THE UNIVERSITY OF CHICAGO

PROTEINS MADE BY IMPROVED CHEMICAL METHODS UNVEIL
THE MOLECULAR BASIS OF PROTEIN STRUCTURE

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
DUHEE BANG

CHICAGO, ILLINOIS
JUNE 2005
IN MEMORY OF MY FATHER
PROTEINS MADE BY IMPROVED CHEMICAL METHODS UNVEIL THE
MOLECULAR BASIS OF PROTEIN STRUCTURE

Duhee Bang

Advisor: Stephen B. Kent, Professor of Chemistry

ABSTRACT

An original set of chemical methods has been developed and was applied to the
total synthesis of crambin to show their utilities. The three peptide segment ligation was
optimized, and the total synthesis of crambin by the three segments ligation gave a ten-
fold increase in yield (Chapter 2). A ‘one-pot’ method was developed to eliminate the
tedious purification steps and consequent handling losses (Chapter 3). Three peptides
could be ligated in one-pot, and folded in the same reaction mixture to give desired
protein molecule with only a single final purification step. The His$_6$-tag assisted chemical
protein synthesis was developed. The presence of the His$_6$ tag enabled the isolation of
peptide or protein products directly from ligation reaction mixtures by Ni-NTA affinity
column purification (Chapter 4). A kinetically controlled convergent synthesis was
devised by applying a very simple and general concept, a kinetic control of reactivity, to
the convergent synthesis of proteins (Chapter 5).

Chemically synthesized proteins were used to improve our understanding of the
functional role of a salt bridge in the crambin molecule, and our understanding of the
chemical basis of the amino acid residues found in the C’ position of C-cap of an α-helix.
The role of a salt bridge in the crambin molecule was investigated, and I showed that the
salt bridge of crambin molecule guides the formation of correct disulfide bonds (Chapter
6). The total synthesis of a protein analogue of novel topology, ‘crambin cyclotide’ was


explored, and I observed that cysteine residues found their unique disulfide bond partners in the presence of an artificial covalent bond (Chapter 7). The enantiomeric protein D-crambin was crystallized with exactly opposite handedness compared to the L-crambin (Chapter 8). A striking conservation of crystal structure between a ubiquitin molecule containing D-amino acid and a corresponding wild type ubiquitin molecule was observed (Chapter 9). By the use of ubiquitin diastereomers, I showed that the preference for the C’ Gly termination at the C-cap of protein α-helix is based primarily on conformation effects and to a much less extent on solvation effects (Chapter 10).
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Introduction to Total Synthesis of Proteins &amp; Understanding Proteins by Chemistry</td>
<td>1</td>
</tr>
<tr>
<td>1.1.</td>
<td>Chemical protein synthesis</td>
<td>1</td>
</tr>
<tr>
<td>1.2.</td>
<td>Present thesis</td>
<td>4</td>
</tr>
<tr>
<td>1.3.</td>
<td>Formulating ideas towards the practical synthesis of proteins</td>
<td>5</td>
</tr>
<tr>
<td>1.4.</td>
<td>Chemistry applied to the study of the folded structure of protein molecules</td>
<td>16</td>
</tr>
<tr>
<td>2.</td>
<td>Total Chemical Synthesis of Crambin by Ligation of Two Segments &amp; by Optimized Sequential Ligation of Three Segments</td>
<td>24</td>
</tr>
<tr>
<td>3.</td>
<td>A One-Pot Total Synthesis of Crambin</td>
<td>49</td>
</tr>
<tr>
<td>4.</td>
<td>His$_6$ Tag-Assisted Chemical Protein Synthesis</td>
<td>62</td>
</tr>
<tr>
<td>5.</td>
<td>Chemical Protein Synthesis by Convergent Ligation of Unprotected Peptide Segments</td>
<td>86</td>
</tr>
<tr>
<td>6.</td>
<td>Exploring the Effects of a Salt Bridge on the Folding of Crambin: Synthesis and Xray Crystal Structure of a Protein Analogue, Crambin‘carboxamide</td>
<td>113</td>
</tr>
<tr>
<td>7.</td>
<td>Design and Synthesis of a Protein Molecule of Novel Topology Based on Crambin</td>
<td>134</td>
</tr>
</tbody>
</table>
8. Crystal Structures of Synthetic Crambin (L-crambin), Enantiomeric Crambin (D-crambin), and Racemic Crambin (D&L-crambin) 148

9. How a Protein Adopts a D-amino Acid to Its Structure: Total Chemical Synthesis and X-ray Crystal Structure of a Protein Diastereomer: [D-Gln$^{35}$]Ubiquitin 169

10. Using Protein Diastereomers to Probe the Chemical Basis of Protein Stability:
    Why an $\alpha$-Helix Terminates with a C’ Gly Cap 186

11. Summary and Significance 202

   11.1. Better synthetic methods made total synthesis of proteins easier 202
   11.2. Proteins made by chemistry unveiled the chemical basis of protein structures 204
LIST OF FIGURES

Figure 2.1 Molecular structure of crambin 25
Figure 2.2 Target crambin amino acid sequences 27
Figure 2.3 Optimization of peptide segment synthesis 29
Figure 2.4 Ligation reaction of Cram[Cys\textsuperscript{16-31}]\textsuperscript{\textasciitilde}thioester with Cram[Cys\textsuperscript{32-46}] 34
Figure 2.5 Ligation reaction of Cram[1-V15A]\textsuperscript{\textasciitilde}thioester with Cram[Cys\textsuperscript{16-46}] 35
Figure 2.6 Folding of the polypeptide Cram[1-46] 36
Figure 2.7 Two dimensional-TOCSY \textsuperscript{1}H NMR of synthetic crambin 37
Figure 3.1 Molecular structure of crambin and target amino acid sequence. 50
Figure 3.2 A one-pot synthesis of crambin 54
Figure 4.1 His\textsubscript{6} tag-assisted total chemical synthesis of crambin-His\textsubscript{6} 68
Figure 4.2 His\textsubscript{6} tag-assisted total chemical synthesis of the modular repeat protein TPR 73
Figure 5.1 Convergent chemical ligation of a protein molecule 89
Figure 5.2 Key steps for the realization of convergent synthesis of proteins 90
Figure 5.3 Our devised scheme for the native chemical ligation of two peptide thioesters 92
Figure 5.4 Titration strategy applied model kinetically controlled ligation 94
Figure 5.5 A further revised scheme for kinetically controlled chemical ligation 96
Figure 5.6 Comparison of the ligation rates 97
Figure 5.7 Total chemical synthesis of the model protein crambin by the fully convergent ligation of six unprotected peptide segments 99
Figure 5.8 Ligation of [Thz\textsuperscript{4}-Ala\textsuperscript{15}]-\textsuperscript{\textasciitilde}thiophenylester and [Cys\textsuperscript{16}-Leu\textsuperscript{25}]-\textsuperscript{\textasciitilde}thioester 101
Figure 5.9 Crude product mixtures from kinetically controlled ligation reactions 103
Figure 5.10 Final ligation between [Thr\textsuperscript{1}-Leu\textsuperscript{25}]-\textsuperscript{\textasciitilde}thioester and [Cys\textsuperscript{36}-Asn\textsuperscript{46}] 105
Figure 6.1 Citrulline has a \textsuperscript{\alpha}urea group in the place of \textsuperscript{\delta}guanidinium of Arg 117
Figure 6.2 Stereoview of the crystal structure of crambin-\textsuperscript{\alpha}carboxamid e 117
Figure 6.3 Ribbon diagram of crambin molecule and crambin-\textsuperscript{\alpha}carboxamide 118
Figure 6.4 LC-MS profiles of purified polypeptide Cram[1-46]-\textsuperscript{\alpha}carboxamide 120
Figure 6.5 Folding crambin and crambin-\textsuperscript{\alpha}carboxamide 122
Figure 6.6 Stereoview of superimposed crystal structures of crambin-\textsuperscript{\alpha}carboxamide and crambin 125
Figure 6.7 Comparison of main chain B-factor values between crambin-\textsuperscript{\alpha}carboxamide and native crambin 127
Figure 7.1 The side chain of Arg10 and \textsuperscript{\alpha}carboxylate of Asn46 of crambin 136
Figure 7.2 Retrosynthetic analysis of the total synthesis of our novel topological analogue 139
Figure 7.3 Our improved retrosynthetic design for the total synthesis of novel crambin topological analogue 141
Figure 7.4 UV chromatogram and MS profile from LC-MS analysis of crude mixture upon folding and disulfide bond formation 143
Figure 7.5 Purified novel topological analogue by preparative HPLC 144
Figure 8.0 Target crambin amino acid sequences: the ‘PL’ form 150
Figure 8.1 The crystal structure of synthetic crambin molecule with I4\textsubscript{1} space
LIST OF SCHEMES

Scheme 2.1 Synthetic strategy for the preparation of crambin by ligation of two peptide segments 29
Scheme 2.2 Synthetic strategy for the crambin preparation by ligation of three peptide segments 32
Scheme 3.1 Synthetic strategy for a one-pot synthesis of crambin 52
Scheme 3.2 Comparison between N-terminal Cys residue protection strategies 53
Scheme 4.1 Synthetic strategies for the consecutive ligation of several peptide segments 64
Scheme 9.1 Target aa sequence for human erythrocytic [Met1Leu]ubiquitin 172
LIST OF TABLES

Table 2.1 Yields of Synthetic Peptide Segments 33
Table 3.1 Yields from one-pot syntheses 57
Table 6.1 Data collection and refinement statistics for crambin-carboxamide 124
Table 6.2 The B values of the gamma S atoms from six cysteines 126
Table 8.1 Data collection and refinement statistics for crambin 151
Table 8.2 Data collection and refinement statistics for enantiomeric crambin 159
Table 8.3 Data collection and refinement statistics for racemic crambin 163
Table 9.1 Data collection and refinement statistics 179
Table 9.2 φ, ϕ angle in each molecule 181
Table 9.3 Root mean square differences 182
Table 10.1 ΔΔG values calculated from calorimetric measurements of the stabilities of synthetic ubiquitins. 194
Table 10.2 Data collection and refinement statistics for ubiquitin molecules with P432 space group 197
ACKNOWLEDGMENTS

I am very grateful for the incredible opportunities that I have been able to enjoy while receiving my doctoral training at the University of Chicago. These opportunities always stimulated my scientific curiosity, and guided my successful achievements in science. Looking back, I was really lucky to come here, Chicago, and to meet a wonderful group of people.

Most of all, I thank Steve Kent, Brad Pentelute, and Valentina Tereshko. Not only Steve believed in my scientific talent but also he guided me to do good science, to broaden my perspective on science and to shape my own ideas on research. Brad was the one who turned my ideas to real work. He always encouraged me to try hard. He spent many nights with me in the lab, and gave me excellence advice to make my experiments better. Valya encouraged me to make crystals. Many times, she went to the synchrotron at Argonne National Lab with me, and showed me how to do crystallography. Without her encouragement and efforts, I would not have learned X-ray crystallography and would not have any protein crystal structure data for my thesis work.

I also thank people in the Kent lab, Kossiakoff lab, Sosnick lab, and Makhatadze lab. Especially, I thank Neeraj Chopra, Wendy Gordon, Dave Boerema and Alex Gribenko. These people really helped my experiments in various ways.

I am very thankful to Tony Kossiakoff and Chuan He together with my advisor Stephen Kent who served as my thesis committee and afforded very insightful criticisms.

Finally, I thank my family for spending uncountable days to pray for my health.
**ABSTRACT:** An original set of chemical methods has been developed and was applied to the total synthesis of crambin to show their utilities. The three peptide segment ligation was optimized, and the total synthesis of crambin by the three segments ligation gave a ten-fold increase in yield (Chapter 2). A ‘one-pot’ method was developed to eliminate the tedious purification steps and consequent handling losses (Chapter 3). Three peptides could be ligated in one-pot, and folded in the same reaction mixture to give desired protein molecule with only a single final purification step. The His$_6$-tag assisted chemical protein synthesis was developed. The presence of the His$_6$ tag enabled the isolation of peptide or protein products directly from ligation reaction mixtures by Ni-NTA affinity column purification (Chapter 4). A kinetically controlled convergent synthesis was devised by applying a very simple and general concept, a kinetic control of reactivity, to the convergent synthesis of proteins (Chapter 5).

Chemically synthesized proteins were used to improve our understanding of the functional role of a salt bridge in the crambin molecule, and our understanding of the chemical basis of the amino acid residues found in the C’ position of C-cap of an α-helix. The role of a salt bridge in the crambin molecule was investigated, and I showed that the salt bridge of crambin molecule guides the formation of correct disulfide bonds (Chapter 6). The total synthesis of a protein analogue of novel topology, ‘crambin cyclotide’ was explored, and I observed that cysteine residues found their unique disulfide bond partners in the presence of an artificial covalent bond (Chapter 7). The enantiomeric protein D-crambin was crystallized with exactly opposite handedness compared to the L-crambin (Chapter 8). A striking conservation of crystal structure between an ubiquitin molecule containing D-amino acid and a corresponding wild type ubiquitin molecule was observed.
(Chapter 9). By the use of ubiquitin diastereomers, I showed that the preference for the C’ Gly termination at the C-cap of protein α-helix is based primarily on conformation effects and to a much less extent on solvation effects (Chapter 10).
Chapter 1. Introduction to the Total Synthesis & Understanding Proteins by Chemistry

1.1 Chemical protein synthesis

Proteins are an abundant class of macromolecules that have been selected through the evolutionary process to perform many and varied biological functions.[1] Because of the important roles that they play, understanding proteins is a critical step to understanding living systems. Biological methods, such as recombinant DNA expression[2] and site-directed mutagenesis,[3] have played a major role for the manipulation of protein molecules to elucidate the molecular basis of their biological functions.

Recently, proteins have been recognized as a main target molecule in the interface between biology and chemistry. Chemistry is playing a more important role in the study of proteins in an effort to overcome the difficulties associated with conventional biological methods used to study proteins. Many chemical tools are being developed and applied to the study of protein molecules. Especially, systematic chemical manipulation of proteins with an atomic resolution is starting to shows utility for the dissection of the molecular basis of protein function.

In principle, chemistry could allow us to introduce biophysical probes and chemical functionalities to any desired positions in the protein molecule.[4-6] Historically, and to the present day, typical modification of proteins by chemistry has been carried out by, for example, alkylation of cysteine or acylation of lysine side chain. In some cases, and with heroic effort, even non-natural amino acids have been incorporated directly into
the site of interest by nonsense suppression mutagenesis.\textsuperscript{[5]} It is becoming obvious that the more we want to dissect the principles underlying the properties of proteins the more we need to have good chemistries to use to this end. Ultimately, we can imagine the total synthesis of proteins by chemical means.

The total chemical synthesis of proteins is an effort to enable true understanding of the protein molecule, and is a promise for the unlimited variation of their chemical structures. For chemists, the deepest level of understanding for a phenomenon also comes through the synthesis of the molecules involved. Thus, one of the tenets of chemistry is that to prove true understanding of a structure one must be able to synthesize it. The total synthesis of proteins is an attractive prospect for more practical reasons. It would enable us to answer biological questions that could not be solved by conventional methods, and in particular would allow us to introduce changes into a target protein molecule that can otherwise usually be manipulated within only the naturally occurring 20 amino acids.

There are a number of advantages that can be obtained from the total synthesis of proteins.\textsuperscript{[4, 7]} These opportunities may include the total synthesis of homogenous protein constructs for use as drugs, the production of protein molecules that could not be produced by cell-based methods, and the variation of protein covalent structure with an unlimited range of chemical functionalities for the better understanding of the molecular basis of protein function, including how the protein molecules form their three dimensional structures. Indeed, atom-by-atom control over protein molecular structure, and the incorporation of 'non-coded' analogue structures has already been used to dramatically increase our level of understanding about protein molecules.\textsuperscript{[8-10]}
The goal of total chemical synthesis of proteins is not a new dream in the chemical sciences— it existed throughout the twentieth century. Indeed, the great German chemist Emil Fischer proposed chemical protein synthesis for the first time one hundred years ago.\textsuperscript{[11]} However, in efforts to realize the practical synthesis of proteins, both solid phase peptide synthesis\textsuperscript{[12]} or convergent solution condensation of protected peptides\textsuperscript{[13]} have been pursued over long periods of time with very limited success. Only proteins the size of the insulin molecule were successfully prepared, by reproducible practical synthetic means and to acceptable chemical standards of molecular homogeneity and structural identity. Using these now-classic methods, the total synthesis of a protein the size of insulin was and remains a heroic undertaking for teams of highly skilled synthetic chemists. Clearly, for practical total chemical synthesis of proteins a new approach was needed.

About ten years ago, the intellectual history of the total synthesis of proteins gave rise to a remarkable concept, ‘chemical ligation’ i.e. the chemoselective covalent joining of unprotected peptides, devised and reported by the Kent group in the early 1990’s.\textsuperscript{[14]} The chemical ligation principle, later extended to ‘native chemical ligation’ i.e. native peptide bond forming synthesis from the chemoselective ligation of unprotected peptides, for the first time opened an era of practical chemical protein synthesis.\textsuperscript{[15]} A wide variety of proteins has been successfully prepared by total synthesis using the chemical ligation method.\textsuperscript{[4]}

Although we have experienced unprecedented success for access to the protein constructs by total chemical synthesis since the first introduction of the chemical ligation method, significant challenges still remain. Some proteins are easier to synthesize than
the others. Most reported total syntheses are still limited to relatively small proteins in
the 6.5-13kiloDalton size range.[4] Protein syntheses are still very dependent on the
success of synthesis of the peptide building blocks.[16] Obtaining high quality synthetic
peptides longer than 50 amino acid residues is not always straightforward. Indeed,
proteins larger than 200 amino acids have been produced by total synthesis only rarely.[10]
In other words, there is still plenty of room for improvement in the ‘total synthesis of
protein’.

1.2 Present thesis

In my thesis research, these questions were asked: What limits the routine
synthesis of chemical protein constructs? How can we improve the synthesis in the
matters of required time, quality of protein, and routine reproducibility? As well, with the
protein construct(s) in hand, how can chemistry be used as an approach to understand
questions in biology?

Throughout my Ph. D. work, I aimed (1) to realize the total synthesis of proteins
in a very practical manner, and (2) to understand specific questions relating to the
formation of protein the three dimensionnal structures.

The first part of this thesis presents systematic approaches towards the realization
of proteins of any size by total chemical synthesis. Thorough analysis of the difficulties
of the current approaches led me to the concept ‘divide & conquer’, that is to say, divide
the full-length polypeptide until peptide segments can be readily made, and then join
them in a fast/practical manner. Several original methodologies, covering sequential as
well as convergent ligation of peptide segments, have been developed to realize the
concept. Robustness of the methodologies has been proven using the model protein crambin, and the methods have been applied to real protein targets such as ubiquitin. More importantly, the methods development led to deep insight into the chemistry of ligation of peptide segments, and this insight has opened new approaches for the chemical synthesis of proteins.

The second part of this thesis presents the application of chemistry to the dissection of rules underlying the formation of folded protein structure. The physico-chemical basis of protein α-helix termination, and the salt bridge effect for protein stability were investigated. Studies on protein α-helix termination reported here promise to resolve the long-term controversy on how and why most proteins choose a Gly residue at the C’ position in the C-terminus of a protein α-helix. Studies on the salt bridge of crambin molecule experimentally verified the role of a key salt bridge on the protein folding & stability, as proposed 15 years ago. Use of synthetic protein constructs made possible by the application of my novel methods was key to the success of the research reported in this second part of the thesis.

In the following two introduction sections, I describe detailed background information guiding my thesis work, how I defined the questions I addressed, the difficulties I encountered, and how I approached the solution of these questions.

1.3 Formulating ideas for the practical synthesis of proteins

Lessons from the history of chemical protein synthesis
The modern era of chemical protein synthesis emerged from the ‘chemical ligation principle’. The chemical ligation principle states: (1) the use of unprotected peptides as building blocks for the total synthesis of long polypeptide chains found in protein molecules; and, (2) the use of chemoselective reaction, i.e. to covalently join these peptide building blocks. This powerful concept was first realized by the introduction of uniquely reactive functional moieties at the C-termini of one peptide and the N-termini of the other peptide; the first report of the use of chemical ligation to make a protein molecule appeared in 1992 and used the reaction between a peptide-thioacetyl and α-bromoacetyl-peptide to form a thioester-linked ligation product.

The elegance of the chemical ligation principle for the synthesis of proteins is clear when we consider traditional approaches to obtaining a full-length synthetic polypeptide, the essential precursor to the folded protein molecule. As Anfinsen vividly demonstrated, a functional protein such as ribonuclease A could be denatured and then be refolded to give its full catalytic activity; from this observation, he inferred that the folding of a protein from the corresponding polypeptide is a spontaneous process in general. Thus, great effort has been devoted to prepare a long synthetic polypeptide chain that is then to fold spontaneously to form a target protein molecule.

The length of the target polypeptide for a protein construct is however usually overwhelmingly large to be approached by conventional organic synthesis. For example, proteins of average size have 300 amino acids in their sequence, and domains-the molecular building blocks of proteins are ~150 amino acids in size. Even the smallest enzymes are usually larger than 100 residues.
Historically, two strategies were conventionally employed to realize the synthesis of target molecules of this enormous size. First, the length of peptides prepared by stepwise solid phase synthesis was simply extended to the size of small proteins, i.e. ~100+ amino acids.\textsuperscript{[18]} Although this strategy led to a reported total synthesis of ribonuclease A (124 aa),\textsuperscript{[19]} and to the synthesis of crystalline HIV-1 protease (one enzyme molecule = a homodimer of 99 residue polypeptide chains),\textsuperscript{[20]} problems from the accumulation of by-products formed by side reaction during hundreds of peptide chain elongation reactions could not be avoided. Many of these byproducts are resin-bound, and thus accumulate in a statistical fashion to contaminate the final product polypeptide when it is deprotected and released from the resin support in the last step of a synthesis.\textsuperscript{[21]} The power of the solid phase synthesis principle together with modern highly optimized chemistry for the stepwise synthesis of peptides gives recovered yields that average 96-97% per amino acid residue in the target peptide chain; thus, for a 15-20 residue peptide, an isolated yield of 50% of high purity peptide may be obtained, while for a 40-45 residue peptide a yield of 20% of high purity product may be obtained.\textsuperscript{[22]} However, for longer peptides even these extraordinary yields would not be sufficient to be able to get a highly homogeneous peptide product of defined covalent structure, even with modern purification tools such as preparative reverse phase HPLC and with the use of modern analytical tools such as electrospray mass spectrometry to guide the purification.

From studies using current state-of-the-art analytical tools, we know that obtaining high quality polypeptide constructs longer than ~50 amino acids is always difficult. In addition to the statistical accumulation of byproducts, sequence dependence
of chain elongation can cause difficulties for the synthesis of even small peptide segments,\cite{23} and this limits the utility of stepwise SPPS for the synthesis of even small proteins. Stepwise synthesis is simply not practical for the preparation of proteins of typical size (i.e. hundreds of amino acid residues).

The second conventional approach is the use of maximally protected peptides and their condensation\cite{13, 24} This strategy has resulted in the synthesis of crystalline ribonuclease A,\cite{24} as well as good syntheses of several other small proteins.\cite{13} The solution condensation of protected peptide segments has the advantage in principle that it makes possible the purification of the intermediate product from each condensation reaction. However, in practice the blocked nature of the functional groups on the peptide intermediates leads to ineffective purification, as well as difficult analytical control, and poor solubility of the side chain-protected polypeptides\cite{25} results in low concentrations and consequent slow reactions and poor yields. Furthermore, this approach is labor intensive and unusually great chemistry skills are required to successfully undertake solution synthesis by condensation of maximally-protected peptides. For these reasons, the solution synthesis of long peptides from protected peptide segments is no longer practiced.

By contrast, the chemical ligation principle shows a conceptual leap from the conventional approaches for the chemical synthesis of proteins. The chemical ligation principle recognizes the utility of \textit{unprotected} peptides as building blocks for the synthesis of long polypeptide chains.\cite{14} Because we know how to make unprotected peptides, up to about 40-50 residues, and how to purify and characterize them, chemoselective ligation of two of peptides enables us to immediately double the size of
synthetically accessible polypeptide chains, to dramatically improve the yields obtained for the chemical synthesis of proteins.

Many ligation chemistries were developed to realize the chemical ligation principle. The chemistries include: (1) the original thioester-forming chemical ligation forming ligation from the reaction of a peptide-thiocarboxylate with a bromacetyl-peptide;[14] (2) oxime-forming ligation, from the reaction between a peptide-aldehyde or -ketone and an aminoyacetyl-peptide;[26] and, (3) thiazolidine-forming ligation from the reaction between a peptide-aldehyde and N-terminal cysteinyl-peptide.[27] These methods were used to prepare synthetic protein molecules, but the synthetic proteins contained the non natural structures formed at the ligation site.

An ingenuous perfection of the chemical ligation principle was accomplished in 1994 by the introduction of ‘native chemical ligation’, i.e. native peptide bond-forming ligation of fully unprotected peptides. Ligation between a peptide-"thioester and a Cys-peptide is carried out by an initial reversible transthioesterification, resulting from the nucleophilic attack of N-terminal cysteine thiol on the thioester, followed by an irreversible intramolecular acyl transfer from sulfur to nitrogen.[15] This results in the formation of a native amide bond at the ligation site, and the regeneration of the Cys side chain. The synthetic polypeptide/protein contains no unnatural structure.

The elegance of the native chemical ligation comes from its simplicity in practice. To perform the reaction, two lyophilized unprotected peptides are simply mixed in aqueous buffer at neutral pH, in the presence of a thiol catalysts.[28] Overnight reaction leads to near-quantitative formation of the desired ligation product, and the product can be purified by HPLC to give high purity polypeptide construct in excellent yield. Native
chemical ligation has been applied to the preparation of a wide variety of synthetic protein molecules. The synthetic proteins include: enzymes such as HIV1-protease,\textsuperscript{[10]} phospholipase A2,\textsuperscript{[29]} and barnase\textsuperscript{[28]}; glycoprotein mimetic molecules for drug use, such as synthetic erythropoiesis protein\textsuperscript{[7]} and polymer-modified AOP-Rantes\textsuperscript{[30]}; integral membrane proteins, such as influenza M2 proton channel\textsuperscript{[31]} and a mechanosensitive gated ion channel\textsuperscript{[32]}; heme proteins such as cytochrome b562\textsuperscript{[33]}; and, photocycle proteins such as the photoactive yellow protein (W. Gordon, unpublished data); and various other classes of proteins (zinc fingers\textsuperscript{[34]}, etc.).

There are clear lessons from the brief history of chemical protein synthesis using modern ligation methods. These are: (1) the quality of the final protein molecule is highly dependent on the quality of the precursor full length synthetic polypeptide; (2) to obtain high quality full-length polypeptide, it is ideal to use high quality synthetic peptide building blocks; (3) for the practical assembly of peptide building blocks, unprotected peptides should be joined in a chemoselective fashion using the native chemical ligation reaction.

**Formulating ideas from the lessons: ‘divide & conquer’**

Let’s imagine that we want to chemically synthesize proteins of typical size (300 amino acid residues, and this has been never reported), or for that matter proteins of any size. Let’s further imagine that our goal is learning biology through total synthesis and routine chemical manipulation of proteins, in other words, we want to have synthetic proteins and their chemical analogues every month or week or even day.
Here are the two truths in the real world. *Truth one:* one hundred years of history for the peptide synthesis is now converging to give reasonable quality crude peptides up to about 50 amino acid residues from a two to three day synthesis, but with a high failure rate depending on the person’s skill sets (or instrument used), on the peptide sequence, or on the rigorous quality control of the individual chemicals used. For a 40 a.a. residue peptide, it usually takes two days and ends up with good quality peptide with a good success rate. For a 30 a.a., we can almost always succeed to obtain high quality peptides in a day or two.

*Truth two:* there is a great need for ligation of multiple peptides. Yet, the state-of-the-art chemical protein synthesis laboratories still tend to use ligation of as few peptide segments as possible: e.g. two 50 residue peptides for 100 residue proteins, or two 60 residue peptides for 120 residue proteins. For 150 residue proteins, they may consider three 50 residue peptides, but they may start with two 75 residue peptides just to see if they can get decent building blocks. Most of the proteins synthesized by the use of the chemical ligation were prepared from the ligation just two peptides. There are several reasons for avoiding the use of more than two peptides, and one of the main reasons is that methods for the ligation of more than two peptides are not robust enough.

Coming back to the synthesis of proteins of typical size, what would be the ideal approach? To answer this question, I reached the concept, ‘divide and conquer’. This means that any size of target protein can be synthesized by breaking the full-length polypeptide to the comfortable length of peptide segment building blocks to be synthesized. The length of individual building would vary. However, as I pointed out, high quality 30 residue peptides can be synthesized with great success rate in a day or
two. Now, we can divide our target polypeptide into multiple small peptide segments, and it will be easy to make all of these building blocks. We can even make ten different 30 residue peptides in a day or two simultaneously (i.e. the entire set of high quality building blocks for the synthesis of proteins of typical size in a day!). Now, if we have methods for the ligation of ten unprotected peptide building blocks in a reasonable fashion, we should be able to prepare our target protein.

Then, how will we ‘conquer’, in other words how can we join these peptide building blocks together? My thesis research started with the firm belief for the ‘divide and conquer’, and I continuously formulated ideas to solve the challenge, ‘conquer’. As I report below, there are several difficulties that need to be overcome to connect multiple unprotected peptide segments in a practical fashion. There was a very important rule I tried to follow: the synthetic methods have to be simple and easy.

**Sequential and convergent ligation methods; challenges and my approaches**

Here I briefly describe the challenges for the ligation of many segments. The challenges are mainly focused on the realization of practical chemical synthesis of proteins of any size.

First, there is a difficulty associated with the protecting group of an N-terminal cysteine. When we deal with the assembly of a polypeptide chain by the chemical ligation of more than two peptides using native chemical ligation, we are not dealing with *fully* unprotected peptides any more. We have to protect the N-terminal cysteine of middle peptide segments to prevent intramolecular cyclization through the reaction of N-terminal Cys and C-terminal thioester of the middle segment itself during the attempted
sequential ligation. This problem will be discussed in a great detail in Chapter 2. Here, I want to note that in principle we can solve the problem of ‘middle segments’ in two distinct ways: (i) by the use of a highly reliable N-terminal Cys protecting group; or, (ii) by the use of mechanistically distinct ligation chemistries, each of which can be performed with impunity in the presence of either of the mutually reactive functionalities for the other ligation chemistry.

There is a second practical difficulty in the chemical synthesis of large polypeptides. It comes from the laborious and wasteful procedures to isolate and purify the intermediate product after each ligation, and from similar losses after the removal of any cysteine residue protecting group. Using existing methods, significant handling losses during HPLC purification cannot be avoided at each step. In addition to this, each intermediate product has to go through a time-consuming lyophilization step for the solvent exchange to prepare for the next chemical ligation reaction. These handling procedures are very time and material consuming procedures if we have to do sequential ligation of several peptides.

Third, convergent chemical ligation is a major challenge to be solved for the chemical synthesis of proteins.\textsuperscript{[35, 36]} The utility of convergent chemical ligation would be compared to parallel computing or to organic synthesis with a fully convergent strategy (simple calculation is enough to give an appreciation of the practical importance of this convergent strategy). Previous approaches to the convergent synthesis of fully unprotected peptides have been reported.\textsuperscript{[35, 37]} However, these reports used non-peptide bond-forming ligation chemistries. Thus, products have an unnatural moiety at the ligation site.
To solve the listed challenges – practical protection of N-terminal Cys; minimal handling steps; convergent ligations- I took a somewhat different path. The originality of my thesis partly results from this new path. My synthetic journey starts with the native chemical ligation of two peptides to prepare a very small model protein, crambin (46 aa). However, one of the two peptides was very difficult to synthesize. As I pointed out above, syntheses of even small peptides are not always trivial. The synthesis was redesigned, and the 46 residue polypeptide divided into three small peptide segments to be used as building blocks. The syntheses of the smaller peptides were dramatically improved. However, in the ligation of the three segments, I ran into all kinds of problems including difficulties associated with the protecting group of cysteine residue, handling losses from the purification of intermediate products, and tedious and time consuming steps for all the necessary procedures (for ligation of only three peptides).

