THE STUDY OF CAVITAND-BASED DE NOVO HELICAL BUNDLE PROTEINS

by

EMILY SATOKO SEO

B.Sc., The University of British Columbia, 2000

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Chemistry)

THE UNIVERSITY OF BRITISH COLUMBIA

August 2006

© Emily Satoko Seo, 2006
Abstract

The template assembly approach for the *de novo* design of proteins provides a useful tool for studying protein structure and folding. The template employed here is a cavitand, which is a rigid macrocycle used to organize helical bundles, resulting in a structure called a cavitein (derived from cavitand and protein).

The linker length between the individual helical peptides and the cavitand has been shown to have a dramatic effect on the structure and properties of the caviteins: this previous work studied a set of caviteins using a sequence designed to link from the hydrophobic/hydrophilic interface of the peptides. Here, a new series of four-helix bundle caviteins was synthesized using a sequence designed to link from the hydrophobic face (AEELLKKLEELLKKG). By changing the attachment point, the linker length requirement was slightly reduced, and in turn, the packing between the helices was improved. The optimal linker for this new sequence was found to be two glycine residues, as this linker resulted in a cavitein with the most native-like characteristics.
Molecular dynamics simulations were carried out on the same series of caviteins studied experimentally. The computer modelling results generally agreed with the experimental data in terms of helical content and conformational specificity. These simulation results were used to better comprehend the behaviour of the caviteins in solution.

A peptide sequence designed for a four-helix structure (CGGGEELLKKLEELLKKG) was linked onto various-sized [n]cavitands. The four-helix bundle was found to be the most stable and native-like compared to the five- and six-helix bundles, as the design intended, which shows that the same sequence exhibits different native-like properties depending on the number of helices in a bundle.

*De novo* proteins simplify the interactions involved in protein folding by allowing subtle changes in the design to be made. It has been demonstrated that slight modifications in the cavitein design have dramatic effects on their stability and structural properties. With improved understanding of how the different forces interact to influence the overall structure, it should be possible to design more complex caviteins with function.
Table of Contents

Abstract ............................................................................................................................................... ii

Table of Contents ................................................................................................................................ iv

List of Tables ........................................................................................................................................ ix

List of Figures ........................................................................................................................................ xi

List of Schemes ....................................................................................................................................... xvii

List of Abbreviations ............................................................................................................................ xviii

Acknowledgments ................................................................................................................................. xx

CHAPTER ONE: Introduction ................................................................................................................. 1

1.0 General Introduction and Overview ............................................................................................ 1
  1.0.1 The Importance of Proteins ................................................................. 1
  1.0.2 The Protein Folding Problem ............................................................ 2
  1.0.3 Methods for Studying the Protein Folding Problem ....................... 3
  1.0.4 Thesis Overview ................................................................................. 3

1.1 Peptide and Protein Structure ................................................................................................. 5
  1.1.1 General Structural Features ......................................................... 5
  1.1.2 Secondary Structure .................................................................... 6
  1.1.3 The α-Helix .................................................................................... 7
  1.1.4 α-Helical Motifs ............................................................................ 9
    1.1.4.1 Coiled Coils .......................................................................... 9
    1.1.4.2 Helical Bundles .................................................................. 11
    1.1.4.3 Four-Helix Bundles ............................................................. 12
    1.1.4.4 Five- and Six- Helix Bundles .............................................. 13
  1.1.5 Conformational Mobility of Side Groups Within Proteins: Native Proteins and Molten Globule Structures ........................................................................................................... 13

1.2 Factors Influencing Protein Structure and Folding .................................................................. 16
  1.2.1 Conformational Entropy and Non-Covalent Interactions ............. 16
    1.2.1.1 Conformational Entropy ..................................................... 16
    1.2.1.2 Electrostatic Interactions ................................................. 17
    1.2.1.3 Van der Waals Interactions ............................................ 18
    1.2.1.4 Hydrogen Bonding ......................................................... 18
    1.2.1.5 Hydrophobic Effects ....................................................... 19
  1.2.2 Factors that Influence α-Helical Structures .................................... 19
    1.2.2.1 Helical Propensity ......................................................... 20
    1.2.2.2 Helix Macrodipole Effects ............................................. 22
    1.2.2.3 Helix Capping Effects .................................................... 23
1.2.2.3.1 N-Capping ................................................. 24
1.2.2.3.2 C-Capping ................................................. 25
1.2.2.3.3 Hydrophobic Capping ..................................... 26
1.2.2.3.4 Capping Motifs ............................................ 26
1.2.2.4 Helix Chain Length Effects .................................. 27
1.2.2.5 Intrahelical Interactions ..................................... 29
1.2.3 Interhelical Interactions ........................................ 30
1.2.3.1 Packing Pattern .............................................. 30
1.2.3.2 Interhelical Electrostatic Interactions and Orientation of the Helices .......................... 31
1.2.3.3 Factors that Influence Peptide Oligomerization in Multi-Helical Proteins ..................... 32

1.3 Various Views on Protein Folding ................................... 36
1.3.1 Thermodynamic and Kinetic Control ................................ 36
1.3.2 Classical View versus New View ................................ 37
1.3.2.1 Classical View ............................................... 38
1.3.2.2 New View .................................................... 39
1.3.2.3 Current Examples of the New View and the Classical View ........................................ 40

1.4 De Novo Protein Design ............................................. 42
1.4.1 DeGrado- Minimalist and Incremental Approach ..................... 42
1.4.2 Combinatorial Techniques ........................................ 46

1.5 Template Assembled Synthetic Proteins ............................... 48
1.5.1 The Purpose and Advantages of the “Template-Assembly’ Approach ....................... 48
1.5.2 Examples of Templates ........................................... 50
1.5.2.1 Peptide-Based Templates ..................................... 50
1.5.2.2 Metal-Ligand Complex-Based Templates .......................... 51
1.5.2.3 Porphyrin-Based Templates .................................... 52
1.5.2.4 Aromatic Ring-Based Templates ................................. 54
1.5.2.5 Carbohydrate-Based Templates ................................ 56
1.5.2.6 Cholic Acid-Based Templates .................................. 58
1.5.3 Examples of Template Assembled Synthetic Proteins with Function ............................ 60

1.6 Computer Modelling ............................................... 61

1.7 Chapter Conclusion and Thesis Objectives ............................ 62

1.8 References ....................................................... 64

CHAPTER TWO: The Effect of the Linker Length Between the Template and the Peptides on the Structural Properties of the Caviteins ................................................................. 74

2.0 Introduction ......................................................... 74
2.0.1 Effect of the Linkers ............................................... 79
2.0.2 Goals for the Second Generation of Caviteins ..................................................... 87
2.0.3 Nomenclature ..................................................... 88
2.1 Results and Discussion

2.1.1 Design of the Second Generation Caviteins

2.1.1.1 Template Choice and Synthesis

2.1.1.2 Sequence Design and Synthesis

2.1.1.3 Cavitein Synthesis

2.1.2 Characterization of the Caviteins

2.1.2.1 Far UV Circular Dichroism (CD) Spectra

2.1.2.2 Near UV CD Spectra

2.1.2.3 Effect of Guanidine Hydrochloride

2.1.2.4 Oligomeric State

2.1.2.5 $^1$H Nuclear Magnetic Resonance (NMR) Spectra

2.1.2.5.1 One-Dimensional (1D) $^1$H NMR Spectroscopy

2.1.2.5.2 Two-Dimensional (2D) Homonuclear $^1$H NMR Spectroscopy

2.1.2.5.3 Hydrogen/Deuterium Exchange

2.1.2.6 ANS Binding Studies

2.2 Chapter Summary and Conclusion

2.3 Experimental

2.3.1 Arylthiol Cavitand Synthesis

2.3.1.1 General

2.3.1.2 Synthesis of the Arylthiol Cavitand

2.3.2 Peptide and Cavitein Synthesis

2.3.2.1 General

2.3.2.2 Peptide Synthesis

2.3.2.3 Thiocresol-Based Peptide OGS1

2.3.2.4 Cavitein Synthesis

2.3.3 Circular Dichroism (CD) Experiments

2.3.3.1 Far and Near UV CD Spectra

2.3.3.2 Denaturation Studies

2.3.4 Sedimentation Equilibria Studies

2.3.5 NMR Experiments

2.3.5.1 1D $^1$H NMR Spectra

2.3.5.2 2D $^1$H NMR Spectra

2.3.5.3 N-H/D Exchange 1D $^1$H NMR Spectra

2.3.6 ANS Binding Studies

2.4 References

CHAPTER THREE: Computer Modelling Study on the Second Generation Linker Caviteins

3.0 Introduction

3.1 Methods

3.1.1 Helical Content

3.1.2 Conformational Specificity
3.1.3 Perimeter Distance ........................................ 197
3.1.4 Supercoiling .............................................. 198
3.1.5 Tilt of Helices with Respect to the Cavitand ........ 199
3.1.6 Linker Elbow Orientation ................................ 200

3.2 Results and Discussion ...................................... 201
3.2.1 Helical Content .......................................... 202
3.2.2 Conformational Specificity ............................... 207
3.2.3 Relative Motion and Compactness of the Helices .... 215
3.2.4 Supercoiling .............................................. 223
3.2.5 Tilt of Helices with Respect to the Cavitand ....... 226
3.2.6 Linker Elbow Orientation ................................ 233

3.3 Chapter Summary and Conclusion ........................... 235

3.4 References ..................................................... 237

CHAPTER FOUR: The Study of Different-Sized TASP and Reversible Cavitein Systems .................................... 239

4.0 General Introduction ......................................... 239

4.1 The Study of Different-Sized Template Assembled Synthetic Proteins ............................................. 240
4.1.1 Goals ...................................................... 246
4.1.2 Nomenclature ............................................. 247

4.2 Results and Discussion ...................................... 249
4.2.1 Synthesis of the Larger TASP ............................ 249
4.2.1.1 Template Choice and Synthesis .................... 249
4.2.1.2 Sequence Choice and Synthesis ................. 249
4.2.1.3 Synthesis of the Larger Caviteins ............... 251
4.2.2 Characterization of the Different Sized TASP ....... 253
4.2.2.1 Far UV Circular Dichroism (CD) Spectra ....... 253
4.2.2.2 Near UV Circular Dichroism (CD) Spectra .... 255
4.2.2.3 Effect of Guanidine Hydrochloride ............. 256
4.2.2.4 Oligomeric State ................................... 259
4.2.2.5 1H Nuclear Magnetic Resonance (NMR) Spectroscopy ................................................. 261
4.2.2.5.1 One-Dimensional (1D) 1H NMR Spectroscopy 261
4.2.2.5.2 Hydrogen/Deuterium Exchange ............... 263
4.2.2.6 ANS Binding Studies ................................. 267
4.2.3 Study of Peptide Sequence Designed for a Five-Helix Bundle ............................................. 270
4.2.3.1 Design and Synthesis of the S5 Peptide Sequence 272
4.2.3.2 Synthesis of the Various-Sized TASP using the S5 Sequence ......................................... 274

4.3 Reversible Cavitein Systems .................................. 275

4.4 Chapter Summary and Conclusion .......................... 284
4.5 Experimental

4.5.1 Synthesis of the Cavitand Templates


4.5.1.2 Synthesis of Sodium Benzyliothiolate Cavitand

4.5.1.3 Synthesis of Sodium Arylthiolate Cavitand

4.5.1.4 Synthesis of the Phosphated Footed Benzyliothiol [4]Cavitand

4.5.2 Peptide and Cavitein Synthesis

4.5.2.1 Peptide Synthesis

4.5.2.2 Cavitein Synthesis

4.5.3 Circular Dichroism (CD) Experiments

4.5.3.1 Far and Near UV CD Spectra

4.5.3.2 Denaturation Studies

4.5.4 Sedimentation Equilibria Studies

4.5.5 NMR Experiments

4.5.5.1 1D $^1$H NMR Spectra

4.5.5.2 N-H/D Exchange 1D $^1$H NMR Spectra

4.5.6 ANS Binding Studies

4.6 References

CHAPTER FIVE: Thesis Conclusion and Future Outlook

5.0 Thesis Conclusion

5.1 Future Goals

5.1.1 Continuation of the Reversible Cavitein Systems

5.1.2 Continuation of the Effect of Different-Sized [n]Cavitands

5.1.3 Structure Determination

5.1.4 Proteins with Function

5.2 References

Appendix A. The Twenty Commonly Occurring Amino Acids

Appendix B. Computer Modelling Definitions

Appendix C. Steps for Setting up the Molecular Dynamics Simulations
List of Tables

| Table 1.1. | Stabilizing intrahelical interactions | 29 |
| Table 1.2. | GCN4-pl mutants and their oligomeric states | 33 |
| Table 1.3. | The sequences used in DeGrado's design of de novo four-helix bundles | 45 |
| Table 2.1. | Summary of the experimental results of the first generation caviteins with different types and lengths of linkers | 85 |
| Table 2.2. | First generation peptide sequence, S0, and the second generation peptide sequence, S2. The term cavitein follows the peptide sequence name when referring to the four-helix TASP | 89 |
| Table 2.3. | Fraction α-helix, $f_{\alpha}$, of the caviteins in the second generation series using the equation from the experimental section | 99 |
| Table 2.4. | Guanidine hydrochloride-induced denaturation data of the caviteins in the second generation series | 106 |
| Table 2.5. | Molecular weight estimations by sedimentation equilibria for caviteins at 20 °C in 50 mM phosphate buffer, pH 7.0, when fit to a single, ideal species. Solvent density was estimated to be 1.00 g/mL | 109 |
| Table 2.6. | The resonance assignments in ppm for the 2GS2 cavitein in 50 mM acetate buffer and 10 % D2O at pH 4.62 at 20 °C | 125 |
| Table 2.7. | The resonance assignments in ppm for the 1GS2 cavitein in 50 mM pH 4.62 acetate buffer and 10 % D2O at 20 °C | 129 |
| Table 2.8. | N-H/D exchange data on the caviteins in the second generation series in 50 mM deuterated acetate buffer, pD 5.02 at 20 °C. The calculated data are the averages of three estimates at three different times | 139 |
| Table 3.1. | Summary of the differences between the three sets of simulations for the second generation caviteins. ‘No’ and ‘Yes’ refer to if the position restraints were on during the specified step | 195 |
| Table 3.2. | Clustering results for the second generation caviteins, showing the degree of conformational spread over 20 ns for Simulation 1 | 207 |
| Table 3.3. | Clustering results for the second generation caviteins showing the degree of conformational spread over 20 ns for the 1GS2 cavitein and 5 ns for the rest of the caviteins in the second generation for Simulation 2 | 214 |
| Table 3.4. | Clustering results for the second generation caviteins showing the degree of conformational spread over 20 ns for the 1GS2 cavitein and 5 ns for the rest of the caviteins in Simulation 3. The caviteins sampled a smaller conformational space compared to the caviteins in Simulations 1 and 2 | 214 |
| Table 3.5. | Average perimeter distance ($P_x(t)$) values using the central carbonyl carbon atom, the central carbonyl nitrogen atom, and the terminal carbonyl carbon atom calculated from Simulation 1 | 216 |
| Table 3.6. | Average $P_x(t)$ values of the central carbonyl carbon atom and the terminal carbonyl carbon atom calculated from Simulation 2 and Simulation 3 | 221 |
| Table 3.7. | Qualitative degrees of tilt of the caviteins in Simulations 1, 2 and 3 | 232 |
| Table 3.8. | The linker elbow conformation of the caviteins in the second generation series | 233 |
Table 4.1. Names used for the peptide sequences (to refer to the template assembled synthetic proteins, the term TASP or cavitein follows the sequence name)...... 248

Table 4.2. Guanidine hydrochloride-induced denaturation data of the different-sized helical bundle TASPs. The fraction folded is given by normalizing the $[\theta]_{222}$ value at the different concentrations of GuHCl with the $[\theta]_{222}$ value of the completely folded, non-denatured state................................................................. 258

Table 4.3. Molecular weight estimations by sedimentation equilibria studies for the S3-[5] and S3-[6] caviteins at 20 °C in 50 mM phosphate buffer, pH 7.0 when fit to a single, ideal species. A monomer-dimer fit was also carried out for the the S3-[5] cavitein to determine the association constant, $K_a$. Solvent density was estimated to be 1.0 g/mL ................................................................. 259

Table 4.4. N-H/D exchange data on the set of different-sized helical bundle caviteins in 50 mM deuterated acetate buffer, pD 5.02 at 20 °C. ................................................................. 266

Table 4.5. Summary of the conditions used for the synthesis of a cavitein under reversible conditions. The reactions were carried out under nitrogen at rt for 8 h, except where indicated. The last two entries were carried out in the absence of any redox buffers to examine the linkage between the cavitand and the peptides.............. 279
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>(a) L-amino acid and (b) peptide with L-amino acids.</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>Ramachandran plot.</td>
</tr>
<tr>
<td>Figure 1.3</td>
<td>Right-handed α-helix.</td>
</tr>
<tr>
<td>Figure 1.4</td>
<td>(a) Two-stranded coiled coil structure and (b) helical wheel diagram.</td>
</tr>
<tr>
<td>Figure 1.5</td>
<td>Square bundle interaction.</td>
</tr>
<tr>
<td>Figure 1.6</td>
<td>Native structure versus molten globule structure.</td>
</tr>
<tr>
<td>Figure 1.7</td>
<td>α-Helix showing the backbone hydrogen bond pairs.</td>
</tr>
<tr>
<td>Figure 1.8</td>
<td>Interhelical side chain packing: (a) knobs into holes and (b) ridges into grooves.</td>
</tr>
<tr>
<td>Figure 1.9</td>
<td>Various knobs into holes packing: perpendicular, parallel and acute.</td>
</tr>
<tr>
<td>Figure 1.10</td>
<td>Protein under (a) thermodynamic control and (b) kinetic control.</td>
</tr>
<tr>
<td>Figure 1.11</td>
<td>Three classical models.</td>
</tr>
<tr>
<td>Figure 1.12</td>
<td>Folding funnel of a protein.</td>
</tr>
<tr>
<td>Figure 1.13</td>
<td>Family of DeGrado’s four-helix bundle proteins.</td>
</tr>
<tr>
<td>Figure 1.14</td>
<td>Schematic representation of the ‘template-assembly’ approach.</td>
</tr>
<tr>
<td>Figure 1.15</td>
<td>Mutter’s TASP with a peptidic template.</td>
</tr>
<tr>
<td>Figure 1.16</td>
<td>Ghadiri’s TASP using a metal-ligand template.</td>
</tr>
<tr>
<td>Figure 1.17</td>
<td>A porphyrin template used in the synthesis of a TASP.</td>
</tr>
<tr>
<td>Figure 1.18</td>
<td>Fairlie’s aromatic-based templates.</td>
</tr>
<tr>
<td>Figure 1.19</td>
<td>Methyl α-D-Galp template.</td>
</tr>
<tr>
<td>Figure 1.20</td>
<td>TASP’s derived from (a) Glcp and (b) Altp templates.</td>
</tr>
<tr>
<td>Figure 1.21</td>
<td>Cholic acid with maleimide functionalized groups.</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>Resorcinarene with R₁ = feet and R₂ = rim.</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>(a) Arylthiol and (b) benzythiol cavitands.</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>Potential hydrogen bonding interactions between the amide hydrogen of a residue, and the bridged oxygen of the arylthiol cavitand (left) or the arylthiol sulfur (right).</td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>Effect of the linker length between the template and peptides.</td>
</tr>
<tr>
<td>Figure 2.5</td>
<td>Helical wheel diagram of the peptide sequence, S₀ = EELLKKLEELLKKG, forming a four-helix bundle. Helices are oriented in a parallel fashion. Viewer is looking down the helical axes from C- to N-termini.</td>
</tr>
<tr>
<td>Figure 2.6</td>
<td>Methylene variant, 4(CH₂) S₀ cavitein versus glycine variant, 1GS₀ cavitein. Note both caviteins have the same number of atoms between the sulfur atom of the cavitand and the peptide.</td>
</tr>
<tr>
<td>Figure 2.7</td>
<td>Helical wheel diagram of the sequence, S₂ = AEELLKKLEELLKKG. Helices are oriented in a parallel fashion to form the four-helix bundle. Viewer is looking down the helical axes from C- to N-termini.</td>
</tr>
<tr>
<td>Figure 2.8</td>
<td>CD spectra of the second generation cavitein series. Samples are 40 μM in 50 mM phosphate buffer, pH 7.0 at 25 °C.</td>
</tr>
<tr>
<td>Figure 2.9</td>
<td>Near UV CD spectra of the second generation caviteins. Samples are 40 μM in pH 7.0 phosphate buffer at 25 °C.</td>
</tr>
<tr>
<td>Figure 2.10</td>
<td>Guanidine hydrochloride-induced denaturation curves for the second generation caviteins and the 0GS2 peptide, all at 40 μM concentrations in pH 7.0 phosphate buffer at 25 °C.</td>
</tr>
</tbody>
</table>
Figure 2.11. Expanded amide regions (9.5 to 5.5 ppm) of the second generation caviteins (1.5 mM) in pH 7.0 phosphate buffer and 10 % D<sub>2</sub>O at 25 °C on a 600 MHz spectrometer. 112

Figure 2.12. Expanded aliphatic regions (4.5 to 0.0 ppm) of the second generation caviteins (1.5 mM) in pH 7.0 phosphate buffer and 10 % D<sub>2</sub>O at 25 °C on a 600 MHz spectrometer. 114

Figure 2.13. 2D COSY spectrum of 2 mM 2GS2 cavitein in 50 mM pH 4.62 acetate buffer and 10 % D<sub>2</sub>O at 20 °C, expanded in the amide region. 118

Figure 2.14. 2D NOESY spectrum of 2 mM 2GS2 cavitein in 50 mM pH 4.62 acetate buffer and 10 % D<sub>2</sub>O at 20 °C, expanded in the amide region. 120

Figure 2.15. 2D NOESY spectrum of 2 mM 2GS2 cavitein in 50 mM pH 4.62 acetate buffer and 10 % D<sub>2</sub>O at 20 °C, expanded in the aliphatic/amide region. 122

Figure 2.16. 2D COSY spectrum of 2 mM 2GS2 cavitein in 50 mM pH 4.62 acetate buffer and 10 % D<sub>2</sub>O at 20 °C, expanded in the alpha/amide region. 124

Figure 2.17. 1D <sup>1</sup>H NMR spectrum of the 2GS2 cavitein in 50 mM pH 4.62 acetate buffer and 10 % D<sub>2</sub>O at 20 °C, expanded in the amide region. Residues are labelled based on information from the 2D NMR spectra. 126

Figure 2.18. 2D NOESY spectrum of 1.1 mM 1GS2 cavitein in 50 mM pH 4.62 acetate buffer and 10 % D<sub>2</sub>O at 20 °C, expanded in the amide region. 127

Figure 2.19. 2D TOCSY spectrum of 1.1 mM 1GS2 cavitein in 50 mM pH 4.62 acetate buffer and 10 % D<sub>2</sub>O at 20 °C, expanded in the alpha/amide region. 128

Figure 2.20. 1D <sup>1</sup>H NMR spectrum of the 1GS2 cavitein in pH 4.62 acetate buffer and 10 % D<sub>2</sub>O at 20 °C, expanded in the amide region. Residues are labelled based on information from the 2D NMR spectra. 130

Figure 2.21. Stack plot of 1D <sup>1</sup>H NMR N-H/D exchange spectra for 2.2 mM 1GS2 cavitein in 50 mM deuterated acetate buffer at pH 5.02 on a 500 MHz NMR Varian Unity spectrometer at 20 °C. The reference spectrum was taken in 50 mM pH 4.62 acetate buffer and 10 % D<sub>2</sub>O. 132

Figure 2.22. Stack plot of 1D <sup>1</sup>H NMR N-H/D exchange spectra for 1.9 mM 2GS2 cavitein in 50 mM deuterated acetate buffer at pH 5.02 on a 500 MHz NMR Varian Unity spectrometer at 20 °C. The reference spectrum was taken in 50 mM pH 4.62 acetate buffer and 10 % D<sub>2</sub>O. 133

Figure 2.23. Stack plot of 1D <sup>1</sup>H NMR N-H/D exchange spectra for 1.9 mM 4GS2 cavitein in 50 mM deuterated acetate buffer at pH 5.02 on a 500 MHz NMR Varian Unity spectrometer at 20 °C. The reference spectrum was taken in 50 mM pH 4.62 acetate buffer and 10 % D<sub>2</sub>O. 135

Figure 2.24. Stack plot of 1D <sup>1</sup>H NMR N-H/D exchange spectra for 1.9 mM 3GS2 cavitein in 50 mM deuterated acetate buffer at pH 5.02 on a 500 MHz NMR Varian Unity spectrometer at 20 °C. The reference spectrum was taken in 50 mM pH 4.62 acetate buffer and 10 % D<sub>2</sub>O. 137

Figure 2.25. Fluorescence emission spectra of 2 μM ANS in the presence of 100 % methanol, 95 % ethanol, and 50 μM of each cavitein: 1GS2, 2GS2, 3GS2 and 4GS2 in 50 mM phosphate buffer, pH 7.0 at 25 °C. 141

Figure 2.26. Distance discrepancy between the cavitation template and helices of the four-helix bundle protein when the effective helix diameter is on the lower end (a) and higher end (b). Linkage is represented by a dot for the first generation caviteins, and by a star for the second generation caviteins. 146
Figure 2.27. Schematic representation of the FastMoc™ protocol on the ABI 431A peptide synthesizer.

Figure 2.28. The purification of the 3GS2 cavitein monitored by analytical reversed-phase HPLC using a gradient of 30 to 60% acetonitrile (with 0.05% TFA) in water (with 0.1% TFA) over 50 minutes: (a) after the first purification (pre-peak still remains) and (b) after the second purification (pre-peak is removed).

Figure 2.29. Sedimentation equilibrium analysis of 20 μM 1GS2 cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 27,000 rpm. The solid line in the upper plot represents the theoretical fit to a monomer-dimer in equilibrium. The lower plot represents the residuals to the fit.

Figure 2.30. Sedimentation equilibrium analysis of 20 μM 1GS2 cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 35,000 rpm. The solid line in the upper plot represents the theoretical fit to a monomer-dimer in equilibrium. The lower plot represents the residuals to the fit.

Figure 2.31. Sedimentation equilibrium analysis of 20 μM 1GS2 cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 35,000 rpm. The solid line in the upper plot represents the theoretical fit to a monomer in equilibrium. The lower plot represents the residuals to the fit, showing the scatter of data points around the best-fit line (note that the residuals are unevenly spread when the 1GS2 cavitein is forced to be a monomer).

Figure 2.32. Sedimentation equilibrium analysis of 40 μM 2GS2 cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 27,000 rpm. The solid line in the upper plot represents the theoretical fit to a single non-interacting species. The lower plot represents the residuals to the fit.

Figure 2.33. Sedimentation equilibrium analysis of 40 μM 2GS2 cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 35,000 rpm. The solid line in the upper plot represents the theoretical fit to a single non-interacting species. The lower plot represents the residuals to the fit.

Figure 2.34. Sedimentation equilibrium analysis of 40 μM 2GS2 cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 40,000 rpm. The solid line in the upper plot represents the theoretical fit to a single non-interacting species. The lower plot represents the residuals to the fit.

Figure 2.35. Sedimentation equilibrium analysis of 20 μM 3GS2 cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 27,000 rpm. The solid line in the upper plot represents the theoretical fit to a monomer-dimer in equilibrium. The lower plot represents the residuals to the fit.

Figure 2.36. Sedimentation equilibrium analysis of 20 μM 3GS2 cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 35,000 rpm. The solid line in the upper plot represents the theoretical fit to a monomer-dimer in equilibrium. The lower plot represents the residuals to the fit.

Figure 2.37. Sedimentation equilibrium analysis of 20 μM 3GS2 cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 40,000 rpm. The solid line in the upper plot represents the theoretical fit to a monomer-dimer in equilibrium. The lower plot represents the residuals to the fit.
Figure 2.38. Sedimentation equilibrium analysis of 20 μM 4GS2 cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 27,000 rpm. The solid line in the upper plot represents the theoretical fit to a single non-interacting species. The lower plot represents the residuals to the fit.

Figure 2.39. Sedimentation equilibrium analysis of 40 μM 4GS2 cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 27,000 rpm. The solid line in the upper plot represents the theoretical fit to a single non-interacting species. The lower plot represents the residuals to the fit.

Figure 2.40. Sedimentation equilibrium analysis of 60 μM 4GS2 cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 27,000 rpm. The solid line in the upper plot represents the theoretical fit to a single non-interacting species. The lower plot represents the residuals to the fit.

Figure 3.1. Dimensions measured using the ‘helipad’ program: height (H) and range (R).

Figure 3.2. Ramachandran plot of the 1GS2 cavitein.

Figure 3.3. Ramachandran plot of the 2GS2 cavitein.

Figure 3.4. Ramachandran plot of the 3GS2 cavitein.

Figure 3.5. Ramachandran plot of the 4GS2 cavitein.

Figure 3.6. Average energy of the 3 most predominant clusters as a function of simulation time for Simulation 1 of the 1GS2 cavitein: Cluster 0 (73%), Cluster 1 (18%) and Cluster 2 (6%).

Figure 3.7. Average energy of the 3 most predominant clusters as a function of simulation time for Simulation 1 of the 2GS2 cavitein: Cluster 0 (40%), Cluster 1 (27%) and Cluster 2 (16%).

Figure 3.8. Average energy of the 3 most predominant clusters as a function of simulation time for Simulation 1 of the 3GS2 cavitein: Cluster 0 (31%), Cluster 1 (28%) and Cluster 2 (26%).

Figure 3.9. Average energy of the 3 most predominant clusters as a function of simulation time for Simulation 1 of the 4GS2 cavitein: Cluster 0 (54%), Cluster 1 (42%) and Cluster 2 (4%).

Figure 3.10. Normalized histogram from Simulation 1 of the sum of the interhelical distances between the carbonyl carbons of the central leucine.

Figure 3.11. Normalized histogram from Simulation 1 of the sum of the interhelical distances between the nitrogen atoms of the central leucine.

Figure 3.12. Normalized histogram from Simulation 1 of the sum of the interhelical distances between the carbonyl carbons of the terminal glycine.

Figure 3.13. Normalized histogram from Simulation 3 of the sum of the interhelical distances between the carbonyl carbons of the central leucine.

Figure 3.14. Normalized histogram from Simulation 3 of the sum of the interhelical distances between the carbonyl carbons of the terminal glycine.

Figure 3.15. Coiled-coil phase yield (Δωr) per residue number for cluster 0 of 1GS2, 2GS2, 3GS2 and 4GS2 from Simulation 1, calculated using the TWISTER program.

Figure 3.16. 1GS2 height distances from the sulphur plane to the α-carbon of alanine (Ala 2) for all four helices in Simulation 1, calculated using the helipad program.

Figure 3.17. 1GS2 range distances from the sulphur plane to the α-carbon of alanine (Ala 2) for all four helices in Simulation 1.

Figure 3.18. 2GS2 height distances from the sulphur plane to the α-carbon of alanine (Ala 3) for all four helices in Simulation 1.
Figure 3.19. 2GS2 range distances from the sulphur plane to the α-carbon of alanine (Ala 3) for all four helices in Simulation 1.

Figure 3.20. 3GS2 height distances from the sulphur plane to the α-carbon of alanine (Ala 4) for all four helices in Simulation 1.

Figure 3.21. 3GS2 range distances from the sulphur plane to the α-carbon of alanine (Ala 4) for all four helices in Simulation 1.

Figure 3.22. 4GS2 height distances from the sulphur plane to the α-carbon of alanine (Ala 5) for all four helices in Simulation 1. Height distances may not be significant due to the longer linkers from the template to the helical bundle.

Figure 3.23. 4GS2 range distances from the sulphur plane to the α-carbon of alanine (Ala5) for all four helices in Simulation 1. All four helices show fairly sharp peaks with the exception of helix A.

Figure 4.1. Templates used to synthesize a three helix bundle (a) cyclotribenzylene (CTB), and a four-helix bundle (b) benzylthiol [4]cavitand.

Figure 4.2. Energy minimized structures of [n]cavitands obtained using MM2 force field: (a) [4]cavitand, (b) [5]cavitand, (c) [6]cavitand.

Figure 4.3. Byproduct of the S3-[5] cavitin synthesis.

Figure 4.4. CD spectra of the S3-[5] and S3-[6] cavitins. Each sample contains 40 μM protein in 50 mM phosphate buffer, pH 7.0 at 25 °C.

Figure 4.5. Near UV spectra of the S3-[5] and S3-[6] cavitins. Each sample contains 40 μM protein in 50 mM phosphate buffer, pH 7.0 at 25 °C.

Figure 4.6. Guanidine hydrochloride-induced denaturation curves for the S3-[5] and S3-[6] cavitins, both 40 μM concentrations in pH 7.0 phosphate buffer at 25 °C.

Figure 4.7. Hypothetical arrangement of the helices for the (a) six-helix bundle cavitin and the (b) five-helix bundle cavitin.

Figure 4.8. Expanded amide regions of the 1D 1H NMR spectra of 0.57 mM S3-[5] and 0.42 mM S3-[6] cavitins, each in 50mM phosphate buffer, pH 7.0 and 10 % D2O at 25 °C.

Figure 4.9. Expanded aliphatic regions of the 1D 1H NMR spectra of 0.57 mM S3-[5] and 0.42 mM S3-[6] cavitins, each in 50mM phosphate buffer, pH 7.0 and 10 % D2O at 25 °C.

Figure 4.10. Stack plot of the 1D 1H NMR N-H/D exchange spectra for 0.3 mM S3-[4] cavitin in 50 mM deuterated acetate buffer, pD 5.02 on a 500 MHz NMR Varian Unity spectrometer at 20 °C. The reference spectrum was taken in 50 mM acetate buffer, pH 4.62 and 10 % D2O.

Figure 4.11. Stack plot of the 1D 1H NMR N-H/D exchange spectra for 0.3 mM S3-[5] cavitin in 50 mM deuterated acetate buffer, pD 5.02 on a 500 MHz NMR Varian Unity spectrometer at 20 °C. The reference spectrum was taken in 50 mM acetate buffer, pH 4.62 and 10 % D2O.

Figure 4.12. Stack plot of the 1D 1H NMR N-H/D exchange spectra for 0.3 mM S3-[6] cavitin in 50 mM deuterated acetate buffer, pD 5.02 on a 500 MHz NMR Varian Unity spectrometer at 20 °C. The reference spectrum was taken in 50 mM acetate buffer, pH 4.62 and 10 % D2O.

Figure 4.13. Fluorescence emission spectra of 50 μM of S3-[4], S3-[5], S3-[6] TASPs, 100 % methanol and 95 % ethanol in the presence of 2 μM ANS in pH 7.0, 50 mM phosphate buffer at 25 °C.

Figure 4.14. Dimensions of the [n]cavitands in Å.
Figure 4.15. Possible cavity formed within the hydrophobic core of the five-helix bundle... 270
Figure 4.16. Helical wheel diagram of a five-helix bundle using the peptide sequence......... 273
Figure 4.17. Guanidine hydrochloride-induced denaturation curves of the S3-[4] and S4-[4] caviteins......................................................... 277
Figure 4.18. Sedimentation equilibrium analysis of 20 µM S3-[5] cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 27,000 rpm. The solid line in the upper plot represents the theoretical fit to a monomer-dimer in equilibrium. The lower plot represents the residuals to the fit........................................... 279
Figure 4.19. Sedimentation equilibrium analysis of 20 µM S3-[5] cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 35,000 rpm. The solid line in the upper plot represents the theoretical fit to a monomer-dimer in equilibrium. The lower plot represents the residuals to the fit........................................... 289
Figure 4.20. Sedimentation equilibrium analysis of 20 µM S3-[5] cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 27,000 rpm. The solid line in the upper plot represents a poor fit to a monomer. The lower plot represents the residuals to the fit, showing the uneven distribution of data points around the best-fit line.............................................................................. 301
Figure 4.21. Sedimentation equilibrium analysis of 20 µM S3-[6] cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 27,000 rpm. The solid line in the upper plot represents the theoretical fit to a single non-interacting species. The lower plot represents the residuals to the fit.......................................................... 302
Figure 4.22. Sedimentation equilibrium analysis of 20 µM S3-[6] cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 35,000 rpm. The solid line in the upper plot represents the theoretical fit to a single non-interacting species. The lower plot represents the residuals to the fit.......................................................... 303
Figure 4.23. Sedimentation equilibrium analysis of 20 µM S3-[6] cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 40,000 rpm. The solid line in the upper plot represents the theoretical fit to a single non-interacting species. The lower plot represents the residuals to the fit.......................................................... 304
Figure 5.1. Structure of ANS (left) and quinine (right)................................................................................. 319
List of Schemes

Scheme 1.1. General depiction of the protein folding problem ................................................. 2
Scheme 2.1. Four-helix cavitein synthesis .................................................................................... 79
Scheme 2.2. Synthesis of model de novo proteins ....................................................................... 81
Scheme 2.3. Synthesis of the methyl-footed arylthiol cavitand .................................................. 91
Scheme 2.4. Thiocresol reacts with peptide sequence S1 ................................................................. 93
Scheme 2.5. Synthesis of the Second Generation Caviteins ......................................................... 96
Scheme 4.1. Synthesis of the [n]cavitands .................................................................................... 243
Scheme 4.2. Activation of the peptide by reaction with 2,2'-dipyridyl disulfide (DPDS) .............. 250
Scheme 4.3. Synthesis of the S3-[5] and S3-[6] caviteins ................................................................. 251
Scheme 4.4. Synthesis of the sodium arylthiolate cavitand .......................................................... 278
Scheme 4.5. Reaction of arylthiol cavitand with the activated peptide in a solvent mixture of tris buffer and DMF ......................................................................................... 280
Scheme 4.7. Synthesis of a four-helix cavitein in tris buffer, pH 8.7, under nitrogen ................. 283
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>å</td>
<td>angstroms</td>
</tr>
<tr>
<td>AIBN</td>
<td>2,2'-azo-bis-isobutyronitrile</td>
</tr>
<tr>
<td>ANS</td>
<td>1-anilinonaphthalene-8-sulfonate</td>
</tr>
<tr>
<td>Calcd</td>
<td>calculated</td>
</tr>
<tr>
<td>CCl₄</td>
<td>carbon tetrachloride</td>
</tr>
<tr>
<td>CD</td>
<td>circular dichroism</td>
</tr>
<tr>
<td>CHCl₃</td>
<td>chloroform</td>
</tr>
<tr>
<td>cm</td>
<td>centimetres</td>
</tr>
<tr>
<td>COSY</td>
<td>correlation spectroscopy</td>
</tr>
<tr>
<td>CTB</td>
<td>cyclotribenzylene</td>
</tr>
<tr>
<td>d</td>
<td>day(s)</td>
</tr>
<tr>
<td>DCC</td>
<td>dynamic covalent chemistry</td>
</tr>
<tr>
<td>DCL</td>
<td>dynamic combinatorial library</td>
</tr>
<tr>
<td>DCM</td>
<td>dichloromethane</td>
</tr>
<tr>
<td>DDP</td>
<td>di-t-butyl N,N-diethylphosphoramidite</td>
</tr>
<tr>
<td>DHB</td>
<td>2,5-dihydroxybenzoic acid</td>
</tr>
<tr>
<td>DIPEA</td>
<td>diisopropylethylamine</td>
</tr>
<tr>
<td>DMF</td>
<td>N,N-dimethylformamide</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethyl sulfoxide</td>
</tr>
<tr>
<td>DMA</td>
<td>N,N-dimethylacetamide</td>
</tr>
<tr>
<td>DPDS</td>
<td>2,2'-dipyridyl disulfide</td>
</tr>
<tr>
<td>DQF</td>
<td>double quantum filter</td>
</tr>
<tr>
<td>ESI</td>
<td>electrospray ionization</td>
</tr>
<tr>
<td>ESMS</td>
<td>electrospray ionization mass spectrometry (or spectrum)</td>
</tr>
<tr>
<td>Et</td>
<td>ethyl (-CH₂CH₃)</td>
</tr>
<tr>
<td>EtOAc</td>
<td>ethyl acetate</td>
</tr>
<tr>
<td>EtOH</td>
<td>ethanol</td>
</tr>
<tr>
<td>fₓH</td>
<td>fraction α-helix</td>
</tr>
<tr>
<td>GSH</td>
<td>glutathione (reduced)</td>
</tr>
<tr>
<td>GSSG</td>
<td>glutathione (oxidized)</td>
</tr>
<tr>
<td>GuHCl</td>
<td>guanidine hydrochloride</td>
</tr>
<tr>
<td>h</td>
<td>hour(s)</td>
</tr>
<tr>
<td>HBTU</td>
<td>2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate</td>
</tr>
<tr>
<td>HPLC</td>
<td>high performance liquid chromatography</td>
</tr>
<tr>
<td>HOBT</td>
<td>1-hydroxybenzotriazole</td>
</tr>
<tr>
<td>HOHAHA</td>
<td>homonuclear Hartmann-Hahn</td>
</tr>
<tr>
<td>HRMS</td>
<td>high resolution mass spectrometry (or spectrum)</td>
</tr>
<tr>
<td>IR</td>
<td>infrared</td>
</tr>
<tr>
<td>J</td>
<td>coupling constant</td>
</tr>
<tr>
<td>LSIMS</td>
<td>liquid secondary ionization mass spectrometry (or spectrum)</td>
</tr>
<tr>
<td>m/z</td>
<td>mass-to-charge ratio</td>
</tr>
<tr>
<td>M</td>
<td>parent mass (mass spectra) or Molarity (concentration)</td>
</tr>
<tr>
<td>MALDI</td>
<td>matrix laser desorption ionization</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>MD</td>
<td>molecular dynamics</td>
</tr>
<tr>
<td>Me</td>
<td>methyl (-CH₃)</td>
</tr>
<tr>
<td>MeOH</td>
<td>methanol</td>
</tr>
<tr>
<td>min</td>
<td>minute(s)</td>
</tr>
<tr>
<td>MM</td>
<td>molecular mechanics</td>
</tr>
<tr>
<td>mon</td>
<td>month(s)</td>
</tr>
<tr>
<td>ms</td>
<td>millisecond(s)</td>
</tr>
<tr>
<td>MS</td>
<td>mass spectrometry (or spectrum)</td>
</tr>
<tr>
<td>NaOMe</td>
<td>sodium methoxide</td>
</tr>
<tr>
<td>n-BuLi</td>
<td>n-butyllithium</td>
</tr>
<tr>
<td>NBS</td>
<td>N-bromosuccinimide</td>
</tr>
<tr>
<td>NMP</td>
<td>N-methylpyrrolidone</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>NOESY</td>
<td>nuclear Overhauser enhancement spectroscopy</td>
</tr>
<tr>
<td>ns</td>
<td>nanosecond(s)</td>
</tr>
<tr>
<td>1D</td>
<td>one-dimensional</td>
</tr>
<tr>
<td>pep.</td>
<td>peptide</td>
</tr>
<tr>
<td>Ph</td>
<td>phenyl (-C₆H₅)</td>
</tr>
<tr>
<td>ppm</td>
<td>parts per million</td>
</tr>
<tr>
<td>ps</td>
<td>picosecond(s)</td>
</tr>
<tr>
<td>rel intensity</td>
<td>relative intensity</td>
</tr>
<tr>
<td>RMS</td>
<td>root mean square</td>
</tr>
<tr>
<td>Rop</td>
<td>repressor of primer</td>
</tr>
<tr>
<td>rpm</td>
<td>revolutions per minute</td>
</tr>
<tr>
<td>rt</td>
<td>room temperature</td>
</tr>
<tr>
<td>[θ]</td>
<td>molar ellipticity per residue</td>
</tr>
<tr>
<td>t</td>
<td>tertiary</td>
</tr>
<tr>
<td>t₁/₂</td>
<td>half-life</td>
</tr>
<tr>
<td>TBDPS</td>
<td>t-butyl diphenylsilyl</td>
</tr>
<tr>
<td>TFA</td>
<td>trifluoroacetic acid</td>
</tr>
<tr>
<td>THF</td>
<td>tetrahydrofuran</td>
</tr>
<tr>
<td>TLC</td>
<td>thin layer chromatography</td>
</tr>
<tr>
<td>TOCSY</td>
<td>total correlation spectroscopy</td>
</tr>
<tr>
<td>tris</td>
<td>tris(hydroxymethyl)aminomethane</td>
</tr>
<tr>
<td>2D</td>
<td>two-dimensional</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>VWF</td>
<td>von Willebrand factor</td>
</tr>
<tr>
<td>W</td>
<td>watt(s)</td>
</tr>
</tbody>
</table>
Acknowledgments

I would like to thank my supervisor, John Sherman, for his guidance throughout my graduate studies, and for asking questions that taught me to think more deeply both scientifically and non-scientifically. I would like to extend my gratitude to Suzana Straus and Walter Scott for their guidance and assistance, particularly with the computer modelling work.

I wish to thank the past and present members of the Sherman group, especially Diana Wallhorn, who taught me the ropes when I first started, and Heidi Huttunen for the fun times in and outside the lab. I would also like to thank Jon Freeman and Jay Read for proofreading parts of my thesis, and Ali Izadi-Najafabadi for his helpful hints.

This thesis employed a number of biophysical techniques that required specific expertise. I would like to thank Mark Okon for acquiring the NMR spectra and sharing his NMR knowledge, Elena Polishchuk for her help with the equipment in the biological services, and Les Burtnick for sharing his knowledge on protein characterizing techniques.

Finally, thank you to my family and friends, especially my parents for their unconditional love and support over the years, and Jakub Wasiela for his continual encouragement and believing in me.
CHAPTER ONE: Introduction

1.0 General Introduction and Overview

Proteins are vital for the existence of every living species. The importance of proteins and their main structural properties are described in this section. Studying protein folding is a major challenge, and the approaches used to study this topic are discussed. In addition, the thesis overview is outlined.

1.0.1 The Importance of Proteins

Proteins play crucial roles in many biological and cellular processes including respiration, digestion, disease combat, growth, and movement. For example, in humans, hemoglobin transports oxygen from the lungs. Enzymes help speed up chemical reactions such as aiding in the digestion of food. Antibodies recognize invading elements and fight off infectious diseases. Furthermore, proteins such as collagen, keratin, and myosin are major structural components of the body as they are the main constituents of skin, bone, hair and muscle. Proteins fold into a variety of three-dimensional shapes, which are responsible for their unique functions. By studying the relationship between the sequence and structure, a better understanding of the factors involved in folding can be achieved. The biological function of a protein depends on its conformation, and thus, it is important to understand these factors.
1.0.2 The Protein Folding Problem

In the late 1950s, Anfinsen and his coworkers made an impacting discovery. They studied the reversible folding and unfolding of ribonuclease A, and came to the conclusion that the primary amino acid sequence contains all the information necessary to specify the three-dimensional structure of a protein. This finding led to the notion that any folded conformation can be predicted from its primary amino acid sequence (1). However, the task of determining the fold of a protein from its amino acid sequence has proven difficult as there are so many possible conformations, and has come to be known as the protein folding problem. Scientists are still searching for rules that govern protein folding, but predicting the overall structure from a simple sequence still remains to be elucidated (2). Scheme 1.1 illustrates the protein folding problem, showing a flexible peptide chain which folds into a unique structure.

Scheme 1.1. General depiction of the protein folding problem.
1.0.3 Methods for Studying the Protein Folding Problem

The protein folding problem has been approached in a number of ways. In site-directed mutagenesis, certain residues are mutated in a natural protein in order to examine their effects on the structure (3). However, a natural protein can contain many hundreds of amino acids, and therefore it is difficult to determine the effect of a single amino acid on the overall fold. Alternatively, researchers have used statistical analysis of data to determine the preference of an individual amino acid for a given secondary structure. Such studies led to empirical secondary structure prediction schemes, but have only been useful for certain sequences (4). Another approach to studying protein folding is to synthesize and characterize proteins that are designed 'from scratch' (5). These structures have been termed *de novo* proteins. They are smaller than their natural counterparts, but retain the same features involved in folding, and therefore act as simple models to investigate the factors that govern protein folding. Researchers have found that only a small fraction of the residues (less than 10%) in a protein sequence are critical for dictating the overall structure (6, 7), which indicates that there are fundamental forces that regulate protein folding even with large variations in the sequence. Furthermore, computer modelling has been a rapidly emerging technique used to study protein folding (8), and is discussed in Chapter 3.

