Figure 3.1B.

The last step in the convergent synthesis was the native chemical ligation of the two halves of the polypeptide chain, followed by folding and disulfide formation to give the synthetic crambin protein. The data are shown in Figure 3.2. A final purification gave high purity crambin (Figure 3.2C) in good yield.
Figure 3.2. The final step of the fully convergent synthesis of crambin, the conventional native chemical ligation of the two halves of the target polypeptide chain: [Thr1-Leu25]-
athioalkylester and [Cys26-Asn46]. (A) Shows the starting point (i.e. before the addition of
(Figure 3.2 cont.) thiophenol). Overnight reaction under standard native chemical ligation condition gave the full-length crambin polypeptide, which was then folded without purification. (B) Shows the total crude products from folding and disulfide formation of the unpurified full-length polypeptide ligation product by dilution to 1M guanidinium HCl in the presence of 8mM Cys and 1mM cysteine. In one hour, correctly folded crambin molecule was formed in near quantitative yield. (C) shows the folded crambin molecule after purification by preparative HPLC (observed mass = 4702.0 ± 0.8 Da calculated mass = 4702.4 Da). Reverse phase HPLC analyses were performed on a Vydac C4 column using a linear gradient (1-61%) of buffer B in buffer A over 15 min (buffer A = 0.1% TFA in water; buffer B = 0.08 % TFA in acetonitrile).

Several aspects of the kinetically-controlled ligation reactions used in the convergent synthesis of crambin are worthy of comment. Ligation #2 (Scheme 3.2) involves reaction of a peptide-thiophenylester that contains an internal Cys with a free thiol group. In this instance, the presence of a free thiol did not interfere with the formation of the desired product (see Figure 3.3); however, the ligation product underwent internal cyclization with the C-terminal thioester to form a thiolactone; this could be readily converted to a thioalkyl ester, as described in the legend to Figure 3.1. The nature of the C-terminal amino acid residue in a peptide-αthioester affects the relative reactivity of that peptide in native chemical ligation: hindered C-terminal amino acids such as Ile, Thr, or Val are much less reactive than, for example, a C-terminal Ala-αthioester. This needs to be taken into account in the design of a convergent synthesis, and can be used to advantage to help control the outcome of the reaction as we have done in the example presented here. In other work, we have successfully used kinetically-controlled ligation to join two peptides both of which have a C-terminal alanine (i.e. a peptide-Ala-αthiophenylester with a Cys-peptide-Ala-αthioalkylester), with results equally as good as those shown here (unpublished data).

The new tactics and chemistries described here are a major advance in the chemical synthesis of proteins. In our own laboratory, we are successfully using fully convergent synthesis based on kinetically-controlled ligation for an increasing number of protein targets, including erythropoietin, the HIV-1 protease, and a KChIP accessory protein. We believe that kinetically-controlled ligation will form the basis for a truly practical convergent chemical synthesis of proteins.[18]
**Experimental:**

**Peptide Synthesis.** Peptides were prepared manually by “in situ neutralization” Boc chemistry stepwise solid phase peptide synthesis,[19] on -OCH2-Pam-resins (free α-carboxyl peptides) or on HSCH2CH2CO-Leu-OCH2-Pam-resin (α-thioalkylester peptides).[13] Side-chain protection for amino acids was as follows: Arg(Tos), Asn(Xan), Asp(OcHex), Cys(4-CH3Bzl), Glu(OcHex), Lys(2-Cl-Z), Ser(Bzl), Thr(Bzl), Tyr(Br-Z). The 1,3-thiazolidine-4-carboxy (Thz) group was introduced to protect the N-terminal Cys of the middle peptide segments, and Boc-L-thiazolidine-4-carboxylic acid was used for peptide synthesis. After chain assembly was complete, peptides were deprotected and simultaneously cleaved from the resin support by treatment with anhydrous HF containing p-cresol (90:10, vol/vol) for 1 h at 0 °C. After evaporation of the HF under reduced pressure, crude products were precipitated and triturated with chilled diethyl ether, and the peptide products were dissolved in 50% aqueous acetonitrile containing 0.1% TFA. Peptide compositions were confirmed by LC-MS. [Thr1-Cys3]-α-thioalkylester (observed mass (ob.) 525.4±0.4 Da, calculated average mass (ca.) 524.6 Da); [Thz4-Ala15]-α-thioalkylester, (ob. 1491.4±0.5 Da, ca. 1491.8 Da); [Cys16-Leu25]-α-thioalkylester, (ob. 1257.2±0.5 Da, ca. 1257.6 Da); [Thz26-Gly31]-α-thioalkylester, (ob. 827.5±0.5 Da, ca. 828 Da); [Cys32-Thr39]-α-thioalkylester, (ob. 987.8±0.5 Da, ca. 988.3 Da); and [Cys40-Asn46] (ob. 738.4±0.5 Da, ca. 738.8 Da).

**Preparative reverse phase HPLC purification.** To purify synthetic peptides and ligation products, preparative HPLC was performed on a Waters Prep LC 4000 system by use of preparative Vydac C8 column (12μm, 2.2 × 25 cm) at flow rate of 10mL/min and by use of semi-preparative Vydac C4 and C8 columns (1cm x 25cm) at a flow rate of 5mL/min, with a gradient of 10-40% buffer B in buffer A over 60 min. Buffer A: 0.1% TFA in water; Buffer B: 0.08% TFA in acetonitrile. Fractions were collected across the expected elution time, combined based on HPLC and LC-MS analysis, and lyophilized. Conversion of a peptide-α-thioalkylester to a peptide-α-thiophenylester. Lyophilized peptide-α-thioalkylester from HF cleavage was used for the exchange reaction. The crude peptide was dissolved in pH 6.8 aqueous buffer containing 6 M guanidine hydrochloride and 0.2 %
thiophenol (vol/vol). Peptide concentration in the solution was ca. 0.2 mM. The reaction mixture was vigorously stirred for 12 hours with exclusion of air, and analyzed by LC-MS. After the conversion, the reaction was quenched (acidified to pH 3) by addition of aqueous 6N HCl. Residual thiophenol and oxidized thiophenol (diphenyl disulfide) were removed by extraction with diethyl ether. Acid quenching was necessary before the diethyl ether extraction procedure in order to prevent transthioesterification with residual HSCH2CH2CO-Leu that could not be extracted. The peptide-α-thiophenylester was purified by preparative HPLC, confirmed by mass spectrometry, and lyophilized.

Native chemical ligation was carried out under standard conditions:[8] pH 6.8, 200mM sodium phosphate buffer containing 6M guanidine hydrochloride, at a concentration of 2mM for each peptide, 1% (v/v) thiophenol was added to the reaction mixture.

Kinetically-controlled ligation was performed in pH 6.8, 200mM sodium phosphate buffer containing 6M guanidine hydrochloride, at a concentration of 2mM for each peptide in the absence of thiophenol. Experimental procedures for kinetically-controlled ligations in the crambin synthesis:

Ligation #1. Kinetically-controlled ligation of [Thz4-Ala15]-α-thiophenylester and [Cys16-Leu25]-α-thioalkylester was performed in pH 6.8, 200mM sodium phosphate buffer containing 6M guanidine hydrochloride, at a concentration of 2mM for each peptide. The reaction was complete within one hour. The ligation gave almost exclusively the [Thz4-Leu25]-α-thioalkylester. Observed mass of the major peak (2529.8±0.4 Da) was consistent with calculated mass for the desired product [Thz4-Leu25]-α-thioalkylester (2530.0 Da, average isotopes). In Figure 1 B, arrow (i) indicates the cyclized byproduct (cyclic-[Cys16-Leu25], observed mass = 1137.7±0.6 Da calculated mass = 1138.3 Da); arrow (ii) indicates the byproduct [Thz4-Leu25]-[Cys16-Leu25]α-thioalkylester (observed mass = 3567.6±0.9 Da calculated mass = 3568.2 Da) Quantitative conversion of to [Cys4-Leu25]-α-thioalkylester was carried out without affecting the thioester moiety, using 0.2M methoxyamine·HCl added directly to the ligation reaction mixture. The conversion was complete within two hours. The
desired [Cys4-Leu25]-αthioalkylester (observed mass of 2517.8±0.3 Da, calculated mass = 2518.0 Da) was purified by preparative HPLC as described above for use in ligation #2.

**Ligation #2.** Kinetically controlled ligation between [Thr1-Cys3]-αthiophenylester and [Cys4-Leu25]-αthioalkylester was performed in pH 6.8, 200mM sodium phosphate buffer containing 6M guanidine hydrochloride, at a concentration of 2mM for each peptide. The ligation product formed a peptide-αthiolactone from the transthioesterification with an internal cysteine residue. Addition of 200mM mercaptoethanosulfonate sodium salt (MESNA) into the reaction mixture converted the peptide-αthiolactone to the [1-25]peptide-αthioalkyl ester for use in the final native chemical ligation of the two halves of the target polypeptide. See Figure S1 for details.

**Ligation #3.** Kinetically controlled ligation reaction of [Thz26-Gly31]-αthiophenylester and [Cys32-Thr39]-αthioalkylester was performed in pH 6.8, 200mM phosphate buffer containing 6M guanidine hydrochloride, at a concentration of 2mM for each peptide. Reaction was complete within one hour. Native chemical ligation with the C-terminal peptide [Cys40-46] was carried out without purification of the product [Thz26-Thr39]-αthioalkylester. See Figure S2 for details.

**Native Chemical Ligation to give [1-46]crambin.** Native chemical ligation was carried out under standard conditions at pH 6.8, 200mM sodium phosphate buffer containing 6M guanidine hydrochloride, at a concentration of 2mM of each peptide, 1% (v/v) thiophenol was added to the reaction mixture. After overnight reaction, the reaction mixture was diluted 3-fold and allowed to stand for one hour. The correctly folded crambin synthetic protein containing three disulfides was isolated by preparative reverse phase HPLC as described above.

**Yields.** Isolated yields were: 71% for ligation #1 & subsequent conversion of Thz to Cys in the same reaction mixture; 50% for ligation #2; 80% for ligation #3 alone; 40% for ligation
#3 & subsequent ligation #4 in one-pot; 62% for ligation #5 and subsequent folding/disulfide formation in the same reaction mixture.

![Diagram](image-url)

**Figure 3.3.** The crude ligation product from ligation #2, Scheme 3.2, i.e. reaction of [Thr1-Cys3]-αthiophenylester and [Cys4-Leu25]-αthioalkylester. The ligation product formed a peptide-αthiolactone from the transtioesterification with an internal cysteine residue. Addition of 200mM mercaptoethanosulfonate sodium salt (MESNA) into the reaction mixture converted the peptide-αthiolactone to the peptide-αthioester shown, [Thz1-Leu25]-αthio(ethanosulfonate)ester, (Observed mass = 2746±1.0 Da; calculated mass = 2746.9 Da).
Figure 3.4. Crude products from ligation #3, Scheme 3.2, obtained from the reaction of [Thz26-Gly31]-cρthiophenylester and [Cys32-Thr39]-cρthioalkylester, (Observed mass = 1596.5±1.0 Da; calculated mass = 1596.6 Da.
References:


11. Under native chemical ligation reaction conditions, a Cys-peptide2-α-thioester will itself undergo multiple reactions to form cyclic{Cys-peptide2}, Cys-peptide2-Cys-peptide2-α-thioester, and other products of oligomerization and cyclization.


Chapter 4. Selective desulfurization of Cys in the presence of Cys(Acm) in polypeptides obtained by native chemical ligation

The most effective way to synthesize large peptides and proteins involves the use of native chemical ligation — the chemo- and regio-selective reaction of a peptide-thioester and a Cys-peptide, with both peptides in side chain unprotected form, resulting in a single product with a native amide bond at the ligation site.[1] The development of a synthetic strategy for the preparation of a protein by means of native chemical ligation relies on location of suitable Xxx-Cys ligation sites, spaced at intervals no greater than ~40 residues throughout the target amino acid sequence. However, Xxx-Cys sites in a protein’s polypeptide chain are oftentimes limiting: Cys residues are rare or even absent in many proteins, and when present may often be in unusable positions (e.g. not suitably spaced throughout the target sequence). For this reason, considerable efforts have been devoted to developing chemistry for amide-forming ligations at sites other than Xxx-Cys.[2-8]

A useful way to extend the utility of native chemical ligation to peptides and proteins without Cys residues was recently introduced by Yan & Dawson.[9] In their approach, native chemical ligation is done at Xxx-Ala sites, with a Cys residue used in place of the native Ala residue. Subsequent desulfurization of the ligation product with freshly prepared Raney nickel is used to regenerate the native Ala residue at the site(s) of ligation.[9] The method of Yan & Dawson has proven to be a very reliable and useful one.[10-12] We have used this native chemical ligation-global desulfurization approach to synthesize a number of sulfur-free proteins with great success.[13, 14] In all cases, we have obtained excellent yields of the target sulfur-free polypeptide chains.

Because of its great practical utility, we wanted to extend the use of native chemical ligation followed by desulfurization to a more versatile synthesis of Cys-containing peptide and protein targets. We set out to develop a set of chemistries that would allow native chemical ligation followed by selective desulfurization of Xxx-Cys to Xxx-Ala in the presence of side chain protected Cys (Scheme 4.1).

60
**Scheme 4.1.** General synthetic strategy for Cys-containing targets using native chemical ligation combined with selective desulfurization. (‘COSR’ represents an alkylthioester moiety).

This is an important current objective for chemical protein synthesis.[15] Although it has been generally assumed that selective desulfurization of Cys in the presence of side chain protected Cys is not possible,[9-12, 16] it has been reported that under the appropriate conditions organic thiols can be reduced in the presence of thioethers.[17] We chose to focus our efforts on the –CH$_2$NH-COCH$_3$
(‘acetamidomethyl’, Acm) side chain protecting group for Cys because it is amenable to Boc chemistry, is compatible with ligation chemistry, and conditions for Acm protecting group removal are well established.[18] Furthermore, we anticipated that
under the right conditions Cys could be desulfurized in the presence of Cys(Acm), as suggested in work published by Hilvert et al in their studies of selenoCys reduction.[19] Here we report a set of chemical tactics that enable the synthesis of Cys-containing peptides and proteins by means of native chemical ligation, at sites that will become Xxx-Ala, combined with selective desulfurization.

First, we did model studies with a small peptide to test if it was possible to selectively desulfurize Cys in the presence of Cys(Acm). We found that Cys residues can in fact be selectively desulfurized in the presence of Cys(Acm), in a practical and straightforward manner. Then, we combined native chemical ligation and selective desulfurization for the total chemical synthesis of the peptide hormone amylin[20], and the small protein EETI-II[21]. To our knowledge this is the first time that the selective desulfurization of Cys-containing peptides in the presence of Cys(Acm) has been used in the practical synthesis of peptides and proteins.

The peptide LYRACys(Acm)FGCKI (Figure 4.1) was prepared in a stepwise fashion using Boc chemistry solid phase synthesis. It was subjected to reduction using freshly prepared Raney nickel, under the following conditions: 0.2 M aqueous phosphate, 6M guanidine·HCl, pH 4, room temperature, 20 mM triscarboxyethylphosphine·HCl (TCEP). Cys was converted to Ala quantitatively in less than 2 hours as determined by LCMS. Only trace amounts of desulfurization of Cys(Acm) were observed (Figure 4.1).
**Figure 4.1.** Model study for the selective desulfurization of Cys to Ala in the presence of Cys(Acm). Analytical reverse phase HPLC chromatograms of: (i) starting peptide; (ii) products after treatment with Raney nickel. The peak labelled (a) is desulfurized Cys(Acm) and (b) is starting material. The chromatographic separations were done on a Vydac C4 2.1 X 100 mm column using a linear gradient of 5-65 % buffer B over 15 min at 40 °C with a flow rate of 0.5 mL/min (buffer A = 0.1% TFA in H2O; buffer B = 0.08% TFA in acetonitrile). These conditions were used in all subsequent analyses. The eluent was monitored at 214 nm, with on-line electrospray MS (Insets). The observed and calculated (using average isotopes) masses for each principal component were: i) observed (ob) = 1244.1 ±0.5 Da, calculated (ca) = 1244.5 Da; ii) ob = 1212.2 ±0.5 Da, ca = 1212.5 Da.

Next, we explored the synthesis of a 37 residue peptide hormone, [A24P,S27P,S28P]amylin; the three Pro substitutions have been shown to minimize aggregation.[22] Amylin has a C-terminal caboxamide and has one disulfide bond between Cys2 and Cys7, and is thus a suitable biologically active peptide in which to illustrate selective reduction of Cys in the presence of Cys(Acm). The synthetic strategy for the total chemical synthesis of this target peptide is shown in **Scheme 4.2.**
**Scheme 4.2.** Synthetic strategy used for the synthesis of [A24P, S27P, S28P] amylin. ['COSR' = alkylthioester]

**Figure 4.2** shows the data for the synthesis of the [A24P, S27P, S28P] amylin polypeptide. Native chemical ligation of the two peptides amylin[1-12]-cthioester and [Cys13-37]amylin gave the full-length polypeptide in high yield. The purified ligation product, that contained a single Cys residue in the presence of two Cys(Acm) residues, was subjected to Raney nickel reduction. After 6 hours, the reduction was complete and the product appeared as a single symmetrical peak with a mass decrease of 32.1 Da. Solid phase extraction (SPE) was used to isolate the desulfurized product, which was eluted with 0.1% TFA 50/50 acetonitrile/water. The Cys(Acm) protecting groups were removed and the disulfide bond formed by oxidation with iodine[23], giving a high yield of [A24P, S27P, S28P] amylin (see experimental).
Figure 4.2. Analytical HPLC data for the synthesis of [A24P,S27P, S28P]amylin. (Insets) Electrospray MS data for the principal component. A i) time = 0 h for the ligation reaction, ii) time = 8 h reaction complete. The observed and calculated masses for the product: ob = 4125.9 ±0.5 Da, ca = 4125.6 Da. B Data for the selective desulfurization of amylin: i) time = 0 for the reaction, ii) time = 6 h, showing complete desulfurization (peak (a) is an unrelated column contaminant). See Supporting Information for detailsof HPLC configurations used. Observed and calculated masses at
(Figure 4.2 cont.) each step were: i) \(ob = 4125.9 \pm 0.5 \text{ Da}, ca = 4125.6 \text{ Da}; ii) ob = 4093.8 \pm 0.5 \text{ Da}, ca = 4093.6 \text{ Da}.

Next, we undertook the total chemical synthesis of the protein EETI-II using native chemical ligation and selective desulfurization. EETI-II is a small trypsin inhibitor protein that contains three disulfide bonds.[21] The synthetic route for [Met7Nle]EETI-II[1-28] involves the selective desulfurization of residue Cys17 to give Ala17, in the presence of six Cys(Acm) residues, after the ligation of the EETI-II peptide segments [1-16]-c-thioester and Cys17-28 (Scheme 4.3).

![Scheme 4.3. Synthetic Strategy used for the total synthesis of EETI-II[1-28].](image)

Thus, this system is an excellent model to demonstrate the exquisite selectivity of the desulfurization reaction. Subsequently, the Acm s were removed and the product folded to form three disulfide bonds (see experimental). The analytical data for the synthesis of
the EETI-II polypeptide chain by native chemical ligation followed by selective desulfurization is shown in Figure 4.3. The ligation between EETI-II[1-16]-thioester and [Cys17-28]EETI-II took approximately 8 hours, and resulted in a 58% yield.

The purified ligation product was then desulfurized and isolated in 89% yield. The six Acm groups were removed with iodine, after which the kinetically-formed disulfides were reduced with DTT and the protein folded with formation of native disulfides using a glutathione redox couple. The folding reaction was monitored with LCMS and the mass of the oxidized product decreased by six Daltons, corresponding to the formation of three disulfides. The folded protein eluted earlier on reverse phase HPLC as previously reported (see experimental).[24]
Figure 4.3. Analytical HPLC data for the synthesis of [Met7Nle]EETI-II[1-28]. A) Ligation reaction: (i) time = 0; ii) time = 8 h, reaction complete. (Inset) Electrospray MS data for the principal product. Observed and calculated masses were ob = 3344.5 ±0.5 Da, ca = 3345.0 Da. B Selective desulfurization: i) time = 0 ; ii) time = 6 h 30 min; peak (a) is trace residual starting material. (Inset) Electrospray MS data for the
(Figure 4.3 cont.) principal components. Observed and calculated masses at each step were: i) \( \text{ob} = 3344.5 \pm 0.5 \text{ Da}, \text{ca} = 3345.0 \text{ Da} \); ii) \( \text{ob} = 3312.6 \pm 0.4 \text{ Da}, \text{ca} = 3312.9 \text{ Da} \).

In other studies we investigated the selective reduction of Cys in the presence of Met. We found that Cys can be quantitatively desulfurized to Ala in the presence of Met and Cys(Acm) using the mild conditions reported here (see experimental). This extends the original observations of Yan and Dawson.[9]

Efficient and selective desulfurization of Cys in the presence of Cys(Acm) is feasible as demonstrated by syntheses of the peptide hormone amylin and the small model protein EETI-II. With selective desulfurization, proteins can be built by native chemical ligation from disconnections at the more common Xxx-Ala ligation sites, even in peptide and protein targets that contain other Cys residues. Furthermore, as pointed out by Yan and Dawson, native chemical ligation can be used with any amino acid that has been modified to have a thiol moiety on the \( \beta \)-carbon, e.g. \([\beta\text{-SH}]\text{Phe}[9]\)

Subsequent desulfurization regenerates the naturally occurring amino acid at the ligation site. In this way, a variety of Xxx-Yyy sequences can be used as ligation sites. Native chemical ligation in combination with selective desulfurization for the total synthesis of Cys-containing polypeptides thus represents a very useful expansion of the scope of the chemistries for the total synthesis of large peptides and proteins.
**Experimental**

**Peptide Segment Synthesis (peptide-"carboxylate or peptide-"thioester)** Peptides were prepared manually by “in situ neutralization” Boc chemistry stepwise solid phase peptide synthesis[25], on -OCH2-Pam-resins (free "carboxyl peptides) or on HSCH2CH2CO-Leu-OCH2-Pam-resin ("thioester peptides)[26], or on 4-methylbenzhydrylamine-resin. Side-chain protection for amino acids was as follows: Arg(Tos), Asn(Xan), Asp(OhHex), Cys(4-CH3Bzl), Cys(Acm), His(Bom), Glu(OhHex), Lys(2Cl-Z), Ser(Bzl), Thr(Bzl), Tyr(Br-Z). After completion of the chain assembly, peptides were deprotected and cleaved from the resin support by treatment with anhydrous HF containing p-cresol (90:10, vol/vol) for 1 h at 0 °C. After complete evaporation of the HF under reduced pressure, crude peptide products were precipitated and tritutated with chilled diethyl ether, and the peptide products were dissolved in 50% aqueous acetonitrile containing 0.1% TFA.

**Analytical HPLC.** Peptide compositions were confirmed by analytical reverse phase LC-MS using a gradient of acetonitrile versus 0.1% trifluoroacetic acid (TFA) in water. For **ALL** the work reported in this paper, analytical HPLC was carried out as follows: Vydac C4 2.1 X 100 mm column using a linear gradient of 5-65 % buffer B over 15 min at 40 °C with a flow rate of 0.5 mL/min (buffer A= 0.1% TFA in H2O; buffer B = 0.08% TFA in acetonitrile). The eluent was monitored at 214 nm, with on-line electrospray mass spectrometry (MS). [*Note that the elution times in chromatograms A(ii) and B(i) in Figure 4.2 do not align because the capillary tubing on the HPLC instrument was replaced between these runs.*]

**Preparative HPLC:** Peptides were purified on C4, C8, or C18 silica with columns of dimension 22 X 250 mm, 10 X 250 mm, or 10 X 100 mm. The silica used was TP Vydec or self-packed Varian Microsorb. Crude peptides were loaded onto the prep column in ~10% acetonitrile/90% {0.1%TFA in water}, and eluted at a flow rate of 10
mL per minute with a shallow gradient (e.g. 20%B-40%B over 60 minutes) of increasing concentrations of solvent B (0.1% TFA in acetonitrile) in solvent A (0.1% TFA in water). Fractions containing the pure target peptide were identified by analytical LC and MALDI MS, and were combined and lyophilized. Because each peptide behaves differently, general procedures for preparative HPLC cannot be given.

**Native Chemical Ligation:** Ligation reactions were carried out under previously published conditions[27]: 200 mM sodium phosphate buffer containing 6 M guanidine hydrochloride, 20 mM TCEP, pH = 6.8, 2-4 mM for each peptide, 0.5% (v/v) thiophenol, purged and sealed under argon.

**Selective Desulfurization:** Raney nickel was prepared by the slow addition while stirring of 100 mg of NaBH₄ to 600 mg of Ni(OAc)₂(H₂O)₄ dissolved in 3 mL of deionized water. After 5 minutes, the reaction was filtered using a medium sintered glass frit and the solid activated Raney nickel washed with 200 mL of ultrapure water (18 megOhm-cm). The full amount of wet Raney nickel was added to 2-4 mg of peptide dissolved in 3 mL of 200 mM sodium phosphate pH 7 buffer containing 6 M guanidine hydrochloride and 30 mg of TCEP. The final pH ranged from 3-5 and was not adjusted. Each reaction was monitored by analytical HPLC and worked up once the starting material was consumed. The identity of the product was verified by LCMS. The product was isolated by spinning down the Raney nickel and recovering the supernatant. The pellet was washed three times with 200 mM sodium phosphate buffer containing 6 M guanidine hydrochloride to insure maximum yields.

**Figure 4.4.** Analytical HPLC of crude [A24P,S27P,S28P]amylin after removal of Acm groups and oxidation. (Inset: electrospray MS of the principal component. Masses: obsd = 3949.2±0.2 Da, calcd= 3949.4 Da (average isotopes)). The covalent structure of the target peptide is shown in the Figure (Top).
EETI-II Folding[24].
Reduced [Met7Nle]EETI-II(1-28) polypeptide was dissolved in 50 mM TRIS buffer containing 6 M guanidine hydrochloride at pH 7.8. The solution was diluted over 5 minutes with 20 mM TRIS buffer, 10 mM oxidized glutathione, 2 mM reduced glutathione, until a final concentration of 0.7 mg/mL of protein was reached. The folding was followed by LC-MS. The mass of the folded product had decreased by 6 Da, reflecting the formation of three disulfides.

**Figure 4.5.** Analytical HPLC data for folding EETI-II. (i) Reduced [Met7Nle]EETI-II(1-28) polypeptide. The shoulder labelled (a) is residual oxidized material from Acm removal. (ii) Folded EETI-II. (Insets: electrospray MS data for the principal components. The observed and calculated masses were: i) obsd = 2886.2±0.4 Da, calcd = 2886.4 Da, ii) obsd = 2880.0±0.4 Da, ca = 2880.4 Da.
Selective Desulfurization of Cys in the Presence of Met: The model sequence LACAPMLF was used as the crude peptide. Standard desulfurization conditions were employed as described above; the reaction was complete in less than 3 hours.

Figure 4.6. Analytical HPLC for the selective desulfurization of Cys in the presence of Met. (i) Time = zero. (ii) Time = 3 hours. The peaks labelled (a) are crude peptide adducts and peak (b) is starting material. (Insets) Electrospray MS of the main component in each chromatogram. The observed and calculated masses were: i) obsd = 864.8±0.2 Da, calcd = 865.1 Da, ii) obsd = 833.7±0.2 Da, calcd = 833.1 Da.
Desulfurization of EETI-II(1-30): the selective desulfurization reaction of Cys\textsuperscript{17} to native Ala\textsuperscript{17} in the presence of both Met and Cys(Acm). EETI-II(1-30) was prepared by the ligation of EETI-II(1-16)-\textsuperscript{16}thioester-Leu-Leu-Leu-Pro-COOH and Cys\textsuperscript{17}-30. The ligation product was purified by reversed phase HPLC and subjected to desulfurization under the standard conditions given above. After 3 hours reaction, the desulufirazation was complete, as shown by a mass change of 31.3 ± 0.9 Da.

\textbf{Figure 4.7.} Analytical HPLC of the selective desulfurization of Cys in the presence of both Cys(Acm) and Met. (i) is Time = 0. Peak (a) is +16 Da which is Met oxidation. (ii) Time = 3 hours. Peak (b) does not give an identifiable mass and peak (c) is an unrelated column artefact. (Insets) Electrospray MS of the principal components. The observed and calculated masses were: i) ob = 3547.1 ±0.4 Da, ca = 3547.2 Da, ii) ob = 3515.8 ±0.8 Da, ca = 3515.2 Da.
References:


Chapter 5. Total chemical synthesis by selective desulfurization of the mirror image form of the hypothetical protein Rv1738 from Mycobacterium tuberculosis

Mycobacterium tuberculosis (Mt) is a deadly human pathogen and infection can persist under the most aggressive treatments available.[1] A hypothetical protein called Rv1738 is highly expressed when Mt is latent after infection, which occurs under hypoxic conditions (personal communication Jessica Chaston). Rv1738 is also thought to be a part of the DosR regulation system—a two-component system in Mt involved in regulating the genetic response of Mt to hypoxia and nitric oxide. Scant scientific information exists concerning the role of Rv1738 (personal communication Jessica Chaston).

We set out to establish total chemical synthesis access to Rv1738 from Mt, and to determine the Xray crystal structure of this hypothetical protein, as part of a collaboration with Ted Baker’s lab at the University of Auckland, New Zealand. In particular, we wanted to use chemistry to prepare the mirror image form of Rv1738 (D-Rv1738), so that we could crystallize the protein from a racemic solution containing equal amounts of the L- and D-proteins. We adopted the racemic method for crystallization, which can enhance the ability to get useful crystals, because during the last two years recombinantly expressed Rv1738 (i.e. L-Rv1738) has proven to be recalcitrant to crystallization in the Baker laboratory. The racemic protein method has been shown, in some instances, to produce crystals under conditions when an L-protein does not (Chapter 7).

The total chemical synthesis of a protein requires the development of a synthetic strategy centered by the use of native chemical ligation, the amide-forming ligation reaction at Xxx-Cys sites of unprotected peptide segments.[2] The sequence of Rv1738 does not contain any of the Cys residues required for the native chemical ligation reaction (Figure 5.1a). We then considered preparing the protein by the use of selective desulfurization of polypeptides obtained from native chemical ligation. Selective desulfurization allows for native chemical ligation at an Xxx-Cys site, followed by the conversion of Cys to Ala in the presence of Met or Cys(Acm) residues (Chapter 4). With this approach, we developed a synthetic strategy by disconnecting the full-length
amino acid sequence at suitable Xxx-Ala sites into three polypeptides of roughly equal length (Figure 5.1b). The three peptides are to be ligated in a one-pot fashion[3] and later selectively desulfurized to give the desired Rv1738.

![Image](image.png)

**Figure 5.1.** (A) The 94-residue amino acid sequence for Mtb Rv1738. (B) The synthetic strategy used for the preparation of Mtb Rv1738. The first step includes native chemical ligation followed by the conversion of the N-terminal thiazolidine (1,3-thiazolidine-4-carboxylic acid) to Cys. This is followed by a second native chemical ligation step. Once the full-length polypeptide is obtained, selective desulfurization is used for the conversion of Cys to Ala, in the presence of Met.

To verify the practicality of the above synthetic strategy, and to optimize the synthesis, we prepared L-Rv1738; data is presented in the Experimental Section.
(below). The synthesis of D-proteins requires the use of optimized chemical reactions that provide high yield and purity of products. This is important because the building blocks (D-amino acids) are more expensive and not as pure as the corresponding L-amino acids.

We then undertook the total chemical synthesis of the D-enantiomorph of Rv1738. Peptide segments were synthesized, purified, and characterized as described in the Experimental section. Figure 5.2 shows the analytical data for the native chemical ligation of Rv1738 [Thz30-Arg65]-“thioalkylester and [Cys66-Tyr94]-COOH. The reaction was carried out in ligation buffer at pH = 7, 6 M Gu-HCl, 30 mM mercaptophenylacetic acid (MPAA)[4], 0.2 M phosphate, 20 mM TCEP-HCl—using 4.4 mM concentration of each peptide. The ligation was complete after overnight reaction, and the product [Thz30-Tyr94]-COOH was converted to [Cys30-Tyr94]-COOH at pH 4.0 by adding 0.2M MeONH₂·HCl directly to the reaction mixture.[3] After confirming that the reaction was complete by LCMS (Fig 5.2(C)), the peptide [Cys30-Tyr94]-COOH was isolated by SPE (solid-phase extraction) and lyophilized.

The synthesis was continued, by reacting Rv1738 [Met1-Lys29]-“thioalkylester with [Cys30-Tyr94]-COOH, at the same peptide concentrations and in the same ligation buffer used for the first ligation (above). After overnight reaction, full-length peptide was obtained as confirmed by LCMS. At this stage, Trp¹³(CH2) and Trp³⁴(CH2) were converted to Trp by the addition of 1 volume piperidine and 1 volume β-mercaptoethanol to the reaction mixture, and allowed to react for 1 hr at 0°C. The resulting reaction mixture was then diluted, adjusted to pH = 2, and purified by reverse phase HPLC (Fig 5.2(E)). The above reaction sequence was carried out on a 17.55 micromole of each reactant scale, and after purification 5.6 micromole was isolated (32 % yield).

The last step for the synthesis of D-Rv1738 requires selective desulfurization to convert Cys³⁰ and Cys⁶⁶ to Ala³⁰ and Ala⁶⁶ in the presence of Met¹ and Met⁷⁴ (Figure 5.2). 10 mgs of D-[A30C, A66C]Rv1738 was dissolved in 5 mL of desulfurization buffer (see experimental for composition) and allowed to stand while preparing Raney nickel. Raney nickel was prepared by adding NaBH₄ to Ni(II)acetate in water. The
black Raney nickel was then washed with water three times, and the solution of D-[A30C, A66C]Rv1738 was added. The reaction was monitored by LCMS. Desulfurization was complete in approximately 9 hours as indicated by a mass change of *minus ~*64 Daltons (Fig 5.2(F & G)). Upon close examination of the ESMS we detected Met desulfurization and an Asn deletion, in minor amount. This should not present a problem. A separate selective desulfurization reaction was done in parallel on the same scale. For both reactions, the Raney nickel was spun down and washed to isolate the reaction products. The two reactions (20 mgs) were then combined and purified by reverse phase HPLC to give full-length D-Rv1738 (16 mgs; 80 % yield).
**Figure 5.2.** Analytical LCMS traces for synthesis of D-Rv1738. The starting pooled purified peptides for the ligation between (B) [Thz30-Arg65]-thioalkylester (ob = 4329.4 ±0.5 Da, ca = 4329.7) and (A) [Cys66-Tyr94]-COOH (ob = 3295.6 ±0.5 Da, ca = 3295.5). These two peptides have the same retention time, so LCMS was taken for the peptides separately. C) The crude ligation product after the conversion of Thz to Cys to
(Figure 5.2 cont.) give [Cys30-Tyr94]-COOH (ob = 7393.9 ±0.8 Da, ca = 7394.0). The peak labeled (&) in (C) is the ligation catalyst MPAA. After solid phase extraction, the product in (C) was reacted with the peptide [Met1-Lys29]-thioalkylester (ob = 3504.5 ±0.5 Da, ca = 3504.6) shown in (D).

Figure 5.2 continued. (E) The purified Cys-containing (A30C, A66C) deprotected full-length material (ob = 10622.5 ±1.5 Da, ca = 10623.7). (F) After selective desulfurization with Raney nickel. The LC trace is shown for the purified full-length D-Rv1738. G) The ESMS averaged over the entire LC-peak shown in (F) (ob = 10559.7 ±1.5 Da, ca = 10559.9). All chromatographic separations were carried out on a Agilent
Conclusions. The total chemical synthesis of the D-enantiomorph of the hypothetical protein Rv1738 from *Mycobacterium tuberculosis* was achieved in reasonable yield and purity by a three segment, one-pot, selective desulfurization convergent strategy. This is to the best of our knowledge the first experimental verification that selective desulfurization of Cys residues works well with larger polypeptides. Full-length D-Rv1738 was prepared to give ample amounts of material for crystallization trials. The synthesis of D-Rv1738 resulted in 30 or more mgs of protein that is needed for the structure determination. Crystallization experiments are currently underway.
Experimental:

Peptide Segment Synthesis (peptide-“carboxylate or peptide-“thioester) Protected D-amino acids were purchased from Peptides International. D-peptides were prepared manually by stepwise Boc chemistry ‘in situ neutralization’ solid phase peptide synthesis [5] on a 0.2 mmol scale. on -OCH₂-Pam-resins (free “carboxyl peptides) or on HSCH₂CH₂CO-(L-)Xaa-OCH₂-Pam-resin (“thioester peptides). Side-chain protection for the D-amino acids was as follows: Arg(Tos), Asn(Xan), Asp(OcHex), Cys(4-CH3Bzl), Cys(Acm), His(Bom), Glu(OBzl), Lys(2Cl-Z), Ser(Bzl), Thr(Bzl), Trp(CHO), Tyr(Br-Z). After completion of the chain assembly, peptides were deprotected and cleaved from the resin support by treatment with anhydrous HF containing p-cresol (90:10, vol/vol) for 1 h at 0 °C. After complete evaporation of the HF under reduced pressure, crude peptide products were precipitated and triturated with chilled diethyl ether. Then, the peptide products were dissolved in 50% aqueous acetonitrile containing 0.1% TFA and lyophilized.

Table 5.1: The D-peptides prepared for the synthesis of the D-enantiomorph of Rv1738 and the observed and calculated masses for each peptide, as determined by LCMS, and yields for the purified peptides.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Calc. MS (average isotopes)</th>
<th>Obs. MS (Da)</th>
<th>Isolated** (mgs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MGGDQSDHVLQHW(formyl)TVDISIDEHE-GLTRAK-CO-S-CH₂-CH₂-(L)Leu-COOH</td>
<td>3504.56</td>
<td>3504.5 ±0.5</td>
<td>80.6</td>
</tr>
<tr>
<td>Thz-RLRW(CHO)REKELVGGLARL-NPADRNVEIGDELSVAR-CO-S-CH₂-CH₂--(L)Leu-COOH</td>
<td>4329.7</td>
<td>4329.4 ±0.5</td>
<td>73</td>
</tr>
<tr>
<td>Cys-LSDLGKRMLKVSTHD-IEAVTHQPARLLY- COOH</td>
<td>3295.5</td>
<td>3295.6 ±0.5</td>
<td>61.6</td>
</tr>
</tbody>
</table>

**Amount isolated after purification of peptide from the 0.2 mmol scale synthesis.
**Analytical HPLC.** Peptide compositions were confirmed by analytical reverse phase LC-MS using a gradient of acetonitrile versus 0.1% trifluoroacetic acid (TFA) in water. For **ALL** the work reported in this paper, analytical HPLC was carried out as follows: Agilent C3 4.6 X 150 mm column using a linear gradient of 5-65 % buffer B over 15 min at 40 °C with a flow rate of 1.0 mL/min (buffer A= 0.1% TFA in H2O; buffer B = 0.08% TFA in acetonitrile). The eluent was monitored at 214 nm, with on-line electrospray mass spectrometry (MS).

**Preparative HPLC:** Peptides were purified on C4, C8, or C18 silica with columns of dimension 22 X 250 mm, 10 X 250 mm, or 10 X 100 mm. The silica used was TP Vydac or self-packed Varian Microsorb or Agilent Zorbax. Crude peptides were loaded onto the prep column in ~10% acetonitrile/90% {0.1%TFA in water}, and eluted at a flow rate of 10 mL per minute with a shallow gradient (e.g. 20%B-40%B over 60 minues) of increasing concentrations of solvent B (0.1% TFA in acetonitrile) in solvent A (0.1% TFA in water). Fractions containing the pure target peptide were identified by analytical LC and MALDI MS, and were combined and lyophilized. Because each peptide behaves differently, general procedures for preparative HPLC cannot be given.

**Native Chemical Ligation:** Ligation reactions were carried out under previously published conditions: 200 mM sodium phosphate buffer containing 6 M guanidine hydrochloride, 20 mM TCEP, pH = 6.8, 2-4 mM for each peptide, 30 mM MPAA, purged and sealed under argon.[4] The one-pot sequential ligations were carried out on a 17.55 micromole scale of each of the three peptide segments. The Trp(CHO) formyl side chain protecting groups were removed by adding 1 volume piperidine and 1 volume β-mercaptoethanol. After purification, 5.6 micromole (60 mgs, 32% yield) of full length polypeptide was obtained.

**Selective Desulfurization:** Raney nickel was prepared by the slow addition while stirring of 200 mg of NaBH₄ to 1200 mg of Ni(OAc)₂(H₂O)₄ dissolved in 8 mL of deionized water in a 50 mL pp centrifuge tube. After 5 minutes, the reaction was spun
down and the solid activated Raney nickel washed with 120 mL of ultrapure water (18 megOhm-cm). The full amount of wet Raney nickel was added to 10 mg of full length polypeptide dissolved in 5 mL of 200 mM sodium phosphate pH 7 buffer containing 6 M guanidine hydrochloride, 100 mg of TCEP, and 20 mg of EDTA. The final pH was 3 and was not adjusted. The reaction was monitored by analytical HPLC and worked up once the starting material was consumed. The identity of the product was verified by LCMS. The product was isolated by spinning down the Raney nickel and recovering the supernatant. The pellet was washed three times with 200 mM sodium phosphate buffer containing 6 M guanidine hydrochloride to insure maximum yields. Two 10 mg reaction were done in parallel; the crude products were combined after work-up, and then purified. A total yield of 15 mgs was isolated and used for crystallization.
**Figure 5.3:** Purified synthetic product of L-Mtb Rv1738. This material was prepared on a test scale, before preparing D-Mtb Rv1738. Note the excessive levels of deletion peptide byproducts detected in the mass spectrum.
References:


**Chapter 6.** Total chemical synthesis of snow flea antifreeze protein (sfAFP) using modern chemical ligation methods: preparation of the protein enantiomers L-sfAFP and D-sfAFP

**Introduction.** A glycine-rich snow-flea antifreeze protein (sfAFP) was recently discovered.[1] sfAFP has no sequence homology with any known proteins, and no experimental structure has been reported for this interesting molecule. Scientific investigation has been limited by lack of material; difficulties have been reported for both the recombinant expression and for the isolation of the protein from natural sources. The absence of an experimental structure has lead researchers to propose a theoretical 3D model of the molecular structure of sfAFP.[2]

Here, we report the total chemical synthesis of the mirror image forms of sfAFP (i.e. L-sfAFP, the native protein; and D-sfAFP, the native protein’s mirror image). Total chemical synthesis is a useful and versatile way to prepare multiple-milligram quantities of highly pure protein.[3] Synthesis of sfAFP by native chemical ligation[4] is efficient, reproducible, and allows for the facile production of sfAFP analogues. We show that chemically synthesized sfAFP has the expected anti-freeze activity, and that the mirror image D-sfAFP protein displays the same anti-freeze properties. ‘Mirror image’ D-proteins are currently accessible only through chemical synthesis, and to date only a handful of D-proteins have been prepared.[5-8] The present work is the first synthesis of a D-protein to utilize modern chemical ligation methods.[3]

**Synthetic Design.** The 81-residue sfAFP polypeptide (Figure 6.1a) contains three Gly-Cys sites, each of which represents a potential retrosynthetic disconnection for assembly by native chemical ligation. Peptide building blocks were prepared by stepwise Boc chemistry solid phase peptide synthesis (SPPS).[9] We initially set out to assemble the sfAFP polypeptide from 3 peptide building blocks. However, this strategy was precluded by chronic side reactions in the synthesis by SPPS of segments containing Asp-Gly and Asn-Gly sites.[10] Such sites are abundant in the sfAFP sequence, and the observed level of byproducts complicated the preparation of larger peptide segments in acceptable purity. As a consequence, a four-segment approach
(Figure 6.1b) was used in order to give reasonable yields of purified peptide segments to be used as building blocks for the preparation of sfAFP.