The three segment ligation was optimized[23] (Chapter 2) after a series of quantitative analyses. However, it could not be improved beyond a certain point, largely due to the intrinsic problems of the existing methods for the ligation of three peptides. A ‘one-pot’ method[38] (Chapter 3) was developed to solve the tedious purification steps and handling losses. In brief, three peptides could be ligated in the same solution without any intermediate purification steps, and even folded in the same reaction mixture to give a near quantitative yield of the desired protein molecule. The high utility of the ‘one-pot’ method now is evident from the experiences of several of my colleagues.

During the development of the one-pot method, I systematically studied known protecting groups for the N-terminal Cys of middle peptide segment in the three segment ligation,[23] and showed the utility of the 1,3-thiazolidine-1-carboxyl- (Thz) group for a
one-pot synthetic protocol. Also, for rapid purification and high recoveries in the sequential ligation of many peptide segments, I devised the use of His$_6$ tag in the chemical protein synthesis by attaching it to the C-terminus of protein construct (Chapter 4). The method has proved its utility in the synthesis of photoactive yellow protein by our colleague, Dr. Wendy Ryan Gordon (unpublished data).

Convergent synthesis was devised to in order to make a challenging target protein, a novel topological analogue of crambin, reminiscent of the cyclotides, that will be illustrated in the main text. Since my advisor’s first introduction of this target molecule to me, I was very interested in it, but I could not simply start for two years because I realized that the target could be made only by convergent synthesis. The question was always with me. After two years of wondering, I devised an original convergent synthetic strategy, termed ‘kinetically controlled ligation’ (to be submitted). The target protein was easily made by the use of the strategy, and the generality of the novel approach was proved by the fully convergent synthesis of the crambin protein molecule from six peptide segments (Chapter 5).

The synthetic methods developed in the course of this work have several advantages: facile access to chemical protein analogues; synthesis of highly homogeneous protein constructs; time savings; and, application to large target proteins. The facile sequential ligation as well as convergent synthesis serve as cornerstones for the next level of chemical protein synthesis. The methods will be described in a great detail from Chapter 2 to Chapter 5.
Transition; Chemical methods applied to dissect the principles underlying the folded structures of proteins.

My eventual goal is to understand biology using chemistry. My interest in the development of facile methods for chemical protein synthesis has been prompted by my desire to make chemical analogues of the proteins that are the subject of my interest. Now, it is possible to make my desired target proteins in a month or even within a week. In addition to the time saving, the ‘divide’ concept makes possible for the use of previously made peptide building blocks to produce whole libraries of chemical protein analogues by simple changes in one or more building blocks. Thus, it is very natural to think of what we can do better than conventional biological approaches with the protein analogues in hand. Here I describe a brief introductory overview to the use of chemistry to investigate some important questions pertaining to the formation of protein three-dimensional structure.

1.4 Chemistry applied to the study of the folded structure protein molecules

How proteins use salt bridges to stabilize their structures (Chapter 6 & 7)

Detailed understanding of how nature uses the salt bridge to form and maintain the folded structure of a protein molecule will be very valuable for the de novo design of stable proteins. Fifteen years ago, a salt bridge between $\delta$guanidinium of Arg10 side chain and carboxylate of Asn46 of the crambin molecule was proposed to be a very important factor to give the crambin protein molecule its exceptionally high stability.\cite{40} Despite the proposed importance of the salt bridge effect in crambin molecule, there has been no
experimental work to understand the salt bridge of crambin molecule in molecular level due to the difficult expression of recombinant crambin.\[41\]

By the use of chemical protein synthesis, I wanted to control the chemical structure of crambin at an atomic resolution and to precisely perturb the salt bridge without affecting other structural motifs in the crambin molecule. The replacement of the protein- carboxylate by - carboxamide was made possible by the facile total chemical synthesis of crambin. In this study, I focused on (1) how the perturbation of the salt bridge affects the correct disulfide bond formation of the crambin molecule, and (2) how the perturbation affects the thermal fluctuations of the folded structure of the crambin protein molecule.

Series of experimental studies strongly suggested that the salt bridge of crambin molecule guides the formation of correct disulfide bonds and validated the important role of the protein- carboxylate to Arg10 -guanidinium salt bridge in the exceptional stability of the crambin protein molecule.

Furthermore, as mentioned above, an extension of this study led to the design and total synthesis of a very interesting protein analogue of novel topology, so-called ‘crambin cyclotide’, by asking the question: what happens if we use chemistry to replace the salt bridge with a covalent bond?

**Crystal structure of synthetic crambin, of enantiomeric crambin, and of racemic crambin (Chapter 8)**

The crystal structures of natural crambin isolated from plant seeds has been the highest resolution protein structures obtained to date by X-ray diffraction. We
crystallized each of the synthetic L-crambin and D-crambin products, and we co-
crystallized the mixture of D- and L- crambin. X-ray diffraction data for the crystals were
collected using synchrotron radiation, and the crystal structures of the synthetic proteins
were solved.

Using these structures, I intended to investigate two questions: (1) Would the D-
enantiomer crystallize in the same unit cell and with the same symmetry as the L-
enantiomer with opposite handedness? This question was initially asked by Sung-Ho Kim
at Berkeley, because all known D-enantiomeric protein crystals formed with different
unit cell and symmetry from their L-enantiomers;[42] And, (2) can we obtain more
accurate phase information by forming centrosymmetric protein crystals.[43] In the
centrosymmetric crystal, the phase of the reflections are either 0 or π, and the simplicity
of the possible phases would reduce the significant effort to obtain an accurate phase for
each protein crystal. Even where adequate phases are available, the more precise phases
potentially obtained racemic crystals may yield higher quality electron density maps. In
the main text, I discuss the D-, L- and DL crambin structures in detail.

How a protein α-helix terminates? & How a protein adopts a D-amino acid to its
structure? (Chapter 9 & 10)

The chemical basis of α-helix termination was explored. The exact chemical basis
for how the α-helix terminates is still unclear and controversial. The controversy is
particularly focused on a Gly residue that is predominantly occupying the C’ position of
the C-cap region of an α-helix in most protein structures. Even though it is believed that
only Gly adopts a left-handed conformation without paying a significant energetic
penalty.\textsuperscript{[44]} others argue that the peptide backbone of the Gly in that position is more solvent exposed with favorable solvation energetic effects.\textsuperscript{[45, 46]} Studying a series of protein molecules that incorporate different D-amino acids at the C’ position of an \(\alpha\)-helix was proposed to be a unique way to understand the chemico-physical basis of the C-cap by separating conformational effects from solvation effects.

Ubiquitin molecules that incorporate D-amino acids in the place of Gly\textsuperscript{35} were designed, and six different chemical analogues of ubiquitin as well as native ubiquitin were prepared by total synthesis

Before I study thermodynamic aspects of the chemically prepared ubiquitin molecule, I carefully examined whether ubiquitins that had a D-amino acid change in local structure at the C-capping position. I did this to rule out other thermodynamic energies that could be involved for the overall stability of ubiquitin. To eliminate the possibility of structure perturbation caused by D-amino acid incorporation, I crystallized D-amino acid incorporated ubiquitins. Using high-resolution X-ray crystal structures, I have experimentally verified a striking conservation of molecular structure between the folded conformations of a protein diastereomer and the corresponding wild type protein.\textsuperscript{[47]} This preliminary study illustrated that replacement by a D-amino acid of a native Gly residue in the \(\alpha\)-helix C’ position does not significantly perturb the local or global conformations of a protein molecule.

Thermodynamic data from differential scanning calorimetric experiments (DSC) were obtained, and in combination with X-ray crystallography these studies clarified the physico-chemical basis of C-capping on the protein helical motif, and may resolve the long-term controversy about how \(\alpha\)-helix terminates.
References.


Chapter 2. Total Chemical Synthesis of Crambin by Ligation of Two Segments & by Optimized Sequential Ligation of Three Segments

Abstract. Crambin is a small (46 amino acids) protein isolated from the seeds of the plant Crambe abyssinica. Crambin has been extensively used as a model protein for the development of advanced crystallography and NMR techniques, and for computational folding studies. We set out to establish synthetic access to crambin. Initially, we synthesized the 46 amino acid polypeptide by native chemical ligation of two distinct sets of peptide segments (15+31, and 31+15 residues). The synthetic polypeptide chain folded in good yield to give native crambin containing three disulfide bonds. The chemically synthesized crambin was characterized by LC-MS and by 2D-NMR. However, the 31 residue peptide segments were difficult to purify, and this caused an overall low yield for the synthesis. To overcome this problem, we synthesized crambin by the native chemical ligation of three segments (15+16+15 residues). Total synthesis using the ligation of three segments gave more than a ten-fold increase in yield and a protein product of exceptionally high purity. This work demonstrates the efficacy of chemical protein synthesis by the native chemical ligation of three segments and establishes efficient synthetic access to the important model protein crambin for experimental studies of protein folding and stability.
**Introduction**

Our aim is to apply chemistry to the study of the principles governing protein folding. Facile total chemical synthesis of a model protein will enable a wide range of experimental folding studies by providing chemical control of the protein structure. For our model protein, we chose crambin (Figure 2.1).

![Molecular structure of crambin](image)

**Figure 2.1.** Molecular structure of crambin isolated from seeds of the plant *Crambe abyssinica*. Positions of the disulfides are shown in yellow. Coordinates were obtained from the Protein Data Bank (PDB accession no. 1AB1)

Crambin is a small (46 amino acid) protein isolated from the seeds of the plant *Crambe abyssinica*.\(^1,2\) Although no biological function has been ascribed to crambin, it is homologous with membrane-active plant toxins.\(^3\) The crambin molecule contains three disulfide bonds and displays β-strand, β-turn, and helical elements of protein secondary structure. Crambin has been widely used for the development of advanced
Crambin has also been used for ultra-high resolution X-ray crystallography\cite{7,8,9} because crambin crystals are unusually well ordered and diffract to remarkably high resolution. Because it is stable in organic media, crambin has been a useful model protein for designing enzymes to work in non-aqueous solvents.\cite{10} Despite the importance of this protein, the expression of recombinant crambin has been problematic and results in low yields.\cite{11} These difficulties have been attributed to the hydrophobic nature of the crambin molecule.

As the first step to understanding the molecular basis of crambin folding, we set out to establish an efficient total chemical synthesis of crambin. Using modern synthetic methods, the 46 amino acid residue polypeptide chain was prepared by the native chemical ligation of unprotected peptide segments.\cite{12,13} Here we report a systematic exploration of synthetic routes to crambin. First, we describe synthetic schemes involving the ligation of two segments that resulted in low overall yields. Then, we describe the efficient total chemical synthesis of crambin by means of the native chemical ligation of three unprotected peptide segments. For both synthetic routes, efficient folding of the 46-residue polypeptide to form native crambin is described. The covalent and folded molecular structures of the synthetic protein products were rigorously characterized. The optimized synthetic route can provide several hundred milligrams of synthetic crambin protein from a single lab scale synthesis.

**Results**

**Synthetic Design.** The amino acid sequences of the target 46 residue polypeptide chain are shown in Figure 2.2. Crambin isolated from nature has two isomers: one has
Ser$^{22}$ and Ile$^{25}$ (‘SI’ form) and the other has Pro$^{22}$ and Leu$^{25}$ (‘PL’ form). Therefore, we explored the synthesis of both the SI form and the PL form; the SI form by ligation of Cram[1-31]$^{\text{thioester}}$ and Cram[Cys$^{32-46}$], and the PL form by ligation of Cram[1-V15A]$^{\text{thioester}}$ and Cram[Cys$^{16-46}$].

**SI-form Crambin**

Thr-Thr-Cys-Cys-Pro-Ser-Ile-Val-Ala-Arg-
Ser-Asn-Phe-Asn-Val-Cys-Arg-Leu-Pro-Gly-
Thr-Ser-Glu-Ala-Ile-Cys-Ala-Thr-Tyr-Thr-
Gly-Cys-Ile-Ile-Pro-Gly-Ala-Thr-Cys-
Pro-Gly-Asp-Tyr-Ala-Asn$^{46}$

**PL-form [V15A]Crambin**

Thr-Thr-Cys-Cys-Pro-Ser-Ile-Val-Ala-Arg-
Ser-Asn-Phe-Asn-Ala-Cys-Arg-Leu-Pro-Gly-
Thr-Pro-Glu-Ala-Leu-Cys-Ala-Thr-Tyr-Thr-
Gly-Cys-Ile-Ile-Pro-Gly-Ala-Thr-Cys-
Pro-Gly-Asp-Tyr-Ala-Asn$^{46}$

**Figure 2.2.** Target crambin amino acid sequences: the ‘SI’ form has Ser$^{22}$ and Ile$^{25}$; the ‘PL’ form has Pro$^{22}$ and Leu$^{25}$. In the latter case, a Val15Ala substitution was made in the target sequence at the Val$^{15}$-Cys$^{16}$ ligation site to preempt anticipated slow reaction, arising from the β-branched nature of the C-terminal residue in the Cram[1-15]$^{\text{thioester}}$ segment.

**Initial Syntheses by Ligation of Two Segments (Scheme 2.1).** Native chemical ligation of two peptide segments went essentially to completion in 24 hours for both the SI and PL forms. In both cases, the synthetic polypeptide chain folded in good yield to give native crambin containing three disulfide bonds. The folded crambin proteins were characterized by reverse phase high performance liquid chromatography-mass
spectrometry (LC-MS) and by two-dimensional \(^1\)H NMR. However, overall yields of purified folded protein were low (~10 mg per synthesis). This was not because of the ligation reactions or the folding of the linear polypeptide chain, which proceeded in good yield. The most difficult part of the synthesis of crambin by this two-segment approach was synthesis of the purified 31-residue peptides. The crude products resulting from synthesis of Cram[1-31]thioester and Cram[Cys\(^{16}\)-46] contained substantial amounts of impurities eluting very close to the main peak in analytical reverse phase-high performance liquid chromatography (HPLC). The presence of these impurities caused purification by preparative HPLC to be very difficult, even with repeated purification steps. For example, from several consecutive purifications of the crude peptide Cram[Cys\(^{16}\)-46] (977 mg of crude), only 46 mg of purified peptide of acceptable purity was recovered (yield = 4.7\%). Because of the low yields of this peptide, the overall yield for synthesis of the folded protein was very low. Therefore, we decided to study the nature of the impurity peak, with the goal of avoiding it.
Scheme 2.1. Synthetic strategy for the preparation of crambin by ligation of two peptide segments.

Figure 2.3.
**Figure 2.3 continued** Optimization of peptide segment synthesis. Analytical HPLC profiles of crude peptides, monitored at 214 nm. (A) 31 residue crude peptide, Cram[Cys\textsuperscript{16-46}] showed a large des-Thr byproduct peak (arrow). (B) 16 residue crude peptide, Cram[Cys(Acm)\textsuperscript{16-31}]\textsuperscript{thioester} also showed a large des-Thr peak. (C) Mass spectrometric daughter ion analyses. The main peak and the des-Thr peak from Cram[Cys(Acm)\textsuperscript{16-31}]\textsuperscript{thioester} were analyzed by LC-MS-MS. The data showed that of the three Thr residues present in the target sequence, Thr\textsuperscript{21} was missing. (C-a) shows trapping of +2H\textsuperscript{+} ion from the main peak and its MS-MS singly charged daughter ions (Blue). C-b) shows +2H\textsuperscript{+} ion of the des-Thr peak and its MS-MS data (Red). B6 and B8 ions did not occur in the des-Thr-byproduct. (D) Crude peptide Cram[Cys(Acm)\textsuperscript{16-31}]\textsuperscript{thioester} obtained after triple coupling of the Thr\textsuperscript{21} showed no des-Thr peak.

**Optimization.** LC-MS analysis of the synthetic Cram[Cys\textsuperscript{16-46}] showed that a des-Thr peptide was the major impurity present in the crude peptide (Figure 2.3A). Because of the presence of four Thr residues in the sequence of this relatively long (for
MS-MS analysis) peptide, it was not possible to identify where the des-Thr occurred. Thus, we synthesized Cram[Cys\(^{16-31}\)]\(^{-}\)thioester to systematically investigate the nature of the des-Thr byproduct and to test the feasibility of a three segment ligation. Interestingly, we also found a des-Thr impurity as a major byproduct in synthetic Cram[Cys\(^{16-31}\)]\(^{-}\)thioester. (Figure 2.3B). We applied LC-MS-MS (i.e. daughter ion analysis of the byproduct peak in the LC-MS run) to find which of the three Thr residues (Thr\(^{21}\), Thr\(^{29}\) or Thr\(^{30}\)) was missing in the des-Thr byproduct. LC-MS-MS showed that the culprit was Thr\(^{21}\) (Figure 2.3C). Subsequent systematic studies showed that triple coupling (!) of Thr\(^{21}\) was necessary to remedy this problem. An optimized synthesis of Cram[Cys\(^{16-31}\)]\(^{-}\)thioester is shown (Figure 2.3D). By contrast, we found that synthesis of Cram[Cys\(^{16-46}\)] gave a product that still contained substantial des-Thr byproduct even with triple coupling of the Thr\(^{21}\). Consequently, we decided to apply a three-segment ligation strategy to the synthesis of crambin.

**Crambin Synthesis by Ligation of Three Segments (Scheme 2.2).** Since the middle segment has both an N-terminal cysteine and a C-terminal thioester moiety, either the \(\alpha\)-amino or the side chain thiol of the N-terminal cysteine must be protected until after the first ligation, in order to prevent undesired cyclization.\(^{[14]}\) The acetamidomethyl (Acm) group was used to protect the side chain thiol of the N-terminal cysteine of the middle peptide segment.\(^{[15,16]}\) Three peptide segments, Cram[1-V15A]\(^{-}\)thioester, Cram[Cys (Acm)\(^{16-31}\)]\(^{-}\)thioester and Cram[Cys\(^{32-46}\)] were synthesized. Purified synthetic peptides were recovered in excellent yield (~50% for each peptide) (Table 2.1); in each case only a single preparative HPLC run was needed to obtain high purity peptide.
Scheme 2.2. Synthetic strategy for the preparation of crambin by ligation of three peptide segments.
Table 2.1. Yields of Synthetic Peptide Segments

<table>
<thead>
<tr>
<th>Peptide Segment</th>
<th>Resin Scale / Crude Peptide amount</th>
<th>Purified Peptides</th>
<th>Yield (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cram[16-46]</td>
<td>0.4 mmol 977 mg</td>
<td>46 mg</td>
<td>4.7</td>
</tr>
<tr>
<td>Cram[16-31]-&quot;thioester</td>
<td>0.5 mmol *</td>
<td>474 mg (0.246 mmol)</td>
<td>49.2</td>
</tr>
<tr>
<td>Cram[32-46]</td>
<td>0.4 mmol *</td>
<td>340 mg (0.226 mmol)</td>
<td>56.4</td>
</tr>
<tr>
<td>Cram[1-V15A]-&quot;thioester</td>
<td>0.4 mmol 572 mg</td>
<td>280 mg</td>
<td>49.0</td>
</tr>
</tbody>
</table>

* Crude peptides from HF cleavage mixture were directly injected onto the preparative HPLC column, so no crude peptide weight is available.

Ligation of Cram[Cys (Acm)\textsuperscript{16-31}]"thioester and Cram[Cys\textsuperscript{32-46}], followed by Acm group removal. The ligation reaction of Cram[Cys (Acm)\textsuperscript{16-31}]"thioester and Cram[Cys\textsuperscript{32-46}] was essentially complete after 6 hours (Figure 2.4) and the ligation product was readily purified. Initially, the Acm group was removed by the widely used deprotection protocol using mercury acetate (Hg(OAc)\textsubscript{2}) and the reaction quenched with β-mercaptoethanol.\textsuperscript{15,16} However, the quenching reaction took several hours even with 20% β-mercaptoethanol (v/v), and resulted in significant loss of deprotected peptide as Cys(SH)-mercury adducts. In order to optimize the removal of the Acm group, systematic studies were performed. The deprotected peptide was prepared with minimal problems with a shorter overall reaction time by the use of Ag(OAc)\textsubscript{2} with DTT quenching.\textsuperscript{17} The recovered yield after the first ligation and Acm-removal steps was improved to 46% (details in Experimental section).
Figure 2.4. Ligation reaction of Cram[CysAcm]$^{16}$(-31)$^\text{thioester}$ with Cram[Cys$^{32}$-46] was monitored by analytical HPLC (UV profiles at 214 nm are shown). (A) At $t=5$ min, Cram[Cys(Acm)$^{16}$-31]$^\text{COS-Ph}$ ($\star$) was formed. (B) At $t=6$ hours, formation of the product Cram[Cys(Acm)$^{16}$-46] was essentially complete.

Ligation of Cram[1-V15A]$^\text{thioester}$ and Cram[Cys$^{16}$-46], followed by folding of the synthetic polypeptide chain. Ligation of Cram[1-V15A] “thioester and Cram[Cys$^{16}$-46] was essentially complete after ~24 hours (Figure 2.5). Yield of the purified ligation product Cram[1-46] was ~70%. Folding of the polypeptide using a customized variant of standard folding/disulfide formation conditions in the presence of 2M guanidine.HCl was complete in less than one hour and was very efficient giving a near-quantitative yield of folded protein (Figure 2.6). The synthetic protein was readily
purified by preparative HPLC to give highly purified crambin in good yield (~80% based on linear polypeptide).

**Characterization.** Electrospray MS of the linear polypeptide gave an observed mass (Observed: 4708.6±0.2 Da; Calculated: 4708.4 Da) consistent with the covalent structure of the desired 46 residue target sequence. The measured mass of folded crambin was consistent with formation of three disulfide bonds (Observed: 4702.8±0.4 Da; Calculated: 4702.4 Da; i.e. 5.8±0.4 Da decrease, corresponding to the loss of six protons).

![Diagram of ligation reaction](image)

**Figure 2.5.** Ligation reaction of Cram[1-V15]“thioester with Cram[Cys16-46] was monitored by analytical HPLC (UV profiles at 214 nm are shown). (A) Reaction mixture at time=0. (B) At t=24 hours, formation of the product Cram[1-46] was essentially complete.
Figure 2.6. Folding of the polypeptide Cram[1-46] was monitored by LC-MS: (A) & (B) UV profiles at 214 nm. (A) Purified polypeptide, Cram[1-46] prior to the folding reaction. (B) Folding mixture at t=one hour. Folded crambin eluted later than the linear polypeptide. (C) Mass data corresponding to the peak in (A). (D) Mass data corresponding to the peak in (B). On folding, the mass decreased by 6 Da reflecting formation of three disulfide bonds. (★ = cysteine and cystine peak)

The shape of the circular dichroism spectrum of synthetic crambin was identical to that of the natural material. The folded (tertiary) structure of crambin was characterized in detail by two dimensional ¹H NMR (Figure 2.7). Each amino acid peak from the TOCSY fingerprint region was assigned by comparison from reported data,¹⁴,¹⁸ and the observed pattern of cross peaks and the chemical shifts were identical with those
previously reported. Thus, chemically synthesized crambin was shown to be structurally identical to crambin isolated from nature.

Figure 2.7.
**Figure 2.7. continued** Two dimensional-TOCSY $^1$H NMR spectra showing the fingerprint region of folded synthetic crambin. (A) PL-[V15A]crambin; (B) SI-crambin. Assignments were by comparison with previously published data. $^{[4,18]}$

**Discussion**

Past attempts to produce the protein crambin by recombinant means have not met with success. The only published description of the recombinant expression of crambin reported a yield of only $\sim$600 µg/L of purified folded crambin, derived from a fusion
protein construct. Attempts to directly express crambin in multiple expression systems led to unacceptably low yields. To the present day, an efficient production of crambin by expression in recombinant DNA engineered cells has not been achieved (Teeter, M., private communication; Rienstra, C., private communication).

In the work reported here, high purity crambin has been prepared in good yield by optimized total chemical synthesis using modern ligation methods. Although crambin is a small protein, a synthetic approach to the 46- mino acid polypeptide chain by ligation of two unprotected peptide segments gave poor yields because of difficulties associated with the stepwise solid phase synthesis of the longer peptide segments. Systematic studies showed that a three segment ligation strategy enabled the optimization of the synthesis of all peptide segments and is a much more efficient synthetic approach. Systematic optimization of the chain assembly and peptide purification by a single preparative HPLC step gave good recovered yields of the three short peptide segments in high purity (Table 2.1). About 300 mg of each purified peptide segment was obtained from a 0.4 mmol scale solid phase peptide synthesis.

The three segment ligation strategy, comprising two ligation reactions and the associated deprotection, purification, and lyophilization steps, required several days to carry out, starting from the three purified peptide segments. We several times conducted chemical syntheses of crambin at a multi-tens-of-milligrams-of-peptide scale. From this systematic approach, we were able to improve to 46% the overall yield for the {first ligation, removal of the Acm protecting group, and preparative HPLC}. Effective removal of the Acm group from the ligation product was key to this optimization. Deprotection by mercury acetate took several hours for complete reaction and required
the addition of a large amount (20% v/v) of β-mercaptoethanol to quench the reaction and disrupt the Cys(SH)-mercury complexes. Even then, some mercuric adducts of the peptide thiols still remained. The silver acetate method was easier to handle, and a moderate amount (3% v/v) of β-mercaptoethanol or dithiothreitol (DTT) could be used to quench the reaction and completely disrupt the thiol-metal adducts, because of the formation of a metal-thiol precipitate. This made it possible to avoid a lyophilization step after preparative HPLC of the first ligation reaction. That is, after the first ligation and subsequent preparative HPLC step, silver acetate could be added directly to the pooled fractions in HPLC solvent, and after brief (1 hour) reaction the addition of β-mercaptoethanol or DTT was enough to immediately quench the reaction. Furthermore, simple centrifugation to remove the precipitate was followed by preparative HPLC of the supernatant solution to give a significantly improved yield of deprotected ligation product compared to the use of mercury acetate.

For the second ligation reaction, in repeated trials the average recovered yield after preparative HPLC purification of full length 46 residue polypeptide was around 70%.

The folding reaction was very efficient. Beginning with purified polypeptide, folding gave a single protein product as a sharp and symmetrical peak by analytical HPLC, with only trace amounts of co-products (Figure 2.6). Interestingly, the folded crambin protein molecule eluted later than the linear polypeptide in the analytical HPLC profile (Figure 2.6). For typical globular proteins, folding of the polypeptide chain results in earlier elution on HPLC analysis, because of the burial of hydrophobic side chains in the interior of the protein molecule. The contrary behavior observed for crambin
on folding is consistent with the hydrophobic nature of the folded protein. Average recovered yield from the folding reaction and subsequent preparative HPLC step was 80%. Thus, if all of the peptide was used for ligation and folding, it would be possible to make \( \sim 250 \, \text{mg} \) (\( \sim 0.05 \, \text{mmol} \)) of high purity crambin protein from a single lab scale synthesis (see Scheme 2.2). Overall, the three segment strategy gave substantially increased yield of high purity synthetic protein.

Despite the intense interest in crambin, there have been no reports of the study of this molecule by protein engineering. The inability to use protein engineering to systematically study this protein has led to a significant gap in our understanding of this protein model system, and is the consequence of the failure to produce crambin by recombinant means.\textsuperscript{12} Efficient chemical synthetic access to crambin will enable systematic studies of the molecular basis of a range of protein properties. To our knowledge, there have been no experimental folding studies done on crambin, in contrast to the numerous theoretical folding studies of crambin as a model system. Because of the facile synthetic access we report here, the molecular basis of crambin folding and the role of the disulfide bonds in its high stability can now be addressed by systematic variation of the covalent structure of the protein. In addition, synthesis of crambin analogues will be greatly facilitated by this three segment approach: preparation of each analogue protein molecule will involve the straightforward (re)synthesis of only the variant small segment(s). Chemically synthesized crambin will also enable structure determinations using methods such as protein solid state NMR (in conjunction with site-specific isotope labeling), and will provide for the further exploration of racemic protein crystallography.\textsuperscript{19,20}
Conclusion.

We have developed an efficient three segment ligation approach to the total chemical synthesis of high purity crambin in good yield. Hundred milligram amounts of the hydrophobic protein crambin can be readily prepared in a single lab scale synthesis. For the first time, this provides complete experimental control over the protein molecule so that any desired chemical modifications of crambin can be readily achieved. This control over the covalent structure of the molecule will enable the application of chemistry to the experimental elucidation of the principles of protein folding in this important model protein that has been the focus of numerous structural and theoretical studies.

Experimental Section

Abbreviations for peptides. Cram[1-V15A]“thioester: TTCCPSIVARSNFNA-SCH₂CH₂CO-Leu; Cram[Cys(Acm)₁⁶-31]“thioester: C(Acm)RLPGTPEALCATYTG-SCH₂CH₂CO-Leu, Cram[Cys₃²-46]: CIIIPGATCPGDYAN.

Materials Boc-amino acids, S-trityl-β-mercaptopropionic acid and 2-(1H-Benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) were obtained from Peptide Institute Inc (Osaka, Japan). Boc-Asn-OCH₃-Pam-resin, Boc-Leu-OCH₃-Pam-resin, and N,N-diisopropylethylamine (DIEA) were obtained from Applied Biosystems (Foster City, California). N,N-dimethylformamide (DMF), dichloromethane(DCM), and acetonitrile were purchased from Fisher (Chicago, Illinois).
Thiophenol and p-cresol were purchased from Sigma-Aldrich (St. Louis, Missouri). Trifluoroacetic acid (TFA) was from Halocarbon (New Jersey).

**HPLC and LC-MS.** Analytical reverse phase HPLC and LC-MS were performed on an Agilent 1100 Series chromatography instrument equipped with an MSD ion trap, using Vydc C4 columns (5µm, 0.46 × 25 cm). Chromatographic separations were performed using a linear gradient (10-60%) of buffer B in buffer A over 25 min at a flow rate of 1mL/min. Buffer A = 0.1% TFA in water; buffer B = 0.08% TFA in acetonitrile. For MS data, observed masses were derived from the m/z values for all protonation states of a molecular species. Calculation of theoretical masses was based on average isotope composition. Preparative HPLC was performed on a Waters Prep LC 4000 system using Vydc C4 column (12µm, 2.2 × 25 cm) at flow rate of 10mL/min, with a gradient of 20-40% buffer B in buffer A over 60 min. Fractions were pooled based on LC-MS analysis.

**Peptide Segment Synthesis.** Peptides and peptide-α-thioesters were made manually by ‘in situ neutralization’ Boc chemistry stepwise solid phase peptide synthesis,[21] on –OCH₂-Pam-Resins (free α-carboxyl peptides) or on HSCH₂CH₂CO-Leu-OCH₂-Pam-resin (α-thioester peptides).[22] Side chain protection for amino acids was as follows: Arg(Tos), Asn(Xan), Asp(OcHex), Cys(4-CH₃Bzl) or Cys(Acm), Glu(OcHex), Lys(2-Cl-Z), Ser(Bzl), Thr(Bzl), Tyr(Br-Z). For triple coupling of Thr³¹, two additional coupling steps were performed with only DMF washes intervening. 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, v/v) for 1 hour 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. Optimally, the peptide solution was diluted with water containing 0.1% TFA, and injected into the preparative HPLC system without prior lyophilization. Peptide compositions were confirmed by LC-MS. Cram[1-V15A]"thioester (observed mass (ob.) 1785.3±0.4 Da, calculated average mass (ca.) 1785.9 Da), Cram[Cys(Acm)16-31] "thioester (ob.1925.0±0.3 Da, ca.1925.3 Da), Cram[Cys32-46] (ob. 1507.3±0.5 Da, ca. 1507.7 Da)

Native Chemical Ligation of Cram[Cys(Acm)16-31]"thioester and Cram[Cys32-46]. Native chemical ligation reactions[12] were performed in aqueous solution containing 6M GnHCl, 100 mM phosphate, pH 7.5, 1% thiophenol (v/v), at concentration of ~2 mg/ml for each peptide. The ligation reaction was complete in 6 hours. The product was characterized by LC-MS analysis: Cram[Cys(Acm)16-46] (ob. 3213.3±0.3 Da, ca.3213.7 Da).