1.0.4 Thesis Overview

The prediction of the secondary structural elements has been successful, but the prediction of tertiary structures from the amino acid sequence is still far from being established.
The goals of this thesis are to determine the effects of various constituents on a protein’s three-dimensional structure and to test our knowledge of peptide design by synthesizing and characterizing a series of proteins. In this thesis, the relationship between the primary sequence and the overall tertiary structure is probed using de novo proteins composed of a number of α-helices linked onto a rigid template called a cavitand. The cavitand template is an organic macrocycle that is rigid and serves to promote protein folding. Chapter 1 gives a general introduction to protein structure and describes the interactions involved in folding. It explains the de novo protein approach, with special focus on template assembled synthetic proteins. In Chapter 2, the effect of various linker lengths between the peptides and the template on a series of four-helix bundles is examined. Chapter 3 further investigates the relationship between the cavitand template and the peptides using molecular dynamic simulations on the same series of four-helix bundles studied experimentally. In Chapter 4, different-sized bundle proteins are analyzed to determine the effect of the number of helices on a specific sequence. In addition, Chapter 4 describes an approach used in an attempt to screen for native-like proteins. Finally, Chapter 5 presents the thesis conclusions and the future goals.
1.1 Peptide and Protein Structure

1.1.1 General Structural Features

Proteins are made up of building blocks called amino acids, also referred to as residues. The 20 commonly occurring amino acids can be written as either a three-letter code or one-letter code (listed in Appendix A). Despite the limited types of amino acids, the variations in the sequence order, number of amino acids in a sequence, and possible conformations can result in an extensive variety of proteins. Each amino acid consists of an amino group, a carboxyl group, and an α-carbon as shown in Figure 1.1(a). The α-carbon is attached to a hydrogen atom and a side group, R that differs in each amino acid. All the amino acids are chiral with the exception of glycine, which has two hydrogen atoms attached to its α-carbon. By convention, enantiomers of amino acids are designated D and L. Natural proteins consist of L amino acids, although both D and L amino acids exist in nature.

![Diagram](attachment:image.png)

**Figure 1.1.** (a) L-amino acid and (b) peptide with L-amino acids.
X-ray crystallographic data show that the bond between the carbonyl carbon and the 
nitrogen is slightly shorter than a typical C-N bond, but is slightly longer than a double bond 
between C=N; the bond between the carbonyl carbon and oxygen is slightly longer than a typical 
C=O, but is shorter than a typical C-O bond. These data show that the group of atoms involved 
in the peptide bond is in a resonance hybrid. Rotation about the peptide bonds is hindered 
because of their double bond character with the NH-CO dihedral angle, \( \omega \), of around 180° (9). 
However, rotation about the N-C\( \alpha \) bond, designated \( \phi \), and the C\( \alpha \)-C bonds, designated \( \psi \), in the repeating N-C\( \alpha \)-C backbone is possible. This rotation is largely influenced by steric 
between the main chain and the side groups on the adjacent residues. Peptides are written from 
left to right from the N- to C-terminus direction as shown in Figure 1.1(b).

There are four levels of protein structure. The primary structure of a protein is the linear 
sequence of amino acids. The secondary structure refers to the regularities in local conformations 
maintained by hydrogen bonds between amide protons and carbonyl oxygens of the peptide 
backbone. The tertiary structure refers to the folded polypeptide chain stabilized by interactions 
between side groups of non-neighbouring residues. The quaternary structure is the association of 
two or more polypeptide chains into a multisubunit or oligomeric protein.

1.1.2 Secondary Structure

There are two main classes of secondary structures for proteins: the \( \alpha \)-helix and the \( \beta \)- 
sheet, with the former being the most common type. The average globular protein contains 
around 30 % \( \alpha \)-helicity (10). Ramachandran used computer models of small peptides to 
determine which values of phi and psi are permitted sterically in a polypeptide chain of a given
secondary structure (11). Figure 1.2 shows a Ramachandran plot of psi versus phi along with the allowed regions of the secondary structures. Only the α-helices are discussed in further detail due to their significance in the thesis.

![Ramachandran plot](image)

**Figure 1.2.** Ramachandran plot.

### 1.1.3 The α-Helix

Pauling and Corey first proposed the α-helix structure in the early 1950s (12), and found a repeating element along the axis of 0.54 nm called the pitch (axial distance per turn). Soon after, Perutz confirmed the existence of the α-helix and observed in the X-ray diffraction pattern of α-keratin, a secondary repeating unit of 0.15 nm, corresponding to the rise (advance per
The L amino acids typically form right-handed $\alpha$-helices that contain 3.6 residues per turn (see Figure 1.3). They have backbone dihedral angles, $\phi$ and $\psi$, of approximately $-60^\circ$ and $-40^\circ$ (10, 14), respectively.

A key pattern observed by Pauling and Corey was the backbone hydrogen bonding between each amide proton donor and the carbonyl oxygen acceptor located on an amino acid four residues apart as shown by the dashed lines in Figure 1.3. Although, the $i, i+4$ hydrogen bonds along an $\alpha$-helix is the most common pattern, more tightly wrapped $3_{10}$-helix ($i, i+3$ hydrogen bonds), and more loosely wrapped $\pi$-helix ($i, i+5$) also occur. The $\alpha$-helices mentioned in this thesis are considered to have the $i, i+4$ hydrogen bond pattern as shown in Figure 1.3.

Figure 1.3. Right-handed $\alpha$-helix (figure adapted from reference (15)).
1.1.4 α-Helical Motifs

Common α-helical motifs found in natural proteins include coiled coils and helical bundles. Generally, coiled coils are found in fibrous proteins, and helical bundles are found in globular proteins.

1.1.4.1 Coiled Coils

Coiled coils are generally an ensemble of two to five α-helices that pack together in either a parallel or antiparallel orientation with a helix crossing angle near 20° (16). The coiled coil model was first proposed by Crick in 1953 (17). Coiled coils mainly consist of two parallel α-helices, and are commonly found in DNA binding proteins. Figure 1.4 shows a two-stranded coiled coil and its helical wheel diagram. Although less common, three stranded coiled coils such as extracellular α-fibrous proteins (18), influenza hemagglutinin (19) and α actinin (20), and four-helix coiled coils such as repressor of primer, Rop (21, 22) and Lac repressor, LacR (23) also occur. Even less common are the five helix coiled coils (24). The cartilage oligomeric matrix protein (COMP) contains extremely stable five-stranded parallel α-helical coiled coils (25).
Figure 1.4. (a) Two-stranded coiled coil structure and (b) helical wheel diagram showing interactions between the nonpolar residues a and d, and ion pairs between residues e and g.

The first de novo designed α-helical coiled coil was studied by Hodges and coworkers using tropomyosin as a model (26, 27). Generally, coiled coils are stabilized by interaction of right-handed helices that wind around one another in a left-handed supercoil. Coiling maximizes the interchain packing of the hydrophobic residues, and also allows for interchain ionic interactions, which define relative chain alignment and direction (28). The axis of the α-helices bends in a supercoiled structure to maintain systematic side chain interactions over the entire length. This supercoiling reduces the effective number of residues per turn from 3.6 to 3.5 for optimal knob into hole packing. The structural features of a number of coiled coils can be seen in their crystal structures (21, 29, 30).

Left-handed coiled coil structures are characterized by a repeating heptad pattern, [abcdefg]n, where a and d are usually nonpolar residues (see Figure 1.4(b)). Hydrogen bonds are systematically shortened in the hydrophobic region, and therefore these nonpolar residues usually lie in the concave side of the curved helices. Leucine is the most common nonpolar
residue, but other nonpolar residues also occur with preference in either the \textbf{a} or \textbf{d} position. For example, alanine is favoured in the \textbf{d} position and isoleucine and valine are favoured in the \textbf{a} position of a two-stranded coiled coil (31). Polar residues that are solvent exposed are located in the \textbf{e} and \textbf{g} positions with glutamate preferentially in the \textbf{e} position and lysine favoured in the \textbf{g} position. These residues provide specificity between the helices through electrostatic interactions. The remaining three positions, \textbf{b}, \textbf{c}, and \textbf{f} are also hydrophilic as these form helical surfaces that are exposed to the solvent. Occasionally, there is a break in the heptad pattern called a skip (32) or a stutter (33), characterized by a phase shift of one residue or an insert of several residues, defined respectively. These disruptions in the heptad periodicity may produce a 'kink' or local distortion in the axis of the coiled coil.

\subsection*{1.1.4.2 Helical Bundles}

In globular proteins, similar helix-helix interactions as coiled coils occur in short regions, but the interactions are more complex and often do not display simple systematic packing such as knob into holes (28). The interacting helices in a square helical bundle cross at diverse angles and are not regularly inclined to one another (34). In \(\alpha\)-helical bundles, the helix crossing angles can either be 20° or larger and can vary considerably (16). However, unlike coiled coils, the helices do not bend as significantly. Helical bundles tend to be antiparallel as the dipole moment influences packing, whereas coiled coils tend to be parallel. Although helical proteins have similar design principles as coiled coils, one major difference is that the latter contains a narrow hydrophobic face whereas the former contains a wider hydrophobic region owing to a greater number of nonpolar residues in the \textbf{e} and \textbf{g} positions. Furthermore, the helices of a bundle
diverge at the ends due to the absence of significant supercoiling (16). Figure 1.5 shows this type of helical interaction.

![Figure 1.5. Square bundle interaction.](image)

1.1.4.3  **Four-Helix Bundles**

One of the simplest of the globular folding motifs is the four-helix bundle (35, 36). This structure occurs in natural proteins such as cytochrome c, apoferritin, myohaemerythrin, and the tobacco mosaic virus. Because of its abundant occurrence in nature, its simplicity and its apparent symmetry, the four-helix bundle has been the motif of choice for many laboratories in the design and study of *de novo* proteins. Richardson and colleagues have designed a four-helix bundle called ‘Felix’ (37). This designed protein incorporates 19 of the 20 commonly occurring amino acids in order to obtain a native-like sequence; however, it was only shown to be marginally stable and non-cooperative in its unfolding transition. DeGrado’s group synthesized and characterized a series of four-helix bundles, each one designed to improve stability and native-like character (see section 1.4.1).
1.1.4.4 Five- and Six- Helix Bundles

Larger bundles of α-helices are also present in nature such as the M2 transmembrane segment of the nicotinic acetylcholine receptor (five-helix bundle) (38), and HIV gp-41 protein (six-helix bundle) (39). Trypanosome variable surface glycoproteins (VSGs), contain both six- and four-helix bundles (40, 41). Chapter 4 investigates the larger helical bundles and examines the structural differences and sequence specificity of these proteins.

1.1.5 Conformational Mobility of Side Groups Within Proteins: Native Proteins and Molten Globule Structures

Protein folding has been found to be a highly cooperative process in some cases, but in other cases, partially structured species have been observed before the formation of the native-state. Ptitsyn was the first to suggest that such a structure was an intermediate in the protein folding process. Around the same time, Wong and Tanford studied the near UV region of carbonic anhydrase under denaturing conditions, and observed the existence of an intermediate in folding that had a loosely packed core as compared to the native state (42). The term molten globule was introduced in 1983 to describe a structure that contains native secondary structure, but lacks specific tertiary contacts. Ptitsyn made an analogy of the molten globule to the liquid state between the gas and solid phases, equating the unfolded state to the gas and the native state to the solid phase (43). The correct usage of the term ‘molten globule’ has been a topic of debate (44). ‘Pre-molten globule’ has been used to describe both the intermediate between the unfolded state and the molten globule (45), as well as to describe a more structured state of a molten globule (46). In addition, the term ‘highly-ordered molten globule’ has been used to describe an
intermediate between the molten globule and the native protein (45), while others call this structure the native protein with dynamic disorder due to the significant tertiary contacts (47). The question as to whether the molten globule is really a kinetic intermediate or a non-productive species further complicates the use of this term. Creighton studied the protein α-lactalbumin, and concluded that the molten globule state is not a key in the rapid folding process (48). However, others have shown that the molten globule structure of α-lactalbumin does indeed contain a native-like tertiary fold and serves as an intermediate to folding (49, 50).

For the purpose of this thesis, the term molten globule is used to describe a structure that contains an intact secondary structure, but lacks the specific interactions between the side groups of the tertiary structure. Throughout the thesis, native-like proteins are referred to as well-defined structures or proteins that exhibit high conformational specificity. Figure 1.6 shows the differences in side chain packing between the native-like structure and the molten globule structure.

![Native State vs. Molten Globule State](image)

**Figure 1.6.** Native structure versus molten globule structure. The secondary structure is maintained in the molten globule state, but it lacks the specific tertiary contacts that the native structure possesses.
Molten globule structures are not amenable to X-ray crystallography because they have dynamic and flexible character; however, NMR spectroscopy has been a useful method to detect the presence of the molten globule structures as they lack both the chemical shift dispersion characteristic of the native state, and the narrowness of the signals characteristic of a completely unfolded protein (51). In addition, a hydrophobic dye such as ANS has been used to detect molten globules. Due to their fluctuating nature, molten globules preferentially bind to the dye (52). Experiments to probe native-like and molten globule characteristics are discussed further in the next chapter.
1.2 Factors Influencing Protein Structure and Folding

1.2.1 Conformational Entropy and Non-Covalent Interactions

There are many interactions involved in the folding of proteins that give rise to their stability and unique structures. Free energies of unfolding have been estimated to range from 5-20 kcal/mol (53, 54). Protein folding and stabilization depend on conformational entropy and several noncovalent forces such as electrostatic interactions, hydrogen bonding, van der Waals interactions, and hydrophobic effects. The noncovalent interactions may be weak individually, but collectively, they provide stability for the native conformations of the proteins.

1.2.1.1 Conformational Entropy

Side chain entropy of the individual residues plays a role in the packing of the proteins (55). In the open conformation, the side chain degrees of freedom and hence their entropy is high. As the protein starts to fold, the side-chain entropy decreases, which opposes the folding process. Several groups have shown that the entropic cost of restricting a single side chain bond is approximately 0.5 kcal/mol (56). Other interactions can overcome the decrease in entropy to promote folding.
1.2.1.2 Electrostatic Interactions

The classical electrostatic effects are defined as unfavourable non-specific repulsions that result due to highly charged species (57). This traditional view of electrostatic interactions only predicts the destabilization of a protein with an increase in charge. However, specific stabilizing ion pairing effects also exist. Favourable electrostatic interactions can occur between oppositely charged side groups. Ion pair interactions between negatively charged Glu and positively charged Lys have been found to contribute approximately 0.6 kcal/mol of stabilizing free energy per pair in a heterotetrameric coiled coil (58). In general, it has been estimated that favourable ion pairs provide 0.5-3 kcal/mol of stability per pair (59, 60).

Although there are strong arguments that the hydrophobic interactions represent the main folding force, Anderson and coworkers showed that by removing only one salt bridge, the stability of the T4 lysozyme protein was significantly affected by 3-5 kcal/mol (61). Most ionic side chains occur on the surface of the proteins where they can be solvated. Two oppositely charged side groups may occasionally form an ion pair in the interior of a protein (62). Such interactions are significantly stronger than those exposed to the water, since there is no competition with the solvent molecules. Barlow and Thornton found that only about five interactions of ions pairs exist per 150 residues on average, and that the number of ion pairs that are buried is even smaller (one per 150 residues) (62). Furthermore, the protein stabilities were found to be affected only slightly by changing the charged residues of several different residues in various proteins (63). Therefore, it is unlikely that charge-charge interactions are the dominant force for protein folding.
1.2.1.3 Van der Waals Interactions

Van der Waals interactions are composed of London forces and dipole-dipole interactions, and are the result of contacts between induced or fixed dipoles, respectively. They are essentially weak electrostatic interactions that arise due to localized distribution of electrons in the atomic orbitals. The force is generally small, and is in the range of 0.3 kcal/mol (64).

1.2.1.4 Hydrogen Bonding

Hydrogen bonds help stabilize the native conformations of proteins and also contribute to the cooperativity of folding. They can form between an electropositive hydrogen atom and an electronegative atom. A common hydrogen bond in proteins exists between a backbone amide hydrogen and a backbone carbonyl oxygen. Hydrogen bonds can also exist between the peptide backbone and water, between the peptide backbone and a polar side chain, between two polar side chains, and between a polar side chain and water. The strength of this bond can range from 2-10 kcal/mol depending on the electronegativity and orientation of the bonding atoms (57). Hydrogen bonds stabilize the secondary structures such as α-helices, especially in the hydrophobic regions where there is less competition with water. Hydrogen bonds can be essential in specifying a protein’s conformation. For example, DeGrado and colleagues showed that a buried polar residue promoted the formation of a native structure by forming favourable hydrogen bond interactions, but it was energetically costly when compared to the burial of a hydrophobic side chain (65, 66).
1.2.1.5 Hydrophobic Effects

Proteins fold in such a way that the hydrophobic side groups interact with one another in the protein interior to minimize their exposure to the water molecules (67). There is an abundance of evidence that shows that hydrophobic interactions are the main driving force for the tertiary fold of both coiled coil structures and globular proteins in aqueous solutions (54, 68-71). A large number of crystal structures showed that the predominant feature of globular proteins was the nonpolar residues in the hydrophobic core (57, 68). Site-directed mutagenesis experiments demonstrated that replacing a nonpolar residue with an alternate amino acid destabilized the structure (57). Hydrophobic interactions seemingly provide the dominant driving force that stabilizes the protein structures; however, the other non-covalent interactions are also important for added stability and conformational specificity.

1.2.2 Factors that Influence α-Helical Structures

The stability of α-helices affects the stability of the overall proteins they comprise. There are various factors that stabilize the α-helix structures. Certain residues are more likely to occur in α-helices than others, and this tendency is known as the helical propensity. Other factors that affect the α-helices include helix macrodipole effects, helix-capping effects, and helix lengths.
1.2.2.1 Helical Propensity

The helical propensity of different amino acids has been estimated using a variety of experimental and theoretical studies on natural and synthetic peptides and proteins. These studies include statistical surveys of occurrence in crystal structures (72, 73), host-guest analysis (74, 75), model peptides (76-78), site-directed mutagenesis of model proteins (79, 80), and molecular dynamics calculations (81, 82). Determining helical propensity has been an active area of research for about 30 years, and can now be predicted with reasonable certainty.

Pace and Scholtz derived a helix propensity scale based on a range of experimental data (83). The general trend of helical propensity is as follows:

Ala > Arg > Leu > Met > Lys > Gln > Glu > Ile > Trp > Ser > Tyr > Phe > Val > His > Asn
> Thr > Cys > Asp > Gly >> Pro

The scale is in excellent agreement with scales based on data that were derived using different approaches (84, 85). With the exception of proline, the difference in helical propensities between Ala and Gly is 1 kcal/mol. A major determinant of whether a side chain is likely to be in a helix is the loss of conformational entropy (86). Alanine is a strong helix former because of the absence of unfavourable entropy brought upon by inserting a methyl side chain into the helix, and because of its minimal capacity to interact with other side chains (87).

To address any differences in helical propensities between model peptides and proteins, Myers and coworkers studied both the α-helices of the ribonuclease T1 protein, and a helical peptide of an identical sequence (88). They demonstrated that the scales were in good agreement for all the amino acids with the exception of some polar residues. Regan also found that helix
propensity can make the same energetic contributions whether the residue is in a peptide or in a protein (89).

Residues that contain long hydrophobic side chains such as Leu and Met, and polar groups that are several carbons away from the carbon backbone such as Arg and Lys are known to be strong helix formers (83, 90). Leucine and isoleucine are better than valine because the δ carbon of leucine and isoleucine seems to have favourable interactions in the helix that bury some nonpolar surface. Hodges and coworkers investigated the effect of α-helical propensity and side chain hydrophobicity, and found no correlation with each other, although both were shown to stabilize amphiphilic α-helices (91). β-branched residues such as Val, Ile, and Thr have lower helix propensities because of the loss in sidechain conformational entropy upon helix formation, and because the β-branched methyl groups can cause distortions in the local helix backbone (92). Aromatic residues such as Phe, Trp and Tyr also have poor helical propensities due to a similar argument.

Glutamine and glutamic acid display higher helical propensities compared to asparagine and aspartic acid even though the polar groups are the same. The extra methylene in Glu or Gln, as compared to Asp and Asn, allows for better interaction with the solvent and less interaction with the backbone. A side chain to backbone hydrogen bonding is entropically unfavourable, and having shorter side chains interact in a random coil is less entropically costly. Uncharged glutamic acid and aspartic acid have higher helical propensities than their charged counterparts. The charged carboxyls are thought to form stronger hydrogen bonds with the amide hydrogens in the peptide backbone of a random coil.

Glycine has a low helix propensity because it is inherently flexible, and it would be entropically costly to restrict its mobility in a helix backbone. Proline is the worst helix former for two main reasons: it lacks an amide hydrogen atom for main chain hydrogen bonding, and
because of its rigid geometry that can distort the helix backbone (93). Therefore, glycine and proline are often found at the ends of the helices.

It has been demonstrated that the residues at the ends of the helices have different propensities than the residues in the middle (94). Recently, DeGrado and coworkers showed the position-dependent propensities throughout the length of the helix (95).

Knowing the role of individual side chains in forming a specific secondary structure such as the α-helix is a useful tool in *de novo* protein design. However, there are many factors that contribute to the helix stability. For example, the solvent environment can override the preference for a particular secondary structure. Waterhous and coworkers showed that equivocal sequences that formed β-strands in SDS (sodium dodecyl sulphate), formed α-helices in alcohol solvents (96). Examples by Hodges’s group (97), and Kim’s group (98) have also demonstrated similar solvent dependent effects on the predominating secondary structure.

1.2.2.2 Helix Macrodipole Effects

In α-helices, macrodipole effects occur as a result of highly ordered hydrogen bonding between the carbonyl oxygen and the amide hydrogens of the helix backbone (99, 100). The peptide bonds of an α-helix align in such a way that there is an effective +0.5 or -0.5 charge near the N- and C-termini, respectively (101-103). Protein engineers have considered this effect when designing protein sequences. For example, a four-stranded coiled coil was designed to be arranged in an antiparallel fashion to minimize the macrodipole effects of the helices (104). Furthermore, negatively charged residues are frequently incorporated near the N-terminus and positively charges residues are found near the C-terminal region to stabilize the helices (74, 105).
1.2.2.3 Helix Capping Effects

In an α-helix, the backbone amide hydrogen of a residue is capable of forming a hydrogen bond with the backbone carbonyl group of another residue four amino acids apart in an $i, i+4$ manner to stabilize its secondary structure. However, the first four NH groups at the N-terminus and the last four carbonyl CO groups at the C-terminus lack intrahelical hydrogen bond partners as seen in Figure 1.7.

![Diagram of a-helix showing non-hydrogen bonded carbonyls and amide hydrogens](image)

**Figure 1.7.** α-Helix showing the backbone hydrogen bond pairs. Note that the first four amide hydrogens and the last four carbonyl oxygens lack hydrogen bonding partners.
The helix hypothesis, introduced by Presta and Rose, states that a necessary condition for helix formation is the presence of amino acids at the helix termini that possess side chains that can provide hydrogen bond partners for the unpaired main chain NH and CO groups (14). This helix capping effect has also been shown by others to stabilize the helical structure of peptides and proteins (106-109). Moreover, certain residues were more likely to be favoured at either the N- or C-terminal ends. These capping residues often have nonhelical dihedral angles (110). The nomenclature for helices and their flanking residues uses the following notation:

\[ \cdots -N''-N'-(N\text{-cap})-N_1-N_2-N_3- \cdots -C_3-C_2-C_1-(C\text{-cap})-C'-C''- \cdots (111). \]

### 1.2.2.3.1 N-Capping

Serrano and Fersht synthesized a series of mutated proteins by modifying two N-capping residues of barnase, and found that the N-cap can stabilize the protein by as much as 2.5 kcal/mol (107). The most common residues found at the N-cap position include Ser, Asn, Gly, Asp and Thr, and the least common residues were Ala, Leu, Val, Ile, Trp, Arg, Gln and Glu (106, 107). These preferences can be rationalized by several factors. The main factor is the ability of the N-cap side group to backbone hydrogen bond, but other aspects include solvation of the non-bonded backbone sites and the electrostatic interactions with the helix macrodipole. The small residues such as Ser, Asn, Asp, and Gly make suitable N-caps because they can either easily form hydrogen bonds with the backbone NH groups, or they are small enough that they allow solvent molecules to hydrogen bond in their place (112). Asn is an especially good N-cap because its side chain mimics a residue so that the entire amino acid is like a dipeptide (106). In
other words, it can stabilize the first helical turn by providing an interaction comparable to that of another residue. In addition, it discourages further helix propagation due to its low helical propensity. Asp is also an efficient N-cap because it can interact favourably with the helix dipole and also hydrogen bond to one of the unfulfilled backbone NH groups (113). Larger residues are rarely found as N-caps because they do not have the correct geometry for hydrogen bonding to the main chain NH groups (106).

### 1.2.2.3.2 C-Capping

The glycine residue is the most preferred residue at the C-cap position (111). Richardson and Richardson sampled 215 α-helices from 45 globular proteins and found that Gly is located at the C-terminal end of 34% of the helices (106). Gly is a suitable C-cap because it can satisfy the hydrogen bond requirements of two successive carbonyl oxygens, as well, it can stop the continuation of a helix as it has very low helical propensity. Positively charged residues such as His, Lys and Arg are also frequently found at the C-cap position, which can be attributed to minimize the dipole effect of the helix (114). Other residues found as the C-cap include leucine, phenylalanine, valine and alanine because they have an ability to form hydrophobic caps (115).
1.2.2.3.3 Hydrophobic Capping

Initially, only hydrogen bonding was considered to be important in determining capping effects. It was not till years later, that the phenomenon of hydrophobic capping was introduced (110, 111). It involves two hydrophobic residues that are close in proximity to each other that stabilizes the terminal ends of the helices. For example, one residue in the first or last turn of the helix, and the other located outside the helix. Out of a set of 1316 helices, 1309 and 1313 helices included a hydrophobic residue in the first and last turn, respectively (108). The specific positional preferences for the hydrophobic cap were found to be as follows: N3~N4>N2>(N-cap)~N1 at the N-terminal and C3>C2>(C-cap)>C1>C4 at the C-terminal end.

1.2.2.3.4 Capping Motifs

A set of residues can make specific interactions at the capping sites and are referred to as capping motifs. A common capping motif found at the N-terminal end of the helices is the capping box (116, 117). This box caps two of the initial four backbone amide hydrogens. In particular, the N-cap side group forms a hydrogen bond with the backbone of N3, and simultaneously, the N3 side group hydrogen bonds back with the backbone of the N-cap. The preferences in a capping box are as follows: Thr>Ser>Asn at the N-cap position and Glu>Gln at the N3 position (108). The capping box was expanded to include the hydrophobic interaction between residues N' and N4. This lengthened motif has been referred to as both the “expanded capping box” (118) and the “hydrophobic staple” (119).
The two most common motifs at the C-terminal end of the helices are the Schellman and the $\alpha_L$ motifs (111). The Schellman motif consists of a doubly hydrogen-bonded pattern between the NH at C" and C=O at C3, and between NH at C' and C=O at C2. There is also a associated hydrophobic interaction between C3 and C" (108). The C' is typically a glycine. The $\alpha_L$ motif consists of a hydrogen bond between the N-H at C' and C=O at C3. The C' is also a Gly as in the Schellman motif, but the C" is polar therefore does not take part in a hydrophobic interaction with C3.

1.2.2.4 Helix Chain Length Effects

Helix chain lengths have a substantial effect on the stability of the helices and hence the resulting proteins. Typically, at least eight to twelve residues are required to form an isolated $\alpha$-helix, and the stability of the helix increases with length. Several studies have shown that the helicity and stability of coiled coils increase with an increase in chain lengths due to an enhancement in stabilizing contacts (120-123). For conformational stability, generally, a minimum of 5 heptad repeats were required (28). For a heptad sequence such as (KIEALEG)$_n$, it was found that a minimum of 28 amino acids corresponding to 4 heptads, were required to form a stable two-stranded $\alpha$-helical coiled coil (120). However, it was unclear if this four-heptad requirement was sequence dependent. Therefore, a study on a series of polypeptides using a different heptad repeat, (EIEALKA)$_n$, with the number of residues ranging from 9 to 35 amino acids was used to study the effects of chain lengths on the formation of two-stranded coiled coils (121). The sequence was designed to maximize hydrophobic interactions as well as intra- and interchain ionic interactions. Here, a minimum of 3 heptads corresponding to six helical turns
was required for the peptides to adopt a two-stranded coiled coil in aqueous medium, showing the chain length requirements are sequence dependent.

For four-helix bundles, helices with four or more turns were found to pack optimally in elongated bundles, whereas short helices packed in a number of other geometries (34). Fairman and coworkers found that increasing the chain length from three 3 to 5 heptad repeats increased the stability of the four-helix bundle by 9 kcal/mol.

Despite the abundance of evidence showing that increasing chain length increases the stability of the protein structures, a recent study showed the opposite effect; increasing the chain length decreased the overall stability of the coiled coils (124). Here the heptad repeat consisted of \((\text{KAEALEG})_n\). Although a stabilizing nonpolar residue, Leu, was placed in the d position, and stabilizing ionic interactions existed, there was still insufficient enthalpic stabilization to overcome the entropic destabilization from increasing chain length. This study implied that the stability of the coiled coil structure is dependent on the properties of the heptad repeat.
1.2.2.5 Intrahelical Interactions

The previous sections described the various factors that influenced the α-helical structures. The helical propensity of the residues, capping effect, macrodipole effect, hydrogen bonds between residues apart by \( i+4 \), electrostatic interactions between residues apart by \( i+3 \) or \( i+4 \) (corresponding to one helical turn), all have been found to play a role in stabilizing the α-helix. A common electrostatic interaction exists between Glu and Lys, although any two complementary charged residues can generally stabilize a helix by up to 0.5 kcal/mol (101, 125). Table 1.1 highlights the stabilizing intrahelical interactions.

Table 1.1. Stabilizing intrahelical interactions.

<table>
<thead>
<tr>
<th>Stabilizing Effects</th>
<th>Stabilizing Energy (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range of Helix Propensities</td>
<td>~1 (not including Pro)</td>
</tr>
<tr>
<td>N-cap</td>
<td>~1-2.5</td>
</tr>
<tr>
<td>C-cap</td>
<td>~0.5</td>
</tr>
<tr>
<td>Macrodipole Effect</td>
<td>~0.5</td>
</tr>
<tr>
<td>Hydrogen Bonding</td>
<td>~2</td>
</tr>
<tr>
<td>Electrostatic Interactions Between Side Groups</td>
<td>~0.5</td>
</tr>
</tbody>
</table>
1.2.3 Interhelical Interactions

1.2.3.1 Packing Pattern

In 1953, Crick proposed the 'knobs into hole' model for interhelical interactions (17). When the crossover angle is about 20°, the α-helices interact favourably for approximately six turns before they begin to diverge. In coiled coils, the close packing is maintained for many more turns by supercoiling. In the 'knobs into holes' model, the side chain of one residue acts as the 'knob', and points into the 'hole' made by the side chains of an adjacent helix as seen in Figure 1.8(a) (126). However, as more protein structures became available, it was certain that 'knobs into holes' could not solely explain the packing. A few decades later, Chothia proposed the 'ridges into grooves' model, in which side groups of the residues are arranged in a row along the surface of the helix to form ridges that are separated by grooves as shown in Figure 1.8(b) (127, 128). Generally, the ridges are formed by side chains three \((i-3, i, i+3)\) or four \((i-4, i, i+4)\) residues apart, and they pack into the grooves formed by the adjacent helix. Almost a decade ago, Bowie suggested that the two-dimensional models described above do not fully describe the helix packing as the helices are not flat (129). He proposed that packing can be achieved at a variety of interhelical angles as the side chains are flexible (130, 131). As of now, no single model can explain the way all helices interact.
1.2.3.2 Interhelical Electrostatic Interactions and Orientation of the Helices

Interhelical electrostatic interactions also exist and have been extensively studied by Hodges and coworkers using α-helical two-stranded coiled coils (132-134). The net energetic contribution of interhelical electrostatic attractions to coiled coil stability was quantified and found to be approximately 0.4 kcal/mol from three independent comparisons (135). In this same study, it was found that a protonated Glu contributes 0.65 kcal/mol more stability when compared to the ionized Glu, which shows that the interhelical ion pair contribution was outweighed by some other interaction. HPLC results show that the protonated Glu residues are more hydrophobic than the ionized form, and this increased hydrophobicity at lower pH could explain the higher stability of the structure (136). Electrostatic interactions have also played a role in stabilizing hetero-stranded coiled coils over homo-stranded coiled coils that were destabilized by repulsive interactions (137).
The importance of interhelical electrostatic interactions in determining the orientation of the two-stranded coiled coils was demonstrated as the preferred arrangement was the one which provided the favourable interchain electrostatic interactions between oppositely charged side groups (138). As mentioned earlier, two-stranded coiled coils are usually parallel, whereas larger helical bundles are usually antiparallel to minimize the helix macrodipole. In one study, the antiparallel orientation of a four-helix bundle was found to be one kcal/mol more stable than when the helices were arranged in a parallel fashion (139).

1.2.3.3 Factors that Influence Peptide Oligomerization in Multi-Helical Proteins

An important aspect of design is to control the number of helices in the assembly of a helical protein. There have been a lot of studies on the GCN4 leucine zipper, which is a DNA binding protein. The N-terminal domain of GCN4 containing 33 residues, called GCN4-pl and its mutants have been examined to study the packing preferences of the hydrophobic residues on the 'multiple-strandedness' of the proteins (140-142). The wild type GCN4-pl forms a parallel dimer and contains Leu at all the d positions and mostly Val at the a positions with the exception of Met 2 and Asn 16 (29). The wild type sequence is as follows:

```
ad ad a d a d a a
RMKQLEDK VEELSK NYHLENE VARLKKL VGER
```

The mutants consisted of changing all the a residues to Ile, Val or Leu, and also changing the d residues to Ile, Val, and Leu. For example, GCN4-p-IL consists of Ile residues in the a
positions and Leu residues in the d positions. The oligomeric states of GCN4-p-IL, -p-II, and-p-LI were found to be dimeric, trimeric and tetrameric, respectively. The mutant peptides -p-VI, -p-VL, -p-LV, -p-LL, all resulted in multiple self-associative states. Table 1.2 shows the resulting mutants and their respective oligomeric states.

Table 1.2. GCN4-pl mutants and their oligomeric states.

<table>
<thead>
<tr>
<th>GCN4-pl Mutants</th>
<th>a residue</th>
<th>d residue</th>
<th>Oligomeric State</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCN4-pIL</td>
<td>I</td>
<td>L</td>
<td>Dimer</td>
</tr>
<tr>
<td>GCN4-pII</td>
<td>I</td>
<td>I</td>
<td>Trimer</td>
</tr>
<tr>
<td>GCN4-pLI</td>
<td>L</td>
<td>I</td>
<td>Tetramer</td>
</tr>
<tr>
<td>GCN4-pVI</td>
<td>V</td>
<td>I</td>
<td>Multiple</td>
</tr>
<tr>
<td>GCN4-pVL</td>
<td>V</td>
<td>L</td>
<td>Multiple</td>
</tr>
<tr>
<td>GCN4-pLV</td>
<td>L</td>
<td>V</td>
<td>Multiple</td>
</tr>
<tr>
<td>GCN4-pLL</td>
<td>L</td>
<td>L</td>
<td>Multiple</td>
</tr>
</tbody>
</table>

Coiled coils pack as knobs into holes either in a perpendicular or a parallel fashion. Perpendicular packing is when the Ca-Cβ bond makes a 90° angle with the Ca-Cα bond, while parallel packing is when the Ca-Cβ bond is oriented parallel to the Ca-Cα bond. The dimeric mutant above shows parallel packing in the a layer and perpendicular packing in the d layer. The tetrameric mutant shows the parallel arrangement in the d layer and the perpendicular arrangement in the a layer. From these observations, β-branched residues (i.e. Ile) has a geometric preference for parallel packing and Leu has the preference for perpendicular packing. The trimer shows a third type of knobs into holes packing called the acute arrangement (142). Acute packing is when the Ca-Cβ bond makes an angle of less than 90° (usually 60°) with the Ca-Cα bond. This type of packing may be a result of the Ile in the GCN4-pl-II trimer trying to avoid the perpendicular arrangement. Figure 1.9 shows these different packing arrangements.
The effect of a single amino acid substitution in the hydrophobic core on both the stability and the oligomeric state of $\alpha$-helical coiled coils were demonstrated by Wagschal and colleagues (143, 144). The importance of a single amino acid substitution in triggering an oligomerization switch was also observed in the $\alpha_2D$ protein. When His 26 was changed to a nonpolar Phe, the dimer switched to a trimer (65).

The importance of Asn 16 in GCN4-pl was shown when it was changed to Val (145). This modification changed the coiled coil structure from dimer into a trimer. Alber and coworkers have also shown the importance of Asn 16 in dimer formation by substituting it with Gln 16, which resulted in a mixture of dimer and trimer (146). Furthermore, two designed peptides containing Leu at all the $a$ and $d$ positions were found to form a heterotetramer, but when a Asn was substituted into the $a$ position of each peptide, a heterodimer resulted (147). These results suggest that asparagine is an important residue in regulating the oligomeric state as it has the ability to form specific interhelical hydrogen bonds. Such an interaction has been observed in the X-ray structure of GCN4 (29).

The formation of a cavity in the hydrophobic core resulting from Ala substitutions can also change the oligomeric state of a protein (148). Two disulfide-bridged antiparallel coiled coils, which differed only in the position of a single Ala residue in the middle heptad were synthesized. When the four Ala residues were spread over two adjacent planes, the resulting

![Perpendicular, Parallel, Acute Knob-Hole Packing](image-url)
structure was a four-stranded protein. However, when all four Ala residues resided in the same packing plane, a cavity resulted, which destabilized the four-stranded coiled coil, and a two-stranded protein formed instead. Thus far, the control of oligomerization states has been attributed mostly to the hydrophobic residues at the a and d positions. However, it appears that the e and g positions also influence the 'multi-strandedness' of the proteins (149). For example, by replacing charged residues at positions e and g of the GCN4 leucine zipper with Ala, the oligomeric state changed from a dimer to a tetramer (150). In another case, a single residue substitution at the e or g position changed a three-stranded coiled coil to a four-stranded structure (151).

Despite the many studies that have been carried out to identify the structural elements that are responsible for the formation of different association states, it is still not completely clear how the combination of hydrophobic packing and electrostatic interactions work together to control the oligomerization of the peptides.
1.3 Various Views on Protein Folding

The prediction of protein folding rates and mechanisms is of great interest, but is associated with many difficulties. Anfinsen and coworkers showed that proteins fold and unfold reversibly; they stated that the native structure of small globular proteins is thermodynamically stable and at the global minimum of their accessible free energies (1). This statement has been called the “thermodynamic hypothesis”. About a decade later, Levinthal argued that this hypothesis alone could not explain protein folding because there are too many possible protein conformations to find the native structure in the conformational space by random searching. This argument has come to be known as “Levinthal’s paradox” (152). To solve this problem, the classical view of protein folding emerged. It stated that proteins fold via a specific pathway, and that the thermodynamic and kinetic components involved in determining a protein’s three dimensional are mutually exclusive.

1.3.1 Thermodynamic and Kinetic Control

In terms of protein folding, thermodynamic control refers to a protein reaching its global minimum in energy, and the folding pathway is considered to be independent. In other words, the native structure is only determined by the final native conditions, and not by the initial states (153). However, a process under thermodynamic control may take a long time because it requires an extensive search. On the other hand, kinetic control means that the folding takes a shorter time (on biological time scales) because it is pathway dependent. Some conformations are kinetically inaccessible, and the protein may only reach a local minimum. Here, the final
structure may differ depending on the initial conditions (153). A hypothetical protein under either thermodynamic or kinetic control is depicted in Figure 1.10.

![Thermodynamic and Kinetic Control](image)

**Figure 1.10.** Protein under (a) thermodynamic control and (b) kinetic control. Under thermodynamic control, the global energy minimum is accessible from any point on the energy curve, whereas, under kinetic control, the energy surface is more convoluted with more than one minimum. A protein starting from the right side of the curve could get trapped in a local minimum and never reach the global minimum. The actual native state could correspond to either the local or global minimum.

### 1.3.2 Classical View versus New View

Although Levinthal solved the random search problem by including folding pathways, there still exists the problem of thermodynamic versus kinetic control (pathway independent versus pathway dependent), which is known as Levinthal dichotomy (154). The classical view only involves the folding from one specific conformation, which limits all other starting points. Thus, a new view emerged to replace the ‘tunnel’ pathway of the old view with a ‘funnel’ pathway. Baldwin compares the classical view with the new view (155).
1.3.2.1 Classical View

The classical view is based on three simple models as shown in Figure 1.11:

\[
\begin{align*}
U \rightarrow I \rightarrow N & \\
U \rightarrow N & \\
U \rightarrow I_1 \rightarrow I_2 \rightarrow \ldots \rightarrow N
\end{align*}
\]

On-Pathway Model \hspace{1cm} Off-Pathway Model \hspace{1cm} Sequential Model

**Figure 1.11.** Three classical models: The on-pathway, the off-pathway, and the sequential model, where U=unfolded, I=intermediate, and N=native structure.

If the optical properties of the protein folding follow a single-exponential time decay, the folding and unfolding of the protein is considered two-state because only the native state N and denatured state U are involved. In a two-state model, the protein folds or unfolds in a cooperative manner, meaning that the first few interactions initiate the subsequent interactions. If multiple exponentials are observed, a more complex model is required to include the intermediate I. Models are chosen based on whichever best fits the experimental data. The classical model shows the average behaviour of the protein and does not give much atomic detail.
1.3.2.2 New View

The new view replaces the sequential models of the classical view with a funnel concept of parallel multi-pathways (156-158). Dill and colleagues compared the new view of the folding process with water trickling down the multiple mountain sides of various shapes and less like the flow through a single gulley as in the classical view (154). The classical view focused on specific structures, whereas the new view uses an ensemble perspective that recognizes that random processes and off-pathway steps also contribute to the folding speed. The funnels used to demonstrate the new view are described in terms of energy landscapes (156-159). The proteins with a smooth funnel folding landscape is said to be under thermodynamic control, whereas rough landscapes are said to be under kinetic control (153). Figure 1.12 shows a folding funnel for a protein folding.

The width of the funnel represents the entropy and the depth represents the energy. As a protein moves down the slope, both the energy and the entropy decrease. The shape of the funnel is dependent on the amino acid sequence. The folding is thought to occur through molten globule structures and a bottleneck region where it is possible for a conformation to get trapped in a local minimum, or discrete intermediates can form that lead the protein to the native structure.
Figure 1.12. Folding funnel of a protein.

1.3.2.3 Current Examples of the New View and the Classical View

For the most part, the new view has been accepted to describe protein folding. A recent article discussed the folding process of hen lysozyme from a ‘new view’ perspective (160). The folding of this protein was found to occur through two distinctive paths by kinetic binding studies and hydrogen exchange labelling. Although the new view is accepted by some
researchers, others have shown that the classical view is sufficient to describe protein folding. A recent article supports the classical pathway. Three forms of cytochrome c were found to possess significantly different thermodynamic stability, but the folding barrier for all three proteins were sizable in energy and were of the same magnitude, indicating that the folding rates were independent of the thermodynamic driving force (161).

Even with the abundance of research that has attempted to elucidate the folding mechanism of proteins, this topic is still not completely understood. Furthermore, the folding process has been described as reaching the native-structure that has the lowest energy conformation, but this is not always the case. In this thesis, the mechanism of protein folding is not examined, but rather the relationship between sequence and the overall structure is investigated.
1.4 De Novo Protein Design

*De novo* protein design involves the engineering of a protein from scratch using an artificial sequence intended to adopt a predetermined three-dimensional structure. Designing a sequence that folds into a given structure would be easier than predicting the unique structure adopted by a given amino acid sequence. This approach provides a powerful tool to address the conformational specificity of a protein because it allows for the examination of the relationship between the forces involved in folding. The goal is to evaluate the principles that influence the structure and ultimately the function of a protein. The success or failure of a design tests the underlying concepts and principles involved in folding.

1.4.1 DeGrado- Minimalist and Incremental Approach

William F. DeGrado is a pioneer of *de novo* protein design. Traditionally, *de novo* design has been studied using a minimalist approach (162) from first principles (163). These proteins contain highly repetitive sequences with high symmetry and limited covalent interactions, so that the folding process can be broken up into fundamental forces. The key features were designed for the four-helix bundle as it is a common motif in nature. The early design employed very few residue types with high helical propensity. Leu was the only hydrophobic residue, and Glu and Lys were the only hydrophilic residues. When the protein folded into a four-helix bundle, the nonpolar leucine residues pointed toward the centre forming a hydrophobic core to minimize their exposure to the aqueous solvent. The Glu and Lys amino acids were placed four residues apart so that they could form intrahelical salt bridges. Both the N- and C- termini were capped.
with Gly as this residue is commonly found at the end of the helices in natural proteins (111). To reduce any unfavourable macrodipole effects, the N- and C- termini were capped with acetyllys and amides, respectively. Furthermore, the Glu residue was placed closer to the N-terminus and the Lys residues was incorporated closer to the C-terminus. The antiparallel arrangement of the helices resulted in some potential interhelical repulsions which may have slightly destabilized the structure.

A purely minimalist approach has resulted in proteins that lack the thermodynamic characteristics of a native protein, and thus, a hierarchic design strategy was required to obtain more well-defined proteins (102). In collaboration with Eisenberg and coworkers, DeGrado’s group designed a collection of four-helix bundles using in an incremental approach to obtain idealized four-helix bundles (164, 165). The first step towards this goal was to design peptide sequences, namely $\alpha_1A$ and $\alpha_1B$, which would tetramerize into a four-helix bundle. Both sequences were designed using the minimalist approach to obtain amphipathic helices that allowed the nonpolar residues to interdigitate most effectively within a four-helix bundle. $\alpha_1B$ was improved from $\alpha_1A$, in which three residues were modified to increase the burial of the nonpolar residues and to offset the destabilizing dipole effect (166).

The second stage involved incorporating a Pro-Arg-Arg loop to connect the two helices ($\alpha_2$). This loop permitted the helices to dimerize, which resulted in a four-helix bundle, $\alpha_2B$. In addition, a single polypeptide chain called $\alpha_4$, composed of four $\alpha_1B$ and three Pro-Arg-Arg loops, was also synthesized. The $\alpha_4$ polypeptide was the first example of a designed protein that was shown to adopt a folded, globular conformation in aqueous solution (163). It was found to be monomeric in solution, highly helical and very stable toward denaturants. The stabilities of the four-helix bundles were enhanced with the addition of a loop connecting the helices.
However, both the $\alpha_2B$ and $\alpha_4$ proteins lacked the tight chain interactions required for a native-like conformation. Figure 1.13 shows the iterative design of the four-helix bundle proteins.

**Figure 1.13.** Family of DeGrado's four-helix bundle proteins.
In the design of the next generation, α₂C, the diversity of just the nonpolar residues was increased to improve the native-like character of the protein (167). Half of the Leu residues were replaced by aromatic and more conformationally restricted β-branched nonpolar side groups. α₂C showed an improvement in its ability to form a unique conformation, but it still showed characteristics of a molten globule structure.

The next generation of the α₂ family was designed to further induce conformational specificity (168). This new protein, α₂D was engineered to possess a metal binding site by modifying three more residues, with two of the three being His. Table 1.3 shows the designed sequences of DeGrado’s four-helix bundle family.

**Table 1.3.** The sequences used in DeGrado's design of *de novo* four-helix bundles. The bold letters illustrate the changes from the previous generation.

<table>
<thead>
<tr>
<th>Sequence Name</th>
<th>Sequence order (see Appendix A for one-letter code abbreviations)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helix1</td>
<td>[GKLEELLKKLLEELKG]</td>
</tr>
<tr>
<td>Helix2</td>
<td>[GELEELLKKLKEELKG]</td>
</tr>
<tr>
<td>Loop</td>
<td>[PRR]</td>
</tr>
<tr>
<td>α₁A</td>
<td>Ac-[Helix1]-COOH</td>
</tr>
<tr>
<td>α₁B</td>
<td>Ac-[Helix2]-CONH₂</td>
</tr>
<tr>
<td>α₂B</td>
<td>Ac-[Helix2]-[Loop]-[Helix2]-CONH₂</td>
</tr>
<tr>
<td>α₂C</td>
<td>M-[Helix2]-[Loop]-[Helix2]-[Loop]-[Helix2]-[Loop]-[Helix2]-COOH</td>
</tr>
<tr>
<td>α₂D</td>
<td>Ac-[GEVEELKKFKELWKG]-[Loop]-[GEIIELFKKFKELIKG]-CONH₂</td>
</tr>
</tbody>
</table>

α₂D was found to bind to Zn(III). Even in the absence of the metal, this structure behaved like a native-like protein. However, the improved conformational specificity occurred at the cost of decreased stability as nonpolar residues were replaced by polar residues. The three dimensional structure of α₂D was solved by high resolution NMR (169), and rather than the
expected *syn* or *anti* conformation, an unexpected topology called the bisecting U-motif was observed (see Figure 1.13).