**Figure 6.1.** (A) The 81-residue glycine-rich amino acid sequence for sfAFP. (B) The sequential condensation of four peptide segments synthetic strategy used for the preparation of sfAFP. Each step involved 1) native chemical ligation of two unprotected peptide segments, followed by 2) the conversion of the N-terminal thiazolidine (Thz) to Cys. After purification, the full-length synthetic polypeptide chain was folded with concomitant formation of two intramolecular disulfide bonds to form the sfAFP protein.

**Results.** Typical ligation conditions were as follows: 5-10 mM peptide, 10-30 mM 4-mercaptophenylacetic acid (MPAA),[11] 20 mM triscarboxyethylphosphine.HCl (TCEP), 6 M guanidinium chloride, 200 mM sodium phosphate, pH 6.8-7.0. Ligations were monitored by LC-MS, and were typically complete within 4 hours. Upon completion of each ligation, the N-terminal thiazolidine (1,3-thiazolidine-4-R-carboxylic acid) moiety was converted to L-cysteine by treatment with 0.2 M methoxyamine HCl at pH 4.0 for approximately 2 hrs.[12] Intermediate ligation products were either isolated by solid phase extraction (SPE) or purified by reverse phase HPLC prior to subsequent ligations, in order to avoid potential methoxyaminolysis of thioester peptides. The full-length reduced 81 residue
polypeptide chain was purified and then subjected to folding and formation of disulfide bonds.

We screened approximately ten sets of conditions to optimize folding and oxidation of the reduced sfAFP polypeptide. Folding conditions screened included varying temperature, redox systems, and chaotropes. We found that the best folding conditions were to treat the purified, full-length polypeptide with a redox couple consisting of 8 mM cysteine/1 mM cystine in pH 7.8 buffer at 4 °C for ~24 hours. These conditions reproducibly gave a good yield of a single product that co-eluted with the reduced polypeptide on reversed phase HPLC but that had a mass 4 Daltons lower, consistent with the formation of two intramolecular disulfide bonds (see Experimental). LCMS analyses of folded, purified L-sfAFP (prepared by ligation of peptides synthesized from L-amino acids) and D-sfAFP (prepared by ligation of peptides synthesized from D-amino acids) are shown in Figure 6.2. Overall synthetic yields ranged from 15 to 30%, depending on the number of intermediate purifications performed. Typical amounts prepared in a single synthesis were 35-100 milligrams. Analytical data for all stages of a representative synthesis is given in the Experimental section (below).
Figure 6.2. Analytical LCMS traces for synthetic folded L- and D-sfAFP. The chromatographic separations were carried out on a self-packed Varian microsorb C₄ 2.1 X 50 mm column using a linear gradient of 1-61 % buffer B over 15 min (buffer A= 0.1% TFA in H₂O; buffer B = 0.08% TFA in acetonitrile). Flow rate 0.5 mL per min with detection at 214nm. The inset is the on-line electrospray MS summed over the entire LC peak. Calculated masses were based on average isotope composition.

The CD spectrum obtained for synthetic L-sfAFP corresponded to that reported for the sfAFP isolated from natural sources.[1] As expected, the sfAFP enantiomers had equal and opposite CD spectra, within experimental error (Figure 6.3). The CD spectra suggest that the synthetic sfAFP consists of either random coil or polyproline type II helices (PP-II). We have since determined the Xray structure of sfAFP (which will be reported elsewhere) and found that the protein contains only PP-II secondary structure.
Figure 6.3. The CD spectra of the protein enantiomers D-sfAFP and L-sfAFP. CD spectra were recorded using a Jasco 715 instrument at room temperature by dissolving 0.02 mgs of D- or L-sfAFP protein in 200 microL of 50 mM phosphate buffer pH = 6.9. A 1 mm path length cell was used.

The antifreeze activity of our folded, synthetic materials was verified by an ice recrystallization inhibition assay.[13] This assay distinguishes antifreeze proteins from the rest of nature’s proteins by their unique ability to bind to ice surfaces and prevent the grain migration that causes ice recrystallization. Figure 6.4 shows that, as would be expected for an achiral phenomenon, native L-sfAFP and its mirror image D-sfAFP display identical ice recrystallization inhibition activity. We also conducted a control recrystallization inhibition functional assay that showed that the reduced sfAFP is devoid of antifreeze activity in this assay (see experimental). sfAFP must be folded for antifreeze activity.
Figure 6.4. Ice recrystallization inhibition assay (for details see experimental) of D and L-sfAFP at A) t = 0 and B) t = 20 hrs. The lanes are as follows: lane 1 is 100 mM NaCl, 10 mM phosphate, pH = 7.5, lane 2 is 25 mg/mL L-sfAFP, 100 mM NaCl, 10 mM phosphate, pH = 7.5, and lane 3 is 25 mg/mL D-sfAFP, 100 mM NaCl, 10 mM phosphate, pH = 7.5.

Discussion. The total synthesis of sfAFP reported here was carried out numerous times during a 12-month period to supply material for crystallization trials. More than 100 mg of folded, purified sfAFP was prepared from a single synthesis. We encountered difficulty in obtaining useful crystals of the L-sfAFP; consequently, we wanted to increase the possibility of crystallization by the use of a racemic mixture, as suggested by Yeates.[14] To that end, we undertook the total synthesis of the protein enantiomers D-sfAFP and L-sfAFP. Efficient, reproducible chemical access to a protein molecule is an absolutely critical prerequisite for the cost-effective preparation of mirror image proteins (D-amino acids are ~four times more expensive than L-amino acids, for a protein of typical composition). Preparation of multiple analogues was facilitated by the modular nature of the synthesis. In addition to the preparation of D-sfAFP, we have successfully applied our total synthesis to the preparation of two

---

§ It has been predicted that racemic protein mixtures will crystallize more readily than natural single enantiomer proteins.[14] We have experimentally verified this result for several protein systems (manuscripts in preparation).
distinct sets of site-specifically isotope-labelled sfAFP preparations for use in NMR experiments (to be reported elsewhere), and to the preparation of two different selenium-containing sfAFP analogues for use in anomalous dispersion Xray experiments (to be reported elsewhere). Both selenium analogues were prepared by ligation of non-native N-terminal peptides to the native 13-81 peptide, eliminating the need to re-synthesize unchanged portions of the sequence in each synthesis.

Efficient chemical access to sfAFP analogues, combined with knowledge of sfAFP’s crystal structure, will enable the systematic study of the molecular basis of sfAFP antifreeze activity. Finally, D-sfAFP antifreeze activity may have important practical applications because D-proteins are expected to be non-immunogenic and resistant to degradation by natural proteases.[5] Therefore, D-AFPs could potentially be used to prevent tissue damage that occurs during the freezing of organs for long-term storage.[15]
**Experimental:**

**Peptide Segment Synthesis (peptide-α-carboxylate or peptide-α-thioester)** Both D- and L-peptides were prepared by manual stepwise solid phase peptide synthesis using ‘in situ neutralization’ Boc chemistry.[9] Peptides were synthesized on a 0.4 mmol scale on -OCH$_2$-phenylCH$_2$CONHCH$_2$(`Pam’)-resins (α-carboxyl peptides) or on HSCH$_2$CH$_2$CO-Xaa-OCH$_2$-Pam-resin (α-thioester peptides).[16] Side-chain protection for amino acids was as follows: Arg(Tos), Asn(Xan), Asp(OcHex), Cys(4-CH$_3$Bzl), His(Bom), Lys(2Cl-Z), Ser(Bzl), Thr(Bzl). Where appropriate (i.e. for the three peptide-thioester segments), N-terminal cysteine was incorporated as 1,3-thiazolidine-4-R-carboxylic acid (Thz). After completion of the chain assembly, the N-alphaBoc group was removed by treatment with trifluoroacetic acid (TFA), the N-alpha-deprotected peptide-resin was thoroughly washed with DMF and DCM, then dried under a stream of nitrogen, and then the peptides were cleaved from the resin support and the side chain protecting groups simultaneously removed by treatment with anhydrous HF containing p-cresol (90:10, vol/vol) for 1 h at 0 °C. After complete evaporation of the HF under reduced pressure, crude peptide products were precipitated and triturated with chilled diethyl ether, and the peptide products were dissolved in 50% aqueous acetonitrile containing 0.1% TFA.

**Analytical LCMS.** Peptide compositions were confirmed by analytical reverse phase LC-MS using a gradient of acetonitrile versus 0.1% trifluoroacetic acid (TFA) in water. For all the work reported, unless otherwise noted, analytical HPLC was carried out as follows: Vydac C4 2.1 X 150 mm column using a linear gradient of 1-61 % buffer B over 15 min with a flow rate of 0.5 mL/min (buffer A= 0.1% TFA in H$_2$O; buffer B = 0.08% TFA in acetonitrile) at 40 °C. The eluent was monitored at 214 nm, and by on-line ion trap electrospray mass spectrometry (MS).
Preparative HPLC: Peptides were purified on C4, C8, or C18 silica with columns of dimension 22 X 250 mm, 10 X 250 mm, or 10 X 100 mm. The silica used was TP Vydac or self-packed Varian Microsorb or Agilent Zorbax. Crude peptides (50-300 milligrams) were dissolved in 5% acetonitrile/95% (0.1% TFA in water) to a concentration of ~ 20 mgs/mL and loaded onto the prep column by pumping at a flow rate of 5-10 mL/min. After the non-peptidic material had eluted, as judged by the re-establishment of the 214nm baseline, the peptidic components were eluted at a flow rate of 10 mL per minute using a shallow gradient (e.g. 20%B-40%B over 60 minutes) of increasing concentrations of solvent B (0.1% TFA in acetonitrile) in solvent A (0.1% TFA in water). The exact gradient used was determined by the elution behaviour of the desired peptide, as assessed by prior analytical HPLC and confirmed by preliminary runs at low loading on the preparative column being used. Fractions containing the pure target peptide were identified by analytical LC and MALDI MS, and were combined and lyophilized.

Native Chemical Ligation: Ligation reactions using purified synthetic peptide segments were carried out as previously described:[11] 200 mM sodium phosphate buffer containing 6 M guanidine hydrochloride, 20 mM TCEP, pH = 6.8, at a concentration of 5-10 mM for each peptide segment, using 30 mM 4-(carboxymethyl)thiophenol (‘MPAA’) as catalyst.[11] The ligation buffer had previously been purged with helium and the ligation reaction was carried out under argon. After the completion of each ligation, as judged by LC-MS of aliquots, methoxylamine hydrochloride (0.2 M) was directly added to the reaction mixture; the pH was lowered to 4.0. This chemical step converts the N-terminal Thz to Cys and is essentially complete in 2-4 hrs, as judged by analytical LC-MS of aliquots.

Representative synthesis of sfAFP: The synthesis described below was carried out on a 9.9 micromole scale of each peptide segment, and after folding/disulfide formation and purification 1.54 micromol (10.0 milligrams) of the final product was isolated (16 % yield).
**Synthesis of Peptide Segments:** The peptide building blocks (and corresponding masses) used in this synthesis were as follows: Thz$^1$-Lys-Gly-Ala-Asp-Gly-Ala-His-Gly-Val-Asn-Gly$^{12}$-CO-S-CH$_2$-CH$_2$-CO-Ile-Pro-COOH (ob $= 1395.3 \pm 0.5$ Da, ca $= 1395.4$ Da), Thz$^{13}$-Pro-Gly-Thr-Ala-Gly-Ala-Ala-Gly-Ser-Val-Gly-Gly-Pro-Gly$^{27}$-CO-S-CH$_2$-CH$_2$-CO-Leu-Pro-COOH (ob $= 1468.4 \pm 0.5$ Da, ca $= 1468.5$ Da), Thz$^{28}$-Asp-Gly-Gly-His-Gly-Gly-Asn-Gly-Asn-Pro-Gly$^{42}$-CO-S-CH$_2$-CH$_2$-CO-Ile-COOH (ob $= 1481.9 \pm 0.5$ Da, ca $= 1482.4$ Da), and Cys$^{43}$-Ala-Gly-Gly-Val-Gly-Gly-Ala-Gly-Gly-Ala-Ser-Gly-Gly-Thr-Gly-Val-Gly-Gly-Arg-Gly-Gly-Lys-Gly-Gly-Ser-Gly-Thr-Pro-Lys-Gly-Ala-Asp-Gly-Ala-Pro-Gly-Ala-Pro$^{81}$-COOH (ob $= 3025.0 \pm 0.5$ Da, ca $= 3025.2$ Da).

**L-[Thz$^1$-Gly$^{12}$]-thioester:** The peptide Thz$^1$-Lys-Gly-Ala-Asp-Gly-Ala-His-Gly-Val-Asn-Gly$^{12}$-CO-S-CH$_2$-CH$_2$-CO-Ile-Pro-COOH was prepared on a 0.4 millimole scale using standard SPPS protocols, as described above. For a typical preparative HPLC run, 240 mg of crude peptide was obtained after HF cleavage, and after purification and lyophilization 120 mg (50% yield) of the desired peptide was isolated: mass ob $= 1395.3 \pm 0.5$ Da, ca $= 1395.4$ Da (average isotopes).
Figure 6.5 A) LCMS analysis of the crude peptide after HF cleavage; the peak labeled (S) is p-cresol. The LC trace in (B) is the peptide after purification and (C) is the electrospray MS for the major peak. The retention times for (A) and (B) do not match because different gradients were used. For the LC trace in (A) a linear gradient of 5-65 \% buffer B over 15 min and in (B) a linear gradient of 1-61 \% buffer B over 15 min.

L-[Thz\textsuperscript{13}-Gly\textsuperscript{27}]-thioester: The peptide Thz\textsuperscript{13}-Pro-Gly-Thr-Ala-Gly-Ala-Ala-Gly-Ser-Val-Gly-Gly-Pro-Gly\textsuperscript{27}-CO-S-CH\textsubscript{2}-CH\textsubscript{2}-CO-Leu-Pro-COOH was prepared on a 0.4 millimole scale using standard SPPS protocols as given above. The 1.25 g of peptide-resin was cleaved and purified directly without lyophilization. Precise isolation yields were not recorded; however, at least 200 mgs of pure peptide was recovered after isolation: mass ob = 1468.4 ±0.5 Da, ca = 1468.5 Da (average isotopes).
**Figure 6.6.** A) LCMS analysis of the crude peptide after HF cleavage; the peak labeled (S) is p-cresol. The LC trace in (B) is the peptide after purification and (C) is the electrospray MS for the major peak. The retention times for (A) and (B) do not match because different gradients were used. For the LC trace in (A) a linear gradient of 5-65 % buffer B over 15 min and in (B) a linear gradient of 1-61 % buffer B over 15 min.

**L-[Thz$^{28}$-Gly$^{42}$]-thioester:** The peptide Thz$^{28}$-Asp-Gly-Gly-His-Gly-Gly-Asn-Gly-Gly-Asn-Gly-Asp-Pro-Gly$^{42}$-CO-S-CH$_2$-CH$_2$-CO-Ile-COOH was prepared on a 0.4 millimole scale using standard SPPS protocols as given above. For a typical preparative HPLC run, 200 mg of crude peptide was purified and after lyophilization 70 mg (35 % yield) of desired peptide was isolated: mass ob = 1481.9 ±0.5 Da, ca = 1482.4 Da (average isotopes).
Figure 6.7. A) LCMS analysis of the crude peptide after HF cleavage; the peak labeled (S) is p-cresol. The side product peak on the backside is minus 17 or 18 Da; this arises from the several Asp-Gly and Asn-Gly in the target peptide sequence. The LC trace in (B) is the peptide after purification and (C) is the electrospray MS for the major peak. The additional ion series in (C) is from fragmentation of the peptide in the MS.

**L-[Cys43-Pro81]COOH:** The peptide Cys<sup>43</sup>-Ala-Gly-Gly-Val-Gly-Gly-Ala-Gly-Gly-Ala-Ser-Gly-Gly-Thr-Gly-Val-Gly-Gly-Arg-Gly-Gly-Lys-Gly-Gly-Ser-Gly-Thr-Pro-Lys-Gly-Ala-Asp-Gly-Ala-Pro-Gly-Ala-Pro<sup>81</sup>-COOH was prepared on a 0.4 millimole scale using standard SPPS protocols as given above. For a typical preparative HPLC run, 200 mg of crude peptide was purified and after lyophilization 70 mg (35 % yield) of desired peptide was isolated mass ob = 3025.0 ±0.5 Da, ca = 3025.2 Da (average isotopes).
Figure 6.8. A) LCMS analysis of the crude peptide after HF cleavage; the peak labeled ($S$) is p-cresol. The LC trace in (B) is the peptide after purification and (C) is the electrospray MS for the major peak.

Ligation of [Thz$^{28}$-Gly$^{42}$]-thioester and [Cys$^{43}$-Pro$^{81}$]-COOH: Reaction was carried out at room temperature, using concentrations of 5 mM for each peptide, at pH 6.8 and 10 mM MPAA thiol catalyst.
**Figure 6.9.** Analytical HPLC traces of aliquots of the ligation of [Thz\(^{28}\)-Gly\(^{42}\)]-
\(\text{thioester and [Cys}^{43}\)-Pro\(^{81}\)-COOH: A) } t = 0 \text{ and B) } t = \text{ overnight. In (B), the product shown was obtained after treatment with 0.2 M methoxylamine hydrochloride for 2 hrs to give [Cys}^{28}\)-Pro\(^{81}\)-COOH (ob} = 4276.3 \pm 0.8 \text{ Da, ca} = 4276.4 \text{ Da). The peak labeled (\&) is the ligation catalyst MPAA. The reaction product was isolated by solid phase extraction (SPE).}

**Ligation of [Thz\(^{13}\)-Gly\(^{27}\)]-
\(\text{thioester and [Cys}^{28}\)-Pro\(^{81}\)-COOH: Ligation was carried out as described above. The product of this reaction was purified by RP-HPLC; in other syntheses we continued without isolation of this intermediate, thereby increasing overall yields.**
Figure 6.10. Analytical HPLC traces of aliquots from the ligation of [Thz^{13}-Gly^{27}]-a-thioester and [Cys^{28}-Pro^{81}]-COOH: A) t = 0 and B) t = overnight. In (B), the product shown was obtained after treatment with 0.2 M methoxylamine hydrochloride for 2 hrs to give [Cys^{13}-Pro^{81}]-COOH (ob = 5416.6 ±0.5 Da, ca = 5416.8 Da). The peak labeled (&) is the ligation catalyst MPAA.

Ligation of [Thz^{1}-Gly^{12}]-a-thioester and [Cys^{13}-Pro^{81}]-COOH: Ligation was carried out as described above. The full length reduced polypeptide was purified by reverse phase HPLC.
Figure 6.11. Analytical HPLC traces of aliquots taken from the ligation of [Thz\(^1\)-Gly\(^{12}\)]-"thioester and [Cys\(^{13}\)-Pro\(^{81}\)]-COOH: A) \(t = 0\) and B) \(t = \) overnight. In (B) the product shown was obtained after treatment with 0.2 M methoxylamine hydrochloride for 2 hrs to give [Cys\(^1\)-Pro\(^{81}\)]-COOH (ob = 6483.2 ±0.8 Da, ca = 6483.8 Da). The peak labeled (\&) is the ligation catalyst MPAA.
**Folding/Disulfide Formation for Synthetic sfAFP(Cys\(^1\)-Pro\(^{81}\)):** HPLC purified polypeptide (A) was folded by dissolving 2.2 micromol (14.6 mg) in 30 mL of pH = 7.8 folding buffer containing 8 mM cysteine, 1 mM cystiene·2HCl, and 50 mM phosphate. A single product (B) containing two disulfide bonds was formed within approximately 24 hours as confirmed by LCMS indicating a loss of 4 Da. All manipulations were carried out at 4 °C. After completion of the folding reaction, dialysis or HPLC was used to isolate the product. For the case reported here, the folding buffer was added to a 3500 MW cut-off dialysis bag and dialyzed extensively against water at 4 °C and then lyophilized to give 2.16 micromol (14 mgs) of material. If HPLC was used, standard purifications were carried out using a Vydac C18 10 X 250 mm column as described above.

*Figure 6.12. Folding/Disulfide formation for Synthetic sfAFP(Cys\(^1\)-Pro\(^{81}\)).*
**Recrystallization inhibition activity assays:** The activity assay done by Jennifer Dashnau at the University of Pennsylvania. Samples were loaded into 25 ul microcapillary tubes (Drummond Microcaps, Drummond Scientific Co., Broomall, PA) and each end was flame sealed. Then, samples were flash frozen for about 10 seconds in 2,2,4-trimethylpentane cooled with dry ice and immediately placed in a bath of the same solvent cooled to -6 degrees C by a jacketed beaker connected to a Fisher Isotemp 1016S circulating bath. Images were taken at 40X total magnification using a Nikon SMZ-2B microscope (Melville, NY) and a DCM35 Digital Microscope Camera (Hangzhou Huaxin IC Technology, Silicon Valley, CA) utilizing the software ScopePhoto 1.0 (Sopetek). Data are shown in Figure 6.4.

In a separate ice recrystallization inhibition assay (Figure 6.13) we used the following conditions with AFP I as a positive control. Conditions used were: 1) 10 mM sodium phosphate, pH 7.5 with 100 mM NaCl, 2) approximately 25 mg/ml AFP I, 3) approximately 25 mg/ml sfAFP reduced form, and 4) approximately 25 mg/ml snow flea AFP oxidized (i.e. folded/disulfide-containing) form. Solutions were examined at 0 hours and at 4 hours; ice crystal formation was observed at 4 hours for 1) and 3), but not for 2) and 4).

![Figure 6.13](image)

**Figure 6.13.** Ice recrystallization inhibition assay series at a) 0 hours, b) 4 hours, and c) difference image. Samples from left to right are: 1) 10 mM sodium phosphate, pH 7.5 with 100 mM NaCl, 2) approximately 25 mg/ml AFP I, 3) approximately 25 mg/ml snow flea AFP reduced form, and 4) approximately 25 mg/ml snow flea AFP oxidized form.
References:


cryoprotective properties of ice-active substances associated with Antarctic sea
chemical ligation: Expanded scope by using straightforward methodology.*
Proceedings of the National Academy of Sciences of the United States of
Chapter 7. Total chemical synthesis and racemic crystallography used to determine the novel X-ray structure of a snow flea antifreeze protein (sfAFP)

Abstract. Crystal growth is often the bottleneck for the modern determination of a protein structure.[1] A potential solution to this problem is provided by the prediction that a racemic mixture made up of a native L-protein molecule and its enantiomer (i.e. the mirror image form of the protein, the ‘D-protein’) would crystallize more readily than the native L-protein: racemic protein solutions are hypothesized to favor crystal growth because the space group P<1bar> is accessible and presents the highest probability of crystal formation for soluble globular proteins.[2] Here we report a test of this hypothesis by the use of total chemical synthesis[3] and racemic crystallization for determination of the X-ray structure of the snow flea antifreeze protein (sfAFP)[4]. Crystal formation from a racemic solution containing equal amounts of D-sfAFP and L-sfAFP occurred much more readily than for L-sfAFP alone. More facile crystal formation also occurred from a quasi-racemic mixture of D-sfAFP and L-Se-sfAFP, a chemical analogue that contains a selenium (as an additional -SeCH2- moiety) at one residue and thus differs slightly from the true enantiomer. X-ray diffraction data and multiple wavelength anomalous dispersion (MAD) phasing[5] from these quasi-racemate crystals were then used to determine the X-ray structure of the sfAFP protein molecule to a resolution of 1.2Å. The novel structure of the L-sfAFP was made up of six left-handed PII helices, stacked in two sets of three to form a compact brick-like structure with one hydrophilic face and one hydrophobic face. These results illustrate the utility of total chemical synthesis combined with racemic crystallography for determining the X-ray structure of a protein.

While X-ray structures have been determined for many thousands of distinct protein molecules[6], it is not infrequently observed that a particular protein may be difficult to crystallize or form poorly-ordered crystals. This has become one of the principal obstacles to structural genomics, the high throughput determination of protein X-ray structures.[1] Proteins found in nature (L-proteins) can only crystallize in one of
65 space groups (those lacking inversion and/or mirror symmetry), out of the total of 230 unique space groups that describe all possible crystal symmetries. The frequency of protein crystallization in different space groups is markedly non-uniform (and different than the distribution for small molecule crystals explained by the classic molecular close-packing theory of Kitagorodski), with most protein crystals belonging to one of just a few space-groups and with P2₁2₁2₁ being by far the most commonly observed. This preference for certain space groups has been explained by an entropic model. This same model predicts that for globular proteins, a racemic mixture will crystallize more readily. This is so because the space-group most favored on theoretical grounds, P<1bar>, contains inversion symmetry and is thus not accessible to natural L-proteins, but is in principle accessible to the achiral entity formed by an enantiomeric pair made up of a D-protein and the corresponding L-protein. The more facile crystallization of protein racemates has never been experimentally verified: only two proteins have been crystallized as racemates[7, 8], and in neither case was crystallization of the natural L-protein limiting.

The novel sfAFP protein studied here is the smaller of two protein isoforms isolated by ice-affinity purification of snow-flea homogenate.[4] Antifreeze proteins inhibit ice crystal growth and owe their antifreeze properties to processes still under investigation.[9] The 81 residue amino acid sequence of sfAFP (Chapter 6) is glycine rich, and is characterized by a repetitive Gly-Xaa-Xaa motif, the locations of the several Pro residues, and by two intramolecular disulfides. sfAFP has no known homologues (antifreeze protein or otherwise)[4], and no experimental molecular structure has been reported for the protein. Experimental structure determination has been precluded in the first instance by a simple lack of material – isolation of the natural product gives only microgram quantities, and recombinant expression is reportedly complicated[10].

Total chemical synthesis using modern ligation methods[3] reproducibly gave up to 100 milligram quantities of high purity sfAFP from a single laboratory-scale synthesis. The activity of our folded, synthetic material was verified by ice
recrystallization inhibition. Using the synthetic sfAFP, we set up extensive crystallization trials carried out over a period of six months with L-sfAFP alone, varying temperature, protein concentration, and the concentration of (NH₄)₂SO₄ or other precipitants. These screens provided no useful crystallization leads. Crystals that were obtained were of poor quality as evidenced by twinning. By screening ~30 carefully selected of these crystals grown by vapor diffusion from well solutions containing 1.7M ammonium sulfate as the precipitant, we were able to collect and index synchrotron diffraction data to sub-angstrom resolution. However, exhaustive attempts to optimize this result and to crystallize heavy atom derivatives for phase determination failed.

With the goal of obtaining the Xray structure of the sfAFP protein, we set out to explore the use of racemic crystallization[2, 11, 12] as an approach to obtaining high quality crystals for Xray diffraction studies. The mirror image forms of a protein molecule necessary for racemic crystallization can only be obtained by total chemical synthesis.[3, 13-15] Using chemistry to prepare a protein molecule allows not only for the preparation of the enantiomeric protein molecule but also for an almost unlimited range of analogues. Modern methods based on chemical ligation have enabled straightforward synthesis of proteins up to ~20kDa,[16] and recent innovations such as fully convergent chemical synthesis promise to significantly extend the size of proteins accessible by total synthesis.[17] The total chemical synthesis of the two enantiomorphs, D-sfAFP and L-sfAFP was accomplished using modern chemical ligation methods.

In contrast to the attempted crystallization of L-sfAFP, when using a racemic solution containing equal amounts of D-sfAFP and L-sfAFP we were amazed to find that crystals were obtained in ~50% of the 194 conditions initially explored using the commercially available Hampton index, crystal screen and crystal screen II. Figure 7.1 shows simultaneous crystallization experiments under representative conditions for L-sfAFP alone and for a racemic solution of D- and L-sfAFP. Diffraction data from a sampling of the crystals formed in our initial screen revealed that the protein racemate
crystallized in space-group P<1bar>, with unit cell parameters a = 28.6 Å, b = 32.4 Å, c = 59.8 Å, α = 88.7°, β = 89.2°, γ = 73.4° and with two L-sfAFP molecules in the asymmetric unit (Figure 7.2).

**Figure 7.1** Racemic crystallization of [D-sfAFP + L-sfAFP]. Protein crystals appeared rapidly for a racemic mixture of D- sfAFP and L-sfAFP, but not for L-sfAFP alone, when screened under a standard set of crystallization conditions at 23°C. Representative examples are shown in this Figure. Xray diffraction was used to verify that the crystals were in fact protein. The crystallization conditions are a subset of the main findings reported here and are as follows: A. 0.1 M Bicine pH = 9.0, 2% v/v 1,4-Dioxane, 10% w/v polyethylene glycol 20,000 B. 0.1 M Bis-Tris pH = 6.5, 1.15 M ammonium sulfate C. 0.1 M Tris pH = 8.5, 25% w/v polyethylene glycol 3350.

The next step was to solve the structure of the sfAFP molecule. We chose to solve the structure using experimental phases from the established multiple wavelength anomalous dispersion (MAD) method. To measure anomalous differences, we collected data from a selenium-containing quasi-racemate, in which only one enantiomer contained a selenium atom. (Footnote: For centric structures, Friedel’s law still holds in the presence of anomalous scattering, and it is not possible to measure anomalous
differences. However, if only one enantiomer in a quasi-racemate contains an anomalously scattering atom, anomalous differences can be observed). To obtain the quasi-racemate, we co-crystallized D-sfAFP and L-Se-sfAFP, an sfAFP analogue in which Asn$^{11}$ was replaced with seleno-Cys alkylated with bromoacetamide (‘pseudo-Se-Asn’; see Experimental for details). The facile production of crystals by racemic crystallization was critical for obtaining the quasi-racemate - we made extensive efforts to grow crystals of the L-Se-sfAFP protein alone, with no success. By contrast, quasi-racemic D-sfAFP and L-Se-sfAFP crystallized readily under a variety of conditions from a solution containing equal amounts of D-sfAFP and L-Se-sfAFP. Diffraction data were collected to a resolution of 1.2Å. The data had pseudo-P1(bar) symmetry and the structure was solved in space group P1 using MAD data collected at three wavelengths (see Experimental). Figure 7.2 shows the D-sfAFP and L-Se-sfAFP molecules that form the asymmetric unit in the quasi-racemate.
Figure 7.2. The quasi-racemate crystal packing, showing the D-sfAFP molecule (purple ribbon) and the L-Se-sfAFP (green ribbon), with the ‘pseudo-Se-Asn’ at position 11 shown as CPK solid spheres colored by atom type (C, green; O, red; N, cyan; Se, yellow). A. Unit cell with two D-sfAFP/L-sfAFP pairs located at positions 0, 1/2b, 1/2c, and 0,0,0, in the asymmetric unit. The D-sfAFP contains right-handed polyproline type II (PPII) helices; the L-Se-sfAFP contains left-handed PPII helices. B. View of one D-sfAFP/L-sfAFP located at 0, 1/2b, 1/2c. Side chains are included in ball-and-stick mode.

The resulting high resolution Xray structure of sfAFP is shown in Figure 7.3. The protein fold observed in the Xray structure of sfAFP appears to be generally similar to the structural model recently proposed by Davies et al[10]. It consists of six left-handed PP-II helices, stacked in two groups to form an oblong brick-shaped molecule. The
proline residues define turns linking the helices, and the N-terminal of the protein is stabilized by two intramolecular disulfides. One face of molecule is apolar, and the other polar. The implications of this novel protein fold for antifreeze activity have been discussed[10]. It has been hypothesized that in antifreeze proteins a flat, apolar face interacts snugly with the highly ordered water molecules found at the surface of an ice crystal, presenting the polar (hydrophilic) face to the surrounding solution and inhibiting further growth of the ice crystal[20, 21]. Interestingly, we observed an array of highly ordered water molecules on the apolar face of the sfAFP protein (Figure 7.3). This observation is consistent with the proposed mechanism for inhibition of ice crystal growth.[10]

Figure 7.3. The xray structure of snow flea antifreeze protein (sfAFP). Atomic
(Figure 7.3 cont.) resolution (1.2 Å) Xray structure in the P1 space-group of sfAFP, obtained from quasi-racemate crystals of D-sfAFP and L-Se-sfAFP. A. SigmaA-weighted 2Fo-Fc electron density map (blue, 1.5 sigma level) of the proposed ice-binding surface encompassing residues 17-21, 47-51, 74-78. Five ordered water molecules are shown (magenta spheres) with electron density shown at 1 sigma. B. Cartoon of the backbone fold of sfAFP. Amino acid side chains are shown as sticks. The dashed box indicates the region shown in A (above). C. The main structural elements of sfAFP are six anti-parallel polyproline type II (PPII) (left-handed) helices, stacked in two groups and joined by five reverse turns and interlocked by a complex hydrogen bond network. The N-terminal half of the protein contains two intramolecular disulfide bonds between residues Cys$^1$-Cys$^{28}$ and Cys$^{13}$-Cys$^{43}$. D. Surface map highlighting sfAFP’s amphiphilic character. A 180 degree rotation (along the long axis) shows the contrasting surfaces: apolar (left); polar (right). E. Two views related by a 90° rotation of highly-ordered first shell water molecules (shown in magenta) interacting with the sfAFP apolar face backbone amide carbonyls and nitrogens (dashed magenta lines). Water molecules (magenta spheres) are within hydrogen bonding distance (2.7-3.0 Å) of the sfAFP backbone and each other (solid magenta lines). Xray statistics are given in Table 7.1. Backbone RMSD deviations between molecules for L and D-sfAFP are given in Table 7.2.

The resulting structural model from the quasi-racemate was used to solve by molecular replacement the structure of L-sfAFP using the original 0.98Å diffraction data. The L-sfAFP molecular structure in the P2$_1$ crystal is identical to that observed in the quasi-racemate. Table 7.3 gives the Xray statistics for the P2$_1$ L-sfAFP data set. The 2Fo-Fc electron density map obtained at 0.98 Å is shown in Figure 7.4 (A). Diffraction data were also collected from crystals of the true racemate P1<bar> {D-sfAFP + L-sfAFP}. Table 7.4 gives the Xray statistics for the {D-sfAFP + L-sfAFP} data set. The 2Fo-Fc electron density map obtained for the true racemate {D-sfAFP + L-sfAFP} at 1.0 Å is shown in Figure 7.4(B).
**Figure 7.4.** SigmaA-weighted 2Fo-Fc electron density map (blue, 1.5 sigma level) of the proposed ice-binding surface encompassing residues 17-21, 47-51, 74-78 for a) P2$_1$ L-sfAFP and b) P1<bar> {D-sfAFP + L-sfAFP}. Five ordered water molecules are shown (magenta spheres) with electron density shown at 1 sigma.

**Conclusions.** What is the likely impact of racemic protein crystallography? Racemic protein crystallography, enabled by modern synthetic methods, has the potential to overcome the crystal growth bottleneck. This approach only requires the knowledge of the amino acid sequence and the ability to chemically synthesize the protein. After preparation of the mirror image forms of sfAFP, we were able to determine the structure within weeks, in contrast to the six months we had spent trying to crystallize L-sfAFP alone. Moreover, we were able to grow diffraction quality crystals from a simple sparse matrix screen thereby eliminating the need for extensive fine screen optimization. Racemic protein crystallography, using a quasi- or true racemate approach, also provides an effective way to crystallize novel protein analogues that may be difficult to crystallize. It is not uncommon for a protein analogue to present a formidable crystallization challenge, even more so than the native protein.
In principle, the presence of inversion symmetry should extend the size of molecular structures accessible by modern dual-space iteration direct methods (for introduction to direct methods, see Stout and Jensen). The largest protein xray structure solved by direct methods is (insert largest structure), and the modern Shake and Bake algorithm has a reported size limit of \(~1000\) non-hydrogen atoms in the asymmetric unit. Neither of the racemic protein structures reported to date were solved by direct methods [7,8], and further experimental work is necessary to determine the practical size limits for the direct solution of centric protein crystal structures.

Since our initial success with racemic and quasi-racemic sfAFP, we have applied the method of racemic crystallization to the preparation of 5 additional protein racemates. While it is impossible to prove complete generality of the method, we believe that preparation of mirror image proteins for racemic crystallization will prove to be a powerful tool in structural biology.
**Experimental:**

**Total chemical synthesis of sfAFP and analogues.** The 81 residue sfAFP polypeptide chain was synthesized by native chemical ligation of four unprotected peptides (Thz$^1$-12-thioester; Thz$^{13}$-27-thioester; Thz$^{28}$-42-thioester; Cys$^{43}$-81) as described in Chapter 6. L-Se-sfAFP was prepared by incorporation of ‘pseudo-Se-Gln’ into the Thz$^1$-12-thioester peptide. The LC-MS of L-Se-sfAFP is shown in **Figure 7.5.** Boc-‘pseudo-Se-Gln’-OH for use in peptide synthesis was prepared by reducing L-selenocystine with sodium borohydride, alkylating at selenium with 2-bromoacetamide, and N$^\text{"{a}}$-acylating with Boc$_2$O.

![Analytical LCMS traces for synthetic folded L-Se-sfAFP](image)

**Figure 7.5.** Analytical LCMS traces for synthetic folded L-Se-sfAFP. The chromatographic separations were carried out on a Vydac C$_4$ 2.1 X 150 mm column using a linear gradient of 1-61 % buffer B over 15 min (buffer A= 0.1% TFA in H$_2$O; buffer B = 0.08% TFA in acetonitrile). The inset is the ESMS summed over the entire LC peak. Calculated masses were based on average isotope composition.

**sfAFP activity assay.** Full details are given in Chapter 6. Ice recrystallization inhibition assays were carried out at minus 6 °C, using AFP I as a positive control. Conditions used were: 1) 10 mM sodium phosphate, pH 7.5 with 100 mM NaCl, 2) approximately 25 mg/ml AFP I, 3) approximately 25 mg/ml sfAFP reduced form, and 4) approximately 25 mg/ml snow flea AFP oxidized (i.e. folded) form. Solutions were
examined at 0 hours and at 4 hours; ice crystal formation was observed at 4 hours for 1) and 3), but not for 2) and 4).

**Protein crystallization.** Lyophilized sfAFP was dissolved in water at 10-40 mg/mL and used directly for crystallizations. Crystallization screening conditions were based on the commercially available Hampton index, crystal screen and crystal screen II. Crystals were grown by vapor diffusion in hanging drops near 23°C. The drops were generated by mixing 1 µl of protein solution with 1 µl of reservoir solution and placed against 1ml of reservoir solution. Racemic solutions of sfAFP were prepared in water at 38 mg/mL (19 mgs of D-sfAFP and 19 mgs of L-sfAFP) and 19 mg/mL (9.5 mgs of D-sfAFP and 9.5 mgs of L-sfAFP). Simultaneous sparse matrix room temperature screens were carried out for the racemic protein solutions at both concentrations. Crystals appeared after 1 day, and after 10 days, 93 of the 194 sets (~48%) of the conditions explored had crystals suitable for use in diffraction experiments. No crystals were formed from the L-sfAFP solutions using the same set of conditions. The same approach was used for crystallization of the quasi-racemate in which suitable crystals for diffraction appeared after 1 day in a number of conditions. Diffraction data for the quasi-racemate was collected from a crystal grown from 0.2 M sodium malonate, pH 7, and 20% w/v polyethylene glycol 3350. Diffraction data for L-sfAFP was collected from a crystal grown from 0.1 M Bis-Tris pH = 5.5, 1.8 M ammonium sulfate. Diffraction data for the racemate was collected from a crystal grown from 0.1 M Bis-Tris pH = 6.5, 1.15 M ammonium sulfate.

**Data collection for the quasi-racemate.** For low temperature data collection, selected crystals were transferred to the cryoprotectant (reservoir solution plus 20% (v/v) glycerol) for a few seconds and flash-frozen in liquid nitrogen. Three-wavelength anomalous (L-Se /D quasi-racemate) X-ray data was collected at 100K at APS (Argonne National Laboratory beamline 23ID equipped with a MARCCD 300 detector). Images were processed and scaled with HKL2000.[23]
**Data collection for the racemate.** For low temperature data collection, selected crystals were transferred to the cryoprotectant (reservoir solution plus 20% (v/v) glycerol) for a few seconds and flash-frozen in liquid nitrogen. Three-wavelength anomalous (L-Se /D quasi-racemate) X-ray data was collected at 100K at APS (Argonne National Laboratory beamline 23ID equipped with a MARCCD 300 detector). Images were processed and scaled with HKL2000.[23]

**Data collection for L-sfAFP.** For low temperature data collection, selected crystals were transferred to the cryoprotectant (reservoir solution plus 20% (v/v) glycerol) for a few seconds and flash-frozen in liquid nitrogen. X-ray data was collected at 100K at APS (Argonne National Laboratory beamline 19BM equipped with a SBC2 detector). Images were processed and scaled with HKL2000.[23]

**Determination of X-ray structure from the quasi-racemate.** Data were collected to a resolution of 1.2 Å. The L-Se-sfAFP /D-sfAFP quasi-racemate structure was solved by MAD.[5] The heavy atom search was performed in SHELXD[24] using X-ray data collected at 12.6607 eV peak energy and 3.5 Å resolution cut-off. MAD phasing and density modification were performed with use of SOLVE and RESOLVE, respectively.[25] The model building performed with the RESOLVE basic script was complete with use of ARP/WARP program.[26] The electron density and model examinations were done using TURBO-FRODO.[27] The restrained positional and anisotropic B-factor refinement was done in REFMAC5.[28] R_free was monitored by setting aside 5% of the reflection as a test set. Crystal data and X-ray data statistics for the quasi-racemate data set are listed in Table 7.1. Backbone RMSD (Å) deviations between molecules in the quasi racemate are presented in Table 7.2.