Removal of the Acm Protecting Group from Cram[Cys(Acm)16-46]. a. Acm-group removal by Hg(OAc)2 and β-mercaptoethanol. The ligated peptide Cram(Cys(Acm)16-46) was purified and lyophilized. Then, 78 mg (0.024 mmol) of purified peptide Cram[Cys(Acm)16-46] was dissolved in 39 ml of 20% aqueous acetonitrile containing 0.1% TFA. A six fold excess of Hg(OAc)2 (0.146 mmol) was added to the solution. Reaction was complete in 30 min by LC-MS, and Cram[Cys16-46]–mercury complex was formed. For the quenching reaction, 20% β-mercaptoethanol (v/v) was added, and after 3 hours, no further reaction occurred. The solution was injected onto preparative HPLC, fractions were identified by LC-MS, and the purified peptide was lyophilized (26 mg, 33% yield).
b. Acm-group removal by Ag(OAc)$_2$ and DTT or β-mercaptoethanol. Acm-group removal by Ag(OAc)$_2$ was performed directly in solution containing HPLC gradient buffer (no lyophilization step) after purification of the peptide from the ligation reaction of 74 mg (0.038 mmol) of Cram[Cys(Acm)$_{16}$-31]$^{\alpha}$thioester and 57 mg (0.0378 mmol) of Cram[Cys$_{32}$-46]. Ag(OAc)$_2$ (2 mmol) was added to ~100 ml of the solution from preparative HPLC. Reaction was complete (LC-MS) in one hour, and Cram[Cys$_{16}$-46]$^{\alpha}$–silver complex was formed. DTT (2.4 mmol) was added to the solution for quenching, and a precipitate formed immediately. After centrifugation, the supernatant was filtered and used for preparative HPLC. Fractions containing purified peptide were identified by LC-MS, combined and lyophilized (55 mg, 46.3% yield). Another Acm-group removal reaction was carried out on a similar scale under identical conditions except that 3% β-mercaptoethanol (v/v) was used in the quenching reaction. Precipitation occurred and the suspension was centrifuged. After filtration, the solution was used directly for preparative HPLC purification. Purified peptides were lyophilized (40 mg, 41.5% yield).

Native Chemical Ligation of Cram[1-V15A]$^{\alpha}$thioester and Cram[Cys$_{16}$-46].

Native chemical ligation reactions were performed in 6M GnHCl, 100 mM sodium phosphate, pH 7.5, 1% thiophenol (v/v), at a concentration of 2 mg/ml for each peptide. Ligation reactions were complete in 24 hours. The reactions were performed at 0.013-0.017 mmol scale, and purified yields obtained varied from 64% to 73%. Cram[1-46] (ob: 4708.6±0.2 Da; ca: 4708.4 Da).

Folding. The purified polypeptide chain Cram(1-46) was folded in 2M GnHCl, 100 mM Tris, 8 mM cysteine, 1 mM cystine, pH 8.0, at concentration of ~0.3 mg/ml with
exclusion of air. During the folding reaction, no stirring was performed. The reactions were performed at 0.003-0.008 mmol scales, and purified yields varied from 69% to 88%. Folded crambin was characterized by LC-MS (ob: 4702.8±0.4 Da; ca: 4702.4 Da).

2D ¹H-NMR. 2D-TOCSY spectra were taken on a Varian Inova 600 spectrometer by a previously reported protocol.²⁴ 3.5 mg of crambin was dissolved in 500 µl of 75% d₅-acetone/20% H₂O/5% D₂O. The spectra (45 millisec mixing time) was acquired with 256 complex points in the t₁ dimension and 4096 complex points in the t₂ dimension using a sweep width of 7654 Hz. Water suppression was achieved by pre-saturation. Zero-filling (×2) was applied in the t₁ dimension. Acquired data were processed with a 90° shifted sine-bell function.

References.


Chapter 3. A One-Pot Total Synthesis of Crambin

Abstract. We developed a one-pot synthesis of crambin that showed significant advantages over our previous crambin synthesis reported in Chapter 2. Total synthesis of crambin was carried out by the native chemical ligation of three unprotected peptide segments, and the folding of the synthetic polypeptide chain, all in the same reaction mixture without purification of intermediates. To accomplish this one-pot synthesis, we explored the use of the 1,3-thiazolidine-4-carboxo- (Thz) group to protect the N-terminal Cys of the middle peptide segment. The use of Thz was a key to the success of this strategy. The one-pot synthesis with only a single final purification step gave the protein molecule of exceptional purity in only two days elapsed time, with an overall yield of ~40%.
Crambin (Figure 3.1) is a small protein isolated from the plant Crambe abyssinica that has no known biological function.\cite{11, 21} It contains three disulfides, α-helix, β-sheet, and a reverse turn: all features typical of a globular protein molecule, and for that reason crambin has been extensively studied as a model. Ultra-high resolution (0.54 Å) structures have been obtained by x-ray crystallography.\cite{3, 4} Also, crambin has been used as a model system for novel protein NMR techniques,\cite{5} and for computational studies of protein folding.\cite{6, 7} Because of difficulties with recombinant DNA expression in microbial systems,\cite{8} presumably arising from its hydrophobic nature, experimental studies on the crambin molecule have been severely limited: no protein engineering of this interesting molecule has been reported.

**Figure 3.1.** Molecular structure of crambin (PDB accession no. 1AB1) and target amino acid sequence.
Our goal is to use crambin as a model system for experimental studies. We envisioned that a complete control of the chemistry of the crambin molecule would enable us to illuminate fundamental principles of protein folding. Recently, we reported an effective total chemical synthesis of [V15A]crambin. This synthesis used chemical ligation of unprotected peptides in a three segment strategy to give the full length 46 residue polypeptide chain that folded efficiently to form native crambin. However, the chemical tactics we used necessitated multiple intermediate purification and lyophilization steps, and was thus arduous and time consuming. A more effective synthesis with fewer purification steps would facilitate rapid studies of chemical variants of the crambin molecule, and would enable site-specific isotopic labeling for SS-NMR, FT-IR and other studies.

Here, we report a one-pot synthesis of crambin from three peptide segments that makes use of novel chemical tactics that enable all the ligation steps, and the folding of the full-length polypeptide chain, to be performed without purification or lyophilization of intermediate products. The synthesis makes use of a convergent strategy, with an improved protecting group for the Cys residue at the N-terminal of the middle segment. The resulting synthetic protein is correctly folded, is of exceptional purity, and is obtained rapidly and in high yield.

The synthetic strategy for the one-pot synthesis is shown in Scheme 3.1. In the one-pot synthesis, only a single, final purification step will be necessary for the production of high purity crambin. In order to make the one-pot synthesis possible, we investigated and optimized the necessary reactions. In a three-segment ligation strategy (Scheme 3.1), the N-terminal Cys residue of the middle peptide segment must be
protected in order to prevent cyclization caused by intramolecular reaction with the thioester moiety.\textsuperscript{[10, 11]} In our previous synthesis of crambin we used the acetamidomethyl (Acm) group to protect the side chain thiol of the N-terminal Cys. However, removal of the S-Acm group has proved to be problematic. For crambin we were not able to obtain a yield of more than 46 % for the combined \{ligation & deprotection\} steps (\textbf{Scheme 3.2B}), even with a highly optimized protocol.\textsuperscript{[9]}

\textbf{Scheme 3.1.} Synthetic strategy for a one-pot synthesis of crambin by ligation of three unprotected peptides segments.
Scheme 3.2. Comparison between N-terminal Cys residue protection strategies. Thiazolidine group protection strategy (A) resulted in better yield with less manipulation. Acm group protection (B) required a preparative HPLC step before removal of Acm group, and gave a lower overall yield.

We explored the use of the 1,3-thiazolidine-4-carboxo- (Thz) group (Scheme 3.2A) to protect the N-terminal Cys of the middle peptide segment. Facile conversion of a Thz-peptide to a Cys-peptide,\[12] and use of the Thz group to protect an N-terminal cysteine in native chemical ligation\[13] have been reported. Optimized stepwise solid phase synthesis\[9] of the segment Cram[Thz\[16\]-31]thioester on a 0.4 mmol scale gave a good yield (~50 %, ~ 350 milligrams) of purified middle peptide
**Figure 3.2.** A one-pot synthesis of crambin. Reactions were monitored by LC-MS. UV profile at 214 nm and mass data corresponding to major product peaks are shown. The chromatographic separations were performed using a linear gradient (10-50%) of buffer B in buffer A over 20 min (buffer A = 0.1% TFA in water; buffer B = 0.08 % TFA in acetonitrile). At t=20 h, formation of the first ligation product [Thz\textsuperscript{16}-46]Cram (retention time (rt) =14.4 min) was essentially complete. At t=24 h, conversion of Thz- to Cys-peptide (rt=14.2 min) by methoxyamine.HCl was complete. Note that the 12 dalton mass decrease and slightly earlier elution time. At t=45 h, formation of the second ligation product Cram[1-46] was essentially complete (rt=15.2 min, observed mass = 4707.7 ±0.6 Da. calculated mass = 4708.4 Da.). At t=46 h, the folding reaction was essentially complete to give crambin (rt=15.8 min, observed mass = 4702.7 ±0.4 Da. calculated mass = 4702.4 Da.). Note that folded crambin eluted later than the linear polypeptide, reflecting the more hydrophobic nature of the protein molecule # = thiophenol. * = diphenyldisulfide. & = Cram[1-V15A]-NHOCH\textsubscript{3}

The ligation of Cram[Thz\textsuperscript{16}-31]“thioester and [Cys\textsuperscript{32}-46]Cram went to completion in 20 hours (Figure 3.2, t=20 h). Addition of 0.2 M methoxyamine.HCl to the crude ligation mixture reduced the pH to ~4, and reaction for 2 hours at ambient temperature resulted in quantitative conversion of the [Thz\textsuperscript{16}-46]Cram to the desired ligation product [Cys\textsuperscript{16}-46]Cram (Figure 3.2, t=24 h). In order to evaluate the efficiency of this procedure, in preliminary studies we isolated the ligated Cys-peptide product by preparative HPLC. Recovered yields were 65–67 % (two trials). Thus, protection of the N-terminal Cys as a Thz residue was much more efficient, both in terms of recovered
yield, manipulations required and time, than the commonly used Acm group (Scheme 3.2).

Next, we explored the possibility of carrying out the second ligation reaction directly on the crude product mixture after methoxyamine·HCl treatment, i.e. without purification of the deprotected peptide [Cys\textsuperscript{16-46}]Cram. The question was whether the thioester moiety would be sensitive to the presence of methoxyamine. Model studies showed that a thioester peptide reacts only very slowly with methoxyamine in pH ~7 buffer, to form a small amount of the peptide-\textsuperscript{N}(N-methoxy)carboxamide that does not take part in the ligation reaction. Thus, we conjectured that the second ligation could be conducted by adding to the crude product mixture a slight excess of Cram[1-15]thioester peptide over [Cys\textsuperscript{16-46}]Cram. In order to effect the ligation reaction in the same batch, we had to increase the pH to ~7, because the conversion of Thz to Cys by methoxyamine·HCl had been carried out at pH ~4. Several trials were performed to optimize the pH adjustment procedure (see Experimental). Addition of the third peptide segment Cram[1-15]thioester (in ~10mole% excess) to the solution resulting from the first ligation and Thz-conversion (Figure 3.2, t=25 h) at pH ~7 resulted in quantitative ligation within 20 hours (Figure 3.2, t=45 h).

We expected that efficient folding of the product 46 residue polypeptide chain could be performed directly in this crude mixture, simply by adjusting the concentration of guanidinium·HCl (Gn·HCl) and the pH, and by addition of suitable redox reagents (if necessary). Thus, we added two volumes of pH 8.5, 0.1 M Tris buffer containing 8 mM Cys·HCl/1 mM cystine\textsuperscript{14} to dilute the Gn·HCl to 2 M and to adjust the pH to ~8, conditions previously shown\textsuperscript{9} to be optimal for the folding of crambin and formation of
disulfides. As anticipated, folding and disulfide formation (Figure 3.2, t=46 h) gave an essentially quantitative yield of the correctly folded crambin.

Now that we have optimized each reaction, can we carry out a ‘one-pot’ total synthesis of the crambin molecule? Bringing it all together, we performed a one-pot, three-segment ligation and folding on a tens-of-milligrams scale: the first ligation, in pH 7.5 phosphate buffer containing 6 M Gn.HCl; addition of methoxyamine.HCl to convert Thz- to Cys-peptide at pH 3.8; readjust the solution pH to 7.3 and addition of the third segment to effect the second ligation; and, addition of two volumes of Tris buffer to dilute the Gn.HCl, followed by addition of cysteine/cystine redox reactants\textsuperscript{[14]} to fold and form disulfides. The folded crambin molecule was purified directly from the total crude products, by prep-HPLC (Figure 3.2). The optimized crambin synthesis took 48 hours total elapsed time and was successfully reproduced three times at different scales (Table 3.1). Overall yields from starting peptide segments were 35~45 %. The covalent and tertiary structures of the synthetic protein were determined by LC-MS (Figure 3.2, t=46 h), and x-ray crystallography (Chapter 8).

Table 3.1: Yields from one-pot syntheses

<table>
<thead>
<tr>
<th></th>
<th>[1-V15A]\textsuperscript{thi}</th>
<th>Thz\textsuperscript{31-31}\textsuperscript{thi}</th>
<th>[32-46]</th>
<th>Purified crambin</th>
<th>Yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>1\textsuperscript{st} synthesis</td>
<td>50 mg (0.028 mmol)*</td>
<td>48 mg (0.026 mmol)</td>
<td>41 mg (0.026 mmol)</td>
<td>42 mg (0.009 mmol)</td>
<td>35 %</td>
</tr>
<tr>
<td>2\textsuperscript{nd} synthesis</td>
<td>28 mg (0.016 mmol)*</td>
<td>28 mg (0.015 mmol)</td>
<td>22 mg (0.015 mmol)</td>
<td>26 mg (0.0055 mmol)</td>
<td>37 %</td>
</tr>
<tr>
<td>3\textsuperscript{rd} synthesis</td>
<td>12 mg (0.007 mmol)*</td>
<td>11.2 mg (0.006 mmol)</td>
<td>9.5 mg (0.006 mmol)</td>
<td>12.5 mg (0.0027 mmol)</td>
<td>45 %</td>
</tr>
</tbody>
</table>

* Note \textasciitilde10 mole\% excess Cram[1-V15A]\textsuperscript{thi} was required for the second ligation.

This one-pot synthesis of crambin has significant advantages over our previously reported synthesis. Starting from the three purified peptide segments, the previous
synthesis took approximately ten days of uninterrupted manipulations, including multiple purifications and lyophilizations to obtain purified crambin molecule in ~25 % overall yield.\(^1\) However, the current one-pot synthesis with only one final purification step gave a synthetic crambin product of identical purity in only two days elapsed time, with an improved overall yield of ~40 %. The one-pot synthesis also vividly demonstrated the synthetic utility of the Thz residue as a reversibly protected form of N-terminal Cys.

In summary, \textit{for the first time}, we have demonstrated the total synthesis of a protein molecule by the native chemical ligation of three unprotected peptide segments, and the folding of the synthetic polypeptide chain, all in the same reaction mixture without purification of intermediates. This novel one-pot synthesis proved to be a substantially more efficient way of producing the crambin molecule. The synthetic work reported here may serve as a useful model for the synthesis of more complex protein targets. Successful use of this one-pot approach to total synthesis depends on near quantitative ligation reactions to avoid residual reactants interfering with subsequent steps. In our experience, native chemical ligation reactions proceed in essentially quantitative yield, even with only very modest excesses of one reactant. Carry over of residual amounts of unreacted segments containing Cys thiol groups could potentially interfere with folding of the full-length polypeptide product. However, reaction with such residual segments would not normally be expected to be part of a productive folding pathway, and will be reversed by the excess thiol reactant used in the folding/disulfide formation step. For these reasons, using the approach described here we anticipate that it will be straightforward to make molecules containing up to 120-150 amino acids (i.e. from three 40-50 residue segments). In principle a similar approach could be used to
assemble larger proteins from more than three unprotected peptide building blocks but, at some as yet undefined point, increasing amounts of coproducts will interfere.

**Experimental Section**

Peptides and peptide-“thioesters were made manually by ‘in situ neutralization’ Boc chemistry stepwise solid phase peptide synthesis\[^{15}\] on -OCH\(_2\)-Pam-Resins (free “carboxyl peptides) or on HSCH\(_2\)CH\(_2\)CO-Leu-OCH\(_2\)-Pam-resin\[^{16}\] (“thioester peptides). Detailed procedures for peptide syntheses and for analytical and preparative HPLC were as described\[^{9}\] The first native chemical ligation reaction\[^{17, 18}\] was performed in 6 M Gn.HCl, 100 mM sodium phosphate, pH 7.5, 1% thiophenol (v/v), at a concentration of 2 mg/ml for each peptide. Conversion of Thz- to Cys-peptide was performed by adding 0.2 M methoxyamine.HCl, causing the pH of the reaction mixture to drop to ~4. Subsequently, 0.5 M NaOH, 6 M Gn.HCl, 100 mM sodium phosphate, 1% thiophenol (v/v) solution was carefully added to the reaction mixture for adjustment to pH 7.0~7.3. The second ligation was performed by adding the peptide Cram[1-15]“thioester to the reaction mixture at room temperature. The folding reaction was performed by adding two volumes of 0.1 M Tris, pH 8.5 buffer containing 8mM Cys.HCl/1mM cystine to give a solution that was pH ~8 and 2 M Gn.HCl.

**References**


[11] Cysteine residues located in *endo*-positions of either segment are side-chain unprotected but do not prevent the correct ligation reaction even if they react with the peptide thioester, since the resulting Cys side chain thioesters can not rearrange and are transesterified to the productive N-terminal cysteine thioester.


[14] In the case of crambin, experiment showed that the thiophenol/diphenyldisulfide mixture from the prolonged second ligation reaction forms an effective redox couple that leads to efficient folding and formation of the correct disulfides, even in the absence of added cysteine/cystine. However, it may be that for other globular proteins the more polar, charged cysteine/cystine redox couple may be essential to prevent trapping of misfolded mixed disulfide forms.


Chapter 4. **His$_6$ Tag-Assisted Chemical Protein Synthesis**

**Abstract:** In order to make more practical the total chemical synthesis of proteins by the ligation of unprotected peptide building blocks, we have developed a method to facilitate the isolation and handling of intermediate products. The synthetic technique makes use of a His$_6$ tag at the C-terminal of the target polypeptide chain, introduced during the synthesis of the C-terminal peptide segment building block. Presence of a His$_6$ tag enables the isolation of peptide or protein products directly from ligation reaction mixtures by Ni-NTA affinity column purification. This simple approach enables facile buffer exchange to alternate reaction conditions and is compatible with direct analytical control by protein mass spectrometry of the multiple ligation steps involved in protein synthesis. We used syntheses of crambin and a modular tetraticopeptide repeat protein of 17kiloDaltons as models to examine the utility of this affinity purification approach. The results show that His$_6$ tag-assisted chemical protein synthesis is a useful method that substantially reduces handling losses and provides for rapid chemical protein syntheses.
**Introduction**

Chemical ligation\cite{1} enables the synthesis of large polypeptide chains by the chemoselective reaction of *unprotected* peptide segments. A variety of ligation chemistries has been developed for this purpose.\cite{2-11} Application of the chemical ligation principle\cite{1} has led to practical chemical syntheses of a wide variety of different classes of proteins.\cite{12} Synthetic access to protein molecules has been used to elucidate the molecular basis of protein folding and stability,\cite{13} to elucidate the molecular basis of protein function\cite{14}, to design and build proteins of novel structure,\cite{15} and to determine the molecular structure of proteins by both NMR\cite{16} and by X-ray crystallography.\cite{17} Chemical protein synthesis has also been used to develop candidate protein therapeutic molecules with improved properties.\cite{17,18}

Thus far, most research on chemically synthesized protein molecules\cite{12} has been focused on relatively small proteins in the size range of 50 ~ 150 amino acids (a.a.) (i.e. made from two or three peptide segments). This size limitation is the result of two phenomena: (i) practical constraints on the size of the unprotected peptide segment building blocks:\cite{19} and, (ii) technical challenges in the chemical ligation of more than two or three peptide segments. Even highly optimized stepwise solid phase peptide synthetic procedures max out at ~50 amino acid residues for the practical preparation of high purity unprotected peptides.\cite{20} Thus, the chemical synthesis of a protein of typical size (~300 amino acids)\cite{21} would require the use of at least six synthetic peptide segments as building blocks.
Scheme 4.1. Synthetic strategies (1A for conventional and 1B for our novel His₆ tag-assisted) for the consecutive chemical ligation of several peptide segments to give a target protein/polypeptide chain. His₆ tag-assisted chemical protein synthesis (1B) dramatically speeds up and facilitates the isolation and handling of intermediate ligation products.

The most effective way of covalently joining unprotected peptide segments to form a protein molecule is native chemical ligation[2]. Native chemical ligation involves the reaction of a peptide-^4^thioester with a Cys-peptide; reversible thioester-thiol exchange with the N-terminal Cys residue gives a thioester-linked intermediate that undergoes an irreversible intramolecular rearrangement, to give a near quantitative yield of a single product linked by a native peptide bond at the ligation site. The native chemical ligation
reaction is both practical and highly effective.\(^a\) Each ligation product can be purified by reverse phase HPLC and characterized with great precision by electrospray mass spectrometry. In principle, the ability to purify and characterize intermediate products at each successive stage of construction of a protein molecule is one of the major advantages of the chemical ligation approach to total protein synthesis; it insures accurate construction of high purity protein molecules. However, in practice such purification & characterization is arduous: the consecutive chemical ligation of several peptide segments involves multiple laborious purifications carried out by reverse phase HPLC (see Scheme 4.1A). The repetitive HPLC purifications result in significant handling losses. Moreover, each of these purification steps entails time-consuming lyophilization of the product to enable solvent exchange for subsequent reactions.

In an attempt to overcome these problems, we and others have introduced methods to facilitate the handling of the intermediate products formed in the course of chemical protein synthesis. These methods include solid phase chemical ligation,\(^{[23, 24]}\) and ‘one-pot’ methods for the synthesis of proteins.\(^{[25]}\) In the solid phase approach,\(^{[23, 24]}\) unprotected peptide building blocks are consecutively ligated onto a water-compatible polymer support. After each ligation reaction, excess soluble reactants and coproducts are removed by filtration and washing. When assembly of the target sequence is complete, the product polypeptide is liberated from the polymer for analysis and purification.

\(^a\) Note that the native chemical ligation reaction itself has been shown to have undiminished effectiveness for the reaction of even very large protein molecule.\(^{[22]}\)
However, significant additional chemistry is required for the preparation of the cleavable peptide-polymer linker for each new target protein,[23, 24] Also, analytical control of solid phase chemistry is not straightforward, and significant amounts of unreacted peptide byproducts accumulated after as few as three ligation steps, even with use of a large excess of soluble peptide-thioester reactants.[23] In the one-pot synthesis approach,[25] a three-segment ligation synthesis and subsequent folding for the preparation of a protein molecule were carried out in one-pot without purification of intermediate products; only a single, final HPLC purification step was needed to produce high yields of synthetic protein of exceptional purity. The one-pot approach significantly reduced the number of manipulations required for the total synthesis of a protein molecule. However, one-pot synthesis requires a near quantitative yield for each ligation reaction, and at some as yet undefined point amounts of unreacted peptide segments (and byproducts derived from these segments) will accumulate sufficient to interfere with the effective purification of the final product.

Here, we report an alternative, more effective approach to the facilitation of chemical protein synthesis: the use of a C-terminal His$_n$ “tag”, to facilitate the isolation and handling of intermediate products formed in the course of chemical protein synthesis. The principles of His$_n$ tag-assisted chemical protein synthesis are shown on Scheme 4.1B. Synthetic case studies that are reported include the total synthesis of the model protein crambin, and the total synthesis of a 17kiloDalton modular repeat protein, ‘TPR’.

66
Results

Model 1: Crambin

We have previously developed synthetic routes to the protein crambin that involve the ligation of three unprotected peptide segments. Chemical ligation of the peptide segments and the folding of the full length polypeptide has been highly optimized\textsuperscript{[25,27]}. Thus, crambin was chosen as a suitable target with which to examine His\textsubscript{6} tag-assisted chemical protein synthesis. The synthetic Scheme is shown in Figure 4.1A.

Three unprotected peptide segments (Cram[1-15A]“thioester, [Thz\textsuperscript{16-31}]Cram“thioester and [Cys\textsuperscript{32-46}]Cram-His\textsubscript{6}) were prepared as described\textsuperscript{[27]} except that [Cys\textsuperscript{32-46}]Cram-His\textsubscript{6}“amide was made on MeBHA resin. The first ligation reaction between 6.0 \( \mu \)mol of Cram[Thz\textsuperscript{16-31}]“thioester and 4.7 \( \mu \)mol of [Cys\textsuperscript{32-46}]Cram-His\textsubscript{6} amide was carried out in 4 ml pH 7.3 100 mM phosphate buffer containing 6 M guanidinium chloride and 1 % thiophenol (analytical data shown in Figure 4.1B (i)). After completion of the ligation reaction, a very small amount of Cram[Cys\textsuperscript{32-46}]-His\textsubscript{6} remained (Figure 4.1B (ii); elution time 8 min) together with a large amount of unreacted [Thz\textsuperscript{16-31}]Cram“thioester. The pH of the solution was reduced to \( \approx \) 4 by an addition of 0.2 M methoxyamine-hydrochloride to effect the conversion of Thz-peptide (mass 3976.2 \( \pm \) 0.3 Da) to Cys-peptide (mass 3964.4 \( \pm \) 0.3 Da) (Figure 4.1B(iii)).
Figure 4.1. His<sub>6</sub> tag-assisted total chemical synthesis of crambin-His<sub>6</sub>. Numbers ((i) to (viii)) from synthetic strategy (1A) correspond to the same numbers in the chromatographic data (1B (i) to (viii)). Reactions were monitored by HPLC. The chromatographic separations were performed using a linear gradient (10-50 % of buffer B in buffer A over 20 min); buffer A = 0.1 % trifluoroacetic acid (TFA) in water; buffer B = 0.08% TFA in acetonitrile. See text for a detailed description of each panel in 1B.

For affinity purification on a Ni-agarose column, the pH of the ligation solution was adjusted to ~6.1-6.5.<sup>b</sup> After binding of His<sub>6</sub>-tagged peptides to the Ni-NTA agarose

---

<sup>b</sup> At the low pH used in the Thz to Cys conversion reaction, the peptide-His<sub>6</sub> does not adsorb to the Ni-Argarose column.
column, unreacted peptide-thioester and non-tagged coproducts were washed away with buffer. The column effluent was analyzed by HPLC to make sure that no ligated product was present (data not shown). Subsequent elution of the {product peptide}-His₆ was carried out using the ligation buffer containing 200 mM imidazole. The eluted {product peptide}-His₆ solution has all the necessary components for the next ligation step.

The second ligation reaction was initiated by the addition of 5.8 μmol of Cram[1-V15A]thioester to the mixture eluted from the Ni-column (Figure 4.1B (v)). Reaction for 20 hours resulted in formation of the second ligation product (Figure 4.1B (vi); elution time 12.5 min). The Cram[1-46]-His₆ was recovered by passage over a fresh Ni-agarose column. After adsorption to the Ni-agarose and washing, folding/disulfide formation was carried out on the column by buffer exchange to a folding buffer (pH 8, 100 mM Tris containing 2 M guanidinium chloride, 8 mM cysteine and 1 mM cystine). After thirty minutes standing in folding buffer, elution was performed with the same buffer containing 200 mM imidazole (Figure 4.1B (vii); note the purification effected by adsorption and washing). The folded, disulfide-containing protein crambin-His₆ eluted at 14.5 min on reverse phase HPLC, later than the reduced polypeptide crambin(SH)₆-His₆ at 14 min, reflecting the more hydrophobic nature of the folded protein molecule (Figure 4.1B(vii))[25]. Crambin-His₆ had a measured mass 6.1±0.2Da less than that of the reduced polypeptide, reflecting the formation of the expected three disulfide bridges in the folded protein. The synthetic crambin-His₆ protein was purified by reversed phase HPLC (Figure 4.1B(viii)), and 0.75 μmol of the His₆ tagged crambin was obtained. Overall yield was 16 %, based on the limiting C-terminal peptide Cram[1-46]-His₆.
Previously, we reported a 40% yield for the synthesis of crambin by a three segment ligation one-pot strategy(30). The relatively low yield for the synthesis of crambin by His₆ tag-assisted ligation is largely due to the less efficient folding step c (Figure 4.1B(vii))

Model 2: Modular Repeat Protein TPR

To examine the utility of His₆ tag-assisted chemical protein synthesis for the chemical ligation of more than three peptide segments, we chose a ‘repeat protein’ as a synthetic target. Repeat proteins are formed from multiple copies of a motif of between

---

c Folding of the crambin[1-46]-His₆amide polypeptide is significantly less efficient than the folding of crambin[1-46]’COO’, presumably because of the absence of the salt bridge between crambin”carboxylate and the side chain guanidium of Arg10 which is believed to be important for crambin folding and stability. With the “carboxylate, crambin folds quantitatively. We have previously observed poor folding yields (~50%) arising from the perturbation of the salt bridge by constructing a crambin”carboxamide polypeptide. We note the low folding yield is not coming from the folding in Ni-agarose resin. To show that the low folding yield is not due to on-resin folding, we prepared reduced polypeptide crambin(SH)₆-His₆ as described above. The reaction mixture was divided to two equal portions. We folded/disulfide formed and purified the half of the polypeptide product mixture in the Ni-agarose resin as described above. The other half of the polypeptide product mixture was folded/disulfide formed in solution as described in one-pot synthesis. Folding efficiencies were poor for the both conditions, and were essentially the same.
~20-40 a.a.\textsuperscript{d}. After examination of several repeat proteins, we chose to make a repeat protein based on the 34 amino acid residue peptide repeat (tetra trico peptide repeat, TPR)\textsuperscript{e}. The synthetic strategy is shown in Figure 4.2A.

**Peptide Design and Synthesis.** Our initial TPR motif target sequence was the 34 amino acid residue peptide Thz–AWYNLGNAYKQGDYDEAEYYQKALELDPNNA–“thioester. An N-terminal Thz\textsuperscript{\textdagger} was incorporated to enable the use of native chemical ligation at cysteine, with the intent that the resulting Cys residues at the ligation sites would be carboxamidomethylated to give ‘pseudo-Gln’ residues (cf. Ref.\textsuperscript{[18]}) after construction of the TPR. Boc chemistry SPPS of this initial peptide sequence gave a product with a series of -17 dalton byproducts. Subsequent analysis identified dehydration of Asn side chain as the culprit. Thus, several Asn residues were replaced by

\textsuperscript{d} There are ~20 classes of repeat proteins that are found throughout nature. Usually, the number of repeats range from 3~20. Thus, repeat proteins are good model systems in which to examine His tag-assisted modular chemical protein synthesis, because the synthesis of a repeat motif can be performed from a single peptide-thioester. Also, any desired number of ligation reactions can be performed.

\textsuperscript{e} The Regan group has successfully constructed one, two and three TPR repeats by recombinant means. In their work, the amino acid sequence of the repeat motif was selected by statistical analysis of 1837 different TPR motifs from 107 proteins. The global amino acid propensity for each position in the TPR motif was assigned. We used this propensity data to design the peptide sequences used in the current work.
the next most common amino acid in the global propensity data for that amino acid position of the TPR motif (see Footnote 6). The sequence was redesigned for synthetic convenience and was Thz-AYYNLGAYYKLDGYDEAILEYYQKALELDPDNA-"thioester (Thz-TPR[W]-"thioester). Using this sequence we were able to obtain a much improved synthetic product. We also prepared a second TPR motif peptide Thz-
AYYNLGAYYKLDGYDEAILEYYQKALELDPDNA-"thioester (Thz-TPR[Y]-"thioester). A third peptide containing the necessary C-terminal His_6 tag CAWYRLGHAYYKLDGYDEAILEYYQKALELDPDHHHHHH (Cys-TPR[W]-His_6) was also synthesized. All peptides were purified by reverse phase HPLC and characterized by electrospray mass spectrometry.

\[ \text{Thz-TPR[Y]-"thioester was prepared because a mass difference of TPR[Y] and TPR[W] enabled us to determine by mass spectrometry that each desired ligation reaction had been carried out.} \]
Figure 4.2.
Figure 4.2. continued. **His**$_6$ tag-assisted total chemical synthesis of the modular repeat protein TPR. Numbers ((i) to (x)) in the synthetic strategy (2A) correspond to the same numbers in the chromatographic data (2B (i) to (x)). Reactions were monitored by HPLC. The chromatographic separations were performed using the same gradient as in the crambin synthesis. CD spectra (2B (xi)) were recorded in the far UV wavelength region (250 nm – 190 nm) at 5°C.
Ligations and Ni-NTA agarose column purifications (Figure 4.2). Data from the ligation reactions and the Ni-column purifications are shown in Figure 4.2. The peptide segment Thz-TPR[Y]-α-thioester was reacted with Cys-TPR[W]-His$_6$ under standard ligation conditions as described above for crambin (Figure 4.2B (i)) (see Experimental for more details about reaction scale and conditions). After the reaction was complete, methoxyamine-hydrochloride was added to effect the conversion of Thz- to Cys-peptide at pH 4. After readjustment to pH 6.3, the His$_6$-tagged peptides were adsorbed onto the Ni-agarose column, and unreacted peptide-thioester and non-tagged coproducts were washed away with buffer. Elution of the Cys-{product peptide}-His$_6$ was carried out using the ligation buffer containing 200 mM imidazole. (Figure 4.2B (iii)).