Changing just three residues enhanced the native-like character of a protein. This finding promoted the idea of exploring the determinants for conformational specificity. Hydrophobic interactions are known to have an important role in driving protein folding (170); hydrogen bonds (65, 66) and salt bridges (171) have less substantial roles in stabilizing the proteins, but nevertheless, they are more sensitive to distance and geometry of the interacting groups, and therefore play an important role in specifying a protein's conformation. For example, DeGrado and coworkers showed the preference of heterodimers over homodimers by incorporating a buried salt bridge at the cost of thermodynamic stability (171). In another example, they showed that hydrogen bonds not only help to enhance the native-character, but they also help to specify the oligomeric state (65). In α2D, when His 30, which is capable of hydrogen bonding was changed to Lys 30, which can not participate in H-bonds, its native-like character was reduced as it adopted a more molten globule-like structure shown by NMR spectroscopy. When His 26 was replaced with Phe (isosteric, but nonpolar), there was a switch from a dimer to a trimer. These studies show that slight modifications have a significant effect on the resultant structures.

### 1.4.2 Combinatorial Techniques

Hecht and coworkers have used combinatorial methods to generate *de novo* proteins with native-like character (172). To create the libraries, a binary code strategy was employed (173). In this technique, the sequence locations of the polar and nonpolar amino acids are constrained, however, the identities of these side chains are allowed to vary. The initial attempt to test the
binary code strategy focused on the design of four-helix bundles composed of 74 residue sequences (174). However, most of the proteins from this initial collection failed to form native-like proteins. Since previous work showed that longer helices in α-helical bundles enhanced stability, and in some cases, increased structural specificity, longer α-helices of 102 residues were designed in the second generation library (175). Characterization of five arbitrarily chosen structures from this new library revealed that these structures were more stable and more well-ordered compared to the proteins from the first collection.

Libraries contain an abundance of structures that make differentiating between native-like and non native-like proteins a difficult task. Therefore, the combinatorial method was combined with electrospray mass spectrometry (ESMS) to monitor the hydrogen-deuterium exchange kinetics in semicrude samples to screen for native properties (176). The advantages of using this ESMS screen are that it is quick, requires little sample, and tolerates impurities. Hecht's group have shown that sequences generated by the binary system form well-ordered four-helix bundles with native-like character that have neither been selected by evolution nor designed by computation (172). Moreover, a fraction of the four-helix bundle library showed function such as binding heme (177) and peroxidase activity (178). These studies show that it is possible to obtain functional proteins from random sequence libraries.

Pluckthun's group also used a combinatorial approach to synthesize sequences that nature has not explored (179). In contrast to Hecht, they constructed a library that did not restrict the chain topology using the binary pattern strategy. Their main goal was to determine strategies for selecting a meaningful sequence. This library was found to contain both native-like and molten globule structures.
1.5 Template Assembled Synthetic Proteins

Although the folding problem is still far from being solved, considerable progress has been made towards understanding the fundamental forces involved in protein stability and specificity by using de novo proteins. A further approach to using de novo proteins is to design, synthesize and analyze template assembled synthetic proteins (TASPs). This concept was introduced by Mutter and coworkers in the mid eighties (180).

1.5.1 The Purpose and Advantages of the "Template-Assembly" Approach

The TASP approach consists of covalently attaching the peptides to a template or carrier molecule to construct the proteins. The template is characterized as a synthetic device to organize and direct structural units in a defined spatial arrangement. In non-templated de novo proteins, the folding is dictated by non-covalent interactions mentioned in section 1.2.1. In the case of the TASPs, the template offers an extra driving force, helping overcome the large entropic barrier of bringing inherently flexible peptides together. The 'template assembly' approach reduces the complexity of the folding process of a linear polypeptide chain by reducing the number of loops and turns. In addition, this approach allows for individual elements to be designed and synthesized before linking them together. Figure 1.14 shows a schematic representation of the 'template assembly' approach.
The template pre-organizes the peptide strands, promotes the folding process and stabilizes the overall branched structure. The use of a template can also control the orientation of the peptides and determine the size of a protein bundle, established by the number of reactive sites open for chemical ligation. The space between the reactive functional groups can dictate the packing topology of the protein structure. The use of a template may avoid certain pathways in the natural folding process since it explicitly directs the folding. The template may also force the peptides to fold into a three-dimensional structure that may otherwise be inaccessible.
1.5.2 Examples of Templates

A variety of templates have been used to study the folding of de novo proteins consisting of α-helices. Examples include peptidic, metal-ligand complex-based, porphyrin-based, aromatic ring-based, carbohydrate-based, and cholic acid-based scaffolds or carrier molecules. All template assembled synthetic proteins were found to be more stable and have enhanced helicity when compared to the non-templated peptides.

1.5.2.1 Peptide-Based Templates

Mutter and coworkers used peptides themselves as templates for the assembly of the helices to form four-helix bundle motifs. Earlier efforts comprised of using acyclic templates with a -ProGly- segment to induce a type II β-turn (181, 182). The turn was required for preorganizing the peptides for assembly. To further stabilize the TASPs, a cyclic template was synthesized by covalently linking two cysteine side groups via a disulfide bond (183). Figure 1.15 below shows a structure of a TASP that encorporates a peptidic template.

![Figure 1.15. Mutter's TASP with a peptidic template.](image)
This template sequence was employed because molecular modelling studies showed a low-energy conformation with the tripeptide, LysAlaLys, adopting an antiparallel β-sheet conformation. More importantly, the Lys side chains were found to be located on the same side of the template, which allows their ε-amino groups to attach to the peptides from the same face.

The design of the amphipathic helical bundles followed established principles, including using residues with high helical propensity and incorporation of residues with oppositely charged side chains in positions, $i, i+4$, for favourable intrachain interactions (184).

1.5.2.2 Metal-Ligand Complex-Based Templates

Ghadiri and coworkers used a metal-ligand coordination complex as their template to direct protein folding (185). For synthesis of the template assembled proteins, pyridyl groups were first attached onto the N-termini of the peptides, and then subsequently coordinated to ruthenium(II) metal in a square planar arrangement. The resulting protein was a parallel four-helix bundle as seen in Figure 1.16.
The peptide sequence was designed to be amphiphilic using Leu and Ala as the nonpolar residues. Glu and Lys were placed four residues apart to allow for intrahelical salt-bridge interactions. Furthermore, the C-terminus was amidated to compensate for any unfavourable helix dipole effects. The ruthenium-based template using pyridyl groups may have oriented the helices close in space to promote folding, but may not be ideal for the pre-organization of the helices in a bundle, since a fairly long linker group was required to join the peptides to the template.

1.5.2.3 Porphyrin-Based Templates

Sasaki and Kaiser were first to use a porphyrin-based template for protein assembly (186). The porphyrin template has a number of advantages (187). Firstly, the available functional groups provide electrophilic sites for the attachment of the peptides. The synthetic schemes for
the preparation of modified tetraphenylporphyrins are well developed. Secondly, porphyrin-based templates are resistant to aggregation and auto-oxidation. Thirdly, a potential catalytic site for oxidation of organic substrates is available. The first artificial hemeprotein was synthesized using a coproporphyrin template as shown below:

![Figure 1.17. A porphyrin template used in the synthesis of a TASP.](image)

This hemeprotein was named 'helichrome', and was designed to mimic the structure and function of cytochrome P-450 (188). The four propanoic acid linkers of the template were coupled to identical amphipathic peptides to afford a four-helix bundle. The special feature of a catalytic centre was made possible as the metallo-porphyrin prevented the collapse of the protein bundle, thereby creating a hydrophobic binding pocket inside the molecule for substrate binding. As for the peptide sequence, Leu and Glu were initially used as the hydrophobic and hydrophilic amino acids. Several of the Glu were substituted with Gln to reduce unfavourable charge-charge repulsions between the carboxyl groups of the Glu residues. Finally, alanine was added to the beginning and the centre of the sequence to reduce steric crowding above the porphyrin ring.
The free energy of folding of the 'helichrome' protein was found to be comparable to that of native globular proteins. Kinetic studies demonstrated that the catalytic activity of helichrome had similar hydroxylase activity as native hemeproteins. DeGrado and coworkers also used porphyrin-based templates to assemble the peptides into a well-defined tertiary structure (189). They showed an approach to linking unprotected peptides containing Cys residues to the bromoacetamido group of the porphyrin-based template by displacing the bromide and forming a thioether linkage.

1.5.2.4 Aromatic Ring-Based Templates

Templates used for protein assembly generally have been synthetically complex to easily vary its dimensions. Fairlie and coworkers synthesized a series of simple templates composed of rigid aromatic units in attempt to elucidate the influence of template size, shape and directionality (see Figure 1.18) (190). The templates were each reacted with four equivalents of a cysteine terminating peptide. Unlike a and b, template c had the added advantage of orienting all four peptides in the same direction. Templates a-c were found to promote \( \alpha \)-helix formation whereas d did not. The proteins resulting from linking peptides onto templates a-c were found to be monomeric by sedimentation equilibria studies. These proteins were also found to have similar stabilities to each other. Therefore, it was concluded that the flexible linker between the template and the peptides was sufficiently long enough that the shape, size and directionality of the template did not have substantial effects on the overall structure of the four-helix bundles.
Figure 1.18. Fairlie's aromatic-based templates.

The research carried out for this thesis uses another aromatic ring-based template to pre-organize the peptides for assembly and promote folding. These templates are known as cavitands, and are discussed in more detail throughout the thesis.
1.5.2.5 Carbohydrate Based Templates

Jensen and coworkers developed the use of carbohydrates as templates for the *de novo* design of proteins (191). The resulting structures were named *carbopeptides* (192, 193) and *carboproteins* (194, 195), depending on their size. Carbohydrates make suitable templates for assembling peptides because: monosaccharides are polyfunctional molecules, the pyranose ring is quite rigid, epimers of the sugars are often attainable, and the regioselective manipulation of their functional groups is possible. Earlier work used shorter peptide sequences such as H-Ala-Leu-Ala-Lys-Leu-Gly-β-Ala taken from the C-terminal end of a sequence reported by Mutter. D-galactopyranose (D-Galp) was chosen as the original carbohydrate-based template since four non-anomeric hydroxyl groups are above or in the plane defined by C2-C3-C5-O5 (see Figure 1.19). The anomeric position was left available for anchoring to a solid support as the carbopeptides were made by solid-phase synthesis (191, 196).

![Methyl α-D-Galp template](image)

**Figure 1.19.** Methyl α-D-Galp template. All non-anomeric hydroxyl groups for attachment of the peptides are in (C2, C3 and hydroxymethyl on C5) or above (C4) the plane defined by C2-C3-C5-O5.

The second generation of carbohydrate-based proteins were synthesized using longer peptides that were based on Mutter's sequences (197). The templates and the peptides were purified separately before coupling them together, which resulted in the preparation of
carboproteins with higher yield and purity. Furthermore, this approach allowed for attachment of non-identical or non-parallel helices to the template. The hydroxyl groups of the α-D-Galp template were functionalized as aminooxyacetyle groups and the various length peptide aldehydes were linked onto this template. The resulting four-helix carboprotein with the longer sequence was found to be quite helical (67 %), but the bundles with shorter peptide chains were much less helical (29 %), and a single-stranded carbopeptide contained 45 % helicity (194). The longer sequence four-helix bundle was also found to be the most stable, and was the only species that showed concentration independent CD spectra, which is indicative of non-aggregation.

In the recent research, different carbohydrate templates (hydroxyls differing in the axial or equatorial orientation) were used to examine the effects of various stereochemistry (195). Two new templates which were based on D-glucopyranoside (Glcp) and D-altropyranoside (Altp) were compared to the original D-Galp template.

![Figure 1.20. TASP](image)

R= Ac-YEELKKLEELKKKA-NHCH2CH=NOC_2CH3CO-

The peptide sequence was based on a sequence used by Mezo and Sherman with the addition of Tyr at the N-terminus for concentration determination purposes, and an Ala residue at the C-terminus for enhanced helicity (198). The C-terminal Gly enabled oxime ligations with
no risk of racemization. All the carboproteins using the three different templates were found to be \( \alpha \)-helical. The protein using the Alt\( p \) template contained the highest helicity, which was explained by better side-chain packing as a result of different linkage stereochemistry. However, the different templates did not vary the thermodynamic stability of the various carboproteins, as they all displayed similar denaturation curves.

1.5.2.6 Cholic Acid-Based Templates

Cholic acid is an amphiphilic steroid that contains three hydroxyl groups on one face, and a hydrophobic backbone with methyl groups on the other face. Because of its unique properties, cholic acid has been used for artificial ion channels (199), for drug targeting (200), and for a scaffold in the assembly of combinatorial libraries (201). Cholic acid is readily available, rigid and amphipathic, which make it a good candidate as a template for preorganizing the peptides for folding. Li and Wang reported the use of cholic acid as a new template for peptide assembly (202). The hydroxyls can be functionalized with groups such as maleimide to permit linkage with the peptides (see Figure 1.21).
The peptides used for linkage were the HIV-1 peptide inhibitor DP178, and the HIV-neutralizing antibody 2F5. Cysteine was added to either the N- or C-terminus for ligation purposes. The resulting three-helix bundles had higher helical content compared to a single peptide like all other TASP examples. The longer maleimide linker was found to yield proteins with higher helical content compared to the shorter linker. The longer term goal of using cholic acid-based proteins is to develop mimics of the trimeric gp41 proteins that are exposed during viral membrane fusion, so that an effective HIV-1 vaccine can be developed.
1.5.3 Examples of Template Assembled Synthetic Proteins with Function

For over a decade, attempts have been made to design artificial proteins mimicking some structural and functional properties of natural proteins. In the case of TASPs, the template provides synthetic access to an increasing number of chain topologies, which makes a variety of potential applications and biomimetic chemistry possible.

Mutter and coworkers found that several of their earlier designed four-helix TASPs showed ion channel forming potential (203). The original sequences were designed to be amphipilic so that the nonpolar side chains could bury themselves in a hydrophobic core. However, the relative orientation of the amphipilic segments depends on the polarity of the surrounding media. Thus, in lipid bilayers, these TASPS formed ionic channels through the reorientation of the helical modules so the nonpolar residues exposed themselves to the membrane, and the charged residues formed the hydrophilic pore.

A template is useful in that it can control the number of helices in a bundle. A TASP of Vpu, an ion channel forming membrane protein has been synthesized to study its mechanism (204). The precise number of Vpu monomers in the channel is unknown. To address this problem, both the tetrameric and pentameric proteins were synthesized with the assistance of a peptidic template with four and five attachment points, respectively. It was concluded that the five-helix bundle better modelled the structural motif of the conductive channel as its conductance was similar to the native Vpu.

A template of a TASP is also useful in controlling the topology of the protein structure. Mutter and coworkers developed a template assembled synthetic protein that inhibits platelet aggregation (205). Platelet adhesion is mediated by the interaction of the von Willebrand factor
(VWF) with its platelet receptor. The TASP was designed to mimic the active domain of the VWF and block the binding site for platelet activation.

Despite that present knowledge gap between the secondary and tertiary structures, the ability to construct proteins that mimic some essential features of natural proteins gives hope in understanding the relationship between structure and function.

1.6 Computer Modelling

The computation field has progressed to a point where important chemical, physical and biological questions can be addressed. Examples of such topics include protein folding and dynamics (8, 206). The first molecular dynamics simulation of a protein was reported in 1977 and consisted of a 9.2 ps trajectory for a small protein in vacuum (207). Since then, computer technology has become much more advanced, and molecular modelling using these high-speed computers has become significantly more popular (208). Dahiyat and Mayo used a computer program to design a de novo protein (209). They came up with an algorithm that predicted an optimal sequence for a given fold, and when this algorithm was put to the test, it worked. This success opened the door further to understanding the elements affecting protein folding. Chapter 3 discusses the use of molecular dynamics simulations to investigate a set of protein systems that were experimentally studied in Chapter 2.
1.7 Chapter Conclusion and Thesis Objectives

Decoding the protein folding problem that relates protein sequence to structure is one of the major goals of protein chemistry. As explained in this chapter, it is a significant challenge to understand this relationship as many factors are involved that affect the fold. *De novo* sequence design provides a means of investigating the principle structural features associated with protein architecture as they are simplified versions of real proteins. Designing, synthesizing and characterizing novel structures facilitate the identification of properties that increase our understanding of the rules associated with the folding process and overall structure. A further route of *de novo* design is using the ‘template assembly’ approach, which provides an additional driving force by lowering the entropic cost of protein folding. The template also provides more control options such as regulating the number of helices in a bundle protein, and directing the orientation of the helices. Although a great deal of research has gone into designing and studying *de novo* template assembled synthetic proteins, there is still much to learn about the relationship between the template and the protein. This thesis presents an investigation into the effects of various factors on the overall protein structure using both experimental and computational approaches. In particular, the relationship between the template and the tertiary fold of the protein are probed to understand what features are responsible for obtaining a native-like protein.

In Chapter 2, experimental results are presented on a series of template assembled four-helix bundle proteins, differing in linker lengths between the template and the peptides, using a cavitand as the template. The effects of the linker length on the structural properties of the proteins are examined to gain a better perspective of the effective diameter and position of the helices with respect to the cavitand template. The peptide sequence was designed in order to obtain a protein with a high degree of native-like character. An optimal linker length should
yield a protein with a well-defined structure. Chapter 3 presents the study of the same set of proteins as in Chapter 2 using molecular dynamics. The simulation and experimental results are compared, and the computer modelling data is used to explain the behaviour of the proteins studied experimentally. Chapter 4 describes a set of template assembled synthetic proteins, in which a sequence designed for a four-helix bundle is forced to form a five- and six-helix bundle, using a template. These proteins are characterized to test the peptide design and to determine if the sequence preferentially forms the four-helix bundle over the five- and six-helix bundles. Furthermore, not only are we concerned with learning the factors which affect protein folding, but we are also interested in new methods for obtaining native-like proteins. The last part of Chapter 4 explores the use of dynamic covalent chemistry with the hopes of one day being able to screen for native-like caviteins. The chapter goals and hypotheses are explained further in the individual chapters. Chapter 5 reviews the overall successes of the thesis, and discusses the final conclusions and the future direction of this work.
1.8 References


CHAPTER TWO: The Effect of the Linker Length Between the Template and the Peptides on the Structural Properties of the Caviteins*

2.0 Introduction

As mentioned in Chapter 1, the use of de novo proteins offers an attractive way to investigate protein structure and folding because they provide simplified models that are intended to maintain the fundamental structure of natural proteins. The examination of their structures can be further simplified by reducing or eliminating turns and multitudes of arrangements between the secondary structural units, such as their orientation, with the use of template assembled synthetic proteins (TASPs). The ‘template-assembly’ approach is known to enhance the helicity and the stability of the resulting proteins as compared to a single-stranded peptide (1). In all the examples demonstrated in Chapter 1.5.2, the template promoted folding and provided stability to the helical bundle proteins. However, in these cases, the effects of the linker type and lengths between the template and the peptides were not analyzed at all or not in much detail. For example, Chapter 1.5.2.4, Fairlie and coworkers examined the effects of different templates that varied in size, shape and directionality, and stated that when the linker length is sufficiently long, the formation of the TASP is less sensitive to the dimensions of the template (2). The proteins possessed similar helicities and stabilities regardless of the different types of templates. However, they had different slopes in their chemical denaturation curves, which were not accounted for. Moreover, the stability and the properties of these proteins were

* "A version of this chapter will be submitted for publication. Seo, E. and Sherman, J.C. The Effect of the Linker Length on the Structural Properties of the Caviteins."
expected to be affected with shorter linkers, but were not further explored. In another example (see Chapter 1.5.2.6), the linker length between the cholic acid-based template and the maleimide groups coupled to the peptides was found to affect the helicity. The longer linker gave a protein with a higher helical content, but this finding was not examined further.

Other past work examining the effect of residues outside the helix has focused on non-templated proteins. The effect of loops connecting the helices in a bundle have been studied rather than the linkers, due to the absence of a template in these examples (3-5). For instance, Regan investigated the length of the loops in the ROP protein and found that increasing its length resulted in decreased stability toward chemical and thermal denaturation (6). In contrast, Simons and coworkers showed that increasing loop flexibility enhanced the stability of their de novo proteins (6). Whether the studies involved loops between peptides, or linkers between the template and peptides, the contribution of the functional groups and residues outside the helix have been limited.

In our group, the template assembled synthetic proteins have been employed to examine the effects of the linker length between the template and the peptides in more detail. The template used by our group is a rigid macrocycle with an enforced cavity called a cavitand. The term cavitand was first introduced by Cram in 1982 (7), and is derived from a *resorcinarene*, which is shown in Figure 2.1. *Resorcinarenes* have been synthesized with different functional groups at the feet (R₁) and rim (R₂) positions (8). Cram discovered that a *resorcinarene* could be made rigid by adding methylene bridging groups between the adjacent phenolic groups to yield a cavitand.
Since *resorcinarennes* can be synthesized with a variety of feet and rim functional groups, these positions can be manipulated to create a cavitand with the desired functional groups. For the foot position of the cavitand, the effect on solubility was considered since the caviteins are examined in aqueous solvents. Typically, long alkyl chains were used in the feet positions of the cavitands in order to enhance their solubility in organic solvents (9). To increase their solubility in aqueous solvents, methyl groups were used in the feet positions. The methyl footed cavitands showed concentration independent $^1$H NMR spectra in D$_2$O, whereas cavitands with longer, more hydrophobic phenyl-ethyl feet showed concentration dependent spectra (10). Since concentration independent spectra are indicative of a monomeric species in solution and concentration dependent spectra are indicative of self-aggregation, the cavitands with methyl feet were used for the assembly of the proteins.

It is important to consider the reactivity of the functional groups at the rim positions of the cavitand. When an electrophilic aryl halide was incorporated at the rim positions, and reacted with nucleophilic thiol groups on the peptides, the yield was poor. Therefore, the reactivities
were reversed. Highly nucleophilic thiol groups were incorporated at the rims of the cavitand, and were reacted with peptides that contained electrophilic functional groups, which resulted in a better yield. Two synthetically available thiol-rimmed cavitands are the arylthiol and the benzylthiol templates (see Figure 2.2). For protein synthesis, the arylthiol cavitand has been found to generate more native-like structures compared to the benzylthiol cavitand (11). Figure 2.2 shows the difference in the linker region between the arylthiol and benzylthiol cavitand templates. In both cases, the preferred conformation was found to be the one in which the lone pairs on the sulfur atom were pointed away from the lone pairs of the oxygen atom to minimize the dipole-dipole interactions. The difference in the structural properties of the resulting caviteins using these two different cavitand templates was attributed to the difference in the attachment positions.

Figure 2.2. (a) Arylthiol and (b) benzylthiol cavitands. Note that for both cavitands, the preferred conformation has the lone pairs of sulfur directed away from the lone pairs on the bridged oxygen to minimize the dipole-dipole repulsions (from ‘MM2’ molecular mechanics calculations).
Furthermore, the attachment points of the arylthiol cavitand allow the peptides to be spaced at an almost ideal interhelical distance, which is typically about 7 to 14 Å (12). An added advantage of the cavitand template is its ability to hydrogen bond with an unsatisfied amide hydrogen of the peptide that it links onto (see Figure 2.3). The cavitand template acts as a good N-cap, providing hydrogen bond acceptors to the unsatisfied amide hydrogen atoms of the first residue at each of the four N-termini. The presence of a hydrogen bond was observed using $^1$H NMR and IR spectroscopy (13). The concentration independence of these molecules suggested that the hydrogen bonding was most likely due to intramolecular interactions. Two possible hydrogen bond patterns were found to exist and the model caviteins are shown below in Figure 2.3. The amide proton that is hydrogen bonded to the bridged oxygen, (see left side of Figure 2.3) has been observed in other cavitand-based systems (14). The amide proton that is hydrogen bonded to the rim sulfur atom (see right side of Figure 2.3) was postulated from the evidence that thiocresol, which contains no bridged oxygens, also showed weak hydrogen bonding. Although not as common, sulfur has been shown to form hydrogen bonds in some proteins (15).

![Figure 2.3. Potential hydrogen bonding interactions between the amide hydrogen of a residue, and the bridged oxygen of the arylthiol cavitand (left) or the arylthiol sulfur (right).](image-url)
Our group studies the four-helix bundle motif using the 'template assembly' approach, which differs from other groups who use this method, due to the use of a cavitand template. When the peptides are covalently attached onto the cavitand, the resulting protein has been termed cavitein (cavitand + protein). Scheme 2.1 shows the general synthesis of a four-helix bundle cavitein.

**Scheme 2.1.** Four-helix cavitein synthesis.

In previous work carried out by our group, different types and lengths of linkers between the cavitand template and the peptides were examined (11). This early work is briefly summarized, and the goals of the next step toward learning about the effects of the linker lengths are outlined. In addition, the nomenclature used for the peptide sequences in this thesis is described.

### 2.0.1 Effect of the Linkers

It has been shown that the linker lengths between the cavitand template and the peptides have a substantial effect on the protein's overall structure (11). If the linker is too short, steric
crowding at the template can destabilize the first turn, promoting exposure of its hydrophobic core, and/or lead to self-association. If the linker is too long, the stabilizing effect and the directing ability of the template are minimized. An ideal linker is somewhere in between these two extremes in which a well-defined, monomeric protein can be obtained without compromising its overall stability. Figure 2.4 illustrates the effect of the linker length in between the template and the peptides.

![Too Short](image1) ![Ideal](image2) ![Too Long](image3)

**Figure 2.4.** Effect of the linker length between the template and peptides.

The preliminary study of the linker length was carried out by incorporating different numbers of methylene units between the arylthiol cavitand and a phenylalanine residue protected as an ethyl ester (13). The incorporation of one to four methylene groups produced the corresponding products with high efficiency. Scheme 2.2 shows the synthesis of the model *de novo* proteins.
Scheme 2.2. Synthesis of model *de novo* proteins.

It was demonstrated that as the number of methylene groups increased in the linker region, the electrophilicity of the carbon next to the leaving group decreased. Furthermore, only the product with one methylene unit between the template and the phenylalanine ethyl ester was found to have the ability to form any substantial hydrogen bonding interactions with the cavitand template (see Figure 2.3).

To investigate the effect of the linker lengths on the template assembled synthetic proteins, one, two and four methylene linkers were incorporated between the arylthiol cavitand and the peptides. Concurrently, Gly residues were also used to study the effect of a different
linker type and length. The linker lengths between the arylthiol cavitand and the peptides were modified by varying the number of Gly residues. The peptide was designed to be amphiphilic so that hydrophobic bundling would be the primary driving force for folding. The sequence was designed to link from the hydrophilic/hydrophobic interface of the helices to the template, and is called S0 (see nomenclature section 2.0.3). The helical wheel of this sequence, which shows the arrangement of the hydrophobic and hydrophilic residues in a parallel four-helix bundle, is shown in Figure 2.5. Note that the hydrophobic residues bury themselves in the core of the bundle to minimize their exposure to the solvent.

Figure 2.5. Helical wheel diagram of the peptide sequence, S0 = EELLKKLEELLKKG, forming a four-helix bundle. Helices are oriented in a parallel fashion. Viewer is looking down the helical axes from C- to N-termini.
The sequence design followed a minimalist approach, which uses a minimum number of different residues (16). Apart from Gly, only three residues, Leu, Lys and Glu were incorporated in the design, which all possess high helical propensities. Oppositely charged amino acids were placed three to four residues apart for potential intrahelical interactions. Negatively charged Glu residues were incorporated closer to the N-terminus and positively charged Lys residues were incorporated closer to the C-terminus in order to reduce the macrodipole effect. The sequence was C-capped with a Gly, as it commonly occurs at the end of helices in natural proteins (17). Gly residues are suitable amino acids at the terminal ends of the peptides because they can satisfy the hydrogen bond requirements of the amide backbone, and also act as ‘helix breakers’ (see Chapter 1.2.2.3).

In earlier de novo work by other groups, peptides with protected side chains were only removed in the final step of the protein synthesis. However, protected peptides are fairly insoluble, which limits the length of the peptides. Purification of these peptides has also proven difficult, and linking the protected peptide segments has been found to be slow (18). Therefore, methods were developed to couple non-protected peptides segments to other non-protected peptides or templates (19). A nucleophilic thioacid at the C-terminus of one peptide, and a bromoacetyl group at the N-terminus of another peptide, were incorporated by Schnolzer and Kent in order to achieve chemoselective ligation for a thioester bond formation (20). The formation of thioester linkages have also been used by others (21, 22), but these bonds are quite unstable under basic conditions. Other methods of ligation were developed including disulfide (23), oxime (24, 25), and thioether (26, 27) bonds. In our group, both thioether bonds and disulfide bonds have been used to link the peptides to the cavitands due to their ease of formation and stability under basic conditions. In this chapter, thioether bonds were employed to
form the caviteins; in Chapter 4, disulfide bonds were used to synthesize a series of multi-stranded template assembled synthetic proteins.

The caviteins with various numbers of both methylene and Gly linkers were synthesized and characterized. The linker length was found to affect the cavitein's stability, native-like characteristics, and propensity to self-associate (11). Table 2.1 shows the results of work carried out by a previous member of our group, Dr. Adam Mezo. This earlier work is referred to as the first generation caviteins, and the new work presented in this thesis is referred to as the second generation series. The nomenclature used in this thesis is explained in section 2.0.3. The experimental detail and interpretation of the results are described in the later sections when comparing to the second generation caviteins, as analogous steps were carried out and similar concepts were considered with both studies.
Table 2.1. Summary of the experimental results of the first generation caviteins with different types and lengths of linkers carried out by Dr. Adam Mezo. Here, all of the peptide sequences are considered to be part of the four-helix bundle, and therefore referred to as caviteins. Note that the dimerization of the 1GS0 cavitein has been taken into account for the calculation of the stability whereas for the 4(CH$_2$)$_2$SO and the Bzl 0GS0 caviteins, no dimer was detected at the denaturation point and therefore their stabilities were based on a monomeric species.

<table>
<thead>
<tr>
<th>Caviteins</th>
<th>Far UV CD Spectra</th>
<th>AUC data</th>
<th>Stabilities (kcal/mol)</th>
<th>Near UV Absorption in CD Spectra</th>
<th>$^1$H NMR Spectra</th>
<th>ANS Binding Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>0GS0 = 1(CH$_2$)$_2$SO</td>
<td>$\alpha$-Helical</td>
<td>Monomer/Dimer</td>
<td>-11.9 ± 1.5</td>
<td>Weak</td>
<td>Poorly dispersed</td>
<td>No binding</td>
</tr>
<tr>
<td>1GS0</td>
<td>$\alpha$-Helical</td>
<td>Dimer</td>
<td>-22.9 ± 1.9</td>
<td>Very strong</td>
<td>Very well-dispersed</td>
<td>No binding</td>
</tr>
<tr>
<td>2GS0</td>
<td>$\alpha$-Helical</td>
<td>Monomer</td>
<td>-9.9 ± 0.3</td>
<td>Strong</td>
<td>Well-dispersed</td>
<td>No binding</td>
</tr>
<tr>
<td>3GS0</td>
<td>$\alpha$-Helical</td>
<td>Monomer</td>
<td>-10.8 ± 0.5</td>
<td>Strong</td>
<td>Poorly dispersed</td>
<td>No binding</td>
</tr>
<tr>
<td>2(CH$_2$)$_2$SO</td>
<td>$\alpha$-Helical</td>
<td>Monomer</td>
<td>-9.9 ± 0.9</td>
<td>Very weak</td>
<td>N/A</td>
<td>Weak binding</td>
</tr>
<tr>
<td>4(CH$_2$)$_2$SO</td>
<td>$\alpha$-Helical</td>
<td>Dimer</td>
<td>-11.9 ± 1.0</td>
<td>Weak</td>
<td>N/A</td>
<td>Weak binding</td>
</tr>
<tr>
<td>Bzl 0GS0</td>
<td>$\alpha$-Helical</td>
<td>Dimer</td>
<td>-8.6 ± 0.3</td>
<td>Very weak</td>
<td>Very poorly dispersed, broad signals</td>
<td>Weak binding</td>
</tr>
</tbody>
</table>

In general, the caviteins composed of the methylene linkers were found to possess more molten globule-like structures, and the caviteins composed of the Gly linkers were found to possess more native-like proteins. Figure 2.6 depicts the 4(CH$_2$)$_2$SO cavitein and the 1GS0 cavitein to show their differences. These two caviteins are shown for comparison because they contain the same number of atoms between the template and the peptide, and therefore the distance to the template is not a factor. Although both linkers are fairly flexible, the Gly linker is slightly more rigid due to the planarity of the peptide bond. The Gly linker also contains an
amide hydrogen atom and a carbonyl oxygen atom that have potential to hydrogen bond with a bridged oxygen of the cavitand, and with an unsatisfied backbone hydrogen, respectively.

Figure 2.6. Methylene variant, \(4(CH_2)_2S0\) cavitein versus glycine variant, 1GS0 cavitein. Note both caviteins have the same number of atoms between the sulfur atom of the cavitand and the peptide.

Out of the Gly linker variants, at least two flexible glycine residues in the linker region were required to obtain a well-defined, monomeric cavitein. Although the 2GS0 cavitein possessed a more-dispersed \(^1\)H NMR spectrum, which is indicative of a native-like protein, the 3GS0 cavitein contained sharper peaks in the amide region, which is also indicative of a native-like protein. In addition, the 3GS0 cavitein was slightly more stable than the rest of the monomeric species using the first generation S0 sequence. From these results, it can be stated that the peptide sequence alone was insufficient to generate native-like caviteins, and that an optimal linker was required to induce a well-defined structure.
2.0.2 Goals for the Second Generation of Caviteins

The second generation of caviteins was designed and synthesized to further examine the role of the linkage between the template and the peptides by linking the helices onto the template from a different point of the bundle with respect to the helix faces. The first generation sequence used in the earlier work was linked onto the template from the hydrophobic/hydrophilic interface (see Figure 2.5). The work presented here for the second generation of caviteins uses a series of peptides that was designed to attach from the hydrophobic face of the helical bundle to the template. This change in linkage point was achieved by adding a hydrophobic residue to the beginning of the original sequence, which would result in linkage occurring 1/3.6 of a turn closer into the hydrophobic core of the helical bundle. The idea was to determine if changing the attachment point would allow the peptides to link closer to the cavitand template. If the linker length requirement could be reduced, then the packing between the helices could be improved, which in turn could enhance the native-like character of the caviteins. Furthermore, although it has been found that the optimal interhelical distances range from 7 to 14 Å (12), the distance between the helices in the caviteins has not been investigated. By determining the linker length requirement for the second generation of caviteins and comparing it with the first generation series, it should be possible to achieve a better understanding of the distances between the adjacent helices. Two to three Gly linkers were necessary to obtain a well-defined cavitein in the first generation series. If the linker length requirement for the second generation of caviteins was found to be the same, the interhelical distances should be smaller than if the linker length requirement was shorter (this concept is explained in more detail in the conclusion section). The second generation of caviteins were characterized in terms of helical content, stability, propensity to self-aggregate, and the ability to form a native-like structure. The caviteins
consisting of different linker lengths are compared to each other, as well as to the caviteins of the first generation series. Only the Gly residues were used to further study the linker length effect, since this linker type was found to produce more native-like caviteins.

### 2.0.3 Nomenclature

The nomenclature that is used in this thesis is described here. The sequence used for the first generation of caviteins is denoted S0. The results of this previous work are shown in Table 2.1. The number before the sequence denotes the number of glycine residues at the beginning of the sequence starting from the N-terminus (i.e. 1GS0 is the peptide S0 with 1 Gly on the N-terminus). For the methylene variants, the number in front denotes the number of methylene linkers, CH₂, and the sequence S0 is still used. For example, the 2(CH₂)S0 cavitein has two methylene linkers between the cavitand template and the peptide sequence S0. Unless otherwise specified, all caviteins in this chapter were synthesized with the arylthiol cavitand. The Bzl term in the last entry of Table 2.1 refers to a cavitein synthesized with a benzylthiol cavitand. The Bzl 0GS0 cavitein represents a cavitein made from a benzylthiol cavitand and the peptide sequence 0GS0.

The intermediate sequence S1, was designed to link from the hydrophobic centre of the helices. A leucine was added to the beginning of the sequence S0 in order to minimize the number of changes. The new sequence S2, incorporates Ala at the beginning of the S0 sequence rather than Leu. The set of caviteins using the sequence S0 is referred to the first generation, and the caviteins using the sequence S2 is referred to as the second generation series for the rest of this thesis. Table 2.2 shows the peptide names and sequences used in this thesis. To refer to the
template assembled synthetic protein (TASP), the word cavitein is placed after the sequence name.

**Table 2.2.** First generation peptide sequence, S0, and the second generation peptide sequence, S2. The term cavitein follows the peptide sequence name when referring to the four-helix TASP.

<table>
<thead>
<tr>
<th>Peptide Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First Generation</strong></td>
<td></td>
</tr>
<tr>
<td>S0 = 0GS0</td>
<td>EELLKKLEELLKKKG</td>
</tr>
<tr>
<td>1GS0</td>
<td>GEELLKKLEELLKKKG</td>
</tr>
<tr>
<td>2GS0</td>
<td>GGEELLKKLEELLKKKG</td>
</tr>
<tr>
<td>3GS0</td>
<td>GGGGEELLKKLEELLKKKG</td>
</tr>
<tr>
<td><strong>Intermediate</strong></td>
<td></td>
</tr>
<tr>
<td>S1 = 0GS1</td>
<td>LEELLKKLEELLKKKG</td>
</tr>
<tr>
<td>2GS1</td>
<td>GGLEELLKKLEELLKKKG</td>
</tr>
<tr>
<td>S2 = 0GS2</td>
<td>AEEELLKKLEELLKKKG</td>
</tr>
<tr>
<td><strong>Second Generation</strong></td>
<td></td>
</tr>
<tr>
<td>1GS2</td>
<td>GAEELLKKLEELLKKKG</td>
</tr>
<tr>
<td>2GS2</td>
<td>GGAEEELLKKLEELLKKKG</td>
</tr>
<tr>
<td>3GS2</td>
<td>GGGAEEELLKKLEELLKKKG</td>
</tr>
<tr>
<td>4GS2</td>
<td>GGGGAEEELLKKLEELLKKKG</td>
</tr>
</tbody>
</table>
2.1 Results and Discussion

2.1.1 Design of the Second Generation Caviteins

The choice of the template and design of the peptide sequence used to study the linker length effect of the second generation caviteins is explained. Subsequently, the synthesis of these caviteins using the chosen template and newly designed sequence is described.

2.1.1.1 Template Choice and Synthesis

Cavitands are useful templates for the assembly of proteins because they are rigid structures that can provide stability, but are also flexible in that they can react with peptides differing in linker lengths. Cavitands can be synthesized with a variety of functional groups which can regulate their solubility and reactivity. They can also regulate the number of peptides in a bundle and control the orientation of these helices. In particular, the arylthiol cavitand is an excellent template for the assembly of the peptides because it can promote the formation of native-like proteins, space the helices apart similar to natural proteins, and provide potential binding sites for small molecules and ions.

Past literature describes the synthesis of the arylthiol cavitand (7, 13). Here, the synthesis is briefly described with a couple of modifications in the last step. Resorcinol 1 and acetaldehyde were reacted in methanol and HCl. Resorcinarene 2 slowly precipitated out of the mixture over
the course of a week (78 % yield). In the next reaction, N-bromosuccinimide (NBS) in 2-
butanone was added to give tetrabromoresorcinarene 3 (64 %). This macrocycle was then
bridged with bromochloromethane in the presence of potassium carbonate to afford
bromocavitand 4 (47 %). Tetrabromocavitand was treated with n-butyllithium at -78 °C and
quenched with S₈ to give a mixture of both tris-cavitand (not shown) and tetra-arylthiol cavitand
5 (~ 95 %) (13). Scheme 2.3 shows the synthesis. The final step of this synthesis was slightly
modified from the literature procedure and is explained in the experimental section. Furthermore,
the original procedure separated the cavitand mixture using silica gel column chromatography to
yield the pure arylthiol cavitand. The separation of this mixture was difficult and unnecessarily
time consuming as it was much easier to separate the products once the peptides were linked
onto the cavitand template.

Scheme 2.3. Synthesis of the methyl-footed arylthiol cavitand.
2.1.1.2 Sequence Design and Synthesis

The sequence employed in this study was based on the first generation sequence, S0, used to examine the effect of the linkers between the peptides and the template (11). In order to link the cavitand template from a different point of the peptide than the first generation sequence, but to maintain all other factors (intrahelical and interhelical salt bridges, macrodipole reducing effects), a logical modification was to add a hydrophobic residue at the beginning of the N-terminus. This new sequence, S1, consisted of adding Leu to the N-terminus of the first generation sequence. This way, the S1 sequence would remain amphiphilic and still consist of a minimum number of different types of amino acids. Leu, which has a high helical propensity, was the only nonpolar residue used in this new sequence. Glu and Lys, the negatively and positively charged residues, respectively, were placed three to four amino acids apart for potential intrahelical salt bridge formation. Glu residues were placed closer to the N-terminus and Lys residues were placed near the C-terminus in order to minimize the macrodipole effect. In addition, Gly was used as the last residue at the C-terminus end. Finally, the C-terminus was amidated to remove any potentially charged groups, mimicking more closely the environment of an α-helix within a protein, and also minimizing the repulsive interactions of the macrodipole.

Since the chosen cavitand template has nucleophilic thiol groups at the rim positions, an electrophilic group was incorporated at the end of the peptide. Linkage of the peptides from the N-terminus rather than the C-terminus was more appropriate because the arylthiol cavitand serves as a better N-cap due to its hydrogen accepting ability. A suitable peptide activating group was the chloroacetyl functional group because it is synthetically simple to incorporate at the N-terminus of the peptide, and it provides a suitable electrophilic site for attack by the nucleophilic thiol groups on the cavitand template.
However, when this peptide with no Gly linkers (0GS1 peptide) was reacted with the cavitand template under the cavitein synthesis conditions (see experimental section), no desired four-helix bundle product was obtained. In fact, no observable product resulted, and the recovered starting peptide was significantly reduced. To test the linkage capability of this new sequence, S1 was reacted with thiocresol under the same conditions as above, and the linked product was obtained as shown in Scheme 2.4.

![Scheme 2.4](image)

**Scheme 2.4.** Thiocresol reacts with peptide sequence S1.

Two glycine residues were also incorporated at the N-terminus of the S1 sequence (2GS1 peptide), and this peptide linked onto the cavitand template to give the four-helix bundle. If the incomplete linkage of the 0GS1 peptide to the cavitand was solely due to steric crowding near the bowl, one would still expect a mono-, di- or tri-substituted product; however, none of these species were obtained. A reasonable explanation for the failure of the cavitein synthesis using the 0GS1 peptide sequence could be attributed to aggregation. If for example, there were too many steric interactions for all four peptides to link onto the bowl, a partially substituted product
may have resulted instead. Consequently, the hydrophobic face of the peptides could have been exposed, resulting in aggregation and loss of solubility. Since the S1 sequence was unsuccessful in forming a four-helix bundle without any Gly residues, a newer sequence was designed. This newer sequence S2, also called the second generation sequence, employed an Ala residue at the beginning of S0 rather than a Leu residue. Ala was a good candidate because it is highly helical, still moderately hydrophobic and is smaller than Leu, and therefore may reduce the steric problem during linkage to the template. This S2 peptide linked onto the cavitand template without any Gly linkers to form a four-helix bundle, demonstrating that the N-terminal Leu in the S1 sequence was probably too large for all four helices to link directly onto the template. Thus, the S2 sequence was chosen to probe the linker length effects in this new study. Figure 2.7 shows the helical wheel diagram of the S2 sequence in a four-helix bundle. An arrow points to where the first generation sequence would link (i.e. from the hydrophobic/hydrophilic interface of the peptides) compared to the second generation sequence which was designed to link closer to the centre of the hydrophobic face.
Figure 2.7. Helical wheel diagram of the sequence, $S_2 = AEELLKLEELKKG$. Helices are oriented in a parallel fashion to form the four-helix bundle. Viewer is looking down the helical axes from C- to N-termini.

2.1.1.3 Cavitein Synthesis

The second generation caviteins were synthesized using the arylthiol cavitant template and the newly designed peptide sequence with a varying number of Gly linkers. The synthetic scheme of the four-helix cavitein using the new sequence, $S_2$, is shown in Scheme 2.5.
For each cavitein synthesis, one equivalent of the methyl footed arylthiol cavitand and excess (8 equiv.) peptide were reacted in the presence of diisopropylethylamine (DIPEA) in dimethylformamide (DMF) for 5 hours at room temperature to afford the corresponding caviteins in high yield. The purity of all the caviteins was assessed by analytical reversed-phase HPLC, and their masses were confirmed by MALDI mass spectrometry.

2.1.2 Characterization of the Caviteins

The structure and properties of the caviteins in the second generation series were analyzed in several ways. These techniques included circular dichroism (CD) spectroscopy, analytical ultracentrifugation, one- and two-dimensional \(^1\)H NMR spectroscopy, and hydrophobic dye binding studies.
Note that the non-templated peptides were also examined in terms of stability and helicity. They were found to be less stable and less helical than their corresponding four-helix bundle, which was consistent with the rest of the TASP examples presented in Chapter 1. However, the analysis of the single-stranded peptides are kept to a minimum because they have been studied extensively by other groups (28-31).

2.1.2.1 Far UV Circular Dichroism (CD) Spectra

Circular dichroism spectroscopy is a useful technique in determining the type and amount of secondary structure present in a peptide or a protein (32, 33). In the far UV region, which is between 190 and 240 nm, the amide chromophore of a peptide bond dominates the CD structure. An α-helix contains one positive band near 195 nm and two negative bands near 208 and 222 nm. The negative band at 222 nm is due to the n → π* amide transition resulting from the strong hydrogen bonding environment of this type of secondary structure, and the negative band at 208 nm and the positive band at 195 nm are due to the π → π* amide transitions (32). The transition at 222 nm is relatively independent of the helix chain length, whereas the bands at 208 nm and 195 nm are reduced in shorter helices (34). CD spectra of low (4 μM) and high (40 μM) sample concentrations were obtained for all the caviteins, and both concentrations gave overlapping spectra, thus, only the high concentration data are shown. Figure 2.8 shows the CD spectra of the second generation caviteins in the linker series. Typically, the secondary structure changes as a protein aggregates. Therefore, this concentration independence suggests that the caviteins exist as monomers, or as very stable dimers. However, this statement is only true if the helicity of the
caviteins increased with aggregation, which may have not been the case. In other words, a concentration dependent curve would indicate the presence of a self-aggregating species. Lack of this dependence merely suggests that the species may have been monomeric. Sedimentation equilibria studies were carried out to determine the oligomeric state of the caviteins in solution (see section 2.1.2.4).

Figure 2.8. CD spectra of the second generation cavitein series. Samples are 40 µM in 50 mM phosphate buffer, pH 7.0 at 25 °C.
From the CD spectra, all of the caviteins in the second generation series, 1GS2, 2GS2, 3GS2 and 4GS2 were found to be α-helical like the first generation series, as they contained the three bands characteristic of this type of secondary structure. The fraction α-helix, $f_H$, for each of the caviteins, excluding the Gly linkers, was calculated using the mean residue ellipticity at 222 nm. The calculated values are shown for each of the caviteins in Table 2.3.

Table 2.3. Fraction α-helix, $f_H$, of the caviteins in the second generation series using the equation from the experimental section.

<table>
<thead>
<tr>
<th>Caviteins</th>
<th>$[\theta]_{222}$ (deg cm$^2$ dmol$^{-1}$)</th>
<th>$f_H$ (without Gly linkers)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1GS2</td>
<td>-14900</td>
<td>0.379</td>
</tr>
<tr>
<td>2GS2</td>
<td>-17700</td>
<td>0.447</td>
</tr>
<tr>
<td>3GS2</td>
<td>-25500</td>
<td>0.638</td>
</tr>
<tr>
<td>4GS2</td>
<td>-19500</td>
<td>0.492</td>
</tr>
</tbody>
</table>

There was no general trend in the amount of helicity as the linker length increased. From the calculations, the 1GS2 cavitein contained the least amount of helicity, and the 3GS2 cavitein contained the highest amount of helicity.