**X-ray structure determination of L-sfAFP.** Data were collected to a resolution of 0.98 Å. The structure was determined by molecular replacement using the program MOLREP[29] with L-sfAFP from the quasi-racemate as a starting model. The rigid body refinement, restrained positional and anisotropic B-factor refinement, was done in
REFMAC5.[28] The electron density and model examinations were done using TURBO-FRODO.[27] \( R_{\text{free}} \) was monitored by setting aside 5% of the reflection as a test set. Crystal data and X-ray data statistics for the L-sfAFP data set are listed in Table 7.3. The coordinates and structure have been deposited in the Protein Data Bank with entry code 2PNE.

**X-ray structure determination of the racemate.** Data were collected to a resolution of 1.0 Å. Since the crystal packing of the quasi-racemate is the same as the true racemate, we proceeded to rigid-body refinement with two L-sfAFP monomers in the asymmetric unit (chains A and B) from the quasi-racemate. The rigid body refinement, restrained positional and anisotropic B-factor refinement, was done in REFMAC5.[28] The electron density and model examinations were done using TURBO-FRODO.[27] \( R_{\text{free}} \) was monitored by setting aside 5% of the reflection as a test set. Crystal data and X-ray data statistics for the racemate data set are listed in Table 7.4.
Table 7.1. The X-ray data collection and refinement statistics for the quasi-racemate sfAFP. *Highest resolution shell is shown in parenthesis. **Based on maximum likelihood.

<table>
<thead>
<tr>
<th>Crystal 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
</tr>
<tr>
<td>Space group</td>
</tr>
<tr>
<td>(a, b, c) (Å)</td>
</tr>
<tr>
<td>(a, b, c) (°)</td>
</tr>
<tr>
<td>Wavelength</td>
</tr>
<tr>
<td>Resolution (Å)</td>
</tr>
<tr>
<td>(R_{sym} or R_{merge})</td>
</tr>
<tr>
<td>(I/I)</td>
</tr>
<tr>
<td>Completeness (%)</td>
</tr>
<tr>
<td>Redundancy</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
</tr>
<tr>
<td>Resolution (Å)</td>
</tr>
<tr>
<td>No. reflections</td>
</tr>
<tr>
<td>(R_{work} / R_{free})</td>
</tr>
<tr>
<td>No. atoms</td>
</tr>
<tr>
<td>Protein</td>
</tr>
<tr>
<td>Ligand/ion</td>
</tr>
<tr>
<td>Water</td>
</tr>
<tr>
<td>B-factors</td>
</tr>
<tr>
<td>Protein</td>
</tr>
<tr>
<td>Ligand/ion</td>
</tr>
<tr>
<td>Water</td>
</tr>
<tr>
<td>R.m.s. deviations</td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
</tr>
<tr>
<td>Bond angles (°)</td>
</tr>
</tbody>
</table>

Ramachandran plot statistics:
- Most favored (%) 93.2 (L), 91.9 (D)
- Additionally allowed (%) 6.8 (L), 8.1 (D)
- Generally allowed (%) 0 (L), 0 (D)
Table 7.2. Backbone RMSD (Å) deviations between molecules in the quasi-racemate: Molecules A and B are L-sfAFP and molecules C and D are inverted D-sfAFP.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C-inverted</th>
<th>D-inverted</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.520</td>
<td>0.165</td>
<td>0.514</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>0.564</td>
<td>0.203</td>
<td></td>
</tr>
<tr>
<td>C-inverted</td>
<td></td>
<td></td>
<td></td>
<td>0.540</td>
</tr>
</tbody>
</table>

Table 7.3 The X-ray data collection and refinement statistics for L-sfAFP:

<table>
<thead>
<tr>
<th></th>
<th>L-sfAFP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P2₁</td>
</tr>
<tr>
<td>Cell parameters</td>
<td>a = 16.70 Å, b = 74.28 Å, c = 17.69 Å, β = 102.2°</td>
</tr>
<tr>
<td>Beamline/Detector</td>
<td>APS 19BM/SBC2</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.97918</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>37 - 0.95 (1.02 - 0.98, 0.98 - 0.95) *</td>
</tr>
<tr>
<td>I/σ(I)</td>
<td>17.9 (3.3, 3.0)</td>
</tr>
<tr>
<td>Rmerge</td>
<td>0.059 (0.26, 0.25)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>90.6 (86.8, 46.0)</td>
</tr>
<tr>
<td>Average redundancy</td>
<td>3.3</td>
</tr>
</tbody>
</table>

| Reflection Statistics          | 23,939                     |
| R_work                         | 0.143 (0.29, 0.33)         |
| R_free                         | 0.166 (0.25, 0.25)         |
| Rms deviations                 |                            |
| Bonds (Å)                      | 0.017                      |
| Angles (degrees)               | 1.675                      |
| Asymmetric Unit                |                            |
| No. protein residues           | 81                         |
| No. waters                     | 113                        |
| Average B factor (Å²)          | 6.5                        |

| Ramachandran plot statistics  |                            |
| Most favored (%)              | 94.6                       |
| Additionally allowed (%)       | 5.4                        |
| Generally allowed (%)          | 0                          |

*Values for highest resolution shells are given in parentheses
Table 7.4. The X-ray data collection and refinement statistics for racemic sfAFP.

<table>
<thead>
<tr>
<th>Description</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>D, L-sfAFP</td>
</tr>
<tr>
<td>Cell parameters</td>
<td>P1&lt;bar&gt;</td>
</tr>
<tr>
<td>a</td>
<td>28.603 Å</td>
</tr>
<tr>
<td>b</td>
<td>32.404 Å</td>
</tr>
<tr>
<td>c</td>
<td>59.849 Å</td>
</tr>
<tr>
<td>α</td>
<td>88.69°</td>
</tr>
<tr>
<td>β</td>
<td>89.18°</td>
</tr>
<tr>
<td>γ</td>
<td>73.41°</td>
</tr>
<tr>
<td>Beamline/Detector</td>
<td>APS 23ID/MARCCD300</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.97918</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>27- 0.98 (1.02 - 0.98) *</td>
</tr>
<tr>
<td>I/σ(I)</td>
<td>17.4 (1.18)</td>
</tr>
<tr>
<td>Rmerge</td>
<td>0.084 (0.27)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>100 (38.9)</td>
</tr>
<tr>
<td>Average redundancy</td>
<td>4.3</td>
</tr>
<tr>
<td>Refinement Statistics</td>
<td></td>
</tr>
<tr>
<td>Reflections used</td>
<td>92,050</td>
</tr>
<tr>
<td>Rwork</td>
<td>0.21</td>
</tr>
<tr>
<td>Rfree</td>
<td>0.23</td>
</tr>
<tr>
<td>Rms deviations</td>
<td></td>
</tr>
<tr>
<td>Bonds (Å)</td>
<td>0.014</td>
</tr>
<tr>
<td>Angles (degrees)</td>
<td>1.613</td>
</tr>
<tr>
<td>Asymmetric Unit</td>
<td></td>
</tr>
<tr>
<td>No. protein residues</td>
<td>162</td>
</tr>
<tr>
<td>No. waters</td>
<td>295</td>
</tr>
<tr>
<td>Average B factor (Å²)</td>
<td>6.8</td>
</tr>
<tr>
<td>Ramachandran plot statistics</td>
<td></td>
</tr>
<tr>
<td>Most favored (%)</td>
<td>93.2</td>
</tr>
<tr>
<td>Additionally allowed (%)</td>
<td>6.8</td>
</tr>
<tr>
<td>Generally allowed (%)</td>
<td>0</td>
</tr>
</tbody>
</table>

*Values for highest resolution shells are given in parentheses.
**Acknowledgements** We thank W. DeGrado (U Penn), T. Sosnick, V. Pingali, V. Torbeev, Li-Wei Hung, C. Weeks, S. Anderson, N. Sanishvili, M. Becker and A. Moglich for discussions and technical advice. The use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under Contract No. W-31-109-Eng-38. Portions of this work were performed at the SBC-CAT and GM/CA CAT located at Sectors 19BM and 23ID, respectively, of the Advanced Photon Source. SBC-CAT is supported by the U.S. Department of Energy, Office of Biological and Environmental Research, under Contract No. W-31-109-ENG-38.
References:


Chapter 8. Significance and Future Perspectives

The underlying theme of my research has been to improve methods for the total chemical synthesis of proteins, so that these important and interesting ‘natural products’ could be prepared in an efficient and modular manner, and to then use this enhanced tool-kit to elucidate the structural and chemical basis of protein function. A robust chemical synthesis tool-kit allows for the preparation of proteins and protein analogues, by a single investigator in a period of weeks-to-months, in excellent purity and with good yields. Highly resolving analytical techniques, notably LCMS, are used to characterize the purity and covalent structure of each starting material, synthetic intermediate, and final product, and modern Xray and NMR structural analysis is used to confirm the folded structure of the final protein product. Efficient and reproducible synthetic access to a target protein molecule enables the facile preparation of a wide range of analogues for structure-function studies.

Kinetically Controlled Ligation (Chapters 2-3). Using kinetically controlled ligation, we can prepare linear and cyclic proteins (Chapter 2) in a modular, fully convergent fashion. This chemical tactic allows for the more efficient preparation of proteins in terms of larger size, higher yields, greater flexibility and reduced time for a synthesis. To illustrate the utility of a convergent synthetic strategy enabled by kinetically controlled ligation, consider the hypothetical synthesis of a protein from eight peptide segments, as shown in Figure 8.1. Both strategies are convergent, but the fully convergent approach shown in (B) is much more useful in that it enables the rapid & versatile preparation of chemical analogues, because each peptide segment is the same number of synthetic steps away from the final product.[1] For the same reason, the fully convergent approach makes far more efficient use of all the starting peptide segments and maximizes the yield of final product. Perhaps most importantly for future chemical protein synthesis, the fully convergent approach will enable the total synthesis of much (at least two-fold) larger polypeptide/protein targets containing up to ~350 amino acids.
As useful as kinetically controlled ligation is, further chemical optimizations are needed. First, improved chemistry is needed to prepare peptide-**aryl**/thioesters more conveniently and in higher yields. Currently, there are two routes to prepare peptide-**aryl**/thioesters as shown in **Chapters 2-3** and **Appendix D**. Both of the methods suffer from the losses during purification and or synthesis of a peptide.

In convergent syntheses, with either kinetically controlled ligations or native chemical ligations, a peptide-thioester with an *unprotected* internal Cys side chain thiol can sometimes be very prone to formation of an (internal) thiolactone, which is much less reactive and cannot be used for a kinetically controlled ligation step. To avoid this problem, we use Cys(Acm) protection of such troublesome side chain moieties. Improved chemistry is needed to eliminate the need for internal Cys protection with Acm. Cys(Acm) is not an ideal protecting group: oftentimes, after HF cleavage and deprotection a crude peptide contains unacceptable levels of products in which the Acm
has been removed (i.e. Cys is generated); these byproducts are hard to remove, resulting in lower yields. And, when using Cys(Acm) during a synthesis, extra chemical steps are needed to remove this protecting group from the final product, also resulting in lower overall synthetic yields.

Finally, in a fully convergent synthetic strategy, new chemistry is needed for the kinetically controlled ligation of more than two peptides in the N-to-C direction (Figure 8.1(B)), because of a strong tendency for thiolactone formation involving the side chain thiol at the ligation site after conversion of the initial ligation product to a thioaryl ester (unpublished observations). Currently, this problem has been circumvented by use of Cys(Acm) at all internal Cys residues and the alkylation of the thiol at the ligation site (e.g. with bromoacetamide, to form a ‘pseudo-Gln’ residue).[2] One possible approach that would overcome this problem is the selective desulfurization of the ligation site Cys in the presence of the thioester moiety at the C-terminal of the same peptide; in a model peptide–thioester we have shown that a Cys can be converted to an Ala, without affecting the thioester (unpublished observations).

Selective Desulfurization (Chapters 5-6). Selective desulfurization was developed to increase synthetic versatility for the synthesis of proteins by native chemical ligation. In this thesis research, it has been demonstrated that freshly prepared Raney nickel can be used under standard conditions for the selective desulfurization of Cys, i.e. conversion of Cys to Ala, even in the presence of Cys(Acm). This simple and practical tactic enables the more common Xxx-Ala junctions to be used as ligation sites for the chemical synthesis of Cys-containing peptides and proteins. Furthermore, as originally pointed out by Yan and Dawson,[3] and subsequently reported by others,[4] native chemical ligation can be used with any amino acid that has been modified to have a thiol moiety on the β-carbon, e.g. [β-SH]Phe (Figure 8.2). Subsequent desulfurization regenerates the naturally occurring amino acid at the ligation site. In this way, a variety of Xxx-Yyy sequences can be used as ligation sites.
Despite the great promise of selective desulfurization as reported in this work, a number of developments are still needed. The reaction between a polypeptide and the activated Raney nickel is heterogeneous, and for some targets this leads to lower recoveries of products because peptide gets trapped in the spongy nickel. One way to overcome this issue would be to develop a *homogenous* desulfurization catalyst for use in selective desulfurization reactions of the type studied here. Another problem is the use of Acm to protect side chain thiols. The same problems associated with Acm in kinetically controlled ligation (see above) also hold for selective desulfurization. One immediate way to solve this problem would be to try the use of Tacm (S-trimethylacetamidomethyl) and Phacm (S-phenylacetamidomethyl) for thiol side chain protection in conjunction with selective desulfurization.

*Racemic Protein Crystallography* (*Chapter 7 and Appendices E-F*). Preparation of the mirror image form of a protein molecule is only possible by the use of chemical methods. Efficient total chemical synthesis of a protein target enables the facile preparation of a wide range of analogues; notably this includes preparation of the protein enantiomer - simply by using D-amino acids in the synthesis of the peptide segments used. Using the improved synthetic chemistry tool kit developed in the course of this thesis research, a number of mirror image proteins were prepared and used to explore the utility of racemic crystallography as shown in *Chapters 5-7* and *Appendices E & F*. We found that, in some instances, difficult-to-crystallize proteins can be readily crystallized from racemic mixtures. We also found new and exciting
ways to determine the phases by de novo methods and quasi-racemate MAD (Figure 8.3).

Figure 8.3. The multiple potential advantages of the use of mirror image forms of a protein molecule in Xray structure determination.

Racemic protein crystallography has proved to have exciting applications for protein Xray structure determination. A lot more work needs to be done to explore the full potential of the novel racemic protein crystallography approach. Work is needed to determine the impact on the quality of electron density data of the inherently quantized phases in diffraction data obtained from centrosymmetric crystals.[5] We also need to explore how best to optimize the use of direct methods for solving the structures of proteins using diffraction data obtained from centrosymmetric crystals. In order to gather further data on these questions, we are systematically applying racemic protein crystallography to proteins of ever larger size. For example, we are currently exploring the racemic crystallography of plectasin, a novel fungal ‘microprotein’ of forty amino acid residues, for which no crystal structure has been reported;[6] we have been able to use direct methods to solve the structure of the plectasin racemate, and we are in the process of comparing the direct method solution obtained from crystals of only L-plectasin with that obtained for {D-plectasin+L-plectasin}. We will also reexamine the use of direct methods and the impact of quantized phases in diffraction data obtained
from racemic sfAFP (81 amino acids). Similar work will be undertaken with the enantiomers of the 94 residue predicted protein Rv1738, described in Chapter 5, and larger proteins.

Summary. In the course of my thesis research, improved chemical methods for the synthesis of proteins were developed. Together with efficient stepwise solid phase synthesis of peptide-thioesters[7] and the use of native chemical ligation,[8] these new methods enable the straightforward preparation of target protein molecules, and the preparation of an almost unlimited range of analogues, including mirror image protein molecules. Ready synthetic access to protein enantiomers was then used to expedite the determination of protein X-ray structures by the use of racemic crystallography. This fruitful area of research, which is in its infancy, allows for more facile crystal growth and enables the use of direct methods for phasing. Much work remains to be done to define the full impact of racemic protein crystallography.
References:


THE UNIVERSITY OF CHICAGO

NEW CHEMICAL METHODS FOR THE SYNTHESIS OF PROTEINS AND THEIR APPLICATION TO THE ELUCIDATION OF PROTEIN STRUCTURE BY RACEMIC PROTEIN CRYSTALLOGRAPHY

VOLUME TWO

A DISSERTATION SUBMITTED TO THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES IN CANDIDACY FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMISTRY

BY BRAD L. PENTELUTE

CHICAGO, ILLINOIS
MARCH 2008
Appendix A. Synthesis of peptides and peptide-thioester building blocks for use in convergent chemical ligation


[Draft manuscript, prepared by Durek, Boerma, Pentelute, and Kent. Pentelute did the experiments.]
**Peptide Synthesis:**

The ability to make proteins by modern chemical ligation methods is *absolutely dependent* on being able to chemically synthesize, in good yield and with high purity, the peptide- and especially the peptide-“thioester-building blocks. Unprotected peptides and peptide-“thioesters up to ~50 residues in length can be prepared in a general fashion by highly-optimized, ‘in-situ’ neutralization Boc chemistry\(^1\) followed by HF cleavage from the solid support and concurrent side-chain deprotection\(^2,3\). The advantages of the solid phase method have been thoroughly reviewed (\(^4\) and references therein). Because ours is a teaching laboratory, we do all peptide syntheses *manually* in simple and inexpensive apparati. Since the advent of the native chemical ligation technique, SPPS protocols have been adapted for the direct synthesis of “thioester peptides. Conversion of aminoacyl-\(\text{OCH}_2\)-Pam-resin\(^5\) to the corresponding trityl-associated mercaptopropionic acid-aminoacyl-\(\text{OCH}_2\)-Pam-resin, followed by stepwise chain assembly, gives an “alkylthioester peptide upon HF cleavage\(^6,7\) (note that even standard Boc chemistry SPPS is not suitable for the preparation of peptide-“thioesters; it is *essential* to use the ‘in situ neutralization’ protocols to make peptide-“thioesters). “Arylthioester peptides for use in kinetically controlled ligation can be prepared directly on resin by a similar approach\(^8\), but are best obtained post-cleavage by treatment of the corresponding “alkylthioester peptide with a large excess of aryl thiol\(^9,10\). The preparation of the basic building blocks needed for convergent chemical protein synthesis is schematically illustrated in Figure 1.

Success at this demanding level of peptide synthesis depends *critically* on the use of high quality, chemically stable resin. In these protocols, we describe proven methods\(^11\) for the preparation of such resin from commercially-available starting materials (Figure 1 and 2, steps 1-13). Additionally, we outline documented side reactions in Boc chemistry SPPS\(^1\), and provide a troubleshooting guide for minimizing them. To complement the usually acceptable purity of the crude peptide products prepared by the solid phase method\(^4\), a preparative scale HPLC purification step is standard after HF cleavage/deprotection. In the accompanying paper we provide general guidelines that
should enable other researchers to perform routine purification of crude synthetic peptides\textsuperscript{12}.
Materials

Reagents

Bio-Rad Bio-beads S-X1 (styrene-divinylbenzene copolymer, 1% crosslinkage, 200-400 mesh size, cat # 152-2150)
Methanesulfonic acid (Aldrich cat # 471356-100mL) ! CAUTION Corrosive
Dichloromethane (DCM, LC/GC grade, EMD cat. no. DX0831-1)
Absolute ethanol !CAUTION Flammable
Methanol (Fisher cat. no. BP1105-4) ! CAUTION: Toxic, flammable
Hydrazine (Aldrich cat # H2761-50mL) ! CAUTION: Hydrazine is highly toxic and dangerously unstable, especially in the anhydrous form.
N-Hydroxymethylphthalimide (Aldrich cat # H41803-100g)
N,N-Diisopropylethylamine (DIEA, Applied Biosystems part no. 400136) ! CAUTION Corrosive
N,N-Dimethylformamide (DMF, B&J BioSyn -Peptide Synthesis Grade, Burdick & Jackson cat. no. BB075-4)
Trifluoroacetic acid (TFA, Biograde, Halocarbon) ! CAUTION Corrosive, causes severe burns
Aminomethyl-copoly(styrene-divinylbenzene) resin (100-200 mesh, 1% divinylbenzene, Rapp Polymer no. 123.773)
Boc-L-leucyl-(4-carboxymethyl)benzyl ester (Boc-Leu-OCH2-phenyl-CH2-COOH, NeoMPS part no. LP01102) CRITICAL Some of these compounds are supplied as cyclohexylamine or dicyclohexylamine salts. MUST be converted to the free acid by cation-exchange on …
S-Trityl-β-mercaptopropionic acid (Peptides International cat. no. ASX-5047-PI)
2.5% (v/v) Triisopropylsilane (Aldrich cat. no. 23378-1-10G), 2.5% (v/v) water in TFA
Boc-protected amino acids (as needed for the synthesis, Peptides International) See Box 1 for side chain protecting group recommendations
1,3-Diisopropylcarbodiimide (DIC, Sigma-Aldrich cat. no. D12,540-7) ! CAUTION Toxic, flammable, allergenic
N-(Benzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate (HBTU,
Peptides International cat. no. KHB-1065-PI) **CAUTION** HBTU can cause an allergic reaction in some people. HBTU often forms fine particles and should be handled in a hood. Avoid skin contact\textsuperscript{13}. **CRITICAL** Two isomers of HBTU exist, the uronium salt (O-HBTU) and the guanidinium salt (N-HBTU), that differ in amino acid activation efficiency. O-HBTU activates amino acids faster than N-HBTU, and has been shown to give better results in the synthesis of difficult peptides\textsuperscript{14}. Commercial sources of HBTU, even those listed as O-HBTU are usually the N-HBTU form.

Ultra high purity hydrogen fluoride (Matheson-Trigas code no. G1534165) ! **CAUTION** Highly toxic, avoid contact, do not breathe. Always handle under supervision. Even minor exposures to HF require immediate clinical treatment\textsuperscript{15}. Take no risk!

p-Cresol (Aldrich cat. no. C85751-5G)

Cold anhydrous ethyl ether, 4°C (Fisher cat. no. E138-1) ! **CAUTION** Highly flammable

50% (v/v) Acetonitrile, 0.1% (v/v) TFA in H2O solution

Calcium oxide (Sigma-Aldrich cat. no. 12076)

Dry ice ! **CAUTION** Causes burns

Deionized water

Guanidine hydrochloride (GuHCl, Fisher cat. no. BP178-1)

4-Mercaptophenylacetic acid (Sigma-Aldrich cat. no. 653152, 97% purity)

**CRITICAL** -Commercially available sources of MPAA contain impurities that might interfere with peptide analysis and purification. In cases where this becomes an issue, we recommend purification of MPAA prior to use in ligation reactions. MPAA is conveniently purified by RP-HPLC on C18 columns.

Sodium phosphate, dibasic, anhydrous (Na\textsubscript{2}HPO\textsubscript{4}, Fisher cat. no. BP332-1)

Nitrogen cylinder

Helium cylinder

Argon cylinder

**Equipment**

Fume hood

1000mL Round bottom flask
Teflon coated stir bar
Magnetic stir plate
Heating mantle
250mL Glass fritted funnel with a coarse frit
4000mL Thick-walled, plastic coated vacuum collection flask as waste trap
Vacuum source
Reflux condenser
500mL Round bottom flask
CaCl₂ Drying tube
Glass graduated cylinder size???
Hot plate
1000mL Erlenmeyer Flasks (5)
50mL Glass peptide synthesis vessel with coarse porosity frit, teflon stopcock, and cap (ChemGlass part no. CG-1860-12, see Figure 6a). **CRITICAL** A high flow rate of solvents through the frit of the synthesis vessel is critical for the synthesis. It is important to choose a vacuum source that can provide sufficient suction and a synthesis vessel with “coarse” porosity frit. Frits tend to clog over time and it might be necessary to change synthesis vessels during a SPPS. We recommend cleaning of synthesis vessels after the synthesis by soaking in chromic acid cleaning solutions for 12-24 hours, then thoroughly rinsing the frit with water and drying. **CAUTION** Chromic acid reacts violently with organic solvents and is extremely corrosive. Use appropriate safety precautions.

4000mL Thick-walled, plastic coated vacuum collection flask as waste trap connected to synthesis vessel outlet
Vacuum source to provide suction to the synthesis vessel through the waste trap (house vacuum line or vacuum pump)
Low temperature trap
Glass stirring rod
20mL Disposable glass scintillation vials for amino acid activation
0.5-1L Squirt bottles for wash solvents (DMF and DCM)
10mL Graduated cylinder for TFA measurement
100mL Volumetric flask for HBTU solution preparation
Automatic pipettes for measuring 3.8mL 0.5M HBTU and 1 mL DIEA
HF Reaction apparatus Type 1 (Peptide Institute of Japan code no. 001, see Figure 7)
200mL Reaction vessels (with 3.5cm inner diameter, part 1 in Figure 7)
Stainless steel pressure gauge and outlet valve fitted to the top of the HF tank
Calcium oxide trap (Peptides Institute of Japan code no. 013)
Teflon shavings for CaO trap (Peptides International cat. no. HFI-0423) Vacuum pump for evacuation of apparatus For each cleavage:
100mL Round bottom glass flask with 24/40 joint (ChemGlass cat. no. CG-1506-05)
60mL Buchner funnel with course frit (ChemGlass cat. no. CG-1402-14)
Vacuum adaptor with 24/40 joint (ChemGlass cat. no. CG-1045-05)
1" x 5/16" Octagonal teflon coated magnetic stir bar
500mL Polypropylene Erlenmeyer filter flask
Neoprene filter adaptor (ChemGlass cat. no. CG-1401-13)
Lab jack
Magnetic stir plate
1000mL Beaker of ice water with stir bar
Hand operated vacuum pump (Fisher cat. no. 01-071)
Dewar flask sufficient to accept the HF cleavage vessel
Safety gear (lab coat, full face shield, neoprene apron, and long neoprene gloves)

Reaction vessels (e.g. 1.5, 2.0mL Eppendorf tubes, 20mL Scintillation vials, depending on scale)
Spatulas, pipettes, disposable syringes, magnetic stir bars and magnetic stirrer, shaker,
0.2μm sterile filters
pH-Meter (with probes suitable for measuring large and small volumes (μL to mL)
Balance (mg to g scale)
High Performance Liquid Chromatography (HPLC) system for analytical and preparative purifications equipped with a UV detector.
HPLC analytical (e.g. 2.1-4.6x50-250mm) and preparative (e.g.10-22x250mm) reversed phase (RP) columns with different alkyl substitutions for hydrophobic/hydrophilic
peptide analysis and purification (e.g. C4, C8, C18). See the accompanying paper for
general guidelines regarding peptide analysis and purification.
Mass spectrometer (e.g. MALDI-TOF-MS, ESI-MS, LC-MS, etc.)
Lyophilizer (freeze-drying apparatus)

Reagent Setup

**HBTU Solution.** Make only as much 0.5M HBTU solution as you need for your
synthesis. Excess solution should be stored at 4°C under argon atmosphere. For 100mL of
0.5M HBTU solution in DMF, dissolve 18.962g (50mmol) of HBTU to a final volume of
100mL with DMF in a volumetric flask. ! **CAUTION** HBTU can cause an allergic
reaction in some people. **Handling of the solid powder should be performed in well-
ventilated fume hood.** Avoid skin contact^{13}.

Equipment Setup

**Peptide synthesis vessel.** In a fume hood, connect the peptide synthesis vessel to the
waste trap. Connect the waste trap to the low temperature trap and the vacuum source.
The vacuum in the waste trap is used to separate solvent from resin by filtration. !
**CAUTION** Use thick-walled polymer-coated glassware suitable for evacuation. !
**CAUTION** The following protocols use a number of caustic reagents including TFA and
DIEA. The following personal protection is recommended: safety glasses, chemically
resistant gloves, and a lab coat.

**HF Reaction apparatus.** - Set up as directed by the manufacturer in a fume hood (see
Figure 7). The HF tank (3, see figure 7) is connected to one end of the HF apparatus. The
vacuum pump (5) is attached to the other end of the apparatus. The calcium oxide trap
(4), filled with sufficient calcium oxide to neutralize the amount of HF used in multiple
reactions (1.2kg of CaO can theoretically neutralize about 900mL of HF, but in order to allow for a safety margin should only be used to neutralize 600mL of HF), mixed with teflon shavings, is connected in-line between the apparatus and vacuum pump. For each cleavage reaction prepare a dry ice/ethanol bath in the Dewar flask and an ice water bath in the 1L beaker. Add a large stir bar to the ice water bath.
Procedure

Preparation of aminomethyl-copoly(styrene-divinylbenzene) resin with a target loading of 1mmol/g\textsuperscript{11} • TIMING ~48h

! **CAUTION**: All steps should be carried out in well-ventilated fume hood.

1| Weigh out 25g of Bio-Rad Bio-beads SX-1 (styrene-divinylbenzene copolymer, 1% crosslinkage, 200-400mesh size) into a 1000mL round bottom flask.

2| Add 4.9g (27.6mmol) of N-hydroxymethylphthalamide to the 1000mL round bottom flask.

3| Add an egg-shaped stir bar to the flask and add 450mL of dichloromethane (DCM).

4| Add 50mL of methanesulfonic acid (MSA) and equip the flask with a CaCl\textsubscript{2} drying tube. **CAUTION**: Methanesulfonic acid is highly corrosive.

5| Gently stir the reaction for precisely 5h at room temperature.

6| Transfer the slurry to a glass fritted funnel (coarse frit, 250mL), wash the resin with DCM (4-5 portions, totaling 1500-2000mL) and ethanol (4-5 portions totaling 1500mL).

7| Dry the resin under vacuum for 1h or until the resin does not clump and is fluid (i.e. flows like fine dry sand).

**PAUSE POINT** The phthalimidomethyl-resin can be stored in this form indefinitely.

8| Transfer the resin to a 500mL round bottom flask and suspend it in 200mL of a solution of 5% hydrazine in absolute ethanol. **CAUTION**: Hydrazine is highly toxic and dangerously unstable, especially in the anhydrous form. Alternatively, hydrazine hyrdate may be used in 95% ethanol. It is important that the ethanol used be free of interfering additives, for example isobutylketone is added as a ‘denaturant’ by some suppliers; this
would cause a side reaction with amine functionalities.

9) Equip the flask with a reflux condenser and reflux the slurry for 8h. Do not attempt to stir.

10) The resulting gelatinous material needs to be transferred hot to a glass fritted funnel (coarse frit, 250mL).

11) Wash the gelatinous material with 6-10 portions totalling 1000-2000mL of boiling absolute ethanol.

12) Wash the resin with hot methanol (3-4 aliquots totalling 1000mL), until the white precipitate is washed away and only resin is left in the glass fritted funnel. **CRITICAL STEP** The white precipitate that forms the gel is the highly insoluble phthalhydrazide reaction product, and must be completely removed. Then wash with dimethylformamide (DMF, 3-4 aliquots totalling 800mL), DCM (3-4 aliquots totalling 800mL), 1:1 DCM:TFA (3-4 aliquots totalling 600mL), DCM (3-4 aliquots totalling 600mL), 10% DIEA:DMF (v/v) (2-3 aliquots totalling 500mL), DMF (2-3 aliquots totalling 600mL), DCM (3-4 aliquots totalling 1000mL), and then dry the resin under vacuum and store in an airtight screw cap container. The aminomethyl-resin can be stored in this form indefinitely (years). **CAUTION** TFA is corrosive and causes severe burns.

13) Determine the loading of the resin (i.e. the amount of amino groups in mmol per gram of resin; the above protocol will give **1.0 mmol/gram aminomethyl styrene divinylbenzene (S-DVB)**) with elemental analysis for nitrogen content. If the facility for this measurement is not available, send your resin to [delete :Desert Analytics, Inc. in Tucson, Az. USA or an alternative] a service provider. Additionally, a sample of the aminomethyl-resin should be examined by infrared spectroscopy to confirm the absence of carbonyl absorption bands (  cm\(^{-1}\)).

**PAUSE POINT**
**Preparation of Boc-Leu-OCH2-Pam-resin**

11 TIMING ~13h.

14| Weigh 0.5mmol of aminomethyl-copoly(styrene-divinylbenzene) resin into the peptide synthesis vessel. Calculate the weight of 0.5mmol of resin by dividing the desired amount of resin (0.5mmol) by the substitution of the aminomethyl-resin (in mmol/g) (i.e. for a substitution of 1.0mmol/g: 0.5mmol / (1.0mmol/g) = 0.500g, i.e. 500mg).

15| Weigh 1mmol (379.4mg) of Boc-Leu-OCH2-phenyl-CH2-COOH into a glass scintillation vial. Please note that to obtain peptides carrying a free “COOH replace Boc-Leu-OCH2-phenyl-CH2-COOH with Boc-Xaa-OCH2-phenyl-CH2-COOH where Xaa is the C-terminal amino acid residue, and the weight used adjusted according to the formula weight of the compound used.

16| Dissolve the Boc-Leu-OCH2-phenyl-CH2-COOH in 10mL DCM. Transfer the solution to the peptide synthesis vessel containing the aminomethyl-resin. Allow the resin to swell in this solution for ~5-15min. D we?

17| Initiate the coupling by adding 1mmol (156.6µl) DIC and mixing well. ! CAUTION DIC is toxic, flammable and allergenic.

PAUSE POINT Cap the vessel and allow the coupling reaction to proceed overnight {this is nuts – the reaction will be completed in a few minutes!} \{Only initial mixing is needed\}

18| Coupling completeness can be monitored using the ninhydrin test for the presence of free amines (see Box 2). Any remaining free amines can be acylated by repeating the coupling steps 15 through 17 (for 2h instead of overnight).

19| Drain the coupling solution from the vessel.
20| DCM flow wash. With the stopcock closed, rinse the sides of the synthesis vessel with DCM from a squirt bottle until the solvent covers the resin. Slowly open the stopcock and continue adding DCM to maintain the solvent level. Continue flowing DCM through the resin rapidly for 30sec using a total volume of ~50-75mL of DCM. Drain the remaining solvent from the vessel.

21| With the stopcock closed, add ~10mL of DCM to the drained resin. Wait for 30sec, mixing occasionally. Drain the solvent from the vessel.

22| Repeat step 20, ‘DCM flow wash’.

PAUSE POINT Loaded resin can be stored for days in this state. For prolonged storage, evaporate remaining solvent under vacuum and store the dried resin in a sealed container.

Synthesis of trityl-associated mercaptopropionic acid-leucine (TAMPAL) resin \(^6\) TIMING \(\sim\) 30min

23| Swell the drained Boc-Leu-OCH2-Pam-resin (0.5mmol) from step 22 in 5-10mL DMF for \(\sim\)30min. Drain the resin. Do not draw air through the drained resin.

24| Boc removal. Start the removal of the N\(^\text{\textprime}\)-Boc-protecting group with neat TFA: with the stopcock open, add 10mL of TFA to the resin. Allow the TFA to drain rapidly.

CAUTION When mixed, TFA and DMF will generate heat and the mixture should be removed as quickly as possible.

25| Close the stopcock and add 10mL of TFA to the resin. Wait 1min, stirring occasionally. Drain the resin.

26| Treat the resin with an additional 10mL of TFA for 1min. Drain the resin.
27| DMF flow wash. With the stopcock open to allow rapid DMF flow, rinse the sides of the synthesis vessel with DMF from a squirt bottle maintaining a ~2-5mm head of solvent above the resin. Continue flowing DMF through the resin for 30sec using a total volume of ~50-75 mL of DMF. Drain the remaining solvent from the vessel.

28| With the stopcock closed, add ~10mL of DMF to the drained resin. Wait 30sec, mixing occasionally. Drain the solvent from the vessel.

29| Repeat step 27, ‘DMF flow wash’.
CRITICAL STEP Incomplete removal of TFA from the deprotected resin (washing steps 27-29) can result in the incomplete coupling of S-trityl-β-mercaptopropionic acid. See the Troubleshooting section for more information.

30| Weigh 2mmol (696.9mg) S-trityl-β-mercaptopropionic acid into a scintillation vial, and dissolve in 3.8mL 0.5M HBTU solution (1.9mmol). Add 1mL DIEA (~5.7mmol), cap the vial, and mix by shaking vigorously then wait 30sec-1min.
CRITICAL STEP The ratio of S-trityl-β-mercaptopropionic acid to HBTU must be greater than one to prevent tetramethylguanidinium termination of the growing peptide chain. See the Troubleshooting section for more information.
CRITICAL STEP Activation of Boc-amino acids with N-HBTU in the presence of a tertiary amine generates a mixture of O-linked and N-linked activated species. Over time activated Boc-amino acids isomerize from the more active O-linked species to the less active N-linked species\textsuperscript{16}. For that reason, pre-activation times should be kept short (30sec to 1min) to ensure that the most reactive form is used for coupling.

31| Add the activated S-trityl-β-mercaptopropionic acid to the N\textsuperscript{α}-Boc-deprotected, washed, and drained resin. Perform the coupling reaction by incubating the resulting mixture for 16min, stirring occasionally. Drain the coupling solution from the vessel. Coupling completeness can be monitored using the ninhydrin test for the presence of free amines (see Box 2). Any remaining free amines can be acylated by repeating the coupling steps 30 through 31.
32| Repeat the ‘DMF flow wash’ from steps 27-29.

**PAUSE POINT** TAMPAL-resin can be stored for days in DMF solution. For longer storage, wash with DCM as in steps 20-22, evaporate remaining solvent under vacuum, and store in a sealed container at room temperature.

**Coupling of the first Boc-amino acid to the TAMPAL resin**

**TIMING ~ 15-20min**

33| If resin has been dried after coupling of S-trityl-β-mercaptopropionic acid, swell in ~10mL DMF. Drain the resin.

34| Remove the trityl group by treatment with a deprotection solution consisting of 2.5% (v/v) triisopropylsilane, 2.5% (v/v) water in TFA: with the stopcock open, add 10mL of deprotection solution to the resin. Allow the solution to drain rapidly. The trityl group will turn the deprotection solution bright yellow. **CAUTION** When mixed, TFA and DMF will generate heat and the mixture should be removed as quickly as possible.

35| Close the stopcock and add 10mL of deprotection solution to the resin. Wait 1min, stirring occasionally. Drain the resin.

36| Treat the resin with an additional 10mL of deprotection solution for 1min. Drain the resin.
If any yellow color is observed after this treatment with deprotection solution, repeat step 35 until the yellow color is no longer observed. Drain the resin.

**CRITICAL STEP** Avoid pulling excessive air through the deprotected TAMPAL resin when draining the resin with vacuum. It is best to perform the operation at this stage under an atmosphere of nitrogen gas in order to prevent oxidation of the thiol6 (with the stopcock open drain the resin by blowing nitrogen gas from a cylinder into the gently-sealed opening of the synthesis vessel).

37| Flow wash the resin with DMF in the same way as detailed in steps 27-29.
**CRITICAL STEP** Incomplete removal of TFA from the deprotected resin can result in incomplete coupling of the next Boc-amino acid. See the Troubleshooting section for more information.

38| Activate the Boc-amino acid to be coupled to the thiol-resin: weigh 2mmol of the protected amino acid into a scintillation vial. Dissolve the protected amino acid in 3.8mL 0.5M HBTU solution (1.9mmol), add 1mL DIEA (~5.7mmol), cap the vial, and mix by shaking vigorously. Wait 30sec-1min.

**CRITICAL STEP** The ratio of protected amino acid to HBTU must be greater than 1 to prevent tetramethylguanidinum termination of the growing peptide chain. See the Troubleshooting section for more information.

**CRITICAL STEP** Activated Boc-amino acids isomerize from the more active O-linked species to the less active N-linked species over time\(^1\). Activation time should not exceed 1min.

**CRITICAL STEP** Transthioesterification reactions involving unprotected internal cysteine residues present in both polypeptide segments to be coupled (see below) can have a deleterious effects on a KCL reaction (see steps 93-102). Therefore, protection of these non-ligation site cysteines with S-Acm is recommended. Acm protecting groups will then be removed after assembly of the full-length polypeptide chain, as will be described further on in the procedure (steps 103-110).

39| Add the activated amino acid to the trityl-deprotected, washed, and drained resin. Perform the coupling reaction by incubating the resulting mixture for 1h, stirring occasionally. Drain the coupling solution from the vessel.

40| Repeat the DMF wash from step 27-29.

**PAUSE POINT** Boc-aminoacyl-TAMPAL-resin can be stored for days in DMF solution. ? Can it be stored dry for longer periods?

**Peptide assembly using manual 'in-situ' neutralization Boc chemistry stepwise solid phase peptide synthesis\(^1\)**•**TIMING** ~20min per amino acid residue
41| Remove the N"-Boc protecting group from the Boc-aminoacyl-TAMPAL-resin/growing peptide with neat TFA following the procedure described in steps 24-26.

42| Wash the resin with DMF following the procedure described in steps 27-29.

CRITICAL STEP Incomplete removal of TFA from the deprotected resin can result in incomplete coupling of the next Boc-amino acid. See the Troubleshooting section for more information.

43| Activate the next amino acid as in step 38.

44| Add the activated Boc-amino acid to the deprotected, washed, and drained resin. Run the coupling reaction for 10min, mixing occasionally. Drain the coupling solution from the vessel.

CRITICAL STEP Glutamine residues require special care both before and after N"-Boc protecting group removal to prevent pyrrolidinone carboxylic acid formation and peptide chain termination. See the Troubleshooting section for more information.

45| Repeat the DMF wash following the procedure described in steps 27-29.

PAUSE POINT Boc-peptide-resin can be stored for days in DMF solution at room temperature.

46| For each residue in the sequence repeat steps 41-45

PAUSE POINT Full length Boc-peptide-resin can be stored for days in DMF solution. For longer storage, wash the resin with DCM (steps 20-22) and remove any remaining solvent under vacuum. The dried resin can be stored in a sealed container for months, preferably at 4°C or in a desiccator.

CRITICAL STEP Removal of the final N"-Boc-protecting group is carried out immediately prior to HF cleavage and side chain deprotection. Do not deprotect, if you intend to store the resin for more than a few days.
Preparation of resin for final deprotection and cleavage •TIMING ~20minutes
47) If peptide-resin was dried following completion of chain assembly add 10mL DMF to resin and allow to swell for 5min. Drain resin.

48) Remove the final Nα-Boc-protecting group from the completed peptide with neat TFA as described in steps 24-26. 
CRITICAL STEP Failure to remove the final Nα-Boc group before HF deprotection and cleavage can result in tert-butylation of methionine residues. See the Troubleshooting section for more information.

49) Wash the resin with DMF following the procedure described in steps 27-29.

50) Wash the resin with DCM following the procedure described in steps 20-22.