The second and third ligation reactions were performed in similar fashion by the addition of Thz-TPR[W]-α-thioester and Thz-TPR[Y]-α-thioester peptides to the Cys-peptide-His$_6$ products eluted from Ni-agarose column (Figures 2B (iv) for the second ligation, and 2B (vii) for third ligation), after prior conversion of the Thz-peptide products to the Cys-peptides by the addition of 0.2 M methoxyamine-hydrochloride, and ten-fold dilution to reduce the imidazole concentration to 20 mM prior to Ni-agarose column purification as described above. The products obtained after the second ligation and Thz- to Cys- conversion are shown in Figure 4.2B(v), (vi). After completion of the third ligation (Figure 4.2B (viii)), the full length 142 residue polypeptide was carboxamidomethylated (i.e. converting Cys to ‘pseudo-Gln’ (‘ΨQ’)) by adding one volume of pH 7.3 phosphate buffer containing 0.1 M 2-bromoacetamide to the crude
ligation mixture. The desired carboxamidomethylated repeat protein Thz-TPR[Y]-ψQ-TPR[W]-ψQ-TPR[Y]-ψQ-TPR[W]-His₆ was recovered by adsorption to a Ni-agarose column and subsequent elution (Figure 4.2B (ix)). The product was further purified by preparative HPLC (Figure 4.2B (x)) and lyophilized.

The overall yield of purified repeat protein was 8%, based on the limiting C-terminal peptide Cys-TPR[W]-His₆. The TPR protein had the expected molecular mass (observed mass 16941±2 Da, calculated mass 16942.2 Da). Circular dichroism spectroscopy of the synthetic protein exhibited a typical α-helical signature in the far UV region (Figure 4.2B (xi)).

Discussion

Synthesis of crambin-His₆ illustrates the straightforward nature of the manipulations required to ligate peptide segments using His₆ tag-assisted protocols. However, the crambin synthesis made use of only three peptide segments (two ligations). Thus, we examined the broader potential of this novel technique for the modular assembly of proteins by synthesis of a repeat protein ‘TPR’ from four unprotected peptide segments.

Synthesis of TPR illustrates several important features of our novel His₆ tag-assisted synthesis methodology. First, the modular assembly of an ~17 kDa protein could be performed with simple manipulations in a matter of days. The main time savings is in the rapid affinity purification (2 hours) compared with HPLC/lyophilization (2 days) – for each ligation or other reaction in the synthesis (see Scheme 4.1). The Ni-NTA agarose column purification was significantly less laborious than preparative HPLC and
also did not require a time-consuming lyophilization step to effect solvent exchange for the next reaction. Second, the use of simple affinity column purification after each ligation allowed us to drive the reactions to completion by use of a modest (1.2-to-1.5-fold) excess of peptide-thioester reagent. Overall recoveries of synthetic proteins were good. The synthesis of TPR included: three native chemical ligation reactions; two conversion reactions of Thz-peptide to Cys-peptide; one alkylation reaction of three Cys residues; three Ni-agarose column purification steps; and, one preparative HPLC purification; for a total of 10 steps. The overall recovery of final product corresponds to ~80% yield for each reaction/handling step. The final yield of desired product (~8%) was particularly impressive because it was obtained from a synthesis carried out on an unusually small scale (only ~5 mg of limiting peptide starting material). Conventional chemical protein synthesis on such a scale would quickly result in complete loss of product because of low recovery at each step. The third advantage of His₆ tag-assisted synthesis is that analytical control by HPLC on each reaction step was readily performed because the desired products in the each step were present in the solution phase (in contrast to the situation in solid phase chemical ligation²³,²⁴).

Conclusions

The His₆ tag-assisted chemical protein synthesis method described here works and is useful, as illustrated by the total syntheses of crambin and the repeat protein TPR. The combination of native chemical ligation and simple purification on Ni-agarose columns enables facile buffer exchange to new reaction conditions, and is compatible with direct
analytical control by modern protein mass spectrometric methods. And, the reduced handling losses from affinity purifications give significantly improved overall yields.

Consecutive assembly of unprotected peptide building blocks by means of His\textsubscript{6} tag-assisted chemical ligation has enabled the straightforward preparation of a \(~17\) kDa protein from four peptide segments. We expect that the assembly of proteins from up to six peptide segments will be straightforward with His\textsubscript{6} tag-assisted consecutive chemical ligations, giving high purity products with an estimated overall yield of 3\~4 \%. If needed, a single intermediate HPLC purification can be used at an intermediate point in the synthesis. Synthesis of molecules in this size range will provide general access to the world of protein domains, the modular building blocks of function in biology, and will enable the synthesis of protein molecules of the typical size (\(~300\) amino acid residues) found in nature.

The work described here is not the final word in chemical protein synthesis. Even with His\textsubscript{6} tag-assisted methods, there are limits to the size of polypeptide chain that can be prepared by consecutive assembly of peptide segments. For a truly practical total synthesis of proteins of typical size (\(~35\) kDa) from six-to-eight peptide building blocks, a convergent synthetic approach\textsuperscript{[29]} will be needed in order to avoid unacceptably low yields and consequent ambiguity of product structures and identities.

**Methods & Experimental**

**Peptide Synthesis.** Peptides and peptide-\textsuperscript{4}thioesters were made manually using stepwise Boc chemistry ‘in situ neutralization’ solid phase peptide synthesis,\textsuperscript{[20]} on –OCH\textsubscript{2}\textsubscript{2}Pam-Resins (for free \textsuperscript{N}carboxyl peptides), on 4-methylbenzhydrylamine (MeBHA) resin (for
carboxamide peptides), or on HSCH₂CH₂CO-Leu-OCH₂-Pam-resin[26] (for thioester peptides). Products were purified by reverse phase HPLC and were characterized by electrospray mass spectrometry.

**Synthesis of His₆ Tag-Peptide** Test His₆-Leu peptides were synthesized on the Boc-Leu–OCH₂-Pam-Resins. We examined three different commercially available Boc-His’s with different imidazole side chain protecting groups. These were: Nα-Boc-Nβ-benzylloxymethyl-His [His(Bom)]; Nα-Boc-Nβ-dinitrophenyl-His [His(Dnp)]; and, Nα-Boc –Nβ-tosyl-His [His(Tos)]. Each crude peptide, HHHHHHHL, from HF cleavage was analyzed by LC-MS. Analytical reverse phase HPLC data showed that peptides made using His(Bom) and His(Dnp) were comparably good, but that the peptide product from the use of His(Tos) had very little of the desired product. The product peptide made using His(Dnp) was obtained as the side chain protected form, (His(Dnp))₆L, because the Nβ-Dnp group is stable to HF cleavage. This causes the product peptide to be hydrophobic and the peptide elutes much later in HPLC analysis. We conjectured that the hydrophobic nature of a (His(Dnp))₆ tag would cause solubility problems for a peptide-(His(Dnp))₆. In contrast, the HHHHHHHL made using His(Bom) turned out to be side chain unprotected and hydrophilic, so we expected that if anything, this His₆-tag would improve the solubility of hydrophobic peptides.

For this reason, we used His(Bom) for the synthesis of the His₆-tag. However, SPPS of a Cys-peptide-His₆ results in a product consisting almost exclusively of the target peptide plus a mass of 12Da. Further analysis reminded us that the +12Da product was caused by conversion of the N-terminal Cys to a 1,3-thiazolidine-4-carboxyl (Thz) moiety, by formaldehyde generated during HF cleavage of a peptide containing the

79
N^{3}-benzyloxymethyl group. Thz-HHHHHH peptide was converted to Cys-HHHHHH by adding 0.2 M methoxyamine.HCl to the aqueous/50 % acetonitrile solution of the peptide from the work-up procedure following HF cleavage. Conversion of Thz- to Cys- peptide was essentially complete in one day and the solution was directly injected to the preparative HPLC system for the purification of the peptide-His_6. Also, use of 0.2 M cysteine.HCl as a scavenger helped minimize the conversion of Cys-peptide to Thz-peptide during HF cleavage.

**Preparation of Ni-NTA Agarose Resin for adsorption of His_6 tagged peptides** Prior to adsorption of the {product peptide}-His_6 the Ni-NTA agarose resin was equilibrated with ligation buffer (phosphate buffer at pH 7.3 containing 6 M guanidinium chloride and 0.5 % thiophenol). The same buffer was used for washing to remove non-tagged coproducts.

**His_6 Tag-Assisted Chemical Protein Synthesis.** The His_6-tag was attached to the C-terminal of the first (i.e. C-terminal) peptide segment. The peptide-His_6 construct was prepared either by stepwise assembly of the peptide on the His_6-Resin, followed by cleavage/deprotection, or by native chemical ligation of a peptide-“thioester with Cys-His_6 to give peptide-Cys-His_6. Subsequent ligation reactions were carried out under standard conditions on the His_6-tagged peptide (Scheme 4.1B). After each ligation reaction, the {product peptide}-His_6 was recovered from excess reactant and non-tagged co-products by adsorption onto a Ni-NTA agarose column, followed by thorough washing. Where appropriate, the {product peptide}-His_6 was eluted in ligation buffer containing 200 mM imidazole, for use in the next ligation reaction. Prior to the next Ni-column purification, we diluted the imidazole concentration of 200 mM to 20 mM.
Approximately 1 milliliter of Ni-NTA agarose\(^8\) resin was used per ~5 mg of ligated peptide. (see Experimental for the Chemical Synthesis of TPR for details of experimental procedures). At any desired stage of the synthesis, aliquots can be removed for analytical control by LC-MS.

**Protein Folding/Disulfide Formation** While adsorbed to the Ni-agarose support, the final full-length reduced crambin polypeptide was folded with concomitant formation of disulfides. After adsorption, the support was washed with folding buffer (pH 8 Tris buffer containing 2 M guanidinium chloride, 8 mM cysteine and 1 mM cystine) to remove excess reactants and non-tagged co-products. The \{crambin(SH)\(n\)-His\(n\)-Ni-agarose equilibrated in folding buffer was allowed to stand for 30 min at room temperature, after which the folded protein \{crambin-His\(n\}\ was eluted with the same buffer containing 200 mM imidazole.

**Experimental Details for the Chemical Synthesis of TPR** First ligation reaction: 1.1 μmol of Thz-TPR[\(W\)-His\(n\] (5.5 mg) and 1.3 μmol of Thz-TPR[\(Y\)-\(thioester\] (5.5 mg) were mixed in 3 ml of ligation buffer (phosphate buffer at pH 7.3 containing 6 M guanidinium hydrochloride and 1 % thiophenol (v/v)) (See Figure 4.2B (i) in the manuscript for \(t=0\) analytical data). After completion of the ligation reaction (Figure 4.2B (ii)), conversion of the ligated Thz-peptide-His\(n\) to Cys-peptide-His\(n\) was carried out by an addition of 0.2 M methoxyamine.hydrochloride to the ligation mixture, at pH 4. The ligation mixture was re-adjusted to pH 6.3 prior to Ni-NTA agarose column

\(^8\) Binding capacity of Ni-NTA agarose (Qiagen, CA) is 5-10 mg (0.3-0.5 μmol for ~20 kDa protein) per milliliter of swollen support.
purification. The Ni-column was pre-equilibrated as described above. After binding of His-tagged peptides to the Ni-agarose column, unreacted peptide-thioester and non-tagged co-products were washed away with buffer. The ligated peptide Thz-TPR[Y]-Cys-TPR[W]-His₉ was eluted with 3 ml of elution buffer (phosphate buffer at pH 7.3 containing 200 mM imidazole, 6 M guanidinium hydrochloride and 1 % thiophenol (v/v)) (Figure 4.2B (iii)).

The second ligation was initiated by adding 1.1 µmol of Thz-TPR[W]-“thioester (4.7 mg) to the mixture eluted from Ni-column (Figure 4.2B (iv)). The Thz-TPR[W]“thioester co-eluted with Thz-TPR[Y]-Cys-TPR[W]-His₉ in the analytical HPLC profile; the co-elution was verified by LC-MS. Ligated Thz-peptide-His₉ (Figure 4.2B (v)) was converted to Cys-peptide-His₉ at pH 4 by addition of 0.2 M methoxyamine-hydrochloride, and the mixture was diluted 10 fold to reduce the imidazole concentration from 200 mM to 20 mM prior to the Ni-column purification. The Ni-column purification procedure was the same as described above. The desired peptide Thz-TPR[W]-Cys-TPR[Y]-Cys-TPR[W]-His₉ was eluted from the affinity column (Figure 4.2B (vi)), and its identity verified by LC-MS.

The third ligation was performed by addition of 1 µmol of Thz-TPR[Y]-“thioester (4.2 mg) to the elution mixture (Figure 4.2B (vii)). After completion of the ligation reaction (Figure 4.2B (viii)), carboxamidomethylation reaction (Cys to ψQ) was carried out by adding one volume of phosphate buffer at pH 7.3 containing 0.1M 2-bromoacetamide and 10mM β-mercaptoethanol to the crude ligation mixture. The desired carboxamidomethylated TPR, Thz-TPR[Y]-ψQ-TPR[W]-ψQ-TPR[Y]-ψQ-TPR[W]-His₉, was purified by affinity chromatography as described above and eluted from Ni-column.
(Figure 4.2B (ix)). The product was further purified by preparative reverse HPLC and lyophilized to give 1.5 mg (0.09 µmol) of purified product.

**Thermal measurements.** For thermal denaturation experiments, the ellipticity was monitored at 222 nm while the temperature was increased from 5°C to 100°C.

**Circular Dichroism Spectroscopy and Thermal Denaturation**

CD spectra were recorded in the far UV wavelength region (250 nm – 190 nm) at 5°C. The sample was 10 µM protein, 10 mM phosphate (pH 8.0) in a 1 mm path length cuvette. For thermal denaturation experiments, the ellipticity was monitored at 222 nm while the temperature was increased from 5°C to 100°C.

**Preparative reverse phase HPLC purification.** Synthetic peptides and ligation products were purified by semi-prep HPLC on a 1cmx25cm Vydac C4 column using a gradient from 20-50% buffer B over 60 minutes at a flow rate of 5ml per minute. Buffer A: 0.1% TFA in water; BufferB: 0.08% TFA in acetonitrile. Fractions were collected across the expected elution time and combined based on ESMS analysis.

**References**


83


Chapter 5. Chemical Protein Synthesis by Convergent Ligation of Unprotected Peptide segments

Abstract. We accomplished a practical convergent ligation of multiple unprotected peptide segments through a combination of improved synthetic approaches. We believed that the convergent ligation of fully unprotected peptides must meet two requirements. First, ligation of two peptides, \{Thz-peptide2-“thioester”\} and \{Cys-peptide3-“thioester”\} must give our desired product, \{Thz-peptide2-peptide3-“thioester”\}. The product peptide can be used for the ligation with a second \{Cys-peptide\}. Second, in addition to the first step, if we are able to convert the \{Thz-peptide2-peptide3-“thioester”\} to \{Cys-peptide2-peptide3-“thioester”\}, the resulting \{Cys-peptide2-peptide3-“thioester”\} would be suitable for the ligation, in the other direction, with a second \{peptide-“thioester”\}. A major challenge for convergent ligation is the intrinsic reactivity of the species \{Cys-peptide3-“thioester”\}; the attempted reaction with \{peptide2-thioester\} will lead to the formation of not only the desired \{peptide2-peptide3-“thioester”\} but also \{cyclic peptide3\}, \{peptide3-peptide3-“thioester”\}, \{peptide2-peptide3-peptide3-“thioester”\} as major products. Series of synthetic strategy were developed to overcome these challenges. Finally, we investigated the reactivity of \{peptide-“thiophenylester”\} with a \{Cys-peptide\} compared to the reactivity of \{peptide-“thioester”\} with \{Cys-peptide\}, in the absence of thiophenol in the ligation mixture. A remarkable difference of the ligation rates was observed. The efficacy of this kinetically controlled convergent ligation approach was vividly demonstrated by the great efficiency in each ligation. The robustness of our
strategy was illustrated by the convergent synthesis of crambin from six peptide segments.
Proteins are the key natural products involved in every life process. Current advances in chemical protein synthesis are making a significant contribution to the atom-by-atom manipulation of proteins by chemistry, and to understanding the properties of protein molecules. Most of all, the preparation of proteins via native chemical ligation of two unprotected peptide segments or via sequential native chemical ligation of more than two peptide segments has dramatically increased the level of access to a wide range of synthetic proteins.

Convergent chemical ligation (schematically shown in Figure 5.1) is a major challenge to be solved for the chemical synthesis of proteins. A convergent approach to the synthesis of proteins has clear advantages including facile access to chemical protein analogues, synthesis of highly homogeneous protein constructs, improved yields, time savings, and application to large target proteins. However, current state of the art synthetic protein chemistry does not make use of fully convergent ligation methods for the synthesis of full-length polypeptides, a precursor to the functional protein molecules. Even though convergent synthesis of fully unprotected peptides has been reported, these works used two different ligation chemistries (usually native chemical ligation in conjunction with unnatural non-peptide bond-forming ligation chemistry) each of which can be performed with impunity in the presence of the mutually reactive functionalities for the other ligation chemistry. Thus, the synthetic products have at least one unnatural moiety at the ligation site.
Figure 5.1. Convergent chemical ligation of a protein molecule. Each bar represents a peptide fragment.

We anticipate that the convergent ligation of six to ten unprotected peptide segments will be essential for the routine access to proteins of typical size (300 amino acids), for the chemical synthesis of glycosylated proteins, and for the coupling of multiple expressed proteins with chemically synthesized peptides. Here we report a highly practical strategy that enables a fully convergent synthesis of a target protein by a novel principle in total protein synthesis – ‘kinetically controlled convergent ligation’.

Key steps to the convergent ligation of peptide segments are illustrated in Figure 5.2. For the joining of two unprotected peptide segments, we use the most practical ligation chemistry, native chemical ligation.[5] We believe that the convergent ligation of fully unprotected peptides must meet two requirements. First, ligation of two peptides,
{Thz-peptide2-“thioester”} and {Cys-peptide3-“thioester”} must give our desired product, {Thz-peptide2-peptide3-“thioester”}. The product peptide can be used for the ligation with the other {Cys-peptide}. Second, in addition to the first step, if we are able to convert the {Thz-peptide2-peptide3-“thioester”} to {Cys-peptide2-peptide3-“thioester”}, the resulting {Cys-peptide2-peptide3-“thioester”} would be suitable for the ligation, in the other direction, with another {peptide-“thioester”}. Accomplishing both the first and the second requirement will enable the convergent synthesis of proteins.

A major challenge for convergent chemical ligation is the intrinsic reactivity of the species {Cys-peptide3-“thioester”}; the attempted reaction with {peptide2-thioester} will lead to the formation of not only the desired {peptide2-peptide3-“thioester”} but also {cyclic peptide3}, {peptide3-peptide3-“thioester”}, {peptide2-peptide3-peptide3-“thioester”} as major products.

![Diagram of peptide ligation](image-url)
**Figure 5.2.** Key steps for the realization of convergent synthesis of proteins. Ligation of {Thz-peptide2-“thioester”} and {Cys-peptide3-“thioester”} should give only the desired product, {Thz-peptide2-peptide3-“thioester”}. The product peptide can be used for the ligation with another {Cys-peptide}. Second, conversion of the {Thz-peptide2-peptide3-“thioester”} to {Cys-peptide2-peptide3-“thioester”} should be easily performed. The resulting {Cys-peptide2-peptide3-“thioester”} is suitable for the ligation with another {peptide-“thioester”} from the C- to N- direction.

---

**Kinetic control of product formation by titration.** To suppress the formation of the undesired by-products, we devised a synthetic scheme in which a high concentration of {peptide1-“thioester”} under ligation conditions is titrated with a solution of {Cys-peptide2-“thioester”} (as shown in Figure 5.3) During the titration, a high concentration of {peptide1-“thioester”} is dissolved in pH 6.7, 0.5M Bis-Tris (Bis(2-hydroxyethyl)iminotris(hydroxymethyl)methane) aqueous buffer containing 6M guanidinium hydrochloride and 1% thiophenol (vol/vol). Also, {Cys-peptide2-“thioester”} is maintained in unbuffered aqueous 6M guanidinium hydrochloride (pH 4~5) to prevent any reaction before its addition to the (peptide1-“thioester”)-containing ligation reaction buffer. Ligation of the two peptides by titration was controlled to be finished in several hours.

In this synthetic strategy, (i) we used the high dependence \(^{12}\) of the rate of the native chemical ligation reaction on the identity of the C-terminal residue (Xxx) of peptide-“thioester” {i.e. peptide-Xxx-“thioester”}. For example, the ligation between
{peptide1-Gly-"thioester"} and {Cys-peptide2-Thr-"thioester"} would exclusively give {peptide1-Gly-Cys-peptide2-Thr-"thioester"} because a Cys-peptide reacts with a peptidyl-Gly-"thioester" many times faster than with a peptidyl-Thr-"thioester"; (ii) during the titration, {peptide1-thioester} is exchanged with thiophenol to form an activated peptide-thiophenylester, to favor fast reaction with {Cys-peptide2-"thioester"} before the intramolecular cyclization of {Cys-peptide2-"thioester"}; (iii) peptides are maintained in a high concentration but {Cys-peptide2-"thioester"} is slowly titrate to the reaction mixture to prevent homo-ligation (i.e. ligation between two {Cys-peptide2-"thioester"}).

**Figure 5.3.** Our devised scheme for the native chemical ligation of two peptide thioesters that prevents the formation of the undesired by-products. During the titration, high
concentration of {peptide1-“thioester”} is dissolved in aqueous buffer containing 6M guanidinium hydrochloride and 1% thiophenol (vol/vol), and the ligation mixture was stirred. Also, {Cys-peptide2-“thioester”} is maintained at low pH in unbuffered aqueous solution containing 6M guanidinium hydrochloride to prevent any reaction before its titration (i.e. slow addition) into the ligation mixture. The titration rate was controlled to be finished in several hours.

A model kinetically controlled ligation was performed in this fashion, with controlled titration of a {Cys-peptide-“thioester”} into the reaction mixture containing a high concentration of peptide-thioester in the presence of thiophenol, and the data were largely consistent with our proposal. For example, a model ligation between {peptide1-Gly-“thioester”} with {Cys-peptide2-Thr-“thioester”} using the titration strategy exclusively gave the desired {peptide1-Gly-Cys-peptide2-Thr-“thioester”} product. Also, a model ligation between {peptide1-Ala-“thioester”} and {Cys-peptide2-Leu-“thioester”} with the titration strategy produced the desired {peptide1-Ala-Cys-peptide2-Leu-“thioester”} as the major product as shown in Figure 5.4. The suppression of {cyclic peptide2} and {peptide2-peptide2-“thioester”} was highly effective. This ligation was particularly impressive because the reaction rate of peptidyl-Ala-“thioester and peptidyl-Leu-“thioester ligation with a Cys-peptide are similar to each other, so that the formation of the desired product as the major product would be very difficult without our titration strategy.

However, our titration approach required two unfavorable aspects for the ligation of unprotected peptides. Obtaining a high concentration of the first unprotected peptide-
thioester (20–50 mM for the reaction event of Figure 5.4) would not be always possible for aggregating peptides. Second, after a certain amount of time, the desired product, {peptide1-peptide2-‘thioester}, also underwent transthioesterification with thiophenol to form {peptide1-peptide2-‘thiophenylester}. This peptide became reactive for the ligation with the {Cys-peptide2-‘thioester} being titrated into the reaction mixture to form a significant amount of {peptide1-peptide2-peptide2-‘thioester} as shown in Figure 5.4. Thus, the use of excess (1.5 to 2 times) of {peptide1-‘thioester} was necessary to suppress formation of {peptide1-peptide2-peptide2-‘thioester}. The titration tactic was practically useful. However, we believed that these two undesirable features would limit the routine use of the titration tactic in a fully convergent synthetic strategy. Further refinement of our strategy was necessary to enable convergent chemical ligation protein synthesis in a practical manner.

**Figure 5.4** Titration strategy applied model to a model ligation. Two separate peptide solutions were prepared. A {peptide1-Ala-‘thioester} (Thz-PSIVARSNFNA-‘thioester)
(9µmol) was dissolved in 0.5ml of pH 6.7, 0.5 M Bis-Tris buffer containing 6M guanidinium hydrochloride and 1% (v/v) thiophenol (18mM peptide concentration). A {Cys-peptide2-“thioester} (Cys-RLPGTPEAL-“thioester) (10µmol) is maintained in 0.2ml water containing 6M guanidinium hydrochloride (pH 4~5) to prevent any ligation reaction before its titration (50mM peptide concentration). The {Cys-peptide2-“thioester} solution was slowly added to the {peptide1-Ala-“thioester} solution mixture over 4 hours. The ligation products spontaneously formed a peptide-“thiolactone by transthioesterification with the internal cysteine residues at the ligation site. Thus, the resulting mixture was treated with 200mM sodium 2-mercaptopethanesulfonate (MES-Na) to convert the peptide-“thiolactone to a peptide-“thioester. Desired product, Thz-PSIVARSNFNA-Cys-RLPGTPEAL-“thio(ethanesulfonate)ester was eluted at 9.2min. (observed mass of 252.5±0.9 Da. calculated mass = 2451.7 Da). Undesired product, Thz-PSIVARSNFNA-Cys-RLPGTPEAL-Cys-RLPGTPEAL-“thio(ethanesulfonate)ester was eluted at 10.3min. (observed mass = 3490.6±0.9 Da. calculated mass = 3489.9 Da). The chromatographic separations were performed using a linear gradient (5-65%) of buffer B in buffer A over 15 min with a flow rate of 0.5 ml/min (buffer A = 0.1% TFA in water; buffer B = 0.08 % TFA in acetonitrile).

**Kinetic control of product formation based on the nature of the thioester.** We suspected that, in the absence of thiophenol in the ligation mixture, the reactivity of a {peptide-“thiophenylester} with a Cys-peptide would be much greater than the reactivity of a standard {peptide-“thioester}. We routinely prepare relatively unreactive peptide“-SCH₂CH₂CO-Leu, i.e. alkyl thioesters, for use in native chemical ligation in the presence
of thiophenol, according to Dawson. If we can prepare the peptide-"thiophenylester" prior to the ligation reaction, reaction of this more activated peptide-"thiophenylester with another peptide-"thioester" was anticipated to give our desired product of intermolecular reaction. In the absence of thiophenol additive in the ligation mixture, we expected that a peptide-"thioester" would be much less reactive compared to the peptide-"thiophenylester.

In the event, a remarkable result was achieved by the introduction of preformed peptide1-"thiophenylester" prepared prior to the ligation reaction (as shown in Figure 5.5). The peptide1-"thiophenylester" was obtained by exchanging a ca. 0.2mM crude peptide1-"thioester" with 0.2 % thiophenol (vol/vol) in pH 6.8 aqueous buffer containing 6 M guanidinium hydrochloride. Then, the converted peptide-"thiophenylester was purified by use of preparative HPLC. Procedures for the quantitative conversion and the purification are described in Experimental.

**Figure 5.5.** A further revised scheme for kinetically controlled convergent chemical ligation. Reaction between the preformed peptide1-"thiophenylester" and Cys-peptide2-"thioester" under normal native chemical ligation conditions (aqueous solution,
neutral pH) except in the absence of any thiophenol additive; only the desired ligation product \{peptide1-Cys-peptide2-thioester\} was formed.

In a model study, in separate reactions, we compared the rates of reaction of the two peptide-thioesters, \{peptide1-Ala-thiophenylester\} and \{peptide1-Ala-thioester\}, with a \{Cys-peptide\} in pH 6.8 aqueous buffer without any thiol additives. In this model ligation reaction, we observed a large difference for the ligation rates of the two thioesters, as shown in Figure 5.6. We anticipated that the rate difference would make the standard -thio as a ‘kinetic synthon’ as compared to -thiophenylester; that is, it would be effectively unreactive under these modified reaction conditions where the -thiophenylester would react rapidly and quantitatively.

![Graph showing ligation rates over time](image)

**Figure 5.6.** Comparison of the ligation rates for the reaction of either \{peptide1-Ala-thiophenylester\} (solid line) or \{peptide1-Ala-thioester\} (dashed line) with a \{Cys-peptide\} under identical reaction conditions. Two separate ligation reactions between
1mM concentrations of the peptides were carried out in 200 mM phosphate buffer, pH 6.8, containing 6 M guanidinium hydrochloride. No thiol additive was present during the ligation reactions. See **Experimental** for more details.

Motivated by the discovery of this large rate difference, we performed a trial ligation reaction, with the same sequence peptides (\{peptide1-Ala-’thiophenylester\} and \{Cys-peptide2-Leu-’thioester\}) used in **Figure 5.4**, in pH 6.6 aqueous buffer in the absence of thiophenol additive (see **Figure 5.8A & B**). The ligation reaction gave exclusively the desired ligation product \{peptide1-Ala-Cys-peptide2-Leu-’thioester\} in essentially quantitative yield in one hour. The suppression of all the possible undesired by-products was highly effective. As well, the reaction was performed at a normal native chemical ligation concentration of 2mM of each peptide; high concentrations of peptide reactants are not required. This high efficiency of differential reaction obviated the use of the titration strategy, even though the titration could be introduced again whenever its use is found to be necessary.

**Fully convergent chemical synthesis of crambin.** To show the robustness of a convergent chemical protein synthesis strategy based on this refined kinetically controlled chemical ligation approach, we designed a fully convergent synthesis of the model protein crambin \[^{9, 11, 15-17}\] by ligation of **six unprotected peptide segments**. The synthetic design and sequence of the target molecule are shown in **Figure 5.7**. In the synthetic design, three peptide segments from N-terminus would be joined from C- to N-direction. By performing these ligations, we wanted to examine how efficiently the
peptide prepared by the first ligation (assigned as ligation #1 in the Figure 5.7) can be used for the next set of convergent ligation (assigned as ligation #3 in the Figure 5.7).

The other three peptides were designed to be ligated in the N- to C- direction. Ligation in the N- to C- direction requires no protecting group for the N-terminal cysteine of the middle peptide segment for the first ligation (assigned as ligation #2 in the Figure 5.7). Thus, the ligation of three peptides is possible to be carried out in ‘one-pot’. Finally, the two peptides representing the two halves of the target polypeptide each prepared by a convergent route, were to be joined by a native chemical ligation reaction, carried out under normal conditions (assigned as ligation #5 in the Figure 5.7). Without purification, the product full length polypeptide could be folded to give native crambin by adjusting concentration of denaturant and redox reagents.

Figure 5.7. Total chemical synthesis of the model protein crambin by the fully convergent ligation of six unprotected peptide segments. Three peptide segments from N-terminus are joined from C- to N- direction. Ligation of the other three peptides are carried out from N- to C- direction. Finally, the two peptides prepared by convergent
routes are joined under normal ligation conditions (assigned as ligation #5), and the product full length polypeptide is folded, with formation of disulfides to give native crambin, by adjusting concentrations of the denaturant and the redox reagents. Recovered yield from each step in the synthesis is described in the parenthesis (Detail for each reaction will be describe in the text).

We prepared six unprotected peptide segments by manual solid phase syntheses. These peptides are [Thr\(^1\)-Cys\(^3\)]-\(^\text{thiophenylester} [\text{Thz}\(^2\)-Ala\(^{15}\)]-\(^\text{thiophenylester}, [\text{Cys}\(^{16}\)-Leu\(^{25}\)]-\(^\text{thioester}, [\text{Thz}\(^{26}\)-Gly\(^{31}\)]-\(^\text{thiophenylester}, [\text{Cys}\(^{32}\)-Thr\(^{39}\)]-\(^\text{thioester}, and [\text{Cys}\(^{40}\)-Asn\(^{46}\)]. We used the 1,3-thiazolidine-4-carboxo- (Thz) group to protect the N-terminal Cys of peptide segments #2 and #4. Each peptide-\(^\text{thiophenylester} was prepared prior to the ligation reaction. The solid phase peptide synthesis and the preparation of peptide-\(^\text{thiophenylester} are described in the Experimental section.