A CD spectrum is the linear sum of the individual secondary structures plus the contributions from aromatic chromophores. Aromatic side chains are known to contribute significantly to the far UV region of a CD spectrum (35, 36). Woody reported that Tyr contributes positively 5000 deg cm$^2$/dmol to the $[\theta]_{222}$ value, and that Phe contributes positively 5800 deg cm$^2$/dmol (35). Baldwin and coworkers showed the same contribution for a Tyr residue, and stated that a Trp residue contributes negatively 3000 deg cm$^2$/dmol to the $[\theta]_{222}$ value when incorporated at the N-terminus of a protein (36). Although the designed caviteins in this thesis do not contain aromatic side groups, they contain a cavitand template at the N-terminus which is composed of aromatic rings. Therefore, the cavitand chromophore could have
contributed significantly to the far UV spectrum, and varied for each cavitein. This contribution is expected to be maximal with the least number of Gly residues, which could explain the lower \([\theta]_{222}\) value of the 1GS2 cavitein. The smaller content of \(\alpha\)-helicity in the 1GS2 cavitein could have been due to some distortion caused by a shorter linker near the cavitand template. The 2GS2 cavitein may have an artificially low helical content due to the contribution from the near UV region (~240-300 nm).

Generally, the ratio of \([\theta]_{222}\) to \([\theta]_{208}\) offers some information about the protein structure. Protein scientists have found that a value of \([\theta]_{222}/[\theta]_{208}\) greater than 1.0 indicates that the \(\alpha\)-helices are stabilized by interacting with each other in a tertiary structure, and that a value of less than 1.0 indicates the presence of a single non-interacting \(\alpha\)-helix (37, 38). The \([\theta]_{222}/[\theta]_{208}\) ratio of all the caviteins were less than 1.0, but this value was most likely skewed by the cavitand template, as the \(\alpha\)-helices in the cavitein were found to interact from other experiments. Therefore, the value of \([\theta]_{222}/[\theta]_{208}\) was not used here to analyze the data.

2.1.2.2 Near UV CD Spectra

The far UV region of a CD spectrum gave information about the secondary structure. Native-like structures and molten globule proteins contain the same amount of secondary structure, and therefore, they cannot be distinguished by the far UV region. Conversely, an aromatic absorption in the near UV region can be observed in the presence of any non-averaged structural elements near this chromophore (39). In the case of the caviteins, the near UV CD spectrum was used to differentiate between the two types of tertiary structures as the cavitand
chromophore absorbs in the near UV region. Molten globules normally show an absence or reduction of the signals in the near UV region because of their time-averaged fluctuating structures (40). Native-like structures possess more intense absorptions in this region due to their specific packing, which causes the aromatic chromophores to be detected in an asymmetric environment. Figure 2.9 shows the expanded near UV region of the caviteins in the second generation series.

![Figure 2.9. Near UV CD spectra of the second generation caviteins. Samples are 40 μM in pH 7.0 phosphate buffer at 25 °C.](image)

The 1GS2 and 4GS2 caviteins show the most reduced signals in the near UV region, indicating the absence of any non-averaged structural elements near the cavitand. The short linker in the 1GS2 cavitein may have restricted the hydrophobic core from forming specific side
chain interactions thereby resulting in a molten globule structure. The lack of signal in the near UV region of the 4GS2 cavitein is most likely due to the protein structure being far away from the cavitand. The 2GS2 cavitein displays the most enhanced signal compared to the rest of the caviteins in both the first (not shown) and second generation series, implying that it is less dynamic, and packed more specifically near the cavitand chromophore. However, this statement is only directly relevant for the shorter linker variants (i.e. caviteins with one Gly or two Gly linker residues), since the effect of the cavitand chromophore is reduced as the linker length increases. The 1D $^1\text{H}$ NMR spectra of these caviteins were obtained and inspected to further examine their native-like character (see section 2.1.2.5).

The sign of the signals in the near UV region were used to gain some information about the supercoiling ability of the helices near the cavitand. Both the 2GS2 and the 3GS2 caviteins show a positive absorption at around 248 nm, and a negative absorption at around 265 nm for the 3GS2 cavitein and 280 nm for the 2GS2 cavitein. This effect may suggest the presence of a supercoiled structure. The 2GS2 and 3GS2 caviteins are likely supercoiling in the same direction due to the same pattern of the positive and negative signals. A reversal of the positive and negative absorbances may indicate a structure that is supercoiling in the opposite direction. The 1GS2 cavitein shows a slight reversal from the 2GS2 and 3GS2 caviteins. The 4GS2 cavitein may not be supercoiling, or may be supercoiled, but too remote from the cavitand to be observed.
2.1.2.3 Effect of Guanidine Hydrochloride

The stability of the proteins can be assessed using chemical denaturants such as urea and guanidine hydrochloride (GuHCl) (41). The precise mechanism of how these denaturants work is still not well understood, but there is evidence that both can bind to the surface of a protein and disrupt the hydrogen bonding and hydrophobic interactions. Earlier work in our group showed that using urea up to 10.0 M changed the helicity of a cavitein less than 5 %, demonstrating the high stability of these TASP's. Urea only affects the hydrogen bonding and hydrophobic interactions, but not the electrostatic interactions, whereas GuHCl is a salt that can disrupt all of these interactions (42). Therefore, in the case of the robust caviteins, GuHCl was used for the stability studies as it is a more effective denaturant.

Figure 2.10 shows the denaturation curves of the caviteins and a single-stranded control peptide 0GS2. GuHCl was found to completely denature the single-stranded peptide, and partially denature the caviteins. The denaturation experiments were carried out using cavitein samples of both low (4 μM) and high (40 μM) concentrations. All the caviteins gave concentration independent denaturation curves suggesting that the caviteins exist as monomers under denaturating conditions. Since the data from the high and low concentrations overlapped, only the high concentrations curves are shown in Figure 2.10. The fraction folded (FF) was determined by dividing the [θ]222 at a specific concentration (from 1.0 M to 8.0 M in 0.25 M increments) by the [θ]222 at 0 M GuHCl.
Figure 2.10. Guanidine hydrochloride-induced denaturation curves for the second generation caviteins and the 0GS2 peptide, all at 40 μM concentrations in pH 7.0 phosphate buffer at 25 °C. Error bars calculated as explained in the experimental section have been omitted for clarity. In addition, although the samples were measured at 0.25 M increments of GuHCl, only points for every 1M increment of GuHCl are shown here.

From Figure 2.10, it is apparent that the caviteins with varying linker lengths possess different stabilities. A crude way of visually examining the stabilities of the caviteins is to look at the concentration of GuHCl required to unfold half the protein, [GuHCl]_{1/2}. A more accurate way for calculating the free energies of folding is to use nonlinear least-squares fitting (43). This method is used assuming that the unfolding transition is a reversible, two-state process, and the free energy of unfolding varies linearly with the concentration of the denaturant. The following
equation relates the observed free energy of folding, $\Delta G_{\text{obs}}^\circ$, with the free energy of folding in the absence of the denaturant $\Delta G_{H_2O}^\circ$.

$$\Delta G_{\text{obs}}^\circ = \Delta G_{H_2O}^\circ - m[\text{GuHCl}]$$

where $m$ is the change in $\Delta G_{\text{obs}}^\circ$ as a function of [GuHCl], which gives an estimate of the cooperativity during the unfolding process. The denaturation midpoint ([GuHCl]$_{1/2}$), the $m$ value, and the calculated $\Delta G_{H_2O}^\circ$ values for the caviteins in the second generation linker series are presented in Table 2.4.
Table 2.4. Guanidine hydrochloride-induced denaturation data of the caviteins in the second generation series. Note that the calculations were performed on cavitein samples at 4 and 40 μM, and the values for both concentrations were the same within error.

<table>
<thead>
<tr>
<th>Caviteins</th>
<th>[GuHCl]_{1/2} (M)</th>
<th>$m$ (kcal/molM)</th>
<th>$\Delta G_{H_2O}^o$ (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1GS2</td>
<td>5.4 ± 0.1</td>
<td>-1.4 ± 0.1</td>
<td>-8.3 ± 0.3</td>
</tr>
<tr>
<td>2GS2</td>
<td>6.9 ± 0.1</td>
<td>-2.1 ± 0.1</td>
<td>-14.1 ± 0.7</td>
</tr>
<tr>
<td>3GS2</td>
<td>5.3 ± 0.1</td>
<td>-1.8 ± 0.1</td>
<td>-10.6 ± 0.3</td>
</tr>
<tr>
<td>4GS2</td>
<td>4.9 ± 0.1</td>
<td>-1.5 ± 0.1</td>
<td>-7.9 ± 0.3</td>
</tr>
</tbody>
</table>

The stabilities for the 1GS2 and 4GS2 caviteins are the lowest, again indicating that the optimal linker for the S2 peptide sequence is likely between these two linker lengths. Compared to caviteins in the first and second generation series (with the exception of the stable 1GS0 dimer), the 2GS2 cavitein has the most negative $\Delta G_{H_2O}^o$ value, hence the greatest stability. Similarly, the 2GS2 cavitein has the highest $m$ value of all the caviteins, again with the exception of the first generation 1GS0 cavitein. Compared to other template assembled synthetic proteins (44, 45), the caviteins were found to be generally more stable. The $m$ values obtained for the caviteins are comparable to those found for natural proteins (46). A high $m$ value corresponds to high cooperativity in folding, and high cooperativity has been related to native-like character (47). The high $m$ value of the 2GS2 cavitein shows that this species exhibits the most native-like properties. This value has also been found to correlate with solvent accessible surface areas (48). A larger $m$ value means a larger protein surface is exposed. For the caviteins shown in Table 2.4, the 2GS2 cavitein appears to have the highest cooperativity and/or have the most exposed surface area.
The $\Delta G^o_{\text{H}2\text{O}}$ values should be interpreted with some scepticism because there is no theoretical proof for assuming that the free energy of unfolding varies linearly with denaturant concentration. At high concentrations of denaturants, this assumption may fail and introduce uncertainty. Additional errors are associated when the caviteins do not demonstrate a full unfolding transition. The 2GS2 cavitein has the largest error for its $\Delta G^o_{\text{H}2\text{O}}$ value due to the lack of a complete post-denaturation curve. Furthermore, errors significantly increase as the length of extrapolation is expanded. In other words, larger errors are incorporated when the denaturation midpoint is higher (43).

All the caviteins demonstrated concentration independent curves, which may be indicative of a monomeric species; however, the presence of the GuHCl salt may have prevented self-aggregation. Thus, sedimentation equilibria studies were carried out in the absence of GuHCl to determine the oligomeric state of the caviteins in solution.
2.1.2.4 Oligomeric State

The circular dichroism (CD) spectra and the guanidine denaturation curves of the caviteins showed concentration independent results, which suggests that these proteins may be monomeric in solution. However, certain assumptions were made to make these suggestions. For example, it was assumed that the self-association of a cavitein was associated with a change in helicity, which may not necessarily be the case. In addition, the presence of guanidine hydrochloride (GuHCl) may have inhibited aggregation that could have otherwise taken place in the absence of this salt. Thus, for the reasons above, a more rigorous method of determining the self-associative state of the caviteins in solution was required. Sedimentation equilibria studies using an analytical ultracentrifuge can be used to determine molar mass, association constants, stoichiometries, and non-ideality (49, 50). Here, it was used to approximate the molecular weight and determine the oligomeric state of the caviteins in solution.

A centrifugal force was applied to the solution until equilibrium was reached. At this point, the flux of the sedimenting molecules was balanced exactly by the flux of diffusing molecules at each radial position. As a result, a time-invariant concentration gradient developed. Analysis of the concentration gradient at equilibrium gave information about the molecular weight of the protein in solution, and thus determined the degree of self-association (see experimental section).

The caviteins were studied at various concentrations (20, 40, 60 μM) in 50 mM pH 7.0 phosphate buffer at 20 °C and at three different rotor speeds (27000, 35000, 40000 rpm). The experimental section describes the experiments in more detail and also contains the fits of the raw data to the theoretical curves. For all the caviteins, the best fit to the data was obtained using a model describing a single, ideal species. If a cavitein was found to possess a substantially
larger molecular mass than its monomeric weight (i.e. aggregating), the self-associating model was used to determine the association constant. Table 2.5 gives the results of the sedimentation equilibria studies.

**Table 2.5.** Molecular weight estimations by sedimentation equilibria for caviteins at 20 °C in 50 mM phosphate buffer, pH 7.0, when fit to a single, ideal species. Solvent density was estimated to be 1.00 g/mL. A monomer-dimer fit was also carried out for the species that were found to self-aggregate to determine the association constants.

<table>
<thead>
<tr>
<th>Caviteins</th>
<th>Calculated Monomer Molecular Weight (Da)</th>
<th>Experiment Estimated Molecular Weight (Da)</th>
<th>Partial Specific Volume (mL/g)</th>
<th>Predominant Species</th>
<th>Association Constant, $K_{a,2}$ (Absorbance)</th>
<th>Association Constant, $K_{a,2}$ (M$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1GS2</td>
<td>8069</td>
<td>12200 ±900</td>
<td>0.78</td>
<td>Monomer/Dimer</td>
<td>$3.8 \pm 0.3$</td>
<td>$(3.4 \pm 0.4) \times 10^4$</td>
</tr>
<tr>
<td>2GS2</td>
<td>8297</td>
<td>9300 ±700</td>
<td>0.78</td>
<td>Monomer</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>3GS2</td>
<td>8525</td>
<td>16100 ±900</td>
<td>0.78</td>
<td>Dimer</td>
<td>$150 \pm 20$</td>
<td>$(1.8 \pm 0.3) \times 10^6$</td>
</tr>
<tr>
<td>4GS2</td>
<td>8754</td>
<td>9700 ±300</td>
<td>0.77</td>
<td>Monomer</td>
<td>-------</td>
<td>-------</td>
</tr>
</tbody>
</table>

From fitting to the single ideal species model, the estimated molecular weights demonstrated that the 1GS2 cavitein existed as a monomer/dimer, the 2GS2 and 4GS2 caviteins existed as monomers, and the 3GS2 cavitein existed as a dimer. When the 1GS2 cavitein raw data was fit using the self-associating model, the association constant for a monomer/dimer equilibrium, $K_{a,2}$, was found to be $(3.4 \pm 0.4) \times 10^4$ M$^{-1}$ (when $\varepsilon_{270} = 15000 \pm 700$ cm$^{-1}$ M$^{-1}$). When the 3GS2 cavitein raw data was fit to a monomer/dimer equilibrium, $K_{a,2}$ was found to be $(1.8 \pm 0.3) \times 10^6$ M$^{-1}$ (when $\varepsilon_{270} = 20000 \pm 1000$ cm$^{-1}$ M$^{-1}$).
The solution molecular weight of a protein is dependent on the solution density ($\rho$) and the partial specific volume ($\bar{v}$) of the solute (51). The density of the solvent is the mass in grams of one mL of solvent and is dependent on its temperature and composition. The partial specific volume of a solute is defined as the change in volume of the solution per gram of solute added, which corrects for hydration, solute binding and electrostriction effects (52), and is also dependent on the temperature. An accurate estimation of both these parameters is essential for an accurate molecular weight determination. The solution density was estimated readily from the buffer composition (51). The estimation of the partial specific volume ($\bar{v}$) was much more difficult. The $\bar{v}$ value is typically estimated from the amino acid composition. However, this method results in significant errors if the proteins have a non-globular structure or exhibit preferential hydration. Moreover, proteins with non-amino acid components could lead to additional errors. In the case of the caviteins, the non-amino acid cavitand component is approximately 10% of the protein's molecular weight. Small errors in $\bar{v}$ can manifest themselves as large errors in the molecular weight (51). For example, if the $\bar{v}$ value for the 2GS2 cavitein is changed from 0.78 to 076, the molecular weight would change from 9300 Da to 8500 Da, which is a difference of 800 Da. Since the cavitand contribution to the partial specific volume is unknown, the estimated molecular weights in solution may not be entirely accurate. Nevertheless, the molecular weight range was still within that of a monomer, and thus valuable information can be obtained from the sedimentation equilibria studies.
2.1.2.5 \(^1\)H Nuclear Magnetic Resonance (NMR) Spectra

Nuclear magnetic resonance (NMR) spectroscopy has been used to study protein structures for several decades. For this project, one-dimensional (1D) \(^1\)H NMR spectroscopy (53, 54) and two-dimensional (2D) \(^1\)H NMR spectroscopy (5, 55) were used to study the tertiary structure of the caviteins, and to assign resonances to the protons in the amide region. In addition, hydrogen/deuterium exchange experiments were carried out using 1D \(^1\)H NMR to determine the protection factor of the amide protons. The spectra were acquired by Okon from the laboratory of L. McIntosh (UBC Department of Biochemistry).

2.1.2.5.1 One-Dimensional (1D) \(^1\)H NMR Spectroscopy

One-dimensional \(^1\)H NMR is a well known technique used to investigate protein structures (53, 54), and here, it was used to probe the tertiary structure of the caviteins. De novo proteins with native-like character generally display high chemical shift dispersion and sharp signals (56). On the other hand, molten globule structures are not as specifically packed and are conformationally heterogeneous, and therefore give rise to more broad, less dispersed signals (57). The 1D \(^1\)H NMR spectra of the caviteins in the second generation linker series are shown in Figure 2.11 and 2.12. Figure 2.11 shows the expanded amide region, and Figure 2.12 shows the expanded aliphatic region of the spectra.
Focusing on the amide region (see Figure 2.11), it can be seen that the spectrum of the 2GS2 cavitein contains the sharpest and most well-separated signals, compared to the other caviteins, which indicate the presence of a well-organized hydrogen bond network of the amide protons.

![Expanded amide regions (9.5 to 5.5 ppm) of the second generation caviteins (1.5 mM) in pH 7.0 phosphate buffer and 10 % D2O at 25 °C on a 600 MHz spectrometer.](image)

**Figure 2.11.** Expanded amide regions (9.5 to 5.5 ppm) of the second generation caviteins (1.5 mM) in pH 7.0 phosphate buffer and 10 % D2O at 25 °C on a 600 MHz spectrometer.

In addition, the cavitand signal around 6.1 ppm is a doublet for the 2GS2 cavitein whereas it is a singlet for the rest of the caviteins. This signal corresponds to the H_out peak. For distinction between the H_out and H_in protons, see Scheme 2.5. The H_out signal should be a doublet.
because it is coupled to the $H_{in}$ proton. The 2GS2 cavitein is the only species that has this observable doublet out of the linker series, which implies that it has the most well-defined packing between the side chains. Furthermore, the $H_{out}$ signal of the 2GS2 cavitein is slightly shifted downfield. This proton may be more deshielded than the $H_{out}$ signals of the other caviteins because it was able to more efficiently form a hydrogen bond. The 3GS2 cavitein shows the next best spectrum in terms of sharpness and dispersion; however, it was found to change overtime (see N-H/D exchange section 2.1.2.5.3), and therefore it is not compared directly with the other caviteins. The 1GS2 and 4GS2 caviteins both show broad, less-dispersed spectra, which may be due to the lack of conformationally specific interactions within the caviteins. The 4GS2 has the least chemical shift dispersion and the most broad signals in the amide region, which suggests that the linker containing four Gly residues was too long for the peptides to specifically interact with one another. Next, looking at the signals in the aliphatic region (see Figure 2.12), proton signals in the $\alpha$-carbon region (around 4 ppm) and the rest of the aliphatic protons ($\beta,\gamma,\delta,\varepsilon$) can be seen.
Once again, the 2GS2 cavitein contains the sharpest, most-dispersed signals, followed by the 3GS2 cavitein, then the 1GS2 and 4GS2 caviteins. The side chain packing within the hydrophobic core can be seen by observing the signals arising from the δ methyl protons of the nonpolar Leu residues at around 0.8 ppm. This region shows especially broad signals for the 1GS2 and 4GS2 caviteins, demonstrating their relatively poor packing within the core of the protein bundles. For the 1GS2 cavitein, the broad spectrum may be due to the monomer/dimer equilibrium.

**Figure 2.12.** Expanded aliphatic regions (4.5 to 0.0 ppm) of the second generation caviteins (1.5 mM) in pH 7.0 phosphate buffer and 10 % D₂O at 25 °C on a 600 MHz spectrometer.
For the most part, the extent of chemical shift dispersion has been used to indicate the degree of native-like character; however, it should also be noted that the limited dispersion may be a result of a supercoiled structure (58). For example, the 4GS2 cavitein has the least-dispersed spectrum, which may be an indication of supercoiling. However, for the most part, the discussions concerning dispersion of the chemical shifts in the $^1$H NMR spectra are limited to characterizing the extent of native-like structure. Supercoiling is quantified in Chapter 3 where these caviteins were examined using molecular dynamics.

From the 1D $^1$H NMR spectral data, it can be concluded that the 2GS2 cavitein is the most well-defined, native like species out of all the caviteins in its series, and the 1GS2 and 4GS2 caviteins show broad 1D $^1$H NMR spectra, which are indicative of molten globule structures. Compared to the first generation series (data not shown), the 2GS2 cavitein displays the overall best $^1$H NMR spectrum in terms of dispersion and sharpness of the signals.
2.1.2.5.2 Two-Dimensional (2D) Homonuclear $^1$H NMR Spectroscopy

2D $^1$H NMR spectroscopy is a useful tool in investigating the tertiary structure of proteins (59), and to assign proton resonances (5). 2D contour plot spectra have two axes corresponding to chemical shift. For this study, 2D Double-Quantum-Filtered (DQF)-COrelation Spectroscopy (COSY) (60), Nuclear Overhauser Enhancement Spectroscopy (NOESY) (61), and HOMonuclear HArtmann-HAhn (HOHAHA), also known as TOTal Correlation Spectroscopy (TOCSY) (62, 63) were used to examine the caveins.

A COSY spectrum shows which pairs of nuclei are coupled to one another through bonds in a molecule. The signals on the diagonal correspond to the 1D $^1$H NMR spectrum. The cross peaks off the diagonal occur at positions where there is coupling between a proton of one axis and a proton of the other axis. In a single COSY spectrum, all the spin-spin coupling pathways that are three or less bonds apart can be identified, if the signals are sufficiently dispersed.

A TOCSY (or HOHAHA) spectrum is similar to the COSY spectrum in that it is based on spin-spin propagation via scalar coupling. But unlike COSY where only a single nucleus is irradiated at a time, in TOCSY, the resonance of an isolated spin multiplet is selectively inverted and then the inverted magnetization is propagated through the $^1$H coupling network. A TOCSY spectrum can give information about the all the protons directly or indirectly coupled to the inverted resonance, whereas as COSY spectrum only shows scalar coupled protons that are generally no further than three bonds away.

A NOESY spectrum relies on the Nuclear Overhauser Effect and shows the pairs of nuclei that are close together in space by irradiating one resonance while observing another. The NOESY spectrum is also symmetrical and has a diagonal set of peaks corresponding to the 1D $^1$H NMR spectrum. The off diagonal signals occur at positions where a proton of one axis is
close in space to a proton on the other axis that is no more than 5-6 Å away. The intensity of the signal is proportional to $1/r^6$ where $r$ is the distance between the interacting nuclei.

2D NMR spectroscopy was used to assign the chemical shifts of the amide protons for the 1GS2 and 2GS2 caviteins. The amide proton signals of the 3GS2 and 4GS2 caviteins could not be assigned due to their time-dependent behaviour (see N-H/D exchange section), and their broad 1D $^1$H NMR spectrum (see 1D NMR section), respectively.

The 2GS2 cavitein shows sharp, well-dispersed signals in the 2D $^1$H NMR spectra, which confirms the results from the 1D $^1$H NMR spectrum that it has a large degree of native-like character. The high dispersion also allowed for the assignment of the resonance signals in the amide region. To assign the chemical shifts of the 2GS2 cavitein $^1$H NMR spectrum, both 2D DQF-COSY and 2D NOESY were used. Figure 2.13 shows the 2D COSY spectrum and Figure 2.14 shows the 2D NOESY spectrum of the 2GS2 cavitein, both expanded in the amide region.
Figure 2.13. 2D COSY spectrum of 2 mM 2GS2 cavitein in 50 mM pH 4.62 acetate buffer and 10 % D$_2$O at 20 °C, expanded in the amide region.
Looking at the expanded amide region of the COSY spectrum (Figure 2.13), there is only one set of cross peaks. Examining the sequence, the only possible NH/NH through bond interaction that would have shown a strong COSY signal would be from the amide cap protons at the C-terminus of the peptides.

To determine the identity of the remaining peaks in the amide region, the NOESY spectrum of the expanded amide region was examined more closely. Figure 2.14 shows this expanded region with the cross peaks labelled.

The assignment of the cross peaks started with the amide cap protons that were identified using the COSY spectrum. These protons show a NOE interaction with the C-terminal residue, G17. The only other cross peak for G17 corresponds to G17/K16, which determines the resonance for K16. For the K16 residue, other possible cross peaks can be seen. This ambiguity was deciphered using NMRView, which can expand certain regions of the spectrum and differentiate between signals using the alignment of crosshairs. Employing this program, only one other cross peak from the K16 chemical shift was observed, which was used to identify the K15 resonance. The only other cross peak to K15 corresponds to K15/L14, which allowed for the determination of the L14 resonance. From the L14/L13 cross peak, the L13 resonance was found. This process continued until all the amide proton chemical shifts were determined.
Figure 2.14. 2D NOESY spectrum of 2 mM 2G5S2 carried in 50 mM Ph 4.62 acetate buffer and 10% D2O at 20°C, expanded in the amide region.
Alternatively, the assignment of the amide proton resonances was carried out from the other end, the N-terminus. In the amide region, one can usually assign the amide proton of the terminal residues because they usually have only one associated cross peak. In other words, the amide protons of the terminal residues have a cross peak relating to the previous residue, \(i-1\) (in the case of the C-terminal residue), or the next residue, \(i+1\) (in the case of the N-terminal residue), but not both. However, in the case of the peptides linked onto the caviteins, the C-terminal residue was capped with an amide group; therefore, the amide proton of the C-terminus residue should show cross peaks from both the amide protons of the previous residue, and the amide protons from the cap. The only other terminal residue remaining was from the N-terminus. There is only one chemical shift in the amide region that has only one cross peak, and it was identified to be G1. From the G1 resonance, the only other cross peak corresponds to G1/G2, which was used to find the G2 resonance. The other cross peak from G2 is for G2/A3, which was used to determine the A3 resonance. Again, if there were more than a couple cross peaks corresponding to a resonance peak, that region of the spectrum was expanded and signals were distinguished using NMRView. The process was continued until all the amide protons were assigned. The assignments from the N-terminus converged with those starting from the C-terminus, thereby validating the resonance determination of the amide protons.

Another way to verify the accuracy of the resonance assignments was to examine the region of the NOESY spectrum that relates the amide protons to the aliphatic protons. Figure 2.15 shows this expanded region.
Figure 2.15. 2D NOESY spectrum of 2 mM 2GS2 cavitein in 50 mM pH 4.62 acetate buffer and 10 % D$_2$O at 20 °C, expanded in the aliphatic/amide region.
By looking at Figure 2.15, one can see that the amide proton at 8.51 ppm has aliphatic proton cross peaks in common with the amide proton at 8.63 ppm, which shows that these two amide protons belong to residues that are next to each other in the sequence. The aliphatic protons corresponding to the amide proton at 8.51 ppm are characteristic of a Glu residue in an α-helix, and the aliphatic protons corresponding to the amide hydrogen at 8.63 ppm are characteristic of a Leu residue (5). The amide protons can interact through space with aliphatic protons from itself, $i$, and from the aliphatic hydrogen from its previous, $i-1$, and occasionally from $i-3$ and $i-4$ residues, but not from the next residue, $i+1$. Therefore, the Leu residue is before the Glu residue in the sequence. Scanning the S2 sequence, the only place where a Leu residue comes before a Glu residue is L10 to E11. The identification of these two resonances agrees with the rest of the assignments starting from either end of the sequence. This finding further confirms the accuracy of the resonance assignments. Finally, the $\alpha$-protons corresponding to the amide protons were assigned on the expanded alpha/amide region of the 2D COSY spectrum as shown on Figure 2.16.
Figure 2.16. 2D COSY spectrum of 2 mM 2GS2 cavitein in 50 mM pH 4.62 acetate buffer and 10 % D$_2$O at 20 °C, expanded in the alpha/amide region.
Table 2.6 shows the resonance assignments for the amide protons (NH) of the 2GS2 cavitein along with the corresponding α-protons (αCH). These data indicate that the four-helix bundle has a four-fold symmetry.

Table 2.6. The resonance assignments in ppm for the 2GS2 cavitein in 50 mM acetate buffer and 10 % D$_2$O at pH 4.62 at 20 °C.

<table>
<thead>
<tr>
<th>Label</th>
<th>NH (ppm)</th>
<th>αCH (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH$_A$</td>
<td>7.19</td>
<td>-------</td>
</tr>
<tr>
<td>NH$_B$</td>
<td>7.42</td>
<td>-------</td>
</tr>
<tr>
<td>G1</td>
<td>9.16</td>
<td>3.96</td>
</tr>
<tr>
<td>G2</td>
<td>7.87</td>
<td>4.15</td>
</tr>
<tr>
<td>A3</td>
<td>7.24</td>
<td>3.62</td>
</tr>
<tr>
<td>E4</td>
<td>8.00</td>
<td>3.95</td>
</tr>
<tr>
<td>E5</td>
<td>7.67</td>
<td>4.02</td>
</tr>
<tr>
<td>L6</td>
<td>7.58</td>
<td>4.10</td>
</tr>
<tr>
<td>L7</td>
<td>7.96</td>
<td>3.87</td>
</tr>
<tr>
<td>K8</td>
<td>7.88</td>
<td>4.05</td>
</tr>
<tr>
<td>K9</td>
<td>7.97</td>
<td>4.08</td>
</tr>
<tr>
<td>L10</td>
<td>8.63</td>
<td>4.01</td>
</tr>
<tr>
<td>E11</td>
<td>8.51</td>
<td>3.78</td>
</tr>
<tr>
<td>E12</td>
<td>7.64</td>
<td>3.99</td>
</tr>
<tr>
<td>L13</td>
<td>7.88</td>
<td>4.07</td>
</tr>
<tr>
<td>L14</td>
<td>8.20</td>
<td>4.08</td>
</tr>
<tr>
<td>K15</td>
<td>7.73</td>
<td>4.14</td>
</tr>
<tr>
<td>K16</td>
<td>7.86</td>
<td>4.24</td>
</tr>
<tr>
<td>G17</td>
<td>8.13</td>
<td>3.91</td>
</tr>
</tbody>
</table>

The resonance assignments of the amide protons have been labelled on the 1D $^1$H NMR spectrum of the 2GS2 cavitein shown in Figure 2.17.

125
Figure 2.17. 1D $^1$HNMR spectrum of the 2GS2 cavitein in 50 mM pH 4.62 acetate buffer and 10 % D$_2$O at 20 °C, expanded in the amide region. Residues are labelled based on information from the 2D NMR spectra.

From Figure 2.17, one can see that the amide proton of G1 was shifted the most downfield, indicating that a hydrogen bond to the cavitand bowl may have been present. The amide proton of A3 was found to be as shielded as the amide protons of the C-terminal cap, suggesting it lacked participation in the hydrogen bond network of the helices.

Compared to the 2GS2 cavitein, the 1GS2 cavitein showed less-resolved 1D and 2D $^1$H NMR spectra, which made the resonance assignments more challenging. Okon assisted with the assignments using the NMRView software. The resonance assignments for the 1GS2 cavitein were carried out similarly to the 2GS2 cavitein. Figure 2.18 shows the expanded amide region of the NOESY spectrum with labelled cross peaks, and Figure 2.19 shows the labelled TOCSY spectrum, expanded in the alpha/amide proton region.
Figure 2.18. 2D NOESY spectrum of 1.1 mM 1GS2 cavitein in 50 mM pH 4.62 acetate buffer and 10 % D$_2$O at 20 °C, expanded in the amide region.
Figure 2.19. 2D TOCSY spectrum of 1.1 mM 1GS2 cavitein in 50 mM pH 4.62 acetate buffer and 10 % D$_2$O at 20 °C, expanded in the alpha/amide region.
Table 2.7 shows the resonance assignments for the amide protons (NH) of the 1GS2 cavitein along with the corresponding α-protons (αCH).

Table 2.7. The resonance assignments in ppm for the 1GS2 cavitein in 50 mM pH 4.62 acetate buffer and 10 % D₂O at 20 °C.

<table>
<thead>
<tr>
<th>Label</th>
<th>NH (ppm)</th>
<th>αCH (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHₐ</td>
<td>7.38</td>
<td>-----</td>
</tr>
<tr>
<td>NHₐ</td>
<td>7.17</td>
<td>-----</td>
</tr>
<tr>
<td>Gl</td>
<td>8.02</td>
<td>3.86, 4.09</td>
</tr>
<tr>
<td>A2</td>
<td>8.36</td>
<td>4.16</td>
</tr>
<tr>
<td>E3</td>
<td>8.68</td>
<td>3.89</td>
</tr>
<tr>
<td>E4</td>
<td>8.42</td>
<td>4.04</td>
</tr>
<tr>
<td>L5</td>
<td>7.74</td>
<td>4.01</td>
</tr>
<tr>
<td>L6</td>
<td>8.12</td>
<td>3.91</td>
</tr>
<tr>
<td>K7</td>
<td>7.87</td>
<td>4.12</td>
</tr>
<tr>
<td>K8</td>
<td>7.81</td>
<td>4.14</td>
</tr>
<tr>
<td>L9</td>
<td>8.49</td>
<td>4.04</td>
</tr>
<tr>
<td>E10</td>
<td>8.37</td>
<td>3.83</td>
</tr>
<tr>
<td>E11</td>
<td>7.69</td>
<td>4.01</td>
</tr>
<tr>
<td>L12</td>
<td>7.86</td>
<td>4.12</td>
</tr>
<tr>
<td>L13</td>
<td>8.15</td>
<td>4.10</td>
</tr>
<tr>
<td>K14</td>
<td>7.75</td>
<td>4.17</td>
</tr>
<tr>
<td>K15</td>
<td>7.80</td>
<td>4.27</td>
</tr>
<tr>
<td>G16</td>
<td>8.07</td>
<td>3.94</td>
</tr>
</tbody>
</table>

The resonance assignments of the amide protons have been labelled on the 1D ¹H NMR spectrum of the 1GS2 cavitein shown in Figure 2.20. Compared to the 2GS2 cavitein, the amide proton of the first residue, Gl was located more upfield for the 1GS2 cavitein. In addition, the amide proton of the third residue, E3 was found to be the most deshielded for the 1GS2 cavitein, whereas the third residue amide proton of the 2GS2 cavitein, A3, was found to be the most shielded of the amide protons. These findings show that the two caviteins varying by only one single Gly residue in the linker region have different folded structures. No 2D NMR spectra were
collected for the first generation series of caviteins, and therefore they can not be directly compared here.

![1D $^1$H NMR spectrum of the 1GS2 cavitein in pH 4.62 acetate buffer and 10 % D$_2$O at 20 °C, expanded in the amide region. Residues are labelled based on information from the 2D NMR spectra.](image)

**Figure 2.20.** 1D $^1$H NMR spectrum of the 1GS2 cavitein in pH 4.62 acetate buffer and 10 % D$_2$O at 20 °C, expanded in the amide region. Residues are labelled based on information from the 2D NMR spectra.

### 2.1.2.5.3 Hydrogen/Deuterium Exchange

Hydrogen exchange experiments have been used to assess the macromolecular structure and dynamic behaviour of proteins (64, 65). It is a useful technique in that the exchange can be studied under many conditions, detected in a residue-specific manner, and is extremely sensitive to conformational changes. Exposed protein amide hydrogens readily exchange with the solvent; if the solvent is deuterated, the process is known as a hydrogen/deuterium (H/D) exchange.
Native proteins generally contain a collection of slow exchanging amides when compared to peptides (66), whereas molten globule have shown intermediate behaviour due to their fluctuating tertiary structures (67).

Exchange of amide protons that are involved in hydrogen bonds can exchange with solvent when they are exposed to the solvent in a closed to open mechanism as shown below:

\[ \text{NH (closed) } \xrightleftharpoons{K_{\text{op}}} \text{NH (open) } \xrightarrow{k_{\text{int}}} \text{exchange} \]

where \( K_{\text{op}} \) is the equilibrium opening constant, and \( k_{\text{int}} \) is the intrinsic exchange rate in the open form. The exchange rate of any hydrogen, \( k_{\text{ex}} \), can be determined from the following equation:

\[ k_{\text{ex}} = K_{\text{op}}k_{\text{int}} \]

If the amide proton is in an unstructured region of the protein, exchange with the solvent occurs relatively quickly. If the amide proton resides in a structured region of the protein, the exchange is slowed or protected. The extent of protection from the solvent is expressed as a protection factor, \( P \), which is dependent on temperature and pH. A large \( P \) value signifies an amide proton that takes longer to exchange with the solvent as it is more ‘protected’ than an amide that has a lower protection factor. Typically, the backbone amides of native proteins have protection factors that range from \( 10^4 \) to \( 10^8 \) (67), whereas molten globule structures have protection factors that range from 10 to \( 10^3 \) (68).

The exchange rates of the caviteins were studied at 20 °C, pD 5.02 acetate buffer. Figures 2.21 to 2.24 show the N-H/D exchange \(^1\)H NMR spectra of the second generation caviteins.
Figure 2.21. Stack plot of 1D $^1$H NMR N-H/D exchange spectra for 2.2 mM 1GS2 cavitein in 50 mM deuterated acetate buffer at pD 5.02 on a 500 MHz NMR Varian Unity spectrometer at 20 °C. The reference spectrum was taken in 50 mM pH 4.62 acetate buffer and 10 % D$_2$O.

From Figure 2.21, it can be seen that the amide proton signals for the 1GS2 cavitein disappeared within a day. The cavitand signals at 6.1 ppm and 7.1 ppm, which correspond to H$_{out}$ and H$_{para}$, respectively, did not change over time (see Scheme 2.5 for proton labels). Therefore, these signals were used to normalize the peak heights of the amide protons. Compared to the 1GS2 cavitein, the 2GS2 cavitein contained amide proton signals that remained for much longer as shown in the next set of spectra (Figure 2.22).
Figure 2.22. Stack plot of 1D $^1$H NMR N-H/D exchange spectra for 1.9 mM 2GS2 cavitein in 50 mM deuterated acetate buffer at pD 5.02 on a 500 MHz NMR Varian Unity spectrometer at 20 °C. The reference spectrum was taken in 50 mM pH 4.62 acetate buffer and 10 % D$_2$O.
One of the first set of signals to disappear was the amide caps at the C-termini, which were identified by 2D NMR spectroscopy. These signals exchanged before the first scan after D$_2$O addition was taken. The 2GS2 cavitein contained amide proton signals even after a few months with the last signal remaining up to three months. The last proton signal to disappear was found to be the central hydrophobic residue, Leu in the peptide sequence of the caviteins, suggesting that this residue plays an important role in the stabilization of the overall structure. In contrast to the 2GS2 cavitein, for the 1GS2 and 4GS2 caviteins, all of the amide proton signals disappeared within a day or two. Figure 2.23 shows the N-H/D exchange spectra of the 4GS2 cavitein.
Figure 2.23. Stack plot of 1D $^1$H NMR N-H/D exchange spectra for 1.9 mM 4GS2 cavitein in 50 mM deuterated acetate buffer at pD 5.02 on a 500 MHz NMR Varian Unity spectrometer at 20 °C. The reference spectrum was taken in 50 mM pH 4.62 acetate buffer and 10 % D$_2$O.

The N-H/D exchange data are consistent with the stabilities of the caviteins determined by the GuHCl denaturation experiments. The more stable caviteins contained amide protons that exchanged more slowly with the solvent. Since the backbone hydrogen bonds had to be broken for exchange to occur with deuterium, these experimental results show that the hydrogen bonds within the helices are important in stabilizing the cavitein systems.
Compared with the rest of the caviteins, the 3GS2 cavitein was found to be an anomaly. The $H_{\text{out}}$ and $H_{\text{para}}$ cavitand signals that remained constant in the spectra of all the other caviteins did not stay constant for the 3GS2 cavitein, as they started to decrease over time, accompanied by the increase of other unidentified signals. The exchange rates could not be calculated for this cavitein due to the absence of a time constant signal that was used to normalize the signals of the other caviteins. Figure 2.24 shows the irregular behaviour of the 3GS2 cavitein exchange data.

The same phenomenon occurred when the experiment was carried out in the non-deuterated buffer. When the solution in the NMR tube was re-lyophilized, and a $^1$H NMR spectrum was acquired, it looked similar to the spectra of the matured sample (8 d). When the contents were repurified using a reverse-phased HPLC, lyophilized, and prepared for NMR spectroscopy, the acquired spectra looked similar to that of the initial spectra taken when the buffer was first added to the sample.
Figure 2.24. Stack plot of 1D \(^1\)H NMR N-H/D exchange spectra for 1.9 mM 3GS2 cavitein in 50 mM deuterated acetate buffer at pD 5.02 on a 500 MHz NMR Varian Unity spectrometer at 20 °C. The reference spectrum was taken in 50 mM pH 4.62 acetate buffer and 10 % D\(_2\)O.

The N-H/D experiment was carried out at a lower concentration of the 3GS2 cavitein in one experiment, and in a neutral buffer for another experiment to see if aggregation was the cause of this abnormal behaviour, but no definite conclusion could be made. Furthermore, electrospray ionization (ESI) mass spectrometer (PE-Sciex API 300 triple quadrupole, acquired
by Shouming He) was used to determine the molecular weights of both the young and mature samples. Both samples gave a molecular weight of the monomeric species. This abnormal behaviour of the 3GS2 cavitein was taken into consideration when analyzed and compared to the rest of the caviteins.

The amide proton chosen for analysis was the last signal to disappear for each cavitein. From the 2D NMR data of the 1GS2 and 2GS2 caviteins, it was demonstrated that the last residue to disappear from the spectra was the central Leu amino acid. The data are consistent with other globular proteins that also exhibit the most highly protected amide protons in the centre of the hydrophobic core (67). Using the equations in the experimental section 2.3.5.3, the exchange rates and the protection factors were calculated for each of the second generation caviteins, and are shown in Table 2.8.
Table 2.8. N-H/D exchange data on the caviteins in the second generation series in 50 mM deuterated acetate buffer, pD 5.02 at 20 °C. The calculated data are the averages of three estimates at three different times (listed errors represent the standard deviation of the estimates).

<table>
<thead>
<tr>
<th>Caviteins</th>
<th>Amide Proton Chemical Shift (ppm)</th>
<th>First-Order Rate Constant (h⁻¹)</th>
<th>Half-Life (h)</th>
<th>Protection Factor*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1GS2</td>
<td>8.48&lt;sup&gt;bc&lt;/sup&gt;</td>
<td>1.4 ± 0.4</td>
<td>0.50 ± 0.08</td>
<td>(1.6 ± 0.2)×10²</td>
</tr>
<tr>
<td>2GS2</td>
<td>8.62&lt;sup&gt;b&lt;/sup&gt;</td>
<td>(4 ± 1)×10⁻³</td>
<td>(1.9 ± 0.7)×10²</td>
<td>(6 ± 1)×10⁴</td>
</tr>
<tr>
<td>3GS2&lt;sup&gt;d&lt;/sup&gt;</td>
<td>8.65</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>4GS2</td>
<td>8.49&lt;sup&gt;b&lt;/sup&gt;</td>
<td>3.5 ± 0.8</td>
<td>0.20 ± 0.04</td>
<td>(6.0 ± 0.8)×10¹</td>
</tr>
</tbody>
</table>

- a these values are based on the half-life of an unprotected proton at pD 5.02 at 20°C to be 3.18 E-3 h.
- b these values correspond to the chemical shift of the last amide proton to disappear during exchange.
- c this value is a combination of two signals in the NMR spectra that could not be resolved.
- d the non-exchange <sup>1</sup>H NMR spectra of the 3GS2 cavitein changed over time and therefore the rate constant, half-life and protection factor could not be calculated for this species.

Overall, the 2GS2 cavitein has the highest protection factor, and has a value characteristic of a native-like protein (67). The 1GS2 and 4GS2 caviteins possess protection factors that are characteristic of molten globule proteins (68). Compared to the monomeric caviteins in the first generation series, the 2GS2 cavitein has a protection factor 10-fold higher than the monomeric 2GS0 and 3GS0 caviteins under the same exchange conditions (69).
2.1.2.6 ANS Binding Studies

1-anilinonaphthalene-8-sulfonate (ANS) is a hydrophobic dye that has been used as a fluorescent probe to detect molten globule proteins (54, 70, 71). ANS is brightly fluorescent in solvents such as methanol and ethanol, but is only weakly fluorescent in water. In the presence of a protein with an exposed hydrophobic pocket, binding of ANS can take place, and fluorescence can be detected. ANS preferentially binds to the molten globule state because it is a dynamic structure that lacks the tight side-chain packing within the core of the structure that the native state possesses (72). The completely unfolded protein wherein all the hydrophobic sites are exposed, however also do not bind ANS. This has been reasoned on the basis that hydrophobic pockets of definite geometry do not exist in the fully unfolded state. In the case of the caviteins, a nonpolar binding site potentially exists in the hydrophobic core of the four-helix bundles. Figure 2.25 shows the relative fluorescence intensities of ANS for the caviteins in the second generation series.
Figure 2.25. Fluorescence emission spectra of 2 μM ANS in the presence of 100 % methanol, 95 % ethanol, and 50 μM of each cavitein: 1GS2, 2GS2, 3GS2 and 4GS2 in 50 mM phosphate buffer, pH 7.0 at 25 °C. Error bars omitted for clarity, but explained in the experimental section.

Like the Gly and the methylene variants of the first generation caviteins (see Table 2.1), the second generation of caviteins using the S2 sequence only weakly bound to ANS, suggesting that these proteins are native-like or only slightly molten globule-like. However, the 1GS2 cavitein bound the most ANS, followed by the 4GS2 cavitein, which agree with the other experimental results in that the 1GS2 and 4GS2 caviteins possess the most molten globule-like characteristics. The 1GS2 cavitein was also accompanied by a slight shift to a lower wavelength, known as a blue shift. This shift is usually an indication of decreased accessibility of the fluorescent entity to the bulk solvent (73). One Gly was insufficient to provide enough flexibility for optimal packing of the helices in the four-helix bundle. On the contrary, the 2GS2 cavitein
bound the least amount of ANS, which is consistent with the rest of the experimental findings that it is the most-native like out of all the caviteins. The reference peptide did not bind any ANS regardless of concentration. The fluorescence intensity remained on the baseline throughout the scan (not shown).

Denaturants such as urea and guanidine hydrochloride have been used in ANS binding studies to detect any structural changes during unfolding (74, 75). The caviteins are sensitive to guanidine hydrochloride (GuHCl), and therefore, ANS experiments were carried out in the presence of this denaturant to see if any unfolding intermediates could be detected. As the concentration of this denaturant increased from 0M to 8M in 1M increments, there were no noticeable changes in the ANS fluorescence spectra, which indicate that there were no observable intermediates during the unfolding of the caviteins that can be detected by ANS binding experiments. In other words, the use of GuHCl did not generate a distinct intermediate structure.
2.2 Chapter Summary and Conclusion

The *de novo* template assembly approach has proven to be a useful tool in understanding protein structure and folding because it simplifies the structure by reducing loops and turns, it provides stability, and it allows for control of many factors including orientation and number of peptides in a bundle. In previous work by other groups, the linker length between the template and the peptides was found to have noticeable effects on a protein's overall structure; however, these results were not further examined. In our group, the effect of the linker length between the cavitand template and the peptides was investigated in more detail using a cavitand as the template. Mezo carried out the first set of experiments on the first generation caviteins and found that the linker length had substantial effects on the structural properties of the proteins. The peptide sequence studied in this first generation series linked onto the template from the hydrophobic/hydrophilic interface. This peptide sequence required at least two to three Gly residues to give a monomeric, well-defined cavitein. The two Gly variant, 2GS0, displayed the most dispersed $^1$H NMR spectrum, but with broad signals. Although the 3GS0 cavitein displayed a less-dispersed $^1$H NMR spectrum compared to the 2GS0 cavitein, the three Gly variant was found to be slightly more stable and showed sharper signals in its $^1$H NMR spectrum. Mezo believed that the two Gly variant gave the most native-like cavitein from its chemical shift dispersion. However, others in our group believe that out of the first generation caviteins, the three Gly variant possessed more native-like properties due to its slightly higher stability and sharper resonances. These different points of view show how the same data can be interpreted differently. A cautious interpretation would be that the linker length requirement for the first generation of caviteins was two to three Gly linkers.
However, there were questions that still remained: 1) is it possible to improve the native-like character of the caviteins, 2) why are a certain number of Gly residues required for a certain sequence to adopt a native-like conformation, and 3) what are the approximate interhelical distances between adjacent helices in the cavitein bundles? The synthesis and characterization of the second generation of caviteins presented in this chapter attempted to answer these questions. The effect of the linker lengths were further studied by designing a sequence that would link onto the template from the hydrophobic face. The goals were to improve the native-like characteristics of the caviteins by linking the peptides closer to the cavitand template, and to learn about the effective diameter and position of the helices relative to this template. This second generation series of caviteins were investigated using CD spectroscopy, chemical denaturation studies, analytical ultracentrifugation, NMR spectroscopy, and ANS binding studies. From the CD spectra, the caviteins in the second generation series were found to be helical like those in the first generation series. From the near UV data, the 2GS2 cavitein displayed the most enhanced signal of all the caviteins, which implied that this species had the most non-averaged structural elements (i.e., native-like characteristics) near the aromatic chromophore. However, the effect of the cavitand chromophore may have been minimal for the caviteins with the longer linkers as the helical bundles were removed further away from the template. The GuHCl-induced denaturation studies gave the free energies of folding. The 2GS2 cavitein was found to have the highest stability of all the monomeric caviteins in both the first and second generation series. Similarly, the $m$ value, which is indicative of high cooperativity and native-like properties, was also the greatest for the 2GS2 cavitein, with the exception of the first generation dimeric 1GS0 cavitein. The analytical ultracentrifuge allowed for sedimentation equilibria studies that were used to determine the molecular weight of the caviteins in solution. From the $^1$H NMR spectral data, the 2GS2 cavitein showed the best spectrum in terms of
dispersion and sharpness, indicating that it contained the most structurally specific components, characteristic of native-like proteins. Compared to the $^1$H NMR spectra of the first generation series, the 2GS2 cavitein demonstrated overall better chemical shift dispersion than the 2GS0 cavitein and sharper signals than the 3GS0 cavitein. From the N-H/D exchange data of the monomeric caviteins in both generations, the 2GS2 cavitein possessed the amide proton with the highest protection factor. Finally, from the ANS binding studies, all the caviteins showed only weak fluorescence, with the 2GS2 cavitein displaying the lowest intensity. From all the experimental findings, the 2GS2 cavitein behaved the most native-like out of both the first and second generation series. Therefore, this project was successful in obtaining a cavitein with improved native-like characteristics.