51) Evaporate residual solvent from the Nα-Boc deprotected resin with a stream of nitrogen or under vacuum. CRITICAL STEP The Nα-Boc deprotected resin is not stable enough to be stored for more than a few days. Try to continue with final HF deprotection as soon as possible.

Final deprotection and cleavage with anhydrous HF2•TIMING ~3h

! CAUTION: Hydrogen fluoride is extremely dangerous. Personal safety measures should include lab coat, full-face shield, neoprene apron, and long neoprene gloves. It is particularly important to avoid contact with the HF transfer lines and the ethyl ether used to wash the precipitated peptide as it may contain dissolved HF. All steps must be carried out in a well-ventilated fume hood.

52) Weigh out 500mg of dried resin from step 51 into the HF cleavage vessel (1, Figure 7).
53| Add 500μl p-cresol to the resin in the cleavage vessel (1μl p-cresol/mg resin).

54| Solidify the p-cresol/resin mixture by suspending the cleavage vessel in the Dewar flask containing the dry ice/ethanol bath, add the small stir bar, and attach the cleavage vessel to the cleavage apparatus (1, Figure 7).

CRITICAL STEP Placing the stir bar on top of the frozen p-cresol/resin mixture ensures the stir bar will turn freely after HF has been condensed.

55| Use the lab jack to raise the Dewar flask until the attached cleavage vessel is submersed in the dry ice/ethanol mixture. Allow the cleavage vessel to cool to dry ice temperatures (~10min), adding additional cold ethanol to keep the vessel submerged.

56| With the HF tank closed (valve 10, Figure 7), evacuate the cleavage apparatus by turning on the vacuum pump, and opening the valves connecting the apparatus to the vacuum pump and to the pressure gauge (7, Figure 7). Evacuate the apparatus including the cleavage vessels until the pressure reading on the pressure gauge stabilizes.

57| Close the valve (6 in Figure 7) connecting the vacuum pump to the apparatus. Ensure that the evacuated apparatus does not leak (indicated by a stable low pressure reading on the gauge). Leave the vacuum pump turned on.

CRITICAL STEP The valve connecting the apparatus to the vacuum pump must be closed before the HF tank is opened.

58| Condense ~10mL of HF into the chilled cleavage vessel (1, Figure 7) for a final concentration of 5% (v/v) p-cresol. Open the valve on top of the HF tank (10, Figure 7). HF gas will flow into the apparatus condensing on the cold walls of the cleavage vessel(s) (1 in Figure 7). To check the HF level in a cleavage vessel, temporarily remove the Dewar flask and shine a bright light (9 in Figure 7) from the rear of the vessel. When 10mL of HF has been condensed, close the valve on top of the HF tank (valve 10, Figure 7) and close the cleavage vessel valve. **CAUTION:** The light source should be spark-proof in order to prevent explosion of ether vapors. ? What ether vapors: in case one
forgets to turn of the lamp and starts work up with ether? **CRITICAL STEP** The HF tank valve must be **closed** before opening the apparatus to the vacuum pump.

59| Open the valve to the vacuum pump to remove the remaining HF gas from the apparatus. Close the valve to the vacuum pump. Turn off the vacuum pump, while venting the CaO trap through valve 6 (Figure 7).

60| Lower the lab jack to remove the cleavage vessel from the dry ice/ethanol bath. Replace the Dewar flask with the stir plate and ice water bath. Use the lab jack to immerse the cleavage vessel in the ice water bath. Turn on the stir plate and ensure both stir bars are stirring. Stir for 1h. Ensure the ice bath remains at 0°C by adding additional ice throughout the reaction.

61| Record the amount of HF used on a log sheet. Determine the remaining capacity of the calcium oxide trap.

62| To evaporate the HF from the cleavage vessel, turn on the vacuum pump and open the valve connecting the pump to the apparatus and evacuate (6 in Figure 7). While continuing to stir the water bath and cleavage reaction, slowly and carefully open the valve to the cleavage vessel to evaporate the liquid HF. Ensure that the HF/p-cresol/resin mixture does not bump, contaminating the apparatus. It may help to open the valve in stages, allowing the HF boil-off to slow down before opening the valve further. Continue pumping for 10-15min after the cleavage vessel valve has been fully opened to ensure complete evaporation of the HF. If bubbles are seen forming in the vessel after this time, continue pumping until all boil-off stops. **CRITICAL STEP** Keep the cleavage vessels at 0°C throughout this process.

63| Close all valves, vent the calcium oxide trap, and turn off the vacuum pump and stirrer(s).

64| Carefully remove the cleavage vessel from the cleavage apparatus.
Precipitate the deprotected and cleaved peptide by adding 20-25mL cold anhydrous ethyl ether to the cleavage vessel. 

! **CAUTION** The ethyl ether solution may contain trace amounts of HF. Handle carefully and avoid contact.

Collect the precipitated peptide and resin from the ether suspension using a glass filter funnel, plastic filter flask, neoprene filter adaptor, and a hand-operated vacuum pump.

! **CAUTION** Ethyl ether fumes are highly explosive. Due to the risk of explosion, do not use a house vacuum line or electric vacuum pump to filter the ether.

Wash the cleavage vessel and precipitated peptide 2-3 additional times with 20-40mL ether. Put the ether waste aside.

Connect the filter funnel to the glass vacuum adaptor and to a clean, empty round bottom flask.

Wash the cleavage vessel with ~10-50mL 50% (v/v) acetonitrile, 0.1% (v/v) TFA in H₂O solution. Transfer the wash to the glass filter funnel to dissolve the precipitated peptide. Use the hand-operated vacuum pump to separate the dissolved peptide from the resin. Repeat until all peptide is dissolved.

**CRITICAL STEP** Most peptides will dissolve well in 10-50mL 50% (v/v) acetonitrile, 0.1% (v/v) TFA in H₂O. If solubility is an issue, try to change the acetonitrile to water ratio (10-75% (v/v) acetonitrile in water, 0.1% (v/v) TFA). Alternatively, increase the percentage of TFA.

Analyze the crude peptide solution by liquid chromatography and mass spectrometry ([http://www.ionsource.com](http://www.ionsource.com)) to confirm the identity of the product and establish the complexity of the product mixture. **TROUBLESHOOTING**

Freeze the crude peptide solution in a dry ice/ethanol bath and lyophilize for storage.
PAUSE POINT Crude and purified lyophilized peptides can be stored for months at 4°C in a sealed container. Peptides obtained from SPPS will in most cases require a purification step before they can be further used. See the accompanying paper for some general guidelines for performing preparative HPLC purifications.

**Transthiosterification of peptide- thioalkylesters with 4-mercaptophenylacetic acid**\(^9,10\) **TIMING ~6h**

72| Prepare a stock solution of 6M GuHCl, 0.2M sodium phosphate: For a 50mL solution, mix together 28.7g (300mmol) of GuHCl and 1.42g (10mmol) of anhydrous Na\(_2\)HPO\(_4\) and add deionized water to 50mL final volume. Stir until fully dissolved. Filter the solution through a 0.2µm sterile filter and degas it for at least 20min by sparging with helium: connect a clean fine needle or a sparging element to a helium cylinder and immerse it into the solution. Let a continuous stream of gas flow through the solution. Adjust the pH to 6.8-7.0.

73| Weigh out the peptide- thioalkylester and place it in an appropriate reaction vessel. Note that in favorable cases (high purity of crude peptides obtained in step 71) this procedure can be carried without the need for intermediate purification of the crude peptide.

74| Dissolve the peptide in buffer (prepared in step 72) so that the final concentration of the peptide is about 1mM. Stir until the peptide has dissolved completely.

75| Add 4-mercaptophenylacetic acid (MPAA) to the solution to give a final concentration of 150mM (25.23mg/mL of peptide solution). Stir until MPAA is completely dissolved. Solubility of MPAA is highly pH dependent; increase the pH slightly by carefully (!) adding a few drops of 2M NaOH if solubilization is difficult. Avoid over-titration.
76| Check the pH of the reaction mixture and readjust if necessary to 6.8-7.0.

**CRITICAL STEP** Avoid over-titration especially when adding bases. Exposure of your peptide- thioalkylester to basic conditions might give rise to side reactions.

77| Remove a small aliquot from the reaction mixture and quench it by diluting it 10 fold with 50% (v/v) acetonitrile, 0.1% (v/v) TFA in H₂O solution. Analyze the sample by your preferred method (we highly recommend LC, MALDI-TOF, LC-MS). Afterwards purge the reaction vessel with argon by letting a stream of argon from a cylinder flow into the vessel for about 10sec. Cap the reaction vessel.

78| Continue stirring at room temperature until the exchange reaction reaches equilibrium (typically 2-6h for 85-95% conversion). From time to time analyze an aliquot from the reaction and check the pH value of your reaction as described in steps 76-77; aim to keep the solution at 6.8-7.0. Please note that 2,4-dinitrophenyl (DNP) histidine protecting groups will be quantitatively cleaved from the imidazole side chain under these conditions (the solution will turn yellowish).

79| Purify the product peptide- thio(4-carboxymethyl)phenyl ester by RP-HPLC (see the accompanying paper for general comments on purification).

**? TROUBLESHOOTING**

Table 1: Trouble shooting section for Boc SPPS (see step 70)

<table>
<thead>
<tr>
<th>Problem</th>
<th>Possible Reason</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trifluoroacetylation</strong> terminated peptide + 96Da</td>
<td>Resin contains hydroxymethyl impurities or has a lower acid stability</td>
<td>Use a high quality acid stable resin (i.e. OCH2-Pam-resin, free of extraneous HO-CH2-groups).</td>
</tr>
<tr>
<td><strong>Tetramethylguanidinium termination</strong></td>
<td>Coupling reaction was performed with more</td>
<td>Perform coupling reactions with excess Boc-amino acid over</td>
</tr>
<tr>
<td>Process</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Pyrrolidinone carboxylic acid formation</strong></td>
<td>Deprotected glutamine residue exposed to a mixture of TFA and DMF.</td>
<td></td>
</tr>
<tr>
<td>Terminated peptide after glutamine -17Da</td>
<td>After coupling glutamine and washing the resin with DMF (steps 27-29), wash the resin with DCM (steps 20-22). Remove the N\textsuperscript{\textalpha}-Boc-protecting group with TFA (steps 24-26), and repeat the DCM wash before washing with DMF.</td>
<td></td>
</tr>
<tr>
<td><strong>Acetylation</strong></td>
<td>Boc-amino acid contaminated with acetic acid.</td>
<td></td>
</tr>
<tr>
<td>Terminated peptide + 42Da</td>
<td>Lyophilize the contaminated Boc-amino acid from 50% (v/v) acetonitrile/water before use to remove acetic acid.</td>
<td></td>
</tr>
<tr>
<td><strong>Diketopiperazine formation</strong></td>
<td>Proline or glycine residue within 1 or 2 residues of an ester or thioester linkage was exposed to weak acid.</td>
<td></td>
</tr>
<tr>
<td>C-terminally truncated peptide</td>
<td>Use neat TFA for deprotection. When changing solvents from DMF to TFA or TFA to DMF use rapid washes to minimize exposure to dilute TFA in DMF. Preactivate amino acids to prevent exposure of the resin to free Boc amino acids.</td>
<td></td>
</tr>
<tr>
<td><strong>Tert-butylation</strong></td>
<td>Failure to remove final N-Bocprotecting group before HF cleavage and deprotection.</td>
<td></td>
</tr>
<tr>
<td>Peptide +56Da</td>
<td>Remove the final N-Boc-protecting group before the final HF deprotection and cleavage.</td>
<td></td>
</tr>
<tr>
<td><strong>p-Cresol adduct formation</strong></td>
<td>Acylium ions formed from carboxylic acids (glutamic acid or C-terminus) during cleavage have been scavenged by p-cresol.</td>
<td></td>
</tr>
<tr>
<td>Peptide +90Da</td>
<td>Perform HF cleavage at 0\degree C to prevent ketone formation.</td>
<td></td>
</tr>
<tr>
<td><strong>Amino acid deletion</strong></td>
<td>Incomplete coupling of Boc-amino acid</td>
<td>Resynthesize the peptide. Use the ninhydrin test to check coupling of the suspect amino acid. Repeat incomplete coupling reaction using a different coupling chemistry. Use stable resins, free of extraneous functional groups.</td>
</tr>
<tr>
<td>------------------------------</td>
<td>--------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Methionine sulfoxide</strong></td>
<td>Pulling air through peptide under acidic conditions.</td>
<td>Avoid pulling air through the precipitated peptide after HF cleavage.</td>
</tr>
<tr>
<td>formation Peptide +16Da</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Details on the Potential Side Reactions in Boc SPPS**

**Trifluoroacetylation** - A side reaction generating a terminated peptide with a mass of 96 Da greater than the free peptide chain. The mass increase is the result of a trifluoroacetyl group added to the N-terminal amine of the growing peptide, preventing further chain elongation. Trifluoroacetylation during the neutralization step of conventional stepwise solid phase synthesis is caused by resin-bound trifluoroacetic acid benzyl ester functionalities. The reactive functionalities are generated from resins containing pre-existing hydroxymethyl sites, or from aminoacylester-resins with insufficient acid stability to repeated TFA treatment; acidolysis of the ester resins directly generates the TFA benzyl ester sites. Prevention of trifluoroacetylation is accomplished by the use of aminoacylester-resins that are highly acid stable and free of hydroxymethyl sites, such as a high quality Pam resin\(^\text{17}\).

**Tetramethylguanidinium termination** - A side reaction generating a terminated peptide with a mass of 99 Da greater than the free peptide chain. The mass increase is the result of a tetramethylguanidinium group added to the N-terminal amine of the growing peptide, preventing further chain elongation. If the ratio of amino acid to HBTU is less than one (i.e. more HBTU than amino acid is present during activation) a fraction of the growing peptide will terminate. To prevent this side reaction make sure to use a slight
excess of amino acid over HBTU as recommended in the protocol\textsuperscript{1}.

**Pyrrolidinone carboxylic acid formation** - A side reaction generating a terminated peptide at a glutamine residue with a mass of 17Da less than the terminated peptide containing a non-cyclized glutamine. In the presence of weak acid (e.g. dilute TFA in DMF) the free amine of an N-terminal glutamine residue will cyclize with the side chain. The residue formed is referred to as pyrrolidinone carboxylic acid (PCA) or "PyroGlu". To prevent PCA formation the first TFA deprotection following coupling of a glutamine must be preceded and followed by the DCM wash described in step \textbf{20-22}. DCM washes before and after TFA deprotection will prevent exposure of the deprotected glutamine to dilute TFA in DMF\textsuperscript{1,18}.

**Acetylation** - A side reaction generating a terminated peptide with a mass of 42Da greater than the free peptide chain. The mass increase is the result of an acetyl group added to the N-terminal amine of the growing peptide, preventing further chain elongation. Presence of acetic acid as an impurity in Boc-amino acids can cause this side reaction. Contaminating acetic acid can be removed before use by lyophilizing the Boc-amino acid. There is some suspicion that the presence of ethyl acetate contamination may also lead to acetylation. As a precaution, amino acids containing residual ethyl acetate should be dried under vacuum before use.

**Diketopiperazine formation** - In the presence of a weak acid such as dilute TFA in DMF, Boc-amino acid, or acetic acid the free secondary amine of proline, or to a lesser extent the primary amine of glycine, will attack the ester link to the resin (at the dipeptide stage), eliminating a diketopiperazine (DKP), and causing chain loss from the resin \{this is not correct-rewrite\}. Thioester peptides are particularly prone to this side reaction when prolines or glycines are 1 or 2 residues away from the thioester; in this case a free thiol is formed on the resin. Chain elongation will begin again on this thiol resulting in C-terminally truncated peptides. The following steps can be taken to minimize or prevent DKP formation: use neat TFA for deprotection, when changing solvents from DMF to TFA or TFA to DMF use rapid washes to minimize exposure to dilute TFA in DMF,
preactivate amino acids to prevent exposure of the resin to free Boc-amino acids\cite{Gisin, 1972 #1071}.

**Tert-butylation** - A side reaction generating a byproduct peptide with a mass of 56Da greater than the free peptide chain. The mass increase is the result of a tert-butyl group added to the side chain of methionine residues. Tert-butylation can be prevented by ensuring that the N\textsuperscript{a}-Boc group is removed from the completed peptide chain before HF cleavage and side chain deprotection. Tert-butylation of the side chain thioether of methionine is slowly reversible on standing at room temperature\cite{19}.

**p-Cresol adduct formation** – A side reaction during HF cleavage generating a byproduct peptide with a mass of 90 Da greater than the free peptide chain. During HF cleavage unprotected carboxylic acids (such as glutamic acids and the C-terminus) can form acylium ions which are scavenged by p-cresol. p-Cresol can be linked to the peptide as either an ester or a ketone\cite{3}. p-Cresol esters can irreversibly rearrange to p-cresol ketones when cleavage temperatures exceed 0\degree C.

**Amino acid deletion** - A side reaction caused by incomplete coupling of an incoming Boc-amino acid generating a byproduct that is lower in mass by the mass of the missing amino acid. Difficult couplings can originate from the chemical properties of the sequence being synthesized, or the nature of the Boc-amino acid being coupled. Once the deleted amino acid has been identified, monitor the coupling of this Boc-amino acid during the next synthesis using the quantitative ninhydrin test (box 2). If the coupling reaction fails to go to completion in the allotted time, re-couple the Boc-amino acid. This may help, depending on the chemical basis of the incomplete reaction.

**Racemization** – Partial racemization of activated Boc amino acids can occur in the presence of base\cite{20}. Activated Boc-amino acids can eliminate the activating ester, forming an oxazalone. Reversible deprotonation of the oxazalone causes loss of configuration at the $\alpha$-carbon. Incorporation of the partially racemized Boc-amino acid into the growing peptide chain gives rise to the formation of a variety of peptide diastereomers, isomers
that can be very closely related and therefore hard to separate from the target chirally pure peptide. Cysteine, aspartic acid, glutamic acid, and dinitrophenyl-protected histidine (His(DNP)) are more susceptible to racemization: all other amino acids typically exhibit less than a few tenths of a percent racemization under the activation and coupling conditions described here. Use of a separate neutralization step and activation and coupling in the absence of base completely avoids racemization, but is incompatible with the synthesis of thioester peptides (see diketopiperazine formation above). Byproducts that arise from partial racemization are peptide diastereomers, i.e. distinct chemical compounds with the same mass as the target peptide. They will in general be separable by HPLC. The D-amino acid content of the crude peptide will document the enantiomeric purity of the starting materials and synthetic methods used, and can be determined by chiral chromatography after acid hydrolysis of the peptide product (with appropriate controls for racemization during the hydrolysis) Refs.HOW IS RACEMIZATION DETECTED

**Methionine sulfoxide formation** – Pulling air through a precipitated peptide under acidic conditions can result in formation of methionine sulfoxide peptide with a mass of 16Da greater than the free peptide. Methionine sulfoxides, but not sulfones can be reduced under a variety of conditions\(^21\).

**N-to-O acyl shift** - In the presence of strong acids such as anhydrous hydrogen fluoride, serine and to a lesser extent threonine containing peptides can undergo an N-to-O acyl shift. The migration can be sequence dependent. In some cases, the O-linked peptide can be converted to the N-linked peptide by exposure to a mild base (i.e. 0.05M sodium bicarbonate for 24hr)\(^22\)-\(^24\). The shift can be detected by subjecting a sample of the peptide to harsh basic conditions (e.g. 50mM NH\(_4\)HCO\(_3\), pH 9.0, T=50°C, 24h), which results in hydrolysis of the ester bond and the cleavage of the peptide into two segments ("thioesters, if present, are also hydrolyzed")\(^24\).
**Anticipated Results**

For demonstration purposes we undertook the synthesis of a small 15 amino acid residue peptide-thioester (TTCCPSIVARSNFNA-thioalkylester corresponding to Crambin segment Cram[Thr-Ala]-thioalkylester; see accompanying paper). Figure 5a shows a representative HPLC chromatogram of the crude 15 amino acid residue peptide-thioester after HF side chain deprotection and cleavage from the resin. Side products of the non-optimized synthesis were identified using HPLC-MS. This information can (and should) be used for developing an optimized synthesis protocol. The crude peptide was purified using preparative HPLC. Figure 5b shows an analytical HPLC chromatogram of the purified peptide. The scale of the synthesis was 0.8mmol and the yield was 444 mg (230mmol) of lyophilized peptide after HF cleavage and purification, which corresponds to a 29% overall yield for synthesis and purification. This peptide was meant to be ligated to Cram[Cys-Leu]-thioalkylester in a kinetically controlled ligation to give Cram[Thr-Leu]-thioalkylester. This requires an activation of Cram[Thr-Ala]-thioalkylester by transthioesterification with MPAA to give the more reactive Cram[Thr-Ala]-thio(4-carboxymethyl)phenyl ester.

A portion of the obtained Cram[Thr-Ala]-thioalkylester (60mg, 0.031mmol) was transthioesterified with 150mM MPAA as described in steps 72-79. Figure 5c shows a HPLC chromatogram of the crude reaction after 4h indicating about 85% formation of the desired Cram[Thr-Ala]-thio(4-carboxymethyl)phenyl ester. The purified reaction product is shown in Figure 5d of which 40mg were isolated (0.021mmol, 68% yield).
References


10. Johnson, E. C. & Kent, S. B. Insights into the mechanism and catalysis of the


Protected Amino Acids

The following amino acid side chain protecting groups are compatible with 'in situ' neutralization Boc chemistry SPPS, and are removable by anhydrous HF treatment unless noted:

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Protecting Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arginine</td>
<td>tosyl (Tos)</td>
</tr>
<tr>
<td>Asparagine</td>
<td>9H-xanthen-9-yl (Xan)&lt;sup&gt;(a)&lt;/sup&gt;</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>cyclohexyl (OcHex)&lt;sup&gt;(b)&lt;/sup&gt;</td>
</tr>
<tr>
<td>Cysteine</td>
<td>4-methylbenzyl (MeBzl), acetamidomethyl&lt;sup&gt;(c), (d)&lt;/sup&gt;</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>cyclohexyl (OcHex)&lt;sup&gt;(b)&lt;/sup&gt;</td>
</tr>
<tr>
<td>Histidine</td>
<td>2,4-dinitrophenyl (DNP)&lt;sup&gt;(c), (e), (f)&lt;/sup&gt;</td>
</tr>
<tr>
<td>Lysine</td>
<td>2-chloro-Z (2ClZ)</td>
</tr>
<tr>
<td>Serine</td>
<td>benzyl (Bzl)</td>
</tr>
<tr>
<td>Threonine</td>
<td>benzyl (Bzl)</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>bromo-Z (BrZ)</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>formyl (CHO)&lt;sup&gt;(e), (f)&lt;/sup&gt;, cyclohexyloxycarbonyl (Hoc)&lt;sup&gt;(g)&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

[Glutamine and methionine are used without side chain protection.]

(a) Unprotected asparagine has been reported to undergo side reactions during activation and can result in low coupling yields<sup>26, 27</sup>. Protection of the asparagine side chain with the 9-xanthenyl group prevents these problems<sup>27</sup>. After coupling, TFA treatment of Asn(Xan) will slowly remove the xanthenyl group, indicated by the appearance of an intense green/yellow color. This color will be strongest during the first deprotection following Boc-Asn(Xan) coupling, and may be observed during the next few deprotections.

(b) Cyclohexyl protecting groups should be used for aspartic and glutamic acid in place of benzyl protecting groups. In the presence of weak acid or base benzyl-protected aspartic acid, especially when next to a glycine, can undergo intramolecular cyclization to form a succinimide. The succinimide can be opened in two ways resulting in the desired -
linked peptide or the -linked isomer.\(^{28}\)

(c) Acetamidomethyl (Acm), 2,4-dinitrophenyl (DNP), 1,3-thiazolidine (Thz) and formyl (CHO) protecting groups are not labile to anhydrous HF and must be removed separately. See the accompanying paper for an Acm and Thz removal protocol.

(d) Boc-L-thiazolidine-4-carboxylic acid is used as a protected form of cysteine on the N-terminus of thioester peptides to prevent intramolecular cyclization.\(^{29}\) Following ligation, thiazolidine can be converted to cysteine by treatment with 0.2 M methoxylamine HCl at pH 4.0 as described in the accompanying paper.\(^{30,31}\)

(e) In the case of non-thioester containing peptides 2,4-Dinitrophenyl (DNP) is removed by thiolysis after chain assembly is complete, prior to HF cleavage.\(^{25}\) Briefly, swell the resin in DMF and treat it for 30min with a solution of 20% 2-mercaptoethanol, 10% DIEA in DMF. Wash the resin with DMF and repeat the treatment (if the solution is still yellowish after the second treatment repeat a third time). In the case of thioester containing peptides DNP protecting groups are removed after HF cleavage by treating the crude or purified peptide with 1000 equivalents of 2-mercaptoethanol at pH 6.8 in 6M GuHCl for 1h, which yields the peptide- thio(2-hydroxyl)ethyl ester. DNP is also quantitatively cleaved under transthioesterification conditions with MPAA as described in steps \(^{72-79}\). We recommend to remove DNP protecting groups prior to any ligation reaction, since the protecting group can irreversibly migrate to amino functionalities present in the polypeptide.\(^{32}\)

(f) In the case of non-thioester containing peptides formyl (CHO) protecting groups are removed after chain assembly is complete prior to HF cleavage by treatment of the resin with 10% piperidine, 5% water in DMF for 2h at 0°C. In the case of thioester peptides, the CHO groups must be left attached to the peptide until the thioester has undergone a NCL/KCL reaction. Cleavage in solution is carried out by treating the peptide with 20% piperidine, 38% 2-mercaptoethanol in aqueous 6M GuHCl solution for 40min at 0°C.\(^{33}\)
(g) Removal of Hoc protecting groups by HF treatment can give rise to side reactions, which can be prevented by adding Fmoc-Leu as a scavenger to the cleavage reaction (10 equivalents for each Trp(Hoc) residue)\textsuperscript{34}.

**Quantitative Ninhydrin Test\textsuperscript{35}**

**Reagents**

76\% (w/w) Phenol/ethanol solution (Applied Biosystems part no. 400463) ! CAUTION Toxic, Flammable
0.2mM Potassium cyanide/pyridine solution (Applied Biosystems part no. 400464) ! CAUTION Harmful, Flammable
0.28M Ninhydrin/ethanol solution (Applied Biosystems cat. no. 400465) ! CAUTION Flammable
60\% (v/v) ethanol solution in water
N,N-Dimethylformamide (DMF, B&J BioSyn -Peptide Synthesis Grade, Burdick & Jackson cat. no. BB075-4)
Dichloromethane (DCM, LC/GC grade, EMD cat. no. DX0831-1)

**Equipment**

Disposable fritted plastic SPE tube
Wastetrap for SPE tube
Vacuum source to provide negative pressure to the waste trap
Glass test tube
Heating block set to 100°C
UV-Vis spectrophotometer
Adjustable pipette capable of measuring 20\textmu L

**Procedure**

1| Using an adjustable pipette with the tip cut off to widen the opening, remove 20\textmu L of
resin/coupling solution slurry. Transfer the resin sample to the fritted SPE tube.

2] While applying vacuum to the waste trap, wash the resin briefly with DMF, wash the resin briefly with DCM, and dry the resin by air aspiration.

3] Transfer the dried resin to a tared test tube. Record the mass of the resin sample. Aim for about 5mg.

4] Add the following, in order, to the dry resin sample: 20μL phenol/ethanol, 40μL KCN/pyridine, 20μL ninhydrin/ethanol.

5] Place the test tube in the heating block at 100°C for exactly 5min.

6] This step can be performed using option A for qualitative results, or option B for quantitative results.

A. Observe the solution in the test tube for a blue/purple color indicating the presence of free primary amines and incomplete coupling.

B. For quantitative determination of coupling completeness:
   i. Add 3mL 60% ethanol solution
   ii. Measure the absorbance of the solution at 570nm
   iii. Calculate the percent coupling using the following equations:

   \[
   \frac{\mu\text{mol amine}}{g \text{ sample}} = \frac{A_{570} \times \text{dilution} \times 10^6}{\varepsilon_{570} \times \ell \times \text{mg resin sample}}
   \]

   • \( A_{570} \) is measured in a cuvette with a path length of 1 cm (commonly 1cm)
   • Dilution is the volume of 60% ethanol used in step 6.B.i (in mL)
   • \( \ell \) converts chromophore concentration from M to μM
   • \( \varepsilon_{570} = 1.57 \times 10^4 \text{ M}^{-1} \text{ cm}^{-1} \)
   • \( \% \text{ coupling} = \frac{1 - \frac{\mu\text{mol amine}}{g \text{ sample}}}{10^3 \times \text{resin substitution}} \times 100 \)
   • μmol amine/g sample is calculated with the previous equation
• resin substitution is in mmol/g. Note that as more peptide is synthesized on the resin, this value decreases-eventually several fold.
• $10^3$ converts mmol of peptide bound resin to $\mu$mol
• 100 converts the number from a decimal to a percent
Figure A.1. Schematic guide to the protocols. Preparation of the basic building blocks for convergent chemical protein synthesis: peptide−"COOH, peptide−"thioalkylester and peptide−"thio(4-carboxymethyl)phenyl ester (=peptide−"thioarylester) are highlighted in blue. Boc =tert-butyloxy carbonyl; Pam=phenylacetamidomethyl; TAMPAL= trityl-associated 3-mercaptopropionic acid leucine.
Figure A.2. Preparation of aminomethyl-copoly(styrene-divinylbenzene) resin.

Figure A.3. Boc solid phase peptide synthesis. (a) Reaction vessel used for manual solid phase peptide synthesis (SPPS). (b) Summary of steps for Boc SPPS.
Figure A.4. HF cleavage apparatus setup. Parts are labeled as followed: 200mL Cleavage vessel (1), reservoir vessel (2, not used), HF cylinder (3), CaO trap (4), vacuum pump (5), three way stop cock (6, vacuum pump vent), manometer (7), two way stop cock (8), light source for monitoring of HF condensate (9, optional), main HF tank valve (10).
Figure A.5. Boc SPPS of peptide– thioesters and transthioesterification with MPAA. HPLC profiles. (a) Crude peptide TTC(Acm)C(Acm)PSIVARSNFNA-COS-CH2CH2CO-Leu.amide (corresponding to Cram[Thr -Ala ]- thioester, compound 4) after HF cleavage from the resin. Side products of the synthesis are labeled as follow: 1: Phe-deletion; 2: HS-CH2CH2COOH-deletion; 3: Cys(Acm)-deletion; 5: Acm deletion; 6/7: p-cresol adducts. (b) Peptide after HPLC purification. The inset shows an ESI-MS spectrum. (c) Transthioesterification of Cram[Thr -Ala ]- alkylthioester (4) with MPAA after 4h. Product Cram[Thr -Ala ]- thio(4-carboxymethyl)phenyl ester (8) is indicated. Asterisks indicate impurities present in the purchased MPAA. (d) Purified Cram[Thr -Ala ]- thio(4-carboxymethyl)phenyl ester (inset: ESI-MS).
Appendix B. Total synthesis of proteins by convergent chemical ligation of unprotected peptides


[Draft manuscript, prepared by Durek, Pentelute, and Kent. Durek, Pentelute, and Torbeev did the experiments.]
Abstract

Preparation of proteins by total chemical synthesis enables almost unlimited variation of the covalent structure of the molecule, and is particularly useful for site-specific labeling of proteins for advanced biophysical studies. In this paper we provide detailed synthetic protocols that have proven generally applicable and useful in our laboratory. These protocols enable the assembly of medium-sized proteins from basic synthetic peptide building blocks (peptide-$\beta$thioesters; the synthesis will be described in an accompanying paper) by modern chemical ligations methods: native chemical ligation (NCL), and the recently developed kinetically controlled ligation (KCL). Subsequent folding, and disulfide formation where applicable, yields the synthetic protein molecule. The combination of these tools permits the assembly of proteins in a fully convergent fashion, giving researchers unprecedented flexibility for designing a synthetic route to target proteins of 200 or more amino acid residues. The time needed for the total synthesis, from scratch, of a target protein of the size of 40-140 amino acid residues is about 2-12 weeks.

**Keywords:** convergent synthesis; native chemical ligation; kinetically controlled ligation; chemical protein synthesis; peptide-thioester
Introduction

Total synthesis of proteins enables us to tailor the structure and the properties of a given target molecule in unprecedented chemical ways. The approach affords control over the polypeptide structure down to the single atom level, which has proven useful for structure-activity studies and has already contributed to an improved understanding of enzyme catalysis, protein folding, and the molecular basis of protein function\(^1\)-\(^3\). Alterations of the protein structure can include subtle changes such as the precise incorporation of isotopically labeled nuclei, defined posttranslational modifications, site-specific incorporation of biophysical probes, and backbone modifications of the polypeptide chain\(^4\). Chemical synthesis also enabled the preparation of novel protein analogs with improved pharmacological properties\(^5\).

Protein structural domains are the building blocks of the natural protein world and are typically in the size range of ~150 amino acid residues. The average protein found in nature is made up of two such domains and has a polypeptide chain of ~300 amino acid residues\(^6\). Thus, in order to prepare proteins of the typical size found in nature, we must be able to make ~150 residue domains and stitch these together to form the polypeptide chain of the target protein molecule. Even highly optimized stepwise solid phase peptide synthesis (SPPS)\(^7\) is only practical up to a size of ~50 amino acid residues. For this reason, it is necessary to make peptides of 50 or fewer amino acids and then stitch them together to make the longer polypeptide chains found in protein molecules. For practical reasons, the stitching together (‘ligation’) of unprotected peptide segments must be used in order to overcome peptide solubility problems, and to insure effective purification and characterization of synthetic intermediates\(^7\). The need for unambiguous covalent joining of unprotected peptide segments led us to develop ‘chemical ligation’ (i.e. chemoselective reaction) approaches\(^4, 8\)-\(^12\). Native chemical ligation\(^11\) – the thioester-mediated reaction of a peptide-\(^3\)thioester with an N-terminal cysteine-peptide to give a native amide bond at the ligation site (Figure 1) – is the most robust and useful amide-forming ligation chemistry developed to date\(^13\).

A general synthetic scheme for the synthesis of proteins by native chemical ligation is shown in Figure 2. In this scheme, unprotected peptides are prepared from
their constituent N\textsuperscript{a}-Boc amino acids by stepwise SPPS, and condensed by native chemical ligation to form the full-length polypeptide chain. The polypeptide is finally folded into the correct tertiary structure, with formation of disulfide bonds (if present).

Early application of the native chemical ligation approach involved a single ligation between just two peptides, the N-and C-terminal segments of the target polypeptide chain, and was used to make small protein molecules such as chemokines\textsuperscript{14}. Native chemical ligation was subsequently extended to the ligation of three or more segments, in which the chemical ligation steps are carried out in a sequential fashion. In such an approach all segments except the C-and N-terminal internal segments have dual reactivity (i.e. a C-terminal \textsuperscript{\text{\textgamma}}thioester and an N-terminal cysteine). In order to prevent side reactions (in particular cyclization\textsuperscript{15} and oligomerization), these synthetic designs involved the use of N-terminal-cysteine protecting groups, S-acetamidomethyl (S-Acm)\textsuperscript{16} or the 1,3-thiazolidine (Thz)\textsuperscript{17,18}, which permitted sequential ligations of three or more peptide building blocks in a “C terminal to N terminal” fashion (Figure 3a). Although C-to-N sequential ligations are no doubt useful\textsuperscript{18}, the linear nature of this inflexible approach is inherently rather limited.

For larger protein targets, a fully convergent synthetic strategy (Figure 3b) is required for total synthesis. The convergent approach requires some form of control over the reactivity of both the cysteine and \textsuperscript{\text{\textgamma}}thioester functionalities. While temporary N-terminal cysteine protection has turned out to be straightforward, robust means for controlling the reactivity of the C-terminal \textsuperscript{\text{\textgamma}}thioester have not been realized until now. We have recently exploited reaction rate differences between peptide-\textsuperscript{\text{\textgamma}}arylthioesters and peptide-\textsuperscript{\text{\textgamma}}alkyldithioesters to realize the total convergent synthesis of proteins (Figure 3b) by ‘kinetically controlled’ ligation\textsuperscript{19}. Kinetically controlled ligation relies on the fact that an \textsuperscript{\text{\textgamma}}arylthioester will react with a Cys-peptide much faster than the \textsuperscript{\text{\textgamma}}alkyldithioester (in the absence of added thiol). This difference in reaction rates can be used to control the outcome of a reaction involving two peptide-\textsuperscript{\text{\textgamma}}thioesters, so that only a single peptide-thioester ligation product is formed. The concept allows more flexible assembly strategies and has already proved to be useful for the synthesis of a variety of proteins in our laboratory (unpublished results).
The preparation of the basic building blocks (peptide-$^a$COOH and peptide-$^a$thioesters) by optimized Boc chemistry stepwise SPPS as well as a procedure for obtaining $^a$arylethioesters from the corresponding $^a$alkylethioesters will be described in an accompanying paper$^{20}$. In this protocol we report practical methodology for the subsequent steps of the convergent chemical synthesis of a protein molecule: native chemical ligation (NCL, steps 1-8) and kinetically controlled ligation (KCL, steps 15-24) of the purified peptide building blocks and folding of the resulting full-length polypeptide to give the synthetic protein molecule (steps 33-41). The entire process is schematically illustrated in Figure 3b.

**Design of a synthetic strategy**

When planning any total chemical protein synthesis strategy, selection of appropriate ligation sites is critical to the success of the synthesis. Linking peptide segments by native chemical ligation requires cysteines, which ideally should be located at appropriate sites in the wild-type protein. This means that without disturbing the wild-type structure, cysteines need to be evenly distributed across the entire target sequence with a spacing of no more than about 30-50 amino acid residues. This is often not the case for the majority of proteins (cysteine is one of the rarest amino acids found in proteins). The reliance of NCL on cysteine ligation sites is clearly a limitation of the methodology, which has provoked the development of a number of approaches attempting to circumvent this restriction. Worth mentioning here are: auxiliary mediated ligations$^{21-23}$, which can in favorable cases extend NCL to Xaa-Ala/Gly ligation sites (where Xaa is any amino acid); the desulfurization of cysteine to alanine following a NCL reaction$^{24}$; and alkylation of the cysteine sulfhydryl group formed in a NCL reaction (to give homo-glutamate/glutamine analogs)$^5$. In practical terms it is often simplest to incorporate cysteine residues into the sequence by replacing amino acids such as alanine, serine or threonine. In most cases the resulting mutations can be expected to remain silent and do not impact protein functioning. The minimal impact of introduced Cys residues has been demonstrated in a number of cases by systematic site-
directed mutagenesis and site-specific chemically labeling of single cysteine residues with a variety of abiotic probes (‘cysteine-scanning mutagenesis’)\textsuperscript{25,26}.

When evaluating different potential ligation sites it is also important to consider the nature of the C-terminal amino acid in the thioester segment (Xaa in a Xaa-Cys ligation). It is well-documented that the identity of the C-terminal amino acid in thioester peptides affects their reactivity\textsuperscript{27}, and until recently use of Val, Ile, Thr and Pro thioesters has been avoided due to their characteristically slow ligation rates. These slow rates can now be taken advantage of in a kinetically controlled ligation to inhibit side reactions of the thioalkylester peptide. Subsequently, native chemical ligation at these traditionally unfavorable sites can be effected by the use of the newly introduced 4-mercaptophenylacetic acid\textsuperscript{28} as a ligation catalyst\textsuperscript{29}. The use of Asp-Cys or Glu-Cys ligation sites requires protection of the side-chain carboxyl group of the corresponding Asp/Glu residues, since they can give rise to side reactions during ligation reactions\textsuperscript{30}. An additional consideration especially when planning a kinetically controlled ligation is the presence of non-igation site cysteines in either “thioester peptide. Transthioesterification with these “internal” cysteine sulphydryls (e.g. thiolactone formation) results in loss of the activated aryl thioester. We therefore recommend to protect non-igation site cysteines as Cys(Acm) to prevent this undesired process.

**Native chemical ligation and kinetically controlled ligation**

NCL and KCL are robust reactions that are usually performed in aqueous buffers at around neutral pH values. In order to obtain high reaction rates it is critical to work at relatively high peptide concentrations (>1mM). For this reason most reactions are carried out in aqueous 6M guanidine HCl (GuHCl). GuHCl is a strong chaotrope and substantially enhances the solubility of unprotected peptide segments.[Need a Ref., NCL, SCIENCE 1994 paper ?] Another critical parameter is the pH value of the reaction medium. The initial reversible transthioesterification step is normally the rate determining step of the reaction and is initiated by a nucleophilic attack of the N-terminal cysteine sulphydryl group on the “thioester carbonyl carbon”\textsuperscript{28}. Since the thiolate
form of the cysteine is the effective nucleophilic species, it is advisable to work at or around a pH of 7.0. Higher pH values should be avoided in order to not compromise the chemoselectivity of the ligation reactions.

In NCL, the observed reaction rate is also dependent on the concentration of exogenous (added) thiol catalyst\textsuperscript{28}. The thiol catalyst retains the cysteine side chains in the reduced state, reverses the formation of unproductive thioesters (involving non-ligation site cysteines) and activates less reactive peptide-“thioesters by transthioesterification”\textsuperscript{29}. Although a number of small molecule compounds have been used as thiol catalysts in the past, a recent thorough study revealed major differences in their ability to accelerate NCL reactions\textsuperscript{28}. 4-mercaptophenylacetic acid (MPAA) for a number of reasons emerged as a far superior thiol catalyst when compared to any of the established and widely used thiol compounds. Beside its beneficial effects on NCL reactions, MPAA has proven also to be highly efficient for separate (preparative) activation of peptide-“thioalkylesters by transthioesterification to the corresponding peptide-“thio(4-carboxymethyl)phenyl esters. The latter peptide-“thioarylestes are extremely useful for KCL reactions or for acceleration of NCL reactions involving unfavorable ligation sites (see above).

Because KCL relies on the difference between reactivities of “thioarylestes versus “thioalkylesters, it is vital to minimize any non-productive transthioesterification reactions. Thus KCL reactions have to be carried out in the absence of exogenous (added) thiol catalyst. In addition we recommend protection of all non-ligation site cysteines (e.g. with S-Acm). However, unproductive transthioesterification reactions cannot be completely avoided, which is typically indicated by the appearance of side products in which the ligation site cysteine of the product has been acylated by an additional transthioesterification reaction (thiolactone and/or branched trimer formation). The resulting non-productive thioester species are stable in the absence of added (exogenous) thiol catalyst; they can be decomposed to reveal the desired linear product by adding a large excess of small molecule thiol compound when the KCL reaction is completed (i.e. after most starting material has been consumed).
Another potential side reaction during KCL is cyclization or oligomerization of the bifunctional polypeptide segment (i.e. the segment that has both an thioalkylester and N-terminal cysteine), through amide bond formation. In our hands these side reactions are of lesser significance (<1 or 2%) and do not pose any major problem. In the methods section below we give detailed protocols enabling researchers to perform NCL (steps 1-8), KCL (steps 15-24), and for removal of temporary protecting groups such as Thz (steps 9-14) and S-Acm (steps 25-32) (see Figure 3b).