Ligation #1, the reaction of peptide segments #2 and #3, [Thz\(^4\)-Ala\(^{15}\)]-\(^\text{thiophenylester} and [Cys\(^{16}\)-Leu\(^{25}\)]-\(^\text{thioester} (shown in Figure 5.8) vividly illustrate the robustness of our kinetically controlled ligation based on differential reactivity of the two peptide-\(^\text{thioesters}. The ligation of [Thz\(^4\)-Ala\(^{15}\)]-\(^\text{thiophenylester} (2mM) and [Cys\(^{16}\)-Leu\(^{25}\)]-\(^\text{thioester} (2.1mM) was carried out in 5 ml of 200mM phosphate buffer containing 6M guanidinium hydrochloride at pH 6.8. Despite the similar reactivity of Ala-\(^\text{thioester} and Leu-\(^\text{thioester} for the ligation with Cys, the much greater reactivity of the thiophenylester exclusively gave a single product, the desired [Thz\(^4\)-Leu\(^{25}\)]-\(^\text{thioester}, and the reaction was completed in one hour (Figure 5.8B). Furthermore, quantitative conversion of Thz ([Thz\(^4\)-Leu\(^{25}\)]-\(^\text{thioester}) to Cys ([Cys\(^4\)-Leu\(^{25}\)]-\(^\text{thioester}) was achieved 100
by adding 0.2 M methoxyamine·hydrochloride (pH = 3.8 in the reaction mixture) without affecting thioester moiety, and the reaction was completed in two hours (**Figure 5.8C**). A preparative HPLC was used to purify the [Cys\(^4\)-Leu\(^{25}\)]-“thioester, and a recovered yield from this purification was 71%, based on the starting peptide segments.

**Figure 5.8.** Ligation of [Thz\(^4\)-Ala\(^{15}\)]-“thiophenylester and [Cys\(^{16}\)-Leu\(^{25}\)]-“thioester, followed by conversion of the [Thz\(^4\)-Leu\(^{25}\)]-“thioester peptide to the [Cys\(^4\)-Leu\(^{25}\)]-“thioester peptide. **8A** shows time zero point. Our ligation exclusively gave [Thz\(^4\)-Leu\(^{25}\)]-“thioester, and the reaction was completed in one hour (**8B**). Observed mass of the
product peak (2529.8±0.4 Da) was consistent with calculated mass (2530.0 Da). Arrow (i) indicates the cyclized [Cys16-Leu25]−“thioester (cyclic-peptide#3, observed mass = 1137.7±0.6 Da. calculated mass = 1138.3 Da.). Arrow (ii) indicates the [Thz2-Leu25]-[Cys16-Leu25]−“thioester (peptide#2-peptide#3-peptide#3−“thioester, observed mass = 3567.6±0.9 Da. calculated mass = 3568.2 Da.) Using 0.2M methoxyamine.HCl, quantitative conversion of Thz ([Thz2-Leu25]−“thioester) to Cys ([Cys4-Leu23]−“thioester) was carried out without affecting the thioester moiety, and the reaction was completed in two hours (8C). Observed mass of 2517.8±0.3 Da. was consistent with calculated mass = 2518.0 Da for the peptide. The chromatographic separations were performed using a linear gradient (1-61%) of buffer B in buffer A over 15 min with a flow rate of 0.5 ml/min (buffer A = 0.1% TFA in water; buffer B = 0.08 % TFA in acetonitrile).

The crambin synthesis by a fully convergent route has two more ligation steps that require the use of our kinetically controlled ligation. These are the ligation of [Thr1-Cys3]−“thiophenylester and [Cys4-Leu25]−“thioester (i.e. ligation #3), and the ligation of [Thz26-Gly31]−“thiophenylester and [Cys32-Thr39]−“thioester (i.e. ligation #2). Each kinetically controlled ligation exclusively gave the desired product (Figure 5.9). The ligation between [Thr1-Cys3]−“thiophenylester and [Cys4-Leu25]−“thioester was completed in one hour. The ligation product spontaneously formed a peptide−“thiolactone by transthioesterification with internal cysteine residues. An addition of 200mM sodium 2-mercaptoethanesulfonate (MES-Na) into the reaction mixture converted the peptide−“thiolactone to a peptide−“thioester shown in a top panel of Figure 5.9. We purified the product [Thz1-Leu25]−“thioester by use of preparative HPLC with a 50 % recovered yield.
The ligation between [Thz\textsuperscript{26}-Gly\textsuperscript{31}]-“thiophenylester and [Cys\textsuperscript{32}-Thr\textsuperscript{39}]-“thioester was completed in one hour as shown in the bottom panel of Figure 5.9. In preliminary studies, the [Thz\textsuperscript{26}-Thr\textsuperscript{39}]-“thioester was purified by use of preparative HPLC with an 80 % recovered yield. However, the product peptide, [Thz\textsuperscript{26}-Thr\textsuperscript{39}]-“thioester, could be ligated with [Cys\textsuperscript{40}-Asn\textsuperscript{46}] simply by adding the C-terminal peptide to the ligation mixture, along with thiopoeol as catalyst, without an intermediate purification step. The resulting peptide [Thz\textsuperscript{26}-Asn\textsuperscript{46}] was converted to [Cys\textsuperscript{26}-Asn\textsuperscript{46}] by adding 0.2 M methoxyamine•hydrochloride (pH =3.8 in the reaction mixture), and the reaction was completed in two hours (data is not shown). This one-pot ligation mixture was purified by use of preparative HPLC to give product [Cys\textsuperscript{26}-Asn\textsuperscript{46}] with an overall 40 % recovered yield.

Figure 5.9. Crude product mixtures from kinetically controlled ligation reactions. Figure 5.9A shows the crude ligation product from ligation #3, i.e. reaction of [Thr\textsuperscript{1}-Cys\textsuperscript{3}]-
“thiophenylester and [Cys⁴-Leu²⁵]-“thioester. The ligation product formed a peptide-
“thiolactone from the transthioesterification with internal cysteine residues. An addition
of 200mM MES-Na into the reaction mixture converted the peptide-“thiolactone to the
peptide-“thioester shown, [Thz¹-Leu²⁵]-“thio(ethanesulfonate)ester, (Observed mass =
2746±1.0 Da.; calculated mass = 2747 Da.). Figure 5.9B shows the ligation products
obtained from the reaction of [Thz²⁶-Gly³¹]-“thiophenylester and [Cys³²-Thr³⁹]-“thioester
(i.e. ligation #2), (Observed mass = 1596.5±1.0 Da.; calculated mass = 1596.6 Da.). The
chromatographic separations were performed using the same gradient as Figure 5.8.

The final ligation between [Thr¹-Leu²⁵]-“thioester and [Cys²⁶-Asn⁴⁶] was carried
out under normal native chemical ligation conditions, pH6.8, 200mM phosphate buffer
containing 6M guanidinium hydrochloride and 1% thiophenol. An analytical HPLC
chromatogram of the reaction at starting point before the addition of thiophenol is shown
in Figure 5.10A. An overnight ligation reaction gave a quantitative yield of the full-
length crambin polypeptide. We then added five volumes of pH 8, 200M phosphate
buffer containing 8mM cysteine and 1mM cystine to dilute the guanidinium
hydrochloride to 1M and to adjust the pH to 7.8, conditions previously shown to be
optimal for the folding of crambin and for the formation of the native disulfides. Folding
and disulfide formation (Figure 5.10B) gave an essentially quantitative yield of the
correctly folded crambin. The folded crambin molecule was purified directly by
preparative HPLC with 62% recovered yield based on the two starting peptide segments
(Figure 5.10C).
**Figure 5.10.** Final ligation between [Thr\(^1\)-Leu\(^{25}\)]-“thioester and [Cys\(^{36}\)-Asn\(^{46}\)]. **Figure 5.10A** shows a starting point before the addition of thiophenol. An overnight reaction under standard native chemical ligation condition gave a full-length crambin polypeptide. **Figure 5.10B** shows the folding and disulfide formation from the full-length polypeptide ligation product. In one hour, formation of the folded crambin molecule was essentially complete. **Figure 5.10C** shows the folded crambin molecule purified directly by preparative HPLC (observed mass = 4702.0 ±0.8 Da. calculated mass = 4702.4 Da.).
Summary & significance. We accomplished a practical convergent ligation of multiple unprotected peptide segments through a series of improved approaches. Prior to our research, the use of \{Cys-peptide2-\"thioester\} was not considered for the convergent synthesis because of the intrinsic reactivity of \{Cys-peptide2-\"thioester\} to form not only \{peptide1-peptide2-\"thioester\} but also \{cyclic peptide2\}, \{peptide2-peptide2-\"thioester\}, \{peptide1-peptide2-peptide2-\"thioester\}. Initially, we wanted to take advantage of the high dependence of native chemical ligation rates on the C-terminal residue of a peptide-\"thioester. However, the rate dependence also meant an intrinsic limitation for the general usage of this concept for the chemical protein synthesis by a convergent route. To avoid the limitation, we developed another scheme that made use of the titration of \{peptide1-\"thioester\} with \{Cys-peptide2-\"thioester\}. This scheme improved our synthesis with a limited success as described above. Finally, we investigated the reactivity of \{peptide-\"thiophenylester\} with Cys compared to the reactivity of \{peptide-\"thioester\} with Cys, in the absence of thiophenol in the ligation mixture. A remarkable difference of the ligation rate was observed. The kinetically controlled ligation was vividly demonstrated by the great efficiency in each ligation. The robustness of our strategy was illustrated by the convergent synthesis of crambin as we described above.

In addition to the great efficiency of our approach, we want to note following: (1) our approach is very practical. It uses easily prepared fully unprotected Cys-peptides and peptide-\"thioesters. Modulation of ligation reactivity was simply achieved by converting one thioester to another by adding different form of thiol additivies (to prepare activated form of thioester from an unactivated form of thioester); (2) Our strategy showed a
possibility for various approaches for synthesis of one protein. It means that our approach

can be used for the peptide segment ligation from N- to C- ligation direction as well as
from C- to N- direction; (3) our synthesis would be compatible with a cleavable auxiliary
for non-Cys ligation [18], or with an alkylation reaction [7] to produce pseudo-Asn/Gln (or
–Asp/Glu) after ligation of middle peptide segments. Our strategy has potential
applications for regioselective conversion of Cys to a desired modified amino acid, in the
presence of many other Cys’s in the target protein molecule.

We believe that our convergent strategy based on kinetically controlled chemical
ligation will have substantial impact for the chemical manipulation of proteins. First,
proteins of typical size will be realized by use of this method. Currently, proteins of size
of ~100-150 amino acids are being routinely produced [2]. A convergent joining of two or
three peptide segments of this size (~100-150 amino acids) can lead to the synthesis of
the proteins of typical size. Second, our strategy would fit for the total synthesis of
glycosylated proteins. Sugar moiety(s) introduced onto short peptides can be joined with
other peptide segments at any point during the synthesis of a glycosylated protein [14].
This approach will be efficient for the saving of the glycosylated peptides that are usually
prepared with significant efforts. Third, we can imagine that our strategy can be adapted
to the proteins made by convergent ligation of expressed polypeptides [3] and chemically
synthesized peptides. This would give a versatile way to produce proteins of any size that
can be chemically mutated at any residues.

In conclusion, kinetically controlled ligation is a general and effective method to
realize fully convergent synthesis of proteins. We believe our strategy will be a
cornerstone for the access to proteins with great efficiency and for the realization for the atomic control of the proteins by chemistry.

Experimental

**Peptide Segment Synthesis (peptide-"carboxylate or peptide-"thioester)**

Peptides were prepared manually by “in situ neutralization” Boc chemistry stepwise solid phase peptide synthesis, on -OCH₂-Pam-resins (free "carboxyl peptides) or on HSCH₂CH₂CO-Leu-OCH₂-Pam-resin ("thioester peptides). 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). 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]-"thioester (observed mass (ob.) 525.4±0.4 Da, calculated average mass (ca.) 524.6 Da) [Thr¹-Ala¹⁵]-"thioester, (ob. 1491.4±0.5 Da, ca. 1491.8 Da), [Cys¹⁶-Leu²²]-"thioester, (ob. 1257.2±0.5 Da, ca. 1257.6 Da), [Thz²⁶-Gly³¹]-"thioester, (ob. 827.5±0.5 Da, ca. 828 Da), [Cys³²-Thr³⁹]-"thioester, (ob. 987.8±0.5 Da, ca. 988.3 Da), and [Cys⁴⁰-Asn⁶⁶] (ob. 738.4± 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 108
by use of semi-preparative Vyde C4 and C8 column (1cm x 25cm) at flow rate of 10mL/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 and combined based on ESMS analysis.

**Preparation of peptide-$^a$thiophenlyester from peptide-$^a$thioester** Lyophilized peptide-$^a$thioester after HF cleavage was used for the exchange reaction. The crude peptide was dissolved in pH 6.8 aqueous buffer containing 6 M guanidinium 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. After the conversion, the reaction was quenched (acidified to pH 3) by HCl. Residual thiophenol and oxidized thiophenol were removed by extraction with diethyl ether. Acid quenching was necessary before diethyl ether extraction procedure to prevent transthioesterification with residual HSCH$_2$CH$_2$CO-Leu that could not be extracted. The solution containing our product peptide was purified by preparative HPLC, and lyophilized.

**Comparison of the ligation rates for the reaction of either {peptide1-Ala-$^a$thiophenylester} or {peptide1-Ala-$^a$thioester} with a {Cys-peptide}**. Three model peptides YKMDFHIAA-$^a$thioester (ob. 1296.0±0.6 Da, ca. 1296.5 Da), YKMDFHIAA-$^a$thioethenylester (ob. 1187.1±0.5 Da, ca. 1187.5 Da), and Cys-NVRSATEPWQL (ob. 1389.3±0.5 Da, ca. 1389.5 Da) were prepared. Two separate ligation reactions between 1mM concentrations of the peptides were carried out in 200 mM phosphate buffer, pH 6.8, containing 6 M guanidinium hydrochloride. No thiol additive was present during the ligation reactions. During the ligation reaction, at each time point, an aliquot (20 µl) from the ligation reaction was quenched by addition of 5 % trifluoroacetic acid (8 µl) The
aliquot was characterized by analytical HPLC. Analysis of ligated species was done by integrating the area from analytical HPLC profiles at each time point.

**Ligation reactions**

*Kinetically controlled ligation (based on the nature of the thioester)* was performed in pH 6.8, 200mM phosphate phosphate buffer containing 6M guanidinium hydrochloride, at a concentration of 2mM for each peptide.

*Kinetically controlled ligation (based on the titration of {peptide1-"thioester} by {Cys-peptide2-"thioester}).* During the titration, high concentration of {peptide1-"thioester} is dissolved in 0.5 M Bis-Tris buffer containing 6M guanidinium hydrochloride and 1% thiophenol (vol/vol), and the ligation mixture was stirred. Also, {Cys-peptide2-"thioester} is maintained in the water containing 6M guanidinium hydrochloride (pH 4~5) to prevent any ligation reaction before its titration (i.e. slow addition) into the ligation mixture. The titration rate was controlled to be finished in several hours.

*Native chemical ligation* was carried out in the same condition for the kinetically controlled ligation except the use of 1% thiophenol.

**References**


Chapter 6. Exploring the Effects of a Salt Bridge on the Folding of Crambin: Synthesis and Xray Crystal Structure of a Protein Analogue, Crambin-\textsuperscript{carboxamide}

**Abstract.** The crystal structure of crambin clearly shows a salt bridge between the \^{\textsuperscript{\textgreek{g}}}\textit{guanidinium} of Arg10 side chain and the \textsuperscript{\textgreek{c}}\textit{arboxylate} of Asn46. The salt bridge was proposed to be a very important factor for the high stability of the folded crambin molecule. We explored (1) how the perturbation of the salt bridge affects the formation of the correct disulfide bonds in the crambin protein molecule; and, (2) how perturbation of the salt bridge affects the folded structure of the crambin molecule. To accomplish our goal, we chemically synthesized a unique protein analogue, crambin-\textsuperscript{carboxamide}. The rate and yield of folding and disulfide bond formation from the fully reduced polypeptides of crambin-\textsuperscript{carboxamide} and the native crambin polypeptide were compared under identical conditions. In addition to this kinetic experiment, to study how the salt bridge perturbation affects the overall structure of the crambin molecule, we solve the crystal structure of crambin-\textsuperscript{carboxamide} by Xray crystallography. Our experimental studies strongly suggested that the salt bridge of crambin molecule guides the formation of correct disulfide bonds and contributes to the tightly folded overall structure of the native crambin molecule.
Introduction:

Being able to systematically improve the properties of a protein molecule is one of the most important goals in current biotechnology. Fundamental to this quest is to understand the principles governing protein folding and structure. Several principles underlie the formation of the folded structure of proteins; these include the packing density, core hydrophobicity, and the stability of each secondary structural element.\textsuperscript{[1]} In addition to these principles, electrostatic interactions from salt bridges between side chains of charged amino acids, such as between Glu/Lys or Arg/Asp, are considered to contribute to the formation of the folded form of protein molecules.

Many experimental studies have been carried out to probe the contribution of salt bridges, using a variety of model peptides and proteins.\textsuperscript{[2-6]} Even though salt bridges of a few proteins showed significant thermodynamic contribution for the stability of proteins (as much as 12-20 kJ/mol\textsuperscript{[7]}), it is believed that the contribution from the formation of a salt bridge is largely context dependent.\textsuperscript{[8]} Thus, a detailed understanding of how nature uses the salt bridge for forming and maintaining the folded structure of a protein molecule will be very valuable for the systematic design of stable proteins.

A small model protein, crambin, has been recognized as a useful model to study various aspects of the formation and stability of protein molecules.\textsuperscript{[9]} Crambin is very stable not only in aqueous media but also in organic media.\textsuperscript{[10, 11]} It is believed that the unusually high stability of the crambin molecule mainly comes from three disulfide bridges and a salt bridge between the -α-carboxylate of the polypeptide chain and the side chain guanidinium of Arg10. Especially, this a -α-carboxylate-Arg10 side chain salt bridge in the crambin molecule was proposed to be a very important factor in the folding and
stability of crambin.\textsuperscript{[9]} The crystal structure of crambin isolated from plant seeds was solved with sub-Angstrom resolution (0.54\textdegree\textsubscript{A}),\textsuperscript{[12]} and the high-resolution structure clearly showed a salt bridge between $^8\text{guanidinium}$ of Arg10 side chain and $^a\text{carboxylate}$ of Asn46.

Despite the proposed importance of the salt bridge effect in the crambin molecule, it has not been explored by site-directed mutagenesis or other recombinant protein engineering techniques. The expression of recombinant crambin has been problematic and has resulted in only low yields of protein.\textsuperscript{[13]} In addition to the difficulty of rec-DNA expression, genetically encoded amino acids would not be the ideal to replace $^8\text{guanidinium}$ of Arg with minimal perturbation of the structure, because there is no non-charged isosteric coded amino acid with which to replace the side chain of Arg (e.g. citruline). The same applies to the $^a\text{carboxylate}$ of protein molecule's peptide chain (e.g. $^a\text{carboxamide}$). To ensure minimal effects on all but the desired perturbation of the salt bridge, we decided to make use of unique capabilities of chemical protein synthesis.

Controlling the chemical structure of proteins through total synthesis of the protein molecule overcomes the intrinsic limitations of cell-based protein engineering methods. We can incorporate any chemical moieties into protein structure by the chemical synthesis of appropriate peptide building blocks.\textsuperscript{[14]} In this case, we want to use chemical synthesis to control the chemical structure of crambin at an atomic resolution, and to perturb the salt bridge without affecting other structural motifs in the crambin molecule.

In this study, we focused on: (1) how perturbation of the salt bridge affects the formation of the correct disulfide bonds in the crambin protein molecule; and, (2) how
perturbation of the salt bridge affects the folded structure of the crambin molecule. The kinetics of disulfide bond formation, together with the crystal structure of a crambin analogue with the salt bridge perturbed in a precise and controlled fashion, strongly suggest that the Arg10{guanidinium}⁻⁻⁻carboxylate salt bridge of crambin guides the formation of correct disulfide bridges and that the salt bridge contributes to holding together the well ordered structure of the crambin molecule.

**Design of the Synthesis:**

We considered two possibilities for the use of chemical analogues designed to perturb the salt bridge between \(^\delta\)guanidinium of Arg10 side chain and \(^a\)carboxylate of Asn46. First, the replacement of the side chain of Arg10 with a non-coded amino acid such as citrulline was considered. Citrulline has a \(^b\)urea group in the place of \(^\delta\)guanidinium of Arg10, and is a reasonably isosteric non-charged replacement for the \(^\delta\)guanidinium group (Figure 6.1). This \(^b\)urea group of a citrulline10 residue would eliminate the salt bridge with the \(^a\)carboxylate of the protein molecule. However, the high-resolution crystal structure of crambin shows that the \(^\delta\)guanidinium of Arg10 also forms hydrogen bond with carbonyl oxygen of Thr2 backbone (Figure 6.2). Thus, replacement of Arg10 \(^\delta\)guanidinium side chain will perturb not only the salt bridge with the \(^a\)carboxylate, but will also affect the interaction with the carbonyl group from the backbone of Thr2. Citrulline may not be the optimal replacement to maintain these interactions.
Figure 6.1. Citrulline has a \(^\delta\)urea group in the place of \(^\delta\)guanidinium of Arg. Thus it is a reasonably isosteric non-charged replacement for the \(^\delta\)guanidinium group.

Figure 6.2. Stereoview of the crystal structure of crambin-\(^\delta\)carboxamide. The Thr2 and Arg10 are highlighted with ball and stick representation. The hydrogen bonding between keto oxygen of Thr2 and \(^\delta\)guanidinium group of Asn10 are marked by red colored arrows (PDB accession code: 1AB1).
Thus, we considered the replacement of the $^{6}$carboxylate of the crambin protein molecule by an $^{6}$carboxamide moiety. This can be uniquely and easily made through chemical synthesis of crambin, simply by making the corresponding peptide building block as the peptide-$^{6}$carboxamide. Moreover, the $^{6}$carboxamide moiety is isosteric with the $^{6}$carboxylate, and it will remove the negative charge to insure no ion pair formation with $^{4}guanidinium$ of Arg10. Thus, the $^{6}$carboxylate by $^{6}$carboxamide replacement must disrupt the salt bridge, but it should not affect other chemical/physical motifs in the crambin molecule. Our designed and anticipated disruption from the chemical crambin-$^{6}$carboxamide analogue is shown in Figure 6.3.

---

**Figure 6.3.** Ribbon diagram of crambin molecule and crambin-$^{6}$carboxamide. A salt bridge between $^{4}guanidinium$ side chain of Arg10 and $^{6}carboxylate$ of Asn46 of the crambin molecule is highlighted with ball and stick representation (left). Our crambin-$^{6}$carboxamide designed to specifically disrupt the salt brigde, i.e. the chemically synthesied crambin-$^{6}$carboxamide protein molecule is shown at right.

---

**Result and Discussion:**

118
Synthesis and characterization of crambin-α-carboxamide. Total synthesis of the crambin-α-carboxamide protein molecule was carried out by the sequential chemical ligation of three peptide segments.[15] Cram[1-V15A]-α-thioester, Cram[Cys(Acm)16-31]α-thioester, and Cram[Cys32-46]-α-carboxamide, were each prepared by Boc chemistry solid phase peptide synthesis.[16] Detailed experimental procedures for the chemical synthesis of crambin-α-carboxamide are described in the Experimental Section. The C-terminal peptide, Cram[Cys32-46]-α-carboxamide, was prepared by solid phase synthesis on p-methylbenzhydrylamine-resin; p-methylbenzhydrylamine-resin generates peptide-α-carboxamide upon cleavage by HF.

The synthetic full-length polypeptide, Cram[1-46]-α-carboxamide (observed mass: 4707.0±0.7 Da; calcd: 4707.4 Da), from the sequential ligation of three peptide segments was purified by preparative HPLC and lyophilized, and was used for folding and disulfide bond kinetics experiment. Like the native 46 residue peptide, the full-length polypeptide-α-carboxamide folded to a unique disulfide-cross linked structure by reaction in pH 8 aqueous buffer containing appropriate redox reagents and a low concentration of guanidinium.HCl acting as a chaotrope to solubilize mis-folded forms. The folded crambin-α-carboxamide (observed mass: 4701.5±0.5 Da; calcd: 4701.4 Da average isotopes). Folded protein species eluted later in analytical HPLC analysis than the reduced polypeptides, reflecting the known hydrophobic nature of the folded protein molecule.[15] In addition to the later elution time, the folded protein species has a mass reflecting the loss of 6 Daltons from formation of three disulfides. The folded proteinα-carboxamide was then purified by preparative HPLC. The analytical data for the reduced and folded/disulfide-containing molecules are shown in Figure 6.4. The
synthetic crambin-\textsuperscript{α}carboxamide was crystallized and its three dimensional molecular structure was determined by X-ray crystallography and was compared with native crambin molecule (see below).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{example.png}
\caption{LC-MS profiles of purified polypeptide Cram[1-46]-\textsuperscript{α}carboxamide in reduced form (above) (ob: 4707.0±0.7 Da; calcd: 4707.4 Da), and the folded disulfide-containing crambin-\textsuperscript{α}carboxamide (below) (ob: 4701.5±0.5 Da; calcd: 4701.4 Da). UV profiles in the left panels are monitored at 214 nm. Electrospray mass spectrometry data.}
\end{figure}
corresponding to the peak in the UV profiles are shown in the right panels; only the [M+3H]$^{3+}$ charge state is shown for each. Chromatographic separations were performed on Vydac C4 column using a linear gradient (30-50%) of buffer B in buffer A over 10 min at a flow rate of 1mL/min. Buffer A = 0.1% TFA in water; buffer B = 0.08% TFA in acetonitrile. For MS data, observed masses were derived from the m/z values for all observed protonation states of a molecular species.

---

*Kinetics of folding/disulfide formation.* The rate and yield of folding and disulfide bond formation from the fully reduced polypeptides of crambin-$α$-carboxamide and native crambin were performed under identical conditions, and compared (Figure 6.5). At certain time points during the course of the folding reactions, we took aliquots of the solution; the reaction was quenched with acid (see Experimental for details), and the extent of folding was determined by analytical reverse phase HPLC. After 30 min of elapsed time in the folding reaction, the amount of folded protein species did not increase.

The folding kinetics experiment showed a striking difference for the yields of formation of correct disulfide bond in the two crambin molecules. The native crambin 46 residue polypeptide-$α$-carboxylate folded and formed disulfide bonds with great efficiency, as reported before.\(^{15}\) However, we observed the folding and the disulfide bond formation of salt bridge perturbed crambin-$α$-carboxamide was less efficient as shown in following Figure 6.5; less than 60 % of reduced polypeptide of crambin-$α$-carboxamide gave the correctly-folded, disulfide-bonded protein molecule under the conditions used in this study (see Figure 6.5 B).
**Figure 6.5.** Folding craminbin and craminbin-α-carboxamide at pH 8, 1M guanidinium hydrochloride, 8mM cysteine and 1mM cystine (Left) at a concentration of 1mg/mL. HPLC analysis of native craminbin (blue) and craminbin-α-carboxamide (red) for the folding kinetics experiment. HPLC profiles were monitored at 214 nm. Chromatographic separations were performed on a Vydac C4 column 4.6x250mm using a linear gradient (10-50%) of buffer B in buffer A over 20 min at a flow rate of 1mL/min. Buffer A = 0.1% TFA in water; buffer B = 0.08% TFA in acetonitrile. (Right) Height of the peak corresponding to correctly folded protein (at 19.5min) plotted as a function of folding reaction time.
Crystal structure of crambin-αcarboxamide. To study how the structure of the crambin molecule is affected by the salt bridge perturbation, we decided to grow crystals of crambin-αcarboxamide, and to solve the structure of this unique analogue protein by X-ray crystallography. Extensive screening of crystallization conditions for crambin-αcarboxamide was performed. Unlike the crystal growth of native crambin (described in Chapter 8), crystal growth of the salt bridge perturbed chemical analogue took much more effort and time, which is also indirect evidence for a more disordered molecular structure. Subsequent rounds of optimization led to X-ray crystallography-quality crystals for crambin-αcarboxamide. Crystal formation could be reproducibly obtained by mixing 2 μl of protein (10 mg/ml in pH 7.5 HEPES buffer containing 100 mM NaCl) and 2 μl of a 0.8 M succinic acid at pH 7.0 (used for X-ray diffraction), or by mixing 2 μl of the protein solution with 0.1 M HEPES buffer containing 15% v/v tacsimate and 2% w/v polyethylene glycol 3350 at pH 7. Elongated single crystals (0.05 x 0.05 x 0.2mm³) were formed in 4-5 weeks.

X-ray diffraction data were collected using synchrotron radiation at the Advanced Photon Source at the Argonne National Laboratory. Crystals diffracted to a resolution of ~1.5 Å. R32 space group for the crystal was identified. The diffraction data set was obtained to 1.6 Å with an 97.7% completeness. The structure of crambin-αcarboxamide molecule was solved by the molecular replacement method using our native crambin crystal structure as a model, and subsequent refinement procedures were carried out by CNS\textsuperscript{17} and Refmac5 protocols\textsuperscript{18} (see Table 6.1 for data collection and refinement statistics).
Table 6.1. Data collection and refinement statistics for crambin-carboxamide.

<table>
<thead>
<tr>
<th>Crystal Space group</th>
<th>crambin-carboxamide</th>
</tr>
</thead>
<tbody>
<tr>
<td>a=b=67.58 Å, c=39.49 Å, 0.05 x 0.05 x 0.2</td>
<td>1.8</td>
</tr>
<tr>
<td>Matthews coefficient</td>
<td>1 protein (46 a.a.)</td>
</tr>
<tr>
<td># Protein in asym.unit</td>
<td>31.4</td>
</tr>
<tr>
<td>Solvent content, %</td>
<td>APS-5ID / MAR225</td>
</tr>
<tr>
<td>Beamline / Detector</td>
<td>1.0000</td>
</tr>
<tr>
<td>Wavelength, Å</td>
<td>25 -1.60</td>
</tr>
<tr>
<td>Resolution range, Å</td>
<td>4568</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>3.8</td>
</tr>
<tr>
<td>Redundancy</td>
<td>97.7</td>
</tr>
<tr>
<td>Completeness, %</td>
<td>6.8</td>
</tr>
<tr>
<td>R-merge, %</td>
<td>21.6%/23.4%</td>
</tr>
<tr>
<td>R factor / R freeb, %</td>
<td>No.of residues in refinement</td>
</tr>
<tr>
<td>43 aa +10H2O</td>
<td></td>
</tr>
</tbody>
</table>

a) R-merge = S_wrap, S1/I(hkl) - <I(hkl)>/S_wrap, S1/I(hkl),> over i observations of a reflection hkl.
b) R factor = Σ | |F(obs) - |F(calc)| |Σ|F(obs)|, 
R-free is the same calculated with 5% data withheld from refinement

Overall structures of crambin and crambin-carboxamide were compared by superimposing the main chain atoms of the two proteins (Figure 6.6).
Figure 6.6. Stereoview of superimposed crystal structures of crambin-α-carboxamide (red) and crambin (cyan). Top panel shows a stereoview of the superposition of synthetic crambin-α-carboxamide (red) and crambin (cyan). The Cα-atom traces of the two proteins are shown. The root-mean-square-difference (rmsd) value for the comparison of the all corresponding main chain atoms (from residue 1 to 43 in the two protein structures was 0.77Å. Bottom panel shows stereoview of the crystal structure of crambin-α-carboxamide. The $^\alpha$guanidinium of Arg10 side chain is highlighted with ball and stick representation.

The crystal structure of crambin-α-carboxamide showed a striking disorder from residue 35 to 46 compared to the native crambin molecule. Comparison of main chain B-
factor values between crambin*-carboxamide and native crambin (Figure 6.7) clearly showed increased flexibility in the 35-46 residue region of the overall folded structure of crambin*-carboxamide, caused by the perturbation of only one salt bridge. More than a third of the crambin*-carboxamide structure became more flexible than native crambin. In particular, ten C-terminal residues of crambin*-carboxamide molecule lost their ordered and defined structure. Building reasonable model of crambin*-carboxamide from residue 44 to 46 was impossible; residues from 41 to 43 were highly disordered (B factors > 50 Å²). This pattern of disorder was also observed in the disulfide bridges. Only the formation of two of three native disulfide bonds in the crambin*-carboxamide protein molecule is clearly evident from the electron density map. Table 6.2 shows the gamma sulfur atoms from six cysteines. The B values of Cys40 and its disulfide partner Cys3 of crambin*-carboxamide molecule showed relatively high B values reflecting their disorder affected by the perturbation on the salt bridge.

Table 6.2. The B values of the gamma S atoms from six cysteines. The values of crambin* were taken from Chapter 8. Disulfide partners are coded with same colors.

<table>
<thead>
<tr>
<th>Residue</th>
<th>B values of gamma S atoms of crambin*</th>
<th>B values of gamma S atoms of crambin*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cys3</td>
<td>35.3</td>
<td>27.5</td>
</tr>
<tr>
<td>Cys4</td>
<td>30.8</td>
<td>28.9</td>
</tr>
<tr>
<td>Cys16</td>
<td>30.8</td>
<td>28.6</td>
</tr>
<tr>
<td>Cys26</td>
<td>32.7</td>
<td>28.0</td>
</tr>
<tr>
<td>Cys32</td>
<td>30.7</td>
<td>27.2</td>
</tr>
<tr>
<td>Cys40</td>
<td>39.3</td>
<td>30.0</td>
</tr>
</tbody>
</table>
Considering that crambin-o-carboxamide crystal diffracted to 1.6 Å with 97 % completeness, the high B-factor for the crambin-o-carboxamide is a clear evidence of increased flexibility in the protein molecule caused by the perturbation of the salt bridge.

**Figure 6.7** Comparison of main chain B-factor values between *crambin-o-carboxamide* (left) and *native crambin* (right). B-factor ranges are shown in different colors; violet, 1-25Å²; blue, 26-30Å²; cyan, 31-33Å²; green, 34-38Å²; yellow, 39-41Å²; orange, 41-44Å²; red, 45-55Å². For crambin-o-carboxamide, 43 residues could be built and refined, and residue 44-46 were not seen in electron-density map.

Our folding/disulfide bond formation kinetic experiment, together with the crystal structure comparison between crambin and crambin-o-carboxamide, provides strong evidence that the salt bridge in the crambin molecule guides the formation of the correct disulfide bonds in crambin molecule. The uniquely precise elimination of a salt bridge by
the replacement of $\text{-}^{\alpha}\text{carboxylate}$ by $\text{-}^{\alpha}\text{carboxamide}$, that is only a single atom change in the protein structure, not only perturbed the salt bridge in the crambin-$^{\alpha}$carboxamide molecule and but also affected the last ten residue of the C-terminus of the molecule.

**Conclusion:**

Contribution of a salt bridge involving the $\text{-}^{\alpha}\text{carboxylate}$ of the crambin protein molecule to folding and disulfide formation of a hydrophobic protein has been explored by design and synthesis of a crambin-$^{\alpha}\text{carboxamide}$ molecule. From the folding kinetic experiment and the crystal structure of the crambin-$^{\alpha}\text{carboxamide}$ molecule, the contribution of the salt bridge to disulfide bond formation and to the formation and stability of the highly ordered structure of native crambin was evident. Our kinetic data showed that disulfide bond formation is highly favored by the salt bridge between $^{\delta}\text{guanidinium}$ of Arg10 side chain and $^{\alpha}\text{carboxylate}$ of Asn46. Comparison of main chain B-factor values between crambin-$^{\alpha}\text{carboxamide}$ and native crambin clearly showed the perturbation of overall folded structure of crambin-$^{\alpha}\text{carboxamide}$ cause by the perturbation of one salt bridge. From these data, we infer that the salt bridge of crambin molecule guides the formation of correct disulfide bonds and contributes to the tightly folded overall structure of the native crambin molecule. Our work on crambin salt bridge provides strong support for the important contribution of salt bridges to the stability of protein structures.

**Transition to the next chapter:** This study led us to a study about using chemistry to engineer the salt bridge as a covalent bond. In the next section, we explored how this led us an interesting protein topological analogue, ‘crambin cyclotide’.

128
Experimental Section

**Abbreviations for peptides.** Cram[1-V15A]"thioester: TTCCPSIVARSNFNA-SCH₂CH₂CO-Leu; (a smaller font was used to distinguish atomic symbols from the single letter code for amino acids) Cram[Cys(Acm)¹⁶-³¹]"thioester: C(Acm)RLPGTPEALCATYTG-SCH₂CH₂CO-Leu, Cram[Cys³²-⁴⁶]"carboxamide: CIIIPGATCPGDYAN-"carboxamide.

**Materials** Boc-amino acids, S-trityl-β-mercaptopropionic acid, p-methylbenzhydramine-resin and 2-(1H-Benzotriazol-1-yl)-1,1,3,3-tetramethylyuronium hexafluorophosphate (HBTU) were obtained from Peptide Institute, Inc., (Osaka, Japan). Boc-Asn-OCH₂-Pam-resin, Boc-Leu-OCH₂-Pam-resin, and N,N-diisopropylethylamine (DIEA) were obtained from Applied Biosystems (Foster City, California). N,N-dimethylformamide (DMF), dichloromethane(DCM), and acetonitrile were purchased from Fisher (Chicago, Illinois). Thiophenol and p-cresol were purchased from Sigma-Aldrich (St. Louis, Missouri). Trifluoroacetic acid (TFA) was from Halocarbon (New Jersey).

**HPLC and LC-MS.** Analytical reverse phase HPLC and LC-MS were performed on an Agilent 1100 Series chromatography instrument equipped with an MSD ion trap, using Vydac C4 columns (5μm, 0.46 × 25 cm). Chromatographic separations were typically performed using a linear gradient (10-60%) of buffer B in buffer A over 25 min at a flow rate of 1mL/min. Buffer A = 0.1% TFA in water; buffer B = 0.08% TFA in acetonitrile. For MS data, observed masses were derived from the m/z values for all protonation states of a molecular species. Calculation of theoretical masses was based on average isotope composition. Preparative HPLC was performed on a Waters Prep LC
4000 system using Vydac C4 column (12μm, 2.2 × 25 cm) at flow rate of 10mL/min, with a typical gradient of 20-40% buffer B in buffer A over 40 min. Fractions were pooled based on LC-MS analysis and lyophilized.

**Peptide segment synthesis.** Peptides and peptide-"thioesters were made manually by ‘in situ neutralization’[^16] Boc chemistry stepwise solid phase peptide synthesis, on p-methylbenzhydrylamine-resin (for peptide-"carboxamide), on –OCH₂-Pam-Resins (free "carboxyl peptides), or on SCH₂CH₂CO -Leu-OCH₂-Pam-resin ("thioester peptides). Side chain protection for amino acids was as follows: Arg(Tos), Asn(Xan), Asp(OcHex), Cys(4-CH₃Bz) or Cys(Acm), Glu(OcHex), Lys(2-Cl-Z), Ser(Bz)l, Thr(Bzl), Tyr(Br-Z). After solid phase chain assembly was complete, peptides were deprotected and simultaneously cleaved from the resin support by treatment with anhydrous HF containing p-cresol (90:10, v/v) for 1 hour 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. Optimally, the peptide solution was diluted with water containing 0.1% TFA, and injected into the preparative HPLC system without prior lyophilization. Peptide compositions were confirmed by LC-MS. Cram[1-V15A]"thioester (observed mass (obs.) 1785.3±0.4 Da, calculated average mass (calcd.) 1785.9 Da), Cram[Cys(Acm)]¹⁶-3¹⁻"thioester (obs.1925.0±0.3 Da, calcd.1925.3 Da), Cram[Cys³²-4⁶]-"carboxamide (obs. 1506.4±0.4 Da, calcd. 1506.7 Da)

**Native Chemical Ligation of Cram[Cys(Acm)]¹⁶-3¹⁻"thioester and Cram[Cys³²-4⁶]-"carboxamide.** Native chemical ligation reactions were performed in aqueous solution containing 6M GdnHCl, 100 mM phosphate, pH 7.5, 1% thiophenol 130
(v/v), at concentration of ~2 mg/ml for each peptide. The ligation reaction was complete in 10 hours at room temperature. The product was characterized by LC-MS analysis: Cram[Cys(Acm)_16-46]-α-carboxamide (ob. 3212.4±0.7 Da, ca.3212.7 Da). The ligated peptide Cram[Cys(Acm)_16-46]-α-carboxamide was purified by prep-HPLC.

**Acm-group removal by Ag(OAc)₂ and DTT.** Acm-group removal by Ag(OAc)₂ was performed directly in solution containing HPLC gradient buffer. Ag(OAc)₂ (20 mM mmol) was added to the solution from preparative HPLC. Reaction was complete (LC-MS) in one hour, and Cram[Cys_16-46]-α-carboxamide–silver complex was formed. DTT (24 mM) was added to the solution for quenching, and a precipitate formed immediately. After centrifugation, the supernatant was filtered and used for preparative HPLC. Fractions containing purified peptide were identified by LC-MS, combined and lyophilized.

**Native Chemical Ligation of Cram[1-V15A]α-thioester and Cram[Cys_16-46]α-carboxamide.** Native chemical ligation reactions were performed in 6M GnHCl, 100 mM sodium phosphate, pH 7.5, 1% thiophenol (v/v), at a concentration of 2 mg/ml for each peptide. Ligation reactions were complete in 24 hours at room temperature. The reactions were performed at 0.013-0.017 mmol scale, and purified yields obtained varied from 64% to 73%. Cram[1-46]-α-carboxamide (ob: 4707.0±0.7 Da; ca: 4707.4 Da).

**Folding.** The purified polypeptide chain Cram(1-46)-α-carboxamide was folded in 2M GnHCl, 100 mM Tris, 8 mM cysteine, 1 mM cystine, pH 8.0, at a concentration of ~0.2 mg/ml with exclusion of air. During the folding reaction, no stirring was performed. Folded crambin-α-carboxamide was characterized by LC-MS (ob: 4701.5±0.5 Da; ca: 4701.4 Da), and purified with preparative HPLC.
Folding kinetics. The purified polypeptide chains Cram[1-46]"carboxamide and Cram(1-46) were folded in 2M GlnCl, 100 mM Tris, 8 mM cysteine, 1 mM cystine, pH 8.0, at concentration of ~0.2 mg/ml. During the folding reaction, no stirring was performed. At each time point, an aliquot from the folding reaction was quenched by addition of 5 % formic acid. Folded crambin products were characterized by analytical HPLC. Analysis folded species was done by integrating the area from analytical HPLC profiles at each time point.

References


Chapter 7. Design and Synthesis of a Protein Molecule of Novel Topology based on Crambin

Abstract. We explored the design, synthesis, and some properties of a novel protein topological analogue using crambin. We used chemistry to replace the salt bridge between the $\delta$guanidinium of the Arg10 side chain and the $\alpha$carboxylate of Asn46 of the crambin molecule with a covalent bond. The novel crambin topology has two distinct features. First, the polypeptide chain of the crambin topological analogue will have a start (N-terminus) but the molecule will have 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 they penetrate the covalent ring made by the rest of polypeptide cyclized through the C-terminal and the side chain of residue 10. For the chemical synthesis of this novel protein construct we used an innovative synthetic approach. We chose to divide the full-length polypeptide into a 30 residue branched peptide and an easily made 16 residue peptide, and we joined these two peptides through a convergent route using kinetically controlled ligation (described in Chapter 5). Total synthesis of the novel protein analogue was carried out with great efficiency. The properties of this novel protein are still being explored.
Introduction:

In Chapter 6, we explored the role of a salt bridge between the $\text{guanidinium}$ of the Arg10 side chain and the $\text{carboxylate}$ of Asn46 at the C-terminal of the crambin polypeptide chain. Our experimental observations suggested that this salt bridge in the crambin molecule guides the formation of correct disulfide bonds and contributes to the tightly folded overall structure of the native crambin molecule.

These studies led us to ask: what if we use chemistry to replace the salt bridge with a covalent bond? Can we design and carry out the synthesis of such a molecule? Will the polypeptide, in the form of a covalent loop, fold to give native disulfides? Whether cysteine residues find their correct disulfide bond partners in the presence of the artificial covalent bond is a significant question. Finally, how will the new covalent bond affect the stability of the protein molecule?

Exploring the possibility of a covalent linkage between the side chain of Arg10 and the $\text{carboxylate}$ of Asn46 at the C-terminal of the protein led us to a novel protein topology (Figure 7.1) that is fascinating in two respects. First, the novel crambin molecule will have a start (N-terminus) but the molecule will have 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 polyanide ring consisting of the polypeptide from residue 10 to residue 46 joined by amide bond between the side chain of residue 10 (Arg10 to be replaced by Lys10) and the $\text{carboxylate}$ of Asn46.
Our challenge was to design the synthesis of such a molecule, to successfully carry out the synthesis, and then to determine the structure and properties of this unprecedented protein.

Figure 7.1 Top panel shows a stereoview of crambin Cα traces. Bottom panel highlights the side chain of Arg10 and the α-carboxylate of Asn46 (side chains are drawn with a licorice method) of crambin. The red colored Cα trace shows the N-terminal nine-residue-linear-polypeptide that penetrates through the covalent polyamide ring consisting of the polypeptide from residue 10 to residue 46 joined by amide bond between the side chain of residue 10 (Arg10 to be replaced by Lys10) and α-carboxylate of Asn46.
This novel protein topology is reminiscent of, but not identical to, the backbone N-to-C covalently cyclized plant proteins, the ‘cyclotides’. Cyclotides are a new protein class that has a very unusual cyclic backbone, i.e. consisting of intramolecular amide bond connection between N- and C- termini of protein. In addition to the cyclic backbone, the cyclotide has an extraordinary cysteine knot topology in which two disulfide bonds and their connecting backbone segments form an embedded ring in the structure that is penetrated by a third disulfide bond. The high stability of cyclotides is thought to come from the exceptional topology, and that along with their interesting biological activities including anti-HIV and anti-microbial activities has made them very interesting class of target proteins to be studied.

**Design:**

We envisioned the covalent attachment of the \(^8\)carboxylate of Asn46 to the \(^9\)amine group of Lys10 (i.e. Arg10Lys) in crambin, as shown in the **top panel of the Figure 7.2.** Synthesis of this novel topology is a first step to understanding the molecule and to verifying our argument. Retrosynthetic analysis (**Figure 7.2**) of the total synthesis of our novel topological analogue showed several challenging aspects of the synthesis. The novel topological analogue will be joined through the side chain of residue10 and \(^8\)carboxylate of residue 46 at the polypeptide level before the folding and disulfide formation step. To obtain the cyclic polypeptide, we need to synthesize a branched full-length polypeptide chain; as shown in the **Figure 7.2**, we chose to connect residue15 and residue16 by native chemical ligation. To be able to have the disconnected full-length polypeptide, we need to synthesize the [1-15]-\(^\text{thioester on the solid phase resin, and then}\)
build residues (from 46 to 16) from the side chain of Lys10. The ‘amine group of Lys10 must be protected with a group stable to the conditions used during the Boc- chemistry solid phase synthesis of the [1-15]-‘thioester, and to be selectively removed without affecting other side chain protecting groups of the peptide attached on resin. For these reasons, we chose to use a base labile Fmoc- protecting group for the ‘amine group of Lys10. After the removal of the Fmoc- group, the solid phase peptide synthesis would be continued on the side chain of Lys10, starting from residue 46 to residue 16.
Figure 7.2 Retrosynthetic analysis of the total synthesis of our novel topological analogue. This initial design was replace by our improved design (shown in Figure 7.3).

This proposed synthetic route (Figure 7.2) is not optimal because the stepwise solid phase synthesis of the full-length branched 46 residue polypeptide is unlikely to give an acceptable product. In the chapter 2, we have shown that obtaining even a nonbranched 31 residue peptide of crambin is very difficult.\[4\]
Synthesis of our novel topological analogue thus required a *convergent* route. In other words, the full-length branched polypeptide must be prepared by ligation of two fragment peptides in a convergent fashion. Thus, we chose to divide the full-length polypeptide to a 30 residue branched peptide and an easily made 16 residue peptide, as shown in Figure 7.3.

As reported in the Chapter 5, kinetically controlled ligation is a very practical method to realize the convergent synthesis of proteins. In the initial kinetically controlled ligation strategy, we used the high dependence of native chemical ligation reaction on the last residue (Xxx) of peptide-“thioester {peptide-Xxx-“thioester}. For the synthesis of our target crambin cyclotide, the ligation between {peptide1-Gly-“thioester} and {Cys peptide2-Val-“thioester} would expected to exclusively form {peptide1-Gly-Cys peptide2-Val-“thioester}, because Cys reacts with a Gly-“thioester many times faster than with a Val-“thioester. The use of kinetic control of this type enabled a reliable synthetic approach to the crambin analogue of novel topology, as shown in Figure 7.3.
**Figure 7.3** Our improved retrosynthetic design for the total synthesis of novel crambin topological analogue by a convergent route

First, kinetically controlled ligation would exclusively result in the ligation between \{peptide-Gly\textsuperscript{31}-thioester\} and \{Cys\textsuperscript{32}-peptide-Val\textsuperscript{15}-thioester\}, but the formation of intramolecular ligation of \{Cys\textsuperscript{32}-peptide-Val\textsuperscript{15}-thioester\} will be prevented by the kinetic control, i.e. the intermolecular ligation is many times faster than the intramolecular ligation in a solution containing a high concentration of both reaction
peptides. The second intramolecular ligation reaction was to be performed at a low concentration of peptide, to minimize intermolecular side reactions. This time, intramolecular chemical ligation would be carried out, but the low concentration will ensure the prevention of intermolecular chemical ligation reaction to form non-controlled oligomerization of the polypeptide building blocks. The resulting full-length branched polypeptide was to be folded with simultaneous formation of disulfide bonds in the presence of appropriate redox agents.

**Result & Discussion**

The branched peptide [Thr$^1$-Lys$^{10}$](Asn$^{46}$-Cys$^{32}$)-Val$^{15}$]$^\alpha$thioester was prepared by stepwise solid phase peptide synthesis,$^{[5]}$ and was purified by preparative HPLC. We used the [Thz$^{16}$-Gly$^{31}$]$^\alpha$thioester from a previous native crambin synthesis. $^{[6]}$ As our design anticipated, the first *intermolecular* ligation between [Thr$^1$-Lys$^{10}$](Asn$^{46}$-Cys$^{32}$)-Val$^{15}$]$^\alpha$thioester and [Thz$^{16}$-Gly$^{31}$]$^\alpha$thioester resulted in full-length polypeptide [Thr$^1$-Lys$^{10}$](Asn$^{46}$-Thz$^{16}$)-Val$^{15}$]$^\alpha$thioester. The Thz$^{16}$ was converted to Cys$^{16}$ using 0.2 M methoxyamine at pH 4, and the product peptide [Thr$^1$-Lys$^{10}$](Asn$^{46}$-Cys$^{16}$)-Val$^{15}$]$^\alpha$thioester was purified by preparative HPLC in a good yield. The second *intramolecular* ligation reaction was performed at a low peptide concentration in an aqueous buffer containing 6M guanidinium hydrochloride; reaction was completed in 24 hours, to give the desired cyclized full-length polypeptide, a precursor to the novel topological protein analogue.

The formation of disulfides and folding was carried out without purifying the cyclized full-length polypeptide. An appropriate mixture containing redox reagents was added to the second ligation reaction, and the solution was diluted to reduce the
concentration of guanidinium hydrochloride and to initiate the folding and disulfide formation reaction.

We had anticipated that the folding of the full-length polypeptide to our novel topological analogue might be difficult because the target molecule has to have the nine residue linear peptide penetrating the new covalent polyamide ring. Surprisingly, folding as described gave a single product containing three disulfide bonds, within one hour. This product had the molecular mass (Calculated Mass = 4726.4 Da, Observed Mass = 4725.9±0.6 Da) consistent with the target structure reflecting the loss of six Da. from the formation of three disulfide bond formation. The novel topological protein analogue showed a later elution time, reflecting the more hydrophobic nature of the folded molecule, than the full-length reduced polypeptide in HPLC analysis. The later elution time of the folded protein molecule is also consistent with previously reported native crambin behavior.[6] The final crude product mixture, shown in Figure 7.4, was subjected to preparative HPLC. The purified novel topological analogue showed a highly homogeneous product shown in Figure 7.5.
**Figure 7.4.** HPLC chromatogram and MS profile (a mass peak of [M+3H]$^{3+}$ is shown) from LC-MS analysis of crude mixture upon folding and disulfide bond formation. Observed mass of the molecule was 4725.9±0.6 Da (Calculated Mass = 4726.4 Da). The reverse phase HPLC chromatographic separations were performed on Vydac C4 column using a linear gradient (5-65%) of buffer B in buffer A over 15 min with a flow rate of 0.5 ml/min (buffer A = 0.1% TFA in water; buffer B = 0.08 % TFA in acetonitrile).

**Figure 7.5.** UV chromatogram and MS profile from LC-MS analysis of purified novel topological analogue by preparative HPLC.

**Characterization of the novel protein analogue.** Work still is in progress. We have crystallized the crambin topological analogue, and have done preliminary X-ray diffraction studies on the crystal. Further work will be required to determine the structure and to measure the stability of the molecule.
Conclusions. In conclusion, design and total synthesis of novel protein cyclotide was carried out with great efficiency using a variation of kinetically controlled convergent synthesis. The high quality of the novel topological analogue synthetic product illustrates the power of chemical protein synthesis.

Experimental

Peptide Segment Synthesis Peptides were prepared manually by “in situ neutralization” Boc chemistry stepwise solid phase peptide synthesis, on HSCH2CH2CO-Leu-OCH2-Pam-resin (thioester peptides). Side-chain protection for amino acids was as follows: Lys(Fmoc), Arg(Tos), Asn(Xan), Asp(OcHex), Cys(4-CH3Bzl), Glu(OcHex), Lys(2-Cl-Z), Ser(Bzl), Thr(Bzl), Tyr(Br-Z). 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 tritutated 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-Lys10(Asn46-Cys32)-Val15]-thioester (N-terminal Thr residue was acetylated, sequence: Acetyl-TTCCPSIVAK(CIIIPGATCPDGYAN)SNFNV-thioester) (observed mass (ob.) 3316.0±1.0Da, calculated average mass (ca.) 3316.8 Da); [Thz16-Gly31]-thioester, (ob. 1865.6±0.8Da, ca. 1866.2 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 column (1cm x 25cm) at flow rate of TFA in water; Buffer B: 0.08% TFA in acetonitrile. Fractions were collected across the expected elution time and combined based on ESMS analysis.

Ligation reactions

Kinetically controlled ligation of [Thr\(^{1}\)-Lys\(^{10}\)(Asn\(^{46}\)-Cys\(^{12}\))-Val\(^{15}\)]-\(^{\alpha}\)-thioester (7mg) and [Thz\(^{16}\)-Gly\(^{31}\)]-\(^{\alpha}\)-thioester (4mg) was performed in 300\(\mu\)l of pH 6.8, 200mM phosphate phosphate buffer containing 6M guanidinium hydrochloride. The ligation reaction resulted in full-length polypeptide [Thr\(^{1}\)-Lys\(^{10}\)(Asn\(^{46}\)-Thz\(^{16}\))-Val\(^{15}\)]-\(^{\alpha}\)-thioester. (sequence: Acetyl-TTCCPSIVAK\(^{10}\)(Thz\(^{16}\)RLPGTPEALCATYTGACIIIPGATCPDGYAN\(^{46}\)SNFNV\(^{15}\)-\(^{\alpha}\)-thioester ob.4962.7 ±1.0Da, ca. 4963.7Da)

Intramolecular ligation reaction of [Thr\(^{1}\)-Lys\(^{10}\)(Asn\(^{46}\)-Cys\(^{16}\))-Val\(^{15}\)]-\(^{\alpha}\)-thioester was performed at a 0.5mM peptide concentration in an aqueous buffer containing 6M guanidinium hydrochloride; reaction was completed in 24 hours, to give the desired cyclized full-length polypeptide, a precursor to the novel topological protein analogue. (ob. 4731.6±1.0Da, ca. 4732.4Da)

References


Chapter 8. Crystal Structures of Synthetic Crambin (L-crambin), Enantiomeric Crambin (D-crambin), and Racemic Crambin (D&L-crambin)

Abstract. Crambin crystals formed from protein isolated from natural sources are unusually well ordered and diffract to remarkably high resolution. This makes crambin a very useful model for X-ray crystallography. Although there are many reports of the crystal structure of crambin, all of them describe the same monoclinic crystal form with P2₁ symmetry. All reported crambin crystals have been obtained from an ethanol/water mixture. The crambin molecules were dissolved in ~80% ethanol/water (v/v) and equilibrated against ~60% ethanol/water. We recognized the utility of our chemically synthesized crambin constructs (synthesis described in Chapter 2) as a subject for further X-ray crystallographic studies that may provide insights into some intriguing questions in protein crystallography. As a first step, we eliminated ethanol from the crystallization solutions and screened aqueous conditions to crystallize L-crambin. From this screening, we obtained crambin crystals that belong to different space group (I4₁32). We also applied these new crystallization conditions to obtain D-crambin crystals (i.e using the enantiomeric form of the crambin protein). The crystal structure of D-crambin was the complete mirror image of L-crambin, within experimental uncertainty. For the first time, this demonstrated that enantiomeric proteins could be crystallized within the same unit cell and symmetry. We also co-crystallized the D- and L-crambin together to form centrosymmetric racemic crystals that belong to the space group P(-1). Interesting features of these D-, L-, D&L-crambin crystal structures are discussed
Crambin is a small (46 amino acid) protein originally isolated from the seeds of the plant Crambe abyssinica.\textsuperscript{[1, 2]} The crambin molecule contains three disulfide bonds and displays β-strand, β-turn, and helical elements of protein secondary structure. Crambin is a widely used model protein for the development of NMR methods,\textsuperscript{[3]} or for the computational studies of protein folding.\textsuperscript{[4]} Most of all, crambin is a very useful model for X-ray crystallography because crambin crystals are unusually well ordered and can diffract to remarkably high resolution.\textsuperscript{[5]}

Previously, we have reported the total synthesis of crambin with a great efficiency.\textsuperscript{[6]} Using the synthetic method, we also prepared an enantiomeric crambin protein, i.e. made entirely of D-amino acids; in the D-amino acids, the side chain chiral centers of both D-Thr and D-Ile are inverted. We crystallized each D- and L-crambin construct and co-crystallized the racemic mixture of D- and L- crambines. X-ray data were collected using synchrotron radiation, and the crystal structures of the synthetic constructs were solved. Here we report the crystal structures of D-crambin, L-crambin, and D\&L- crambin.

**Sequence of synthetic crambin**

Crambin isolated from nature has two isomers: one has Ser\textsuperscript{22} and Ile\textsuperscript{25} (‘SI’ form) and the other has Pro\textsuperscript{22} and Leu\textsuperscript{25} (‘PL’ form).\textsuperscript{[1]} We explored the synthesis of both the SI form and the PL form. We replaced Val\textsuperscript{15} by Ala for synthetic convenience as described in Chapter 2.\textsuperscript{[7]} The chemically synthesized PL form of crambin was used for the crystal formation. The amino acid sequence of the target 46 residue polypeptide chain (‘PL’-[V15A]crambin) is shown in Figure 8.0.
Figure 8.0. Target crambin amino acid sequences: the ‘PL’ form has Pro\textsuperscript{32} and Leu\textsuperscript{25}. A Val\textsuperscript{15} was replaced by Ala15 for synthetic convenience.\textsuperscript{[7]}

8.1 Crystal structure of crambin (L-crambin)

Introduction to crystal structure of crambin

Crystal structures of crambin have been reported many times. In 1965, the first crambin crystal was obtained from an aqueous acetone extract from the seeds of \textit{Crambe abyssinica}.\textsuperscript{[2]} Later, single crystals of this protein were grown from aqueous ethanol solution. X-ray diffraction of the crystal gave the highest resolution (0.88 Å) obtained for any protein to that date.\textsuperscript{[8]} A 0.54 Å resolution structure of crambin using an intense synchrotron source was reported in 2000.\textsuperscript{[5]} The crystal structures of natural crambin isolated from plant seeds has been the highest resolution protein structure obtained to date.

Crystal growth

The reported crambin crystals used for X-ray crystallography were usually grown by vapor diffusion from a 30mg protein/1ml solution containing 80\% ethanol/water (v/v) equilibrated against 60\% ethanol/water.\textsuperscript{[8, 9]} We crystallized our chemically synthesized crambin molecule in aqueous buffer. Our synthetic crambin could be dissolved in pH 8.0, 100mM HEPES buffer containing 150mM NaCl up to 10 mg/ml. We carried out an extensive screening for the crystallization of the crambin molecule using commercially
available crystal screening kits. To our surprise, single crystals were grown in various
different conditions by a hanging drop method. A 2 µl aliquot of a crambin (10 mg/ml)
solution was mixed with 2 µl of various crystallization buffer solutions, and the protein
droplets were equilibrated against crystallization buffer solutions. The crystal forming
buffer solutions include (1) 1.26M (NH₄)₂SO₄ 0.1M HEPES at pH 7.5, (2) 1.26M
(NH₄)₂SO₄ 0.1M MES at pH 6.0, (3) 1.0M K/Na tartrate 0.1M MES at pH 6.0, (4) 1.26 M
(NH₄)₂SO₄ cacodylate at pH 6.5, (5) 0.8M succinic acid at pH 6.5, (6) 1.1M sodium
malonate, 0.1M HEPES, 0.5% v/v Jeffamine ED-2001 at pH 7.0, (7) 15% ethanol 0.1M
MES Zn(OAc)₂ at pH 6.0, (8) 20% 1,4-butanediol 0.1M imizazol at pH 8.0. For our X-
ray data collection, we used single crystals obtained by mixing crambin with 0.8M
succinic acid at pH 6.5 (buffer solution #5).

Data collection & structure determination

X-ray diffraction data were collected using synchrotron radiation at the
Advanced Photon Source at Argonne National Laboratory. Our crambin crystals belong
to cubic I4,32 space group and diffract to1.8 Å with 98% completeness. The structure
was solved by the molecular replacement method using MOLREP program¹⁰ and
previously reported crambin structure (PDB code 1AB1) as a model. The subsequent
refinement procedures were carried out by CNS ¹¹ and Refmac5 protocols¹². Table
8.1 summarizes the data collection and refinement statistics for our synthetic crambin.

**Table 8.1.** Data collection and refinement statistics for crambin

<table>
<thead>
<tr>
<th>Crystal</th>
<th>crambin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>I4,32 (Z=48)</td>
</tr>
<tr>
<td>Cell parameters</td>
<td>a=b=c=104.92 Å</td>
</tr>
<tr>
<td>Crystal size, mm³</td>
<td>0.07 x 0.07 x 0.07</td>
</tr>
<tr>
<td>Matthews coefficient</td>
<td>5</td>
</tr>
<tr>
<td># Proteins in asym.unit</td>
<td>1 protein (46 a.a.)</td>
</tr>
<tr>
<td>-------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Solvent content, %</td>
<td>75.3</td>
</tr>
<tr>
<td>Beamline / Detector</td>
<td>APS-5ID / MAR225</td>
</tr>
<tr>
<td>Wavelength, Å</td>
<td>1.0000</td>
</tr>
<tr>
<td>Resolution range, Å</td>
<td>25 -1.80</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>9257</td>
</tr>
<tr>
<td>Redundancy</td>
<td>4.4</td>
</tr>
<tr>
<td>Completeness, %</td>
<td>97.8</td>
</tr>
<tr>
<td>R-merge(^a), %</td>
<td>6.7</td>
</tr>
<tr>
<td>R factor / R free(^b), %</td>
<td>15.9/16.0</td>
</tr>
<tr>
<td>No.of residues in refinement</td>
<td>46 aa +80 H(_2)O</td>
</tr>
</tbody>
</table>

\(^a\) R-merge = \(S_{hkl}S_{II(hkl)} - \langle I(hkl)\rangle / S_{hkl}\langle I(hkl)\rangle\) over \(i\) observations of a reflection hkl.

\(^b\) R factor = \(\Sigma | |F(\text{obs})| - |F(\text{calc})| |\Sigma |F(\text{obs})|\), R-free is the same calculated with 5% data withheld from refinement.

**Overall features of our structure**

A ribbon representation of the crystal structure of the synthetic crambin molecule is shown in the top panel of Figure 8.1. The superposition of C\(_\alpha\) traces with previously reported crambin structure in P2\(_1\) space group is shown in the bottom panel of Figure 8.1. The two crambin molecules were also superimposed using all main chain atoms, and the root-mean-square-difference (rmsd) value at the corresponding positions is 0.45\(\AA\).
Figure 8.1. **Top panel** shows the crystal structure of our synthetic crambin molecule that was crystallized in I4₁,3₂ space group. The ribbon is colored according to the secondary structure elements: α-helix, 3₁₁ helical turn, β-strand, and loop regions are shown in cyan, blue, green and orange, respectively. **Bottom panel** shows (in stereo) the superposition of our synthetic crambin structure (shown in magenta) with previously reported crambin structure in P₂₁ space group (PDB code 1AB1, shown in yellow). The Cα-traces
are shown. The root-mean-square-difference (rmsd) value of main chain atoms at corresponding positions in the two protein structures is 0.45Å.

Despite very different packing and widely varying solvent content, the crambin molecules adopt essentially the same conformation in cubic I\(_{4_1}32\) and monoclinic P\(_2_1\) crystals. The I\(_{4_1}32\) crystals of our synthetic crambin show very high solvent content (75.3%) as compared to previously reported crambin crystals with P\(_2_1\) symmetry (32%). Considering that the solvent content in protein crystals usually ranges from 25% to 70%, these two structures near the extremes of the range of possible solvent content. One of the rationales for the ultrahigh-resolution diffraction of the P\(_2_1\) crystals was the tightly packed crambin molecules with low solvent contents in the unit cell. Thus, we are very surprised to see that our small (0.07*0.07*0.07 mm\(^3\)) crystals of the synthetic crambin diffracted well despite their extremely high solvent content.

The crystal packing in the I\(_{4_1}32\) crambin crystals is shown in Figure 8.2. The unit cell of synthetic crambin shows large solvent channels running throughout the crystal.
**Figure 8.2.** The packing arrangements in the I4$_1$32 crambin crystals. The view along a crystallographic three-fold axis is shown. **Left panel** shows the module of 24 crambin molecules packed together. Three crambin molecules form a trimer. Each trimer is colored in magenta, cyan, yellow or green. Two trimers pack against each other using their $\alpha$-helical interfaces and form a hexamer. Four neighboring hexamers are shown. One in the center is show in magenta and cyan. Other three are shown in green and yellow. They interact with the central one using their loop regions. **Right panel** shows the triangular shape of the solvent channels formed by crambin molecules. The neighboring modules of 24 crambin molecules (panel 2A) are shown in the different colors (yellow, magenta, cyan).

The water molecules found in the first shell of hydration in the crambin crystals are shown in **Figure 8.3.** The comparison of our I4$_1$32 structure and Teeter’s P2$_1$ crambin structure showed two generally conserved water bridges in the first shell of hydration of the crambin molecules.
Figure 8.3 Top panel shows (in stereo) the comparison of the water bridges between our I4,32 structure (Magenta) and Teeter’s P2₁ crambin structure (Yellow). The water molecules found in the racemic crambin crystals (section 8.3) are shown in green. Bottom panel shows (in stereo) the water bridges in our I4,32 structure.

Very well ordered water polygons were observed between four crambin molecules interacting through their loop regions in the I4,32 crystal, as shown in Figure 8.4. However, none of the higher ordered water structures like pentagons or hexagons were found to be conserved between the cubic and monoclinic crystal forms of crambin.
Figure 8.4. Water polygons found in the $I4_132$ crystal between four crambin molecules (left) and the structure of the water polygons (right, in stereo).

8.2 Crystal structure of D-enantiomeric crambin (D-crambin)

Introduction to crystal structure of D-crambin

Proteins found in nature are made up of L-amino acids and glycine. This results in a L-enantiomeric structure of the chiral protein molecules found throughout nature. A question was asked by Sung-Hou Kim’s group at Berkeley: “Would D-enantiomeric proteins crystallize in the same unit cell and with the same symmetry as L-enantiomeric proteins with opposite handedness?” This question was asked because all known D-enantiomeric protein crystals have different unit cell and symmetry from their L-enantiomers. Xray diffraction studies of synthetic D-monellin showed a unit cell different from that of L-monellin. Also, D-HIV protease showed a different unit cell from that of L-HIV protease. These observations were in conflict with the general notion that the D-proteins should be the exact mirror images of L-proteins, so the same crystal condition should give the same unit cell and symmetry. Binding of one or more
chiral contaminants in crystallization experiments or micro-heterogeneity of the synthetic protein construct was proposed as a culprit.

To answer this question, we chemically synthesized the D-crambin molecule and separately crystallized the D- and L- crambin under the same conditions. The crystal structure of our synthetic D-crambin shows the unit cell and symmetry identical to L-crambin.

In this crystallization experiment, we insured to synthesize highly homogeneous synthetic crambin preparations. Total synthesis of D-crambin was carried out by our optimized three peptide segments ligation protocol.\textsuperscript{[7]} The three peptide segments were chemically synthesized by ‘in situ neutralization’ Boc chemistry stepwise solid phase peptide synthesis\textsuperscript{[15]} using all D-amino acids. LC-MS analysis showed that our D-crambin molecule was highly homogeneous, and of purity comparable to our synthetic L-crambin.\textsuperscript{[6]}

Crystal growth

In the previous Section [8.1], we reported that L-crambin crystals could grow under a variety of crystallization conditions. We used the same conditions to crystallize our D-crambin molecule. The single crystals of D-crambin grew with the same shape as our L-crambin crystal under two different sets of conditions. In particular, D-crambin crystals, formed by mixing 2 µl of D-crambin (10 mg/ml) solution and 2 µl of 0.8M succinic acid at pH 6.5 were used for the X-ray data collection.

Data collection & structure determination

X-ray diffraction data were collected using synchrotron radiation at the Advanced Photon Source at the Argonne National Laboratory. For the D-crambin crystals, the same
*I4,32 space group was identified as for L-crambin crystals.* The D-crambin crystals were smaller in size (0.04*0.04*0.04 mm³) and because of their small size diffracted only to 2.1 Å with 99% completeness. The structure of the D-crambin molecule was solved by the molecular replacement method using the mirror image of our L-crambin structure as a model. The subsequent refinement procedures were carried out by CNS[11] and Refmac5 protocols.[12] **Table 8.2** summarizes the data collection and refinement statistics for our D-enantiomeric crambin. The crystal structure of D-crambin is shown in **Figure 8.5**.

**Table 8.2.** Data collection and refinement statistics

<table>
<thead>
<tr>
<th>Crystal</th>
<th>D-crambin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>I4,32 (Z=48)</td>
</tr>
<tr>
<td>Cell parameters</td>
<td>a=b=c=105.20 Å</td>
</tr>
<tr>
<td>Crystal size, mm³</td>
<td>0.07 x 0.07 x 0.07</td>
</tr>
<tr>
<td>Matthews coefficient</td>
<td>5</td>
</tr>
<tr>
<td># Proteins in asym.unit</td>
<td>1 protein (46 a.a.)</td>
</tr>
<tr>
<td>Solvent content, %</td>
<td>75.3</td>
</tr>
<tr>
<td>Beamline / Detector</td>
<td>APS-5ID / MAR225</td>
</tr>
<tr>
<td>Wavelength, Å</td>
<td>1.0000</td>
</tr>
<tr>
<td>Resolution range, Å</td>
<td>25 -2.1</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>6047</td>
</tr>
<tr>
<td>Redundancy</td>
<td>5.8</td>
</tr>
<tr>
<td>Completeness, %</td>
<td>99</td>
</tr>
<tr>
<td>R-merge&lt;sup&gt;a&lt;/sup&gt;, %</td>
<td>7.8</td>
</tr>
<tr>
<td>R factor / R free&lt;sup&gt;b&lt;/sup&gt;, %</td>
<td>18.1/18.2</td>
</tr>
<tr>
<td>No.of residues in refinement</td>
<td>46 aa+62H₂O</td>
</tr>
</tbody>
</table>

<sup>a</sup>R-merge = S<sub>hkli</sub>S<sub>hkli</sub> - <I<sub>hkli</sub>/I<sub>hkli</sub>> over i observations of a reflection hkl.

<sup>b</sup>R factor = \[ \frac{\sum |F_{\text{obs}}| - |F_{\text{calc}}|}{\sum |F_{\text{obs}}|} \], R-free is the same calculated with 5% data withheld from refinement.
Figure 8.5. Independently determined experimental crystal structures of synthetic D-crambin (left) and synthetic L-crambin (right).

The crystal structure of D-crambin showed that D-enantiomeric proteins can “in the same space group” produce the crystals with the same unit cell dimensions and symmetry as their L-enantiomers. This clearly illustrates that properly prepared D-proteins are the exact mirror image of their corresponding L-proteins, and that Kim’s suggestion of synthetic microheterogeneity (i.e. diastereomer protein contaminants) is the likely culprit for previous observations of non-symmetric crystallization behavior.

8.3 Crystal structure of racemic crambin (D&amp;L-crambin)

Introduction

An X-ray diffraction experiment only gives us information about the amplitudes of structure factors. To solve the crystal structure we need to combine the amplitude information with a correct phase. To obtain the correct phase, people have developed various techniques; these include (1) the isomorphous replacement that makes use of a
strongly diffracting heavy atom to obtain initial estimate of phase and (2) the anomalous scattering that takes advantage of heavy atom’s property to adsorb X-ray of specific wavelength. The anomalous scattering by use of seleno-Met protein constructs is currently very widely used to obtain phase in formation for protein crystals in a very practical and effective manner.

In addition to these methods, there were several attempts to obtain accurate phase by the use of racemic protein crystals that can be formed by crystallization of the D- and L-protein mixture together.[16-19] If during the crystal formation the enantiomeric proteins assemble on each other with an inversion center, we can obtain racemic crystals in a centrosymmetric space group. In a centrosymmetric crystal, the phase of the diffracted X-rays are either 0 or π, and the simplicity of the possible phase would reduce the efforts to obtain an accurate phase for each protein crystal. Preparing centrosymmetric crystals was proposed as an alternative way to obtain more accurate phase during protein structure determination.[16] Because proteins in nature are always present as the L-form, centrosymmetric protein crystals can only be obtained through co-crystallization of D- and L-forms of a protein molecule where the D-protein must be made by chemical synthesis.

We used our chemically synthesized L- and D-crambin molecules to produce racemic protein crystals with a centrosymmetric space group. As we reported in the above sections, the rigorous synthetic and analytical methods that we used insured the high quality of our crambin molecules and the enantiomeric structures of each of the D- and L-crambin molecule were insured.
Crystal growth

The conditions used for crystallization of L-crambin molecules could not be applied to produce racemic crystals. The main difficulty was the solubility of the D- and L-crambin mixture. In the first trials, we separately dissolved our synthetic D- and L-crambin in pH 8.0, 100mM HEPES buffer containing 150mM NaCl up to 10 mg/ml, as we did for the crystal formation of the individual crambin molecule. However, when we mixed D- and L-crambin solutions together, the mixture precipitated very quickly. We lowered the protein concentration by diluting the protein solution to 1 mg/ml before mixing them together. However, the crambin mixture eventually precipitated in only one day. In further crystallization experiments, we used ethanol to solublize the racemic crambin mixture.

We separately dissolved our synthetic D- and L-crambin in pH 8.0, 100mM HEPES buffer containing 150mM NaCl up to 10 mg/ml. Then, we added one volume of crambin and one volume of D-crambin into one volume of ethanol. We used pH 8, 100mM imidazole buffer containing 200mM ZnCl₂ and 10% ethanol as a crystallization buffer. This particular buffer condition was chosen after several rounds of screening for the crystallization of racemic crambin. A 2 μl aliquet of crambin solution was mixed with 2 μl of the crystallization buffer, and the protein droplets were equilibrated against 1 ml of crystallization buffer. Small single crystals were formed.

Data collection & structure determination

The crystal from the co-crystallization diffracted to 1.5 Å even though we used very small size of crystals (0.02 x 0.04 x 0.05 mm³). For the racemic crambin crystals, triclinic space group was identified, and we collected the X-ray diffraction data to 1.8 Å
with an 92% completeness (Table 8.3). Matthews coefficient calculations showed that we have two crambin molecules in the unit cell. The structure determination was carried out by the molecular replacement method in Molrep.\textsuperscript{10} We have used our L-crambin structures a model and tested both P1 and P(-1) space groups. The solution was found in the centrosymmetric space group P(-1) with one molecule in the asymmetric unit. It suggested that out triclinic crystals are formed by a racemic mixture of L- and D-crambin molecules that assemble on each other about an inversion center. The subsequent refinement procedures were carried out in P(-1) space group by CNS\textsuperscript{11} and Refmac5\textsuperscript{12} protocols. Table 8.3 summarizes the data collection and refinement statistics for our racemic crambin.

Table 8.3. Data collection and refinement statistics for racemic crambin

<table>
<thead>
<tr>
<th>Crystal Space group</th>
<th>Racemic-crambin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell parameters</td>
<td></td>
</tr>
<tr>
<td>a=23.987 Å</td>
<td>α=112.6°</td>
</tr>
<tr>
<td>b=27.586 Å</td>
<td>β=108.5°</td>
</tr>
<tr>
<td>c=28.737 Å</td>
<td>γ=99.4°</td>
</tr>
<tr>
<td>Crystal size, mm$^3$</td>
<td>0.02 x 0.04 x 0.05</td>
</tr>
<tr>
<td>Matthews coefficient</td>
<td>1.8</td>
</tr>
<tr>
<td># Protein in asym.unit</td>
<td>1 protein (46 a.a.)</td>
</tr>
<tr>
<td>Solvent content, %</td>
<td>30.3</td>
</tr>
<tr>
<td>Beamline / Detector</td>
<td>APS-5ID / MAR225</td>
</tr>
<tr>
<td>Wavelength, Å</td>
<td>1.0757</td>
</tr>
<tr>
<td>Resolution range, Å</td>
<td>25 -1.80 (could diffract &gt;1.5)</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>5308</td>
</tr>
<tr>
<td>Redundancy</td>
<td>2</td>
</tr>
<tr>
<td>Completeness, %</td>
<td>92.3</td>
</tr>
<tr>
<td>R-merge$^a$, %</td>
<td>10.7 (~5 at low resolution)</td>
</tr>
<tr>
<td>R factor / R free$^b$, %</td>
<td>25.1/25.8</td>
</tr>
<tr>
<td>No.of residues in refinement</td>
<td>46 aa+54H$_2$O</td>
</tr>
</tbody>
</table>

$^a$R-merge = $S_{hkl}S_{I(hkl)} - <I(hkl)>/S_{hkl}$, $<I(hkl)>$ over $i$ observations of a reflection hkl.

$^b$R factor = $\sum |F(\text{obs})| - |F(\text{calc})|/\sum|F(\text{obs})|$. 

163
R-free is the same calculated with 5% data withheld from refinement

The crystal structure of the racemic crambin showed very tight packing between L- and D-crambin (Figure 8.6).

**Figure 8.6.** *(Left panel)* Crystal packing of racemic crambin. D- and L-crambin molecules are shown in yellow and cyan, respectively. **Right panel** shows a stereo projection of one unit cell viewed along the crystal b-axis. Two enantiomeric crambin molecules are tightly packed to each other and generate a dimer via a short β-sheet formed by four H-bonds.

Comparison of the crystal packing of this racemic protein (space group = P(1), solvent content = 30%), our L-crambin (I4,32, 75%), and previously reported Teeter’s L-crambin structure (P21, 32%) is shown in Figure 8.7. The crystal packing of the racemic crystals is similar to the L-crambin crystal of P21 symmetry group rather than our L-crambin crystal structure with I4,32 space group.
Figure 8.7. Comparison of crystal packing of this racemic protein (P(-1), left), our L-crambin (I4,32, middle), and previously reported Teeter’s L-crambin structure (P21, right)

The superposition of three crystal structures of L-crambin are shown in Figure 8.8

Figure 8.8. Superposition (in stereo) of the crystal structures of L-crambin molecules from the racemic (P(-1), green), cubic (I4,32, pink), and previously reported Teeter’s
monoclinic (P2₁, yellow) crystals. The root-mean-square-difference (rmsd) value for the comparison of all main chain atoms at corresponding positions between the L-protein from the racemic crystal (P(-1), green) and our L-crambin (I4,32, pink) was 0.68Å.

During the crystallization experiments we found that the solubility of the racemic mixture in aqueous buffer (less than 1mg/ml) became extremely low as compared to L- or D- crambin alone (10mg/ml). Kim’s racemic monellin was also crystallized using lower protein concentration (2mg/ml) than its L- or D-constructs (10mg/ml). Berg’s crystal formation of racemic rubreoxin (13mg/ml) had to be slowed down by controlling pH and temperature. This trend is also shown in racemic peptide mixtures. Eisenberg group reported that mixing of D- and L- helical peptide turned the mixture to gel. Our observation for the racemic crambin is consistent with these observations, and it leads to an interesting question: what is the molecular mechanism for the low solubility of the racemic mixtures? If L- and D- proteins make dimer and bury hydrophobic surface of the proteins in aqueous solution, then the overall hydrophilic surface area of the proteins should be increased. The increased hydrophilic surface area should contribute to increase the solubility of the proteins. This contradictory observation remains to be understood.

Conclusion.

We crystallized and determined X-ray crystal structures of L-, D-, and D&L-crambin. We found various conditions to crystallize our L-crambin molecules. We obtained a new cubic crystal form of crambin. The cubic crystal structure of L-crambin showed unusually high water content. We also showed that the D-crambin crystal had
exactly the same unit cell dimensions and symmetry as the L-crambin crystal. This unambiguously demonstrated that a D-protein structure is the exact mirror image of the corresponding L-protein structure. Our racemic crambin crystal structure showed a centrosymmetry. The solubility of the racemic protein mixture was very low compared to either L- or D- crambin solution. This low solubility seems to follow general trend for the racemic mixture of proteins, but the origin of low solubility is not understood.

References


Chapter 9. How Protein Adopts D-amino Acid to Its Structure: Total Chemical Synthesis and X-ray Crystal Structure of a Protein Diastereomer: [D-Gln$^{35}$]Ubiquitin

Abstract: We explored how a native protein would incorporate a D-amino acid residue into its overall architecture. To definitively illustrate the local and global conformational perturbations of a protein molecule by D-amino acid incorporation, we decided to synthesize and determine the X-ray crystal structure of a chemically engineered globular protein, ubiquitin containing a D-amino acid residue in the place of Gly$^{35}$ in the C-cap region of an $\alpha$-helix. In order to produce large quantities of high quality protein materials for X-ray crystallography, we developed an efficient synthetic route to a protein diastereomer [D-Gln$^{35}$]ubiquitin by combining ‘one-pot’ three peptide segment native chemical ligation and protein desulfurization for chemical protein synthesis. High resolution X-ray crystal structures showed a striking conservation of molecular structure between the folded conformations of the [D-Gln$^{35}$]ubiquitin and the native ubiquitin molecule.
Proteins found in nature contain polypeptide chains made up of L-amino acids and glycine. The sequence of amino acids in the polypeptide chain defines the folded ‘tertiary’ structure of the protein molecule, and the protein owes its biological activity to that folded structure. To investigate the role of amino acid chemical structure in the formation and stability of the tertiary structure of a protein molecule, non-natural amino acids have been used to design and construct proteins that are not found in nature. In particular, substitution of D-amino acid residues into the protein molecule has been explored for its effects on the binding of insulin to insulin receptors. More recently, D-amino acid substitution has been proposed for enhancing protein stability, and the structure-function relationships in the K+ channel selectivity filter have been explored by incorporation of a D-Ala residue in place of a Gly residue. However, despite their potential importance for dissecting protein structure-function relationships, there is little detailed knowledge of how the incorporation of a D-amino acid residue affects the local and global conformations that define the folding, stability and function of the protein molecule.

We set out to understand how a native protein would incorporate a D-amino acid residue into its overall architecture. To definitively explore the local and global conformational perturbations of a protein molecule by D-amino acid incorporation, we decided to synthesize and determine the X-ray crystal structure of a chemically engineered globular protein, ubiquitin (76 amino acids) containing a D-amino acid residue in the place of Gly35 in the C-cap region of an α-helix. The conformation of the Gly35 residue (ϕ = 81°, ψ = 5°) in native ubiquitin is allowed only for an L-amino acid in a left-handed α-helix, or for a D-amino acid residue. Here we report: (1) an efficient
strategy for the total chemical syntheses of ubiquitin; (2) direct observation of the conservation of amino acid configuration after protein desulfurization by Raney-Ni reduction (i.e. L-Cys to L-Ala); (3) high resolution crystal structures for wild type ubiquitin and for a ubiquitin protein diastereomer, [D-Gln\textsuperscript{35}]UBQ; and, (4) a striking similarity between the molecular structures of wild type ubiquitin and the ubiquitin protein diastereomer. The significance of these results is discussed.

A continuing theme of our research is the investigation of protein folding and stability using chemistry. With that objective in mind, we set out to establish an efficient total synthesis by modern chemical methods of the model protein ubiquitin and its various analogues.\textsuperscript{[5]} We sought to use the recently developed ‘one-pot’ ligation method\textsuperscript{[9]} for the covalent assembly of three unprotected peptide segments. However, the ubiquitin molecule does not have the cysteine residues that are needed for native chemical ligation\textsuperscript{[10]} (the amino acid sequence of ubiquitin is shown in Scheme 9. 1 (a)). We noted that human erythrocytic ubiquitin has two alanine residues, at positions 28 and 46, suitably located for use as ligation sites (Swiss-Prot accession number: P62988).\textsuperscript{[4]} Thus, we adopted a protein desulfurization strategy\textsuperscript{[1]} that enables the use of native chemical ligation at Cys\textsuperscript{28} and Cys\textsuperscript{46}, after which the cysteine residues are converted to the native Ala\textsuperscript{28} and Ala\textsuperscript{46} residues. Our synthetic strategy is shown in Scheme 9. 1 (b).
Scheme 9.1 (a) Target amino acid sequence for human erythrocytic [Met1Leu]ubiquitin.
(b) Peptide segments and strategy used for the total chemical synthesis of ubiquitin

Data for the synthesis of wild type ubiquitin are shown in Figure 9.1. The C-terminal peptide, and the peptide-“thioesters were made by manual stepwise Boc-chemistry ‘in situ neutralization’ solid phase peptide synthesis methods. Several hundred milligrams of high purity peptide were obtained by preparative reverse phase HPLC from each 0.4mmol scale synthesis. The recovered peptide yields were ~30 % for
(1-27)-thioester and (Cys^{46}-76), and ~45 % for (Thz^{38-45})-thioester, based on starting aminoacyl-resin. Using these three peptides, we performed a one-pot, three segment ligation on a tens-of-milligrams scale: the first ligation of (Thz^{38-45})-thioester and (Cys^{46}-76) was performed in pH 6.8 phosphate buffer containing 6M guanidinium hydrochloride and 1% (v/v) thiophenol, at a concentration of 3 mg mL^{-1} for each peptide (Figure 9.1 (a) to (b)); 0.2 M methoxyamine•hydrochloride was added to convert the Thz- to the Cys-peptide ligation product at pH 4.0 (Figure 9.1 (c)); readjustment of the solution to pH 6.8 was followed by addition of the third peptide segment, (1-27)-thioester, to effect the second ligation (Figure 9.1 (d) to (e)). The ligated full-length 76 residue polypeptide was purified by reverse phase HPLC (Figure 9.1 (f)) and lyophilized for the subsequent Raney nickel desulfurization reaction.
Figure 9.1. Data for synthesis of wild type ubiquitin. Reactions were monitored by LC-MS. The UV profile at 214 nm is shown, together with electrospray mass spectrometry data (inset) corresponding to each major product. The chromatographic separations were
performed using a linear gradient (10-50%) of buffer B in buffer A over 20 min (buffer A = 0.1% TFA in water; buffer B = 0.08% TFA in acetonitrile). See text for a detailed
description of each of the panels, a)-g). Observed and calculated masses of the synthetic
intermediates and final products were: b) observed mass (ob.) 5623.1±0.5 Da, calculated
mass using average isotopes (cd.) 5623.4 Da; c) ob. 5611.6±0.7 Da, cd. 5611.4 Da; e) ob.
8610.2±1.0 Da, cd. 8610.9 Da; g) ob. 8546.0±1.1 Da, cd. 8546.8 Da.

Desulfurization with Raney nickel has been applied to proteins to convert cysteine
residues to alanine residues.[13] In model studies on acylated amino acids, no change in
the absolute configuration at the α-carbon of cysteine/alanine is observed.[14,15] In order to
extend the use of native chemical ligation to the chemical synthesis of proteins without
Cys, the Dawson group adopted the use of Raney nickel reduction subsequent to native
chemical ligation at Cys.[11] One downside of the Raney nickel reduction is that the
desulfurization not only affects cysteine residues but also (more slowly) results in the
demethylthiolation of the methionine side chain.[11] The first residue of ubiquitin is a
methionine, so we simply replaced Met¹ in ubiquitin with Leu¹ (Scheme 9. 1 (a).[16] We
prepared the Raney nickel as described.[11] nickel acetate (600 mg) was dissolved in 15 ml
of distilled water, and then sodium borohydride (100 mg) was slowly added. The black
nickel precipitate was collected by filtration and then washed extensively with distilled
water. We carried out the conversion of Cys²⁸ and Cys⁴⁶ to Ala²⁸ and Ala⁴⁶ in the synthetic
polypeptide by adding the freshly prepared nickel precipitate to the peptide solution (~ 20
mg of purified 76 amino acid polypeptide, at a concentration of 3 mg of lyophilized
peptide per 1ml 0.1 M phosphate buffer containing 6 M guanidinium hydrochloride). We
observed that oxidized cysteine residues retarded the rate of the reaction (the same observation has been made previously[13]). To insure the fully reduced state of the two cysteine residues, we added 10 mM TCEP to our reaction mixture. The final pH of the reaction mixture was 6.0. Raney nickel reduction was complete in 12 hours and a mass decrease of 64 Daltons was observed as expected for the loss of two sulfurs (Figure 9.1 (g)). Essentially identical results were obtained for total chemical synthesis of the [D-Gln\textsuperscript{35}]UBQ protein molecule (Figure 9.2). To confirm the folded structure of chemically synthesized ubiquitin, lyophilized polypeptide was dissolved under native conditions and circular dichroism spectra consistent with the native protein structure were obtained (data not shown). To obtain folded protein at the high concentrations necessary for crystallization, the synthetic polypeptide (20 mg/ml) was dissolved in phosphate buffer containing 6 M guanidinium hydrochloride, and dialyzed against distilled water.

Figure 9.2. HPLC profile of purified UBQ(D-Gln) at 214 nm and electrospray mass spectrometry data (inset) corresponding to the product peak (observed mass 8617.4±0.5 Da; calculated mass 8617.9 Da (average isotopes)).
We performed several syntheses of both wild type ubiquitin and the [D-Gln\textsuperscript{35}]UBQ diastereomer protein in order to produce enough protein material for crystal screening and for the optimization of conditions for the growth of high quality crystals. Using the protocols described above, we averaged a 36 % recovered yield up to the preparation of the full-length polypeptide containing Cys\textsuperscript{28} and Cys\textsuperscript{46} (Figure 9.1 (a)-(f)). Recovered yields for the desulfurization reaction averaged 75% (Figure 9.1 (f)-(g)). Ten to twenty milligram quantities of ubiquitin wild type and diastereomer were produced in each synthesis. Overall synthetic yields averaged 27 % based on peptides used. Besides the excellent yields of high quality protein constructs obtained from the synthetic strategy reported here, we could take advantage of the multi-segment strategy for ready synthesis of analogues. In other words, we could reuse the peptides (1-27)-\textsuperscript{\textdagger}thioester and (Cys\textsuperscript{46}-76), and it was only necessary to remake the short [Gly\textsuperscript{35D-Gln}]\textsuperscript{\textdagger}(Thz\textsuperscript{28}−45)peptide-\textsuperscript{\textdagger}thioester for the synthesis of the diastereomer analogue protein.

At first, we considered a protein NMR experiment to investigate the local and global structure of the ubiquitin diastereomer. However, we concluded that high-resolution crystallography would be ideal to explore even small configuration changes caused by D-Gln\textsuperscript{35} incorporation into the protein molecule, and to definitively show local and global structural changes of a protein diastereomer compared to the wild type protein molecule. Ubiquitin is reported to be difficult to crystallize.\textsuperscript{[4, 17-19]} Since the success of the first crystallization,\textsuperscript{[17]} ubiquitin crystal formation has not been reported under the original conditions.\textsuperscript{[4, 17-19]} Instead, most of the X-ray crystallography-quality ubiquitin crystals have been produced by seeding with samples of the original crystals. We carried out extensive screening of crystallization conditions for ubiquitin, but without seeding we
were not able to obtain crystals. However, we were aware that solid-state NMR researchers were using ubiquitin nano-crystals to study protein structures.[20] We carefully examined the conditions reported for the production of nano-crystals, and noticed the use of cadmium salts that are not usually included in commercially available crystal screening kits. Using cadmium salts in combination with the crystal screening kits, we were able to grow two different forms of ubiquitin crystals under different conditions.

Subsequent rounds of optimization led to X-ray crystallography-quality crystals for both the wild type and the diastereomer variant of ubiquitin. Crystal formation could be reproducibly obtained by mixing 2 µl of ubiquitin (20 mg/ml) solution and 2 µl of a solution containing 50 mM CdCl$_2$, 20 % PEG MME 2000 (w/v) in 0.1 M Bis-Tris buffer, pH 6.5.

X-ray diffraction data were collected using synchrotron radiation from the Advanced Photon Source at the Argonne National Laboratory. For both the wild type and diastereomer crystals, cubic (P4$_3$32) and orthorhombic (P2$_1$2$_1$2$_1$) space groups were identified. In particular, the orthorhombic crystals were highly ordered and diffracted well to 1.3 Å (a complete data set was obtained to only 1.5Å because of the use of a rectangular CCD detector). The structures of ubiquitin wild type and diastereomer protein molecules were solved by the molecular replacement method, and subsequent refinement procedures were carried out by CNS and refmac protocols (see Table 9.1 for data collection and refinement statistics for both wild type ubiquitin and [D-Gln$^{15}$]ubiquitin).
Table 9.1 Data collection and refinement statistics

<table>
<thead>
<tr>
<th></th>
<th>Gly-ubiquitin</th>
<th>D-Gln-ubiquitin</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crystal</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P2_12_1</td>
<td>P2_12_1</td>
</tr>
<tr>
<td>Cell parameters</td>
<td>a=43.78 Å, b=50.48 Å, c=92.65 Å</td>
<td>a=43.78 Å, b=50.48 Å, c=92.65 Å</td>
</tr>
<tr>
<td>Crystal size, mm³</td>
<td>0.05 x 0.05 x 0.2</td>
<td>0.05 x 0.05 x 0.2</td>
</tr>
<tr>
<td>Matthews coefficient</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td># Proteins in asym.unit</td>
<td>3 proteins (3x76 = 228 a.a.)</td>
<td></td>
</tr>
<tr>
<td>Solvent content, %</td>
<td>41.9</td>
<td></td>
</tr>
<tr>
<td>Beamline / Detector</td>
<td>APS-5ID / MAR225</td>
<td>APS-17ID / ADSC Q210</td>
</tr>
<tr>
<td>Wavelength, Å</td>
<td>0.97926</td>
<td></td>
</tr>
<tr>
<td>Resolution range, Å</td>
<td>25 -1.40</td>
<td>25 -1.30</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>27222</td>
<td>41144</td>
</tr>
<tr>
<td>Redundancy</td>
<td>5</td>
<td>3.7</td>
</tr>
<tr>
<td>Completeness, % resolution range:</td>
<td>25 -1.75</td>
<td>97.7</td>
</tr>
<tr>
<td></td>
<td>1.75 -1.64</td>
<td>58.5</td>
</tr>
<tr>
<td></td>
<td>1.64 -1.55</td>
<td>39.6</td>
</tr>
<tr>
<td></td>
<td>1.55 -1.40</td>
<td>19.1</td>
</tr>
<tr>
<td></td>
<td>1.40 -1.30</td>
<td>-</td>
</tr>
<tr>
<td>R-merge a, %</td>
<td>7.1</td>
<td>7.4</td>
</tr>
<tr>
<td>R factor / R free b, %</td>
<td>19.3/22.0</td>
<td>20.9/24.6</td>
</tr>
<tr>
<td>No.of residues in refinement</td>
<td>71+ 71+73=215 aa</td>
<td>71+ 71+73=215 aa</td>
</tr>
<tr>
<td></td>
<td>195 HOH</td>
<td>266 HOH</td>
</tr>
<tr>
<td></td>
<td>14 CD²</td>
<td>14 CD²</td>
</tr>
<tr>
<td>Rmsd bonds (Å) / angles (°)</td>
<td>0.021 / 1.9</td>
<td>0.018/1.75</td>
</tr>
<tr>
<td>Average B-factor, Å²</td>
<td>7.4</td>
<td>7.8</td>
</tr>
</tbody>
</table>

a) R-merge = Σₙₖλₙ₀_I(hkl)_i - <I(hkl)>/Σₙₖλₙ₀ <I(hkl)> over i observations of a reflection hkl.

b) R factor = Σ | |F(obs)| - |F(calc)| /Σ|F(obs)|,

R-free is the same calculated with 5% data withheld from refinement.

The wild-type synthetic ubiquitin protein molecule had the same structure as previously reported.[4] The ubiquitin diastereomer incorporated D-Gln⁵ into its overall fold without perturbing the global architecture of the protein molecule (Figure 9.3). The stereochemical configuration of the D-Gln⁵ residue was directly confirmed by the high
resoultion electron density map of the ubiquitin diastereomer, in which the side chain of D-Gln$^{35}$ could be seen pointing in the opposite direction to that expected for the side chain of an L-Gln residue (Figure 9.4). Both Gly$^{35}$ and D-Gln$^{35}$ occupied a left-handed conformation ($\phi=80^\circ$ for Gly$^{35}$ and $\phi=77^\circ$ for Gln$^{35}$; see Table 9.2).

**Figure 9.3.** Superposition of the molecular structures of the three native and three [d-Gln35]ubiquitin protein molecules found in the unit cell for the two different proteins.
**Top panel:** the backbone of Ca-atom traces of the two proteins are shown superimposed; [d-Gln35]ubiquitin (blue); native ubiquitin (red). **Bottom panel:** close up of nine residues near d-Gln35/ Gly 35 (arrow).

![Figure 9.4](image)

**Figure 9.4.** **Left panel:** structure of the [D-Gln35]ubiquitin diastereomer; the sequence Glu34-D-Gln35-Ile36 is highlighted in yellow. **Right panel:** stereoview of the local conformation near the D-Gln35 residue; the residues Glu34-D-Gln35-Ile36 are shown fitted to the 2Fo-Fc electron-density map contoured at 1σ.

**Table 9.2** $\phi$, $\varphi$ angle in each molecule

<table>
<thead>
<tr>
<th></th>
<th>$\phi$ angle</th>
<th>$\varphi$ angle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gly_m1</td>
<td>77°</td>
<td>13°</td>
</tr>
<tr>
<td>Gly_m1</td>
<td>80°</td>
<td>16°</td>
</tr>
<tr>
<td>Gly_m1</td>
<td>82°</td>
<td>12°</td>
</tr>
<tr>
<td>D-Gln_m1</td>
<td>86°</td>
<td>9°</td>
</tr>
<tr>
<td>D-Gln_m1</td>
<td>74°</td>
<td>33°</td>
</tr>
<tr>
<td>D-Gln_m1</td>
<td>72°</td>
<td>34°</td>
</tr>
</tbody>
</table>

**Abbreviations:**
Gly_m1: the first wild-type ubiquitin molecule in unit cell
Gly_m2: the second wild-type ubiquitin molecule in unit cell
Gly_m3: the third wild-type ubiquitin molecule in unit cell
D-Gln_m1: the first [D-Gln35]ubiquitin molecule in unit cell
D-Gln_m2: the second [D-Gln35]ubiquitin molecule in unit cell  
D-Gln_m3: the third [D-Gln35]ubiquitin molecule in unit cell  

Comparison of wild type and diastereomer structures showed a striking similarity in both local and global conformations (Figure 9.3). The root-mean-square-difference (rmsd) value for the comparison of the Cα atoms at corresponding positions in the two protein structures was 0.39Å, and the rmsd value for all corresponding main chain atoms in the local region (from residue 31 to 37) near the D-Gln$^{35}$ substitution was 0.24Å (see Table 9.3). These values are within the experimental variation expected for two proteins of identical structure. This striking similarity (1) validates D-amino acid replacement of natural amino acid residues occupying left-handed conformations in native protein molecules, and (2) provides experimental support for the assumption that replacement of a Gly residue, occupying a left-handed conformation, by a D-amino acid does not perturb the local and global conformations of the protein molecule.

Table 9.3 Root mean square differences  
(There are three ubiquitin molecules in a unit cell)

===================================================================
Global Fold:  
all CA atoms in residues 1 to 71 are used for superposition  
===================================================================

<table>
<thead>
<tr>
<th></th>
<th>Gly_m1</th>
<th>Gly_m2</th>
<th>Gly_m3</th>
<th>D-Gln_m1</th>
<th>D-Gln_m2</th>
<th>D-Gln_m3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gly_m1</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly_m2</td>
<td>0.26</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly_m3</td>
<td>0.61</td>
<td>0.57</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-Gln_m1</td>
<td>0.13</td>
<td>0.28</td>
<td>0.64</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-Gln_m2</td>
<td>0.30</td>
<td>0.12</td>
<td>0.58</td>
<td>0.30</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>D-Gln_m3</td>
<td>0.65</td>
<td>0.57</td>
<td>0.26</td>
<td>0.66</td>
<td>0.57</td>
<td>-</td>
</tr>
</tbody>
</table>

===================================================================
Local Fold near residue number 35:  
all main chain atoms in residues 31 to 37 are used for superposition  
===================================================================

182
The conservation of L-configuration of Ala\textsuperscript{28} and Ala\textsuperscript{46} after the desulfurization reaction could be directly observed in the electron density map of the synthetic proteins (Figure 9.5). This conservation of L-configuration during Raney nickel reduction further validates the synthetic strategy that the Dawson group devised to make cysteine-free proteins, using native chemical ligation followed by a desulfurization reaction.\textsuperscript{[11]}

Figure 9.5. Stereoview showing Ala46 fitted to the 2Fo-Fc electron-density map contoured at 1s. The conservation of configuration in the desulfurization reaction (L-Cys to L-Ala) is apparent.

In summary, in order to produce large quantities of high quality protein materials for X-ray crystallography, we developed an efficient synthetic route to cysteine-free ubiquitin molecules by combining state-of-the-art methods (‘one-pot’ three peptide segment native chemical ligation\textsuperscript{[9]}, and protein desulfurization\textsuperscript{[11]}) for chemical protein
synthesis. Using high resolution X-ray crystallography, we have experimentally verified a striking conservation of molecular structure between the folded conformations of a protein diastereomer and the corresponding wild type protein. This provides experimental support for the assumption that replacement by a D-amino acid of a Gly residue that occupies a left handed conformation does not significantly perturb the local or global conformations of a protein molecule. We are currently using the efficient ubiquitin synthesis and crystallographic methods reported here, in combination with modern biophysical techniques, to precisely dissect the molecular basis of stability in this model protein.

References


[5] The stepwise solid phase synthesis of crystalline ubiquitin has been previously reported. See references [6]-[8].


[16] Leu\(^1\) is the next most abundant position #1 residue in ubiquitins found throughout nature.


Chapter 10. Using Protein Diastereomers to Probe the Chemical Basis of Protein Stability: How an $\alpha$-Helix Terminates.

Summary: We used chemical protein synthesis to study how an $\alpha$-helix terminates. In order to separate conformational effects from desolvation effects, we constructed protein molecules containing a series of D-amino acids in the C’ position at the C-terminus of an $\alpha$-helix (C-capping region). The C’ residue in the C-cap region of a protein invariably has $\phi$, $\psi$ angles characteristic of a left-handed conformation. Only D-amino acids and Gly do not suffer significant energy penalties for adopting such a left-handed conformation. Protein molecule that contains a D-amino acid at the C’-cap position will only be affected by energetic contributions from reduced solvation (interaction with solvent water) of the peptide backbone. By contrast, the energetic penalty for substituting an L-amino acid at the C’-cap position will be the sum of conformational and desolvation effects. We used the protein ubiquitin as a model system in which to study this question, by substituting D-amino acids for the wild-type Gly$^{35}$ at the C’-cap position. Ubiquitin protein molecules containing D-amino acids are protein diastereomers and can only be efficiently prepared by chemical means. We intended to subtract the desolvation energy ($\Delta\Delta G$ [D-aa35]ubiquitin) from the sum of conformation and desolvation effects measured on the protein prepared with traditional mutagenesis methods (i.e. containing the corresponding L-amino acids at C’-cap position) ($\Delta\Delta G$ [L-aa35]ubiquitin). In this way, we successfully dissected the energetic basis of an important aspect of $\alpha$-helix C-capping. Analysis from calorimetric measurements shows that (1) the preference for the C’ Gly termination is based on a combination of a major conformation effect and a minor solvation effect; (2)
an L-amino acid with a β-branched side chain at C’ position affects the conformational energetics two times more than a C’ L-amino acid having no β-branching in the side chain. To support our arguments, we report the crystal structures of a series of chemically synthesized ubiquitin analogues showing minimal perturbations in the local structure.
**Introduction and Design of the experiment.**

The α-helix is one of the most basic structural motifs of protein molecules. The structure was first predicted by Pauling from model building based on known chemical bonding parameters and the planarity of the peptide bond, and the formation of reasonable hydrogen bonds between amide hydrogen and carbonyl oxygen in the backbone of a peptide chain.\(^1, 2\) However, the exact chemical basis for how the α-helix terminates (i.e. as it changes to a non-helix conformation) is still unclear because the helical structure does not provide satisfactory hydrogen bond partners at the C-terminal end.\(^3\) The controversy is particularly focused on a Gly residue that predominantly occupies the C’ position of the C-cap region of an α-helix; some say that only Gly adopts a left-handed conformation without paying an significant energetic penalty;\(^4\) others argue that because of the Gly at the α-helix terminus the peptide backbone is solvent exposed with favorable solvation energetic effects.\(^5, 6\) The substitution of the C’ Gly by other amino acids causes desolvation of peptide backbone, resulting in decreased stability. L-amino acids will also perturb the hydrogen bond between the amide hydrogen of the C’ Gly and the carbonyl oxygen of the fourth residue prior to the C’ position Gly.

Although this question has been studied in several proteins,\(^6, 7\) it is still not completely resolved due to the limitations of protein mutagenesis techniques that can only incorporate L-amino acids. Incorporation of L-amino acids at C’ results in thermodynamic effects that are the sum of conformation and desolvation effects.

Chemical protein synthesis is an attractive alternative to overcome the intrinsic limitation of cell-based protein engineering methods.\(^8\) We can incorporate any chemical
moieties into a protein structure by the chemical synthesis of appropriate peptide building blocks. In particular, for the study of C-capping energetics, incorporation of D-amino acids into a protein molecule at the C’ position would resolve the current question about the physical basis of this aspect of the C-cap, in the following way. D-amino acids and Gly are allowed a left-handed conformation. We reasoned that a protein molecule with a D-amino acid at the C-capping C’ position will be only affected by energetic contribution from the changes in the solvation of the peptide backbone, depending on the size and nature of the side chain at that position. Thus, preparation of a series of protein molecules that incorporate different D-amino acids at the C’ position of an α-helix, and comparison of effects with the corresponding L-amino acid, will be a unique way to separate conformational effects from solvation effects.

To accomplish our goal, we decided to incorporate a series of D-amino acids in place of the Gly$^{35}$ residue at the C’ position of an α-helix in the model protein ubiquitin. Ubiquitin is a widely used model to understand protein folding and protein stability.$^{9}$ Previously, C-capping thermodynamics of ubiquitin had been studied by the use of L-amino acids at the Gly$^{35}$ position incorporated into the ubiquitin molecule by recombinant DNA mutagenesis methods.$^{6}$ Energetic effects for the (L-amino acid)$^{35}$-substituted ubiquitins were measured, and the desolvation effect was thought to be the main reason for terminating the α-heilx with a C’ Gly residue.

**Synthesis of the protein diastereomers.**

Ubiquitin molecules with a series of D-amino acids at the C’ position (i.e. [D-amino acid]$^{35}$)ubiquitins were prepared by ‘one-pot’$^{10}$ native chemical ligation of three
peptide synthetic segments, followed by desulfurization\textsuperscript{111} of the full-length polypeptide with Raney nickel. The target amino acid sequence, ligation sites, and peptide segments used are shown in Figure 10.1. Our chemical protein synthesis methods rely on the thioester-mediated ligation of unprotected peptides at Cys residues.\textsuperscript{112} Ubiquitin has no Cys residues. For this reason, we used the method of Dawson\textsuperscript{111} in which suitably located Ala residues are substituted by Cys residues in the peptide building blocks. Desulfurization converts these Cys residues, used for synthetic convenience in the native chemical ligation method, to wild-type Ala residues.

\begin{enumerate}
\item \textbf{a)}
\begin{verbatim}
LQIFVKTLTG KITLEVEPS DTIENVKAKI QDKEGIPPDQ QRLIFAGKQL EDGRTLSDYNIQKESTLHLVLRLRGG76
\end{verbatim}
\item \textbf{b)}
\begin{verbatim}
LQIFVKTLTGKITLEVEPSDTIENVK-thioester ThzKIQDKEGIPPDQQRLLF-thioester CysGKQEDGRTLSDYNIQKESTLHLVLRLRGG
\end{verbatim}
\end{enumerate}

\textbf{Figure 10.1} (a) Target amino acid sequence for human erythrocytic [Met1Leu]ubiquitin. (b) Peptide segments used for the total chemical synthesis of ubiquitin

Four ubiquitin diastereomers, that incorporated D-Ala\textsuperscript{35}, D-Gln\textsuperscript{35}, D-Thr\textsuperscript{35} and D-Val\textsuperscript{35}, were prepared as described\textsuperscript{113} in Chapter 9. As well, we prepared synthetic ubiquitins with Gly\textsuperscript{35}, L-Gln\textsuperscript{35}, L-Val\textsuperscript{35} as controls. The high quality of the synthetic molecules was verified by LC-MS analyses (electrospray mass spectrometry analysis for each ubiquitin molecule is shown in Figure 10.2)
Figure 10.2
(d) $\text{[D-Thr}^{35}\text{]ubiquitin (8590.6Da.)}$

(e) $\text{[D-Val}^{35}\text{]ubiquitin (8588.1Da.)}$

(f) $\text{[L-Gln}^{35}\text{]ubiquitin (8617.6Da)}$

Figure 10.2. continued.
Figure 10.2. continued. Electrospray mass spectrometry analysis for each chemically synthesized ubiquitin molecule is shown. Mass peaks of [M+8H]$^{8+}$, [M+7H]$^{7+}$, [M+6H]$^{6+}$, and [M+5H]$^{5+}$ are shown for each spectrum. (a) [Gly35]ubiquitin, observed mass (ob.) 8546.0±1.0 Da, calculated mass using average isotopes (cd.) 8546.8 Da; (b) [D-Ala35]ubiquitin, ob. 8560.2±0.7 Da, cd. 8560.8 Da; (c) [D-Gln35]ubiquitin, ob. 8617.4±0.5 Da, cd. 8617.9 Da; (d) [D-Thr35]ubiquitin, ob. 8590.6±0.6 Da, cd. 8590.8 Da; (e) [D-Val35]ubiquitin, ob. 8588.1±0.9 Da, cd. 8588.8 Da; (f) [L-Gln35]ubiquitin, ob. 8617.6±0.7 Da, cd. 8617.9 Da; (g) [L-Val35]ubiquitin ob. 8588.2±0.8 Da, cd. 8588.8 Da.

Stability measurements.

Thermodynamic data (Table 10.1) from differential scanning calorimetric experiments (DSC) shows the effects of D-amino acid incorporation at the Gly$^{35}$ on the stability of ubiquitin. Protein concentration was measured and the standard thermodynamic functions under reference conditions were calculated as described in elsewhere.$^{[6]}$ The differences between the folding free energies of a [(D-amino acid)$^{35}$]ubiquitin and the wild-type ubiquitin ($\Delta\Delta G_{D,aa}$) are are given in Table 10.1.
Column (1). Previously reported corresponding values for L-amino acid incorporated ubiquitins (ΔΔG_{L-aa}) are given in Table 10.1 Column (2).\textsuperscript{[6]}

**Table 10.1.** \(\Delta \Delta G\) values calculated from calorimetric measurements of the stabilities of synthetic ubiquitins.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)*</th>
<th>(2) – (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>([D\text{-}aa35]\Delta \Delta G) (kJ mol(^{-1}))</td>
<td>([L\text{-}aa35] \Delta \Delta G) (kJ mol(^{-1}))</td>
<td>Conformational energy penalty for L-aa(^{35}) (kJ mol(^{-1}))</td>
</tr>
<tr>
<td>Gly35 (‘wild type’)</td>
<td>= 0</td>
<td>= 0</td>
<td>-</td>
</tr>
<tr>
<td>Gln35</td>
<td>- 0.1</td>
<td>+ 4.1</td>
<td>+ 4.2</td>
</tr>
<tr>
<td>Ala35</td>
<td>+ 0.4</td>
<td>+ 4.4</td>
<td>+ 4.0</td>
</tr>
<tr>
<td>Thr35</td>
<td>+ 1.5</td>
<td>+ 9.0</td>
<td>+ 7.5</td>
</tr>
<tr>
<td>Val35</td>
<td>+ 3.5</td>
<td>+11.1</td>
<td>+ 7.6</td>
</tr>
<tr>
<td></td>
<td>Solvation only</td>
<td>Solvation +</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Conformation</td>
<td></td>
</tr>
</tbody>
</table>

*Values taken from Ref.\textsuperscript{[6]}*

**X-ray crystal structures.**

In the Chapter 9, we have shown that incorporation of D-Gln\(^{35}\) in place of the Gly\(^{35}\) residue does not perturb the local and global conformation of the ubiquitin molecule.\textsuperscript{[13]} We however were concerned that the least stable D-Val\(^{35}\) incorporated ubiquitin ([D-Val\(^{35}\)]ubiquitin) might change the local structure at the C-capping position, so that other thermodynamic energies could be involved for the overall stability of ubiquitin. To obviate this possibility, we crystallized chemically engineered [D-Val\(^{35}\)]ubiquitin. As well, we crystallized and solved structure of the [L-Gln\(^{35}\)]ubiquitin as a control.
Comparison among crystal structure of [D-Val\textsuperscript{35}]ubiquitin and previously determined wild-type ubiquitin and [D-Gln\textsuperscript{35}]ubiquitin showed high structural conservation (Figure 10.3). The structure [L-Gln\textsuperscript{35}]ubiquitin adopted the same local conformation as the wild-type and the diastereomeric ubiquitin. (Figure 10.3) The unperturbed backbone conformation observed in the [L-Gln\textsuperscript{35}]ubiquitin crystal structure validates our argument that L-amino acids have to pay energetic penalty to occupy a left-handed conformation in the protein molecule (see below). Superposition of the molecular structures of ubiquitin protein molecules are shown in Figure 10.3. We superimposed the backbone of Cα-atom traces of the five ubiquitin proteins (15 structures; see Figure 10.3 legend). We used previously reported synthetic ubiquitin structures \cite{13} as well for the superimposition, and these structures are denoted with * in the Figure. Data collection and refinement statistics for ubiquitin molecules including [D-Val\textsuperscript{35}]ubiquitin, [L-Gln\textsuperscript{35}]ubiquitin, and [D-Gln\textsuperscript{35}]ubiquitin with P4_32 space group are shown in Table 10.2.
Figure 10.3 Superposition of the molecular structures of the ubiquitin protein molecules.

Top panel: side stereoviews of the backbone of Cα-atom traces of the five proteins (15 molecules) are superimposed; native ubiquitin (three molecules from P2₁,2₁,2₁ space group*, gray). [D-Gln35]ubiquitin (three molecules from P2₁,2₁,2₁*, red & two molecules from P4₃,32 space group, pink); [D-Val35]ubiquitin (two molecules from P4₃,32, green); [L-Gln35]ubiquitin (two molecules from P4₃,32, yellow). Middle panel: top stereoviews
of the backbone of Cα-atom traces of the five proteins are superimposed. **Bottom panel:** close up of ten residues (residue 30-39) near residue 35 (arrow marked on Cα atom). Stereoviews of the backbone of main chain atom traces of the five proteins are superimposed; native ubiquitin (one molecule from P2₁₂,₂₁*, gray); [D-Gln35]ubiquitin (one molecule from P4₁₃₂, pink); [D-Val35]ubiquitin (one molecule from P4₁₃₂, green); [L-Gln35]ubiquitin (one molecule from P4₁₃₂, yellow). The * denotes the ubiquitin structures of P2₁₂,₂₁ space group that are taken from our previously reported structures.[¹³]

---

**Table 10.2.** Data collection and refinement statistics for ubiquitin molecules with P4₁₃₂ space group

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P4₁₃₂</td>
<td>P4₁₃₂</td>
<td>P4₁₃₂</td>
</tr>
<tr>
<td>Cell parameters</td>
<td>a=b=c=105.53 Å,</td>
<td>a=b=c=106.15 Å,</td>
<td>a=b=c=106.58 Å,</td>
</tr>
<tr>
<td>Crystal size, mm³</td>
<td>0.8x0.8x0.8</td>
<td>0.7x0.7x0.7</td>
<td>0.9x0.9x0.9</td>
</tr>
<tr>
<td>Matthews coefficient</td>
<td>3.2</td>
<td>3.3</td>
<td>3.3</td>
</tr>
<tr>
<td># Protein in asym.unit</td>
<td>2 protein (76 a.a.)</td>
<td>2 protein (76 a.a.)</td>
<td>2 protein (76 a.a.)</td>
</tr>
<tr>
<td>Solvent content, %</td>
<td>61%</td>
<td>62%</td>
<td>62%</td>
</tr>
<tr>
<td>Beamline / Detector</td>
<td>APS-5ID / MAR225</td>
<td>APS-5ID / MAR225</td>
<td>APS-5ID / MAR225</td>
</tr>
<tr>
<td>Wavelength, Å</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>Resolution range, Å</td>
<td>20 -2.2</td>
<td>20 -2.2</td>
<td>20 -1.8</td>
</tr>
<tr>
<td>Number of reflections</td>
<td>10092</td>
<td>10299</td>
<td>15995</td>
</tr>
<tr>
<td>Completeness, %</td>
<td>99.55</td>
<td>99.81</td>
<td>99.92</td>
</tr>
<tr>
<td>R factor / R free, %</td>
<td>25.3%/25.7%</td>
<td>25.8%/26%</td>
<td>21.8%/22%</td>
</tr>
<tr>
<td>No.of residues in refinement</td>
<td>72+73 aa</td>
<td>72+73 aa</td>
<td>72+73 aa</td>
</tr>
</tbody>
</table>

**Separation of conformation and solvation effects.**

If we assume that ΔΔG_{D-aa} values contain energetic contribution only from solvation effects, and that ΔΔG_{L-aa} values are due to both solvation and conformational effects as discussed above, the difference values between ΔΔG_{D-aa} and ΔΔG_{L-aa} corresponds to the energetic contribution from the conformational effect: these are shown
in Table 10.1 (3rd column: (2)-(1)). The data are summarized in Figure 10.4. We can see at a glance that the conformational penalty for placing an L-amino acid at the C’ position (cream bars) is much larger than the solvation effect (red bars). This is contrary to previous interpretations based on L-amino acids and Gly only\cite{5,6}, without access to D-amino acid-containing proteins provided by chemical synthesis.

![Bar chart](image)

**Figure 10.4.** Folding free energy values; $\Delta \Delta G_{L, aa}$ values (blue bar, the energetic contribution from the sum of both solvation and conformational effects), $\Delta \Delta G_{D, aa}$ values (red bar, the energetic contribution from the solvation effect), the difference values between $\Delta \Delta G_{D, aa}$ and $\Delta \Delta G_{L, aa}$ (cream bar, the energetic contribution from the conformational effect)
Discussion and Conclusion.

Our data shows that the Gly preference at the C’ position of the C-capping region of an α-helix is mainly due to the ability of a Gly residue, unique among genetically encoded amino acids, to adopt a left-handed conformation without excessive energetic penalty. We also observed the effect of β-branching in an amino acid side chain on the conformational energetics. That is, our data clearly shows that there are two classes of L-amino acids at the C’-cap position in terms of the conformational energetic penalties paid: the conformational effect from amino acids with a β-branched side chain was 7.5 kJ mol⁻¹, and for the amino acids without β-branching was 4.0 kJ mol⁻¹. This is in accord with known effects of β-branching on the conformational energetic effects of the left handed conformation (ϕ ~ minus 80 degrees, ψ ~ +35 degrees): β-branching has a significant unfavorable effect of about the same magnitude as the +3.5kJ mol⁻¹ that we observed.¹⁴ This quantitative agreement validates our assumption that there is little or no conformational penalty for placing a D-amino acid at the C’ position.

Thus, our results provide a quantitative explanation for the Gly preference at the C’ position in the C-cap region of an α-helix in a protein molecule. The Gly preference is mainly due to the conformational flexibility of Gly giving it the ability to occupy a left handed conformation. L-amino acids at the C’ position pay a large conformational penalty for occupying the left handed configuration required at the C’ position. Furthermore, β-branched L-amino acids pay almost twice the energetic penalty from this conformational effect than non β-branched amino acids. Generally, desolvation effects were of small significance, but β-branched amino acids at the C’ position of C-cap region
of an \(\alpha\)-helix make minor contributions to reduced protein stability by desolvating the peptide backbone.

Our study was based on a unique combination of chemical protein synthesis, calorimetric measurements and X-ray crystallography, and unambiguously establishes a physical basis for an important aspect of C-capping the protein \(\alpha\)-helix motif, and resolves the controversy caused by the limitations of classical protein engineering techniques.

References.


Chapter 11. Summary and Significance

11.1. General conclusions on the development of improved methods for the total synthesis of proteins: *Better synthetic methods made the total synthesis of proteins easier.*

In this thesis work systematic analyses and approaches for the improved chemical protein synthesis have been carried out. As a result, an original set of chemical methods has been developed and was applied to the total synthesis of the model protein crambin to show their utility.

In the second chapter, I demonstrated that dividing a difficult-to-make peptide segment into two fragments and joining the divided peptides would improve the chemical synthesis of proteins. The chemical synthesis of crambin by ligation of two peptide segments was not straightforward due to difficulties inherent to the synthesis of one of the two segments. The difficult peptide segment was divided into two short segments. Systematic analyses to optimize the ligation of the three peptide segments were carried out, and an optimized protocol for the synthesis was developed. As a result, total synthesis of crambin by ligation of three peptide segments gave a *ten-fold* increase in yield (from 2% to 20%) compared to crambin synthesis by ligation of two peptide segments.

In the third chapter, I described the development of ‘one-pot’ protein synthesis. In the one-pot method, all the synthetic transformations are carried out in a single reaction mixture: the first ligation of two peptide segments, the conversion of Thz-peptide to Cys-peptide, the second ligation of the product Cys-peptide with the third peptide segment, and the folding & disulfide formation of the full-length polypeptide chain were carried
out simply by addition of new reagents and adjustment of the reaction conditions. A single final purification step was all that was necessary to give synthetic crambin molecule of extraordinary purity with two times improved yield (from 20% to 40%) compared to the previously optimized crambin synthesis described in the chapter 2.

In the fourth chapter, I showed that the His₆-tag could be used to assist the practical synthesis of protein molecules. At the C-terminal of the target polypeptide chain, I introduced a His₆ tag during the synthesis of the C-terminal peptide segment building block. Presence of the His₆ tag enabled the isolation of peptide or protein products directly from ligation reaction mixtures by Ni-NTA affinity column purification. This simple approach enabled facile buffer exchange to alternate reaction conditions and was compatible with direct analytical control by eletrospray mass spectrometry of the multiple ligation steps involved in protein synthesis. This His₆-tag approach combined with the one-pot approach is a highly effective approach to chemical protein synthesis by the sequential ligation of many peptide segments.

In the fifth chapter, through a series of systematic approaches I demonstrated a practical, fully convergent ligation strategy for the condensation of multiple unprotected peptide segments. A very simple and general concept, namely kinetic control of reactivity, was applied to make the convergent synthesis possible. In the initial studies, slow titration of the Cys-peptide2-thioester segment into a high concentration of another peptide1-thioester was used to obtain high yields of the desired peptide1-Cys-peptide2-thioester; this enabled an efficient synthesis of a crambin analogue of unprecedented topology (see below). An improved kinetically controlled convergent synthesis was developed based on differential reactivity of different types of thioesters. By the use of
this simple concept, I illustrated the possibility of the use of an unprotected Cys-peptide-
"thioester segment for the convergent ligation with another peptide-“thiophenylester. Robustness and practicality of this approach was illustrated by a fully convergent total synthesis of crambin from six unprotected peptide segments.

**Significance.** In all, I showed that these improved synthetic methods made the total synthesis of proteins easier: more practical, simpler, and giving higher yields of more pure product proteins. Indeed, the kinetically controlled convergent protein synthesis as well as the practical sequential ligation methods show great promise for the next level of chemical protein synthesis. I imagine that the total synthesis of proteins of typical size (~300 amino acids) would be realized by these methods soon.

11.2. General conclusions on the chemistry applied to the study of proteins:

*Proteins made by chemistry unveiled the chemical basis of protein structures.*

Chemistry played a unique role for the preparation of necessary set of protein analogues to dissect questions relevant to biology. Chemically synthesized proteins were used to improve our understanding of the role of a salt bridge in the crambin molecule, and the chemical basis of amino acid at the C’ position of the C-cap of an α-helix. In addition, I explored the synthesis and properties of a novel protein topological analogue, and the synthesis, crystallization, and Xray structure of synthetic D-crambin, synthetic L-crambin, and synthetic racemic D&L-crambin.

In the sixth chapter, I showed that the salt bridge of crambin molecule guides the formation of correct disulfide bonds. The salt bridge between the guanidinium of Arg10 side chain and the carboxylate of Asn46 at the C-terminal of the polypeptide chain was
perturbed by replacing the \(^8\)carboxylate by an \(^8\)carboxamide. Lower folding yields were obtained for the crambin-\(^8\)carboxamide. As well, I determined the crystal structure of crambin-\(^8\)carboxamide protein and in this way documented the salt bridge’s important role for the exceptional thermal stability of the overall crambin structure.

In the seventh chapter, I explored the total synthesis of a novel protein topological analogue, ‘crambin cyclotide’. Using chemistry, I successfully replaced the salt bridge between the \(^\delta\)guanidinium of Arg10 side chain and the \(^8\)carboxylate of Asn46 of crambin molecule with a covalent bond. In the presence of the artificial covalent bond, I observed that cysteine residues efficiently found their unique disulfide bond partners. Work is still continuing on the structural characterization of this unique protein molecule.

In the eighth chapter, I demonstrated that the D-enantiomeric crambin crystallized with exactly opposite handedness compared to the L-enantiomeric crambin. For the first time, this unambiguously answered a question; “would the D-enantiomeric protein crystallize in the same unit cell and with the same symmetry as the L-enantiomeric protein with opposite handedness?” In addition, I carried out studies of racemic D&L-crambin in a centrosymmetric crystal form.

In the ninth chapter, I experimentally verified a striking conservation of crystal structure between a ubiquitin protein molecule containing a D-amino acid at position 35 and the corresponding wild type ubiquitin molecule(Gly35). This provided experimental support for the generally believed assumption that replacement by a D-amino acid of a Gly residue that occupies a left handed conformation does not significantly perturb the local or global conformations of a protein molecule.
In the tenth chapter, I showed that the preference for the C’ Gly termination at the C-cap of protein α-helix is based on a combination of a major conformation effect and a minor solvation effect. In addition to this, I found that an L-amino acid with a β-branched side chain at the C’ position affects the conformational energetics two times more than a C’ L-amino acid having no β-branching in the side chain.

**Significance.** In conclusion, proteins made by chemistry unveiled novel aspects of the chemical basis of protein structure. Total synthesis, X-ray crystallography and other biophysical tools were used to understand biological questions that could not be answered because of the limitations associated with conventional biological methods.

*The combination of improved synthetic methods and applications of these methods to the study of protein folding, stability, and structure illustrates the great promise that chemical protein synthesis has for the systematic elucidation of the molecular basis of protein function.*