To explain why a certain number of flexible Gly residues was required for a specific sequence to adopt a well-defined structure, and to determine the approximate interhelical distances between the adjacent helices, a diagram of the distance discrepancy between the cavitand template and the effective helix diameter is illustrated (see Figure 2.26). The interhelical distance between adjacent helices can range from 7-14 Å, which is dependent on the particular residues and interactions involved. The variation in the interhelical distances is dependent on the effective diameter of the helix. A well-defined protein should result when the distance discrepancy between the cavitand and the helices is alleviated. The first generation linker series required two to three Gly residues in the linker region to alleviate this strain, whereas the second generation series only required two Gly residues. Figure 2.26(a) shows a situation where the effective helix diameter is comparable to the sides of the cavitand (7 Å), and Figure 2.26(b) shows a situation when the effective helix diameter is much larger than the sides of the cavitand template.
Figure 2.26. Distance discrepancy between the cavitand template and helices of the four-helix bundle protein when the effective helix diameter is on the lower end (a) and higher end (b). Linkage is represented by a dot for the first generation caviteins, and by a star for the second generation caviteins.
If it was found that the same linker length was required for both the first and second generation caviteins, Figure 2.26(a) may be a better representation of the effective helix diameter. On the other hand, if a much longer linker was required for the first generation caviteins compared to the second generation caviteins, Figure 2.26(b), which represents the effective helix diameter on the higher end, may be a better representation. Since the second generation caviteins required only a slightly shorter linker, a rational estimation of the interhelical distance between adjacent helices would be between 8 and 10 Å, which is reasonable for helical bundles containing nonpolar Leu residues.

In Chapter 3, computer modelling is used to further probe the second generation caviteins studied here and to see if better insight into the behaviour of the caviteins can be gained. Although there is still much to be learned about what factors control protein structure and folding, each finding takes us one step closer to constructing an optimal de novo protein. Once a set of fundamental rules for designing native-like proteins can be established, de novo proteins with more complex structure and function can be engineered and probed. Potential applications and future goals are discussed in Chapter 5.
2.3 Experimental

2.3.1 Arylthiol Cavitand Synthesis

2.3.1.1 General

All chemicals were reagent grade from Aldrich. \( N, N \)-dimethylacetamide (DMA) was dried over 5 Å molecular sieves. THF was distilled under nitrogen gas from sodium benzophenone ketyl. \( N \)-bromosuccinimide (NBS) was freshly recrystallized before use. Matrix-Assisted Laser Desorption Ionization (MALDI) mass spectrometry were recorded on a Bruker Biflex IV using 2,5-dihydroxybenzoic acid (DHB) as the matrix. All \(^1\)H NMR spectra were recorded on a Bruker AVANCE AV400 MHz spectrometer at ambient temperature using residual proton signals from deuterated solvents as a reference: CDCl\(_3\) (7.24 ppm), MeOD (3.30 ppm), DMSO-d\(_6\) (2.49 ppm). Silica gel on glass analytical plates (0.2mm) from Aldrich were used for TLC with UV detection. After each step, the product was dried overnight at 0.1 Torr. Bromocavitand 4 was synthesized from resorcinol 1 using the procedures in reference (13). The step from bromocavitand 4 to the arylthiol cavitand 5 was slightly modified from the original procedure as described in the next section.
2.3.1.2 Synthesis of the Arylthiol Cavitand

Bromocavitand 4 (1.0 g, 1.1 mmol) was dissolved in 500 mL dried THF and cooled to -78 °C. n-BuLi (9.0 mL of a 1.6 M solution in hexanes, 14.4 mmol) was then added and the reaction mixture was stirred for 5 minutes. A solution of sulfur (0.5 g, 15.6 mmol) in 30 mL THF was cannulated slowly into the reaction flask over a period of 10 minutes. Then the -78 °C cooling bath was replaced with an ice-bath, and the reaction was stirred overnight; whereas in the literature procedure, the reaction was completed over a course of a few hours. In this original procedure, the -78 °C cooling bath was removed, the reaction mixture was allowed to warm to rt over 1.5 h, and then 2M HCl was added. In the modified procedure, the reaction was left to stir overnight. The next day, the melted ice-bath was removed, and 2 M degassed HCl was added until the solution became acidic. This solution was then evaporated in vacuo. 50 mL of water was added and extracted with 4 x 150 mL of EtOAc. The organic phase was combined, washed with 2 x 100 mL brine, dried over MgSO₄ and left for 15 minutes. The solution was gravity filtered, and the filtrate was concentrated in vacuo. The residue was dissolved in a minimal amount of CHCl₃, transferred to a small vial, and the solvent was left to evaporate under nitrogen to afford a crude light yellow solid (0.45 g, ~ 95 %), which consisted of the desired arylthiol cavitand, and a small amount of tris by-product. In the literature procedure, the residue was dissolved in chloroform and precipitated with hexanes. The precipitate was recrystallized from chloroform/hexanes. Then, the crude product was acetylated and column chromatography was used to separate the tris from the tetrathiol product. The recrystallization and separation steps that were mentioned in literature were not carried out here, as it was found to be much easier to separate the tris from the tetra product after the peptides were linked on to the cavitand. The tris-arylthiol signals on the ¹H NMR spectrum were not listed below since there was only ~10 %.
\[^{1}\text{H} \text{NMR} \ (400 \text{ MHz, } \text{CDCl}_3) \delta \ 6.96 \ (s, 4\text{H, } \text{H}_\text{para}), \ 5.95 \ (d, \ J = 7.0 \text{ Hz, } 4\text{H, } \text{H}_\text{out}), \ 4.94 \ (q, \ J = 7.4 \text{ Hz, } 4\text{H, } \text{H}_\text{methine}), \ 4.36 \ (d, \ J = 7.0 \text{ Hz, } 4\text{H, } \text{H}_\text{in}), \ 3.76 \ (s, 4\text{H, SH}), \ 1.71 \ (d, \ J = 7.4 \text{ Hz, } 12\text{H, CH}_3) \text{ ppm.}\]

\[
\text{MS (MALDI, DHB) } m/z: \ 721 \ (M+H)^+ \\
\]

2.3.2 Peptide and Cavitein Synthesis

2.3.2.1 General

All chemicals used for peptide and cavitein synthesis were reagent grade. The peptides and caviteins were purified by reverse-phase HPLC using the Phenomenex Selectosil C\textsubscript{18} column (250 mm x 22.5 mm, 10 \textmu m particle size, 300 Å pore size). This column was used with a Perkin-Elmer Biocompatible Pump 250 equipped with a PE LC90 BIO Spectrophotometric UV detector and a KIPP and ZONEN chart recorder. UV detection was set to measure at 229 nm for the amide chromophore. The purity of the caviteins was checked by analytical reversed-phase HPLC using two different columns on two different instruments. The first was the Phenomenex Synergi C\textsubscript{18} column (250 mm x 4.60 mm, 4 \textmu m, 100 Å) on a Varian 9010 pump equipped with a Varian 9050 UV detector and a Varian 4290 Integrator. The second was the Waters Delta Pak C\textsubscript{18} column (300 mm x 3.9 mm, 15 \textmu m, 300 Å) on a Waters 600 Controller pump equipped with a Waters 2996 Photodiode Array detector. The gradient was composed of helium-sparged HPLC...
grade acetonitrile with 0.05 % TFA and helium-sparged, filtered, deionized water with 0.1 % TFA. Before injection, all samples were filtered with 0.45 μM Nylon™ syringe filters from Phenomenex. Preparative samples were run at 10 mL/min and analytical samples were run at 1 mL/min. The purified samples were evaporated in vacuo and lyophilized.

Sample concentrations were determined by using the Bradford assay (76) measured on a CARY UV-Visible spectrophotometer. The molecular weights of the peptides and caviteins were determined using a Bruker Biflex IV MALDI mass spectrometer, using saturated cinammic acid as the matrix and samples concentrations of approximately 6×10⁻⁵ M.

2.3.2.2 Peptide Synthesis

All peptides were synthesized on an Applied Biosystems (ABI) 431A automated peptide synthesizer connected to an Apple MacIntosh IIIsi computer. The peptides were synthesized on a 0.25 mmol scale employing FastMoc™ protocols that consisted of using Fmoc protected amino acids. All the Fmoc protected amino acids, solvents, and coupling reagents used for synthesizing the peptides were purchased from Advanced Chemtech, except for the dichloromethane (DCM), dimethylformamide (DMF) and piperidine, which were purchased from Aldrich. The amino acids used are as follows with protected side chains on a couple of the residues: Fmoc-Glu(t-Bu)-COOH, Fmoc-Lys(CO₂-t-Bu)-COOH, Fmoc-Leu-COOH, Fmoc-Ala-COOH, Fmoc-Gly-COOH. The coupling reagents consisted of HBTU (2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate) and HOBT (1-hydroxybenzotriazole) in HPLC grade DMF (77, 78). Rink resin was used to obtain peptides with C-terminal amides (79). The synthesis was carried out in
NMP (N-methylpyrrolidone) solvent, and a typical cycle consisted of several steps: 1) 13 minutes Fmoc deprotection using piperidine, 2) 6 minutes wash with NMP, 3) 30 minutes coupling step with the HBTU/HOBt mixture and 1.0 mmol of the subsequent Fmoc amino acid, and 4) 6 minutes NMP wash for a total cycle time of 55 minutes. The ABI 431A manual includes more details. When the synthesis was complete the peptide was again washed with NMP, then with DCM, and dried. In the case of the reference peptide, the N-terminus was acetylated; and in case of the 1GS2, 2GS2, 3GS2 and 4GS2 peptides, the N-terminus was reacted with chloroacetyl chloride. Each peptide was cleaved from the resin and protecting groups were removed by treatment with 95 % TFA in H₂O. The resin was removed by filtration through a medium frit while washing with DCM. The TFA/DCM mixture was evaporated to ~2 mL in vacuo, and the crude peptide was precipitated with ice-cold ether, and filtered on a fine frit. The peptides were purified by reversed-phase HPLC and were found to be pure by observing the analytical reversed-phase HPLC chromatograms. Figure 2.27 shows the schematic representation of the synthesis.
Figure 2.27. Schematic representation of the FastMoc™ protocol on the ABI 431A peptide synthesizer.
**Peptide 0GS1:** \(\text{ClCH}_2\text{CO-NH-[LEELLKKLEELLKKG]-CONH}_2\).

The 0GS1 peptide on resin (600 mg resin, ~300 mg peptide, 0.160 mmol) was reacted with chloroacetyl chloride (76 µL, 0.96 mmol, 6 equiv.) and DIPEA (167 µL, 0.96 mmol, 6 equiv.) in 5 mL DMF for 1 hour at room temperature under nitrogen. The reaction mixture was filtered using a medium frit and washed with DCM. Subsequent cleavage from the resin and protecting groups by treatment with 95 % TFA in water was carried out, and purification by reversed-phase HPLC afforded peptide 0GS1 as a white solid (100 mg, 22 %)

**MS:** (MALDI, cinnamic acid): \(m/z\) 1860 (M+H)

---

**Peptide 2GS1:** \(\text{ClCH}_2\text{CO-NH-[LEELLKKLEELLKKG]-CONH}_2\)

The 2GS1 peptide on resin (600 mg resin, ~300 mg peptide, 0.152 mmol) was reacted with chloroacetyl chloride (72 µL, 0.91 mmol, 6 equiv.) and DIPEA (159 µL, 0.91 mmol, 6 equiv.) in 5 mL DMF for 1 hour at room temperature under nitrogen. The reaction mixture was filtered using a medium frit and washed with DCM. Subsequent cleavage from the resin and protecting groups, and purification by reversed-phase HPLC afforded peptide 2GS1 as a white solid (120 mg, 24 %)

**MS:** (MALDI, cinnamic acid): \(m/z\) 1974 (M+H)
**Reference Peptide 0GS2: Ac-NH-[AEELLKKLEELLKKG]-CONH₂.**

The last cycle in the synthesis of the reference peptide 0GS2 involved acetylating the N-terminus of the peptide (~100 mg resin, ~50 mg peptide, ~0.03 mmol) using 2 mL of 10% acetic anhydride in 1 mL NMP and stirred at room temperature for 1 hour. The peptide was then cleaved from the resin and the protecting groups by treatment with 95% TFA in water for 2 h. After purification and lyophilization, the afforded peptide was a white solid (15 mg, ~20%).

**MS: (MALDI, cinnamic acid): m/z 1783 (M+H)**

**Peptide 1GS2: ClCH₂CO-NH-[GAEELLKKLEELLKKG]-NH₂.**

The 1GS2 peptide on resin (700 mg resin, ~350 mg peptide, 0.187 mmol) was reacted with chloroacetyl chloride (90 μL, 1.12 mmol, 6 equiv.) and DIPEA (195 μL, 1.12 mmol, 6 equiv.) in 5 mL DMF for 1 hour at room temperature under nitrogen. The reaction mixture was filtered using a medium frit and washed with DCM. Subsequent cleavage from the resin and protecting groups, and purification by reversed-phase HPLC afforded peptide 1GS2 as a white solid (130 mg, 28%).

**MS: (MALDI, cinnamic acid): m/z 1875 (M+H)**
**Peptide 2GS2**: ClCH₂CO-NH-[GGAEELLKKLEELLKKG]-NH₂.

The 2GS2 peptide on resin (700 mg resin, ~350 mg peptide, 0.181 mmol) was reacted with chloroacetyl chloride (74 μL, 0.93 mmol, 6 equiv.) and DIPEA (162 μL, 0.93 mmol, 6 equiv.) in 5 mL DMF for 1 hour at room temperature under nitrogen. The reaction mixture was filtered using a medium frit and washed with DCM. Subsequent cleavage from the resin and purification by reversed-phase HPLC afforded peptide 2GS2 as a white solid (155 mg, 32 %)

**MS**: (MALDI, cinnamic acid): m/z 1932 (M+H)⁺

**Peptide 3GS2**: ClCH₂CO-NH-[GGGAELLKKLEELLKKG]-NH₂.

The 3GS2 peptide on resin (600 mg resin, ~300 mg peptide, 0.151 mmol) was reacted with chloroacetyl chloride (72 μL, 0.91 mmol, 6 equiv.) and DIPEA (159 μL, 0.91 mmol, 6 equiv.) in 5 mL DMF for 1 hour at room temperature under nitrogen. The reaction mixture was filtered using a medium frit and washed with DCM. Subsequent cleavage from the resin and purification by reversed-phase HPLC afforded peptide 2GS2 as a white solid (110 mg, 22 %)

**MS**: (MALDI, cinnamic acid): m/z 1989 (M+H)⁺
**Peptide 4GS2**: ClCH₂CO-NH-[GGGAEELLKKLEELLKKG]-NH₂.

The 4GS2 peptide on resin (600 mg resin, ~300 mg peptide, 0.147 mmol) was reacted with chloroacetyl chloride (70 μL, 0.88 mmol, 6 equiv.) and DIPEA (154 μL, 0.88 mmol, 6 equiv.) in 5 mL DMF for 1 hour at room temperature under nitrogen. The reaction mixture was filtered using a medium frit and washed with DCM. Subsequent cleavage from the resin and purification by reversed-phase HPLC afforded peptide 2GS2 as a white solid (180 mg, 24 %)

**MS**: (MALDI, cinnamic acid): \( m/z \) 2946 (M+H)⁺

---

### 2.3.2.3 Thiocresol-Based Peptide 0GS1

DIPEA (5.6 μL, 32 μmol, 4 equiv.) was added to a solution of thiocresol (1 mg, 8 μmol, 1 equiv.) and 0GS1 peptide (30 mg, 16 μmol, 2 equiv.) in degassed DMF, and the reaction mixture was stirred at room temperature under nitrogen for 5 hours. The crude reaction mixture was monitored by analytical reversed-phase HPLC to observe a product peak after 1 hour. This mixture was evaporated *in vacuo* and purified by reversed-phase HPLC to afford the thiocresol-based peptide as a white solid after lyophilization (5 mg, 32 %).

**MS**: (MALDI, cinnamic acid): \( m/z \) 1948 (M+H)⁺
2.3.2.4 Cavitein Synthesis

The caviteins in the second generation linker series were purified by reversed-phase HPLC. Each cavitein was assessed for purity by observation of an analytical HPLC chromatogram (see Figure 2.28). Their identities were confirmed by MALDI mass spectrometry.

2GS1 Cavitein

DIPEA (2.4 μL, 14 μmol, 10 equiv.) was added to a solution of arylthiol cavitand (1.0 mg, 1.4 μmol, 1 equiv.) and 2GS1 peptide (22 mg, 11 μmol, 8 equiv.) in degassed DMF, and the reaction mixture was stirred at room temperature under nitrogen for 5 hours. The crude reaction mixture was evaporated in vacuo and purified by reversed-phase HPLC to afford the cavitein as a white solid after lyophilization (5 mg, 43%).

MS: (MALDI, cinnamic acid): m/z 8467.0 ± 0.8 (M+H)^+ [calcd 8466.6 Da].

1GS2 Cavitein

DIPEA (2.4 μL, 14 μmol, 10 equiv.) was added to a solution of arylthiol cavitand (1.0 mg, 1.4 μmol, 1 equiv.) and 2GS1 peptide (20 mg, 11 μmol, 8 equiv.) in degassed DMF, and the reaction mixture was stirred at room temperature under nitrogen for 5 hours. The crude reaction mixture was evaporated in vacuo and purified by reversed-phase HPLC to afford the cavitein as a white solid after lyophilization (6 mg, 54%).
2GS2 Cavitein

DIPEA (2.4 µL, 14 µmol, 10 equiv.) was added to a solution of arylthiol cavitand (1.0 mg, 1.4 µmol, 1 equiv.) and 2GS1 peptide (21 mg, 11 µmol, 8 equiv.) in degassed DMF, and the reaction mixture was stirred at room temperature under nitrogen for 5 hours. The crude reaction mixture was evaporated in vacuo and purified by reversed-phase HPLC twice to afford the cavitein as a white solid after lyophilization (5.6 mg, 49%).

MS: (MALDI, cinnamic acid): m/z 8298.5 ± 0.7 (M+H)⁺ [calcd 8298.2 Da].

3GS2 Cavitein

DIPEA (2.4 µL, 14 µmol, 10 equiv.) was added to a solution of arylthiol cavitand (1.0 mg, 1.4 µmol, 1 equiv.) and 2GS1 peptide (22 mg, 11 µmol, 8 equiv.) in degassed DMF, and the reaction mixture was stirred at room temperature under nitrogen for 5 hours. The crude reaction mixture was evaporated in vacuo and purified by reversed-phase HPLC twice to afford the cavitein as a white solid after lyophilization (4 mg, 34 %). For both the 2GS2 and 3GS2 caviteins, the protein had to be purified twice. After the first purification, a pre-peak was still present; but after the second purification, this pre-peak was removed (see Figure 2.28.)

MS: (MALDI, cinnamic acid): m/z 8527.2 ± 0.8 (M+H)⁺ [calcd 8526.6 Da].
Figure 2.28. The purification of the 3GS2 cavitein monitored by analytical reversed-phase HPLC using a gradient of 30 to 60 % acetonitrile (with 0.05 % TFA) in water (with 0.1 % TFA) over 50 minutes: (a) after the first purification (pre-peak still remains) and (b) after the second purification (pre-peak is removed).
4GS2 Cavitein

DIPEA (2.4 μL, 14 μmol, 10 equiv.) was added to a solution of arylthiol cavitand (1.0 mg, 1.4 μmol, 1 equiv.) and 2GS1 peptide (23 mg, 11 μmol, 8 equiv.) in degassed DMF, and the reaction mixture was stirred at room temperature under nitrogen for 5 hours. The crude reaction mixture was evaporated in vacuo and purified by reversed-phase HPLC to afford the cavitein as a white solid after lyophilization (7 mg, 57%).

MS: (MALDI, cinnamic acid): m/z 8755.1 ± 0.8(M+H)^+ [calcd 8754.9 Da].

2.3.3 Circular Dichroism (CD) Experiments

All chemicals were reagent grade except for GuHCl which was electrophoresis grade. The pH of the buffer solution was determined using a Fisher Scientific Accumet pH meter and an AccuTupH electrode. The pH meter was calibrated using buffered standards (pH 4.0, 7.0 and 10.0) purchased from Fisher Scientific.

2.3.3.1 Far and Near UV CD Spectra

The CD spectra were recorded on a JASCO J-710 spectropolarimeter connected to an IBM-compatible PC 80286 DOS. The CD is equipped with a circulating water bath set at 25 °C. Three scans were taken and averaged for every sample, and background correction was done
using 50 mM pH 7.0 phosphate buffer at 25 °C. Parameters were set to a scanning speed of 50 nm/vmin, a step resolution of 0.1 or 0.2 nm, and a bandwidth of 2 nm. The response and sensitivity were set to 2.0 seconds and 20 mdeg, respectively. The instrument was routinely calibrated with d-10-camphorsulfonic acid (80). Error bars were approximately ±5 % and represent the standard deviations for the average of each point. Quartz cuvettes supplied by Hellma were used to hold the samples. Path lengths of 1 mm or 1 cm were used depending on the desired concentration. Each spectrum was acquired at least three times on separately prepared samples on different days.

The raw CD spectra are normalized to mean residue ellipticity [θ] at 222 nm using the following equation:

\[ [\theta]_{222} = \frac{\theta_{\text{obs}}}{10lcn} \]

where \( \theta_{\text{obs}} \) is the observed ellipticity measured in millidegrees, \( l \) is the pathlength of the cell in centimeters, \( c \) is the peptide concentration in mol/L, and \( n \) is the number of residues in the cavitein. The concentrations of the caviteins were determined using the Bradford assay from the UV measurements at 595 nm (76). Errors were typically ±5 %.

The mean residue ellipticity at 222 nm, \([\theta]_{222}\), is usually used to quantify the amount of α-helix present in a structure by taking the ratio of the observed \([\theta]_{222}\) to the mean residue ellipticity of a complete helix, \([\theta]_H\). Chen and coworkers developed an equation to determine the \([\theta]_H\) of a complete helix, taking into account the length dependence of the helix (73). The equation is as follows:
\[ [\theta]_H = X_\infty^H \left(1 - \frac{k}{n}\right) \]

where \( X_\infty^H \) is the molar ellipticity of an infinite helix, \( k \) is the wavelength dependent factor (3.0 at 222 nm), and \( n \) is the number of residues in a helix. The values of \( X_\infty^H \) range from -37,400 deg cm\(^2\) dmol\(^{-1}\) used by Chen and coworkers (73) to -42,500 deg cm\(^2\) dmol\(^{-1}\) used by Scholtz and coworkers (81). Smith and Scholtz derived an overall equation to determine the fraction of an \( \alpha \)-helix, \( f_H \): (81, 82):

\[
f_H = \frac{[\theta]_{222} - [\theta]_C}{[\theta]_H - [\theta]_C}
\]

where the value for the complete random coil is \([\theta]_C = 640 \text{ deg cm}^2 \text{ dmol}^{-1}\) and the chain length dependent full helix is \([\theta]_H \) from above. This above equation was used to calculate the fraction of an \( \alpha \)-helix of the caviteins listed in Table 2.3.

### 2.3.3.2 Denaturation Studies

GuHCl-induced denaturation experiments were carried out on the CD spectrometer using a solution of 8.0 M guanidine hydrochloride (GuHCl). Quartz cuvettes with either 1 mm or 1 cm path lengths were used depending on the desired sample concentration. For the high concentration studies, the cavitein samples were 40 \( \mu \text{M} \), and for the low concentration studies,
the cavitein samples were 4 μM. The exact concentration of the GuHCl stock solution was
determined by refractometry (41), and found to be 8.0 M with less than 0.5 % error. Each
cavitein was analyzed three times with each path length using separately prepared samples, and
the errors were found to be within 5 %. To obtain the most concentrated 8.0 M solution of
GuHCl, the protein was directly dissolved in the 8.0 M stock solution. The other samples of
lower GuHCl concentration were prepared by appropriately diluting the samples with pH 7.0, 50
mM phosphate buffer. The samples were measured from 0 M to 8.0 M in 0.25 M increments.
When GuHCl was removed by reversed phase HPLC, the protein returned to its non-denatured
state, which shows that the unfolding of the caviteins with GuHCl is a reversible process. The
effect of adding the GuHCl is immediate as the results were the same regardless of whether the
measurements were taken 5 minutes or a week after preparation of the samples. The samples
were typically prepared a day ahead and vortexed right before measurements were taken.

The stabilities of the caviteins were determined using the linear extrapolation method
(43). This method assumes that the folding is a reversible, two-state process:

\[ U \rightleftharpoons N \]

where N is the folded native-state of the protein, and U is the fully unfolded state of the protein.
It also assumes that the folding free energy is linearly dependent on the concentration of GuHCl:

\[ \Delta G^\circ_{obs} = \Delta G^\circ_{H_2O} - m[GuHCl] \]
where $\Delta G^o_{obs}$ is the observed free energy of folding at a particular denaturant concentration, $\Delta G^o_{H_2O}$ is the free energy of folding in the absence of denaturant, and $m$ is the change in $\Delta G^o_{obs}$ with respect to the concentration of guanidine hydrochloride, [GuHCl]. At a particular denaturant concentration, $\Delta G^o_{obs}$ can be calculated according to the following equation:

$$
\Delta G^o_{obs} = -RT \ln K_{obs}
$$

where $R = \text{the universal gas constant (1.987} \times 10^{-3} \text{ kcal K}^{-1} \text{ mol}^{-1}$), $T$ is the temperature (298 K), and $K_{obs}$ is the equilibrium constant for folding, which is calculated from:

$$
K_{obs} = [N]/[U] = e^{-\Delta G^o_{obs}/RT} = e^{-((\Delta G^o_{H_2O} - m[GuHCl])/RT)} = f_N / f_U = f_N / (1 - f_N)
$$

where $f_N$ is the fraction of protein present in the folded state and $f_U$ is the fraction of protein present in the unfolded state, which is also equal to $(1 - f_N)$. From the above equations, the following can be derived.

$$
f_N = e^{-((\Delta G^o_{H_2O} - m[GuHCl])/RT)}/[1 + e^{-((\Delta G^o_{H_2O} - m[GuHCl])/RT)}]
$$

For proteins with high stabilities, there may not be sufficient data at high concentrations of GuHCl to approximate the post-transitional baseline. Therefore, the following equation was used to approximate the pre- and post-transitional baselines (83):
\[ \theta_{\text{obs}} = \theta_N (f_N (1 - \alpha [\text{GuHCl}])) + \theta_U (1 - f_N) \]

where \( \theta_{\text{obs}} \) is the ellipticity at 222 nm at a given GuHCl concentration, \( \theta_N \) is the ellipticity of the fully folded state in the absence of GuHCl, \( \theta_U \) is the ellipticity of the unfolded state, and \( \alpha \) is a constant. \( \theta_N, \theta_U, \Delta G_{H_2O}^0, m \) and \( \alpha \) values were determined by nonlinear least-squares analysis using the KaleidaGraph 3.08 (Synergy Software) program. The value of \( \theta_N \) was normalized to 0.9999 in all cases, and \( \theta_U \) was set to 0.0001. The errors reported in this thesis were calculated from this program.

### 2.3.4 Sedimentation Equilibria Studies

Sedimentation equilibria studies were carried out to determine the oligomeric state of the caviteins using a temperature controlled Beckman Coulter Optima™ XL-I analytical ultracentrifuge (located in the NCE building on the UBC campus) with an An60 Ti 4-hole rotor, and an An50 Ti 8-hole rotor. Different initial concentrations of 20, 40, 60 μM solutions were made up with 50 mM pH 7.0 phosphate buffer. To each sample, KCl (0.08 M) was added. The samples were loaded with 125 μL of cavitein solution and 135 μL of reference solution into ultracentrifuge cells containing Epon six-channel centrepieces with 12 mm pathlengths and quartz windows. Data was collected at 20 °C at rotor speeds of 27 000, 35 000 and 40 000 rpm until equilibration was reached across the cell. Samples were allowed to equilibrate for 29 to 35 hours. During this time, scans were taken three hours apart to ensure that equilibrium had been reached. Scans were detected by UV at a wavelength of 270 nm.
The solution density of the samples in aqueous buffer was estimated to be 1.000 g/mL. The partial specific volumes of the caviteins were calculated based on its amino acid composition (51); however, the partial specific volume of the cavitand template portion was not considered in this calculation. It is important to note that a small error in the partial specific volume results in a large error in the molecular weight determination as mentioned in section 2.1.2.4.

The data was analyzed using a nonlinear least-squares analysis using software by Johnson and Yphantis (84). The nonlinear least-squares analysis using the Origin and Beckman (Optima™ XL-A/XL-I Data Analysis Software user’s manual version 6.03) software was also used, and gave similar results. The data was initially fit to a single non-associating ideal species model using the following equation (85):

\[ A_r = \exp[\ln(A_o) + M\omega^2(1 - \bar{v}\rho/2RT)(r^2 - r_o^2)] + E \]

where \( A_r \) is the absorbance at radius \( r \), \( A_o \) is the absorbance at the reference radius \( r_o \) (typically the meniscus), \( \bar{v} \) is the partial specific volume in mL/g, \( \rho \) is the solvent density in g/mL, \( \omega \) is the angular velocity of the rotor in radian/sec, \( M \) is the molecular weight, \( T \) is the absolute temperature in Kelvin, \( R \) is the universal gas constant \( (8.314 \times 10^7 \text{ erg mol}^{-1} \text{ K}^{-1}) \), and \( E \) is the baseline error correction factor.

In cases where the data did not fit well to a single ideal species, best fits were made to a self-associating species using the following equation (85):
\[ A_r = A_{(\text{monomer}, r_0)} \exp[[1 - \nu \rho) \omega^2 / 2RT)(M(r^2 - r_0^2))] + A_{(\text{monomer}, r_0)} \exp[[1 - \nu \rho) \omega^2 / 2RT)n_2(M(r^2 - r_0^2))] + A_{(\text{monomer}, r_0)} \exp[[1 - \nu \rho) \omega^2 / 2RT)n_3(M(r^2 - r_0^2))] + A_{(\text{monomer}, r_0)} \exp[[1 - \nu \rho) \omega^2 / 2RT)n_4(M(r^2 - r_0^2))] + E \]

where \( A_r \) is the absorbance at radius \( r \), \( A_{(\text{monomer}, r_0)} \) is the absorbance of the monomer at the reference radius \( r_0 \); \( n_2, n_3 \) and \( n_4 \) are the stoichiometry for species 2, 3 and 4, respectively; \( K_{a,2}, K_{a,3}, \) and \( K_{a,4} \) are the association constants for the monomer-\( n \)-mer of species 2, 3, and 4, respectively. \( M \) is the molecular weight and \( E \) is the baseline offset.

The values are considered to fit well when the theoretical and experimental data overlay on the absorbance versus \( r^2/2 \) plot. A good fit should result in evenly spread points around the x-axis on the residual plots.

The nonlinear analysis program calculates the association constant in absorbance units. The conversion to units of \( M^{-1} \) requires an extinction coefficient \( \varepsilon \), calculated using Beer's law:

\[ \varepsilon = A / bc \]

where \( A \) is the absorbance of a solution with of concentration \( c \) in mol/L, and \( b \) is the pathlength in cm. This measurement was carried out on a Cary UV-visible spectrophotometer at 270 nm (same \( \lambda \) used for the analytical ultracentrifugation). Once the \( \varepsilon \) was determined, the following equation was used to find the association constant for a monomer-dimer equilibrium (86):

\[ K_{\text{conc}} = K_{abs} \times \frac{\varepsilon l}{2} \]
where $K_{\text{conc}}$ is the association constant in M$^{-1}$, $K_{\text{abs}}$ is the association constant in absorbance units, and $l$ is the pathlength in cm (1.2 cm for a 12-mm centerpiece).

Figures 2.29 to 2.40 show the sedimentation equilibria analyses of the caviteins in the second generation series. The fits were evaluated in terms of the randomness of the residuals and the magnitude of the spread (expressed in terms of standard deviation from the average absorbance collected at each radial position). Not all nine plots (3 rotor speeds, 3 concentrations) are shown for each cavitein since the resulting fits look very similar to one another. However, for the purpose of variety, the following figures are shown: two fits to a monomer-dimer in equilibrium for the 1GS2 cavitein (Figures 2.29 and 2.30), a fit to a monomer for the 1GS2 cavitein to show what a poor fit looks like (Figure 2.31), three fits to a single non-interacting species for the 2GS2 cavitein (Figures 2.32 to 2.34), three fits to a monomer/dimer for the 3GS2 cavitein (Figures 2.35 to 2.37), and three fits to a single non-interacting species for the 4GS2 cavitein (Figures 2.38 to 2.40). Data are shown using various concentrations and rotor speeds.
Figure 2.29. Sedimentation equilibrium analysis of 20 μM 1GS2 cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 27,000 rpm. The solid line in the upper plot represents the theoretical fit to a monomer-dimer in equilibrium. The lower plot represents the residuals to the fit.
Figure 2.30. Sedimentation equilibrium analysis of 20 μM 1GS2 cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 35,000 rpm. The solid line in the upper plot represents the theoretical fit to a monomer-dimer in equilibrium. The lower plot represents the residuals to the fit.
Figure 2.31. Sedimentation equilibrium analysis of 20 μM 1GS2 cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 35,000 rpm. The solid line in the upper plot represents the theoretical fit to a monomer in equilibrium. The lower plot represents the residuals to the fit, showing the scatter of data points around the best-fit line (note that the residuals are unevenly spread when the 1GS2 cavitein is forced to be a monomer).
Figure 2.32. Sedimentation equilibrium analysis of 40 μM 2GS2 cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 27,000 rpm. The solid line in the upper plot represents the theoretical fit to a single non-interacting species. The lower plot represents the residuals to the fit.
Figure 2.33. Sedimentation equilibrium analysis of 40 µM 2GS2 cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 35,000 rpm. The solid line in the upper plot represents the theoretical fit to a single non-interacting species. The lower plot represents the residuals to the fit.
Figure 2.34. Sedimentation equilibrium analysis of 40 μM 2GS2 cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 40,000 rpm. The solid line in the upper plot represents the theoretical fit to a single non-interacting species. The lower plot represents the residuals to the fit.
Figure 2.35. Sedimentation equilibrium analysis of 20 µM 3GS2 cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 27,000 rpm. The solid line in the upper plot represents the theoretical fit to a monomer-dimer in equilibrium. The lower plot represents the residuals to the fit.
Figure 2.36. Sedimentation equilibrium analysis of 20 μM 3GS2 cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 35,000 rpm. The solid line in the upper plot represents the theoretical fit to a monomer-dimer in equilibrium. The lower plot represents the residuals to the fit.
Figure 2.37. Sedimentation equilibrium analysis of 20 μM 3GS2 cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 40,000 rpm. The solid line in the upper plot represents the theoretical fit to a monomer-dimer in equilibrium. The lower plot represents the residuals to the fit.
Figure 2.38. Sedimentation equilibrium analysis of 20 μM 4GS2 cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 27,000 rpm. The solid line in the upper plot represents the theoretical fit to a single non-interacting species. The lower plot represents the residuals to the fit.
Figure 2.39. Sedimentation equilibrium analysis of 40 μM 4GS2 cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 27,000 rpm. The solid line in the upper plot represents the theoretical fit to a single non-interacting species. The lower plot represents the residuals to the fit.
Figure 2.40. Sedimentation equilibrium analysis of 60 μM 4GS2 cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 27,000 rpm. The solid line in the upper plot represents the theoretical fit to a single non-interacting species. The lower plot represents the residuals to the fit.
2.3.5 NMR Experiments

The \textsuperscript{1}H NMR spectra were acquired by Okon in the laboratory of McIntosh, UBC Department of Biochemistry, on either a 500 MHz Varian Unity or a 600 MHz Varian Unity Inova spectrometer.

2.3.5.1 1D \textsuperscript{1}H NMR Spectra

1D \textsuperscript{1}H NMR spectra were obtained using a 600 MHz Varian Unity Inova NMR spectrometer at 25 °C to examine the dispersion and sharpness of the signals. The cavitein samples were prepared by dissolving the proteins in 0.45 mL 50 mM phosphate buffer, pH 7.0 and 0.05 mL D\textsubscript{2}O. The cavitein concentrations were made up to be 1.5 mM. The data were processed using MestReC Version 2.3.

2.3.5.2 2D \textsuperscript{1}H NMR Spectra

To obtain the 2D \textsuperscript{1}H NMR spectra of the 2GS2 cavitein sample, a 500 MHz Varian Unity instrument was used. The 1GS2 cavitein sample was run on a 600 MHz Varian Unity Inova. The concentration of the 2GS2 cavitein sample was 2 mM. The concentration of the 1GS2 cavitein sample was 1.1 mM. The samples were dissolved in 50 mM acetate buffer (90:10, H\textsubscript{2}O:D\textsubscript{2}O) at pH 4.62, and were run at 20 °C, taking 256 scans. 2D spectra were processed by Okon using NMRpipe, and polynomial baseline correction was carried out. No line-broadening
functions were used. The spectra were calibrated to the water signal at 4.78 ppm. The resonance assignments were assisted using NMRView.

2.3.5.3 N-H/D Exchange 1D $^1H$ NMR Spectra

The N-H/D exchange experiments were run on a 500 MHz Varian Unity at 20 °C. The samples were prepared by adding 50 mM pH 4.62 CH$_3$COOH/NaCH$_3$COO buffer, and lyophilizing these caviteins. The experiment was initiated by the addition of 100 % D$_2$O. The sample concentrations were 2.2 mM for the 1GS2 cavitein, 1.9 mM for the 2GS2 cavitein, 1.9 mM for the 3GS2 cavitein, and 1.9 mM for the 4GS2 cavitein. The pD was determined by using the equation below (87):

$$pD = pH_{\text{read}} + 0.4$$

where pD is the pH in D$_2$O, and pH$_{\text{read}}$ is the reading from the electrode measurements using a Fisher Scientific Accumet pH meter. The first scan was taken immediately after shimming the instrument. The next few scans were taken within minutes of each other, and gradually the time between measurements increased as the number of measurements increased. For each spectrum, four scans that took 2 minutes and 16 seconds were acquired. Some protons exchanged immediately, while some other signals remained for a few months. Reference spectra of each of the caviteins in 50 mM pH 4.62 acetate buffer were also taken with only 10 % D$_2$O.
The peak heights were integrated and normalized with the non-exchangeable proton $H_{\text{out}}$ (near 6 ppm) of the cavitand template. To calculate the first-order rate constant of the longest-lived amide proton, the first order rate equation was used:

$$\ln\left(\frac{[H_0]}{[H_t]}\right) = k_{ex}t$$

where $k_{ex}$ is the first order rate constant, $t$ is the time at which the scan was taken, $[H_0]$ is the integration of the amide proton at time zero, and $[H_t]$ is the integration of the amide proton at time $t$. For consistency, the time at which the scans were finished acquiring was used to perform the calculations. From the rate constant, the half-life ($t_{1/2}$) of the last amide proton to disappear was calculated:

$$t_{1/2} = \ln 2 / k_{ex}$$

The intrinsic half-life ($t_{1/2-\text{int}}$) of an unprotected proton was calculated from the equation in reference (64):

$$t_{1/2-\text{int}} = 200/(10^{(pH-3)} + 10^{(3-pH)}) \times 10^{0.05T}$$

where $T$ is in degrees celcius. $k_{\text{int}}$ was calculated in the same manner as $k_{ex}$ except using the intrinsic half-life ($t_{1/2-\text{int}}$). The protection factor $P$ was then calculated using the following equation:

$$P = k_{\text{int}} / k_{ex}$$
2.3.6 ANS Binding Studies

ANS fluorescence spectra were measured on a Varian CARY Eclipse at 25°C. The quartz cell had a pathlength of 10 mm. All samples contained 2 μM ANS in 50 mM phosphate buffer at pH 7.0. Each solution contained either 50 μM of the cavitein, 95 % ethanol, or 100 % HPLC grade methanol. The excitation wavelength was set to 370 nm and the emission was recorded between 400 and 600 nm. For the ANS binding studies using denaturant, each solution contained 50 μM cavitein, 2 μM ANS and GuHCl (ranging from 0M to 8M in 1M increments) in 50 mM phosphate buffer at pH 7.0. The error bars were calculated to be within 5 % using the standard deviation of three individual runs.
2.4 References

CHAPTER THREE: Computer Modelling Study on the Second Generation Linker Caviteins*

3.0 Introduction

As mentioned in Chapter 1, the template assembled synthetic protein (TASP) approach has been a useful tool in studying protein folding (1, 2). The template serves to control the number of helices in a bundle, arrange the helices in a parallel fashion, help overcome the entropic barrier of bringing together the peptides, to stabilize the overall structure. Although the ‘template-assembly’ method has been a useful technique in understanding protein folding, there is still much to be learned about the relationship between the template and the protein structure. To further understand this relationship, a series of proteins differing in the number of N-terminal Gly residues between the template and the peptides were synthesized and characterized experimentally (see Chapter 2). It was found that changing the linker length by one Gly (per peptide) had a substantial effect on the overall structure and properties of the caviteins. In this chapter, computational methods are used to further understand the characteristics of these TASPs. The use of computer modelling has become increasingly popular with advances in modern technology (3). Many chemical and biological processes can be investigated by computational techniques including protein folding (4-6), protein binding (7), and protein dynamics (8). The use of computers to perform calculations and simulations is referred to as in silico. The study of de novo proteins in silico has been a topic of interest due to their simplicity.

* "A version of this chapter will be submitted for publication. Seo, E., Scott, W.R.P., Straus, S.K., Sherman, J.C. Characterization of a New Generation of Caviteins by Molecular Dynamics Simulations."
and native-like features (9). Sikorski and coworkers have studied DeGrado's hierarchic de novo proteins (see Chapter 1.4.1) using Monte Carlos simulations, and found that the computational results were consistent with the experimental evidence (10, 11). The proteins that were found to be molten globule structures experimentally were also found to have non-specific side chain packing by computation. Similarly, the proteins that were found to demonstrate native-like character experimentally were found to pack more specifically by observing the simulation results. The reduced degree of fluctuations in the radius of gyration, and total energy were lower for a native-state protein. In addition, the number of maintained contacts in a native-protein was higher than that of a molten globule structure.

Most molecular modelling consists of three main stages: 1) selection of a model that describes the intra- and inter-molecular interactions in the system, 2) calculations using an energy minimization, a molecular dynamics, or Monte Carlo simulation, and 3) analysis.

The first generation caviteins mentioned in Chapter 2 have been studied by molecular dynamics (MD) simulations (12). The data from the computer modelling showed consistency with the experimental results in terms of helicity and propensity to self-associate. In this chapter, the second generation caviteins that were studied experimentally in Chapter 2 are investigated using MD simulations to further understand the characteristics of these systems. This project was done in collaboration with Scott and Straus (13). The simulations were carried out in explicit water, which means that all the water molecules were considered during the course of the simulation. The experimental data provided a macroscopic picture of the caviteins; however, they were difficult to interpret. Computer simulations can provide information at an atomic level, and should complement the results obtained from the experiments. The calculations carried out during the course of the simulation were used to get a better understanding of how the caviteins behave. However, computer modelling also has its limitations. One fundamental difficulty with
the simulation of macromolecules is the assumptions made in choosing a force field. As Maddox states (14), “so long as the chosen force field is necessarily a crude approximation to the truth, no amount of precision in the computation can give a person confidence that the state with the lowest energy is really the global minimum, let alone that it approximates to that of the native protein with any accuracy.” The other major drawback is due to time limitations as the timescales which can be simulated are on the order of nanoseconds, whereas biologically relevant phenomena, such as protein folding, usually occur on timescales of milliseconds. In the case of the caviteins, another limitation is due to the absence of any X-ray crystal or NMR structures. Because of the lack of a real physical structure, a starting configuration was constructed using all the available experimental data (see Methods section 3.1). Nevertheless, \textit{de novo} proteins make attractive molecules for simulation studies because of their small size and easily modifiable features. They can also be used to generate hypotheses and help fine-tune the design process.

What questions do we want answered and what information could be extracted from the simulation results? What is the optimal linker length between the peptides and the template? Is one Gly too short for linking the peptides to the template in terms of optimal packing of the helices? Why is the 2GS2 cavitein unique in terms of stability and native-like character? Why does the 3GS2 cavitein behave strangely over time in solution as compared to the other caviteins? How much of an effect does the template have on the protein structure when the linker is relatively long like in the case of the 4GS2 cavitein? Answering these questions is a difficult task. Experimental data and computer modelling together may not give all the answers; however, a combination of the investigated results can be used to try to gain insight into the behaviour of the caviteins.
3.1 Methods

The simulations were performed using the GROMOS96 biomolecular simulation package and the 43A1 force field (15, 16). Appendix B includes definitions for some of the terms used in this chapter and Appendix C includes the steps for setting up the simulations for the cavitein systems. The methods explained here are similar to those in the MD simulation studies carried out for the first generation caviteins (12). However, two additional properties are explored and described below: tilt of the helices with respect to the cavitand (section 3.1.5) and the linker orientation (section 3.1.6).

The cavitand structure has been solved by X-ray crystallography (17), but the cavitein as a whole has not been solved by X-ray crystallography nor NMR spectroscopy, therefore the complete cavitein structure was built in silico. For the protein portion of the cavitein, the helix structure was taken from helix D of wild type bacteriorhodopsin (1CWQ.pdb) (18), which is a 7-helix bundle protein. The side chains were replaced with the residues of the desired sequence using the CHIMERA program (19, 20). The addition of the Gly residues to the N-terminus of the helix was carried out manually using ViewerPro software. Using this program, the modified helix was replicated to make four copies and arranged in a parallel fashion. The helices were then interactively brought in close proximity to the corners of a square with 7 Å sides, which consisted of the cavitand template sulphur atoms. The hydrophobic side chains of the amphiphilic helices were oriented toward the centre of the square. In Simulation 1, the helices were placed parallel to each other in a square bundle without any bias towards a left- or right-handed coiled coil structure. TWISTER (21) was used to ensure that the coiled coil phase yield per residue was zero. The N-terminal ends of the helices were interactively brought in close proximity to the arylthiol sulphurs of the template. The system was then minimized to relieve
any strain in the linker region. Care was taken to ensure that the elbow of linkage was in its preferred conformation (22). The system was then solvated, minimized again, and equilibrated for 250 ps without any position restraints. Rectangular periodic boundary condition were imposed. Simulations were performed in the NPT ensemble, (constant number of atoms, constant pressure and constant temperature) with a temperature of 300 K and a pressure of 1 atm using the Berendsen weak coupling methods (23). The SHAKE method (24), was used to constrain the covalent bonds. A reaction field long-range correction (25) was applied to the truncated Coulomb potential. Subsequent simulations of 20 ns were carried out without position restraints for data analysis.

In the second set of simulations, Simulation 2, all of the above conditions were used with the exception that position restraints were switched on during the pre-minimization, minimization, and equilibration steps. The position restraints were enforced on all of the α-carbons of the helices except for the α-carbons of the N-terminal glycines. Simulation 2 was carried out to examine the effects of position restraints for setting up the MD runs, as well as to provide a slightly different starting configuration than Simulation 1.