**Case study: Crambin**

To demonstrate the application of these protocols to the synthesis of proteins, we describe the fully convergent synthesis of the 46 amino acid residue membrane-associated protein crambin (Cram) from four synthetic peptide building blocks (Figure 4). The key transformation in this synthetic scheme is the unambiguous ligation of Cram[Thr1-Ala15]-thioarylester and Cram[Cys16-Leu25]-thioalkylester to give a single product peptide-thioester. Because peptide-arylthioesters react substantially faster with N-terminal-Cys peptides than their less-reactive alkylthioester counterpart, the reaction proceeds to give Cram[Thr1-Leu25] thioalkylester with only trace amounts of side-products. The C-terminal half of the target sequence, Cram[Cys26-Asn46], is assembled from segments Cram[Thz26-Gly31]-thioalkylester and Cram[Cys32-Asn46] by native chemical ligation followed by conversion of Thz to Cys. The two halves are then covalently joined by another native chemical ligation reaction to give the full-length crambin polypeptide. Finally, removal of S-Acm protecting groups and folding with concomitant disulfide formation yields the native protein.
Materials
Lyophilized purified peptide building blocks (peptide-^4^thioesters and peptide-^5^COOH)
The synthesis of these building blocks will be described in an accompanying paper^20.^
Guanidine hydrochloride (GuHCl, Fisher cat. no. BP178-1)
4-Mercaptophenylacetic acid (Sigma-Aldrich cat. no. 653152, 97% purity)
CRITICAL -Commercially available sources of MPAA contain impurities that might interfere
with peptide analysis and purification. In cases where this becomes an issue,
we recommend purification of MPAA prior to use in ligation reactions. MPAA is
conveniently purified by RP-HPLC on C18 columns.
Tris(2-carboxyethyl)phosphine hydrochloride (TCEP, Fluka cat. no. 93284)
Methoxylamine hydrochloride (Sigma-Aldrich cat. no. 226904)
Sodium 2-mercaptoethanesulfonate (Sigma-Aldrich cat. no. M1511)
Silver acetate (Acros Organics cat. no. 202390250)
Sodium phosphate, dibasic, anhydrous (Na$_2$HPO$_4$, Fisher cat. no. BP332-1)
1,4-Dithiothreitol (DTT, Sigma-Aldrich cat. no. 43815)
L-Cysteine (Fluka cat. no. 30089)
L-Cystine-2HCl (Peptides International cat. no. ALC-2706-PI)
Acetonitrile (HPLC grade, aldehyde free) ! CAUTION Flammable, toxic
Trifluoroacetic acid (TFA, Biograde, Halocarbon) ! CAUTION Corrosive, causes severe burns
Deionized water
Helium cylinder
Argon cylinder
**Equipment**

Reaction vessels (e.g. 1.5, 2.0mL Eppendorf tubes, 20mL Scintillation vials, depending on scale)
Spatulas, pipettes, disposable syringes, magnetic stir bars and magnetic stirrer, shaker, 0.2μm sterile filters
pH-Meter (with probes suitable for measuring large and small volumes (μL to mL)
Balance (mg to g scale)
High Performance Liquid Chromatography (HPLC) system for analytical and preparative purifications equipped with a UV detector.
HPLC analytical (e.g. 2.1-4.6x50-250mm) and preparative (e.g.10-22x250mm) reversed phase (RP) columns with different alkyl substitutions for hydrophobic/hydrophilic peptide analysis and purification (e.g. C4, C8, C18).
Mass spectrometer (e.g. MALDI-TOF-MS, ESI-MS, LC-MS, etc.)
Lyophilizer (freeze-drying apparatus)

**Equipment Setup**

**Analytical RP-HPLC**

Analytical separations in our lab are performed on a Agilent 1100 system equipped with a binary pump, autoinjector, column thermostat, UV/Vis diode array detector and Agilent 1100 ion trap mass spectrometer (http://www.ionsource.com) using narrow and standard bore columns. Although coupled LC-MS analysis is no doubt useful, it is not mandatory for the experiments described in this paper; off-line MS analysis of collected peaks works just as well. Analytical runs are always a good starting point for a later preparative purification. We typically use full linear gradients (5-65% buffer B in buffer A over 15min at a flow rate of 0.5mL/min) in order to detect all impurities present in the sample and to identify the concentration of organic modifier at which the desired product gets eluted from the column.
Experimental specifications for the HPLC

In our lab the following configurations are applied: for narrow bore columns (2.1 x 50mm, 3µm) a flow rate of 0.5mL/min; for standard bore columns (4.6 x 150mm, 3-5µm) a flow rate of 1-2mL/min; for semi-preparative columns (10 x 250mm, 5-10µm) a flow rate of 5-10mL/min; for preparative columns (22 x 250mm, 7-15µm) a flow rate of 10-20mL/min.

Peptide solubilization for HPLC

The peptide needs to be dissolved in a minimal volume of a suitable solvent to permit adsorption of the peptide to the stationary phase (i.e. low concentrations of organic modifier in water) and does not interfere with the chromatographic separation (e.g. detergents). Solvents containing high concentrations of strong acids and bases should be avoided, since they reduce column and hardware lifetime. A good starting point to dissolve a peptide, is always a low-percent mixture of buffer B in buffer A (e.g. 5-20% (v/v)) or a low-percent mixture of an alternative organic modifier (ethanol, methanol, 2-propanol) in water. If peptide solubility is rather limited in these solvents, 6M GuHCl in water (pH 2-7) is often an excellent alternative. The sample and all HPLC buffers should be filtered through a 0.22µm filter to remove any insoluble matter.

Preparative HPLC purification using gradient elution

The preparative purification is carried out on a column with very similar or (ideally) identical stationary phase characteristics as the analytical column. In order to obtain a similar elution profile one has to maintain a proportional gradient, which is achieved by keeping the ratio between the gradient volume to the column volume constant and adjusting gradient time or flow rate accordingly. However, in practice the gradient should be made less steep in order to increase resolution. After development of a suitable gradient, the optimal amount of sample that can be loaded onto the column needs to be determined. We typically start out by analyzing gradually increasing quantities of sample under always the same preparative run conditions on the same preparative column (for example by injecting 1, 10, 50, 100, 200, 300mg, etc. of crude peptide dependent on column size). The effect of the increasing sample amount on peak
shape, resolution, and purity (determined by off-line fraction analysis) is used to
determine the optimal amount of sample that can be loaded onto the column. The aim of
this operation should be to find the right balance between throughput, yield, purity and
resolution. Once reproducible conditions have been established several preparative runs
can be performed in a routine and automated fashion in short time.

Collected fractions are then analyzed for desired product either by analytical
HPLC, ESI-MS or by MALDI-TOF-MS, are pooled based on compound purity and
lyophilized.
Procedure

Native chemical ligation\textsuperscript{28} TIMING ~4-16h

1| Prepare a stock solution of 6M guanidine hydrochloride (GuHCl), 0.2M sodium phosphate. For a 50mL solution, mix together 28.7g (300mmol) of GuHCl and 1.42g (10mmol) of anhydrous Na\textsubscript{2}HPO\textsubscript{4} and add deionized water to 50mL final volume. Stir until fully dissolved. Filter the solution through a 0.2µm sterile filter and degas it for at least 20min by sparging with helium: connect a clean fine needle or a sparging element to a helium cylinder and immerse it into the solution. Let a continuous stream of gas flow through the solution.

PAUSE POINT The solution can be stored at room temperature for several days, but should be freshly filtered and degassed prior to continuing with step 2.

2| Prepare the native chemical ligation buffer (6M GuHCl, 0.2M sodium phosphate, 20mM TCEP, 30mM MPAA). For 10mL of solution, mix together 57.3mg (0.2mmol) of tris(2carboxyethyl)phosphine hydrochloride (TCEP) and 50.5mg (0.3mmol) of 4-mercaptophenylacetic acid (MPAA) and add 10mL of the buffer prepared in the previous step. Mix until everything has dissolved (solubility of MPAA is highly pH dependent; increase the pH slightly by adding a few drops of 2M NaOH in case solubilization is difficult). Adjust the pH to 6.8-7.0 and keep the solution under argon atmosphere until further use.

CRITICAL STEP Composition and especially pH of the ligation buffer are critical for a successful native chemical ligation reaction. Make sure the solution is always freshly prepared and take great care when adjusting the final pH. Calibration of the pH meter prior to each measurement is highly recommended.

3| Weigh out equimolar amounts of both peptides and mix peptides as a powder in an appropriate reaction vessel (for ligation reaction volumes smaller than 2mL we typically use Eppendorf reaction tubes and a Eppendorf shaker; for larger volumes disposable
20mL glass scintillation vials and magnetic stirring are employed).

4| Add native chemical ligation buffer (prepared in step 2) so that the final concentration of each peptide is in the range of 1-5 mM. Purge the reaction vessel with argon by letting a stream of argon from a cylinder flow into the vessel for about 10sec. Seal the reaction vessel. Place the solution on a stirrer or shaker and make sure the peptides dissolve completely.

**CRITICAL STEP** The reaction rate of a native chemical ligation reaction is directly proportional to the product of the concentrations of the peptide segments (the higher the concentrations, the higher the reaction rate). However, limited solubility of starting peptides and/or reaction product might require optimization in order to find the best conditions. In most cases concentrations of 1-5mM of each peptide give satisfactory rates.

5| Remove a small aliquot from the reaction mixture and quench it by diluting it 10 fold with 50% (v/v) acetonitrile, 0.1% (v/v) TFA in H₂O solution. Analyze the sample by your preferred method (we highly recommend LC, MALDI-TOF-MS, LC-MS).

6| Check the pH of the reaction mixture and readjust if necessary to 6.8-7.0. Afterwards purge the reaction vessel with argon and seal it.

**CRITICAL STEP** Avoid over-titration especially when adding bases. Exposure of your ligation reaction to basic pH values might result in loss of chemoselectivity and could give rise to other side reactions.

7| Repeat steps 5 through 6 once an hour until most starting material has been consumed and a single product has been formed. As was mentioned before, native chemical reaction rates depend on a number of factors (pH, peptide concentrations, nature of the C-terminal amino acids in the thioester segment, thiol catalyst), however typically reactions are complete within 1-6 hours under the described conditions. **TROUBLESHOOTING.**

8| If your ligation product contains an N-terminal Thz-protected cysteine residue continue
directly with step 10 without further manipulations. Otherwise, purify your ligation product using RP-HPLC (see Box 3 for comments on purification).

**Cysteine deprotection I (Thz → Cys Conversion)** \(^{17,18}\) **TIMING ~3h**

9| (Optional) Prepare a stock solution of 6M GuHCl, 0.2M sodium phosphate (see step 1 for details). Dissolve the Thz-protected peptide at a peptide concentration of about 1-5mM in this buffer and place in a suitable reaction vessel. **CRITICAL STEP** If the peptide is already dissolved in an aqueous buffer, such as 6M GuHCl, 0.2M sodium phosphate (e.g. at the end of a NCL reaction) go directly to step 10.

10| Add solid methoxylamine hydrochloride to the solution to give a final concentration of 0.2M (16.7mg methoxylamine HCl/mL of ligation reaction). Continue stirring at room temperature until the reagent is completely dissolved.

11| Check the pH of the reaction mixture and adjust if necessary to pH 4.0. Continue stirring at room temperature for 2h. **CRITICAL STEP** The pH of the reaction will affect the conversion rate with the optimum at around pH 4.0.

12| Take an aliquot of the reaction and analyze by LC, LC-MS or MALDI-TOF-MS (analysis by a MS method is recommended, since for most peptides the shift in retention time will be marginal). **TROUBLESHOOTING**

13| Repeat steps 11 through 12 until deprotection is complete (reactions typically take 2-3 hours for quantitative conversion). **PAUSE POINT** Reactions can also be left to incubate overnight at room temperature.

14| Purify the reaction product by RP-HPLC (see Box 3 for comments on purification).
**Kinetically controlled ligation**\(^{19}\) **TIMING ~3-6h**

15| Prepare a stock solution of 6M Guanidine hydrochloride (GuHCl), 0.2M sodium phosphate (see step 1 for details). Adjust the pH to 6.3-6.8 and keep the solution under argon atmosphere until further use.

**CRITICAL STEP** Composition and especially pH of the ligation buffer are critical for a successful kinetically controlled ligation reaction. Make sure the solution is always freshly prepared.

16| Weigh out equimolar amounts of both peptides and mix peptides as a powder in a appropriate reaction vessel (for ligation reaction volumes smaller than 2mL we typically use Eppendorf reaction tubes and an Eppendorf shaker; for larger volumes, disposable 20mL glass scintillation vials and magnetic stirring are employed).

17| Add ligation buffer (prepared in step 15), so that the final concentration of each peptide is in the range of 1-5mM. Place solution on a stirrer or shaker and make sure peptides dissolve completely.

**CRITICAL STEP** The reaction rate is directly proportional to the concentration of each peptide segment. In most cases concentrations of 1-5mM of each peptide give satisfactory rates.

18| Remove a small aliquot from the reaction mixture and quench it by diluting it 10 fold with 50% (v/v) acetonitrile in water containing 0.1% (v/v) TFA solution. Analyze the sample by your preferred method (we highly recommend LC, MALDI-TOF, LC-MS).

**TROUBLESHOOTING**

19| Check the pH of the reaction mixture and readjust if necessary to 6.3-6.8. Afterwards purge the reaction vessel with argon and seal it.

**CRITICAL STEP** Avoid over-titration especially when adding bases. Exposure of your ligation reaction to basic pH values might result in loss of chemoselectivity and could give rise to other side reactions.
20] Repeat steps 18 through 19 once an hour until most starting material has been consumed. Reactions can be expected to reach that point within 2-6 hours under the described conditions. Kinetically controlled ligation typically yields complex product mixtures, which are the result of extensive transthioesterification reactions, which are essentially irreversible in the absence of excess exogenous thiol catalyst. These side-products can be resolved with an excess of thiol agent, which is added after most starting material has been consumed.

21] Dilute the reaction mixture with ligation buffer from step 15 so that the peptide concentration is around 1mM. Add sodium 2-mercaptoethanesulfonate (MESNA) as a powder to give a final concentration of 0.2M. Readjust the pH if necessary to 6.3-6.8. **CRITICAL STEP** Avoid over-titration especially when adding bases. Exposure of your ligation reaction to basic pH values might result in loss of chemoselectivity and could give rise to other side reactions.

22] Continue stirring for another 30min at room temperature.

23] Remove a small aliquot from the reaction mixture, quench it by diluting it 10 fold with 50% (v/v) acetonitrile in water containing 0.1% (v/v) TFA solution, and analyze it by your method of choice. Treatment of the crude KCL reaction mixture with excess of thiol agent (e.g. MESNA) should yield a dominant ligation product (see “Anticipated results” section). **TROUBLESHOOTING**

24] Purify the reaction product by RP-HPLC (see Box 1 for comments on purification).

**Cys deprotection II (Acm removal)**[^31][^33] **TIMING ~3h**

25] Dissolve the lyophilized peptide in 50% (v/v) acetonitrile in water containing 0.1% (v/v) TFA to give a final peptide concentration of about 2-3mg/mL (if solubility is an issue, the fraction of acetonitrile can be varied; 50% (v/v) acetonitrile/water, 0.1% (v/v) TFA will work for most peptides).
26| For each mmol of Acm group present, add 5mmol of powdered silver acetate (AgOAc) to the solution.

27| Stir at room temperature for 1h.

28| Take an aliquot and quench it by diluting it 10 fold with 6M GuHCl, 0.2M DTT, pH 7.0. Wait for 10min, centrifuge the sample to remove precipitate and analyze the supernatant by LC, LC-MS or MALDI-TOF-MS (reactions are typically done after 1-2h).

29| If the reaction is not complete go back to step 27.

30| Add 25mmol of solid 1,4-dithiothreitol (DTT) for each mmol of Acm group present in the peptide. A precipitate will form immediately. Stir for another 15min.

31| Transfer the reaction mixture to a centrifugation vial and centrifuge until precipitate and supernatant are clearly separated.

**CRITICAL STEP** In some cases a significant portion of the reaction product co-precipitates together with the Ag-DTT complex. In this case, extract the precipitate several times with a suitable solvent, such as 6M GuHCl in water. Check the supernatant after each extraction for reaction product, combine the extracts that contain the desired peptide and purify them using HPLC.

32| Purify the supernatant by HPLC (see Box 3 for comments on purification).

**Folding TIMING ~4h**

33| Prepare 10mL of a 6M GuHCl in water stock solution. Weigh out 5.74g (60mmol) of GuHCl and add deionized water to 10mL final volume. Stir or shake until everything is dissolved. The pH should be slightly acidic (pH 4.5-5.5) to minimize spontaneous oxidation of cysteines. Adjust the pH if necessary. Filter the solution through a 0.2µm sterile filter and degas it for at least 20min by sparging with helium.
34| Prepare 10mL of the folding buffer, consisting of 100mM Tris-HCl, 9.2mM L-cysteine, 1.2mM L-cystine in water. Weigh out 121.1mg Tris (base), 11.1mg L-cysteine and 3.8mg of L-cystine·2HCl and add deionized water to about 9mL. Stir or shake until everything is dissolved. Adjust the pH to 7.8-8.2 by carefully adding concentrated HCl. Use a volumetric flask to adjust the volume to 10mL and filter the solution through a 0.2μm sterile filter. Degas the solution for at least 20min by sparging with helium: connect a clean fine needle or a sparging element to a helium cylinder and immerse it into the solution. Let a continuous stream of gas flow through the solution. Transfer the solution to a suitable refolding vessel and add a magnetic stir bar. Keep the solution under argon until further use.

**CRITICAL STEP:** Solubility of L-cystine·2HCl might be limited, depending on the buffer composition. If this is an issue, try sonication and/or warming up the suspension to about 40°C.

35| Weigh out ~6mg of lyophilized and fully reduced peptide. Dissolve the peptide in 2mL of denaturation buffer prepared in step 33. Make sure the peptide dissolves completely by slightly agitating the reaction vessel. Filter the solution (or spin down) to remove any insoluble material.

36| Dilute the peptide in denaturation buffer by adding the 2mL solution to the 10mL of folding buffer. Make sure both solutions mix rapidly (stirring).

37| Check the pH of the folding solution and adjust if necessary to pH 7.8-8.2. Keep the mixture under argon and seal the reaction vessel. Continue to stir for 1h.

38| Analyze an aliquot of the reaction mixture by a suitable method (LC/LC-MS/biochemical assays).

**CRITICAL STEP:** In the case of disulfide-containing proteins, formation of the correct tertiary structure is often indicated by an RP-HPLC retention time shift of the folded versus the denatured polypeptide. Correct formation of disulfides can also be detected by means of mass spectrometry. If the protein has a known biological activity, proof for
formation of the correct tertiary structure is achieved by biochemical assays.

39| If folding is not complete go back to step 37.

CRITICAL STEP The success of a folding experiment strongly depends on the nature of the target protein molecule and the exact folding conditions employed. See Box 2 for general guidelines. Please consider the present section of the protocol (steps 33-41) as an excellent starting point for the folding of small disulfide containing proteins that has proven itself in our lab.

40| Transfer the folding solution to a centrifugation vessel and remove aggregated material by centrifugation.

41| Purify the supernatant by your method of choice (See Box 1 and Box 2 for general guidelines).
## TROUBLESHOOTING

**Table B.1.** Troubleshooting section for NCL, KCL

<table>
<thead>
<tr>
<th>Problem</th>
<th>Possible Reason</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slow/No Ligation (NCL and KCL, steps 7, 18, and 23)</td>
<td>Peptide concentration too low</td>
<td>Increase the concentration of each peptide to at least 1mM.</td>
</tr>
<tr>
<td></td>
<td>Unfavorable ligation site (i.e. Ile/Thr/Val/Pro-to-Cys ligation sites)</td>
<td>1. Use alternative ligation sites or exchange the unfavorable residue (e.g. to alanine or 2-aminobutyric acid). Avoid Pro-Cys ligation sites.</td>
</tr>
<tr>
<td></td>
<td>pH too low</td>
<td>Adjust pH to 6.8-7.0 (NCL).</td>
</tr>
<tr>
<td></td>
<td>Oxidation of ligation site cysteine</td>
<td>Adjust pH to 6.3-6.8 (KCL). Avoid exposure of your ligation reaction to air. Degas reaction buffer with helium and keep solution under argon in a sealed container. Add TCEP to reduce disulfides.</td>
</tr>
<tr>
<td>Slow/No NCL (step 7)</td>
<td>No thiol catalyst added/oxidation of thiol catalyst</td>
<td>Add thiol catalyst. Degas buffer prior to reaction start. Keep reaction solution under argon atmosphere in a sealed container.</td>
</tr>
<tr>
<td>Side products during KCL (steps 18 and 23)</td>
<td>Thiol catalyst added</td>
<td></td>
</tr>
</tbody>
</table>
Anticipated Results

For demonstration purposes we undertook the synthesis of a small (46 amino acid residues) disulfide linked protein. The full-length crambin¹⁸,³¹ polypeptide (TTCCPSIVARSNFACRLPGTPEALCATYTGCIIPGATPGDYN; with ligation sites underlined) was assembled from 4 segments using the convergent strategy shown in Figure 3b. The peptide segments are: Cram[Thr-Ala]¹⁵-“thio(4-carboxymethyl)phenyl ester, Cram[Cys-Leu]¹⁶²⁵-“thioalkylester, Cram[Thz-Gly]²⁶³¹-“thioalkylester and Cram[Cys-Asn]³²⁴⁶. It should be noted that due to the diverse chemical nature of polypeptides the anticipated results for other targets can be expected to deviate in detail from the data shown here (such as yields, reaction rates, HPLC profiles, etc.). Still, ongoing work in our and other labs suggests that the devised protocols and the underlying chemistry are generally applicable to a wide range of peptides and proteins. Protocols describing the synthesis of the peptide building blocks including preparation of “thio(4-carboxymethyl)phenyl esters from the corresponding alkylthioesters will be described in the accompanying paper²⁰.

A typical native chemical ligation reaction is shown in Figure 4. The two peptides (Cram[Thz-Gly]²⁶³¹-“thioalkylester and Cram[Cys-Asn]³²⁴⁶) (0.043mmol each) were ligated under native chemical ligation conditions at pH 7.0 and at a concentration of each peptide of 2.2mM (as described in steps 1-8). Figure 4a shows the reaction shortly after mixing of both peptides. After 4 hours the reaction was essentially complete (Figure 4b), after which Thz was quantitatively converted to Cys by alkoxamine treatment at pH 4 (steps 9-14). The reaction product was purified by semi-preparative HPLC (Figure 4c). After lyophilization 60mg (0.028mmol) of purified product Cram[Cys-Asn]²⁶⁴⁶ were obtained, which corresponds to a overall yield of 65% for the native chemical ligation, Thz to Cys conversion and purification.

Data for a kinetically controlled ligation are presented in Figure 5. Peptides Cram[Thr-Ala]¹⁵-“thio(4-carboxymethyl)phenyl ester and Cram[Cys-Leu]¹⁶²⁵-“thioalkylester (each 0.019mmol) were reacted at pH 6.8 in the absence of any added
thiol catalyst at a peptide concentration of 2.1 mM. HPLC analysis of the reaction shortly after mixing of both peptides (Figure 5a) and after 2h (Figure 5b) indicated formation of the desired product (Cram[Thr Leu\textsuperscript{25}]\textsuperscript{-}thioalkylester, compound 5 in Figure 5b), as well as the formation of some side-products, due to unproductive transthioesterification reactions (compounds 3 and 4 in Figure 5b). The reaction mixture was simplified by transthioesterification with MESNA (0.2 M, for 30 min, pH 7.0), as shown in Figure 5c, after which the dominant product was purified by semi-preparative HPLC (Figure 5d). Cram[Thr -Leu\textsuperscript{25}]\textsuperscript{-}MESNA thioester was isolated in good yield (0.011 mmol, 58%).

Cram[Thr -Leu\textsuperscript{25}]\textsuperscript{-}MESNA thioester and Cram[Cys\textsuperscript{1-46} -Asn\textsuperscript{26}] (each 5.2 µmol) were finally linked by native chemical ligation (as described in steps 1-8) at a peptide concentration of 1 mM (see Figure 6a,b). Subsequently, the Acm groups present on cysteines 3, 4 and 40 were removed as described in steps 25-32 (yield for final ligation and Acm removal after purification: 2.7 µmol of Cram[Thr -Asn\textsuperscript{46}], 52%, Figure 6c). The reduced polypeptide was folded in 1 M GuHCl, 0.1 M TrisHCl, 8 mM cysteine, 1 mM cystine, pH 7.8 at a peptide concentration of 0.5 mg/mL as described in steps 33-41, which afforded the disulfide-containing protein. The HPLC chromatogram of the purified final reaction product is shown in Figure 6d, which underscores the high purity of the obtained material (isolated yield for folding and final purification: 92%). Further studies should now be carried in order to demonstrate native protein structure and/or function (e.g. crystallography/X-ray analysis, NMR, biochemical assays).
References:


Reversed phase HPLC of peptides

This section aims at providing some very general guidelines for the analytical and preparative separation of complex peptide mixtures, such as those obtained from SPPS or after a chemical ligation reaction. RP-HPLC is the most often applied and most efficient technique in this regard. Other separation methods, such as ion-exchange chromatography, hydrophobic interaction chromatography, size exclusion chromatography or solid phase extraction might prove useful for certain types of peptides or applications, but for the sake of conciseness will not be covered here. It should be noted that due to the diverse nature and physicochemical properties of polypeptides as well as a number of other factors (see below) we cannot provide a universally valid protocol for preparative purification of peptides, but rather try to summarize a number of critical factors as well as some experimental guidelines that turned out to be useful for analyzing and purifying most peptides in our lab (for more information, see http://www.vydac.com/pub_content.html).

General remarks

A number of critical factors will affect the quality of a RP-HPLC separation experiment. The most important parameters are: Column size; physical parameters of the stationary phase (particle size, particle size distribution, pore size, and pore size distribution); chemical characteristics of the stationary phase (silica/non-silica, bonding (C4, C8, C18, phenyl, etc.), carbon loading, end-capped/non end-capped); mobile phase (ion-pairing agents, organic modifier, pH); flow rate; column temperature and last but not least the sample itself (solvent, volume, concentration and physicochemical properties of the compound of interest). A discussion of all of these parameters is beyond the scope of this paper, but can be found in the cited monographs. Optimization of stationary and mobile phases, flow rate, and column temperature are the most easily controllable parameters in an HPLC experiment. As for the stationary phases, we will focus entirely on silica-based phases.

In HPLC practice, the smaller the particle size the more efficient the column will be, but smaller particles also give rise to an increased column back pressure. The pore size is a critical parameter which will affect the ease of mass transfer (diffusion) of the sample through the pores of the particles. It is important to select the right pore size for a
given separation problem. For short peptides (<15 amino acids) and compact globular small proteins a pore size of 100Å typically gives best results, while longer random-coil polypeptides and proteins require larger pore sizes (200-300Å). Large pore size particles, however, have a comparatively low surface area and therefore lower loading capacity. In addition, large pore size silica are more fragile especially when operated under high pressure conditions, which might result in shorter column lifetime. Choosing an appropriate bonding type (i.e. C4, phenyl, C8, C18) is the next critical step. As a rule of thumb, the higher the molecular weight of the peptide is, the less hydrophobic the alkyl substitution should be (i.e. C18 for hydrophilic peptides shorter than 10 amino acids, C8 and C4 for 10-40 amino acid polypeptides depending on hydrophobicity, and C4 for peptides longer than 40 amino acid residues).

Column temperature can have a decisive effect on chromatographic separations, column backpressure and compound retention (in reversed phase mode compounds tend to elute earlier with increasing temperatures). Although the effects are hardly predictable it is definitely worth trying alternative temperatures (25-60°C) and compare separation efficiency. Standard separations are carried out at 40°C in our lab. It should be noted that higher temperatures also result in shorter column lifetime. The most common mobile phase system used in RP-HPLC of peptides and proteins is a binary mixture of {water containing 0.1% TFA (buffer A)} and {acetonitrile containing 0.08% TFA (buffer B)}. The acid is used to suppress the ionization of free silica silanol groups, which might give rise to undesirable ionic interactions and through ion pairing helps to improve peak shape. Unfortunately TFA is known to suppress MS ionisation, which is the reason why many LC-MS systems utilize formic acid as an alternative. However, formic acid should not be used for preparative peptide purifications since its use might lead to peptide formylation. A practical alternative for TFA for preparative purifications can be triethylammonium phosphate (TEAP), sodium phosphate, or ammonium acetate at different pH values. As an alternative organic modifier 2-propanol, ethanol or methanol may be used instead of acetonitrile, which might prove useful for the separation of very hydrophobic peptides. The flow rate affects mass transfer and hence peak broadening and chromatographic efficiency and in principle needs to be optimized for each separation. When choosing a flow rate, it is important to consider the backpressure
generated by the column (depending mainly on particle size, column dimensions and mobile phase) in order to preserve column lifetime and hardware.

**In vitro protein folding**\(^{42,43}\)

Protein *in vitro* folding, i.e. the process by which an unstructured polypeptide chain acquires its genuine three dimensional (tertiary) structure, is to a high degree dependent on the nature of the target polypeptide chain or protein. Therefore conditions that turn out to work for a given protein might prove unsuccessful for an unrelated protein. Therefore this chapter can only provide some general guidelines rather than a “universal recipe” for folding of synthetic polypeptide chains. Check out the refold database to find out if conditions for your protein of interest have been established previously (http://refold.med.monash.edu.au). **General remarks:** Folding of the polypeptide chain is typically initiated by decreasing the concentration of a denaturant (usually GuHCl or urea) below a certain threshold value. This value is an intrinsic property of a given protein molecule and needs to be determined experimentally. Decreasing the concentration of denaturant can be achieved by a number of ways: dilution, dialysis and on-column buffer exchange. Dilution is by far the easiest and most often applied method. Upon lowering the denaturant concentration the polypeptide chain will start to fold into its tertiary structure. However, a major side reaction during this operation is often the aggregation and precipitation of the polypeptide due to unspecific polypeptide interactions. This (kinetic) competition between folding and aggregate formation can pose a major problem and may result in low folding yields. Aggregate formation is favored at high initial concentrations of unfolded protein, thus the best way to prevent aggregation is to work at low protein concentrations (typically 0.05 – 1mg/mL). Aggregation can also be inhibited by adding special small molecule additives to the refolding buffer. Compounds such as L-arginine, GuHCl and urea (<1M), polyols (e.g. glycerol), Na2SO₄ and some detergents have been shown empirically to increase folding yields of certain proteins. However their mode of action is largely unknown, which is why their effect on the folding of a given target protein cannot be predicted and needs to be determined experimentally\(^{42}\).

Many secreted proteins contain disulfide bonds, which significantly contribute to
the protein’s stability. Folding of such proteins typically requires the addition of a redox system to the folding buffer, such as 8mM cysteine/1mM cystine or 5mM reduced glutathione/0.5mM oxidized glutathione. The redox system facilitates correct disulfide bond formation via thiol-disulfide shuffling. Because the thiolate anion is the reactive species in this disulfide exchange reaction it is advisable to work at slightly elevated pH values (pH 8.0). In the case of non-disulfide containing proteins any unspecific oxidation of cysteines should be prevented by adding reducing agents (such as DTT) to the refolding buffer.

Where the target protein is known to bind to small molecule ligands (such as enzyme substrates, cofactors, inhibitors, etc.) addition of these compound to the refolding buffer can prove useful in order to obtain high refolding yields. In many cases these ligands might even be essential, e.g. in the case of structurally relevant metal ions such as $^{2+} \text{Zn}^+$, $^{2+/3+} \text{Fe}$ etc. or certain cofactors.

After folding, the product protein has to be purified from contaminating compounds (e.g. folding additives, misfolded material, etc.) and in many cases needs to be concentrated for further applications. In the case of highly disulfide crosslinked proteins (such as crambin, lysozyme, RNase, phospholipase A2 etc.) this can be efficiently achieved by RP-HPLC purification and lyophilization. Native and fully active protein material can then be recovered by rehydration of the lyophilized powder in a suitable aqueous buffer (lacking reducing agents). Non disulfide linked proteins, however, are almost guaranteed to denature under standard RP-HPLC conditions, which requires more gentle means for purification and concentration (e.g. gel-filtration, ion-exchange, affinity purification and/or dialysis and ultrafiltration).
**Figure B.1.** Ligation of unprotected peptide segments by means of native chemical ligation. Native chemical ligation is a highly chemo- and regioselective reaction, which enables the precise linking of two polypeptides in the presence of all unprotected functional groups typically found in proteins. Key functionalities are highlighted. Red: -“thioester; blue: N-terminal cysteine
Figure B.2. Chemical protein synthesis overview. Assembling proteins by native chemical ligation of unprotected peptide segments. The ligation site (-NHCH(R’)-CO-NH-CH(2SH)-CO-) is formed by reacting a peptide carrying a C-terminal thioester (red) with a peptide with a N-terminal cysteine (blue).
Figure B.3. Sequential versus convergent assembly of proteins. (a) Sequential ligation strategy. Proteins can be assembled from more than 2 segments in a sequential fashion, starting at the C-terminus and extending towards the N-terminus. Key to the synthesis is the temporary protection of the N-terminal cysteine in form of 1,3-thiazolidine in all segments equipped with two reactive groups (cysteine and "thioester; here peptide 2). This strategy can also be applied to the linking of more than 3 peptide segments. (b) Convergent ligation strategy. Proteins can be assembled in a convergent fashion, in which each starting polypeptide is roughly the same number of chemical transformations away from the final product. Convergent assembly in this manner is not limited to 4 segments: any or all of the peptide building blocks can be built from 2 other peptide segments. The above synthesis scheme was used to assemble the 46 amino acid residue Crambin polypeptide as a demonstration example (see “Anticipated results” section).
Figure B.4. Native chemical ligation of Cram[Thz^{26}-Gly^{31}]-“thioester and Cram[Cys^{32}-Asn^{46}]. (a) HPLC profile of the reaction shortly after mixing of both peptides Cram[Thz^{26}-Gly^{31}]-“thioester (1) and Cram[Cys^{32}-Asn^{46}] (2). (b) Reaction after 4h ligation. Product Cram[Thz^{26}-Asn^{46}] (3) is indicated. (c) Purified reaction product after Cram[Thz^{26}-Asn^{46}] (3) was converted to Cram[Cys^{26}-Asn^{46}] (4) (Inset: ESI-MS).
Figure B.5. Kinetically controlled ligation of Cram[Thr$^1$-Ala$^{15}$]$\alpha$thio(4-carboxymethyl)phenyl ester and Cram[Cys$^{16}$-Leu$^{25}$]$\alpha$thioalkylester. (a) HPLC profile of the reaction shortly after mixing of the peptides Cram[Thr$^1$-Ala$^{15}$]$\alpha$thio(4-carboxymethyl)phenyl ester (1) and Cram[Cys$^{16}$-Leu$^{25}$]$\alpha$thioalkylester (2). (b) Ligation reaction after 2h. (c) In order to reverse the formation of undesired thioester species (namely Cram[Thr$^1$-Leu$^{25}$]$\alpha$thiolactone (3) and the branched thioester species Cram[Thr$^1$-Cys$^{16}$([Thr$^1$-Ala$^{15}$])-Leu$^{25}$]$\alpha$thioalkylester (4)) the reaction was treated with 0.2M MESNA for 30min. (d) Purified Cram[Thr$^1$-Leu$^{25}$]$\alpha$MESNA thioester (6) (Inset: ESI-MS).
**Figure B.6.** Final assembly of Crambin. (a, b) HPLC chromatogram of the final ligation between peptides Cram[Thr$^{1}$-Leu$^{25}$]$_{a}$MESNA thioester (1) and Cram[Cys$^{26}$-Asn$^{46}$]$_{b}$ (2) shortly after mixing of both peptides (a) and after 18h (b). Acm groups were removed by AgOAc treatment for 1.5h and the reaction product was purified by HPLC (c). The full-length reduced Crambin polypeptide (4) was folded and purified (d). ESI-MS analysis indicated the loss of 6 protons upon folding which is in agreement with the formation of 3 disulfides.
Appendix C. Divergent synthesis of pentapeptides
Laboratory Schedule for Sping 2006 – CHEM 22200

**DIVERGENT SYNTHESIS OF PENTAPEPTIDES**

<table>
<thead>
<tr>
<th>Week of</th>
<th>Experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>March 27</td>
<td>First week of class – no labs</td>
</tr>
<tr>
<td>April 3</td>
<td><strong>Part A.</strong> Synthesis of N-Hydroxymethyl phthalimide</td>
</tr>
<tr>
<td></td>
<td><img src="" alt="Chemical Structure" /></td>
</tr>
<tr>
<td>April 10</td>
<td><strong>Part B.</strong> Preparation of Phthalalimidomethyl-resin</td>
</tr>
<tr>
<td></td>
<td><img src="" alt="Chemical Structure" /></td>
</tr>
<tr>
<td>April 17 + April 24</td>
<td><strong>Part C.</strong> Preparation of Aminomethyl resin</td>
</tr>
<tr>
<td></td>
<td><img src="" alt="Chemical Structure" /></td>
</tr>
<tr>
<td>May 1</td>
<td><strong>Part D.</strong> Attachment of Fmoc linker onto aminomethyl resin</td>
</tr>
<tr>
<td></td>
<td><img src="" alt="Chemical Structure" /></td>
</tr>
<tr>
<td>May 8 + 15 + 22</td>
<td><strong>Part E.</strong> Divergent synthesis of Pentapeptides</td>
</tr>
<tr>
<td></td>
<td><img src="" alt="Chemical Structure" /></td>
</tr>
</tbody>
</table>

- Divide mixture in half
- Continue with deblocking, washing, coupling and washing as appropriate
- H-Tyr-Gly-Gly-Phe-Leu-NH₂
- H-Tyr-Ala-Ala-Phe-Leu-NH₂
Experimental Procedure

The goal of this project is to isolate and characterize a specific enzyme that catalyzes the synthesis of a particular peptide. This enzyme is isolated from a microorganism that is known to produce the enzyme. The enzyme is then purified and characterized using a variety of techniques, including SDS-PAGE and Western blotting.

The enzyme is then used to synthesize a specific peptide using solid-phase peptide synthesis. The peptide is then analyzed using mass spectrometry to verify its identity.

Following the isolation and characterization of the enzyme, the peptide is synthesized and purified. The peptide is then analyzed using mass spectrometry to verify its identity.

A summary of the results is then prepared, including the enzyme's purification and characterization, as well as the peptide synthesis and characterization. The summary is then submitted for publication in a scientific journal.

References

Experiment 1: Synthesis of 3-Methoxyacetoxyphenol

**Introduction**

The synthesis of 3-methoxyacetoxyphenol is a kinetic step used to prepare compounds that will be used to test a patient for tuberculosis, which involves the reaction of 3-methoxyacetoxyphenol with a reagent to form a product. This product is then isolated from the reaction mixture to obtain the desired compound. Additionally, it is changed to another product because it is the starting material for the next experiment.

**Understanding the Experiment**

In this experiment, you will synthesize 3-methoxyacetoxyphenol by the following reaction:

\[
\text{PhOH} + CH_3COOH \rightarrow \text{PhOCH}_3 + \text{CO}_2 + \text{H}_2\text{O}
\]

**Procedure**

The original procedure for this reaction in a synthesis lab. Please read the procedure below and in your organic lab notebook. The amount of solvents and chemicals used in the reaction will be described in your lab notebook. The procedure includes the amount of solvents and chemicals used in the reaction, but you should prepare a detailed reaction scheme for your notebook. To estimate the amount of each solvent, it is not possible to weigh a precise amount. However, the procedure should outline the amount of each solvent required for the reaction. Then, follow the procedure as outlined.

**2-Methoxyacetophenone**

The compound was prepared analytically according to the procedure. The compound was then purified by recrystallization. The solvent used was methanol (1:1) and it was used to dissolve the compound. The reaction mixture was then filtered through a column of silica gel and the solvent was removed. The pure compound was then dried in vacuo and stored in a desiccator.

**Analysis of 3-Methoxyacetoxyphenol**

Determine the purity of the synthesized product by recrystallization. This will ensure that the reaction was successful and the product was isolated with high purity.

**Notes**

- Follow the procedure closely and record all observations.
- Keep all equipment clean and tidy for the next experiment.

Your Lab Report for Experiment Part A
In the **Introduction** section of your report:

- State the reaction that is performed on the experiment and its purpose prior to the procedure.

In the **Results** section of your report:

1. Determine the percent yield of the product after drying and/or crystallization.
2. Report the mass percent of the product before and after your calculations.
3. Identify all errors made in the % of yield and % of purity of the product.

In the **Discussion** section of your report:

1. Identify the errors of the reaction using your percent yield and melting point data. Compare the % error and explain how the data can be used to determine the required reaction for the errors.
2. Write the approximate ratio of all hydrogen in (phenylthiol), bromoethane, and % by approximate phenol. Please show the major reaction (where the phenol in solution gives the product) that we have written.
3. Why is water added to the mixture? Calculate the solubility of % molar (moles) phenol and phenol.
**Part B Preparation of Phthalimidomethyl-Resin**


**Introduction**

For this experiment you will use last week’s reaction product to modify a chemically inert crosslinked styrene-divinylbenzene resin with a strong acid. Modifying the resin with a chemical handle allows us to synthesize a peptide rapidly and efficiently as discussed in the global outline. It is important to be careful and minimize the handling loss of your resin because it will be used in all subsequent labs.

**Understanding the Experiment**

In this experiment you will prepare phthalimidomethyl resin by the following reaction:

![Chemical Reaction Diagram]

You will be reacting N-hydroxymethyl phthalimide, Part A’s reaction product, with a resin that is 1% divinylbenzene crosslinked polystyrene noted as Bio-beads. This reaction is catalyzed by methanesulfonic acid (MSA) in dichloromethane at room temperature.

**Procedure for a target loading of approximately 1 mmol/g:**

In a 100 mL round bottom flask equipped with a drying tube suspend 500 mg of Bio-Rad beads styrene-divinylbenzene copolymer, 1% crosslinkage, 200-400 mesh [In the Reagent table section of your notebook, copy and answer the questions: The resin is reported to be 200-400 mesh, what does this mean? If one resin bead has a diameter of 50 um with 10% of the resin is substituted, and a density of approximately 1 mg/mL, what is the concentration of the phthalimido group?] and 100 mg of N-hydroxymethyl phthalimide in 10 mL CH$_2$Cl$_2$. [In the Procedure section of your notebook, copy and answer the question: Why is a drying tube used?]

Add 2 mL of MSA to the reaction. [In the Procedure section of your notebook, copy and answer the question: What safety precautions should one take when handling MSA?] Gently stir this reaction for 3 hours at room temperature. Wash the resin with CH$_2$Cl$_2$ (4-5 aliquots totaling 50 mL), then EtOH (4-5 aliquots totaling 50 mL). Dry the resin under vacuum.

**Analysis of Phthalimidomethyl resin:**

Measure the IR of your product and starting resin by placing resin swollen in DCM in-between IR plates.

**Waste Disposal**

- Ethanol, DCM and MSA? should be disposed of in the “Organic Waste Bottle”.
- Aqueous waste can be washed down the sink with running water.
- All solid waste should be put into the “Solid Waste Jar”.


Your Lab/Report for Experiments Part II

In the Purpose section of your report, write out the overall scientific questions for the reaction explained in the experiment.

In the Results section of your report:
1. Sketch your 3D Mechanism for the reaction and identify all major intermediates. Do you expect the reaction products to have a high enough energy to be detectable? How do you prove/disprove it?
2. Based on the number of moles of X (hydroxide) theoretically you added, do you expect to get a yielding of X almost more than you expected?

In the Discussion section of your report:
1. What is a product? How is XCO3 analyzed? What is the affinity of XCO3?
2. Using ionic bond arguments, explain why substitution occurs at the pka position of the reacting trig.
3. Does the chemical structure of XCO3 influence substitution reactions? How is the driving force of nucleophilic substitution affected by conditions for this reaction?
4. For reaction mechanism, what are possible side reactions?
5. Why is it that the complex with XCO3? Please look at the difference in melting of the react of XCO3 and XCO3? Why does some melting depend on the solvent? Please explain why XCO3 is a better solvent than XCO3 for the dissolution of XCO3.?
Part C. Preparation of aminomethyl-copoly(styrene-divinylbenzene) Resin

Reference: Science, Vol 291, Iss. 5508, p.1523

Introduction
Over the next two lab periods you will prepare aminomethyl-copoly(styrene-divinylbenzene) resin and prepare ninhydrin reagents. You can prepare the ninhydrin reagents anytime during the two lab periods. Make sure that you store the ninhydrin reagents in a safe fashion because we will use them later in the quarter.

Understanding the Experiment
In this experiment you will prepare aminomethyl-copoly(styrene-divinylbenzene), or aminomethyl resin, by the following reaction:

![Chemical structure](image)

Procedure
Along with your reagent table answer the following questions:
Is hydrazine a catalyst or a stoichiometric reactant?
What is the chemical structure of trifluoroacetic acid (TFA) and is it a strong acid, weak acid, strong base, or weak base?
What is disopropyl ethyl amine (DIEA) and is it a strong acid, weak acid, strong base, or weak base?
What is the chemical structure of ninhydrin and what is it used for?

Preparation of aminomethyl resin:
In a 100 mL round bottom equipped with a reflux condenser Pthalamidomethyl-resin from the last lab is suspended in 10 mL of 5% hydrazine in EtOH [What safety precautions should one take when handling hydrazine?]. The mixture was heated to reflux for 3 hours and then cooled to room temperature. The flask is stoppered with a Teflon tape wrapped stopper to eliminate evaporation and placed in your drawer until next week.

The resulting gelatinous material is heated and filtered while hot. The resin is then washed with hot EtOH (3-4 aliquots totaling 50 mL), hot MeOH (3-4 aliquots totaling 25 mL), DMF (3-4 aliquots totaling 20 mL), CH₂Cl₂ (3-4 aliquots totaling 20 mL), 1:1 CH₂Cl₂/TFA (3-4 aliquots totaling 10 mL), CH₂Cl₂ (3-4 aliquots totaling 20 mL), 10% DIEA in DMF (3-4 aliquots totaling 20 mL), DMF (3-4 aliquots totaling 20 mL), CH₂Cl₂ (3-4 aliquots totaling 20 mL), and then dried under vacuum. [In your procedure, write out and answer the following question: What is the voluminous precipitate that was formed in your filtration flask during the EtOH and MeOH washes?] The IR of the product is measured.

Ninhydrin Reagents Preparation (make these with a partner to share):
1. Ninhydrin Reagent—Use 500 mg of high purity ninhydrin dissolved in a total volume of 10 mL of absolute ethanol. This solution must be made up at least a day before it is to be used to allow for dissolution of the ninhydrin. Store in a foil-covered volumetric flask or vial.
# Waste Disposal

- All waste, except for glass, should be disposed of in the "Organic Waste Basket".
- Organic waste should be kept separate from metal and paper waste.
- Solid waste should be put into the "Paper/Bio-waste".

## Your Lab Report for Experiment Part C:

1. In the Purpose section of your report, write out the specific group objectives, reagents, and instruments, and also include the products or the other products that are needed.

2. In the Procedure section of your report, include the specific questions of your protocol and take all major points. Describe the differences between the opening protocol questions and explain how you have determined that the result is consistent with your hypothesis.

3. In the Discussion section of your report:
   - What are the conclusions of the reaction that went to completion?
   - What is the link between your minds and the chemical reaction?
Part II Preparation of Vaccine Fluid Cotton Buds

Introduction
In this experiment you will mount the virus-disk inside lumen to the inner tray cap. You will examine the efficiency of the masks by the inactivated test. After removing the virus from the outer filter, you will determine the number of virus.

Understanding the Experiment

Procedure
At the beginning of the Fmic-based Assay (Fig. 1)
All components of vaccine are added to the outer tray and added in triplicate (see Fig. 1) and washed with 5 ml of DMF. The vials are shaken 3 times for 15 minutes and then centrifuged. One vial of each of the vaccines is added to the outer tray and the content is allowed to sit for 25 minutes without sufficient shaking. To prepare 1 vial of vaccine, extract the virus-disk solution as a separate cell of the other 3 vials of vaccine. In each 3 vials of vaccine, the vaccine is washed and 0.5 ml of 0.1 M phosphate buffer is added. To the vaccine, 0.5 ml of NADPH is added and 0.1 ml of the vitamin D3 (1000 units/ml) is added directly. The vaccine is washed with 0.5 ml of phosphate buffer (PBS), and the vaccine is added to the pipette (1000 units/ml) to obtain the assay of the vaccine. After 6 hours, the vaccine is washed 3 times with 0.5 ml of DMF, and the final washing, then the DMF until the oil-cup level is the top of the tube (i.e., to add the reagent to Fig. 1)
[NADH Test]
A true portion of the sample, about 1 mg of a mixture (e.g., placenta, liver, or lung tissue), is placed in a sample tube, 0.5 drops of phenol reagent, 0.5 drops of potassium ferricyanide, and 2 drops of sodium carbonate are added. (Remember that you already made these reagents.) Incubate at a 35°C water bath for 5 minutes and then dilute with 1 ml of 5% phenol. If the reaction mixture is deep blue, note the attachment of the NADH (NADH) enzyme. Otherwise, continue on to the next portion of the lab.

[Starch Loading Permeability]
Using the 0.5 ml sample set in the permeability test, place the sample in a 0.5 ml tube (3x3). Pipet a constant volume from the grinded sample into the sample set. Fill the sample tube with 5 ml of 2% (w/v) potassium ferricyanide and 1 ml of DI-water. Mix each tube and place on the water bath for 15 minutes. Record the absorbance at 595 nm and determine the starch loading using the starch loading table. (Refer to the lab manual for details.)

[Wheat Germ]
- 0.1 g of wheat germ should be obtained either by grinding one-half a wheat ear.
- Place the sample in the sample tube and add 1 ml of 0.5 M NaOH to the sample.
- Mix the sample and place it in the water bath for 10 minutes.

[Your Lab Report for Experiment Part II]
For the Discussion/Conclusion part report, write the conclusions or deductions for the experiment. Do not forget to title your conclusion. Explain why, or what you did to obtain the results. For the next experiment, describe the next step you will take. (Refer to the lab manual for guidance.)

For the Results Section of your report:
Calculate the binding of wheat germ to investigate.

For the Discussion/Conclusion of your report:
1. Why was the wheat germ used in this experiment? Is the reaction one that is eukaryotic in nature? Why, yes or why not? Describe the reaction.
2. Why did we add the sodium hydroxide to the reaction mixture? To increase the pH level of the wheat germ?
3. What is the purpose of DI-water as the control?
4. What does this mean for the results of the experiment? What effect does the sodium hydroxide have on the reaction?
Chapter 1: Enzyme Synthesis of Proteoglycans

Experiment:

Understanding the Experiment:

This experiment involves the purification of enzymes. The purpose is to isolate and purify the enzymes that catalyze specific reactions. In this case, the enzymes are involved in the synthesis of proteoglycans.

Procedures:

1. **Preparation of Cell Lysate**:
   - Collect cell lysates from appropriate cell lines.
   - Centrifuge the lysates to remove cell debris.
   - Wash the resulting supernatant to remove any remaining cell debris.

2. **Enzyme Assay**:
   - Mix the supernatant with appropriate substrates.
   - Incubate the mixture at optimal conditions.
   - Monitor the reaction over time to determine enzyme activity.

3. **Purification**:
   - Use affinity chromatography to isolate the enzymes.
   - Elute the enzymes using an appropriate buffer.

4. **Characterization**:
   - Analyze the purified enzymes using SDS-PAGE and Western blotting.
   - Determine the molecular weight and molecular mass of the enzymes.

5. **Activity Assay**:
   - Use a colorimetric assay to quantify enzyme activity.
   - Compare the activity of the purified enzymes with the activity of the crude lysate.

6. **Storage**:
   - Store the purified enzymes at -80°C for long-term preservation.

References:

231
In the Discussion section of your report:

1. In your own words, define "drug-receptor antagonists." How is it different from a receptor or bond replica? How can it be used to aid in carcinogenic research?

2. What are the limitations of understanding, to write on a more is more of accepting, phase of WOP by applying?
Appendix D. Direct on-resin synthesis of peptide-(alpha)thiophenylesters for use in native chemical ligation

Direct On-Resin Synthesis of Peptide-Thiophenylersters for Use in Native Chemical Ligation

Ziheng Tang, Janet E. Reeves, and Stephen C. Benkovic*

Institute for Biological Dynamics, Department of Chemistry, and Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, Illinois 60637.

Received November 21, 2001.

Abstract

The use of thiophene precursors in solid-phase peptide synthesis has been shown to have several advantages over traditional strategies. Here, we describe a novel method for the direct on-resin synthesis of thiophene-derivatized peptides. The reaction proceeds through the formation of a thiophene-ester bond, which is then hydrolyzed to yield the desired peptide. This method provides a convenient and efficient approach to the synthesis of thiophene-derivative peptides.
Section 6. A Novel Chemical Ligand Targeted To the Cytoskeleton During Thrombosis: Using the Cytoskeleton as a Scaffold: The Structure of Cytoskeleton and Its Interaction with Biologically Active Molecules. Enzymes.
Direct One-Pot Synthesis of Peptoids/Wingdings De novo Bar Use in Nucleo-chemical Ligation

Heleen Hong, Stella L. Fratiello, Zachary J. Smith, and Stephen W. Kent
Supporting Information

Preparation and Characterisation of 3-ethylmorpholinophenylacetic acid

Preparation: The 3-ethylmorpholinophenylacetic acid was prepared by treating 18 mmol (1.07 g) of 4-morpholinophenylacetic acid with 10.8 mmol (1 g) of ethyl chloride in 30 ml of dichloromethane for three hours. Reaction was quenched with 50 ml of water, and white precipitate was formed in the lower organic layer. The water layer was separated and discarded. The dichloromethane layer was evaporated and dried. The dried emulsion was precipitated and the residual water was removed by lyophilisation. The resulting crude 3-ethylmorpholinophenylacetic acid (14.5 g) was used for peptide-thiophenylacetic synthesis without further purification.

NMR (300 MHz, Bruker) MeOH-CD3OD. 400 MHz: 3.7-3.8 (2H, m, 6H), 7.25-7.3 (2H, m, 4H), 4.8 (2H, s), 4.85 (2H, s).

HPLC-MS: LC-MS analysis was recorded on a high-resolution QTOF coupled to an online MS quadrupole time-of-flight mass spectrometer.
Total mass (presumably created during mass spec analysis) varied for C_{10}H_{20}O.

136.1259a, found 136.1259a

(Phosphorophosphate and <15%). varied for C_{10}H_{20}O(85.1239a, found 85.1239a)
Procedures for the synthesis of peptide "Phosphophoryn".

Preparation of methylmesophosphophoryn glycine ester: A mixture of N,N-di-Gly-OCH₂-Phosphophoryn or S,Fos-Gly-MeHCl ester was added and the free group was removed using 70% tetrabutylammonium hydroxide in methanol. The mase was added to the mixture of 1.0 M of DMEA and 1.0 M of DMEA, and coupled to Py to yield Gly-DMA-Dureno acid or Gly-MeHCl ester. For further peptide chain elongation, the ester group was removed with two 1 minute treatments with 30% TFA/ethylamine and 30% TFA in DMSO. The resulting mesophosphophoryn glycine ester was used for polypeptide chain assembly by the use of "modified in vitro transcription" (MIVT) protocols.

Modified In Vitro Transcription Protocols:

Coupling of first amino acid (C-terminal amino acid): 1.0 mmol of first amino acid and 2.2 mmol of HATU were dissolved in 10 mL of H₂O. After adding 0.5 mL of DME, the coupling cocktail was shaken for one minute. The cocktail was added to a reaction vessel containing the mesophosphophoryn glycine ester. The reaction mixture was stirred for 30 minutes and the coupling solution was removed by filtration. Vividly, the coupling cocktail was added to the vessel and coupled for another 30 minutes.

Coupling of second amino acid: The free group was removed using acid 30% for the second amino acid coupling. We used 0.05 to work with. 2.2 mmol of the second free amino acid and 2.2 mmol of HATU were dissolved in 10 mL of DMF. After adding 0.5 mL of DME, the coupling cocktail was shaken for one minute. The cocktail was added to a reaction vessel containing the first amino acid attached moiety. The vessel was stirred for 30 minutes and the coupling solution was removed by filtration.
Coupling of third amino acid: After the second amino acid coupling, the free group was removed using neat TFA. After addition of the activated third amino acid the tube was sealed very quickly by blowing DME as follows: a 25ml reaction vessel containing tube was then sealed with the tube sealed with 3rd half of DME solution in a total of less than 10 seconds, with the added DME being continuously distilled under vacuum. Incubation was as follows: 2.2ml of DME Five minutes and 12.2ml of TFA was added in total of DME. After adding 0.5ml of DME, the coupling solution was shaken for one minute. The coupling solution was added to reaction vessel containing the second amino acid attached to the resin. We stirred the resin for 32 minutes and removed the coupling solution.

Coupling of remaining amino acids: After the removal of the free group by treatment with neat TFA, amino acids were coupled by the standard "on-the-resin" coupling with 12 minutes coupling using 2.2ml of amino acid and 2.2ml of TFA in DME.

UV changes: After chain assembly was complete, peptides were deprotected and simultaneously cleaved from the resin support by treatment with the resin containing 0.05% p-cresol (50:50, v/v) for 3 hours at 50°C. After evaporation of the DME under reduced pressure, crude products were precipitated and triturated with chilled diethyl ether, and the peptide products were dissolved in 80% aqueous acetonitrile containing 0.05% TFA, and purified by preparative reverse phase HPLC. Peptide compositions were determined by reverse phase HPLC analysis.
The chromatoic sepation were performed as describe in Figure 1 of manuscript.

Obvserved mass of the highest peak at T=7 min: 1293.5 Da. Calculated mass of the peptide molecular structure using average isotope: 1293.6 Da. Observed mass of the inveral layer peak at 5 min: correspond to an adduct from the addition of Dimethyl group (+ 92 Da) during HPLC chomatography.

HPLC profile:

[Graph Image]

MS profile corresponding to T=7 min in HPLC profile:

[Graph Image]
Evaluation of Evaluation:

Synthesis and analytical separation of peptide derivatives:

Two peptide derivatives Ala-Lys-Phc-Ala-Phe-Cys-Gly-Pro-Ala-NH2 were synthesized with either D-Phe or L-Phe at position 4. These peptides were made on an Fmoc polymer pre-loaded Boc-NH2 (0.1). Then resin on a 0.1 mmol scale after coupling cysteine, the resin was split into two parts, and either D-Phe or L-Phe was coupled. The two syntheses were carried out separately. Peptides were deprotected and cleaved from the resin by treatment for 4 h with trifluoroacetic acid (TFA). After precipitation of the TFA, the crude peptides were precipitated with an acid methanol mixture and triturated. Then dehydrated in 80% water-acetone containing 0.1% TFA. Analytical separations were performed on a Jasco Amino + syn Link Biotech 4000 using a column gradient (50-80%) of buffer B (aq. buffer A over 60 min buffer A to 0.1% trifluoroacetic acid (TFA) in water; buffer B for 60% TFA in acetonitrile) with a flow rate of 1.5 ml/min. Each derivative was chromatographed separately, and a 123 nm FAB mass spectrometry (Figure 4) in the main text. LC/MS confirmed the masses of the synthesized peptides. Ala-Lys-Phc-Ala-L-
Phe-Cys-Gly-Pro-Ala-NH2, observed mass: 580.4 (C22H32N4O3S); Ala-Lys-Phc-Ala-L-Phe-Cys-Gly-Pro-Ala-NH2, observed mass: 582.4 (C22H32N4O3S); calculated mass for both was 580.5 (C22H32N4O3S).

Synthesis of Ala-Lys-Phc-Ala-L-Phe-Cys-Gly-Pro-Ala-NH2:
The Cys-Gly-Pro-Ala-NH2 was made as described above. A peptide of crude was taken out after coupling the Cys residue; the free group was removed by treatment with TFA, and then without dried and subjected to 0.1% TFA cleavage as described above. LC-MS confirmed the mass of the Cys-Gly-Pro-Ala-NH2 observed: 437.9 (C22H32N4S), calculated: 437.9 (C22H32N4S). The crude Cys-Gly-Pro-Ala-NH2 was used without further purification.

The two fragments Ala-Lys-Phc-Ala-L-Phe-Cys-Gly-Pro-Ala-NH2 were synthesized with a residue, monosaccharides and a peptide as described above. Both L-Phe and D-Phe peptides were purified to greater than 80% on a 25 cm × 1 g. mmHg preparative C18 column using standard methods. LC-MS confirmed the masses of
the peptides: Ala-Leu-Ph-Leu-Ph-β-D-phenylphosphate 773.9 ± 2 Da, Ala-Leu-Ph-Leu-Ph-β-D-phenylphosphate 773.9 ± 2 Da, calculated mass for both peptides: 773.9 Da

Eggnog of Ala-Leu-Ph-Leu-Ph-β-D-phenylphosphate and Cys-Gly-Pro-Ala-Ser (1:1) was used for modifications. A Ligase was performed. Ala-Leu-Ph-Leu-Ph-β-D-phenylphosphate with Cys-Gly-Pro-Ala-Ser (1:20:1) and phosphatic acid (0.01% N, 0.24% H2O) was incubated at 37°C, 0.1 N, for 20 minutes, an aliquot of the reaction mixture was analyzed by reverse phase HPLC using the gradient described above to separate the intermediates. The same experiment was repeated using Ala-Leu-Ph-Leu-Ph-β-D-phenylphosphate. 3 N the AC/AB chromatogram of the reaction mixture on Figure 9 of the main text. In order to quantify the ratio of Ala-Leu-Ph-Leu-Ph-β-D-phenylphosphate and Cys-Gly-Pro-Ala-Ser by GC/MS, the peak area of the major intermediates was divided by the sum of peak areas for both the major and minor products.

Abbreviations:

ABE1: p-Methyl-β-D-phenylalanine
ABE1: Phe
ABE1: Nacyl
disopropylole
DABF: ADE
diarylformamide
PATE: O-(7-Azaindole-2-carboxamido)-phenylisopropyl acetate
PAE: O-(7-Azaindole-2-carboxamido)-phenylisopropyl acetate
Appendix E. Solving the X-ray structure of the scorpion toxin BmBKTx1 by racemic crystallography and direct methods

K. Mandal, B.L. Pentelute, V. Tereschko, A. Kossiakoff, Stephen B.H. Kent

[Draft manuscript, prepared by Brad Pentelute and Kalyaneswar Mandal. Synthesis by Mandal; crystallization by Mandal, with instruction and advice from Pentelute; Xray diffraction data collection, Pentelute with assistance from Mandal and Tereschko; structure solution Tereschko, Mandal, Pentelute.

Inspired by our results on the crystallization of snow flea antifreeze protein by the racemic method (Chapter 7), we set out to extend to other proteins this approach to the crystallization and determination of the Xray structures of proteins that do not crystallize easily. We identified a series of proteins of ever-increasing size, amenable to total chemical synthesis, which had been reported as difficult or impossible to crystallize, or for which no Xray structure had been reported. Difficulty in crystallization is often observed for proteins composed of many charged, surface exposed amino acid residues. One such example is the lysine-rich scorpion toxin BmBKTx1, a microprotein that is a high-conductance calcium-activated potassium channel blocker.[1-3] It was reported that for BmBKTx1 no crystal formation was observed at 100 mg/mL protein concentration over many weeks at room temperature using sparse-matrix crystallization screens.[3] Consequently, in order to obtain an Xray structure, a different crystallization approach had been taken that used reductive dimethylation of lysine residues to enhance the ability of this molecule to form useful crystals. The new strategy resulted in X-ray structure determination of methylated BmBKTx1.[3] Another approach to facilitate crystallization for this target would be to crystallize the protein from a racemic mixture.[4] Here we report the crystallization and Xray structure determination by direct methods of BmBKTx1 using racemic crystallography.

BmBKTx1 is a microprotein of 31 amino acid residues with six cysteine residues that form three disulfides in the folded protein molecule.[1] The mirror image
isomer (D-protein) of the native protein required for racemic protein crystallization can only be prepared by chemical synthesis. We prepared the polypeptide chains of D- and L-BmBKTx1 in good yield and high purity by manual stepwise Boc chemistry solid phase peptide synthesis.[5] The synthetic scales for the D- BmBKTx1 and L- BmBKTx1 polypeptides were 0.1 and 0.2 millimole, respectively. After the crude peptides were folded with concomitant disulfide formation, followed by HPLC purification, we obtained pure D- and L- proteins in multiple tens-of-milligram amounts. (see Experimental). The LCMS of the folded and purified synthetic protein products are shown in Figure E.1.
Figure E.1. The LC-MS profiles of the folded and purified synthetic BmBKTx1 enantiomers. a) D-BmBKTx1 (ob = 3336.0 ±0.7 Da, ca = 3335.98) b) L- BmBKTx1 (ob = 3336.0 ±0.7 Da, ca = 3335.98); the minor, earlier eluting peak corresponds to product containing an oxidized methionine (+16 Da). The chromatographic separations were performed using a linear gradient (1–61%) of buffer B in buffer A over 21 min (buffer A = 0.1% trifluoroacetic acid (TFA) in water; buffer B = 0.08% TFA in acetonitrile), with detection at 214nm and on-line ion trap electrospray MS.

In accord with the literature report, crystallization trials with native protein alone at 100 mg/mL protein concentration did not produce any crystals or precipitate after several months.[3] In contrast to this observation, striking results were obtained when crystallization were performed using a racemic solution containing equal amounts of D-
BmBKTx1 and L-BmBKTx1: we were gratified that using the commercially available Hampton index at 19 °C, crystals appeared from a solution of 25 mg/mL concentration (i.e 12.5 mg of each racemic protein). Under some of the conditions examined, crystals appeared overnight. One set of conditions was optimized to produce crystals suitable for X-ray diffraction as shown in Figure E.2(A). Diffraction data (Figure E.2(B)) was collected to resolution of 1.1 Å at the Advanced Photon Source and indexing revealed that the protein racemate crystallized in a unique tetragonal centrosymmetric space-group I4₁/a, with one enantiomer in the asymmetric unit. To the best of our knowledge, this is the first time a racemic protein mixture has been observed to crystallize in this space group. With high-resolution diffraction data in-hand, we attempted to solve the structure of the racemate by molecular replacement, using the X-ray structure model of methylated BmBKTx1.[3] However, the molecular replacement method failed to provide a solution.

We then explored the use of direct methods[6, 7] to obtain phasing information that could be used to solve the crystal structure of the BmBKTx1 protein racemate. The use of de novo methods to solve structures is a real challenge in modern protein crystallography.[8] There are only a handful of native L-protein structures that have been solved by direct methods.[9] Despite suggestions that direct methods should be more feasible with racemic crystals, it is noteworthy that neither of the centrosymmetric racemic protein structures previously studied has been solved using direct methods, although two different small peptide racemates were solved with some difficulty. [10, 11]

Surprisingly, our initial attempt at direct methods with SHELXS[12] within 8 h provided numerous solutions from a computational run on a standard dual core Xeon processor. The best structure solution had a figure of merit of 0.2286 which after preliminary refinement with SHELXL,[12] resulted in a well-defined contiguous electron density map in which side chain signatures were easily identifiable. The defined electron density allowed the facile production of a preliminary structural model of the BmBKTx1 molecule by manually building the L-polypeptide backbone into the electron density with the program TURBO-FRODO (Figure E.2(C).[13] From our
initial model we continued the standard refinement procedures using REFMAC5.[14] After placement of hydrogen atoms and solvent molecules, the final model was refined to a crystallographic R-factor of 0.193.

A cartoon representation of the resulting high resolution X-ray structure of the BmBKTx1 racemate is shown in Figure E.2(D). The crystal structure reported here is generally similar both to the solution NMR structure[2] and to the Xray structure of methylated[3] BmBKTx1. Slight backbone deviations were found in the β-strand region (Lys₂₁/Cys₂₂ and Lys₂₈/Cys₂₉), between the X-ray structure of BmBKTx1 reported here and the Xray structure of the methylated variant.
Figure E.2. X-ray structure of BmBKTx1, obtained from racemate crystals of D-BmBKTx1 and L-BmBKTx1. (A) Racemic crystals of D- BmBKTx1 and L-BmBKTx1. Protein crystals appeared for BmBKTx1 racemic mixtures at 19°C at a total protein concentration of 25 mg/mL. The crystallization conditions used were: 0.1 M citric acid pH = 3.5, 0.9 M ammonium sulfate. (B) Diffraction pattern. Data were
(Figure E.2 cont.) collected to 1.1 Å resolution. (C) 2Fo-Fc electron density map of BmBKTx1, contoured at 1 sigma, encompassing residues 30-31. (D) Molecular orientation of the D-BmBKTx1 (purple ribbon) and the L-BmBKTx1 (blue ribbon) molecules. (E) A cartoon model of the BmBKTx1 protein molecule solved at 1.1 Å resolution.

Because of the straightforward de novo structural solution, we wanted to investigate the effect of resolution on the likelihood of obtaining a solution using direct methods. We ran a series of SHELXS computational runs by truncating our data set to resolutions of 1.2 Å, 1.3 Å, 1.4 Å and 1.5 Å.[12] We then took the solutions from these calculations and attempted to refine the solutions. We found that for this molecule and data set, the effective cut-off for structure solution by direct methods is a resolution of 1.2 Å.

Conclusions. The work includes several ‘firsts’: racemic crystallization was used for the facile production of diffraction quality crystals for a protein target recalcitrant to crystallization; a protein racemate was observed to crystallize in the space group I4₁/a.; direct methods were used to solve the racemic protein structure; and, the Xray structure of unmodified BmBKTx1 has been solved at a resolution of 1.1 Å.
**Experimental:**

2-(1H-Benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) and protected amino acids (Peptide Institute, Osaka) were obtained from Peptides International. Side-chain protecting groups used were Arg(Tos), Asn(Xan), Asp(OcHex), Cys(4-CH₃Bzl), Lys(2Cl-Z), Ser(Bzl). Boc-L-Lys(2ClZ)-OCH₂-phenylacetic acid was purchased from NeoMPS, Strasbourg France. Boc-D-Lys(2ClZ)-OCH₂-phenylacetic acid was prepared following the literature procedure.[15] N,N-Diisopropylethylamine (DIEA) was obtained from Applied Biosystems. N,N-Dimethylformamide (DMF), dichloromethane, diethyl ether, HPLC-grade acetonitrile, and guanidine hydrochloride were purchased from Fisher. Trifluoroacetic acid (TFA) was obtained from Halocarbon Products. HF was purchased from Matheson. All other reagents were purchased from Sigma-Aldrich.

**Chemical synthesis of BmBKTx1.** The target amino acid sequence of BmBKTx1 is:


The D- and L- BmBKTx1 polypeptide chains were synthesized on Boc-Lys(2Clz)-OCH₂-Pam-resin of the appropriate chirality using manual in situ neutralization Boc chemistry protocols for stepwise SPPS,[5] on 0.1 and 0.2 mmol scale respectively. After removal of the N-terminal Boc group, the peptides were cleaved from the resin and simultaneously deprotected by treatment at 0 °C for 1 h with anhydrous HF containing 5% p-cresol/thiocresol (1:1) as scavengers. After removal of HF by evaporation under reduced pressure, the crude peptide was precipitated and washed with diethyl ether, then dissolved in 50% aqueous acetonitrile containing 0.1% TFA and lyophilized. The crude lyophilized linear D- and L-BmBKTx1 peptides were obtained in 249 mg and 500 mg respectively. 120 mg of the crude linear D- and 150 mg of the crude linear L-BmBKTx1 were directly subjected to folding with simultaneous formation of disulfides, by dissolving the peptide at 3 mg/mL in 6 M guanidine hydrochloride followed by rapid 6-fold dilution with 100 mM tris-hydroxymethyl aminomethane containing 9.2 mM L-
cysteine and 1.2 mM L-cystine hydrochloride at pH 8.0, to give a final concentration of 0.5 mg/mL peptide and 1M guanidine hydrochloride in the folding buffer. The folding was essentially complete within 1 h at room temperature, as evidenced from the decrease in mass of ~6 Daltons for the product formed, as shown by LC-MS analysis. The final product was purified by reverse phase HPLC on a Silicycle spherical C-18, 10 × 250 mm column at 40 °C using a gradient of 1.0%-41.0% acetonitrile over 80 minutes, at a flow rate of 10mL/min. Fractions containing the desired products were identified by LCMS, combined and lyophilized to furnish high purity D- BmBKTx1 and L-BmBKTx1 in 42.0 milligrams (26%, based on the amount (moles) of starting resin used) and 42.5 milligrams (21%, based on the amount (moles) of starting resin used) respectively.

**Crystallization.** Crystallization of racemic BmBKTx1 was performed by mixing equal amounts (by weight) of lyophilized D- BmBKTx1 and L-BmBKTx1 in water at the following concentrations: 25 mg/mL (12.5 mg of D- and 12.5 mg L-), 50 mg/mL (25 mg of D- and 25 mg L-), 100 mg/mL (50 mg of D- and 50 mg L-), and 150 mg/mL (75 mg of D- and 75 mg L-). Each solution was centrifuged to remove minor amounts of particulate matter and then used directly for crystallizations. Crystallization screening was conducted at ~19 °C using the commercially available Hampton index. Crystallization screens were done by the hanging drop vapor diffusion method. The drops were generated by mixing 1 μL of protein solution with 1 μL of reservoir solution, and placed over 1mL of reservoir solution. Crystals appeared after one day; within a week, ~10-20% of the conditions examined had produced microcrystals. One set of conditions (0.1 M citric acid pH = 3.5, 2 M ammonium sulfate) was further optimized by varying the concentration of precipitant solution to get crystals suitable for X-ray diffraction. X-ray diffraction data was collected from a crystal grown from 0.1 M citric acid pH = 3.5, 0.9 M ammonium sulfate.

**Data collection.** For low temperature data collection, selected crystals were briefly transferred to the cryoprotectant (reservoir solution plus 20% (v/v) glycerol) and flash-frozen in liquid nitrogen. The X-ray diffraction data were collected at 100K at the
Argonne National Laboratory (Advanced Photon Source, beamline 23ID equipped with a MARCCD 300 detector). Crystal diffractions images were integrated, scaled, and merged with HKL2000.[16]

**X-ray structure determination.** Molecular replacement was attempted by using the program MOLREP with CCP4i program suite.[17] The structure of racemic BmBKTx1 was solved by direct methods using SHELXS which revealed the positions for most of the atoms in the initial solution after preliminary refinement with SHELXL.[12] Electron density and model examinations were done using TURBO-FRODO. The restrained positional and anisotropic B-factor refinement was performed in REFMAC5.[14] The hydrogen atoms were included in the riding positions. Molecular graphics were generated using Pymol or Chimera. The data collection and refinement statistics are summarized in Table E.1.
### Table E.1. The X-ray data collection and refinement statistics for the crystal structure of BmBKTx1.

**Data collection statistics**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>$I4_1/a$</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td>$a = 56.63$</td>
</tr>
<tr>
<td></td>
<td>$b = 56.63$</td>
</tr>
<tr>
<td></td>
<td>$c = 31.76$</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 90.00$</td>
</tr>
<tr>
<td></td>
<td>$\beta = 90.00$</td>
</tr>
<tr>
<td></td>
<td>$\gamma = 90.00$</td>
</tr>
<tr>
<td>Mol/asymmetric unit</td>
<td>1</td>
</tr>
<tr>
<td>Wavelength</td>
<td>1.54178</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>20 - 1.1 (1.129 - 1.10)</td>
</tr>
<tr>
<td>$R_{	ext{sym}}$ or $R_{	ext{merge}}$</td>
<td>0.045 (0.417)</td>
</tr>
<tr>
<td>$I/\sigma(I)$</td>
<td>41.6 (2.35)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.78 (97.33)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>5.9</td>
</tr>
</tbody>
</table>

**Refinement statistics**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)</td>
<td>20 - 1.1 (1.129 - 1.10)</td>
</tr>
<tr>
<td>No. reflections</td>
<td>17925</td>
</tr>
<tr>
<td>$R_{	ext{work}}$/$R_{	ext{free}}$</td>
<td>0.193 (0.271)/ 0.209(0.268)</td>
</tr>
<tr>
<td>No. atoms</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>269</td>
</tr>
<tr>
<td>Water</td>
<td>20</td>
</tr>
<tr>
<td>Avarage $B$-factor (Å$^2$)</td>
<td>10.1</td>
</tr>
<tr>
<td>R.m.s deviations</td>
<td>Bond lengths (Å)</td>
</tr>
<tr>
<td></td>
<td>0.018</td>
</tr>
<tr>
<td>Bond angles (%)</td>
<td>2.134</td>
</tr>
</tbody>
</table>

**Ramachandran plot statistics**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Most favored (%)</td>
<td>92.3 (L)</td>
</tr>
<tr>
<td>Additionally allowed (%)</td>
<td>3.8 (L)</td>
</tr>
<tr>
<td>Generously allowed (%)</td>
<td>3.8 (L)</td>
</tr>
</tbody>
</table>

*Highest resolution shell is shown in parenthesis.

**Based on maximum likelihood.
Acknowledgement.

We are thankful to XXX for helpful discussions. Use of the GMCA-CAT beamline 23-ID at the Advanced Photon Source is supported by the companies of the Industrial Macromolecular Crystallography Association through a contract with the Center for Advanced Radiation Sources at the University of Chicago. Use of the Advanced Photon Source is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract no. W-31-109-Eng-38.
References:


Appendix F. Modular total chemical synthesis and racemic crystallization of Kaliotoxin

Kaliotoxin (KTX) is a 38-residue disulfide-rich microprotein isolated from spider venom that blocks the passage of ions through membrane spanning protein channels. Residues 26-35 form a beta-sheet and this region is thought to be responsible for channel blocking. However, the exact binding/blocking interaction of KTX with different potassium channels is still under investigation. The Xray structure of KTX is not known. It has been reported that this class of highly charged microprotein is recalcitrant to crystallization, thereby precluding the possibility of determining the crystal structure. Xray crystal structures and NMR structures provide excellent avenues to study docking interactions.

In order to obtain the Xray structure of KTX, we sought to explore the crystallization of KTX from a racemic solution. The racemic method was chosen because we have recently found that crystal formation is more favorable. During our crystallization screens with a snow flea antifreeze protein (Chapter 7), and with the scorion toxin BMBktx1 (Appendix E) we found that racemic protein mixtures produced crystals when L-protein alone did not.

A racemic mixture consists of equal amounts of the protein enantiomers D-KTX and L-KTX. A mirror image D-protein can only be made by chemical means. In order to determine the Xray structure by racemic crystallography, we thus needed to establish synthetic chemical access to both D-KTX and L-KTX. The modular total chemical synthesis strategy for the mirror image forms of KTX is shown in Figure F.1. The strategy reported here relies on the use of one-pot native chemical ligation from three polypeptide segments and subsequent folding. Such a modular synthetic strategy enables the facile preparation of protein analogues that can be used to probe the binding interactions with different ion channels.
Figure F.1 A) The 38-residue amino acid sequence of Kaliotoxin. B) The modular three segment one-pot ligation strategy for the total chemical synthesis of Kaliotoxin. The first native chemical ligation step is followed by the conversion of the N-terminal 1,3-thiazolidine-4-R-carboxylic acid (Thz) to Cys. After the second native chemical ligation, the full length polypeptide is folded with concomitant formation of the three native disulfide bonds.

The chemical syntheses of D-KTX and L-KTX proceeded smoothly. Each of the three peptide segment building blocks was prepared by Boc chemistry ‘in situ neutralization’ stepwise SPPS and purified by reverse phase HPLC.[5] The segments were ligated using standard conditions as published elsewhere.[6] A representative synthesis of L-KTX is shown in the Experimental section. The entire synthetic process, starting with the first ligation, was completed in less than two days. The LCMS and CD
traces of the final purified synthetic D-KTX and L-KTX molecules are shown in Figure F.2.

Figure F.2. Analytical traces for the characterization of synthetic D-KTX and L-KTX. A) The LCMS of purified folded L-KTX (ob = 4151.3 ±0.7 Da, ca = 4150.0 Da (average isotopes)). B) The LCMS of purified folded D-KTX (ob = 4151.8 ±0.7 Da, ca = 4150.0 Da). Analytical HPLC was carried out as follows: Varian Microsorb self-packed C4 2.1 X 50 mm column using a linear gradient of 5-65 % buffer B over 15 min at 40 °C with a flow rate of 0.5 mL/min (buffer A= 0.1% TFA in H2O; buffer B =
(Figure F.2 cont.) 0.08% TFA in acetonitrile). The eluent was monitored at 214 nm, with on-line electrospray mass spectrometry (MS). C) CD spectra were recorded using a Jasco 715 instrument at 23 °C by dissolving 0.02 mg of D-KTX or L-KTX in 200 microL of 50 mM phosphate buffer pH = 6.9. The path length of the cell was 1 mm.

With the mirror image forms of KTX in hand, we set-up the Hampton index sparse crystallization screen, crystal screen, and crystal screen II using the hanging drop method at ~23 °C. The racemic protein mixture used was by prepared by dissolving equal amounts of D-KTX and L-KTX in water and was screened both at 14 mg/mL (i.e. 7mg/ml of each enantiomer) and at 7 mg/mL. Little precipitation was formed during these crystallization screens, suggesting that an even higher concentration of protein should be used for subsequent trials. After a period of time, crystals appeared in two different sets of conditions. Pictures of crystals grown from 0.2 M postassium sodium tartrate tetrahydrate, 0.1 M sodium citrate tribasic dihydrate pH = 5.6, and 2.0 M ammonium sulfate are shown Figure F.3.
We collected preliminary diffraction data from a twinned crystal that grew from a buffer containing 1.6 M sodium citrate tribasic dihydrate pH = 6.5. The crystal diffracted to 1.6 Å and the data collection statistics are presented in Table F.1. We attempted molecular replacement using the reported NMR structure[7], but failed to obtain a structure solution.
We are currently optimizing the crystallization conditions in order to collect diffraction data from a single crystal. We aim to get higher resolution, so we can solve the structure by direct methods. As reported in Appendix E, we have determined the X-ray structure of a similar venom microprotein called BMBkttx1 by de novo methods.

**Experimental:**

**Peptide Segment Synthesis (peptide-^a^carboxylate or peptide-^a^thioester)** D-peptides and L-peptides were prepared manually by ‘in situ neutralization’ Boc chemistry[5] on a 0.4 mmol scale by stepwise solid phase peptide synthesis, on -OCH₂-Pam-resins (free ^a^carboxyl peptides) or on HSCH₂CH₂CO-Xaa-OCH₂-Pam-resin (^a^thioester peptides).[8] Side-chain protection for amino acids was as follows: Arg(Tos), Asn(Xan), Asp(OcHex), Cys(4-CH₃Bzl), His(Bom), Glu(OBzl), Lys(2Cl-Z), Ser(Bzl), Thr(Bzl), Trp(CHO) Tyr(Br-Z). After completion of the chain assembly, peptides were deprotected and cleaved from the resin support by treatment with anhydrous HF containing p-cresol (90:10, vol/vol) for 1 h at 0 °C. After complete evaporation of the HF under reduced pressure, crude peptide products were precipitated and triturated with
chilled diethyl ether, and the peptide products were dissolved in 50% aqueous acetonitrile containing 0.1% TFA.

The peptide building blocks (and corresponding masses) prepared and used in the synthesis were as follows: Gly\textsuperscript{1}-Val-Glu-Ile-Asn-Val-Lys-Cys-Ser-Gly-Ser-Pro-Gln\textsuperscript{13}-CO-S-CH\textsubscript{2}-CH\textsubscript{2}-CO-Ala-COOH (ob = 1476.1 ±0.5 Da, ca = 1476.78 Da (average isotopes)), Thz\textsuperscript{14}-Leu-Lys-Pro-Cys-Lys-Asp-Ala-Gly-Met-Arg-Phe-Gly-Lys\textsuperscript{27}-CO-S-CH\textsubscript{2}-CH\textsubscript{2}-CO-Ala-COOH (ob =1725.0 ±0.5 Da, ca = 1724.92 Da), and Cys\textsuperscript{28}-Met-Asn-Arg-Lys-Cys-His-Cys-Thr-Pro-Lys\textsuperscript{42}-COOH (ob = 1319.8 ±0.7 Da, ca = 1320.63 Da).