Since the three-dimensional structures of the caviteins are unknown, in order to reduce any bias towards a certain configuration, a third set of simulations was setup and run. For Simulation 3, the helices of the initial configuration were splayed out as far as possible using ViewerPro, where there was still slight interhelical interaction when represented as a space-filling model. When the interhelical distances were set too far, numerous water molecules were found to go inside the core of the helices and buffer the hydrophobic interactions between the side chains. Simulation 3 was run with position restraints as in Simulation 2 during the pre-minimization, minimization, and equilibration step. Simulations 2 and 3 were carried out for 20
ns for the 1GS2 cavitein, and were run for 5 ns for the rest of the caviteins, without any position restraints. Table 3.1 shows a summary of the differences between the three different simulations.

Table 3.1. Summary of the differences between the three sets of simulations for the second generation caviteins. 'No' and 'Yes' refer to if the position restraints were on during the specified step.

<table>
<thead>
<tr>
<th>Simulation Number</th>
<th>Starting Configuration</th>
<th>Pre-minimization</th>
<th>Minimization</th>
<th>Equilibration</th>
<th>Molecular Simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Helices parallel to one another, no supercoil</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>Same as above</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>Helices slightly splayed from N-terminus, no supercoil</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

The computer modelling results were analyzed in terms of helical content, conformational specificity, compactness of the helices (relative motion), supercoiling abilities, tilt of helices with respect to the cavitand, and linker elbow orientation.

3.1.1 Helical Content

The helical content during the course of the simulation was analyzed using a modified version of PROCHECK (26). Ramachandran plots were generated to assess the helicity of the caviteins. The allowed region for the right-handed α-helices on a Ramachandran plot is shown in Chapter 1.1.2.
3.1.2 Conformational Specificity

Conformational specificity was characterized in terms of structural spread encountered during the simulation. This distribution was determined using a clustering algorithm. The algorithm partitions $S$, the sequence of all the conformations $x_i$ where $i = 1 \ldots N$ that are encountered during the simulation, into a sequence of subsets (clusters), each containing similar conformational attributes. In other words, starting from $S$, a subset $s_0$, containing the conformation $c_0$ with the largest number of neighbours in $S$ and all of its similar structures was determined. The determined conformation $c_0$ is considered the representative of all of the conformations in $s_0$. Upon removal of $s_0$ from $S$, the same procedure is applied to the remaining conformations in $S$ until all of the conformations have been grouped. In this way, $S$ is partitioned into an ordered sequence of NC sets (clusters) $s_0, s_1, \ldots s_{NC-1}$, decreasing in size, and each conformation $x_i$ is assigned to a single cluster number $C(i) = 0 \ldots NC-1$, where $i = 1 \ldots N$. In a subsequent operation, the average potential energy of the solvated cavitein $E_j$ is determined for each cluster $j = 0 \ldots NC-1$. We plot $\overline{E}_c(i)$ for each $i$ in the set of simulated conformations as a function of time together with the representative structures $c_0, c_1, c_2, \ldots c_{NC-1}$. To determine whether two conformations, $Q$ and $R$ are considered neighbours, the double difference matrix distance metric $D^{QR}$ was used. Two conformations are considered to be neighbours, if $D^{QR}$ using all $\alpha$-carbons is less than 1.5 Å. The equation for the double difference matrix distance metric is as follows (12):

$$D^{QR} = \left[ \frac{2}{N(N-1)} \sum_{j \neq i} (d^Q_{ij} - d^R_{ij})^2 \right]^{\frac{1}{2}}$$
where $N$ is the number of atoms considered in the distance measurements, $d_{ij}^Q$ is the scalar interatomic distance of two considered atoms $i$ and $j$ within one structure $Q$, and likewise $d_{ij}^R$ is the interatomic distance of two atoms in $R$. This measure was chosen over the more popular RMS positional deviation distance measurement, because it avoids the need for a rigid-body superposition of structures that can lead to misleading results for very different conformations.

### 3.1.3 Perimeter Distance

The *perimeter distance* was calculated in order to assess the compactness of the helices. The *perimeter distance* is defined as the sum of the four interatom distances:

$$P^X(t) = d_{x_ex_x}(t) + d_{x_ex_c}(t) + d_{x_cx_o}(t) + d_{x_ax_x}(t)$$

where $d_{x_ex_x}(t)$ denotes the scalar interatom distance between atom number $X$ of two of the four helices A, B, C and D of the caviteins as a function of simulation time $t$. A large number indicates that the helix bundle is loosely packed at the location of atom $X$. A small number indicates that the helices are tightly packed. The distribution of $P^X(t)$ sampled during the simulation is displayed as a normalized histogram, which allows for a visual comparison of the mobility of the different systems. A typical choice for atom $X$ would be the backbone carbonyl carbon of the C-terminal glycine (C16 to C19 for caviteins 1GS2 to 4GS2, respectively), in order to observe the presence or absence of helical splaying. From N-H/D exchange data (see Chapter
2.1.2.5.3), it was found that the central leucine of the helices were last to exchange for all the caviteins. Therefore, another choice for atom $X$ would be the nitrogen of the middle leucine (N9 to N12 for caviteins 1GS2 to 4GS2). The perimeter distance of the carbonyl carbon of the central leucine was also calculated to determine if there is a significant difference between the perimeter distance of a nitrogen atom versus a carbonyl carbon of the same residue.

3.1.4 Supercoiling

To assess supercoiling of the caviteins, the representative structures found using the clustering algorithm were analyzed using TWISTER (21). The coiled coil phase yield was determined for each residue of all the cavitein systems. These results were compared with the near UV CD data (see Chapter 2.1.2.2) and the proton NMR dispersion (see Chapter 2.1.2.5.1) to see if there is any correlation between the simulated and experimental results.

3.1.5 Tilt of Helices with Respect to the Cavitand

To examine the tilt of the helices with respect to the template, a ‘helipad’ program was written by Dr. Walter Scott. The measure of the tilt was calculated based on a plane created by the four cavitand sulphur atoms (‘helicopter pad’), and a residue atom (‘helicopter’) chosen for measurement. Two dimensions were used to determine the tilt: the height and the range of each helix, which are shown in Figure 3.1.
Figure 3.1. Dimensions measured using the 'helipad' program: height (H) and range (R).

The height is the vertical distance of the chosen residue from the plane, and the range is the distance in the plane from the centre of the pad. The $\alpha$-carbon of the alanine was chosen as the 'helicopter' of each helix, since it was the first residue, other than the flexible glycine residues, that initiated the helix and began the tilt of the helices with respect to the cavitand plane.
3.1.6 Linker Elbow Orientation

To investigate the orientation of the helices with respect to the sulphur atoms of the template, the clusters of the representative conformations were analyzed qualitatively. It has been shown that the preferred linker elbow conformation of the arylthiol cavitand was that in which the sulphur lone pairs were pointing away from the oxygen lone pairs of the template (see Figure 2.2 in Chapter 2). This conformation was found to be 5 kcal/mol more stable than when the lone pairs were pointing toward each other, likely due to the minimized dipole-dipole repulsions between the lone pairs. This conformation allowed the peptides to link onto the template with the methylene elbows sticking outside of the cavitand template. Therefore the initial configurations of all the cavitein systems started with this preferred linker conformation. The molecular dynamics data were analyzed to determine if this elbow conformation was maintained throughout the simulation time.
3.2 Results and Discussion

In Chapter 2, the caviteins with varying number of glycine residues in the template-peptide region were characterized by several techniques: 1) CD spectroscopy was used to determine the secondary structure of the caviteins and to see the extent of native-like character in the near UV region, 2) denaturation experiments using guanidine hydrochloride were performed to determine the free energy of folding, 3) analytical ultracentrifugation was used to determine the oligomeric state, 4) NMR spectra were obtained to investigate the native-like character of the caviteins, 5) N-H/D exchange experiments were carried out to monitor the fluctuation of the structures over time, and 6) ANS binding studies were used to probe the molten globular nature of the cavitein cores. From these results, it was concluded that the 2GS2 cavitein was the most well-defined protein in terms of stability, native-like character, and oligomeric state (see Chapter 2 for more details). Here, the experimental results are compared to the simulation results. Do the simulated results agree with the experimental data? To answer this question, the general experimental trends are reviewed and evaluated with the simulation results regarding helical content, conformational specificity, relative motion of the helices, and supercoiling capabilities. In addition, issues that arose from the modelling results, concerning the arrangement of the helices with respect to the bowl, are also examined. These issues include tilting of the helices relative to the bowl, and the elbow orientation at the linkage site of the peptides to the cavitand template. In agreement with the experimental results, the simulations showed that the length and flexibility of the linker region play a substantial role in the structure and dynamics of the cavitein systems.
3.2.1 Helical Content

Figures 3.2 to 3.5 show the average Ramachandran plots of all four helices in all four caviteins over the course of the entire Simulation 1. The helices were maintained throughout the simulations for the most part just like the caviteins in the first generation series (12). The first phi and last psi could not be defined, and therefore the first and last residues are not shown. On average, right-handed helices were present for at least 75% of the time. This finding agrees with the experimental results that showed that the caviteins are highly helical (see Chapter 2.1.2.1). There are a few points that do not fall in the allowed regions of a right-handed helix. These points generally correspond to the Gly residues in the linker region or the residues near the C-terminus of the helices (see residues labels on Figures 3.2-3.5). This is not unusual since glycine is flexible and can therefore adopt any phi/psi angles. The residues other than glycine that are in the sterically disallowed regions are usually from the last residue(s) at the C-terminal end. These transitions represent an uncoiling of the helices at the C-terminus, which is not uncommon from MD simulation results of α-helices (27). For example, for the 1GS2 cavitein, K15 from the C-terminal end is close to or in the β-sheet region of the plot for helices A, B, and C. For helix D, K14 as well as K15 are outside the helical region. Occasionally, certain residues that are near the centre of the helix transgress from the helix region, which may be indicative of a kink in the helix. In helix D of the 1GS2 cavitein, residues K7 and K8 are also in the disallowed area of the right-handed helix. The kink and the C-terminal fraying of the D helix corresponds to the smaller helical content that was observed for the 1GS2 cavitein compared to the rest of the caviteins (see cluster 0 in Figure 3.2 and far UV CD spectrum in Figure 2.8). Frequent excursions of the linker glycines and the C-terminal residues are also observed for caviteins 2GS2, 3GS2 and 4GS2.
Figure 3.2. Ramachandran plot of the 1GS2 cavitein. The average dihedral angles over the course of the simulation for the residues in the four helices A, B, C and D are shown. The helices are found to be stable throughout the simulation time for the most part. Residue K15 at the C-terminal end is close to or in the β-sheet region for helices A, B and C. For helix D, K15 as well as K14, K8 and K7 are outside the right-handed α-helix region.
Figure 3.3. Ramachandran plot of the 2GS2 cavitein. The average dihedral angles over the course of the simulation for the residues in the four helices A, B, C and D are shown. Residues in the linker region (G2) and at the C terminal end (K16, K15, L14, L13) are outside the right-handed helix region.
Figure 3.4. Ramachandran plot of the 3GS2 cavitein. The average dihedral angles over the course of the simulation for the residues in the four helices A, B, C and D are shown. Residues in the linker region (G2, G3) and residues at the C-terminal end (K17, K16, L15) are outside the right-handed helix region.
Figure 3.5. Ramachandran plot of the 4GS2 cavitein. The average dihedral angles over the course of the simulation for the residues in the four helices A, B, C and D are shown. Residues in the linker region (G2, G3, G4) and K18 at the C-terminal end are outside the right-handed helix region.
Very similar results were observed for Simulations 2 and 3 with right-handed helices present for at least 75% of the time. The only excursions from this region were again the residues at the C-terminal regions and the flexible Gly linkers.

### 3.2.2 Conformational Specificity

The conformational specificity of the four caviteins was determined by calculating the extent of spread in the structures during the course of the simulations. The percentage of each cluster was calculated using the clustering algorithm described in the Methods section. Table 3.2 shows the results for Simulation 1.

**Table 3.2.** Clustering results for the second generation caviteins, showing the degree of conformational spread over 20 ns for Simulation 1. The 1GS2 cavitein samples a smaller conformational space compared to the other caviteins with a predominant conformation of 73% during the total simulated time.

<table>
<thead>
<tr>
<th>Caviteins</th>
<th>Total number of clusters</th>
<th>% of cluster 0</th>
<th>% of cluster 1</th>
<th>% of cluster 2</th>
<th>% of cluster 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1GS2</td>
<td>10</td>
<td>73</td>
<td>18</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>2GS2</td>
<td>14</td>
<td>40</td>
<td>27</td>
<td>16</td>
<td>10</td>
</tr>
<tr>
<td>3GS2</td>
<td>16</td>
<td>31</td>
<td>28</td>
<td>26</td>
<td>9</td>
</tr>
<tr>
<td>4GS2</td>
<td>6</td>
<td>54</td>
<td>42</td>
<td>4</td>
<td>&lt;1</td>
</tr>
</tbody>
</table>

For the 1GS2 cavitein, the most dominant configuration adopted in cluster 0 was occupied 73% of the simulated time (see Figure 3.6). For the other 3 caviteins, the extent of conformational spread was larger. For the 2GS2 and the 4GS2 caviteins, the most predominant
structure occurred 40% and 54% of the time, respectively (see Figures 3.7 and 3.9). For the
3GS2 cavitein, the most populated cluster 0 structure was only 31%, which was not vastly
different from cluster 1 (28%) and cluster 2 (26%) (see Figure 3.8).

For all the 4GS2 cavitein clusters (see Figure 3.9), the helices were relatively further
from the template compared to the other three caviteins, as would be expected for a longer linker.
This result implies that the cavitand template likely had minimal influences on the interhelical
interactions.
Figure 3.6. Average energy of the 3 most predominant clusters as a function of simulation time for Simulation 1 of the 1GS2 cavitein: Cluster 0 (73%), Cluster 1 (18 %) and Cluster 2 (6 %).
Figure 3.7. Average energy of the 3 most predominant clusters as a function of simulation time for Simulation 1 of the 2GS2 cavitein: Cluster 0 (40%), Cluster 1 (27 %) and Cluster 2 (16 %).
Figure 3.8. Average energy of the 3 most predominant clusters as a function of simulation time for Simulation 1 of the 3GS2 cavitein: Cluster 0 (31%), Cluster 1 (28 %) and Cluster 2 (26 %).
Figure 3.9. Average energy of the 3 most predominant clusters as a function of simulation time for Simulation 1 of the 4GS2 cavitein: Cluster 0 (54%), Cluster 1 (42 %) and Cluster 2 (4 %).
Although the tendency to oligomerize could not be measured using the current MD calculations, certain qualitative observations may be used to explain how the caviteins could self-aggregate. The analytical ultracentrifugation results showed that the 2GS2 and 4GS2 caviteins were monomeric in solution (see Chapter 2.1.2.4). From the MD simulations, the helices of the 2GS2 cavitein were tilted with respect to the bowl (see Figure 3.7). A possible explanation to why the 2GS2 cavitein exists as a monomer in solution may be due to this tilt, which could facilitate larger burial of the hydrophobic core, which could in turn prevent self-aggregation. The simulation results of the first generation series revealed that the helices were inclined with respect to the cavitand template for the one Gly linker variant (12). However, this species was found to be a dimer experimentally (22). Here, the tilt may have instigated the self-association by causing the helices to be out of register to one another, exposing some of the hydrophobic core; whereas, in the 2GS2 cavitein case, the tilt appears to have allowed the helices to pack most efficiently to prevent self-aggregation. The tilt of the helices with respect to the template is quantified in section 3.2.5. For the 4GS2 cavitein, the longer linker may have relieved any restriction due to the cavitand template and allowed for sufficient packing to avoid self-aggregation.

In cluster 0 of the 1GS2 protein, helix D unwound more at the C-terminus than the other helices (see cluster 0 in Figure 3.6), which agrees with the lower helicity displayed in the far UV CD spectrum of the 1GS2 cavitein (see Chapter 2.1.2.1). For comparison, the clustering results from Simulation 2 and 3 are shown in Tables 3.3 and 3.4, respectively.
Table 3.3. Clustering results for the second generation caviteins showing the degree of conformational spread over 20 ns for the 1GS2 cavitein and 5 ns for the rest of the caviteins in the second generation for Simulation 2. The caviteins sampled a smaller conformational space compared to the caviteins in Simulation 1.

<table>
<thead>
<tr>
<th>Caviteins</th>
<th>Total number of clusters</th>
<th>% of cluster 0</th>
<th>% of cluster 1</th>
<th>% of cluster 2</th>
<th>% of cluster 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1GS2</td>
<td>8</td>
<td>89</td>
<td>6</td>
<td>3</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>2GS2</td>
<td>5</td>
<td>76</td>
<td>16</td>
<td>8</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>3GS2</td>
<td>5</td>
<td>69</td>
<td>28</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>4GS2</td>
<td>6</td>
<td>84</td>
<td>10</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3.4. Clustering results for the second generation caviteins showing the degree of conformational spread over 20 ns for the 1GS2 cavitein and 5 ns for the rest of the caviteins in Simulation 3. The caviteins sampled a smaller conformational space compared to the caviteins in Simulations 1 and 2. The % of cluster 0 is > 90 % for all the caviteins in Simulation 3 with the exception of the 4GS2 cavitein.

<table>
<thead>
<tr>
<th>Caviteins</th>
<th>Total number of clusters</th>
<th>% of cluster 0</th>
<th>% of cluster 1</th>
<th>% of cluster 2</th>
<th>% of cluster 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1GS2</td>
<td>6</td>
<td>97</td>
<td>1</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>2GS2</td>
<td>5</td>
<td>95</td>
<td>3</td>
<td>1</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>3GS2</td>
<td>7</td>
<td>91</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>4GS2</td>
<td>8</td>
<td>72</td>
<td>20</td>
<td>7</td>
<td>&lt; 1</td>
</tr>
</tbody>
</table>

In general, from the tables containing the clustering results, one can see that the caviteins sampled a smaller conformational space when position restraints were turned on during the pre-minimization, minimization and equilibration steps (Simulations 2 and 3) as compared to when the position restraints were absent (Simulation 1). In general, the representative cluster diagrams of the three sets of simulations were similar. The slight differences are addressed in the following sections.
One should be cautious in interpreting the clustering results between the simulations due to the time differences of data collection. Even the data collected during the longest time span of 20 ns was not necessarily enough to make any definite conclusions.

3.2.3 Relative Motion and Compactness of the Helices

To examine the relative motion of the helices with respect to one another, the perimeter distances $P^X(t)$ were calculated for each cavitein. The perimeter distance is the sum of the interhelical distances. From the N-H/D exchange data, it was found that the last hydrogen to exchange was from the central leucine of the sequence regardless of the number of glycine linkers at the N-terminus. Therefore, the nitrogen atom of this central leucine (N9 to N12 for the 1GS2 to 4GS2 caviteins) was chosen for this measurement. The carbonyl carbon of this same hydrophobic residue was also evaluated to determine whether there would be much difference between the $P^X(t)$ values of different atoms in the same residue. In addition, the perimeter distance of the terminal carbonyl carbon was also calculated to determine the extent of mobility at the C-terminal region. Table 3.5 shows the average perimeter distances of the chosen atoms for all the caviteins in Simulation 1.
Table 3.5. Average perimeter distance \( (P_X(t)) \) values using the central carbonyl carbon atom, the central carbonyl nitrogen atom, and the terminal carbonyl carbon atom calculated from Simulation 1. The central residue refers to the middle leucine of the sequence (C9/N9 to C12/N12 for caviteins 1GS2 to 4GS2), and the terminal residue refers to the last glycine in the sequence (C16 to C19).

<table>
<thead>
<tr>
<th>Caviteins</th>
<th>Average ( P_X(t) ) using the Central Carbonyl Carbon Atom (Å)</th>
<th>Average ( P_X(t) ) using the Central Carbonyl Nitrogen Atom (Å)</th>
<th>Average ( P_X(t) ) using the terminal Carbonyl Carbon Atom (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1GS2</td>
<td>37.6</td>
<td>36.9</td>
<td>43.2</td>
</tr>
<tr>
<td>2GS2</td>
<td>40.9</td>
<td>42.3</td>
<td>42.9</td>
</tr>
<tr>
<td>3GS2</td>
<td>40.2</td>
<td>40.9</td>
<td>47.2</td>
</tr>
<tr>
<td>4GS2</td>
<td>37.7</td>
<td>37.6</td>
<td>32.3</td>
</tr>
</tbody>
</table>

The perimeter distances calculated using both the nitrogen and the carbonyl carbon of the same residue were similar (see Table 3.5 and Figures 3.10 and 3.11). For all the caviteins, the perimeter distance spread was wider for the C-terminal carbonyl carbon compared to the atoms of the central leucine (see Figures 3.10 and 3.12). This finding led to the notion that the helices were fluctuating much more at the ends of the C-termini than at the centre of the helices. In the case of the 1GS2 and the 3GS2 caviteins, the \( P_X(t) \) values were higher at the C-terminal end than in the central region of the helices; however, for the 2GS2 cavitein, the values were similar, and for the 4GS2 cavitein, the value was smaller for the perimeter distance at the C-terminus. According to the sedimentation equilibria studies (see Chapter 2.1.2.4), the 1GS2 cavitein behaved as a monomer/dimer in equilibrium, and the 3GS2 cavitein behaved as a dimer. The average perimeter distances at the C-terminal end were higher than that for the central residue for the 1GS2 and 3GS2 caviteins, which suggests that the self-aggregation of these caviteins was occurring at the C-terminal region. The fact that the perimeter distances of the central and
terminal atoms were similar for the 2GS2 cavitein suggests that the packing was optimal throughout the helices.

Figure 3.10. Normalized histogram from Simulation 1 of the sum of the interhelical distances between the carbonyl carbons of the central leucine: a) C9 for 1GS2, b) C10 for 2GS2, c) C11 for 3GS2 and d) C12 for 4GS2.
Figure 3.11. Normalized histogram from Simulation 1 of the sum of the interhelical distances between the nitrogen atoms of the central leucine: a) N9 for 1GS2, b) N10 for 2GS2, c) N11 for 3GS2 and d) N12 for 4GS2.
Figure 3.12. Normalized histogram from Simulation 1 of the sum of the interhelical distances between the carbonyl carbons of the terminal glycine: a) C16 for 1GS2, b) C17 for 2GS2, c) C18 for 3GS2 and d) C19 for 4GS2.
To relate back to the experimental data, namely, the N-H/D exchange experiments (see Chapter 2.1.2.5.3), the average perimeter distance for the caviteins were compared. The experimental data showed that the central leucine amide protons were the last to disappear. The $P^X(t)$ values of the central nitrogen atoms were in a narrow range of 38 to 41 Å for all the caviteins, which does not correlate with the N-H/D exchange data, since the protons were found to exchange much slower in the 2GS2 cavitein, and much faster in the 1GS2 and 4GS2 caviteins. Moreover, the difference between 38 and 41 Å are not that significant when taking the results from Simulations 2 and 3 into consideration.

For the 3GS2 cavitein, the spread of the perimeter distances between the C-terminal carbonyl carbons was much larger than for any of the other caviteins and reached as high as 70 Å (see Figure 3.12). The fluctuation of the perimeter distance may correlate with the strange time dependent behaviour of this protein (see Chapter 2.1.2.5.3). The 3GS2 cavitein was the only dimer in the second generation series. Thus, the $P^X(t)$ values may have some relevance to explaining the self-associative ability of the caviteins.

For Simulations 2 and 3, only the central carbonyl carbon and the terminal carbonyl carbon atoms were analyzed in terms of perimeter distance since in Simulation 1, the $P^X(t)$ values for the nitrogen atoms of the central residue were found to give very similar results to that of the central residue carbonyl carbons. The average $P^X(t)$ values of all the caviteins in Simulations 2 and 3 are shown in the Table 3.6. The average $P^X(t)$ values between the three sets of simulations were within the same range (between 30 and 40 Å), but differed slightly between each simulation. Like in Simulation 1, the perimeter distance spread was wider between the C-terminal carbonyl carbons than the spread of the central carbonyl carbon atoms for all the caviteins in Simulation 2 and Simulation 3 (see Figures 3.13 and 3.14 for Simulation 3 data). However, in some cases, the average $P^X(t)$ values for the terminal carbonyl carbon was smaller
than those of the central carbonyl carbons. Interestingly, it was in Simulation 3, that gave average $P^*(t)$ values of the terminal carbonyl carbon atoms that were lower than those of the central carbon atoms for all the caviteins, even though the starting configuration possessed splayed helices.

**Table 3.6.** Average $P^*(t)$ values of the central carbonyl carbon atom and the terminal carbonyl carbon atom calculated from Simulation 2 and Simulation 3.

<table>
<thead>
<tr>
<th>Caviteins</th>
<th>Central Carbonyl Carbon Atom (Å)</th>
<th>Terminal Carbonyl Carbon Atom (Å)</th>
<th>Central Carbonyl Carbon Atom (Å)</th>
<th>Terminal Carbonyl Carbon Atom (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Simulation 2</strong></td>
<td><strong>Simulation 3</strong></td>
<td><strong>Simulation 2</strong></td>
<td><strong>Simulation 3</strong></td>
<td></td>
</tr>
<tr>
<td>1GS2</td>
<td>34.4</td>
<td>37.3</td>
<td>37.6</td>
<td>31.9</td>
</tr>
<tr>
<td>2GS2</td>
<td>40.2</td>
<td>36.2</td>
<td>36.7</td>
<td>34.7</td>
</tr>
<tr>
<td>3GS2</td>
<td>35.2</td>
<td>38.5</td>
<td>38.4</td>
<td>35.1</td>
</tr>
<tr>
<td>4GS2</td>
<td>38.4</td>
<td>37.2</td>
<td>38.8</td>
<td>29.8</td>
</tr>
</tbody>
</table>

Like Simulation 1, there was no real correlation with the N-H/D exchange experiments. In Simulation 3, the spread of the perimeter distances between the C-terminal carbonyl carbons was larger for the 3GS2 cavitein than for any of the other caviteins like Simulation 1, but not as prominent. This larger spread was not observed for the 3GS2 cavitein in Simulation 2.

An approximate value for the interhelical distance between the adjacent helices can be estimated from the perimeter distance values. Only considering the atoms from the central Leu residues, the $P^*(t)$ values from Simulations 1, 2 and 3 ranged from 34 to 42 Å, which is the sum of the interhelical distances between the helices. This range approximates the distance between two adjacents helices in the caviteins to be between 8.5 and 10.5 Å, which is in good agreement with the interhelical distances predicted in Chapter 2.
Figure 3.13. Normalized histogram from Simulation 3 of the sum of the interhelical distances between the carbonyl carbons of the central leucine: a) C9 for 1GS2, b) C10 for 2GS2, c) C11 for 3GS2 and d) C12 for 4GS2.
3.2.4 Supercoiling

Using the TWISTER program, the caviteins were analyzed in terms of coiled coil formation. Figures 3.15 shows the plot of the coiled coil phase yield, $\Delta \omega_n$, per residue as a function of residue number for cluster 0 from Simulation 1. The coiled coil phase yield is -4 over
the length of the sequence for an ideal left-handed coiled coil. A value of zero corresponds to the absence of supercoiling of the helices. The average $\Delta \omega_n$ was approximately -5 for the 1GS2 cavitein, which indicates that this protein was supercoiling to the left. For the 2GS2, this value was +4, indicating that its helices were in a right-handed supercoil. The coiled coil phase yield per residue for the 3GS2 cavitein averaged close to zero, indicating little or no supercoiling. For the 4GS2 cavitein, the longest segment of the plot was close to -2, which implies that it was supercoiling slightly to the left.

![Coiled-coil Phase Yield Per Residue vs Residue Number for Cluster 0](image)

**Figure 3.15.** Coiled-coil phase yield ($\Delta \omega_n$) per residue number for cluster 0 of 1GS2, 2GS2, 3GS2 and 4GS2 from *Simulation 1*, calculated using the TWISTER program.

From *Simulations 2* and *3*, the 4GS2 cavitein was the only protein out of the second generation linker series that agreed with *Simulation 1*, with an average $\Delta \omega_n$ value of -2. The $\Delta \omega_n$ values of the other caviteins conflicted amongst the three set of simulations, and therefore these numbers were not used to make any conclusions.
In Chapter 2.1.2.2, it was suggested that the signals in the near UV region of the CD spectrum correlate with the supercoiling ability of the helices in a protein bundle. It was mentioned that a reversal in the maximum/minimum peaks shows opposite supercoiling effects. Experimentally, since all the maxima and minima were in the same order for all of the caviteins except the 1GS2 cavitein, but the values of $\Delta \omega_n$ from the simulations ranged from a negative number to a positive number, the near UV signals could not be correlated to the simulation results. The signals in the near UV region were more likely due to the native-like character of the caviteins.

The dispersion of the NMR spectra has been related to coiled coil formation as proteins that form supercoiled helices often display little dispersion in the amide region of their proton NMR spectra (28, 29). From Chapter 2.1.2.5.1, the N-H region for the 1GS2 cavitein ranged from 7.7 to 8.7 ppm (1.0 ppm difference). For the 2GS2 cavitein, the values ranged from 7.2 ppm to 9.2 ppm (2.0 ppm difference), for the 3GS2 cavitein, they ranged from 7.6 to 8.7 ppm (1.1 ppm difference), and for the 4GS2 cavitein, they ranged from 7.7 to 8.6 ppm (0.9 ppm difference). When solely related to the ability to supercoil, the 2GS2 showed the most dispersion, and hence can be considered to be the least supercoiled or non-supercoiled. However, the dispersion of the proton signals is also highly dependent on the degree of native-like character of the protein (see Chapter 2.1.2.5.1). Native proteins show a more dispersed NMR spectrum compared to molten globule-like structures because the side chains form specific interactions with one another, giving a more ordered structure. In molten globules, the side chains of the helices fluctuate giving rise to broad NMR spectra. Therefore, it is difficult to conclude whether a well-dispersed spectrum is a result of a native-like protein, due to lack of supercoiling, or a combination of both.
3.2.5 Tilt of Helices with Respect to the Cavitand

The tilt of the helices with respect to the cavitand template was measured as described in the methods section. The height and the range of the first residue (i.e. Ala), not including the Gly linkers, were used to determine the degree of the tilt. From the height distances of the 1GS2 cavitein (see Figure 3.16), helices B and C have sharp peaks around 3.2 Å and 3.8 Å, respectively; helices A and D have broad peaks around 1 Å or less. For the range distances (see Figure 3.17), the helices B and C have relatively sharp peaks at 3.1 and 4.2 Å, and helices A and D have sharp peaks at 7.2 and 8.1 Å. Here, the height and range distances are inversely proportional to one another; the higher height distances correlates to the lower range distances. These data suggest that helices A and D, which possess lower height distances and higher range distances are slightly more tilted than helices B and C.

In general, the range distances for all the caviteins contain sharper peaks than the height distances (except for helix D of the 2GS2 cavitein and helix A of the 4GS2 cavitein). The broad peaks are indicative of time fluctuating structures and sharp peaks are indicative of less variation during the simulation time. The fact that the height distances give rise to more broad peaks suggests that there is more fluctuation in the vertical motion than in the horizontal direction (with the vertical z-axis being through the centre of the plane defined by the four sulphur atoms on the template). This observation is reasonable since the chosen atom for measurement was close to the template and tilting most likely affected the height more than the range at this position.

For the 2GS2 cavitein (see Figures 3.18 and 3.19 for the height and the range measurements), the heights are relatively low compared to the other caviteins (see Figures 3.16, 3.18, 3.20 and 3.22), and the range distances of helices C and D of the 2GS2 are relatively high.
(see Figures 3.17, 3.19, 3.21 and 3.23), which indicate that the 2GS2 cavitein is the most tilted out of all the caviteins in the second generation linker series. This prominent tilt can also be seen in Figure 3.7 with the representative clusters. From the circular dichroism spectra in Chapter 2.1.2.2, the 2GS2 cavitein possessed the most enhanced signal in the near UV region, which could have been due to the asymmetric environment near the cavitand chromophore created by the tilt of the helices with respect to the template. In addition, the 2GS2 cavitein was found to be the most stable of all the monomeric caviteins in both the first and second generation linker series. The tilt of the helices could have increased the burial of the hydrophobic surface, thereby enhancing its stability.
Figure 3.16. 1GS2 height distances from the sulphur plane to the $\alpha$-carbon of alanine (Ala 2) for all four helices in *Simulation 1*, calculated using the helipad program. Helices B and C show sharp peaks and larger heights than for helices A and D, indicating the tilt is with helices B and C on top.

Figure 3.17. 1GS2 range distances from the sulphur plane to the $\alpha$-carbon of alanine (Ala 2) for all four helices in *Simulation 1*. All four helices show fairly sharp peaks.
Figure 3.18. 2GS2 height distances from the sulphur plane to the α-carbon of alanine (Ala 3) for all four helices in Simulation 1. Helices A and B show sharp peaks, and helices C and D show broad peaks. The overall height distance of all the helices is smaller compared to the other caviteins.

Figure 3.19. 2GS2 range distances from the sulphur plane to the α-carbon of alanine (Ala 3) for all four helices in Simulation 1. All four helices show fairly sharp peaks with the exception of helix D.
Figure 3.20. 3GS2 height distances from the sulphur plane to the $\alpha$-carbon of alanine (Ala 4) for all four helices in *Simulation 1*. Helices A, B and C show broad peaks, indicating fluctuation over time.

Figure 3.21. 3GS2 range distances from the sulphur plane to the $\alpha$-carbon of alanine (Ala 4) for all four helices in *Simulation 1*. All four helices show fairly sharp peaks.
4GS2 Height Distances for Simulation 1

Figure 3.22. 4GS2 height distances from the sulphur plane to the $\alpha$-carbon of alanine (Ala 5) for all four helices in Simulation 1. Height distances may not be significant due to the longer linkers from the template to the helical bundle.

4GS2 Range Distances for Simulation 1

Figure 3.23. 4GS2 range distances from the sulphur plane to the $\alpha$-carbon of alanine (Ala5) for all four helices in Simulation 1. All four helices show fairly sharp peaks with the exception of helix A.
The height distances for the 3GS2 cavitein (see Figure 3.20) show broad peaks, which suggest the fluctuation of the helices in a vertical motion. These broad peaks are also seen in the 3GS2 cavitein of both Simulation 2 and 3. The instability of the helices is consistent with the relative motion of the helices (see range of perimeter distance values in Figure 3.12). For the 4GS2 cavitein, since the helical bundle was the most distant from the template compared to the rest of the caviteins, the tilt distances with respect to the cavitand may have not been significant. The degrees of tilt of the helices relative to the cavitand axis in the three sets of simulations are shown in Table 3.7.

Table 3.7. Qualitative degrees of tilt of the caviteins in Simulations 1, 2 and 3. Terms are bolded if they do not agree with the other two simulations.

<table>
<thead>
<tr>
<th>Caviteins</th>
<th>Simulation 1</th>
<th>Simulation 2</th>
<th>Simulation 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1GS2</td>
<td>Very slight tilt</td>
<td>Very slight tilt</td>
<td>Very slight tilt</td>
</tr>
<tr>
<td>2GS2</td>
<td>Very tilted</td>
<td>Tilted</td>
<td>Very tilted</td>
</tr>
<tr>
<td>3GS2</td>
<td>Slight tilt</td>
<td>Slight tilt</td>
<td>Quite tilted</td>
</tr>
<tr>
<td>4GS2</td>
<td>Quite tilted</td>
<td>Displaced, but non tilted</td>
<td>Very tilted</td>
</tr>
</tbody>
</table>

Overall, the degree of tilt seems to agree amongst the three sets of simulations. The 1GS2 cavitein was found to be the least tilted and the 2GS2 cavitein was found to be the most tilted. Experimentally, the 2GS2 was found to possess the highest degree of native-like character out of the second generation cavitein series. The incline of the helices, determined from the simulation results of the 2GS2 cavitein, could have facilitated better packing between the side chains of the helices.
3.2.6 Linker Elbow Orientation

As mentioned in Section 3.1.6, the preferred linker elbow conformation of the cavitand was with the lone pairs of the sulphur pointing away from the lone pairs of the bridged oxygen as predicted by MM2 calculations (22). The conformation with the lone pairs pointing away from one another was found to be 5 kcal/mol more stable than the conformation with the lone pairs pointing toward each other (22). Thus, the initial configurations of all the cavitein systems started with this preferred elbow conformation (i.e. the methylene elbows of the peptide linkage were placed outside the rim formed by the cavitand template). The MM2 calculations were carried out only on the cavitand template without considering the stability contributed by the packing of the helices. In addition, the 5 kcal/mol value was determined in a vacuum whereas the caviteins exist in water. Therefore, the linker elbow orientation was assessed by molecular dynamics in explicit solvent. To assess the elbow orientation during the simulation time, cluster 0 for each of the caviteins was observed qualitatively. Table 3.6 summarizes the results.

Table 3.8. The linker elbow conformation of the caviteins in the second generation series. 'In' refers to the methylene elbows in the linker region sticking inside the cavitand, and 'out' refers to the methylene elbows sticking outside the cavitand.

<table>
<thead>
<tr>
<th>Caviteins</th>
<th>Simulation 1</th>
<th>Simulation 2</th>
<th>Simulation 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1GS2</td>
<td>All four out</td>
<td>All out</td>
<td>All out</td>
</tr>
<tr>
<td>2GS2</td>
<td>All out</td>
<td>All out</td>
<td>One in, three out</td>
</tr>
<tr>
<td>3GS2</td>
<td>All out</td>
<td>One in, three out</td>
<td>Two in, Two out</td>
</tr>
<tr>
<td>4GS2</td>
<td>All out</td>
<td>All out</td>
<td>All out</td>
</tr>
</tbody>
</table>

In Simulation 1, the elbow orientations remained in the preferred linker conformation for all the caviteins. In Simulation 2, the only exception was the 3GS2 cavitein in which one of the four elbows was oriented toward the centre of the cavitand. In Simulation 3, the 2GS2 cavitein
also had one of the four elbows pointing inward, but the 3GS2 cavitein had two of the four methylene elbows pointing inward. These data show that it is possible that the optimal packing of the helices may require one or more of the elbows to turn inward. It is noted that the 3GS2 cavitein had the most inwardly turned elbows. This observation may have significance to the anomalous behaviour displayed by the 3GS2 cavitein (see Chapter 2.1.2.5.3).
3.3 Chapter Summary and Conclusion

Molecular dynamics simulations were carried out on the second generation caviteins that were analyzed experimentally in Chapter 2. The helical content, conformational specificity, perimeter distance, supercoiling, tilt of the four-helix bundle, and the elbow orientation of the caviteins were assessed based on three different simulations, and compared to the experimental results. In general, the three simulations gave similar results, and agreed for the most part with the experimental data. The caviteins that were determined to be largely helical by CD spectroscopy were also found to be helical in silico. In addition, the calculated perimeter distance values gave an interhelical distance range of approximately 8 to 10 Å, which agrees with the estimated range from the experimental findings. Molecular modelling allowed us to explain some of the experimental behaviour of the caviteins. From the representative clusters obtained from the simulations, the propensity of the caviteins to self-aggregate was rationalized. The average perimeter distances were determined and gave a suggestion as to how the caviteins may aggregate. For the proteins that were found to self-aggregate experimentally (i.e. 1GS2 and 3GS2 caviteins), the average perimeter distance at the C-terminal ends was higher than that for the central residue, which suggested that the self-aggregation of these caviteins was occurring at the C-terminal region. From analyzing the linker orientation, most of the linker elbows of the peptides remained pointed outward as they were positioned in the starting configuration of the caviteins. A noticeable observation was that a couple of the linker elbows of the 3GS2 cavitein were found to turn inward, which may be associated with its anomalous behaviour.

Analyzing the tilt of the helices with respect to the cavitand template, and examining the linker orientations shed light on how the packing within the core of a protein could be improved to increase the stability and native-like character of the proteins. The helices of the 2GS2
cavitein were tilted with respect to the cavitand template. The tilt could have both increased the burial of the hydrophobic core and improved the interhelical packing, which could explain its higher stability and high degree of native-like character, respectively. Although for the most part, the simulation results for both the first generation (12) and second generation caviteins concurred with the experimental results, certain aspects were rationalized differently. For example, the tilt of the helices in the first generation 1GS0 cavitein and second generation 2GS2 cavitein may have been occurring to allow favourable packing in both cases; however, for the 1GS0 cavitein, the tilt favoured dimer formation by exposing the hydrophobic core, and for the 2GS2 cavitein, the tilt favoured monomer formation by further burying the hydrophobic core. It should also be noted that the experimental and simulation data did not always agree. For example, the supercoiling ability of the caviteins was measured, but did not seem to correlate with the experimental results. However, in the absence of an X-ray crystal or NMR structure, assumptions of the initial configuration had to be made, which made interpreting the findings at an atomic level difficult to do with confidence. The simulation results were used to complement and help elucidate the experimental findings, but were not used as concrete explanations. In addition, due to time limitations of running the molecular simulations, the acquired data may have not been sufficient to make accurate conclusions. Current computation methods are not yet chemically accurate for complex systems, but can still be extremely valuable for gaining insight into important chemical and biological processes, and for paving new paths for the design of future experiments. For example, to understand the structure of proteins and the mechanism of folding, it is critical to elucidate the molecular details of the folding pathways. Caviteins fold rapidly in aqueous solution, and therefore it is difficult to study the folding of these systems experimentally; however, computer simulations may allow observation of early or intermediate folding events, which are difficult to observe experimentally.
3.4 References


13. Special thanks to Dr. Suzana Straus (Assistant Professor) and Dr. Walter Scott (Computer Supervisor), UBC Chemistry Department, for helping with the simulations and analyses. The knowledge required to accomplish this chapter largely came from a combination of discussions with the above mentioned, material taught in Chem 507A by Walter Scott and submitted paper based on the first generation caviteins.


237


20. Molecular graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081).


CHAPTER FOUR: The Study of Different-Sized TASPs and Reversible Cavitein Systems

4.0 General Introduction

In Chapters 2 and 3, four-helix bundle motifs were studied due to their abundance in nature, synthetic availability and approximate symmetry. This chapter focuses on synthesizing and analyzing different-sized TASPs and comparing them to each other (Sections 4.1-4.2). Different-sized bundles in this thesis refer to different numbers of helices in a bundle. Subsequently, there is a section on generating reversible cavitein systems, which are thermodynamically controlled (Section 4.3). The examined caviteins so far have consisted of peptides linked onto the cavitand template via thioether bonds. In this chapter, the template assembled synthetic proteins are synthesized using another type of linkage, namely, the disulfide bonds. Due to their stability to basic conditions, and ability to form reversibly, disulfide bonds are ideal candidates to study experimentally. The importance of correct disulfide linkages in protein function was demonstrated by Anfinsen (1). When active ribonuclease was treated with urea and 2-mercaptoethanol, the protein unfolded and the disulfide bonds were reduced. When the reducing agent was removed, and the ribonuclease was exposed to oxygen, the inactive form of the protein resulted with randomly formed disulfide bonds. However, when the urea was removed and a trace of mercaptoethanol was added, the active form of ribonuclease was regenerated.

*A version of this chapter will be submitted for publication. Seo, E.S. and Sherman, J.C. The Study of Different-Sized TASPs.*
4.1 The Study of Different-Sized TASP

Different bundle sizes (with different number of helices) have been studied in coiled coil systems (2, 3). Harbury et al. found that switching the identity of the nonpolar residues in the GCN4 leucine zipper changed the number of helices in a bundle between two-, three-, and four-stranded coiled coils (2). Hodges and coworkers studied the effect of 20 different amino acid substitutions in the hydrophobic core of a coiled coil to determine the stability and number of helices in the systems (3). For the project described here, the designed sequence is kept the same, and is forced into forming different-sized helical bundles. The four-helix bundle proteins have been the main focus of this thesis thus far. Although they are more common in nature and more frequently examined in literature, five-helix (4, 5) and six-helix (6, 7) bundle proteins also exist. The de novo design and study of these larger helical bundles are however limited in literature (8, 9). In this chapter, de novo five- and six-helix bundle caviteins are designed, synthesized and examined using the ‘template assembly’ approach to investigate the structural properties of these more sizeable TASP, and to determine the effect of the various-sized templates on a specific peptide sequence.

From our group, Causton compared the structure and stability of a three-helix bundle TASP with a four-helix bundle cavitein molecule (10). Up until now, the terms caviteins and TASP have been used interchangeably. Here, the term TASP is used to refer to the three-helix bundle because the template used for the synthesis of this motif is not commonly called a cavitand. For the three-helix bundle, cyclotribenzylene (CTB) was used as the template. Furthermore, for the investigation of the different sized TASP, the benzylthiol version of the templates was used rather than the arylthiol cavitand encorporated in the caviteins discussed in
Chapters 2 and 3. The CTB template and the [4]cavitand template, both with benzylthiol functionalities are shown in Figure 4.1.

Figure 4.1. Templates used to synthesize a three helix bundle (a) cyclotribenzylene (CTB), and a four-helix bundle (b) benzylthiol [4]cavitand.

For this study, Causton synthesized the three- and four-helix TASP using the sequence S3: CCGGEE LLKKLEELLKKG (see nomenclature section), which was designed to form a four-helix bundle. These proteins were synthesized using disulfide bonds between the peptides and either the CTB template or the [4]cavitand. The resulting structures were analyzed to determine their stability, native-like characteristics and propensity to self-aggregate. Both the three-helix and four-helix TASP were found to be α-helical by CD spectroscopy. They were also found to be quite resistant to GuHCl-induced denaturation; however, there was a noticeable difference in their stabilities. The four-helix bundle was found to be more stable than the three-helix bundle TASP. As for the conformational specificity of their structures, both the three- and four-helix bundles were found to possess molten globule characteristics by 1D 1H NMR spectroscopy. Finally, the self-associative state of the four-helix bundle was found to be a monomer whereas the three-helix bundle protein was found to be in a monomer/dimer
equilibrium. The conclusion made from these analyses was that the designed peptide sequence that was intended for a four-helix bundle was indeed better for the four-helix bundle compared to the three-helix bundle in terms of stability and structural specificity. The four-helix bundle was twice as stable as the three-helix bundle; however, the stability of the former structure may have been a consequence of its larger hydrophobic core, which could have led to an increase in the hydrophobic effect. Thus, the per helix stabilities were determined and assessed. It was found that the four-helix bundle was still more stable than the three-helix bundle by one and a half fold. Since the sequence was designed to favour the formation of a four-helix bundle, the Leu residues were probably overcrowded in the three-helix bundle, which could have explained its poor stability and self-associating ability. To further study the sequence dependence on the structure of different-sized bundles, larger caviteins, specifically the five- and six- helix bundles, were synthesized and characterized. Naumann, from our group, was the first to synthesize the five- and six- helix bundle caviteins. However, these proteins were not useful in comparing with the previous three- and four-helix bundles because the peptide sequence was not the same. In addition, these larger caviteins were impure. Here, five- and six-helix bundle proteins using the same sequence, S3, as the earlier work was synthesized, completely purified, characterized and compared with the three- and four-helix bundle proteins. The synthesis of the larger TASPs required a cavitand with five and six functional groups at the rim positions. The synthesis of the [n]cavitands with benzylthiol rims is shown in Scheme 4.1.
The synthesis of the various-sized cavitands was kinetically controlled during the step of the resorcinarene formation 7 (11). The synthesis of the methyl rimmed cavitands are previously described in literature (12). Specifically, the benzylthiol rimmed form of the cavitands are described for the [5]cavitand 10a (13) and the [6]cavitand 10b (14). 2-methyl resorcinol 6 was reacted with diethoxymethane in ethanol and concentrated HCl to give a mixture of resorcinarenes 7 with \( n = 4-7 \). The reaction time determined the predominant size of the resorcinarene. Next, this mixture was reacted with bromochloromethane and potassium carbonate in DMA under a nitrogen atmosphere at 60 °C to give a mixture of the various-sized cavitands 8. These cavitands were then separated based on their different solubilities. [5]Cavitand 8a was reacted with NBS and a catalytic amount of AIBN to give the benzylbromide [5]cavitand 9a. The final reaction in this synthesis consisted of reacting 9a with thiourea in DMF to give the benzylthiol [5]cavitand 10a. Similar steps were carried out from the [6]cavitand 8b in order to obtain the benzylthiol [6]cavitand 10b.

It should be noted that the symmetric conformation of the cavitands was not maintained for the larger [6]cavitand template since more aromatic rings caused steric crowding, which resulted in a distorted structure of lower symmetry. In fact, simple molecular mechanics calculations (MM2 force field) showed that the symmetric cone conformation that was observed for the [4]cavitand and [5]cavitand was not observed for the [6]cavitand as shown in Figure 4.2 (12).
Figure 4.2. Energy minimized structures of [n]cavitands obtained using MM2 force field (figure used with permission from reference (12)): (a) [4]cavitand, (b) [5]cavitand, (c) [6]cavitand.

The [4]cavitand and [5]cavitand contained a $C_{4v}$ and $C_{5v}$ axis of symmetry, respectively, but the [6]cavitand did not contain a $C_{6v}$ axis of symmetry. To confirm the cavitand structures, X-ray crystal structures were obtained (12). The crystal structure of the [6]cavitand was very similar to the $C_{2v}$ conformation predicted by molecular mechanics calculations (12).
4.1.1 Goals

Five- and six-helix bundle caviteins were synthesized and characterized using the peptide sequence of the original three- and four-helix bundle TASPs. The four-helix bundle cavitein was also synthesized and analyzed using techniques that were not previously carried out. The purpose of this project was to determine the effect of various-sized templates (that can accommodate a different number of helices) on the overall protein structure using a specific sequence designed for a four-helix bundle. In this thesis, different-sized bundles refer to different number of helices in a bundle. In previous work, a sequence designed for a four-helix bundle adopted a more-defined tetrameric bundle compared to the trimeric bundle. Here, the peptide design was further tested by using the same sequence in the larger bundles. Is this sequence design generic in the sense that it would work for any bundle size larger than a three-helix bundle? Would this sequence produce the most native-like four-helix bundle over the larger pentameric and hexameric bundles? Would the larger TASPs result in more stable structures due to their larger hydrophobic cores? What is the stability per helix in each helical bundle? Is the peptide sequence alone sufficient to dictate a well-defined protein structure or does the sequence display different native-like characteristics depending on the size of the bundle? What kind of effect would the extra asymmetric element of the [6]cavitand have on the structure of the hexameric protein bundle? These questions are addressed in this chapter. The five- and six-helix bundle TASPs were synthesized and analyzed in terms of helicity by CD spectroscopy, stability by chemical denaturation, tendency to self-aggregate by sedimentation equilibria studies, and degree of native-like characteristics by near UV CD spectroscopy, $^1$H NMR spectroscopy and ANS binding studies. The experimental data collected for the five- and six-helix bundles are compared to each other, and to the previous data on the three- and four-helix bundle proteins. It is expected
that the four-helix bundle would be more stable per helix and more native-like than the larger five- and six-helix bundles since the peptide was designed to favour the formation of the four-helix bundle. If this is true, then the de novo design is fairly successful in obtaining a desired protein from scratch.

To further test the efficiency of the design, a peptide sequence that favours the formation of a five-helix bundle was designed in an attempt to determine whether the sequence would result in the most well-defined five-helix bundle as the design intended. Overall, this project examines the structural properties of different-sized bundle TASP$s$ when a sequence that is anticipated to form a well-defined bundle of a specific size, $n$, is linked on to a cavitand template that forces one less, $(n-1)$ or one more $(n+1)$ helix in a bundle.

4.1.2 Nomenclature

The nomenclature of the peptide sequences in this chapter is summarized in Table 4.1. To refer to the template assembled synthetic protein (TASP) or cavitein, the number in brackets follows the sequence name to denote the size of the bundle. For example, S3-[4] is a cavitein that is composed of four S3 peptides and the benzylthiol [4]cavitand template. The S3-[5] cavitein is a TASP that consists of five S3 peptides and the benzylthiol [5]cavitand template. The S3-(3) TASP is composed of three S3 peptides and the CTB template. The round brackets signify the CTB template, which is not commonly called a cavitand. The S4-[4] cavitein consists of four S4 peptides on the benzylthiol [4]cavitand template. To refer to just the templates, the bracketed number is added to the beginning of the term cavitand for $n = 4-6$. For example, a cavitand template with four benzylthiol functional groups is called a [4]cavitand, and a cavitand with five
reactive sites is called [5]cavitand, and so on. All templates are assumed to be benzylthiol
cavitands unless otherwise specified. A protein which consists of an arylthiol cavitand template
and four S6 peptides is referred to as S6-aryl[4] cavitein, and a protein which consists of four S3
peptides and the phosphate footed cavitand template is referred to as S3-phos[4] cavitein.

Table 4.1. Names used for the peptide sequences (to refer to the template assembled
synthetic proteins, the term TASP or cavitein follows the sequence name).

<table>
<thead>
<tr>
<th>Sequence Name</th>
<th>Peptide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>S3</td>
<td>CGGGEELKKLEELKKG</td>
</tr>
<tr>
<td>S4</td>
<td>CGGGEEAAKKEAEEAAKKG</td>
</tr>
<tr>
<td>S5</td>
<td>CGGGEELKKLEELKKG</td>
</tr>
<tr>
<td>S6</td>
<td>GGGEELKKLEELKKG</td>
</tr>
</tbody>
</table>
4.2 Results and Discussion

4.2.1 Synthesis of the Larger TASP

The design and synthesis of the larger S3-[5] and S3-[6] TASP are described in this section.

4.2.1.1 Template Choice and Synthesis

The cavitand templates that were chosen for the synthesis of the larger template assembled synthetic proteins consisted of benzylthiol rims and hydrogen feet since this form of cavitand template was synthetically available in different sizes, and was capable of forming disulfide bonds with the peptides.

4.2.1.2 Sequence Choice and Synthesis

The sequence that was previously used in the study of the three-helix bundle and four-helix bundle TASP (10) was used here in order to investigate this same sequence on the larger [5]cavitand and [6]cavitand templates. The peptide sequence was based on the same design principles used in Chapter 2.1.1.2 and was essentially the same sequence as the first generation peptide sequence with three Gly, 3GS0: GGGEEELLKKEELLKKG, except with the addition of
a Cys residue at the N-terminus, before the Gly linkers. The Cys was incorporated at the N-terminus of the peptide in order to form a disulfide bond between the template and the peptides. Three Gly residues were chosen to be in the linker region, as this number appeared to give the most well-defined structure out of the first generation studies. This peptide is called S3 and its sequence is as follows: CGGGEELLKKLEELLKKG from N- to C-terminus. The nomenclature is described in section 4.1.2.

For the formation of the disulfide bonds between the benzylthiol cavitand and the peptides to occur exclusively, the peptides were first activated. Without this activation, the peptides were found to dimerize, thereby minimizing the assembly of the desired TASP. The peptide was activated by reacting it with 2,2'-dipyridyl disulfide (DPDS) in ethanol as shown in Scheme 4.2.

**Scheme 4.2.** Activation of the peptide by reaction with 2,2'-dipyridyl disulfide (DPDS).
4.2.1.3 Synthesis of the Larger Caviteins

The syntheses of the larger caviteins consisted of reacting the appropriately-sized benzylthiol cavitand with an excess amount of activated S3 peptide. The synthesis of the S3-[5] and S3-[6] caviteins is shown in Scheme 4.3.

![Scheme 4.3. Synthesis of the S3-[5] and S3-[6] caviteins.](image)

For the synthesis of the S3-[5] cavitein, 7 equiv. of activated peptide was reacted with the [5]cavitand, and for the synthesis of the S3-[6] cavitein, 8 equiv. of activated peptide was reacted with the [6]cavitand in the presence of DIPEA in DMF for 5 hours. After the completion of the reaction, the solvent was removed \textit{in vacuo}, and purified by reversed-phase HPLC. Two products were obtained for each cavitein reaction. For the synthesis of the S3-[5] cavitein, these products were the desired five-helix bundle, and a side-product that consisted of the [5]cavitand covalently linked onto four peptides and one S-pyridyl group (see Figure 4.3). The ratio of the products was 2:1 for the desired product to the side-product. Similarly, for the synthesis of the
S3-[6] cavitein, the obtained products were the desired six-helix bundle, and the [6]cavitand with five helices and one S-pyridyl group. However, the ratio of the products was 1:2 (desired:side).

![Diagram of cavitein structure]

**Figure 4.3.** Byproduct of the S3-[5] cavitein synthesis.

After the first purification, the caviteins were assessed for purity by analytical reversed-phase HPLC, but they were found to be impure as they still contained a pre-peak and a post-peak. Therefore, a second purification was performed, which resulted in a pure product by observation of a single peak by reversed-phase chromatography. The identities of these larger caviteins were confirmed by MALDI spectrometry.
4.2.2 Characterization of the Different-Sized TASP

The larger S3-[5] and S3-[6] caviteins were characterized using circular dichroism (CD) spectroscopy, chemically induced denaturation, analytical ultracentrifugation, $^1$H NMR spectroscopy, and ANS binding studies. The S3-[4] cavitein that was previously studied (10) was also investigated here using techniques that were not employed in the original work, such as N-H/D exchange and ANS binding. The data of the various-sized template assembled synthetic proteins were analyzed and compared to one another to determine if the sequence itself was sufficient to dictate their native-like character, or if the cavitand template, which dictates the number of helices in the bundle, had a larger effect on the structural specificity of the caviteins. The effect of the template size on the structural properties of the proteins using the same peptide sequence was investigated.

4.2.2.1 Far UV Circular Dichroism (CD) Spectra

Circular dichroism (CD) spectroscopy is useful in determining the secondary structure of the proteins (15). The amide chromophore of a peptide bond mainly influences the far UV region of the CD spectra. An $\alpha$-helix is characterized by two minima at 208 and 222 nm, and one maximum signal at 195 nm. The CD spectra of the S3-[5] and S3-[6] caviteins are shown in Figure 4.4. The acquired spectra of both the low (4 $\mu$M) and high (40 $\mu$M) concentrations overlapped, and therefore only the high concentration data are shown.
Both the S3-[5] and S3-[6] caviteins possess two minima bands and one maximum band, which are characteristic of α-helical structures. From Chapter 2.1.2.1, it was explained that aromatic choromores have a considerable effect on the far UV region of the CD spectrum. Here, the two different-sized TASPs have two different cavitand templates, and the contribution of each of these chromophores in the far UV is unknown. Since the fraction of α-helix is dependent on the $[\theta]_{222}$ value of the far UV region (see Chapter 2.1.2.1), this value was not calculated for these larger caviteins. The only conclusion that is made from the far UV region of the spectrum is that the S3-[5] and S3-[6] caviteins are similar in helicity. The S3-(3) and S3-[4] TASPs were also found to be helical from the earlier work (10).
4.2.2.2 Near UV Circular Dichroism (CD) Spectra

The near UV circular dichroism spectra have been used to obtain information about the native-like character of the caviteins. The aromatic absorption of the cavitand in the near UV region can be observed if there are any non-averaged structural features near this chromophore. For example, a native-like structure contains more intense signals in this region because of its specific side chain interactions. On the contrary, a molten globule protein lacks the intense absorptions in this region due to their fluctuating structures, which average out the signals. Figure 4.5 shows the near UV CD spectra of the S3-[5] and S3-[6] caviteins.

![Near UV CD Spectra](image)

**Figure 4.5.** Near UV spectra of the S3-[5] and S3-[6] caviteins. Each sample contains 40 μM protein in 50 mM phosphate buffer, pH 7.0 at 25 °C.
From Figure 4.5, the S3-[5] and S3-[6] caviteins appear to possess molten globule-like structures from the lack of absorption in the near-UV region. Both the three-helix and four-helix TASPs have also demonstrated similar behaviour (10). However, the absence of a strong near UV signal does not necessary indicate a molten globule state due to the relatively longer linker length of three Gly residues from the cavitand chromophore. It should also be noted that disulfide bonds give rise to broad weak signals in the near UV region.

4.2.2.3 Effect of Guanidine Hydrochloride

The larger S3-[5] and S3-[6] caviteins were analyzed by GuHCl-induced denaturation to determine their free energies of folding in aqueous solution. Both caviteins displayed concentration independent curves, and therefore only the high concentrations curves are shown in Figure 4.6.
Figure 4.6. Guanidine hydrochloride-induced denaturation curves for the S3-[5] and S3-[6] caviteins, both 40 μM concentrations in pH 7.0 phosphate buffer at 25 °C. Error bars calculated as explained in the experimental section have been omitted for clarity. In addition, although the samples were measured at 0.25 M increments of [GuHCl], only points for every 1 M increment of GuHCl are shown here.

The S3-[5] and S3-[6] caviteins appear to possess similar stabilities as the approximate [GuHCl]$_{1/2}$ values are comparable. However, using the [GuHCl]$_{1/2}$ value only gives an estimation of the stabilities. The ‘nonlinear least-squares method’ is a more accurate way of determining the stabilities of the proteins (described in Chapter 2.1.2.3), which was also used here. Table 4.2 shows the calculated free energies of folding for the TASP in aqueous buffer.
Table 4.2. Guanidine hydrochloride-induced denaturation data of the different-sized helical bundle TASPs. The fraction folded is given by normalizing the [θ]_{222} value at the different concentrations of GuHCl with the [θ]_{222} value of the completely folded, non-denatured state. The data for the S3-(3) and S3-[4] TASPs are taken from earlier work (10) and are shown here for comparison.

<table>
<thead>
<tr>
<th>TASP</th>
<th>[GuHCl]_{1/2} (M)</th>
<th>m (kcal/molM)</th>
<th>ΔG^{o}_{H_2O} (kcal/mol)</th>
<th>Per Helix ΔG^{o}_{H_2O} (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S3-(3)</td>
<td>4.4 ± 0.1</td>
<td>1.0 ± 0.1</td>
<td>-4.7 ± 0.5</td>
<td>-1.6 ± 0.2</td>
</tr>
<tr>
<td>S3-[4]</td>
<td>5.7 ± 0.1</td>
<td>1.7 ± 0.1</td>
<td>-9.4 ± 0.7</td>
<td>-2.4 ± 0.2</td>
</tr>
<tr>
<td>S3-[5]</td>
<td>5.5 ± 0.1</td>
<td>1.06 ± 0.06</td>
<td>-5.4 ± 0.6</td>
<td>-1.1 ± 0.1</td>
</tr>
<tr>
<td>S3-[6]</td>
<td>5.5 ± 0.1</td>
<td>0.88 ± 0.04</td>
<td>-4.6 ± 0.5</td>
<td>-0.77 ± 0.08</td>
</tr>
</tbody>
</table>

From the ΔG^{o}_{H_2O} values, it can be seen that the S3-[4] cavitein is the most stable TASP. Since the size of the helical bundle has an effect on its stability, it is also important to look at the ΔG^{o}_{H_2O} per helix. This value for the S3-[4] cavitein approximately doubles that of the S3-[5] cavitein, and more than doubles that of the S3-[6] cavitein. Therefore, the TASP with four helices is the most stable, which agrees with the peptide sequence originally designed to form a four-helix bundle.

The m value gives an indication of the cooperativity of the protein. High cooperativity has been associated with a high degree of native-like character (16). The m values for the caviteins are comparable to those found for native proteins (17). The S3-[4] cavitein exhibited the largest m value, which implies that it possesses the most native-like characteristics compared to the other sized TASPs using the same sequence. Again, this result is not surprising given that the sequence was designed for a four-helix bundle.
4.2.2.4 Oligomeric State

Sedimentation equilibria studies have been carried out to determine molar masses, association constants, and stoichiometries of protein systems using an analytical ultracentrifuge (18). Here, this method was used to estimate the molecular weight of the larger template assembled synthetic proteins in order to determine their oligomeric states. Both the S3-[5] and S3-[6] caviteins were studied at three different concentrations (20, 40 and 60 µM) and three different rotor speeds (27000, 35000 and 40000 rpm in pH 7.0, 50 mM phosphate buffer at 20 °C). The experimental section contains the fits of the experimentally collected data. Initially, both caviteins were fit to a single non-interacting species. The S3-[5] cavitein was found to possess a molecular weight of a dimer, and therefore fits to a monomer-dimer in equilibrium were carried out. Table 4.3 summarizes the results obtained from the sedimentation equilibria studies.

Table 4.3. Molecular weight estimations by sedimentation equilibria studies for the S3-[5] and S3-[6] caviteins at 20 °C in 50 mM phosphate buffer, pH 7.0 when fit to a single, ideal species. A monomer-dimer fit was also carried out for the the S3-[5] cavitein to determine the association constant, $K_a$. Solvent density was estimated to be 1.0 g/mL.

<table>
<thead>
<tr>
<th>Caviteins</th>
<th>Calculated Monomer Molecular Weight (Da)</th>
<th>Experimentally Estimated Molecular Weight (Da)</th>
<th>Partial Specific Volume (mL/g)</th>
<th>Predominant Species</th>
<th>Association Constant, $K_{a,2}$ (Abs.)</th>
<th>$K_{a,2}$ (M$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S3-[5]</td>
<td>10820</td>
<td>20900 ± 800</td>
<td>0.78</td>
<td>Dimer</td>
<td>187 ± 20</td>
<td>$2.4 \times 10^6$ ± 0.3</td>
</tr>
<tr>
<td>S3-[6]</td>
<td>12980</td>
<td>13900 ± 700</td>
<td>0.77</td>
<td>Monomer</td>
<td>------</td>
<td>------</td>
</tr>
</tbody>
</table>

The S3-[6] cavitein was found to exist as a monomer whereas the S3-[5] cavitein was found to exist as a dimer. The association constant, $K_{a,2}$, for the five-helix bundle was
determined to be $2.4 \times 10^6 \pm 0.3 \text{ M}^{-1}$ (when $\varepsilon_{270} = 21000 \text{ cm}^{-1} \text{ M}^{-1}$) using the equation as explained in the experimental section. From the previous study of the three- and four-helix bundles, the S3-[4] cavitein was found to be a monomer by sedimentation equilibria studies, whereas the S3-(3) TASP was found to be a monomer-dimer in equilibrium (10). Again, since the TASP s were synthesized using a sequence that was designed for a four-helix bundle, it is not unusual that the S3-[4] cavitein existed as a monomer. The S3-(3) TASP may have had an overcrowded hydrophobic core due to its smaller template, and in contrast, the S3-[5] cavitein may have had a loosely packed hydrophobic core; both instances could have led to self-aggregation. As for the S3-[6] cavitein, although its template is larger than that of the S3-[5] cavitein, it also contains an element of asymmetry not present in the other cavitand sizes. This additional asymmetric element of the [6]cavitand could have reduced the interhelical gap that caused self-aggregation in the five-helix bundle, thereby giving rise to a monomeric structure. Figure 4.7 shows how the hexameric helical bundle of the cavitein could have been arranged near the cavitand template, as compared to the pentameric bundle.

Figure 4.7. Hypothetical arrangement of the helices for the (a) six-helix bundle cavitein and the (b) five-helix bundle cavitein.
4.2.2.5  $^1$H Nuclear Magnetic Resonance (NMR) Spectroscopy

$^1$H NMR spectroscopy was used to analyze the various-sized TASPks as described in Chapter 2.1.2.5. The spectra were acquired by Okon from the laboratory of McIntosh, UBC Department of Biochemistry.

4.2.2.5.1 One-Dimensional (1D) $^1$H NMR Spectroscopy

As mentioned in Chapter 2.1.2.5.1, $^1$H NMR spectroscopy is a useful tool in determining the degree of a protein’s native-like character. A native-like protein exhibits sharp, well-dispersed signals, whereas a molten globule structure exhibits more broad, less-dispersed signals. Figure 4.8 shows the expanded amide region of the spectra of the S3-[5] and S3-[6] caviteins, and Figure 4.9 shows the expanded aliphatic region of the spectra.

![Figure 4.8](image-url)

**Figure 4.8.** Expanded amide regions of the 1D $^1$H NMR spectra of 0.57 mM S3-[5] and 0.42 mM S3-[6] caviteins, each in 50mM phosphate buffer, pH 7.0 and 10 % D$_2$O at 25 °C.
The 1D $^1$H NMR spectra of both the S3-[5] and S3-[6] caviteins are broad. The spectra of the S3-[5] and S3-[6] caviteins are similar to one another in terms of sharpness and dispersion. It should also be noted that these spectra are similar to the 1D $^1$H NMR spectrum of the S3-[4] cavitein presented in the earlier studies (10). The broad and poorly dispersed signals of the different-sized TASP imply that these structures possess molten globule characteristics, which is consistent with the original study showing that employing the benzyl cavitand template in the synthesis of a TASP produces structures with less native-like properties compared to caviteins synthesized using the arylthiol cavitand (19).
4.2.2.5.2 Hydrogen/Deuterium Exchange

Hydrogen/deuterium exchange experiments are a useful way to probe the dynamic behavior of the proteins as described in Chapter 2.1.2.5.3. Since amide protons of a protein backbone can exchange with the solvent, this technique can be used to assess the fluctuating ability of the proteins. For example, a molten globule structure is more likely to exchange all of its amide protons faster than a native-like protein that contains specific side chain interactions. Figure 4.10 to 4.12 show the $^1$H NMR hydrogen/deuterium exchange data of the S3-[4], S3-[5] and S3-[6] caviteins, respectively.
**Figure 4.10.** Stack plot of the 1D $^1$H NMR N-H/D exchange spectra for 0.3 mM S3-[4] cavitein in 50 mM deuterated acetate buffer, pD 5.02 on a 500 MHz NMR Varian Unity spectrometer at 20 °C. The reference spectrum was taken in 50 mM acetate buffer, pH 4.62 and 10 % D$_2$O. The * symbols represent the non-exchangeable cavitand signals.
**Figure 4.11.** Stack plot of the 1D $^1$H NMR N-H/D exchange spectra for 0.3 mM S3-[5] cavitein in 50 mM deuterated acetate buffer, pD 5.02 on a 500 MHz NMR Varian Unity spectrometer at 20 °C. The reference spectrum was taken in 50 mM acetate buffer, pH 4.62 and 10 % D$_2$O. The * symbol represents the non-exchangeable cavitand signal.

**Figure 4.12.** Stack plot of the 1D $^1$H NMR N-H/D exchange spectra for 0.3 mM S3-[6] cavitein in 50 mM deuterated acetate buffer, pD 5.02 on a 500 MHz NMR Varian Unity spectrometer at 20 °C. The reference spectrum was taken in 50 mM acetate buffer, pH 4.62 and 10 % D$_2$O. The * symbols represent the non-exchangeable cavitand signals.
Here, the protection factors were estimated using the equation in the experimental section and are presented in Table 4.4.

**Table 4.4.** N-H/D exchange data on the set of different-sized helical bundle caviteins in 50 mM deuterated acetate buffer, pD 5.02 at 20 °C.

<table>
<thead>
<tr>
<th>TASP</th>
<th>Amide Proton Chemical Shift&lt;sup&gt;b&lt;/sup&gt; (ppm)</th>
<th>First-Order Rate Constant (h&lt;sup&gt;-1&lt;/sup&gt;)</th>
<th>Protection Factor&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>S3-[4]</td>
<td>8.4</td>
<td>3.5 ± 0.5</td>
<td>60 ± 5</td>
</tr>
<tr>
<td>S3-[5]&lt;sup&gt;c&lt;/sup&gt;</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>S3-[6]</td>
<td>8.3</td>
<td>15.4 ± 0.9</td>
<td>14.1 ± 0.7</td>
</tr>
</tbody>
</table>

<sup>a</sup> the protection factors are based on the half-life of an unprotected proton at pD 5.02 at 20 °C to be 3.18 E-3 h.

<sup>b</sup> the chemical shift values correspond to the last set of amide protons to disappear during the exchange process.

<sup>c</sup> the protection factor of the S3-[5] cavitein could not be calculated since the amide proton signals disappeared almost instantaneously.

All the amide protons belonging to the different-sized caviteins exchanged with the solvent within an hour. For the S3-[4] cavitein, the protection factor was calculated to be about 60. The protection factor for the S3-[6] cavitein was lower at 14.1, and even lower for the S3-[5] cavitein. Typically, molten globule structures have protection factors that range from 10 to 10<sup>3</sup> and native-like proteins have protection factors that are larger (20). The results show that the various-sized TASP<sub>1</sub>s all have molten globule characteristics, which are consistent with the 1D <sup>1</sup>H NMR spectra. Nevertheless, the S3-[4] four-helix bundle was found to contain the most protected amide protons compared to the larger S3-[5] and S3-[6] caviteins.
ANS binding studies is also used here (see Chapter 2.1.2.6) to determine the extent of native-like character of the caviteins. ANS fluoresces at approximately 480 nm in methanol and 470 nm ethanol, or when bound to a protein with an exposed hydrophobic pocket (21). Figure 4.13 shows the ANS binding studies of the different-sized TASP: S3-[4], S3-[5] and S3-[6], as well as the emission spectra of 100 % methanol and 95 % ethanol.

![Figure 4.13. Fluorescence emission spectra of 50 μM of S3-[4], S3-[5], S3-[6] TASP, 100 % methanol and 95 % ethanol in the presence of 2 μM ANS in pH 7.0, 50 mM phosphate buffer at 25 °C.](image)

The S3-[4] cavitein did not show any significant amount of ANS binding, just like the other four-helix bundles studied in Chapter 2.1.2.6. However, the S3-[5] and S3-[6] caviteins
were found to bind a substantial amount of ANS. Consistent with the other experimental results, the four-helix S3-[4] cavitein displays the highest conformational specificity compared to the S3-[5] and S3-[6] caviteins. These larger TASPs are the first caviteins that have shown a considerable amount of fluorescence by ANS binding studies. A slight blue shift can be observed for the S3-[5] and S3-[6] helix bundles, which is an indication of decreased accessibility of the ANS molecules to the bulk solvent. Although binding of ANS is characteristic of a molten globule state, native proteins such as apomyoglobin have also been found to bind this dye (21). In the case of the S3-[5] and S3-[6] caviteins, other experimental results (i.e. $^1$H NMR spectroscopy and N-H/D exchange data) have shown that they exist as molten globule structures. Therefore, it is safe to conclude that these caviteins exist as molten globules. The question still remains if ANS binding occurred due to poorly packed helices or due to a cavity created within the helical bundle. A possibility is that a larger cavity could have caused the helices to pack poorly. That is, a larger cavity and poorly packed helices may be interrelated and indistinguishable. From Chapters 2 and 3, it was determined that the interhelical distance between two adjacent helices containing nonpolar Leu residues was generally between 8 and 10 Å. However, the sides of the cavitands are smaller than this range. The dimensions of the [4]cavitand, [5]cavitand and [6]cavitand are defined in Figure 4.14.
Although the longest distance of the cavitand becomes larger with an additional aromatic ring, the distance between the adjacent functional groups does not increase. A larger cavitand may not be able to make up for an extra helix that has an effective diameter that is comparable or larger than the size of the cavitand. Therefore, the adjacent helices could have been overpacked, but could have allowed a cavity to form (see Figure 4.15), disrupting the packing of the helices.
Figure 4.15. Possible cavity formed within the hydrophobic core of the five-helix bundle.

It is unknown exactly how the helices pack together, especially for the six-helix bundle cavitein, which consists of the asymmetric [6]cavitand. It is possible that the six-helix bundle cavitein could have existed as two bundles of three-helices or a central four-helix bundle with a straggling helix on each side.

4.2.3 Study of Peptide Sequence Designed for a Five-Helix Bundle

From the study above, it was shown that the sequence designed for a four-helix bundle was indeed superior in forming a well-defined tetrameric bundle in terms of stability and specificity. The larger five- and six-helix bundles could have been more stable due to the increased hydrophobic effect within the core of the bundle. However, the per helix stability was
the highest for the four-helix bundle, and the increased size of the hydrophobic core could not compensate for the lack of specific interactions between the helices of the larger bundles. The importance of sequence design in obtaining a well-defined four-helix bundle protein was demonstrated. To further test our knowledge of peptide design, a sequence that was intended to form a five-helix bundle was designed and synthesized. An abundant amount of research has been put towards the study of de novo four-helix bundles (22-24). Although not as common, investigation of de novo six-helix bundles also exist (8, 25). Yang and coworkers showed the ability of two different amphiphilic octadecapeptides, each containing eight Leu residues, to form a helical hexamer in solution (8). Even less frequent is the study of de novo five-helix bundles (9). In this example, the peptide sequence was not directly designed to form a five-helix bundle; instead, a peptide designed to form a four-helix bundle was modified to promote the formation of a five-helix bundle by incorporating homocysteine residues capable of disulfide bonding at the helical interfaces (9). Both these examples, however, did not investigate the effect of different bundle sizes. Here, a sequence that is intended to form a five-helix bundle was designed from scratch and used in an attempt to synthesize and characterize the different-sized caviteins.

The questions that may be answered include: would the five-helix bundle show the most stable and native-like characteristics as the sequence intends, or would the four-helix bundle form a more-well defined structure as before? Could the hexameric bundle display more native-like character due to the asymmetry of the [6]cavitand, which may be able to adapt to the sequence with an increased number of nonpolar residues? To respond to these questions, the S5 peptide was synthesized in an effort to generate and characterize the various-sized bundle caviteins.
4.2.3.1  Design and Synthesis of the S5 Peptide Sequence

The S5 peptide sequence was designed using a minimalist approach to preferentially form a five-helix bundle. In general, a peptide sequence that forms \textit{de novo} four-helix bundles contains five to six nonpolar residues. \textit{De novo} hexameric bundles generally contain peptides with eight nonpolar residues (8). Therefore it made sense to design a five-helix bundle with seven nonpolar residues. This newly designed peptide S5 has the following sequence: CGGGEELLKLEELLKLG. Figure 4.16 shows the helical wheel diagram of a pentameric bundle using this sequence. The amphiphilic nature of this peptide allows the nonpolar residues to get buried in the hydrophobic core, and the polar residues to get exposed to the solvent. The sequence also contains potential sites for intrahelical salt bridges.
Figure 4.16. Helical wheel diagram of a five-helix bundle using the peptide sequence, S5 = CGGGEELLKKLLEELLKKLG (excluding the linker residues). Helices are oriented in a parallel fashion. Viewer is looking down the helical axes from C- to N-termini.

The S5 peptide contains three Gly linkers and a Cys at the N-terminal end like the previous S3 peptide sequence. The Cys was incorporated so that the peptides could link on the cavitand template via disulfide bonds. The S5 peptide was also activated like the S3 peptide in order to inhibit the formation of disulfide bonds between the peptides.
4.2.3.2 Synthesis of the Various-Sized TASP\textsubscript{s} using the S5 Sequence

When the activated S5 peptide was reacted with the benzylthiol [4]cavitand under the same basic conditions used in the synthesis of the other caviteins, no four-helix bundle was obtained. Furthermore, the recovered peptide was reduced in yield even though no product was observed. The same phenomenon occurred when the activated S5 peptide was reacted with the benzylthiol [5]cavitand. Since the activated S5 peptide failed to produce any cavitein, the sequence was slightly modified by removing the Cys residue, and adding a chloroacetyl linker to the N-terminus. This peptide was named S6 and has the following sequence: CICH\textsubscript{2}CO-NH-[GGGEELLKKLLEELLLKLG]. This peptide was also unsuccessful in obtaining any cavitein product when reacted with the benzylthiol form of the [4]cavitand or the [5]cavitand. However, when the S6 peptide was reacted with the arylthiol form of the [4]cavitand, a four-helix bundle of the expected mass was obtained. Unfortunately, only the [4]cavitand of the arylthiol template form is synthetically available, whereas the [5]cavitand and [6]cavitand are not. For the future, a different sequence can be designed for a five-helix bundle, still with seven Leu residues, but in a different order.
4.3 Reversible Cavitein Systems

A growing topic of interest has been in dynamic covalent chemistry (DCC) (26). DCC has to do with chemical reactions carried out reversibly under equilibrium conditions. Because the formation of the products is controlled thermodynamically, the resultant product distribution only depends on the relative stabilities of the final products, unlike kinetically controlled reactions, where the relative proportions of the products are determined by the free energy differences between the transition states. Dynamic covalent chemistry offers advantages like the potential of synthesizing molecules with complex topologies that may be difficult to synthesize using traditional approaches, and the ability to re-modify the product distribution by changing the reaction conditions (i.e. temperature, pH, concentrations). The idea of DCC is used to create dynamic combinatorial libraries (DCLs) (27). DCLs differ from other conventional combinatorial libraries in that the individual members of the library are constantly interconverting, so that the composition of the library is under thermodynamic control. In an example, the addition of a template increased the concentration of one of the receptors by selectively binding to it at the expense of the other members in the DCL (28). Initial studies on DCLs were based on tranesterification; however, these reactions required harsh conditions (27). Exchange of Schiff bases (29) and hydrazone exchange (30) have also been demonstrated as reversible reactions under thermodynamic control; however, the former can only be turned off by hydrogenation, and the latter requires an acidic environment to switch on the exchange. The reaction of thiols with disulfides has been shown to be highly specific and reversible and has been given the name of thiol-disulfide exchange reactions or equilibra. Kim and Alber employed the thiol-disulfide interchange reactions to determine the most stable orientation of a four-helix
bundle protein (2). For the cavitein systems, it would be most reasonable to employ the thiol-disulfide exchange since the cavitand templates already possess thiol groups, and the peptides can be synthesized with a terminal Cys residue. Moreover, extensive study on thiols and disulfides have shown that disulfides are stable toward many functional groups, disulfide exchange takes place under mild conditions in the presence of a catalytic amount of thiol, and that thiol-disulfide interchange is negligible under acidic conditions. In this section, the concept of DCC and DCL was applied to the template assembled synthetic proteins using thiol-disulfide exchange reactions. The idea was to react the cavitand template with two different peptide sequences in order to obtain the most stable cavitein, whether it consists of one peptide sequence or a mix of the two different sequences. Firstly, the stability of two benzylthiol [4]cavitand-based four-helix bundles were determined. One cavitein was synthesized using solely the S3 peptide sequence, while the other cavitein was synthesized using solely the S4 peptide sequence. The S4 peptide is analogous to the S3 peptide in sequence order (see nomenclature section). The only difference is that all the nonpolar residues in the S3 peptide are Leu, whereas these nonpolar positions are replaced with Ala in the S4 peptide. The free energies of folding were determined as before using GuHCl-induced denaturation curves (see Figure 4.17).
Figure 4.17. Guanidine hydrochloride-induced denaturation curves of the S3-[4] and S4-[4] caviteins.

The stability data of the S3-[4] cavitein is shown in Table 4.2. The stability of the S4-[4] cavitein could not be measured using the method previously mentioned due to the lack of the pre- and post-transition baselines. However, looking at the denaturation curves, it is obvious that the S3-[4] cavitein is significantly more stable than the S4-[4] cavitein, since the former does not start to melt until a much higher concentration of GuHCl as compared to the latter. Therefore, the two different peptides S3 and S4 were reacted with the [4]cavitand template under reversible conditions to determine if the S3-[4] cavitein would be the most dominant species, since it was found to be the most stable. Furthermore, by employing the dynamic combinatorial library approach, a product that is otherwise difficult to synthesize using conventional methods may be readily obtained. For example, a possible cavitein product that may form under reversible conditions is a four-helix bundle consisting of two S3 peptide strands and two S4 peptide strands. Other possibilities include three S3 peptides and one S4 peptide, or vice-versa.
In order to set up a reversible system, the first step consists of synthesizing a cavitand template that could be solubilized in water, since the stability of the caviteins largely depends on the hydrophobic effect. The benzylthiol cavitand was reacted with excess 1 M sodium methoxide in THF under nitrogen to give the water soluble sodium benzylthiolate cavitand; however, this salt form of the template was not stable under the reversible conditions, and therefore the arylthiol cavitand was converted to the sodium arylthiolate cavitand in a similar manner. The synthesis of the sodium arylthiolate cavitand is shown in Scheme 4.4.

![Scheme 4.4](image)

**Scheme 4.4.** Synthesis of the sodium arylthiolate cavitand.

The sodium arylthiolate cavitand was found to be much more stable under the reversible conditions than the sodium benzylthiolate cavitand. The sodium arylthiolate cavitand was reacted with the peptide in an aqueous buffer under many conditions, but for the most part, the reaction was unsuccessful in generating the cavitein products. Some of the different experimental conditions that were attempted are listed in Table 4.5.
Table 4.5. Summary of the conditions used for the synthesis of a cavitein under reversible conditions. The reactions were carried out under nitrogen at rt for 8 h, except where indicated. The last two entries were carried out in the absence of any redox buffers to examine the linkage between the cavitand and the peptides.

<table>
<thead>
<tr>
<th>Cavitand Template + Peptide</th>
<th>Solvent (pH)</th>
<th>Ratio of Cavitand to Peptide to GSH to GSSG</th>
<th>Additional Modifications (ie. temp. atmosphere, and additives)</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sodium Arylthiolate + S3 peptide</td>
<td>0.1 M Tris Buffer, pH 8.7</td>
<td>(1:5:5:5)</td>
<td>Only pep-glutathione adduct &amp; pep-pep dimer.</td>
<td></td>
</tr>
<tr>
<td>Sodium Arylthiolate + S3 peptide</td>
<td>0.1 M Tris Buffer, pH 8.7</td>
<td>(1:5:10:5)</td>
<td>Same as above with less pep-glutathione adduct.</td>
<td></td>
</tr>
<tr>
<td>Sodium Arylthiolate + S3 peptide</td>
<td>0.1 M Tris Buffer, pH 8.7</td>
<td>(1:5:10:5)</td>
<td>Add heat (50 °C) and add light (100 W), separately.</td>
<td>Same as above.</td>
</tr>
<tr>
<td>Sodium Arylthiolate + S3 peptide</td>
<td>0.1 M Phosphate Buffer, pH 8.3</td>
<td>(1:5:10:5)</td>
<td>Same as above.</td>
<td></td>
</tr>
<tr>
<td>Sodium Arylthiolate + S3 peptide</td>
<td>0.1 M Tris Buffer, pH 8.7</td>
<td>(1:5:10:5)</td>
<td>Same as above.</td>
<td></td>
</tr>
<tr>
<td>Sodium Arylthiolate + S3 peptide</td>
<td>0.1 M Tris Buffer, pH 8.7</td>
<td>(1:5:10:5)</td>
<td>Added 1 mM EDTA and 0.2 M KCl as in ref. (31, 32)</td>
<td>Same as above.</td>
</tr>
<tr>
<td>Sodium Arylthiolate + S3 peptide</td>
<td>0.1 M Tris Buffer, pH 8.7</td>
<td>(1:5:10:5)</td>
<td>Under argon gas</td>
<td>Very small yield of desired four-helix bundle (&lt; 1 % yield)</td>
</tr>
<tr>
<td>Arylthiol Cavitand + Activated S3 Peptide</td>
<td>0.1 M Tris Buffer, pH 8.7: DMF (1:3)</td>
<td>None</td>
<td>Full linkage, but low yield of four-helix bundle (~ 8 %).</td>
<td></td>
</tr>
<tr>
<td>Arylthiol Cavitand + Activated S3 Peptide</td>
<td>0.1 M Tris Buffer, pH 8.7: DMF (9:1)</td>
<td>None</td>
<td>Partial Linkage. Mono- and di-substituted cavitein products.</td>
<td></td>
</tr>
</tbody>
</table>
The reactions between the sodium arylthiolate cavitand and S3 peptide in 0.1 M tris buffer, pH 8.7 under nitrogen at rt did not produce the desired four-helix bundle. Only side-products such as the peptide-glutathione adduct and the peptide-peptide dimer formed instead. The redox conditions were modified to minimize the amount of mixed disulfide (between peptide and glutathione), and the peptide dimers. The idea was to try to maximize the amount of the cavitein (disulfide bond between the cavitand and peptide species). Increased temperatures, more basic conditions, different buffers and added reagents were used, but to no success. When argon gas was used instead of nitrogen, a very small amount of the desired four-helix bundle resulted. Different conditions were again tried under argon with longer reaction times, but the reaction did not improve to give any more of the desired cavitein.

As seen in the last two entries in Table 4.5, reactions were carried out under non-reversible conditions between the arylthiol [4]cavitand and the activated S3 peptide in a combination mixture of the aqueous tris buffer and DMF (see Scheme 4.5). The reaction in 1:3 tris buffer to DMF resulted in a low yield of the four-helix bundle. As the amount of tris buffer was increased, the fully linked four-helix bundle was no longer obtained. Only the mono- and di-substituted cavitein products resulted.

![Scheme 4.5](image)

**Scheme 4.5.** Reaction of the arylthiol cavitand with an activated peptide in a solvent mixture of pH 8.7, 0.1 M tris buffer and DMF.
Since the main problem with setting up a reversible system for the synthesis of the caviteins appeared to be associated with the linkage between the sodium arylthiolate cavitand template and the peptides, another water soluble cavitand was considered. Mezo showed that phosphate feet could be incorporated into the benzylthiol [4]cavitand to achieve solubility in water (33). He also found that this phosphate footed bowl could link onto peptides via disulfide bonds in a phosphate buffer (33). The synthesis of the phosphate footed benzylthiol cavitand is shown in Scheme 4.6. The procedure was followed using the original literature (33); however, a modification was made to one of the steps to increase the efficiency of the synthesis. The step that was improved (in increasing the yield) is shown in bold from the methyl rimmed cavitand 14 to the benzyl bromide cavitand 15.
Originally, the bromination step from 14 to 15 was carried out using benzoyl peroxide as the catalyst under refluxing conditions. Reinhoudt’s group found that when benzoyl peroxide was used to catalyze the bromination of a pentyl footed cavitand, only partially brominated products were obtained (34). They found that using AIBN instead of benzoyl peroxide resulted in high yield of the desired completely brominated product (34). In addition, AIBN was used as
the catalyst in the successful bromination reaction for the synthesis of the larger [5]cavitand and [6]cavitand (13, 14). Because of the successful outcome of these examples, AIBN was used here instead of benzoyl peroxide. Furthermore, instead of heat refluxing this step as originally described, the reaction mixture was exposed to a 100 W light bulb to initiate the radical bromination. The overall yield of this step increased 21% compared to the original procedure.

Once the water soluble phosphate footed cavitand was synthesized, it was reacted with the activated S3 peptide in 0.1 M tris buffer at pH 8.7 under irreversible conditions, and the desired four-helix bundle was obtained as shown in Scheme 4.7. It should therefore be possible to carry out the cavitein reaction under reversible conditions (see Table 4.5) using the phosphate footed cavitand template. The continuation of this work is outlined in the next chapter under future goals.

\[ \text{AcNH[Cys-PeptideS}_{3}] \]

**Scheme 4.7.** Synthesis of a four-helix cavitein in tris buffer, pH 8.7, under nitrogen.
4.4 Chapter Summary and Conclusion

The stability and structural properties of caviteins, composed of different number of helices disulfide bonded to the [n]cavitands, were investigated. The peptide sequence S3 that was used in the earlier study of the three- and four-helix bundles, was employed to synthesize and characterize larger five- and six-helix bundles. The different-sized caviteins using the S3 sequence were studied by various techniques, and it was found that this sequence, intended for the formation of a four-helix bundle was indeed the most superior for the tetrameric species in terms of stability and structural specificity. All the TASP's demonstrated high helicity by CD spectroscopy, and molten globule characteristics by observation of the near UV spectra, 1D \(^1\)HMR spectra, and from the N-H/D experiments. Although the TASP's were found to possess characteristics of a molten globule, the four-helix bundle was found to be the most stable by GuHCl-induced denaturation studies and the most native-like as demonstrated by a higher protection factor and the least amount of ANS binding. Furthermore, the four-helix bundle was found to exist as a monomeric species. Therefore, the peptide designed for a four-helix bundle was not a generic sequence that was ideal for all different bundle sizes. The same sequence displayed different degrees of native-like character depending on the size of the bundle it was in. Other groups have studied the ability of a sequence to adopt a specific bundle size when certain residues were mutated. The study in this chapter was the first look at examining a set of proteins that was forced into forming a certain bundle size using the same sequence. The availability of the different-sized [n]cavitands allowed for the synthesis of various-sized TASP's, which were used to quantify how much better a sequence designed for a four-helix bundle was than on a five- or a six-helix bundle.
Disulfide bonded cavitein structures were also used to explore the field of dynamic covalent chemistry, which can be a valuable tool for the synthesis of a desired or a new compound. Dynamic combinatorial libraries can be used to screen for molecules with specific structural or functional properties. In this chapter, thiol-disulfide exchange reactions were attempted to employ the concept of dynamic combinatorial library in an effort to obtain the most stable protein, which is most likely to contain a high degree of native-like character. The arylthiol cavitand was converted to the water soluble sodium arylthiolate form so that the reversible reactions could take place in an aqueous environment. Different conditions were tried, but the reactions were unsuccessful. The main problem was likely the inability of the cavitand template to efficiently link onto the peptides in an aqueous media. Thus, the water soluble phosphate footed cavitand template was synthesized as an alternative to the sodium arylthiolate cavitand. This phosphate footed cavitand template was successful in linking onto the peptides in a basic aqueous buffer, and therefore can be used to carry out the cavitein reactions under reversible conditions. The next chapter discusses the future plans of this work along with other potential projects.
4.5 Experimental

4.5.1 Synthesis of the Cavitand Templates


4.5.1.2 Synthesis of Sodium Benzythiololate Cavitand

A solution of sodium methoxide (1.0 M in methanol, 130 µL, 0.13 mmol) was added to a solution of benzythiol cavitand (20 mg, 0.026 mmol) in THF (20 mL), and the reaction mixture was stirred for 2 h at rt. The reaction mixture was concentrated in vacuo and THF was added. This step was repeated a couple of times to remove the methanol. The suspension was filtered through a fine frit, washed with THF and dried at 78 °C to afford a dark yellow solid (18 mg, 81%).
1H NMR (400 MHz, D2O) δ 7.43 (s, 4 H, H_{para}), 5.89 (d, J = 7.5 Hz, 4 H, H_{out}), 4.81 (q, J = 7.4, 4 H, H_{methine}, 4.21 (d, J = 7.5 Hz, 4 H, H_{in}), 3.11 (s, 8 H, SCH2), 1.73 (d, J = 7.4 Hz, 12 H, CH3) ppm.

MS (LSIMS, thioglycerol) m/z: 864 (M+H)⁺

4.5.1.3 Synthesis of Sodium Arylthiolate Cavitand

A solution of sodium methoxide (1.0 M in methanol, 140 μL, 0.14 mmol) was added to a solution of arylthiol cavitand 5 (20 mg, 0.028 mmol) in THF (20 mL), and the reaction mixture was stirred for 2 h at rt. The reaction mixture was concentrated in vacuo and THF was added. This step was repeated a couple of times to remove the methanol. The suspension was filtered through a fine frit, washed with THF and dried at 78 °C to afford a dark yellow solid (19 mg, 85 %).

1H NMR (400 MHz, D2O) δ 7.02 (s, 4 H, H_{para}), 5.98 (d, J = 7.5 Hz, 4 H, H_{out}), 4.80 (q, J = 7.4, 4 H, H_{methine}, 4.13 (d, J = 7.5 Hz, 4 H, H_{in}), 1.70 (d, J = 7.4 Hz, 12 H, CH3) ppm.

MS (LSIMS, thioglycerol) m/z: 808 (M+H)⁺
4.5.1.4 Synthesis of the Phosphated Footed Benzylthiol [4]Cavitand

The synthesis of the phosphate footed benzylthiol [4]cavitand is described in literature starting from 2-methylresorcinol 11 (33). The step that was modified from literature is from the methyl rimmed cavitand 14 to the benzyl bromide cavitand 15 and is described here. NBS (1.51 g, 8.5 mmol) and AIBN (0.05 g, 0.3 mmol) were added to a solution of cavitand 14 (3.54 g, 2.0 mmol) in CCl₄. The solution was exposed to a 100 W light bulb 30 cm away for 24 h under nitrogen and then cooled to rt. The solution was filtered, and the filtrate was evaporated in vacuo. The crude product was purified by column chromatography (hexanes:EtOAc, 9:1) to afford benzyl bromide 15 as a white solid (3.12 g, 75 %). The characterization of this product by $^1$H NMR and MS is the same as that reported in literature (33).

4.5.2 Peptide and Cavitein Synthesis

4.5.2.1 Peptide Synthesis

All peptides were synthesized on an Applied Biosystems (ABI) 431A automated peptide synthesizer using standard Fastmoc™ procedures as described in Chapter 2.3.2.2. The final step entailed acetylating the N-terminus of the peptide with acetic anhydride before cleavage of the resin and protecting groups. The cleavage mixture contained a mixture of TFA, water and ethane dithiol in a 38:1:1 ratio. The crude peptides were purified by reversed-phased HPLC and lyophilized. Peptides were characterized using MALDI mass spectrometry.
**Peptide S3:** Ac-NH-[CGGGEELLKKLEELLKKG]-CONH$_2$

The S3 peptide on resin (800 mg) was reacted with 3 mL acetic anhydride and 1 mL NMP and stirred at room temperature under nitrogen for 1 h. The reaction mixture was filtered using a medium frit and washed with DCM. Subsequent cleavage of the resin and protecting groups by treatment with the TFA cleavage mixture was carried out. After purification by reversed-phase HPLC and lyophilization, a white solid was obtained (153 mg, 31 %).

**MS:** (MALDI, cinnamic acid): $m/z$ 1986 (M+H)$^+$

**Activated Peptide S3**

The S3 peptide (20 mg, 10 µmol) dissolved in 3 mL ethanol was rapidly added to a stirring solution of 2,2'-dipyridyl disulfide (11 mg, 50 µmol) dissolved in 2 mL ethanol, and the reaction mixture was stirred at room temperature for 1 h. The ethanol was then reduced to less than 1 mL volume *in vacuo*, and the remaining solution was pipetted onto ice-cold diethyl ether. The resulting precipitate was filtered and washed with ice-cold diethyl ether. The solid was re-dissolved in water, filtered using a 0.45 µm nylon filter, and purified by HPLC. Lyophilization resulted in a white solid (16 mg, 76 %).

**MS:** (MALDI, cinnamic acid): $m/z$ 2095 (M+H)$^+$
Peptide S4: Ac-NH-[CGGGEEAAKKAEEAAKKG]-CONH₂

The S4 peptide on resin (830 mg) was reacted with 3 mL acetic anhydride and 1 mL NMP and stirred at room temperature under nitrogen for 1 h. The reaction mixture was filtered using a medium frit and washed with DCM. Subsequent cleavage of the resin and protecting groups by treatment with the TFA cleavage mixture was carried out. After purification by reversed-phase HPLC and lyophilization, a white solid was obtained (160 mg, 36%).

MS: (MALDI, cinnamic acid): m/z 1776 (M+H)⁺

Activated Peptide S4

The S4 peptide (20 mg, 11 µmol) dissolved in 3 mL ethanol was rapidly added to a stirring solution of 2,2'-dipyridyl disulfide (12 mg, 55 µmol) dissolved in 2 mL ethanol, and the reaction mixture was stirred at room temperature for 1 h. The ethanol was then reduced to less than 1 mL volume in vacuo, and the remaining solution was pipetted onto ice-cold diethyl ether. The resulting precipitate was filtered and washed with ice-cold diethyl ether. The solid was re-dissolved in water, filtered using a 0.45 µm nylon filter, and purified by HPLC. Lyophilization resulted in a white solid (15 mg, 71%).

MS: (MALDI, cinnamic acid): m/z 1885 (M+H)⁺
**Peptide S5:** Ac-NH-[CGGGEELLKLLEELLLKKLG]-CONH₂

The S5 peptide on resin (850 mg) was reacted with 3 mL acetic anhydride and 1 mL NMP and stirred at room temperature under nitrogen for 1 h. The reaction mixture was filtered using a medium frit and washed with DCM. Subsequent cleavage of the resin and protecting groups by treatment with the TFA cleavage mixture was carried out. After purification by reversed-phase HPLC and lyophilization, a white solid was obtained (100 mg, 18 %).

**MS:** (MALDI, cinnamic acid): *m/z* 2213 (M+H)^+

**Activated Peptide S5**

The S5 peptide (22 mg, 10 µmol) dissolved in 3 mL ethanol was rapidly added to a stirring solution of 2,2'-dipyridyl disulfide (11 mg, 50 µmol) dissolved in 2 mL ethanol, and the reaction mixture was stirred at room temperature for 1 h. The ethanol was then reduced to less than 1 mL volume *in vacuo*, and the remaining solution was pipetted onto ice-cold diethyl ether. The resulting precipitate was filtered and washed with ice-cold diethyl ether. The solid was redissolved in water, filtered using a 0.45 µm nylon filter, and purified by HPLC. Lyophilization resulted in a white solid (8 mg, 35 %).

**MS:** (MALDI, cinnamic acid): *m/z* 2321 (M+H)^+
Peptide S6: C1CH₂CO-NH-[GGGEELLKKLEELLKKLG]-NH₂

The S6 peptide on resin (900 mg) was reacted with chloroacetyl chloride (122 μL, 1.54 mmol, 6 equiv.) and DIPEA (269 μL, 1.54 mmol, 6 equiv.) in 5 mL DMF for 1 hour at room temperature under nitrogen. The reaction mixture was filtered using a medium frit and washed with DCM. Subsequent cleavage from the resin and purification by reversed-phase HPLC afforded peptide S6 as a white solid (150 mg, 28 %)

MS: (MALDI, cinnamic acid): m/z 2143 (M+H)_+

4.5.2.2 Cavitein Synthesis

Each cavitein was assessed for purity by observation of a single peak by analytical reversed-phase HPLC. Their identities were confirmed by MALDI spectroscopy.

S3-[4] Cavitein

DIPEA (2.2 μL, 12.9 μmol, 10 equiv.) was added to a solution of the [4]cavitand (1.0 mg, 1.3 μmol) and activated S3 peptide (22 mg, 10.3 μmol, 8 equiv.) in degassed DMF (1 mL), and the reaction mixture was left stirring for 5 h under N₂. The reaction mixture was evaporated
in vacuo, dissolved in deionized water, and purified by reversed-phase HPLC to afford the S3-[4] cavitein (4.5 mg, 40 %).

**MS:** (MALDI, cinnamic acid): $m/z$ 8711.0 ± 0.9 (M+H)$^+$ [calcd 8710.6 Da].

S4-[4] Cavitein

DIPEA (2.2 μL, 12.9 μmol, 10 equiv.) was added to a solution of the [4]cavitand (1.0 mg, 1.3 μmol) and activated S4 peptide (21 mg, 10.3 μmol, 8 equiv.) in degassed DMF (1 mL), and the reaction mixture was left stirring for 5 h under N$_2$. The reaction mixture was evaporated in vacuo, dissolved in deionized water, and purified by reversed-phase HPLC to afford the S4-[4] cavitein (5.8 mg, 57 %).

**MS:** (MALDI, cinnamic acid): $m/z$ 7869.5 ± 0.7 (M+H)$^+$ [calcd 7869.0 Da].

S3-[5] Cavitein

DIPEA (3.8 μL, 22 μmol, 20 equiv.) was added to a solution of the [5]cavitand (1.0 mg, 1.1 μmol) and activated S3 peptide (16 mg, 7.7 μmol, 7 equiv.) in degassed DMF (1 mL), and the reaction mixture was left stirring for 5 h under N$_2$. The reaction mixture was evaporated in vacuo, dissolved in deionized water, and purified twice by reversed-phase HPLC to afford the S3-[5] cavitein (2.0 mg, 16 %).
**S3-[6] Cavitein**

DIPEA (3.1 μL, 18.0 μmol, 20 equiv.) was added to a solution of the [6]cavitand (1.0 mg, 0.9 μmol) and activated S3 peptide (15 mg, 7.4 μmol, 8 equiv.) in degassed DMF (1 mL), and the reaction mixture was left stirring for 5 h under N₂. The reaction mixture was evaporated *in vacuo*, dissolved in deionized water, and purified twice by reversed-phase HPLC to afford the S3-[6] cavitein (1.4 mg, 12 %).

**S6-Aryl[4] Cavitein**

DIPEA (1.7 μL, 9.7 μmol, 10 equiv.) was added to a solution of the arylthiol cavitand 5 (0.7 mg, 0.97 μmol) and S6 peptide (17 mg, 7.8 μmol, 8 equiv.) in degassed DMF (1 mL), and the reaction mixture was left stirring for 6 h under N₂. The reaction mixture was evaporated *in vacuo*, dissolved in deionized water, and purified by reversed-phase HPLC to afford the S6-aryl[4] cavitein (2.5 mg, 28 %).
Arylthiol Cavitand + Activated S3 Peptide in DMF:Tris (3:1)

Arylthiol Cavitand 5 (0.5 mg, 0.7 μM) was dissolved in 0.6 mL degassed DMF and degassed 0.1 M tris buffer, pH 8.7 (0.2 mL) was added very slowly. Activated S3 peptide (11.6 mg, 5.5 μmol, 8 equiv.) in degassed DMF (0.15 mL) and degassed 0.1 M tris buffer, pH 8.7 (0.05 mL) was added to the arylthiol cavitand solution. The reaction mixture was stirred at rt under nitrogen until the solution turned from cloudy to clear (~ 16 h). The reaction mixture was concentrated in vacuo, filtered and purified by reversed-phase HPLC. After lyophilization, a white solid was obtained (0.5 mg, 8 %).

MS: (MALDI, cinnamic acid): m/z 8711.0 ± 0.9 (M+H)$^+$ [calcd 8710.6 Da].

Arylthiol Cavitand + Activated S3 Peptide in DMF:Tris (1:9)

Activated S3 peptide (15 mg, 6.9 μmol, 5 equiv.) in degassed DMF (0.05 mL) and degassed 0.1 M tris buffer, pH 8.7 (0.9 mL) was added to a solution of arylthiol cavitand 5 (1 mg, 1.4 μmol) in degassed DMF (0.05 mL) and degassed DIPEA (10 μL, 55.5 μmol, 40 equiv.), and the reaction mixture was stirred for 18 h. The reaction was quenched with the addition of TFA (0.05 mL). The reaction mixture was extracted with deionized water and CHCl$_3$. The aqueous phased was filtered and purified by reversed-phase HPLC, two products were obtained. After separately lyophilizing the products, a monosubstituted cavitein (0.5 mg, 13 %) and a disubstituted cavitein (0.6 mg, 9 %) resulted as a white solid.
**S3-Phos[4] Cavitein**

Activated S3 peptide (13.2 mg, 6.3 μmol, 8 equiv.) and the benzylthiol cavitand with phosphate feet (1 mg, 0.79 μmol) in 0.1 M tris buffer, pH 8.7 (2 mL) was stirred for 2 h at rt under nitrogen. The crude product was purified twice by reversed-phase HPLC and lyophilized to afford the S3-phos[4] cavitein as a white solid (0.4 mg, 5.6 %).

**4.5.3 Circular Dichroism (CD) Experiments**

The CD experiments were carried out using all reagent grade reagents except for GuHCl, which was electrophoresis grade. The pH of the buffer solution was measured using a Fisher Scientific Accumet pH meter, calibrated using buffered standards at pH 4.0, 7.0 and 10.0, and an AccuTrpH electrode.
4.5.3.1 Far and Near UV CD Spectra

The CD spectra were recorded on a JASCO J-710 spectropolarimeter connected to an IBM-compatible PC 80286 DOS. The temperature was maintained at 25 °C with a circulating water bath. Quartz cuvettes from Hellma of 1 mm or 1 cm size were used to hold the samples, depending on the desired concentration. A background consisting of 50 mM pH 7.0 phosphate buffer was subtracted from the sample scans. A minimum of three scans were taken for each sample and averaged out. The error bars, which represent the average standard deviation at each point, were approximately ±5 %. The raw CD spectra were normalized using the equation described in Chapter 2.3.3.1.

4.5.3.2 Denaturation Studies

The JASCO J-710 CD spectrometer was used to perform the GuHCl-induced denaturation experiments. Quartz cuvettes of 1 mm or 1 cm path lengths were used to carry out high and low concentration studies, respectively. Each sample contained 40 μM cavitein (for high concentration studies) or 4 μM cavitein (for low concentration studies), and an appropriate amount of 8.0 M GuHCl solution (to make solutions that contained 0 M to 8 M GuHCl concentrations in 0.25 M intervals) in pH 7.0 phosphate buffer solution. To prepare sample solutions containing 8 M GuHCl solution, the cavitein solutions were lyophilized, and dissolved directly in the 8 M GuHCl stock solution. The samples were vortexed right after preparation and right before taking the measurements. Each sample was measured at 25 °C for a minimum of
three times on three different days. The stabilities of the caviteins were determined by methods described in Chapter 2.3.3.2.

4.5.4 Sedimentation Equilibria Studies

Sedimentation equilibria studies were carried out on a temperature controlled Beckman Coulter Optima™ XL-1 analytical ultracentrifuge with an An50 Ti 8-hole rotor as described in Chapter 2.3.4. Figures 4.18 to 4.23 show the data fits of the S3-[5] and S3-[6] caviteins. Although the sedimentation equilibria studies were carried out at three different concentrations (20, 40 and 60 μM), and three different rotor speeds (27000, 35000 and 40000 rpm), only the low concentration data are shown. The fits are evaluated in terms of the scatter around the x-axis of the residuals plot. Residual plots that show evenly distributed points over a small magnitude represent a good fit. For the S3-[5] cavitein, the sedimentation equilibria data are shown for fits to a monomer-dimer in equilibrium for the low concentration samples at 27000 rpm and 35000 rpm. A fit to a monomer for the S3-[5] cavitein is also shown to demonstrate a poor fit. For the S3-[6] cavitein, the fits to a single non-interacting species are shown for the low concentration samples at 27000, 35000 and 40000 rpm.
Figure 4.18. Sedimentation equilibrium analysis of 20 μM S3-[5] cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 27,000 rpm. The solid line in the upper plot represents the theoretical fit to a monomer-dimer in equilibrium. The lower plot represents the residuals to the fit.
Figure 4.19. Sedimentation equilibrium analysis of 20 µM S3-[5] cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 35,000 rpm. The solid line in the upper plot represents the theoretical fit to a monomer-dimer in equilibrium. The lower plot represents the residuals to the fit.
Figure 4.20. Sedimentation equilibrium analysis of 20 μM S3-[5] cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 27,000 rpm. The solid line in the upper plot represents a poor fit to a monomer. The lower plot represents the residuals to the fit, showing the uneven distribution of data points around the best-fit line.
Figure 4.21. Sedimentation equilibrium analysis of 20 μM S3-[6] cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 27,000 rpm. The solid line in the upper plot represents the theoretical fit to a single non-interacting species. The lower plot represents the residuals to the fit.
Figure 4.22. Sedimentation equilibrium analysis of 20 μM S3-[6] cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 35,000 rpm. The solid line in the upper plot represents the theoretical fit to a single non-interacting species. The lower plot represents the residuals to the fit.
Figure 4.23. Sedimentation equilibrium analysis of 20 μM S3-[6] cavitein in 50 mM phosphate buffer, pH 7.0 at 20 °C using a rotor speed of 40,000 rpm. The solid line in the upper plot represents the theoretical fit to a single non-interacting species. The lower plot represents the residuals to the fit.
4.5.5 NMR Experiments

The $^1$H NMR spectra were acquired by Okon in the McIntosh laboratory, UBC Department of Biochemistry.

4.5.5.1 1D $^1$H NMR Spectra

1D $^1$H NMR spectra were obtained using a 500 MHz Varian Unity NMR spectrometer at 25 °C. The cavitein samples were prepared by dissolving the caviteins in 90 % pH 7.0, 50 mM phosphate buffer and 10 % D$_2$O. The concentrations of the measured protein samples were 0.57 mM for the S3-[5] cavitein and 0.42 mM for the S3-[6] cavitein. The data were processed using MestRec Version 2.3.

4.5.5.2 N-H/D Exchange 1D $^1$H NMR Spectra

The N-H/D exchange experiments were run on a 500 MHz Varian Unity at 20 °C. The experiment was initiated by the addition of D$_2$O to a lyophilized sample containing 0.3 mM of each of the cavitein samples, S3-[4] cavitein, S3-[5] cavitein, or S3-[6] cavitein in pH 4.62 acetate buffer. For the reference spectra, 0.3 mM samples of each cavitein were measured in 90 % pH 4.62 acetate buffer, and 10 % D$_2$O. From the N-H/D exchange experiments, the protection factor $P$, was calculated using the following equation:
\[ P = \frac{k_{int}}{k_{ex}} \]

\( k_{int} \) and \( k_{ex} \) were calculated as explained in Chapter 2.3.5.3. Using the calculated protection factor values, the caviteins were assessed in terms of native-like character.

4.5.6 ANS Binding Studies

ANS binding studies were conducted on a Varian CARY Eclipse Spectrometer at 25 °C. Each measured solution contained either 50 μM of cavitein sample, 95 % ethanol, or 100 % HPLC grade methanol. All the cavitein samples contained 2 μM ANS in 50 mM phosphate buffer at pH 7.0. The 95 % ethanol and 100 % methanol also contained 2 μM ANS. The excitation wavelength was set to 370 nm and the emission was recorded between 400 nm and 600 nm.
4.6 References


CHAPTER FIVE: Thesis Conclusion and Future Outlook

5.0 Thesis Conclusion

The three-dimensional structure of a protein is encoded in its primary amino acid sequence. To understand this relationship, it is important to recognize the various forces that stabilize the structures of the proteins. As there are many variables involved in controlling a protein structure, it is a major undertaking to unravel the protein folding problem by predicting a sequence that will adopt a particular fold. An instrumental approach to studying this problem has been to use de novo proteins, which are designed from scratch, but retain much of the same information as natural proteins. This thesis discussed ways to help elucidate the protein folding problem by using de novo template assembled synthetic proteins (TASPs). The difference between our group and others who have studied TASPs is the use of a cavitand as the template. A cavitand is an organic macrocycle composed of bridged aromatic rings. Other TASP examples include using Mutter's peptidic template and Ghadiri's metal-ligand-based templates. The peptidic templates were non-rigid, resulting in proteins with relatively low stability (1), and the longer linkers required for the metal-ligand-based proteins resulted in proteins with poor packing (2). On the other hand, a cavitand template can provide rigidity, and still allow sufficient helical flexibility (i.e. linkers can be incorporated into the structure), to produce proteins with high stability and specificity. Another advantage is that this template has an enforced cavity that could be used to bind small ions and molecules (see future goals section). In this thesis, a cavitand was used as the template to promote protein folding.
The general objective of the protein field is to contribute to solving the protein folding problem. The main goals of this thesis were to learn what factors are important in obtaining a structurally specific protein, and to be able to synthesize a cavitein with improved native-like characteristics. In particular, the linker length effect between the peptides and the cavitand template was probed (Chapters 2 and 3), and the effect of a specifically designed peptide sequence in different-sized bundles (i.e. different number of helices) was examined (Chapter 4).

In Chapter 1.5.2, different examples of TASPs were presented. In all the cases, the template promoted folding, provided stability and increased helicity. However, the relationship between the template and the peptides were not studied in much detail. For example, Fairlie and coworkers synthesized a series of four-helix bundles using aromatic-based templates that differed in size, shape and directionality. They came to the conclusion that when sufficiently long linkers were used between the template and the peptides, the resulting proteins were not sensitive to the dimensions of the template (3). They expected that the structure of the template should become important with shorter linkers, however, this hypothesis was not tested. In another example, a cholic acid-based template was used in the synthesis of the proteins with different linker lengths. It was shown that the linker length had an effect on the helicity of the proteins, but this finding was not further explored (4).

The length of the linker between the cavitand template and the peptides was found to have a dramatic effect on the structural properties of the caviteins (5). This first generation series of four helix bundles were composed of a peptide sequence that was designed to link from its hydrophobic/hydrophilic interface. In Chapter 2, a new series of four-helix bundle caviteins was synthesized and characterized using a peptide sequence that was designed to link 1/3.6 of a turn closer to the hydrophobic face compared to the first generation sequence. The idea was to reduce the linker length requirement in order to improve the packing between the helices, thereby
increasing the native-like characteristics of the caviteins. The first generation work showed that at least two to three Gly linkers were required to obtain a monomeric, well-defined cavitein (5). Although the first generation 2GS0 cavitein gave the most well-dispersed $^1$H NMR spectrum, the 3GS0 cavitein gave sharper signals in its $^1$H NMR spectrum, and was found to be more stable by GuHCl-induced denaturation studies. The new generation 2GS2 cavitein was found to demonstrate the most native-like characteristics of both the first and second generation linker series. Excluding the first generation 1GS0 cavitein, which was found to be a dimer, the 2GS2 cavitein showed the most-enhanced signal in the near UV region of its CD spectrum, the overall most well-dispersed, sharp signals in its $^1$H NMR spectrum, and possessed the highest protection factor from the N-H/D exchange experiments. In addition, the 2GS2 cavitein was found to be the most stable monomeric cavitein with a free energy of folding approximately 4 kcal/mol more stable than any other monomeric cavitein. Unlike the first generation caviteins, it was clear that no more or no less than two Gly residues were necessary in obtaining a native-like protein using the second generation peptide sequence. Thus, by changing the linkage point closer to the hydrophobic core, packing within the interior of the cavitein was improved and structural specificity was increased. Since a slightly shorter linker was required for the second generation caviteins compared to the first generation series, it was deduced that the interhelical distance between two adjacent helices containing nonpolar Leu residues was in the range of 8 to 10 Å. Since a slightly shorter linker was required to afford a well-defined protein when the linkage point was closer to the hydrophobic core, it would be interesting to see if a slightly longer length linker would be required for a sequence that is designed to link from the hydrophilic face, and to see if this longer linker would still be capable of producing a well-defined cavitein capable of taking advantage of the template’s stabilizing effects. By synthesizing and characterizing this
third generation of caviteins, and comparing the three generations, an even better understanding of the effective size of the helices may be achieved.

In Chapter 3, molecular dynamics simulations were performed to gain better insight into the behaviour of the caviteins in the second generation. Three sets of simulations starting from three different configurations were carried out, and all sets generally agreed with each other, and with the experimental results. All the caviteins were found to be highly helical by simulations as they were found to be experimentally. The interhelical distance between two adjacent helices were calculated from the perimeter distance values, and found to agree with the 8 to 10 Å range estimated in Chapter 2. The propensity of the caviteins to self-aggregate was rationalized by analyzing the representative clusters. The 1GS2 cavitein and the 3GS2 cavitein were found to self-aggregate from the sedimentation equilibria studies. The average perimeter distances at the C-terminal end were higher than that for the central residue, suggesting that self-aggregation of these caviteins was occurring at the C-terminal region. The 2GS2 cavitein was of particular interest due to its high conformational specificity found experimentally. From the simulation results, it was shown that the 2GS2 cavitein was tilting with respect to the cavitand template. This tilt may have been occurring to increase the burial of the nonpolar Leu residues and to improve the interhelical packing, which may explain the high stability and specificity of the 2GS2 cavitein. However, the limitations of this computer modelling work included the absence of a starting configuration and the lack of infinite time. To more accurately and thoroughly understand the properties and behaviour of the caviteins, an X-ray crystal or NMR structure would be ideal.

Although the goals of this project were to determine the linker length effects on the cavitein structures and to find an optimal linker between the cavitand and peptides, the main objective of the thesis was to increase our understanding of the factors which contribute to the
structural specificity and stability of the proteins. An issue that was raised during the study of the linker length effects was related to the middle hydrophobic residue. The 2D $^1$H NMR resonance assignments and N-H/D exchange results revealed that the central hydrophobic Leu was imperative in stabilizing the overall four-helix bundle. Work has been carried out to study the context-dependence of the hydrophobic residues in the four-helix caviteins by individually substituting each of the Leu residues with Ala. This study also revealed that the middle Leu was the most important position in maintaining a stable well-defined cavitein (6).

So how well were we able to design a protein from scratch? In Chapter 2, a native-like four-helix bundle was successfully obtained by slightly reducing the linker length requirement between the cavitand template and the peptides. The fact that a well-defined four-helix bundle was obtained emphasizes the importance of a peptide design. To further study the success or failure of a design strategy, a peptide sequence intended for a four-helix bundle was linked onto the various-sized [n]cavitands. Causton studied this particular sequence on three- and four-helix bundle TASP's (7), and found that the latter was more stable and structurally specific than the former. In Chapter 4, this same sequence was used to synthesize and characterize larger five- and six-helix bundles. Again, the four-helix bundle was the most stable and structurally more defined than the larger helical bundles. This finding could be explained by the packing within the hydrophobic core. Since the sequence contained five nonpolar residues intended for a four-helix bundle, the nonpolar Leu residues were probably able to pack more efficiently than in a three-, five-, or six-helix bundle. In the smaller three-helix bundle, the nonpolar residues could have been overpacked in the hydrophobic core. For the five- and six-helix bundles, the larger [n]cavitands may have created a cavity within the hydrophobic core, which in turn could have resulted in poorly-defined structures. On the other hand, the helices could have been overpacked in these larger helical bundles as the extra aromatic ring in the cavitand template may have not
been able to accommodate the addition of a helix with a slightly larger effective diameter. An interesting finding was the ability of the larger five- and six-helix bundles to bind ANS. This result was the first time any cavitein had been able to bind a small molecule, and therefore these larger caviteins could be used in the study of potential drug binding sites (see future goals section).

Furthermore, it was demonstrated that although the sequence was designed for a four-helix bundle, the six-helix bundle was found to be more well-defined than the five-helix bundle. For example, the S3-[6] cavitein was found to be a monomer and possessed a slightly higher protection factor than the S3-[5] cavitein. These results could have been due to the asymmetric structure of the [6]cavitand, which may have facilitated better packing between the helices compared to the five-helix bundle cavitein.

We have demonstrated the importance of peptide design in dictating a protein's stability and conformational specificity. We were also able to synthesize a cavitein with improved native-like characteristics. But how native-like are these caviteins? In general, de novo proteins composed of an all leucine core have been found to exhibit molten globule structure (8, 9). In Chapter 1.4.1 (DeGrado's family of de novo proteins), an incremental approach was used to improve the structural specificity of a protein. α2B is a four-helix protein containing an all leucine core. When different nonpolar residues were substituted for some of the Leu residues, the packing in the interior of the core was improved to give α2C, but this protein still displayed characteristics of a molten globule structure. In the next step, a potential metal binding site was incorporated to give α2D; this protein was found to resemble a native-like structure (10). In the case of the caviteins, the cavitand template was able to induce native-like character of the attached helical bundle even with an all leucine core (11). The 2GS2 cavitein displayed comparable native-like properties as the α2D protein (in terms of cooperative denaturation, 1H
NMR chemical shift dispersion and lack of ANS binding), although the methyl region (0-1.25 ppm) of the α2D protein was more dispersed than that of the 2GS2 cavitein. However, the α2D protein contains a hydrophobic core with a more diverse set of nonpolar residues, whereas the 2GS2 cavitein only contains Leu residues in its hydrophobic core. It would be interesting to see if the conformational specificity of the 2GS2 cavitein could be increased by incorporating a more diverse set of residues like in the α2D protein? The S3-[5] and S3-[6] caviteins were more comparable to the molten globule structure of the α2C protein (showed cooperative denaturation, but bound ANS and displayed broad, poorly-dispersed signals in its ^1H NMR spectrum).

The stability of the caviteins generally correlated with their native-like character. The higher the stability of a cavitein, the more native-like it was found to be. However, it should be pointed out that a protein that is stable to denaturation is not necessarily native-like (12). For example, α4 is stable four-helix bundle protein (13) with higher stability (-22.5 kcal/mol) and a higher m value (3.57 kcal/molM) compared to any of the caviteins; however, the α4 protein showed poor chemical dispersion in its ^1H NMR spectrum and bound ANS. In general, the caviteins were higher or comparable in stabilities to other TASPks (14, 15) and some natural proteins (16). Although the caviteins with an optimal linker have been found to possess native-like characteristics similar to other de novo proteins, they are still not as native-like as some natural proteins. For example, the protection factor of the highly protected protons in cytochrome c was determined to be in the range of 10^8 (17), whereas the protection factor for the 2GS2 cavitein was found to be in the range of 10^4, which shows that the cavitein design could still be improved to increase their native-like character.

Solving the protein folding problem is an immense task. De novo protein design has provided a powerful tool for investigating protein structure and folding. We have studied the effects of various factors on the protein structure, which increased our understanding of how to
design more native-like proteins. The significance of a precise peptide design was demonstrated; however, in the case of template assembled synthetic proteins, the sequence design alone did not contain all the information necessary in determining the native-like character of the caviteins. The linker length and the number of helices in a bundle (dictated by the size of the [n]cavitand) had substantial effects on the overall structure. Improving our knowledge about how to synthesize native-like proteins is important, but finding new techniques to identify these stable structures with specific interhelical interactions is also important. A strategy using dynamic covalent chemistry was described in Chapter 4, in an attempt to obtain the most stable caviteins, since native-like proteins are reasonably stable. The continuation of this work is described in the next section.

5.1 Future Goals

5.1.1 Continuation of the Reversible Cavitein Systems

Now that a suitable water soluble cavitand has been synthesized, this phosphate footed cavitand template can be reacted with the two different peptides, S3 and S4 under reversible conditions to determine if the most stable four-helix bundle would form. If this procedure is successful, the proof of principle stage can be continued and a mixture of three or four different peptides can be reacted with the cavitand template to find out which combination of peptides would afford the most stable four-helix bundle. Eventually, it should be possible to create a dynamic combinatorial library consisting of different peptide sequences to screen for the most
stable (most likely native-like) cavitein structures. Furthermore, dynamic combinatorial chemistry can be used in synthesizing and isolating molecules that are otherwise difficult using traditional methods. For example, in principle, a [4]cavitand reacted under reversible conditions with four different peptide sequences could result in a four-helix bundle with four different peptides. Since a four-helix bundle with the same sequence is relatively symmetric, it is difficult to study by NMR spectroscopy. However, more information can be obtained from the NMR spectra of helical bundles using sequences with more variety. For example, instead of Leu as the only nonpolar residue, and Glu and Lys as the only polar residues, other amino acids could be incorporated in their place.

5.1.2 Continuation of the Effect of Different-Sized [n]Cavitands

The study of the effects of the different-sized cavitands was carried out using a sequence which resulted in relatively molten globule structures. This sequence was chosen to synthesize the five- and six-helix bundles for comparative purposes, as the previous study used this sequence to synthesize three- and four-helix bundles. A different sequence that has the ability to form more native-like proteins may be more useful in seeing distinct differences between the data obtained experimentally. For example, using the S3 peptide sequence that was disulfide bonded to the cavitand template resulted in very broad $^1$H NMR spectra, regardless of the bundle size. Perhaps linking the 2GS2 peptide sequence via thioether bonds to the [n]cavitands would produce more native-like proteins, which would result in more distinct experimental data between the different-sized helical bundles. It would also be interesting to see if the middle hydrophobic residue still has the same importance in the larger five- and six-helix bundles.
Because the dispersion in the $^1$H NMR spectra of the S3-[5] and S3-[6] caviteins was poor, and the N-H/D exchange occurred within minutes, it was difficult to determine which residue in the sequence was the most important in stabilizing the helical bundle. Again, if a more native-like sequence is used for the formation of the larger helical bundles, then the $^1$H NMR spectra would be more dispersed, and exchange should occur slowly enough to determine the last residue remaining.

5.1.3 Structure Determination

One of the major limitations mentioned in Chapter 3 was the lack of a three-dimensional structure. The molecular dynamics simulation results were biased toward the chosen starting configurations. The simulation time may have not been sufficient in sampling the actual conformations, which limited the accuracy of the results. A three-dimensional structure of the caviteins would be extremely useful in understanding the properties of these template assembled synthetic proteins. Both NMR spectroscopy (18) and X-ray crystallography (19) have been successful in the determination of a protein structure. Since the caviteins are highly symmetric and were synthesized using the minimalist approach (with a minimum number of residue types), determination by NMR spectroscopy would be very difficult. By reducing symmetry or introducing structure diversity as explained above in section 5.1.1., it may be possible to determine the solution structure of a cavitein. For X-ray crystallography to be useful, the proteins must have the ability to crystallize. The caviteins are very soluble in aqueous solutions, which make it difficult for them to crystallize. Using more nonpolar residues in the sequence may help solve this problem.
5.1.4 Proteins with Function

With improved understanding about the relationship between sequence and structure of the \textit{de novo} proteins, it would be possible to design and synthesize more complex template assembled synthetic proteins with a specific function. For example, it was found that the larger five- and six-helix bundles were capable of binding the hydrophobic dye ANS. Therefore, it may be possible to carry out drug binding studies using these larger template assembled synthetic proteins. Quinine, an antimalarial drug, has an aromatic structure similar to that of ANS with an additional alkaloid group (see Figure 5.1). Quinine binding studies can be carried out to determine if the larger caviteins will display a similar fluorescence to ANS.

![ANS and Quinine structures](image)

\textbf{Figure 5.1.} Structure of ANS (left) and quinine (right).

Binding sites have been engineered into a protein to regulate the oligomeric state of the protein. For example, Alber and coworkers demonstrated that a GCN4 leucine zipper mutant favours dimer formation; however upon binding to a hydrophobic ligand such as benzene or cyclohexane, trimer formation was favoured (20). The larger S3-[5] and S3-[6] caviteins bound to ANS can be analyzed by sedimentation equilibria studies to determine if the oligomeric state
was altered. Binding sites have also been engineered to improve the native-like character of the proteins. DeGrado and coworkers enhanced the native-like characteristics of a de novo protein by incorporating a metal binding site at the interfaces of the helices (21). The ANS bound S3-[5] and S3-[6] caviteins could be analyzed by near UV circular dichroism and $^1$H NMR spectroscopy to determine if the bound caviteins possess improved native-like characteristics from the non-bounded species. Since the S3 peptide sequence was designed to form a four-helix bundle, when the sequence was attached to the larger [5]cavitand and [6]cavitand templates, a cavity within the hydrophobic core may have resulted. It is possible that the ANS bound S3-[5] and S3-[6] caviteins would show increased conformational specificity as the hydrophobic dye can fill or reduce the space of this cavity.

*De novo* metalloproteins have been the focus of many studies (22, 23). Schnepf and coworkers showed the ability of a four-helix bundle to bind copper by incorporating a Cys and several His residues at positions necessary for the ligation of the metal (23). CD and 1D NMR spectroscopy were used to study the secondary and tertiary structure, and UV-vis spectroscopy was used to demonstrate the binding of the metal. The metalloprotein examples show that it is possible to design metallocaviteins by incorporating residues in the peptide sequence that are capable of ligating to metals. Metal-binding sites can have structural purposes as explained above, or functional purposes, such as in the case of catalysis.

*De novo* design of catalytic proteins can provide the basis to engineer novel catalysts and also offer increasing knowledge of how enzymes function. The successful designed protein catalyst must be able to bind and orient the substrates, transition states and intermediates that are next to the catalytic groups. The advantages of using *de novo* proteins is that specific structural components can be controlled or modified, and a greater versatility of substrates may be offered in terms of size and shape. Caviteins can be designed to act as catalysts and may offer an added
advantage because they are robust and can be tested under a wide range of temperatures and pH values. A popular topic in the synthetic catalytic protein field has been the hydrolysis of nitrophenyl esters (24, 25). Baltzer and coworkers showed a pair of His residues close in space which was able to catalyze the hydrolysis of p-nitrophenylesters through cooperative nucleophilic and general acid catalysis (26). Thus far, the caviteins have not been synthesized with any His residues. However, in the future, His residues can be incorporated in order to examine their catalytic effect on various chemical reactions. Eventually, with increased understanding in the area of small protein catalysts, synthetic enzymes that are able to speed up chemical reactions that are not catalyzed by nature may be tailor-made.

Another topic of interest has been associated with membrane proteins (27, 28), which play significant roles in many cellular processes and account for approximately one third of the gene products. Moreover, they make up for approximately 70 % of protein targets in the pharmaceutical industry. Although abundant, very little is known about the structure and mechanism of these proteins. One method to help understand their structure and how they function is to make synthetic membrane proteins and study their interaction with the lipid bilayers. Membrane proteins can be classified as peripheral or integral proteins. Peripheral proteins reside on the surface of the bilayer, whereas integral proteins reside span the lipid bilayer and often act as transport facilitators. The protein sequence and the surrounding environment affect the relative orientation of the amphiphilic segments. In a polar environment the protein should bury its nonpolar residues into a hydrophobic core and expose its polar residues to the aqueous solvent to give a peripheral membrane protein; whereas in a nonpolar environment such as the lipid bilayer, the protein should expose its nonpolar residues to the membrane and its polar residues to the core, giving rise to an integral transmembrane protein. Mutter's group demonstrated the latter by designing a four-helix TASP that was able to form an
ion channel by reorientation of the helical modules in the lipid bilayer to form a hydrophilic pore (29). Most protein channels are thought to be alpha helical (30, 31). The original designed caviteins have the potential to form transmembrane proteins due to their amphiphilic nature and high helicity. An important design strategy to consider in creating a synthetic ion channel is to incorporate charged or polar groups at the end of the protein channels. From studies of natural ion channels, most often, charged residues are found at the ends of the helices to provide good interaction with the polar head groups of the lipid bilayer (32). Mutter’s four-helix TASPs that were found to form synthetic ion channels were oriented by a peptidic template consisting of several charged Lys residues. The caviteins, however generally contain a nonpolar cavitand entity, which would lack the anchoring ability to the polar head groups, and hence would not help with the orientation. However, a good template to generate a transmembrane cavitein would be the phosphate footed cavitand described in Chapter 4.3, since the phosphate feet are polar and have the ability to interact with the polar head groups of the lipid bilayer. Thus, a potential transmembrane cavitein can be designed and synthesized using an amphiphilic, relatively nonpolar peptide sequence and the phosphate footed cavitand template.
5.2 References


### Appendix A. The Twenty Commonly Occurring Amino Acids.

![Amino Acid Structure]

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>One Letter Code</th>
<th>Three Letter Code</th>
<th>Side Chain = R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycine</td>
<td>G</td>
<td>Gly</td>
<td>H</td>
</tr>
<tr>
<td>Alanine</td>
<td>A</td>
<td>Ala</td>
<td>CH₃</td>
</tr>
<tr>
<td>Valine</td>
<td>V</td>
<td>Val</td>
<td>CH(CH₃)₂</td>
</tr>
<tr>
<td>Leucine</td>
<td>L</td>
<td>Leu</td>
<td>CH₂CH(CH₃)₂</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>I</td>
<td>Ile</td>
<td>Isobutyl</td>
</tr>
<tr>
<td>Serine</td>
<td>S</td>
<td>Ser</td>
<td>CH₂OH</td>
</tr>
<tr>
<td>Threonine</td>
<td>T</td>
<td>Thr</td>
<td>CH(OH)CH₃</td>
</tr>
<tr>
<td>Cysteine</td>
<td>C</td>
<td>Cys</td>
<td>CH₃SH</td>
</tr>
<tr>
<td>Methionine</td>
<td>M</td>
<td>Met</td>
<td>CH₂CH₂SCH₃</td>
</tr>
<tr>
<td>Proline</td>
<td>P</td>
<td>Pro</td>
<td>See below</td>
</tr>
<tr>
<td>Aspartic Acid</td>
<td>D</td>
<td>Asp</td>
<td>CH₂COOH</td>
</tr>
<tr>
<td>Asparagine</td>
<td>N</td>
<td>Asn</td>
<td>CH₂CONH₂</td>
</tr>
<tr>
<td>Glutamic Acid</td>
<td>E</td>
<td>Glu</td>
<td>CH₂CH₂COOH</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Q</td>
<td>Gln</td>
<td>CH₂CH₂CONH₂</td>
</tr>
<tr>
<td>Lysine</td>
<td>K</td>
<td>Lys</td>
<td>CH₂(CH₂)₂NH₂</td>
</tr>
<tr>
<td>Arginine</td>
<td>R</td>
<td>Arg</td>
<td>(CH₂)₃NHC(NH)(NH₂)</td>
</tr>
<tr>
<td>Histidine</td>
<td>H</td>
<td>His</td>
<td>See below</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>F</td>
<td>Phe</td>
<td>CH₂Ph</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Y</td>
<td>Tyr</td>
<td>CH₂Ph-OH₂</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>W</td>
<td>Trp</td>
<td>See below</td>
</tr>
</tbody>
</table>

R = Proline Histidine Tryptophan

![Proline Structure]

![Histidine Structure]

![Tryptophan Structure]
Appendix B. Computer Modelling Definitions

1. **algorithm**: A set of instructions a computer follows to solve a problem.

2. **coiled coil phase yield (Δ\(\omega_n\))**: A value used to determine the extent of supercoiling of the helical bundles. The coiled coil phase yield is -4 over the length of the sequence for an ideal left-handed coiled-coil. A value of zero corresponds to the absence of supercoiling of the helices. A positive value indicates the presence of a right-handed coiled-coil structure.

3. **in silico**: literally means "in silicon" and means "performed on computer or via computer simulation." The phrase is derived from the latin phrases *in vivo* and *in vitro* that are commonly used in biology and are used to refer to experiments done in living organisms and outside of living organisms respectively.

4. **in explicit water**: all the atoms in the water molecules are treated directly in the calculations.

5. **boundary conditions**: a way to mimic a complex system by simulating one unit. Macroscopic properties can be calculated from the simulations using a relatively small number of particles, in a way that the particles experience forces as if they were in bulk fluid. Imagine a box that contains particles that are replicated in all directions to give a periodic pattern. Only the central box is simulated. The coordinates of the particles in the image boxes can be computed by addition or subtraction of integral multiples of the box sides. Particles interact with the closest periodic images of the surrounding systems. If a particle leaves the box during the course of the simulation, it is replaced by an image particle that enters from the opposite side. Hence, the number of particles in the central box always remains constant. The cube is the simplest periodic system, but other shapes that can fill all the space by translation operations of the 3D shape can be used (i.e. space filling shape). The best choice is to choose a periodic cell that reflects the geometry of the system. In the case of the caviteins, the most appropriate shape would be a rectangular box.

6. **NPT ensemble**: the system is kept using a constant number of atoms (or molecules), constant pressure and constant temperature.

7. **Berendsen weak coupling methods**: A method for maintaining constant temperature or pressure. To keep the pressure of the system constant, the periodic boundary box is expanded or shrunk, and the molecule is scaled accordingly. To keep the temperature constant, the velocities are scaled.

8. **SHAKE Method**: A technique used to put constraints on bonds or angles of the system.

9. **constraint**: A fixed requirement that the system is forced to satisfy (i.e. in constraint dynamics, bonds or angles are forced to adopt specific values throughout the simulation. In constraint dynamics, the total degrees of freedom are reduced.
10. **restraint**: When a bond or angle is restrained, it is able to deviated from the desired value. The restrain only acts to encourage a particular value. Restraints are most easily incorporated using additional terms in the force field potential energy function. Using restraints does not change the degrees of freedom.

11. **reaction field long range**: Compensates for the cut-off which results in artifacts. The system is divided in two parts: an inner region where the atoms are explicitly treated, and an outer region where the atoms are treated as a continuous medium. An analytical solution is used to give a modified Coulomb potential.

12. **double difference matrix distance** ($D^{QR}$): Measurement used to compare different conformations. Two conformations are considered to be neighbours, if $D^{QR}$ using all alpha carbons is less than 1.5 Å.

13. **splay**: When the interhelical distance increases from the N-terminus to the C-terminus.

14. **fray**: When a helix unravels or uncoils from one end.

15. **rigid-body superposition** of structures can lead to misleading results for different conformations because the subset of atoms that are chosen for superposition and comparison is arbitrarily.

16. **PROCHECK**: Program that is used to determine the dihedral ($\phi$ and $\psi$) angles of the residues.
Appendix C. **Steps for Setting up the Molecular Dynamics Simulations**

Most molecular modelling consists of three main stages: 1) to select a model that describes the intra- and inter-molecular interactions in the system, 2) to calculate the values using an energy minimisation, a molecular dynamics, or Monte Carlo simulation, and 3) to analyze the calculations and to check that the simulation has been carried out properly.

The two most common models are quantum mechanics and molecular (classical) mechanics. These models allow for the calculation of energies of the atoms and molecules in different arrangement. For the simulation of the caviteins, the classical mechanics model is chosen because it is simpler and less time consuming than quantum mechanics. Classical mechanics only consider atoms whereas quantum mechanics consider all nuclei and electrons.

Molecular dynamics simulation will be used to calculate the forces and the relative energies. The Molecular Dynamic Algorithm is as follows:
Steps in Setting up and Running the simulation:

1) Pick the Hamiltonian, choose the boundary conditions, and set up the topology

2) Obtain the coordinates

3) Pre-minimise

4) Solvate the proteins

5) Minimise

6) Equilibrate the system

7) Run the simulation

1a) Pick the Hamiltonian:

A molecular model of a classical system must be constructed in the form of a Hamiltonian, which is used to sample phase space. When choosing the Hamiltonian, one must determine the appropriate ensemble (here, we use the isothermal-isobaric ensemble-constant N, P, T). It is also important to determine the pH, which in turn affects the protonation state.

The Hamiltonian is also determined by: i) choice of degree of freedom, ii) choice of functional form, and iii) choice of parameters. The Hamiltonian equation is governed by the kinetic energy, $K$ and the potential energy, $V$:

$$H(q, p) = K(q, p) + V(q, p)$$

where $\vec{q}$ and $\vec{p}$ are the coordinates and the momenta

i) Before setting up an equation, one must consider the relevant degrees of freedom and the choice of the generalised coordinates in space. Larger degrees of freedom may better describe a model, however takes a longer time to simulate. On the other hand, lowering the degrees of freedom may save time, but is further from reality. Classical methods treat all atoms explicitly. A
way of reducing the degrees of freedom is by using a united atom, which combines non-polar hydrogen atoms into heavy atoms.

ii) Functional forms must be chosen for the types of interactions involved. Different functional forms are required for bonded versus non-bonded atoms. For covalent interactions, one must consider bonds, bond angles, dihedral angles (torsional angles), and improper dihedral angles. For non-bonded interactions, the Lennard–Jones and Coulomb potential are considered. (see Gromos Reference for equations). The forces on the atoms can be calculated from the potential energy function:

\[ f_i(t) = -\frac{\partial V(r(t))}{\partial r_i} \]

iii) The appropriate parameters that best reflect reality must be chosen. Force field parameters are influenced by the choice of d.o.f, choice of functional form, choice of test set and choice of reference data. It is also important to achieve proper energetic balance between all the intramolecular and intermolecular interactions.

1b) Choose the boundary conditions:

Using boundaries, macroscopic properties can be calculated from simulations using a relatively small number of particles, in a way that the particles experience forces as if they were in bulk fluid. Here, rectangular periodic boundary conditions were used.
1c) Setting up the topology

Scosim.master is a file that contains the master set of instructions that were written for setting up the simulation for the caviteins. Genjobfile.py reads the scosim.master and template files to generate a job file.

This file contains information such as the SIMCODE (title), GLYNUM (number of glycines), SCO GMT title, WINKIE (type of bowl where A denotes the aryl bowl and B denotes the benzyl bowl), where to find GROMOS 96 programs and files, project directory and names, topology file names, number of solute molecules (NRP), number of solvent molecules (NSM), total number of atoms (NEGRLINE (NRP)), and the job codes used for the simulation runs in order or execution: GENMTB, SCO GMT, GENC RDS, PREMINIMISE, PROBOX, MINIMISE, STARTUP, MDRUN. These variables are from the template file that is copied into the directory of each new cavitein to set-up for simulation (ie. genmtb.templ).

The job codes are briefly described below:

GENMTB: Generates building blocks

SCO GMT: Writes topology file by putting the blocks together

GENCRDS: Generates coordinates

PREMINIMISE: Minimises initial unsolvated configuration
PROBOX: Solvates Protein
MINIMISE: Minimises initial solvated molecule
STARTUP: Equilibrates system, reads initial coordinates, and generates velocites
MDRUN: Continues to run the simulations

Note: The term ‘configuration’ is used to describe the arrangement of the atoms, and should not be confused with its standard meaning of different bonding arrangement of the atoms in a molecule.

GENMTB reads from scosim.master and genmtb.templ to generate a job file called jgenmtbGAcav1.sh.

After the job is run, the building blocks are generated, but the topology is not complete. Force field information such as the bond lengths and force constants, as well as the molecular topology building information are read by SCOGMT to complete the topology. SCOGMT is a modified version of PROGMT, which accounts for the cavitant template.
The topology file name that is generated can be seen in scosim.master. This file should be checked for correct title and correct number of glycines.

2) Obtain the coordinates

Initial coordinates are required to set up a simulation. Usually the structure is solved by X-ray crystallography or by NMR (solution state) spectroscopy. In the absence of any three-dimensional structure, as in the case of the caviteins, the initial configuration is constructed in silico. The cavitand template has a crystal structure, and for the helices, helix D in the wild type bacteriorhodopsin is used to provide the starting coordinates.
The pdb file and topology file are read and a job file is created to obtain the coordinates of the starting configuration.

3) Preminimise

This step is crucial for the cavitein systems since it is responsible for reducing the gap between the template and the helices. The initial arrangement of a molecule can often determine the success or failure of a simulation. It is important that the initial configuration does not contain any high-energy interactions as these may cause instabilities in the simulation. By performing energy minimsation prior to the simulations, these high-energy spots can be reduced.

4) Solvate

In classical mechanics, all atoms are treated explicitly including water. During this step, the cavitein is rotated so that the longest part of its structure is aligned along the Z-axis, and then the water molecules are added to the system.

5) Minimise

Solvation introduces high-energy contacts between the solute and the solvent, and therefore minimisation is required. During this step, the water molecules are permitted to relax around the cavitein molecule.

6) Equilibrate the system

This step enables the system to reach equilibrium. The thermodynamic and structural properties such as energies, temperature and pressure are monitored during the equilibration stage, and should continue until these properties become stable. The initial velocities are generated in this step.
7) Run the simulation

The production phase is when the properties of the system, such as the force on the atoms, are calculated at regular intervals. The new coordinates and energies of the system are saved on a disk file. The analysis step follows the simulations. Properties not calculated during the simulation are carried out (see results in Chapter 3), and the structural changes of the protein are examined. This step also detects any unusual behaviour, indicating problems with the simulation.