**Analytical HPLC.** Peptide compositions were confirmed by analytical reverse phase LC-MS using a gradient of acetonitrile versus {0.1% trifluoroacetic acid (TFA) in water}. For the work reported in this paper, analytical HPLC was carried out as follows: Vydac C4 2.1 X 150 mm column using a linear gradient of 1-61% (5-65%) buffer B over 15 min at 40 °C with a flow rate of 0.5 mL/min (buffer A= 0.1% TFA in H\textsubscript{2}O; buffer B = 0.08% TFA in acetonitrile) or Hypersil C18 4.6 X 50 mm column using a linear gradient of 1-61% buffer B over 15 min at 40 °C with a flow rate of 1.0 mL/min (buffer A= 0.1% TFA in H\textsubscript{2}O; buffer B = 0.08% TFA in acetonitrile). The eluent was monitored at 214 nm, and by on-line electrospray mass spectrometry (MS).

**Preparative HPLC:** Peptides were purified on C4, C8, or C18 silica with columns of dimension 22 X 250 mm, 10 X 250 mm, or 10 X 100 mm. The silica used was TP Vydac or self-packed Varian Microsorb or Agilent Zorbax. Crude peptides were loaded onto the prep column in ~10% acetonitrile/90% {0.1%TFA in water}, and eluted at a flow rate of 10 mL per minute with a shallow gradient (e.g. 20%B-40%B over 60 minutes) of increasing concentrations of solvent B (0.1% TFA in acetonitrile) in solvent A (0.1% TFA in water). Fractions containing the pure target peptide were identified by analytical LC and MALDI MS, and were combined and lyophilized.
Native Chemical Ligation: Ligation reactions were carried out on a scale of 10 micromoles at a peptide concentration of 13 mM (total volume 0.75 mL), under previously published conditions: 200 mM sodium phosphate buffer containing 6 M guanidine hydrochloride, 20 mM TCEP, pH = 6.8, 13 mM for each peptide, 10-100 mM mercaptophenylacetic acid (MPAA) as catalyst,[6] purged and sealed under argon. After the first ligation reaction, 0.2M methoxyamine.HCl was added to convert the Thz-peptide product to Cys-peptide and left overnigt at pH 4. The pH was adjusted to 6.8 and the next peptide-thioester (10 micromoles) was added in equimolar amount and the ligation reaction allowed to proceed at room temperature for 3 hours.

One-Pot Folding reaction: After confirming the completion of the ligation reaction, the mixture was diluted with 6 mL of pH = 7.8 buffer containing 6 M Gu-HCl and 0.1 TRIS-HCl. Then 35 mL of redox buffer was added containing 100 mM TRIS Base, 8 mM cysteine, 1 mM cystine-2HCl, at pH 7.8. The folding and disulfide formation was complete after overnight reaction, as indicated by mass change of minus 6 Da (ob. 4150.1 ± 0.7 Da, ca. 4150.0 Da (average isotopes)). The product was then purified by reverse phase HPLC and lyophilized and 3.3 micromole of material was isolated (33% yield).

Xray data collection: For low temperature data collection, selected crystals were briefly transferred to the cryoprotectant (reservoir solution plus 20% (v/v) glycerol) and flash-frozen in liquid nitrogen. The X-ray diffraction data were collected at 100K at the Argonne National Laboratory (Advanced Photon Source, beamline 23ID equipped with a MARCCD 300 detector).

X-ray structure determination: Crystal diffractions images were processed and scaled with HKL2000.[9] Initial attempt to solve the structure by molecular replacement was attempted by using the program MOLREP with CCP4i program suite.[10]

Representative synthesis of L-KTX: The reaction sequence was carried out on a 10 micromole scale and after purification of the full-length folded material approximately 3.3 micromole (33 %) of material was isolated after purification.
**Figure F.4.** Ligation of L-KTX [Thz14-Lys27]“-thioester and [Cys28-Lys38]-COOH. Analytical HPLC traces of the ligation of A) t = 0 and B) t = 5 hrs. The analytical conditions used for the monitoring of this ligation was Hypersil C18 4.6 X 50 mm column using a linear gradient of 1-61 % buffer B over 15 min at 40 °C with a flow rate of 1.0 mL/min (buffer A= 0.1% TFA in H2O; buffer B = 0.08% TFA in acetonitrile). In part (B) the product shown is [Thz14-Lys38]-COOH (ob = 2868.6 ± 1.0 Da, ca = 2867.5 Da (average isotopes)). The peak labeled (&) is the ligation catalyst MPAA.
Figure F.5. Ligation of L-KTX [Gly1-Gln13]"-thioester and [Cys14-Lys38]-COOH and folding of [Gly1-Lys38]-COOH in one-pot: Analytical HPLC traces of the starting peptides for the ligation reaction. A) Crude [Cys14-Lys38]-COOH from the previous ligation after the conversion of Thz to Cys by treatment of 0.2 M methoxyamine-HCl (ob = 2856.5 ± 1.0 Da, ca = 2855.5 Da (average isotopes)). B) The purified [Gly1-Gln13]"-thioester peptide used for this ligation (ob = 1476.1 ± 0.5 Da, ca = 1476.78 Da (average isotopes)). C) The crude ligation products after approximately 7 hrs reaction. The main peak is [Gly1-Lys38]-COOH (ob = 4156.1 ± 1.0 Da, ca = 4156.0 Da (av. isotopes)). D) Crude products obtained after overnight folding/disulfide formation of the crude reaction mixture without prior purification (ob = 4150.1 ± 0.7 Da, ca = 4150.0 (av isotopes)). The peak labeled (&) is the ligation catalyst MPAA. The analytical conditions used for the monitoring of the ligation and folding was Vydac C4 2.1 X 150 mm column using a linear gradient of 1-61 % buffer B over 15 min at 40 °C with a flow rate of 0.5 mL/min (buffer A= 0.1% TFA in H2O; buffer B = 0.08% TFA in
(Figure F.5. cont.) acetonitrile). Effluent was monitored at 214nm and by online electrospray MS. Note the analytical conditions for the monitoring of these reactions is different from that used for the first ligation.
References:


**Appendix G.** Comparative properties of insulin-like growth factor 1 (IGF-1) and [Gly7D-Ala]IGF-1 prepared by total chemical synthesis.

Comparative Properties of Insulin-like Growth Factor I (IGF-I) and [Gly7D-Ala10]IGF-I Prepared by Total Chemical Synthesis.

Yudkes B. S. (Received), Early S. (Promotion), and Michael W. H. (Letter to the Editor).

Receptors for growth factors...
amount of other minor isomers was also reduced as compared to the literature.\textsuperscript{35} Folding, as monitored by LCMS, was complete in 1 h. Following oxidation, the mass decreased by 0.6 ± 0.7 Da indicating the formation of three disulfide bonds. The principal components of the folding reaction were purified by preparative RP-HPLC, enabling pure IGF-1 and IGF-swap to be obtained (Figure 1A, B). The overall yield of IGF-1 was 6.7%.

Synthetic IGF-1 and IGF-swap were characterized by CD and \textsuperscript{1}H-NMR spectroscopy. Previous studies have established that native IGF-1 and IGF-swap exhibit different mean \textalpha-helix contents due to the segmental unfolding of helix 2 (residues 41-48 in IGF-1, canonical residues 41-48 in the non-native isomer)\textsuperscript{15} (Figure 5S). Structural differences between IGF-1 and IGF-swap are also associated with significant differences in \textsuperscript{1}H-NMR chemical shifts. \textsuperscript{1}H-NMR spectra of synthetic IGF-1 (or IGF-swap) obtained herein (Figure 2A) correspond to control spectra of biosynthetic IGF-1 (or IGF-swap) kindly provided by Eli Lilly and Co (Figure 56). These NMR data show that our synthetic IGF-1 and IGF-swap each had a defined tertiary structure, and also confirm the previously determined disulfide pairing schemes for the two protein isomers.\textsuperscript{35} The synthetic proteins were further characterized by measurement of relative binding affinities to the Type 1 IGF receptor (Table 1). The activity of synthetic IGF-1 is the same as that of biosynthetic IGF-1, and for both recombinant and synthetic proteins the activity of the disulfide `swap' isomer is reduced by about 30-fold.\textsuperscript{57} The natural chemical ligation route thus provides an efficient method for the synthesis and purification of a protein whose structure and activity are indistinguishable from biosynthetic IGF-1.

To illustrate the utility of this synthetic protocol in studies of non-standard analogs, we investigated the substitution of Gly\textsuperscript{7} by D-Ala. This analog was designed based on corresponding studies of human insulin. Gly\textsuperscript{7} (canonical position B\textbeta in insulin) participates in a Type 1 \beta-turn and exhibits absolute conservation among members of the vertebrate insulin-related superfamily. Although not part of the classical receptor-binding surface, the homologous glycine in GLP\textsuperscript{28} participates in a large-scale allosteric reorganization of zinc insulin hexamers,\textsuperscript{35} designated the TR transition.\textsuperscript{40} In this transition the dihedral angles of GLP\textsuperscript{28} move from the right side of the Ramachandran plot (with positive \phi-angle ordinarily forbidden to L-amino acids) to the left side (with negative \phi-angle characteristic of L-amino acids). In insulin, substitution of GLP\textsuperscript{28} by D-Ala markedly augments the thermodynamic stability of insulin (presumably by stabilizing the B\textbeta-related \beta-turn) but impedes binding to the insulin receptor by 1000-fold.\textsuperscript{57} Gly\textsuperscript{7}D-AlaIGF-1 was obtained using the same synthetic procedure as for native IGF-1 (Figure 1C). The overall synthetic yield of Gly\textsuperscript{7}D-AlaIGF-1 was 6.7%. The observed folding rate of Gly\textsuperscript{7}D-AlaIGF-1 was much slower than native IGF-1. It is notable that the amount of Gly\textsuperscript{7}D-AlaIGF-swap was greater in the folding of Gly\textsuperscript{7}D-AlaIGF-1 as compared to native IGF-1 (Figure 57). \textsuperscript{1}H-NMR spectra of synthetic Gly\textsuperscript{7}D-AlaIGF-1 and Gly\textsuperscript{7}D-AlaIGF-swap are shown in Figure 2(B).

Characterization of the present D-Ala\textsuperscript{7} analog of IGF-1 indicates augmented stability (AAk\textsuperscript{7}izzling 0.6 ± 0.1 kJ/mole relative to native IGF-1) as observed by CD-monitored guanidine denaturation experiments. Substitution of Gly\textsuperscript{7} by D-Ala augments the stability of the non-native isomer more substantially (AAk\textsuperscript{7}izzling > 1.2 kJ/mole) than it does the stability of the native isomer, consistent with the greater relative yield of the non-native isomer on oxidative folding of the D-Ala\textsuperscript{7} polypeptide. Although the structural basis for this selective stabilization is unclear, these findings demonstrate that single-amino-acid substitutions in IGF-1 can modulate the fidelity of disulfide pairing by favoring a preferred pairing scheme (even if non-native) rather than by destabilizing competing folds. Although CD and \textsuperscript{1}H-NMR spectra of the D-Ala\textsuperscript{7} analog are similar to those of native IGF-1 (Figure 5S and Figure 2), its binding to the IGF receptor is reduced by 300-fold (Table 1). The similar properties of homologous D-Ala-substituted analogs of insulin and IGF-1 (i.e., higher thermodynamic stability and lower receptor-binding affinity as compared to the native form) provide evidence that the N-terminal segments of these proteins play corresponding roles in their respective receptor-binding mechanisms. It is possible that the IGF-1 monomer, on binding to its receptor, undergoes a similar change in conformation near Gly\textsuperscript{7} as proposed for the insulin monomer. It would be of future interest to characterize sites elsewhere in the insulin and IGF-1 molecules at which corresponding substitutions yield divergent properties, as these sites would identify inequivalent receptor-binding mechanisms. Together, such comparative studies promise to provide a map of the functional similarities and differences between insulin and IGF-1. Such data could inform design of novel ligands with novel biological properties.

In conclusion, we have described herein an efficient synthetic scheme for the production of IGF-1 and the protein diastereomer analog Gly\textsuperscript{7}D-AlaIGF-1. Based on native chemical ligation and an optimized folding protocol, this scheme facilitates the introduction of non-standard amino acids whose biosynthetic incorporation is difficult or not feasible. In addition to the preparation of D-amino acid-containing IGF-1 analogs, this synthetic scheme will enable chemical modification of the polypeptide backbone and the introduction of non-standard side chains, including photocleavable derivatives useful in mapping the interface between IGF-1 and its receptor. Facile synthesis of chemical analogs of IGF-1 thus promises to enable a new generation of structure-activity studies of potential interest in the treatment of cancer and diabetes mellitus.

Received (will be filed in by the editorial staff).
Published online (will be filed in by the editorial staff).

**Keywords:** chemical protein synthesis - insulin - growth factor - native chemical ligation - protein folding

Table 1. Binding affinities of IGF-1, IGF-swap, and [Gly\textsuperscript{7}D-Ala]IGF-1 to the Type 1 IGF receptor.

<table>
<thead>
<tr>
<th>Dissociation Constant (nM)</th>
<th>Relative Affinity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGF-1</td>
<td>0.046 ± 0.0006</td>
</tr>
<tr>
<td>IGF-swap</td>
<td>6.0 ± 0.8</td>
</tr>
<tr>
<td>[Gly\textsuperscript{7}D-Ala]IGF-1</td>
<td>16.7 ± 0.3</td>
</tr>
</tbody>
</table>
References


Figure 5. A series of runs are shown with different mass spectrometry spectra. The mass spectrometry spectra show the presence of specific compounds at different masses. The spectra are labeled with peak numbers and mass values. The spectra are compared to theoretical spectra for identification.
Figure: 405. Cationic extraction of 115-263/486 fibrous staple. The CHTC (cotton/mixed) and 115-263 fibres were subjected to identical fixation procedures. The CHTC fibres were then treated with a NaOH solution of 6N, 15% of HSOH in 80% of water. A 1:15:37-1:50:13 ratio of 1:15:37 to 1:50:13 NaOH solution was used. The CHTC fibres were then treated with a NaOH solution.
Figure 68. Column A is processed to 90% yield of 1,2,3-trihydroxybenzene. Column B is treated with 1 M NaOH at 100°C for 48 h. The columns are eluted with 1 M HCl at 50°C.

The chromatographic separation was performed using a thin-layer chromatography system. The column A and B samples were analyzed using two-dimensional thin-layer chromatography. Column A = 1 M NaOH-acetone, Column B = 1 M HCl. Elution was monitored using ultraviolet light at 254 nm.
Figure 4B: CE groups of amino acids. A: Control (---), 100 µM 755 (----), and 100 µM 755 + 100 µM 545 (-----) show an upward trend and 100 µM 755 + 100 µM 545 (-----) show a downward trend. B: CE groups of amino acids at the toxic concentration and extractable from the animal fragments. Tryptic ribonucleosides concentrations of 5% of the total ribonucleosides at pH 7.4 and pH 7.8.
Figure 10. (A) 20% of the molecules in the sample were of 50% of the 100% map.
(B) 20% of the 50% map were of the 100% map.
(C) 20% of the 50% map were of the 100% map.
(D) 20% of the 50% map were of the 100% map.
Figure 4: Time profiles observed from column D (key details omitted).
Appendix H. Insights from atomic-resolution X-ray structures of chemically-synthesized HIV-1 protease in complex with inhibitors

Insights from Atomic-Resolution X-Ray Structures of Chemically Synthesized HIV-1 Protease in Complex with Inhibitors

Erik G. B. Johnson1,2,3, Enrico Melito1,2, Vaeqan Shen3,4,
Erald Penkule6,7,8,8,8, Sam Kish3,9, Jan Flinker1,
Wei-Jen Tang1,4, and Stephen H. Korgel1,4,9

1Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL 60637, USA
2Department of Biological Chemistry, The University of Chicago, Chicago, IL 60637, USA
3Department of Chemistry, The University of Chicago, Chicago, IL 60637, USA
4Department of Chemistry, The University of Chicago, Chicago, IL 60637, USA
5School of Pharmacy, The University of Wisconsin-Madison, Madison, WI 53706, USA
6Department of Chemistry, Arizona State University, Tempe, AZ 85287, USA
7School of Medicine, University of Wisconsin-Madison, Madison, WI 53706, USA
8Department of Chemistry, Arizona State University, Tempe, AZ 85287, USA
9School of Medicine, University of Wisconsin-Madison, Madison, WI 53706, USA

The structure of the HIV-1 protease (PR) in complex with a broad-spectrum, chemically synthesized inhibitor (RCV-209) revealed by X-ray crystallography. The crystal structure of RCV-209 bound to the active site of the enzyme has been determined by the National Institutes of Health. The complex exhibits a high degree of concordance with the X-ray crystal structure of RCV-209 bound to the active site of the enzyme (RCV-209). The results of this study highlight the importance of understanding the mechanism of action of PR inhibitors and the potential for developing new therapeutic strategies for the treatment of HIV-1.
Introduction

The human immunodeficiency virus 1 (HIV-1) matrix is a protein that plays a central role in the assembly and egress of HIV-1 virions. The matrix protein, also known as the matrix (MA) protein, is a transmembrane glycoprotein that is involved in the formation of the viral envelope and the packaging of the viral RNA. The matrix protein is synthesized as a pre-matrix protein that is cleaved by the viral protease into a matrix (MA) and a capsid (CA) protein. The matrix protein is involved in the formation of the viral envelope and the packaging of the viral RNA. The matrix protein is synthesized as a pre-matrix protein that is cleaved by the viral protease into a matrix (MA) and a capsid (CA) protein. The matrix protein is involved in the formation of the viral envelope and the packaging of the viral RNA. The matrix protein is synthesized as a pre-matrix protein that is cleaved by the viral protease into a matrix (MA) and a capsid (CA) protein. The matrix protein is involved in the formation of the viral envelope and the packaging of the viral RNA. The matrix protein is synthesized as a pre-matrix protein that is cleaved by the viral protease into a matrix (MA) and a capsid (CA) protein. The matrix protein is involved in the formation of the viral envelope and the packaging of the viral RNA. The matrix protein is synthesized as a pre-matrix protein that is cleaved by the viral protease into a matrix (MA) and a capsid (CA) protein. The matrix protein is involved in the formation of the viral envelope and the packaging of the viral RNA. The matrix protein is synthesized as a pre-matrix protein that is cleaved by the viral protease into a matrix (MA) and a capsid (CA) protein. The matrix protein is involved in the formation of the viral envelope and the packaging of the viral RNA.

Preliminary

Chemical synthesis of HIV-1 MA

The strategy used for the chemical synthesis of the HIV-1 MA protein is based on the previous synthesis of the HIV-1 MA protein. The strategy involves the synthesis of the HIV-1 MA protein using solid-phase peptide synthesis (SPPS). The SPPS approach is used because it allows for the efficient synthesis of large quantities of the protein. The synthesis involves the stepwise addition of amino acids to a solid support, followed by cleavage and purification of the final product.

Characterization of synthetic HIV-1 MA

The purity of the synthetic HIV-1 MA protein was determined by amino acid analysis, mass spectrometry, and biochemical assays. The protein was found to be homogeneous and to have the expected molecular weight and amino acid composition. The protein was also found to be stable under physiological conditions and to retain its biological activity.

Conclusion

The chemical synthesis of the HIV-1 MA protein using SPPS is a valuable tool for the production of large quantities of this protein for use in structural and functional studies. The synthetic HIV-1 MA protein has been shown to be pure and to have the expected molecular weight and amino acid composition. The protein has also been shown to be stable under physiological conditions and to retain its biological activity, making it a useful tool for the study of HIV-1 matrix function.
Fig. 1. Synthesis of mPsA C-terminal peptide sequence. The polymer was isolated using HPLC, and the desired polymer was purified using reverse phase HPLC. The purified polymer was then isolated into the active peptide and analyzed by mass spectroscopy. The mass spectra showed high purity. The mass error was calculated by comparing the theoretical and experimental masses. The mass error was 0.02%. The purified polymer was then isolated into the active peptide and analyzed by mass spectroscopy. The mass spectra showed high purity. The mass error was calculated by comparing the theoretical and experimental masses. The mass error was 0.02%.
Fig. 2. Characterization of refolded HIV-1 PR. (a) Cleavage of the p24/p15 substrate peptide by synthetic HIV-1 PR. p24/p15 peptide (1.8 mg, 0.5 mM) was dissolved in 1 mL of acetate buffer (50 mM, pH 5.6), and 16 pmol of enzyme was added to the peptide solution. The solution was incubated for 9 h at 37 °C and then analyzed by matrix-assisted laser desorption/ionization time-of-flight MS using a-cyano-4-hydroxy-cinnamic acid as the matrix (GHKARVL expected mass, 779.9±0.4 Da; observed mass, 779.3±0.4 Da; MQRGPRNQRK expected mass, 1381.6 Da; observed mass, 1381.8±0.2 Da; GHRARVLAEAMSQVTNSATIMQRGPRNQRK expected mass, 2218.2±0.1 Da; observed mass, 2218.2±0.1 Da; AEAMSQVTNSATIMQRGPRNQRK expected mass, 2820.2 Da; observed mass, 2820.2±0.1 Da; AEAMSQVTNSATIMQRGPRNQRK expected mass, 2820.2 Da; observed mass, 2820.2±0.1 Da). The small peaks to the right of the principal peaks correspond to sodium adducts. (b) Steady-state kinetics of synthetic HIV-1 PR. The fluorogenic substrate Ala-Tyr(4)Phe(3)Glu(1)QR-CONH₂, where Ala stands for 2-amino-2-propionic acid, and F(Ph₂) stands for p-nitro-phenylalanine, was incubated at various concentrations with the enzyme, and initial velocities were determined by the initial rate of increase in fluorescence after substrate cleavage. Assays at each concentration were performed in triplicate. Final assay conditions were as follows: 50 mM sodium acetate buffer, pH 5.6, 1% dimethyl sulfoxide, 37 °C. $K_m$ and $V_{max}$ values were determined after fitting the data points to the Michaelis–Menten equation using a nonlinear least squares fitting program in Origin 7.0.

Excellent quality of the electron density maps permitted the placement of several residues into alternative conformations and the introduction of several partially occupied water sites. However, two regions in the structures of both complexes could not be unambiguously modeled; residues 16–18 and residues 34–37 of chain B were disordered. These residues are positioned on the surface of the protein and are involved in contacts with symmetry-related molecules. Final $R_{me}$ and $R_{free}$ values of 12.7% and 16.88% were obtained for the HIV-1 PR-JG-365 complex, whereas those of 13.85% and 19.02% were obtained for the HIV-1 PR-MVT-101 complex. The presence of the nonnatural amino acids in the synthetic enzyme was readily apparent. Methionine and cysteine residues were modified to include carbon atoms instead of sulfur (L-ν-leucine and L-α-amino-n-butyric acid, respectively). In addition, the CG atom of Lys 1 was replaced by a sulfur atom (Slys). The primary structure of the synthetic pro-
Tables 4-9. Values without standard deviations.

### Table 4. Values without standard deviations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value 1</th>
<th>Value 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter 1</td>
<td>Value 1</td>
<td>Value 2</td>
</tr>
<tr>
<td>Parameter 2</td>
<td>Value 1</td>
<td>Value 2</td>
</tr>
</tbody>
</table>

The table above provides data for various parameters, each with two values. The values are not specified in the text provided.

---

### Table 5. Values without standard deviations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value 1</th>
<th>Value 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter 1</td>
<td>Value 1</td>
<td>Value 2</td>
</tr>
<tr>
<td>Parameter 2</td>
<td>Value 1</td>
<td>Value 2</td>
</tr>
</tbody>
</table>

The table above provides data for various parameters, each with two values. The values are not specified in the text provided.

---

### Table 6. Values without standard deviations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value 1</th>
<th>Value 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter 1</td>
<td>Value 1</td>
<td>Value 2</td>
</tr>
<tr>
<td>Parameter 2</td>
<td>Value 1</td>
<td>Value 2</td>
</tr>
</tbody>
</table>

The table above provides data for various parameters, each with two values. The values are not specified in the text provided.

---

### Table 7. Values without standard deviations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value 1</th>
<th>Value 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter 1</td>
<td>Value 1</td>
<td>Value 2</td>
</tr>
<tr>
<td>Parameter 2</td>
<td>Value 1</td>
<td>Value 2</td>
</tr>
</tbody>
</table>

The table above provides data for various parameters, each with two values. The values are not specified in the text provided.

---

### Table 8. Values without standard deviations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value 1</th>
<th>Value 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter 1</td>
<td>Value 1</td>
<td>Value 2</td>
</tr>
<tr>
<td>Parameter 2</td>
<td>Value 1</td>
<td>Value 2</td>
</tr>
</tbody>
</table>

The table above provides data for various parameters, each with two values. The values are not specified in the text provided.

---

### Table 9. Values without standard deviations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value 1</th>
<th>Value 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter 1</td>
<td>Value 1</td>
<td>Value 2</td>
</tr>
<tr>
<td>Parameter 2</td>
<td>Value 1</td>
<td>Value 2</td>
</tr>
</tbody>
</table>

The table above provides data for various parameters, each with two values. The values are not specified in the text provided.
The interaction of antibody X with FFr with MVT-109

We then used MVT-109 to determine the structure of the product/substrate complex. MVT-109 contains a reduced antigenic domain of FFPR including the entire 35-64 plus the 26-49 peptide. The three-dimensional structure of MVT-109 has been determined by X-ray crystallography. The results of these experiments are presented in Figure 5. The overall conformation of the protein is similar to the previously published structure. Reproduction of our results was performed with the program DISCOIL using the previously reported 35-64 peptide: MVT-109 complex also observed FFr monomer with FFPR. The conformation was refined by the program DISCOIL which is consistent with the results for MVT-109. In this way, the antibody bound either to the extracellular or intracellular domain in these two X-ray crystallographic orientations.

The structural groups on the antibodies indicate that the antibody is specific for the antigenic determinants.
In order to obtain more information on the conformational stability of the PTH-2 protein, we have performed molecular dynamics simulations on the complex. The results of these simulations are shown in Figure 1. The simulations were performed using the AMBER software package. The results show that the complex is stable and that the PTH-2 protein is well ordered.

In conclusion, the PTH-2 protein complex is a stable structure that maintains the conformational stability of the PTH-2 protein. The simulations provide valuable insights into the structure and stability of this complex.
296
The present work has been carried out to study the conformational properties of the (OH) group in polypeptides. Theoretical and experimental studies have been combined to determine the conformational preferences of the (OH) group in polypeptides. Theoretical calculations were performed using the Amber force field, and experimental studies were conducted using 2D NMR spectroscopy. The results indicate that the conformational preferences of the (OH) group are influenced by the surrounding amino acid residues.

Table 1: Hydrogen bonding and water coordination in OH groups in polypeptides.

<table>
<thead>
<tr>
<th>OH group</th>
<th>Hydrogen bonding</th>
<th>Water coordination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycine</td>
<td>0.2</td>
<td>0.0</td>
</tr>
<tr>
<td>Alanine</td>
<td>0.2</td>
<td>0.0</td>
</tr>
<tr>
<td>Serine</td>
<td>0.2</td>
<td>0.0</td>
</tr>
<tr>
<td>Threonine</td>
<td>0.2</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Theoretical calculations suggest that the (OH) group in polypeptides is involved in hydrogen bonding with neighboring amino acid residues. The water coordination of the (OH) group is minimal, indicating that the (OH) group does not interact significantly with water molecules in the polypeptide backbone.
Considering that the concentration of some uracils in aqueous solution is 0.5 M, the calculated standard binding free energies of -4.4, 1.5, -3.1, and -1.8 kcal/mol were used for N3O, 3O, 6, and 9, respectively, to represent the structure of the uracils based on the calculated structures. The standard free energies were in agreement with the observed structure of the uracils, but not with the calculated structures. The calculated standard binding free energies for the intramolecular hydrogen-bonding interactions were -2.5 kcal/mol and -1.8 kcal/mol for N3O and 3O, respectively. The calculated standard binding free energies were in agreement with the observed structural features of the uracils.
299

Materials and Methods

Materials

[Materials list and descriptions]...

Pepstatin solutions

The L-4-ATFESB-NHC3 is a highly purified Pepstatin A solution, which is used as a control in the experiments. The solutions are prepared by dissolving the Pepstatin A in distilled water to a concentration of X mg/mL. The solutions are then stored at -20°C until use.

Pepstatin assays

The enzyme activity is measured by a colorimetric method. The reaction mixture contains the substrate, the enzyme, and the Pepstatin solution. After a reaction time of X minutes, the reaction is stopped by adding a distilled water. The absorbance is measured at a wavelength of Y nm and compared to a control without Pepstatin. The activity is expressed as X units/mL.

Results

The results show that the enzyme activity is significantly increased in the presence of Pepstatin A. This effect is dose-dependent, and the optimal concentration is found to be Z mg/mL. The enzyme activity is also affected by other factors such as pH and temperature. The optimal conditions are: pH X and temperature Y°C.

Discussion

The results suggest that Pepstatin A is a potent inhibitor of the enzyme. The mechanism of action is currently under investigation. Further studies are needed to elucidate the detailed mechanisms of inhibition.

Conclusion

In conclusion, Pepstatin A is a promising inhibitor for the enzyme, and further studies are needed to determine its potential for therapeutic use.
Supplementary Data

Supplementary data associated with this article can be found in the online version of the Journal.

References

Oxidative stress and chronic inflammation

1. Oxidative stress is a major contributor to chronic diseases and can be induced by various factors such as inflammation, environmental toxins, and aging.
2. Chronic inflammation is characterized by prolonged activation of inflammatory pathways, leading to tissue damage and disease.
3. The interplay between oxidative stress and chronic inflammation can exacerbate each other, creating a vicious cycle.
4. Understanding the mechanisms underlying this bidirectional relationship is crucial for developing effective therapeutic strategies.
5. Recent advancements in research have shed light on the complex interplay between oxidative stress and chronic inflammation.
6. Pro-inflammatory cytokines and chemokines play a significant role in amplifying oxidative stress, while antioxidants and anti-inflammatory interventions can mitigate this effect.
7. Further studies are needed to elucidate the precise mechanisms and develop targeted therapies to address these chronic health challenges.
Supplementary Material

Figure S1. Total crude products from reaction 50% of (a) 1-49V/1091A: request for mass 4096.3 Da, observed mass 4096.1 ± 0.1 Da. Observed mass does not include the 80-86 Da, which gives a very broad signal, and (b) Cyclo-499: request for mass 3990.7 Da, observed mass 3990.1 ± 0.1 Da. Peptides were analysed by analytical RP-HPLC with online ESI-MS detection, using a 5-67% gradient of acetonitrile B to acetonitrile A, where A = H2O + 0.1% TFA and B = CH3CN + 0.1% TFA. UV detection was at 214 nm. Values are an average over the entire UV peak during LC-MS analysis.
Figure S2. Ligation of L-GLY-GLY to Cys-2. Silica gel by limited degradation. 3) RP-HPLC analysis of the ligation reaction as soon as the pH of the solution was adjusted to 7.0 after the peptides were mixed and a sample could be withdrawn and assayed for analysis. A large fraction of the peptides has already ligated. 5) Analysis of the ligation after 1 hour. The ligation is complete, with the remaining peaks representing undefined/impure/degraded ions. The observed mass 3779.5 Da, observed mass 3772.5 ± 1.4 Da. 6) Analysis after limited depurination of Cys-propeptide. The mass has declined by ~7 Da, representing removal of the two N-terminal groups (expected mass 19653.0 Da, observed mass 19660.1 ± 1.8 Da). Run from same sample by analytical RP-HPLC with online ESI-MS detection using a 1–65% gradient of methanol B vs. solvent A (water: methanol = 95:5). 5) and B = 0.1% ACN + 0.03% TEA. UV detection was at 214 nm. Means are an average over the predominant UV peak during LC-MS analysis.
Figure S3. Alternative orientations of KG-217. The inhibitor molecule is shown as yellow and green stick for A and B orientations, respectively. Final e, which is obtained by difference electron density maps contoured at 4σ, are shown in blue. a, b, c, and d, difference electron density maps contoured at 4σ, a, b, c, and d, are shown in blue. After modeling and refinement of KG-217 in orientation A (green stick), and b, c, d, after modeling and refinement of KG-217 in orientation B (green stick). e, f, g, and h, electron density maps contoured at 4σ, e, f, g, and h, are shown in blue. After modeling and refinement of KG-217 in alternative orientations, A and B, with occupation 50% each.
Figure 8b. Thermodynamic cycle for the calculation of the standard binding free energy ($\Delta G_{\text{bind}}$). The first step ($\Delta G_{\text{prod}}$) corresponds to the removal binding process. The process is calculated on the pathway that involves dissociation of the ligand (L) in solvent (right vertical arrow) and in the protein (E2) (left vertical arrow). The free energies for the inverse process are not because it contains identical left and right states, to which the ligand is turned off. The thermodynamic cycle requires equation: $\Delta G_{\text{bind}} = \Delta G_{\text{prod}}^{\text{reverse}} - \Delta G_{\text{prod}}^{\text{forward}}$. 
**Figure S5.** Direct attack (left) and lytic water (right) mechanisms for the reaction catalyzed by HIV-1 protease. States 1B and 1E correspond to the conformations depicted in Figure 6. For the sake of brevity, only concerted variants of the nucleophilic attack and CN bond cleavage steps are shown. In addition, the hydrolysis of the anhydride intermediate may occur via the nucleophilic attack on either of the two \( sp^3 \) carbon atoms but only one of these pathways is shown.
Appendix I: Total Chemical Synthesis of a Cysteine-Free 10FN3 monobody protein that targets the integrin α,β3

Chemical protein synthesis enabled by native chemical ligation[1] is a robust and practical method that can be used to good effect for the rapid synthesis of protein analogues. The utility of chemical protein synthesis for preparing unique analogues of the covalent structure of a protein is documented by unprecedented investigations of the effect of protein structure on protein stability and function.[2-4] The total chemical synthesis of proteins both complements and extends protein expression using recombinant DNA-based molecular biology, by providing the experimenter with a new and independent set of tools for systematically varying the structure of a target protein molecule. More specifically, chemical protein synthesis enables precise isotopic labeling with single atom precision, the atom-by-atom substitution of functional groups, the incorporation of a much wider variety of nonnatural amino acids and other building blocks into the protein molecule, modification of the polypeptide backbone, the production of homogenous glycoproteins and glycoprotein mimetics, and is versatile and useful for the ‘caging’ of protein function.[2-8]

Recently, a considerable research effort has focused on the biochemical engineering of stable protein domains (‘scaffolds’) to display novel properties. One such protein scaffold is the tenth fibronectin type III domain (10FN3) (Figure I.1). The β-sandwich fold of the 10FN3 resembles immunoglobulin domains and can be modified in the various binding loops to recognize other targets like Src[9], αvβ3 integrin[10], and ubiquitin[11] with high affinity. These novel protein constructs are referred to as 10FN3 ‘monobodies’. 10FN3’s favorable physical attributes include high stability, low

308
molecular weight, an absence of Cys residues, and good water solubility.

Furthermore, the 10FN3’s highly stable β-sandwich scaffold allows for extensive loop mutations with a minimal effect on protein stability.[11-13]

![Molecular structure of 10FN3](image)

**Figure 1.1.** (a) molecular structure of 10FN3 (PDB 1TTG). (b) target amino acid sequence

Here, we report the total chemical synthesis by modern native chemical ligation methods of a 96 residue 10FN3 monobody that was modified in the RGD loop.[10] To synthesize this protein domain we used one-pot native chemical ligation[14] and Raney nickel desulfurization (**Chapter 4**). The optimized synthesis furnished the synthetic
protein with an overall yield of ~30%, based on starting peptide segment building blocks. The final product 10FN3 monobody was folded and characterized by LCMS and by 1D $^1$H-NMR.

**Synthetic Design.** Total chemical synthesis of a protein by modern methods requires the development of a synthetic strategy based upon the use of native chemical ligation, the amide-forming ligation reaction at Xxx-Cys sites of unprotected peptide segments.[15] The sequence of 10FN3 does not contain any of the Cys residues required for the use of the native chemical ligation reaction (Figure I.1b). We then considered preparing the protein by the use of desulfurization of cysteine-containing polypeptide products obtained by native chemical ligation.[16] Desulfurization enables native chemical ligation at an Xxx-Cys site, followed by the conversion of Cys to Ala (Chapter 4); that is, effectively, native chemical ligation at Xxx-Ala sites in the target polypeptide chain. With the desulfurization approach in mind, we developed the synthetic strategy shown in Figure I.2. The synthetic strategy was devised by disconnecting the full-length amino acid sequence of the target 10FN3 polypeptide chain at suitable Xxx-Ala sites, into three peptide segments 38 amino acids, 37 amino acids, and 21 amino acids in length and thus expected to be readily accessible to preparation by optimized stepwise solid phase peptide synthesis (SPPS). The Met residue at the N-terminal of the target sequence was changed to an isosteric norleucine (Nle) residue, to avoid potential issues with desulfurization. However, in some cases Leu was used. The three peptides are then sequentially ligated in a one-pot[14] fashion, that is without intermediate isolation and purification, and the full length product is then
desulfurized to give the desired sequence of the target 10FN3 polypeptide (Figure I.1(b)).

**Figure I.2.** The synthetic strategy used for the preparation of 10FN3. The first step includes native chemical ligation followed by the conversion of the N-terminal thiazolidine (1,3-thiazolidine-4-\(R\)-carboxylic acid) to L-Cys. This is followed by a second native chemical ligation step. Once the full-length polypeptide is obtained, desulfurization is used for the conversion of Cys to Ala to give the desired cysteine-free target sequence.

**Results.** Peptide segments were synthesized, purified, and characterized as described in the Experimental section. **Figure I.3** shows the analytical data for the native chemical ligation of [Thz39-Tyr75]-\(\alpha\)-thioalkylester and [Cys76-Thr96]-COOH. The reaction was carried out in ligation buffer at pH = 6.7, 6 M Gu-HCl, 1 % thiophenol (v/v), 0.2 M phosphate—using 2 mM initial concentration of each peptide.
segment. The ligation was complete after overnight reaction, and the product [Thz39-Thr96]-COOH was converted to [Cys39-Thr96]-COOH at pH 4.0 by adding 0.2M MeONH₂·HCl directly to the reaction mixture. After confirming that the reaction was complete by LCMS (Fig 1.3(B)), the pH was readjusted to 6.7.

The synthesis was continued by reacting 10FN3[Leu1-Tyr38]-αthioalkylester with [Cys39-Thr96]-COOH, at 0.7 mM peptide concentrations in the same ligation solution used for the first ligation (Fig 1.3(C)). The reaction was diluted to 0.7 mM peptide concentration because the [Leu1-Tyr38]-αthioalkylester was not completely soluble at 2 mM. After overnight reaction, full-length peptide was obtained as confirmed by LCMS (Fig 1.3(C)), and the desired full length Cys-containing [A39C,A76C]10FN3(1-96) polypeptide was purified by reverse phase HPLC. The above reaction sequence was carried out on a 5 micromole of each peptide segment scale, and after purification 2.6 micromole of product was isolated (52 % yield).
Figure I.3. Analytical LCMS traces for the one-pot synthesis of 10FN3. (Insets) electrospray MS data for the principal component. A) is time = 0 h for the first ligation reaction between [Thz39-Tyr75]-athioalkylester and [Cys76-Thr96]-COOH. B) The crude ligation product after the conversion of Thz to Cys to give [Cys36-Thr96]-COOH (ob = 6193.3 ±0.8 Da, ca = 6192.8 Da (av isotopes)). The peak labeled (&) is the ligation catalyst thiophenol. C) is time = 0 h for the second ligation reaction between
(Figure I.3 cont.) [Leu1-Tyr38]-a-thioalkylester and [Cys36-Thr96]-COOH. D) 
LCMS analysis of the crude ligation products; the principal component at 12.7min is the 
expected [Leu1-Thr96]-COOH (ob = 10481.8 ±0.8 Da, ca = 10482.7 Da (av isotopes)). 
All chromatographic separations were carried out on a Vydac C4 2.1 X 150 mm column 
using a linear gradient of 5-65 % buffer B over 15 min (buffer A= 0.1% TFA in H2O; 
buffer B = 0.08% TFA in acetonitrile) at 0.5 mL/min and 40 °C.

The final step of the synthesis of 10FN3 involved desulfurization to convert 
Cys39 and Cys76 to Ala39 and Ala76 (Figure I.4). 10 mgs (0.95micromoles) of 
[A39C,A76C]10FN3(1-96) was dissolved in 3 mL of desulfurization buffer (6 M 
Gn•HCl phosphate buffer with 20 mM TCEP at pH = 3-4) and allowed to stand while 
preparing Raney nickel. Raney nickel was prepared by adding NaBH4 to Ni(II)acetate 
in water. The black Raney nickel was then washed with water three times, and the 
solution of [A39C, A76C]10FN3 was added. The reaction was monitored by LCMS. 
Desulfurization was complete in approximately 24 hours as indicated by a mass change 
of minus ~64 Daltons (Figure I.4). At this stage, Trp24(CHO) and Trp83(CHO) were 
deprotected (converted to Trp) by the addition of {1 volume piperidine + 1 volume β-
mercaptoethanol} to the reaction mixture, for 1 hr at 0°C. The resulting reaction mixture 
was then diluted, adjusted to pH = 2, and purified by reverse phase HPLC and 
lyophilized. After purification, 0.67 micromoles of full-length 10FN3 was isolated (7 
mgs; 70 % yield).
Figure I.4. Desulfurization [A39C,A76C]10FN3(1-96). A) t =0. B) t = ~24hrs; the principal peak is 10FN3(1-96) (ob = 10416.8 ±0.8 Da, ca = 10418.6 Da (av isotopes)). C) Purified 10FN3 after conversion of Trp(CHO) to Trp. D) Electrospray MS of the purified 10FN3 synthetic product (ob = 10360.6 ±0.8 Da, ca = 10362.6 Da (av isotopes)). All chromatographic separations were carried out on a Vydac C4 2.1 X 150 mm column using a linear gradient of 5-65 % buffer B over 15 min (buffer A= 0.1% TFA in H2O; buffer B = 0.08% TFA in acetonitrile) at 0.5 mL/min and 40 °C.
The final product 10FN3 monobody was folded simply by dilution into physiologic solution, and characterized by LCMS, and by one dimensional $^1$H -NMR (see Experimental). The $^1$H -NMR spectrum obtained for the synthetic 10FN3(1-96) is similar to the spectrum reported in the literature.[17] Future studies aim to improve the stability of 10FN3 by incorporation of fluorinated amino acids, and to investigate the chemical variation of the binding loop structures as a route to novel monobody reagents.
**Experimental:**

**Peptide Segment Synthesis (peptide-αcarboxylate or peptide-αthioester).** Protected L-amino acids were purchased from Peptides International. Peptides were prepared manually by stepwise Boc chemistry ‘in situ neutralization’ solid phase peptide synthesis[18] on a 0.4 mmol scale on -OCH2-Pam-resins (free αcarboxyl peptides) or on HSCH2CH2CO-Xaa-OCH2-Pam-resin (αthioester peptides).[19] Side-chain protection for the L-amino acids was as follows: Arg(Tos), Asn(Xan), Asp(OcHex), Cys(4-CH3Bzl), Cys(Acm), His(Bom), Glu(OBzl), Lys(2Cl-Z), Ser(Bzl), Thr(Bzl), Trp(CHO), Tyr(Br-Z). After completion of the chain assembly, peptides were deprotected and cleaved from the resin support by treatment with anhydrous HF containing p-cresol (90:10, vol/vol) for 1 h at 0 °C. After complete evaporation of the HF under reduced pressure, crude peptide products were precipitated and triturated with chilled diethyl ether. Then, the peptide products were dissolved in 50% aqueous acetonitrile containing 0.1% TFA and lyophilized.

The sequences for the peptides used in the synthesis were the following: [Leu1-Tyr38] αthioalkylester, Leu-Gln-Val-Ser-Asp-Val-Pro-Arg-Asp-Leu-Glu-Val-Val-Ala-Ala-Thr-Pro-Thr-Ser-Leu-Leu-Ile-Ser-Trp(CHO)-Asp-Ala-Pro-Ala-Val-Thr-Val-Arg-Tyr-Arg-Ile-Thr-Tyr-CO-S-CH₂-CH₂-CO-Phe-COOH (ob = 4542.6 ±0.5 Da, ca = 4543.0 Da(av isotopes)); [Thz39-Tyr75]-αthioalkylester, Thz-Glu-Thr-Gly-Gly-Asn-Ala-Pro-Val-Gln-Gln-Phe-Thr-Val-Pro-Gly-Ser-Lys-Ser-Thr-Ala-Thr-Ile-Ser-Gly-Leu-Lys-Pro-Gly-Val-Asp-Tyr-Thr-Ile-Thr-Val-Tyr-CO-S-CH₂-CH₂-CO-Val-COOH (ob = 3988.2 ±0.5 Da, ca = 3988.4 Da); and [Cys76-Thr96]-COOH, Cys-Val-Thr-Pro-Arg-
Gly-Asp-Trp(CHO)-Asn-Glu-Gly-Ser-Lys-Pro-Ile-Ser-Ile-Asn-Tyr-Arg-Thr (\( \text{ob} = 2421.4 \pm 0.5 \text{ Da}, \text{ca} = 2421.6 \text{ Da} \)).

**Analytical HPLC:** Peptide compositions were confirmed by analytical reverse phase LC-MS using a gradient of acetonitrile versus 0.1% trifluoroacetic acid (TFA) in water. For ALL the work reported in this paper, analytical HPLC was carried out as follows: Vydac C4 4.6 X 150 mm column using a linear gradient of 5-65% buffer B over 15 min at 40 °C with a flow rate of 0.5 mL/min (buffer A= 0.1% TFA in H2O; buffer B = 0.08% TFA in acetonitrile). The eluent was monitored at 214 nm, with on-line electrospray mass spectrometry (MS).

**Preparative HPLC:** Peptides were purified on C4, C8, or C18 silica with columns of dimension 22 X 250 mm, 10 X 250 mm, or 10 X 100 mm. The silica used was TP Vydac or self-packed Varian Microsorb or Agilent Zorbax. Crude peptides were loaded onto the prep column in ~10% acetonitrile/90% {0.1%TFA in water}, and eluted at a flow rate of 10 mL per minute with a shallow gradient (e.g. 20%B-40%B over 60 minutes) of increasing concentrations of solvent B (0.1% TFA in acetonitrile) in solvent A (0.1% TFA in water). Fractions containing the pure target peptide were identified by analytical LC and MALDI MS, and were combined and lyophilized. Because each peptide behaves differently, general procedures for preparative HPLC cannot be given.

**Native Chemical Ligation:** Ligation reactions were carried out under previously published conditions: 200 mM sodium phosphate buffer containing 6 M guanidine
hydrochloride, 20 mM TCEP, pH = 6.7, 1-2 mM for each peptide, 1% (v/v) thiophenol, purged and sealed under argon.[20]

**Selective Desulfurization:** Raney nickel was prepared by the slow addition while stirring of 100 mg of NaBH4 to 600 mg of Ni(OAc)2(H2O)-4 dissolved in 3 mL of deionized water. After 5 minutes, the reaction was filtered using a medium sintered glass frit and the solid activated Raney nickel washed with 200 mL of ultrapure water (18 megOhm-cm). The full amount of wet Raney nickel was added to 2-4 mg of peptide dissolved in 3 mL of 200 mM sodium phosphate pH 7 buffer containing 6 M guanidine hydrochloride and 30 mg of TCEP. The final pH ranged from 3-5 and was not adjusted. Each reaction was monitored by analytical HPLC and/or LCMS, and worked up once the starting material was consumed. The identity of the product was verified by LCMS. The product was isolated by spinning down the Raney nickel and recovering the supernatant. The pellet was washed three times with 200 mM sodium phosphate buffer containing 6 M guanidine hydrochloride to insure maximum yields.

**Folding:** The folding reaction was done by dissolving lyophilized 10FN3[1-96] in 6 M Gn•HCl, 50mM sodium phosphate, pH = 4.5, and 100 mM NaCl. The solution was then diluted to 2 M Gn•HCl with buffer to a final peptide concentration of 1 mg/mL. The peptide was then dialyzed for 18 hours against 50 mM sodium phosphate, 100 mM NaCl, pH = 4.5, by changing the buffer every 6 hours.
**One Dimensional $^1$H -NMR:**

All spectra were acquired with a Varian Unity Inova 600 MHz spectrometer equipped with a 5 mm triple resonance Cold Probe. Pulse sequences were from Varian’s Biopack.

Spectra were processed using NMRPipe and analyzed with NMRViewJ. The sample conditions were 10 mM sodium phosphate, 50 mM NaCl, pH = 4.5, 0.8 mM protein, 5% deuterium oxide.

![1-D NMR of 10FN3](image)

**Figure I.5.** 1-D NMR of 10FN3.
References:


Appendix J. Design, synthesis, and NMR of site-specifically isotope labeled L-sfAFP

In this work, we set out to study the NMR structure of sfAFP (Chapter 6) by the preparation of site-specific isotope analogues. The strategy was developed in collaboration with Josh Kurutz and Zak Gates at the University of Chicago and entails the synthesis of five isotopically-labeled analogues of the sfAFP molecule made by chemical synthesis using distinct labeling patterns of uniformly enriched ($^{15}$N, $^{13}$C$_a$, $^{13}$C$_o$)-glycine (Scheme J.1). The rationale for these labeling schemes focuses on two areas: the ability to distinguish Xxx-Gly residues from Gly-Gly sequences; and, the expected chemical shift of the beta protons of the residue preceding the G being observed.

To distinguish a Gly-Gly sequence from an Xxx-Gly, we can use the HNCO experiment, which correlates the HN and N shifts of residue i to the carbonyl $^{13}$C signal of residue i-1. In polypeptides containing only enriched Gly, these crosspeaks will only occur with Gly-Gly sequences. Second, it is normal to observe Halpha(i-1)-to-HN(glycine, i) crosspeaks, and it is reasonable to expect to observe Hbeta(i-1)-to-HN(glycine, i) crosspeaks as well. Few Halpha chemical shifts are unique to specific residue types, but Hbeta shifts can easily be used to assign crosspeaks under the right circumstances. For instance, if we have two sequence elements, such as His-Gly and Ala-Gly, and we observe Hbeta(His or Ala)-to-HN(Gly) crosspeaks, we know the one with beta $^1$H chemical shift between 2.6 and 3.9 is the one to His, and one between 1.1 and 1.6 is the Ala. Also, it may help to combine information from Gly-Gly and/or Gly signals when just one residue is between them, for example, in the sfAFP sequence - Cys-Asp-Gly-Gly-His-Gly-Gly-Asn-Gly-Gly-Asn-. The contexts of all the GG and G sequences in the sequence were tabulated to come up with the labeling design for the five labeling schemes for sfAFP shown in Scheme J.1. Polypeptides 1 and 2 site-specifically label all the Gly-Gly sequences in the sfAFP molecule, while polypeptides 3, 4, and 5 contain only some of the Gly-Gly sequences in 1 and 2 labeled, to provide redundant information of lower complexity. This may help with assigning some of the coupling patterns observed. Each polypeptide
contains an assortment of residues positioned at the i-1 position relative to Xxx-Gly or Gly-Gly sequences; the assortment was chosen to minimize probable overlap with beta $^1H$ chemical shifts.

**Isotope labeled analogues:**
1: CK GAD GAH GVN GCC GTA GAA GSV GGP GCD GGH GGN GNP GCA GGV GHA GGA GGT GV GGR GKK GGS GTPK GAD GAP GAP
2: CK GAD GAH GVN GCP GTA GAA GSV GGP GCD GGH GGN GNP GCA GGV GGA GGS GGT GV GGR GKK GGS GTPK GAD GAP GAP
3: CK GAD GAH GVN GCP GTA GAA GSV GGP GCD GGH GGN GNP GCA GGV GGA GGS GGT GV GGR GKK GGS GTPK GAD GAP GAP
4: CK GAD GAH GVN GCP GTA GAA GSV GGP GCD GGH GGN GNP GCA GGV GGA GGS GGT GV GGR GKK GGS GTPK GAD GAP GAP
5: CK GAD GAH GVN GCP GTA GAA GSV GGP GCD GGH GGN GNP GCA GGV GGA GGS GGT GV GGR GKK GGS GTPK GAD GAP GAP

**Scheme J.1** Strategy for site-specific isotope labeling of the sfAFP polypeptide chain for the NMR investigation of sfAFP. Gly residues in red are $^{15}N$, $^{13}C_α$, $^{13}C_O$-glycine.

With the design of the labeling scheme in-hand, we went ahead and prepared by total chemical synthesis analogues #1 and #2, as described in the Experimental section. The NMR experiments presented here were done on only on analogue #2 (Figure J.1). This is work that is still in progress, and our initial results are reported here for completeness.
Figure J.1. Sequence and LC-MS analysis of folded, purified site specifically isotope-labeled L-sfAFP (analogue #2). Labeled (\(^{15}\text{N},^{13}\text{C}_a,^{13}\text{C}_o\))-glycine residues are shown in red.

Results: Two site-specifically labeled analogues of L-sfAFP were prepared by total chemical synthesis, using the optimized synthetic protocols described in Chapter 6, for use in heteronuclear shift correlation experiments. Isotopically-enriched (\(^{15}\text{N},^{13}\text{C}_a,^{13}\text{C}_o\))-glycine was incorporated by coupling with a four fold excess. The amino acid sequences designed for the isotope-labeled analogue #2 of L-sfAFP is shown in Figure J.1, where the highlighted residues are (\(^{15}\text{N},^{13}\text{C}_a,^{13}\text{C}_o\))-glycine. After the labeled sfAFP #2 was folded, we expected its \(^{15}\text{N}-^1\text{H}\) HSQC (heteronuclear single quantum correlation spectroscopy)[1] spectrum, which correlates \(^{15}\text{N}\) and \(^{1}\text{H}_N\) chemical shifts, to have fifteen disperse cross peaks – one for each of the labeled glycine residues.

In fact, at 4 degrees C we observed approximately thirty cross peaks(Figure J.2). Of these thirty, about fifteen were well-dispersed, while the remaining ~fifteen cross peaks occurred within a narrow chemical shift range near the values for glycine in a random coil (\(^{1}\text{H}_N = 8.33\) ppm, \(^{15}\text{N} = 108.8\) ppm).[2] This result was interpreted as indicating the presence of at least two structurally-distinct species whose exchange is slow relative to the NMR timescale. Additionally, the clustering of cross peaks near
glycine random coil shift values suggested that we were observing an equilibrium between folded protein (the well-dispersed cross peaks) and unstructured polypeptide with cross peaks typical of a random coil\(^\#\). This hypothesis was supported by comparing the \(^{15}\)N-\(^1\)H HSQC spectra of the isotope-labeled L-sfAFP analogue #2 that were obtained with and without \(^1\)H saturation(Figures J.2 and J.3).[1] In these experiments, fifteen cross peaks exhibited strong heteronuclear NOEs, consistent with their assignment to residues participating in secondary structure, while the remaining cross peaks exhibited weak or negative NOEs, consistent with their assignment to unstructured polypeptide. It is important to note that in our hypothesis used to interpret these data, the “unstructured” cross peaks do not necessarily correspond to a single, distinct structure, but more likely to the average glycine shifts of many structures in rapid exchange.

\(^\#\) Random coils have similar chemical shifts to PPII. This may complicate the interpretation of the data.
**Figure J.2.** $^{15}\text{N}-^{1}\text{H}$ HSQC spectrum of isotope-labeled sfAFP analogue #2. Data were acquired at 4 degrees C, without $^{1}\text{H}$ saturation. Blue contours: positive; red contours: negative.
Figure J.3. $^{15}$N-$^1$H HSQC spectrum of isotope-labeled sfAFP analogue #2. Data were acquired at 4 degrees C, with $^1$H saturation. Blue contours: positive; red contours: negative.
Table J.1. Heteronuclear NOEs (calculated from Figures J.2 and J.3) for the ~30 $^{15}$N-$^1$H HSQC cross peaks the site-specifically isotope-labeled sfAFP analogue #2. Data were obtained at 4 degrees C.

We then performed $^{15}$N-$^1$H HSQC experiments on site-specifically labeled sfAFP analogue #2 as a function of temperature, and found that as the temperature was raised, the “structured” cross peaks actually increased in intensity, while the “unstructured” cross peaks decreased in intensity and in some cases disappeared entirely (see Figure J.4). This qualitative result suggested that we were observing cold denaturation of the sfAFP. Such an interpretation of these NMR data is counterintuitive because it suggests that only a small fraction of the sfAFP is folded in the temperature range where it functions (i.e. near freezing).
Figure J.4. $^{15}$N-HSQC of isotope-labeled sfAFP Data were obtained at 23 degrees C, without $^1$H saturation. Blue contours: positive; red contours: negative.

The $^{15}$N-$^1$H HSQC experiments described above gave qualitative insight into the solution behavior of sfAFP. Their interpretation, however, is complicated by the fact that the structured protein and “unstructured” polypeptide are expected to have significantly different amide $^1$H$_N$ exchange rates. At increased temperature, without further data we cannot say to what extent the disappearance of the “unstructured” peaks is due to $^1$H$_N$ exchange, and to what extent it is due to a decrease in concentration of the “unstructured” polypeptide. Even at 4 degrees C, it is possible that the “unstructured” polypeptide is substantially underrepresented in its apparent concentration due to greater amide $^1$H$_N$ exchange rates. Better probes for determining the relative populations of structured-unstructured sfAFP would be $^{13}$C-$^1$H HSQC cross peaks, which correlate $^{13}$C$_a$ and $^1$H$_a$ shifts and for which we need not consider proton
exchange. The $^{13}$C-$^1$H HSQC spectrum of site specifically isotope-labeled L-sfAFP analogue #2, taken at 4 degrees C, is shown in **Figure J.5**. The spectrum shows two sets of cross peaks: a dominant sets of peaks at the values expected for random coil glycine $^1$H$_a$ shift (3.96 ppm), and a broad, substantially-less intense set of peaks downfield approximately 1 ppm in the proton dimension. Initially, this result was difficult to interpret due to its lack of shift dispersion.[2] In light of the $^{15}$N-$^1$H HSQC results (and the now-known structure of sfAFP which consists almost exclusively of PPII helixes (**Chapter 7**), however, it seems reasonable to assign the two sets of $^{13}$C-HSQC cross peaks to the structured protein (minor peak) and “unstructured” polypeptide (major peak) species. In this interpretation, the ratio of the intensities of the two sets of cross peaks shows that at 4 degrees C, only ~10 percent of the sfAFP molecules are folded in solution (from the $^{15}$N-$^1$H HSQC data at 4 degrees C, this value appears greater than 30 percent). Another interesting feature of the $^{13}$C-$^1$H HSQC spectrum is the $^{13}$C$_\alpha$ chemical shifts. For both the major and minor cross peaks, $^{13}$C$_\alpha$ resonances are shifted up field greater than 20 ppm from the glycine random coil value (45.1 ppm).[2] Not only does this chemical shift seem remarkable in itself, but the fact that the “unstructured” polypeptide exhibits a non-random coil $^{13}$C$_\alpha$ shift is indicative of substantial residual structure in the sfAFP’s unfolded state (See Footnote #, above).
Figure J.5. $^{13}$C-$^1$H HSQC spectrum of isotope-labeled sfAFP analogue #2. Data were acquired at 4 degrees C. Blue contours: positive; red contours: negative.

Discussion. Site-specific labeling of a sub-set of numerous spectrally-overlapping amino acid residues in a protein molecule greatly simplified interpretation of our NMR results. Uniformly-labeled sfAFP would give rise to $^{15}$N-$^1$H HSQC spectra with roughly 162 cross peaks, and it is unlikely that we could have assigned these peaks to two different families as we were able to with our simplified 30-cross peak spectra. NMR experiments revealed that folded sfAFP exists in equilibrium with a less-structured form – a result that eluded our simple activity assay. Surprisingly, our results seemed to indicate that only a fraction of sfAFP molecules are folded near the freezing temperature where the protein is expected to function. Extensive further NMR
experimentation is needed in order to fully understand the structure(s) of sfAFP in solution, and to confirm the tentative interpretations presented above.

**Experimental**

**NMR experiments:** All NMR experiments reported here were performed on a sample of 6 mg isotope-labeled L-sfAFP analogue #2 in 330 μL 10 mM phosphate, 5 % D₂O, pH 6.8 buffer. All spectra were acquired with a Varian Unity Inova 600 MHz spectrometer equipped with a 5 mm triple resonance Cold Probe. Pulse sequences used were from Varian’s Biopack. Spectra were processed using NMRPipe³ and analyzed with NMRViewJ⁸.

**Non-labeled Peptide Segment Synthesis (peptide-"carboxylate or peptide-"thioester):** L-peptides were prepared manually by stepwise Boc chemistry ‘in situ neutralization’ solid phase peptide synthesis[3] on a 0.4 mmol scale on -OCH₂-Pam-resins (free "carboxyl peptides) or on HSCH₂CH₂CO-Xaa-OCH₂-Pam-resin ("thioester peptides). Side-chain protection for the L-amino acids was as follows: Arg(Tos), Asn(Xan), Asp(OcHex), Cys(4-CH₃Bzl), Cys(Acm), His(Bom), Glu(OBzl), Lys(2Cl-Z), Ser(Bzl), Thr(Bzl), Trp(CH₀), Tyr(Br-Z). After completion of the chain assembly, peptides were Nalpha-deprotected and then cleaved from the resin support with simultaneous removal of side chain protection by treatment with anhydrous HF containing p-cresol (90:10, vol/vol) for 1 h at 0 °C. After complete evaporation of the HF under reduced pressure, crude peptide products were precipitated and triturated with chilled diethyl ether. Then, the peptide products were dissolved in 50% aqueous acetonitrile containing 0.1% TFA and lyophilized.

**Isotope labeled (([¹⁵N, ¹³C₆, ¹³C₀]-glycine)) Peptide Segment Synthesis (peptide-"carboxylate or peptide-"thioester):** Labeled peptides were prepared using the same chemistry above unless noted here. Each labeled peptide was prepared on a reduced 0.1 mmol scale to conserve expensive labeled amino acid. We purchased Boc-([¹⁵N, ¹³C₆, ¹³C₀]-glycine from Cambridge Isotopes. Each coupling reaction was carried out with
0.4 millimole of Boc-(₁⁵N, ₁³C₆, ₁³C₀)-glycine, 0.39 millimole HBTU (from 0.4 M solution in DMF), and 0.250 microliters of DIEA. Standard coupling times were used. Amounts obtained and yields from this reduced scale of synthesis are given below for each specific peptide synthesized.

**Analytical LCMS.** Peptide compositions were confirmed by analytical reverse phase LC-MS using a gradient of acetonitrile versus 0.1% trifluoroacetic acid (TFA) in water. For all the work reported, unless otherwise noted, analytical HPLC was carried out as follows: Vydac C4 2.1 X 150 mm column using a linear gradient of 1-61% buffer B over 15 min with a flow rate of 0.5 mL/min (buffer A= 0.1% TFA in H₂O; buffer B = 0.08% TFA in acetonitrile) at 40 °C. The eluent was monitored at 214 nm, and by on-line ion trap electrospray mass spectrometry (MS).

**Preparative HPLC:** Peptides were purified on C4, C8, or C18 silica with columns of dimension 22 X 250 mm, 10 X 250 mm, or 10 X 100 mm. The silica used was TP Vydac or self-packed Varian Microsorb or Agilent Zorbax. Crude peptides (50-300 milligrams) were dissolved in 5% acetonitrile/95% (0.1%TFA in water) to a concentration of ~ 20 mgs/mL and loaded onto the prep column by pumping at a flow rate of 5-10 mL/min. After the non-peptidic material had eluted, as judged by the re-establishment of the 214nm baseline, the peptidic components were eluted at a flow rate of 10 mL per minute using a shallow gradient (e.g. 20%B-40%B over 60 minutes) of increasing concentrations of solvent B (0.1% TFA in acetonitrile) in solvent A (0.1% TFA in water). The exact gradient used was determined by the elution behaviour of the desired peptide, as assessed by prior analytical HPLC and confirmed by preliminary runs at low loading on the preparative column being used. Fractions containing the pure target peptide were identified by analytical LC and MALDI MS, and were combined and lyophilized.

**Native Chemical Ligation:** Ligation reactions using purified synthetic peptide segments were carried out as previously described:[4] 200 mM sodium phosphate
buffer containing 6 M guanidine hydrochloride, 20 mM TCEP, pH = 6.8, at a concentration of 5-10 mM for each peptide segment, using 30 mM 4-(carboxymethyl)thiophenol (‘MPAA’) as catalyst.[4] The ligation buffer had previously been purged with helium and the ligation reaction was carried out under argon. After the completion of each ligation, as judged by LC-MS of aliquots, methoxylamine hydrochloride (0.2 M) was directly added to the reaction mixture; the pH was lowered to 4.0. This chemical step converts the N-terminal Thz to Cys and is essentially complete in 2-4 hrs, as judged by analytical LC-MS of aliquots. 

**Total chemical synthesis of site specifically labeled sfAFP analogue #2:** The synthesis described below was carried out on a 10.9 micromole scale of each peptide segment, and after folding/disulfide formation and purification ~3 micromole (20 milligrams) of the final sfAFP product was isolated (27 % yield).

**[Thz\(^1\)-Gly\(^{12}\)]-thioester:** The unlabeled peptide Thz\(^1\)-Lys-Gly-Ala-Asp-Gly-Ala-His-Gly-Val-Asn-Gly\(^{12}\)-CO-S-CH\(_2\)-CH\(_2\)-CO-Leu-Pro-COOH was prepared on a 0.4 millimole scale using standard SPPS protocols, as described above. For a typical preparative HPLC run, ~240 mg of crude peptide was obtained after HF cleavage of only a portion of the peptide-resin obtained from the chain assembly, and after purification and lyophilization ~120 mg (50% yield) of the desired peptide was isolated: mass ob = 1395.3 ±0.5 Da, ca = 1395.4 Da (average isotopes).
Figure J.6 A) LCMS analysis of the unlabeled peptide Thz¹-Lys-Gly-Ala-Asp-Gly-Ala-His-Gly-Val-Asn-Gly¹²-CO-S-CH₂-CH₂-CO-Leu-Pro-COOH. (A) Crude product obtained after HF cleavage; the peak labeled ($) is p-cresol. (B) Peptide after purification. (C) The electrospray MS for the major peak in C. The retention times for (A) and (B) do not match because different gradients were used. For the LC trace in (A) a linear gradient of 5-65% buffer B over 15 min and in (B) a linear gradient of 1-61% buffer B over 15 min.

[Thz¹³-Gly²⁷]-thioester: The unlabeled peptide Thz¹³-Pro-Gly-Thr-Ala-Gly-Ala-Ala-Gly-Ser-Val-Gly-Gly-Pro-Gly²⁷-CO-S-CH₂-CH₂-CO-Ile-COOH was prepared on a 0.4 millimole scale using standard SPPS protocols as given above. The 1.25 g of peptide-resin was cleaved and purified directly without lyophilization. Precise isolation yields
were not recorded; however, at least 200 mgs of pure peptide was recovered after isolation: mass ob = 1371.1 ±0.5 Da, ca = 1371.4 Da (average isotopes).

**Figure J.7.** A) LCMS analysis of the unlabeled peptide Thz\(^{13}\)-Pro-Gly-Thr-Ala-Gly-Ala-Ala-Gly-Ser-Val-Gly-Gly-Pro-Gly\(^{27}\)-CO-S-CH\(_2\)-CH\(_2\)-CO-Ile-COOH. (A) Crude peptide after HF cleavage; the peak labeled ($) is p-cresol. (B) Peptide after purification. (C) The electrospray MS for the major peak.

**Labeled-(analogue #2)-[Thz\(^{28}\)-Gly\(^{42}\)]-thioester:** The peptide Thz\(^{28}\)-Asp-Gly*-Gly*-His-Gly*-Gly*-Asn-Gly*-Gly*-Asn-Gly*-Asn-Pro-Gly\(^{42}\)-CO-S-CH\(_2\)-CH\(_2\)-CO-Leu-Pro-COOH was prepared on a 0.1millimole scale as described above. (*) indicates isotope labeled Gly. The 240 milligram of peptide-resin was cleaved and purified directly without lyophilization. From this ~25 milligrams of purified peptide was isolated: mass ob = 1600.0 ±0.5 Da, ca = 1600.5 Da (average isotopes).
**Figure J.8.** A) LCMS analysis of the isotope labeled peptide Thz\textsuperscript{28}-Asp-Gly*-Gly*-His-Gly*-Gly*-Asn-Gly*-Gly*-Asn-Pro-Gly\textsuperscript{42}-CO-S-CH\textsubscript{2}-CH\textsubscript{2}-CO-Leu-Pro-COOH. (A) crude peptide after HF cleavage; the peak labeled ($) is p-cresol. The side product peak on the backside is minus 17 or 18 Da; this arises from the several Asp-Gly and Asn-Gly in the target peptide sequence. (B) Peptide after purification. (C) Electrospray MS data for the peak in C.

**Labeled-(analogue #2)-[Cys\textsuperscript{43}-Pro\textsuperscript{81}]COOH:** The peptide Cys\textsuperscript{43}-Ala-Gly-Gly-Val-Gly-Gly-Ala-Gly-Gly-Ala-Ser-Gly-Gly-Thr-Gly-Val-Gly*-Gly*-Arg-Gly*-Gly*-Lys-Gly*-Gly*-Ser-Gly*-Thr-Pro-Lys-Gly-Ala-Asp-Gly-Ala-Pro-Gly*-Ala-Pro\textsuperscript{81}-COOH was prepared as given above. (*) indicates isotope labeled Gly. A total of 577 milligrams of peptide-resin was obtained. For a typical preparative HPLC run, 200 mg
of crude peptide was purified and after lyophilization 70 mg (35 % yield) of desired peptide was isolated mass $ob = 3051.6 \pm 0.5$ Da, $ca = 3049.6$ Da (average isotopes).

**Figure J.9.** A) LCMS analysis of the isotope labeled peptide Cys$^{43}$-Ala-Gly-Gly-Val-Gly-Gly-Ala-Gly-Gly-Ala-Ser-Gly-Gly-Thr-Gly-Val-Gly*-Gly*-Arg-Gly*-Gly*-Lys-Gly*-Gly*-Ser-Gly*-Thr-Pro-Lys-Gly-Ala-Asp-Gly-Ala-Pro-Gly*-Ala-Pro$^{81}$-COOH. (A) crude peptide after HF cleavage; the peak labeled ($) is p-cresol. (B) Peptide after purification. (C) Electrospray MS for the major peak in C.
Ligation of Labeled-(analogue #2)-[Thz$^{28}$-Gly$^{42}$]-“thioester and Labeled-(2)-[Cys$^{43}$-Pro$^{81}$]-COOH: Reaction was carried out at room temperature, using concentrations of ~5 mM for each peptide, at pH 6.8 and 10 mM MPAA thiol catalyst.

Figure J.10. Analytical HPLC traces of aliquots of the ligation of labeled-(2)-[Thz$^{28}$-Gly$^{42}$]-“thioester (LC peak A2) and labeled-(2)-[Cys$^{43}$-Pro$^{81}$]-COOH (peak LC A1): A) t = 0 and B) t = overnight. In (B), the product shown was the desired labeled-(2)-[Thz$^{28}$-Pro$^{81}$]-COOH (ob = 4333.8 ± 0.8 Da, ca = 4333.4 Da). C) Is the ESMS for the product in (B). The reaction product was then treated with MeONH$_2$HCl for two hours to give labeled-(2)-[Cys$^{28}$-Pro$^{81}$]-COOH (LCMS data not shown) (ob = 4321.8 ± 0.8 Da, ca = 4321.4 Da). The peak labeled (&) is the ligation catalyst MPAA. The reaction product was isolated by solid phase extraction (SPE).
Ligation of [Thz\textsuperscript{13}-Gly\textsuperscript{27}]-\textquotesingle thioester and labeled-(analogue #2)-[Cys\textsuperscript{28}-Pro\textsuperscript{81}]-COOH: Ligation was carried out as described above. The product of this reaction was purified by RP-HPLC; in other syntheses we continued without isolation of this intermediate, thereby increasing overall yields. However, a LCMS of T=0 was not recorded.

**Figure J.11.** Analytical HPLC traces of the ligation of [Thz\textsuperscript{13}-Gly\textsuperscript{27}]-\textquotesingle thioester and labeled-(2)-[Cys\textsuperscript{28}-Pro\textsuperscript{81}]-COOH: A) [Thz\textsuperscript{13}-Gly\textsuperscript{27}]-\textquotesingle thioester and B) t = overnight. In (B), the product is the ligation product labeled-(2)-[Thz\textsuperscript{13}-Pro\textsuperscript{81}]-COOH (ob = 5474.4 ±0.5 Da, ca = 5473.8 Da). C) Is the ESMS for the product in (B). The reaction product was then treated MeONH\textsubscript{2}HCl for two hours to give labeled-(2)-[Cys\textsuperscript{13}-Pro\textsuperscript{81}]-COOH (ob = 5462.4 ±0.5 Da, ca = 5461.8 Da). The peak labeled (\&) is the ligation catalyst MPAA.
Ligation of [Thz$^1$-Gly$^{12}$]-thioester and Labeled-(analogue #2)-[Cys$^{13}$-Pro$^{81}$]-COOH: Ligation was carried out as described above. The full length reduced polypeptide was purified by reverse phase HPLC to give ~27 milligrams of the full length labeled-(2)-(Cys$^1$-Pro$^{81}$) sfAFP.

**Figure J.12.** Analytical HPLC traces of aliquots taken from the ligation of [Thz$^1$-Gly$^{12}$]-thioester (LC peak A1) and labeled-(2)-[Cys$^{13}$-Pro$^{81}$]-COOH (LC peak A2): A) t = 0 and B) t = overnight. In (B), the principal component is the desired ligation product, labeled-(2)-[Thz$^1$-Pro$^{81}$]-COOH (ob = 6539.7 ±0.7 Da, ca = 6540.8 Da). C) Is the ESMS for the product in (B). The reaction product was then treated MeONH$_2$HCl for two hours to give labeled-(2)-[Cys$^{13}$-Pro$^{81}$]-COOH (ob = 6527.7 ±0.7 Da, ca = 6528.8 Da). The peak labeled (&) is the ligation catalyst MPAA.
**Folding/Disulfide Formation for Synthetic sfAFP labeled-(analogue #2)-(Cys\textsuperscript{1}-Pro\textsuperscript{81}):** HPLC purified polypeptide (A) was folded by dissolving 2.2 micromol (15 mg) in 30 mL of pH = 7.8 folding buffer containing 8 mM cysteine, 1 mM cysteine\textcdot2HCl, and 50 mM phosphate (ob = 6527.7 ±0.7 Da, ca = 6528.8 Da). A single product (B) containing two disulfide bonds was formed within approximately 24 hours as confirmed by LCMS indicating a loss of 4 Da (ob = 6524.5 ±0.7 Da, ca = 6524.8 Da). All manipulations were carried out at 4 °C. After completion of the folding reaction, dialysis or HPLC was used to isolate the product. For the case reported here, the folding buffer was added to a 3500 MW cut-off dialysis bag and dialyzed extensively against water at 4 °C and then lyophilized to give 2.16 micromol (14 mgs) of material.

![Figure J.13](image)
Synthesis of sfAFP analogue #1: The synthesis described below was carried out on a 9.8 micromole scale of each peptide segment, and after folding/disulfide formation and purification ~3.2 micromole (19 milligrams) of the final product was isolated (32 % yield).

Labeled-(analogue #1)-[Thz¹-Gly¹²]-thioester: The peptide Thz¹-Lys-Gly-Ala-Asp-Gly-Ala-His-Gly-Val-Asn-Gly*¹²-CO-S-CH₂-CH₂-CO-Ile-Pro-COOH was prepared as described above. (*) indicates isotope labeled Gly. The 240 milligram of peptide-resin was cleaved and purified directly without lyophilization and after purification and lyophilization ~50 mg of the desired peptide was isolated: mass ob = 1397.9 ±0.5 Da, ca = 1398.4 Da (average isotopes).
Figure J.14 A) LCMS analysis of the peptide Thz\textsuperscript{1}-Lys-Gly-Ala-Asp-Gly-Ala-His-Gly-Val-Asn-Gly\textsuperscript{*12}-CO-S-CH\textsubscript{2}-CH\textsubscript{2}-CO-Ile-Pro-COOH after HF cleavage; the peak labeled ($) is p-cresol. The LC trace in (B) is the peptide after purification and (C) is the electrospray MS for the major peak.

Labeled-(analogue #1)-[Thz\textsuperscript{13}-Gly\textsuperscript{27}]-thioester: The peptide Thz\textsuperscript{13}-Pro-Gly*-Thr-Ala-Gly-Ala-Ala-Gly-Ser-Val-Gly-Gly-Pro-Gly\textsuperscript{27}-CO-S-CH\textsubscript{2}-CH\textsubscript{2}-CO-Ile-Pro-COOH was prepared described above. (*) indicates isotope labeled Gly. The 240 milligram of peptide-resin was cleaved and purified directly without lyophilization and after purification and lyophilization ~50 mg of the desired peptide was isolated: mass ob = 1471.1 ±0.5 Da, ca = 1471.5 Da (average isotopes).
Figure J.15. A) LCMS analysis of the crude peptide Thz\textsuperscript{13}-Pro-Gly*-Thr-Ala-Gly-Ala-Ala-Gly-Ser-Val-Gly-Gly-Pro-Gly\textsuperscript{27}-CO-S-CH\textsubscript{2}-CH\textsubscript{2}-CO-Ile-Pro-COOH after HF cleavage; the peak labeled ($) is p-cresol. The LC trace in (B) is the peptide after purification and (C) is the electrospray MS for the major peak.

[Thz\textsuperscript{28}-Gly\textsuperscript{42}]-thioester: The peptide Thz\textsuperscript{28}-Asp-Gly-Gly-His-Gly-Gly-Asn-Gly-Gly-Asn-Gly-Asn-Pro-Gly\textsuperscript{42}-CO-S-CH\textsubscript{2}-CH\textsubscript{2}-CO-Ile-COOH was prepared on a 0.4 millimole scale using standard SPPS protocols as given above. For a typical preparative HPLC run, 200 mg of crude peptide was purified and after lyophilization 70 mg (35 % yield) of desired peptide was isolated: mass ob = 1481.9 ±0.5 Da, ca = 1482.4 Da (average isotopes).
Figure J.16. A) LCMS analysis of the crude peptide Thz$^{28}$-Asp-Gly-Gly-His-Gly-Gly-Asn-Gly-Gly-Asn-Gly-Asn-Pro-Gly$^{42}$-CO-S-CH$_2$-CH$_2$-CO-Ile-COOH after HF cleavage; the peak labeled ($) is p-cresol. The side product peak on the backside is minus 17 or 18 Da; this arises from the several Asp-Gly and Asn-Gly in the target peptide sequence. The LC trace in (B) is the peptide after purification and (C) is the electrospray MS for the major peak. The additional ion series in (C) is from fragmentation of the peptide in the MS.

Labeled-(analogue #1)-[Cys$^{43}$-Pro$^{81}$]COOH: The peptide Cys$^{43}$-Ala-Gly*-Gly*-Val-Gly*-Gly*-Ala-Gly*-Gly*-Ala-Ser-Gly*-Gly*-Thr-Gly*-Val-Gly-Gly-Arg-Gly-Gly-Lys-Gly-Gly-Ser-Gly-Thr-Pro-Lys-Gly-Ala-Asp-Gly-Ala-Asp-Gly-Ala-Pro$^{81}$-COOH was prepared as given above. () indicates isotope labeled Gly. For a typical preparative HPLC run, 200 mg of crude peptide was purified and after lyophilization 70 mg (35% yield) of desired peptide was isolated mass ob = 3051.9.6 ±0.5 Da, ca = 3052.6 Da (average isotopes).
Figure J.17. A) LCMS analysis of the crude peptide Cys$^{43}$-Ala-Gly*-Gly*-Val-Gly*-Gly*-Ala-Gly*-Gly*-Ala-Ser-Gly*-Gly*-Thr-Gly*-Val-Gly-Gly-Arg-Gly-Gly-Lys-Gly-Gly-Ser-Gly-Thr-Pro-Lys-Gly-Ala-Asp-Gly-Ala-Pro-Gly-Ala-Pro$^{81}$-COOH after HF cleavage; the peak labeled (S) is p-cresol. The LC trace in (B) is the peptide after purification and (C) is the electrospray MS for the major peak.

Ligation of [Thz$^{28}$-Gly$^{42}$]-α-thioester and Labeled-(analogue #2)-[Cys$^{43}$-Pro$^{81}$]-COOH: Reaction was carried out at room temperature, using concentrations of ~5 mM for each peptide, at pH 6.8 and 10 mM MPAA thiol catalyst.
**Figure J.18.** Analytical HPLC traces of aliquots of the ligation of [Thz$^{28}$-Gly$^{42}$]-
"thioester (LC peak A2) and labeled-(1)-[Cys$^{43}$-Pro$^{81}$]-COOH (peak LC A1): A) t = 0
and B) t = overnight. In (B), the product shown was labeled-(1)-[Thz$^{28}$-Pro$^{81}$]-COOH
(ob = 4314.8 ± 0.8 Da, ca = 4315.4 Da). C) Is the ESMS for the product in (B). The
reaction product was then treated MeONH$_2$HCl for two hours to give labeled-(2)-
[Cys$^{28}$-Pro$^{81}$]-COOH (ob = 4302.8 ± 0.8 Da, ca = 4321.4 Da). The peak labeled (&) is
the ligation catalyst MPAA. The reaction product was isolated by solid phase extraction
(SPE).

**Ligation of Labeled-(1)-[Thz$^{13}$-Gly$^{27}$]-"thioester and labeled-(analogue #1)-[Cys$^{28}$-
Pro$^{81}$]-COOH:** Ligation was carried out as described above. Reaction was carried out
at room temperature, using concentrations of ~5 mM for each peptide, at pH 6.8 and 10
mM MPAA thiol catalyst. The product of this reaction was purified by RP-HPLC; in other syntheses we continued without isolation of this intermediate, thereby increasing overall yields.

Figure J.19. Analytical HPLC traces of aliquots of the ligation of labeled-(1)-[Thz\textsuperscript{13}-Gly\textsuperscript{27}]^{-}thioester (LC peak A2) and labeled-(1)-[Cys\textsuperscript{28}-Pro\textsuperscript{81}]-COOH (peak LC A1): A) t = 0 and B) t = overnight. In (B), the product shown was labeled-(1)-[Thz\textsuperscript{13}-Pro\textsuperscript{81}]-COOH (ob = 5457.6 ± 0.8 Da, ca = 5458.8 Da). C) Is the ESM for the product in (B). The reaction product was then treated MeONH\textsubscript{2}HCl for two hours to give labeled-(1)-[Cys\textsuperscript{13}-Pro\textsuperscript{81}]-COOH (ob = 5445.6 ± 0.8 Da, ca = 5446.8 Da). The peak labeled (&) is the ligation catalyst MPAA.
Ligation of Labeled-(1)-[Thz\(^1\)-Gly\(^{12}\)]-\(^4\)thioester and Labeled-(analogue #1)-[Cys\(^{13}\)-Pro\(^8\)]-COOH: Ligation was carried out as described above. The full length reduced polypeptide was purified by reverse phase HPLC.

**Figure J.20.** Analytical HPLC traces of aliquots taken from the ligation of labeled-(1)-[Thz\(^1\)-Gly\(^{12}\)]-\(^4\)thioester (LC peak A1) and labeled-(1)-[Cys\(^{13}\)-Pro\(^8\)]-COOH (LC peak A2): A) \(t = 0\) and B) \(t = \) overnight. In (B), the product is the ligation product labeled-(1)-[Thz\(^1\)-Pro\(^8\)]-COOH (ob = 6528.4 \(\pm\)0.7 Da, ca = 6528.8 Da). C) Is the ESMS for the product in (B). The reaction product was then treated MeONH\(_2\); HCl for two hours to give labeled-(1)-[Cys\(^{13}\)-Pro\(^8\)]-COOH (ob = 6516.4 \(\pm\)0.7 Da, ca = 6516.8 Da). The peak labeled (&) is the ligation catalyst MPAA.
Figure J.21. Folding/Disulfide Formation for Synthetic sfAFP labeled-(analogue #1)-(Cys<sup>1</sup>-Pro<sup>81</sup>): HPLC purified polypeptide (A) was folded by dissolving 3.2 micromol (21 mg) in 30 mL of pH = 7.8 folding buffer containing 8 mM cysteine, 1 mM cysteine-2HCl, and 50 mM phosphate (ob = 6516.3 ±0.7 Da, ca = 6516.8 Da). A single product (B) containing two disulfide bonds was formed within approximately 24 hours as confirmed by LCMS indicating a loss of 4 Da (ob = 6512.6 ±0.7 Da, ca = 6512.8 Da). All manipulations were carried out at 4 °C. After completion of the folding reaction, dialysis or HPLC was used to isolate the product. For the case reported here, the folding buffer was added to a 3500 MW cut-off dialysis bag and dialyzed extensively against water at 4 °C and then lyophilized to give 2.9 micromole (19 mgs) of material.
References:


