

# **Study of Cavitand-Based De Novo Four-Helix Bundle Proteins**

by

**ADAM ROBERT MEZO**

B.Sc. (Honours), Queen's University, 1993

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

in

THE FACULTY OF GRADUATE STUDIES

(Department of Chemistry)

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

January, 1999

© Adam Robert Mezo, 1999

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Chemistry

The University of British Columbia  
Vancouver, Canada

Date Feb 8/99

## Abstract

The protein folding problem is studied by designing, synthesizing and characterizing the structures of de novo four-helix bundles. Each de novo protein (e.g., **27**) consists of four designed peptides (e.g., **26**) attached to a rigid cavitand macrocycle (e.g., **3**) in order to overcome large entropic barriers to folding. We have named these hybrid de novo proteins “*caviteins*” as a result of their constituent parts (*cavitand* + *protein*).

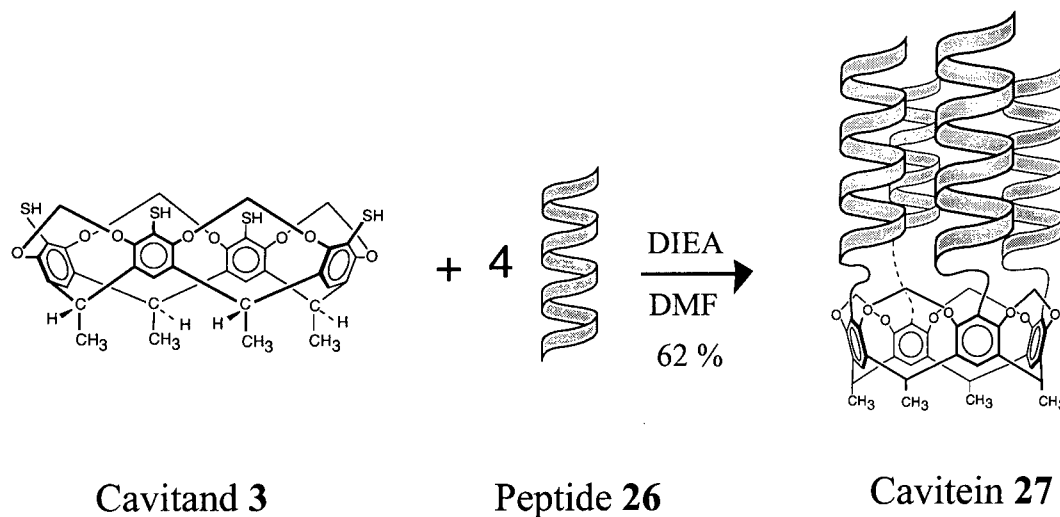
Firstly, a number of different cavitand macrocycles were synthesized bearing reactive thiol or benzylbromide moieties at the cavitand’s “rim” position to allow for peptide attachment. In addition, cavitands bearing methyl or propyl-phosphates moieties at the “foot” position were synthesized in order to study the effect of the cavitand foot on cavitein structure.

As a synthetic model for cavitein synthesis, *N*-activated phenylalanine ethyl ester derivatives were coupled to methyl-footed tetrathiol cavitand **3** in 34-76% yield. Interestingly, the hydrogen bonding characteristics of their amide NHs vary considerably: only the cavitand-phenylalanine hybrid bearing a single methylene linker displays significant hydrogen bonding. This behaviour is attributed to an NH hydrogen bond to a cavitand “bridge” oxygen and the nearby sulfur atom.

The design of each cavitein consists of an *N*-activated amphiphilic amino acid sequence (e.g., **26**) and a cavitand macrocycle (e.g., **3**) bearing sulfur moieties at its rims. The coupling of four activated peptides to each cavitand proceeded efficiently in varying

yields (9-62%). Their structures are highly helical and stable towards guanidine hydrochloride in comparison to peptide **28**, much a result of the cavitand template.

We observe that the cavitand-peptide linker has a profound effect on the structure and oligomeric state of the cavitein. In general, we find that the glycine linker variants possess native-like structural characteristics while the methylene linker variants possess molten globule-like structural characteristics. We attribute the enhanced structural characteristics of the glycine variants to the effect of their added hydrogen bond donors and acceptors to the ends of each helix. The caviteins presented herein represent simple model systems that demonstrate the complexities and subtleties of forces involved in protein folding and allow for further study of the protein folding problem.



Peptide **26**:  $\text{ClCH}_2\text{CO-NH-[EELLKKLEELLKKG]-CONH}_2$

Peptide **28**:  $\text{CH}_3\text{CO-NH-[EELLKKLEELLKKG]-CONH}_2$



## Table of Contents

<b>Abstract.....</b>	<b>ii</b>
<b>List of Schemes.....</b>	<b>ix</b>
<b>List of Figures.....</b>	<b>x</b>
<b>List of Tables.....</b>	<b>xv</b>
<b>List of Abbreviations.....</b>	<b>xvi</b>
<b>Acknowledgments.....</b>	<b>xviii</b>

### Chapter One:        Introduction

A. General Introduction .....	1
i. Why Study Proteins? .....	1
ii. The Protein Folding Problem .....	2
iii. Solving the Protein Folding Problem with De Novo Proteins .....	3
iv. Thesis Goal .....	5
v. Thesis Overview .....	6
B. Protein Structure .....	7
i. General Aspects.....	7
ii. The $\alpha$ -Helix.....	9
iii. $\alpha$ -Helical Motifs.....	11
a. Coiled-Coils.....	12
b. Square Bundles.....	13
iv. Molten Globules vs. Native Structures: Conformational Mobility Within Proteins.....	13
C. Current Views on the Mechanism of Protein Folding .....	14
i. Thermodynamic and Kinetic Control .....	14
ii. Folding Pathway and Funnels .....	16
iii. Chaperones and Assisted Folding .....	18
D. Factors Controlling Protein Folding .....	19
i. General Aspects Controlling Protein Folding .....	20
a. Conformational Entropy .....	20
b. Electrostatic Interactions .....	20
c. Van der Waals Interactions .....	21
c. Hydrogen Bonding .....	22
d. Hydrophobic Effects .....	22
ii. Factors Controlling $\alpha$ -Helical Structure .....	23
a. Helix Propensities and Factors Influencing Helix Propensities .....	24
b. Helix Macrodipole Effects .....	27
c. Helix-Capping Effects .....	28
1. N-Capping .....	29
2. C-Capping .....	30

d. Side-Chain to Side-Chain Interactions Within the Helix .....	31
e. Helix Chain Length .....	32
iii. Factors Determining the Stabilities of Four-Helix Bundles .....	33
a. Inter-Helical Side-Chain Interactions .....	33
1. Packing .....	33
2. Other Side-Chain to Side-Chain Interactions Within Four-Helix Bundles .....	35
b. Helix Orientation .....	35
E. Examples of De Novo Protein Design .....	37
i. De Novo Designed Four-Helix Bundles .....	38
a. DeGrado .....	38
b. Hecht .....	45
c. Baltzer .....	46
d. Dutton .....	46
e. Farid .....	47
ii. Examples of Redesign of Naturally Occurring Sequences Forming Four-Helix Bundles .....	47
a. Alber .....	48
b. Regan .....	48
c. Fairman .....	49
F. Template Assembly in De Novo Protein Design .....	50
i. What is Template Assembly? .....	50
ii. Examples of Template Assembled Four-Helix Bundles .....	52
a. Peptide Templates .....	53
b. Pyridine- and Bipyridine-Based Templates .....	54
c. Porphyrin Templates .....	56
G. Conclusions .....	60
H. References .....	61

## Chapter Two:       **Synthesis of Cavitands Suitable For Incorporation                           Into De Novo Proteins**

A. Introduction to Cavitands .....	70
i. Resorcin[4]arenes .....	70
ii. Methylene-Bridged Resorcin[4]arenes .....	72
B. Results and Discussion .....	74
i. Considerations and Requirements for Peptide Incorporation Onto Cavitands .....	74
ii. Synthesis of Methyl-Footed Tetrathiol <b>3</b> .....	76
iii. Synthesis of Hydroxyl- and Phosphate-Footed Tetrathiols <b>4</b> and <b>5</b> .....	79
iv. Synthesis of Hydroxyl- and Phosphate-Footed Benzylbromides <b>19</b> and <b>21</b> .....	83
C. Conclusion .....	85
D. Experimental .....	87

i. General .....	87
ii. Synthetic Procedures .....	88
E. References .....	102

### **Chapter Three: Model Syntheses for De Novo Proteins**

A. Introduction .....	104
i. Synthesis Using Protected Peptides .....	104
ii. Chemoselective Ligation .....	105
B. Results and Discussion .....	111
i. Synthesis .....	111
ii. Hydrogen Bonding .....	116
C. Conclusion .....	121
D. Experimental .....	122
i. General .....	122
ii. Synthetic Procedures .....	122
E. References .....	133

### **Chapter Four: Design, Synthesis and Characterization of the First Cavitien: A De Novo Designed Four-Helix Bundle Templated by a Cavitand Macrocycle**

A. Introduction .....	135
B. Results and Discussion .....	136
i. Design of Cavitein <b>27</b> .....	136
a. Tetrathiol Cavitands .....	136
b. Peptide Sequence .....	138
ii. Synthesis of Cavitein <b>27</b> .....	140
iii. Structural Characterization of Cavitein <b>27</b> .....	143
a. Far-UV CD Spectrum .....	144
b. Oligomeric State .....	146
c. Effect of Guanidine Hydrochloride and Urea .....	153
d. Effect of Salt and Methanol .....	158
e. Effect of Temperature .....	162
f. Near-UV CD Spectrum .....	163
g. One-Dimensional <sup>1</sup> H NMR Chemical Shift Dispersion .....	165
h. Hydrogen/Deuterium Amide Exchange .....	170
i. ANS Binding .....	172
C. Conclusion .....	174
D. Experimental .....	176
i. General .....	176
ii. Synthesis of Peptides <b>26</b> and <b>28</b> .....	177
iii. Synthesis of Cavitein <b>27</b> .....	179

iv. CD Studies .....	180
v. NMR Studies .....	182
vi. ANS Binding Studies .....	184
vii. Analytical Ultracentrifugation Studies .....	184
E. References .....	188

**Chapter Five:           Effect of Peptide-Cavitand Linkage and Cavitand Foot on Cavitein Structure and Stability.**

A. Introduction .....	192
i. Effect of Cavitand-Peptide Linkages .....	192
ii. Effect of Cavitand Feet .....	195
iii. Nomenclature .....	196
B. Results and Discussion .....	198
i. Synthesis .....	198
a. N-terminal Glycine Cavitein Variants .....	198
b. N-terminal Methylene Cavitein Variants .....	200
c. Phosphate-Footed Cavitein Variants .....	204
ii. Structural Characterization .....	206
a. Comparison of Glycine Variants .....	206
1. Far-UV CD Spectroscopy .....	206
2. Oligomeric State of the Gly Variants .....	209
3. Stabilities Towards Guanidine Hydrochloride .....	215
4. Near-UV CD Spectroscopy .....	219
5. ANS Binding .....	221
6. <sup>1</sup> H NMR Chemical Shift Dispersion .....	222
7. Hydrogen/Deuterium Exchange of N1G/Ar/Me .....	229
b. Comparison of Methylene Variants .....	232
1. Far-UV CD Spectroscopy .....	232
2. Oligomeric State and Stability Towards Guanidine Hydrochloride.....	233
3. Near-UV CD Spectroscopy .....	240
4. ANS Binding .....	241
5. <sup>1</sup> H NMR Chemical Shift Dispersion of N1/Bzl/Me .....	242
c. Comparison of Glycine and Methylene Variants: Particularly N1G/Ar/Me and N4/Ar/Me .....	245
d. Comparison of Phosphate- and Methyl-Footed Cavitein Variants .....	248
1. CD Spectroscopy .....	248
2. Oligomeric State and Stability Towards Guanidine Hydrochloride .....	250
C. Conclusions .....	253
D. Future Work .....	255
E. Experimental .....	258
i. General .....	258

ii.	Synthesis of Cavitand <b>40</b> .....	258
iii.	Synthesis of Peptides .....	259
iv.	Synthesis of Cavitein Variants .....	262
v.	Guanidine Hydrochloride Denaturation Analysis .....	267
vi.	NMR H/D Exchange on N1G/Ar/Me .....	269
vii.	Analytical Ultracentrifugation Studies .....	270
viii.	Computer Modeling .....	281
F.	References .....	282
<b>Appendix A:</b>	The Common Naturally Occurring Amino Acids .....	284
<b>Appendix B:</b>	Amino Acid Sequences of Various Naturally Occurring and De Novo Designed Four-Helix Bundles and Coiled-Coils as Discussed in Chapter One, Section E .....	285

## List of Schemes

Scheme 1.1.	General Illustration of the Protein Folding Problem .....	2
Scheme 1.2.	General Synthesis of a Cavitand-Based De Novo Four-Helix Bundle .....	5
Scheme 2.1.	General Synthesis of Methylene-Bridged Resorcin[4]arenes or Cavitands .....	73
Scheme 2.2.	Synthesis of Bromocavitand <b>8</b> .....	77
Scheme 2.3.	Synthesis of Tetrathiol Cavitand <b>3</b> .....	78
Scheme 2.4.	Synthesis of Hydroxyl-Footed Tetrathiol <b>4</b> .....	80
Scheme 2.5.	Synthesis of Phosphate-Footed Tetrathiol <b>5</b> .....	82
Scheme 2.6.	Synthesis of Hydroxyl- and Phosphate-Footed Cavitands <b>19</b> and <b>21</b> .....	84
Scheme 3.1.	Illustration of the Chemoselective Synthetic Methodology Employed by Schnölzer and Kent .....	106
Scheme 3.2.	DeGrado's Ligation of Unprotected Peptides to a Porphyrin Template Using a Bromoacetyl Group and a Cysteine Residue .....	109
Scheme 3.3.	Synthesis of Thiocresol-Based Control Compounds <b>23a-d</b> .....	113
Scheme 3.4.	Synthesis of 2,6-Dimethoxythiocresol-Based Control Compound <b>24</b> .....	114
Scheme 3.5.	Synthesis of Model De Novo Proteins <b>25a-d</b> .....	115
Scheme 4.1.	Synthesis of Cavitein <b>27</b> . The Linkage Between the Cavitand and the Peptides is "cavitand-(S-CH <sub>2</sub> CO-NH-peptide) <sub>4</sub> " .....	142
Scheme 5.1.	Syntheses of Cavitein Variants Including One ("N1G/Ar/Me"), Two ("N1GG/Ar/Me") and Three ("N1GGG/Ar/Me") N-terminal Glycine Residues .....	199
Scheme 5.2.	Syntheses of Cavitein Variants Including Two ("N2/Ar/Me") and Four ("N4/Ar/Me") Methylene Linker Units .....	202
Scheme 5.3.	Schematic Representation of the Synthesis of Cavitein <b>41</b> ("N1/Bzl/Me") .....	203
Scheme 5.4.	Synthesis of Phosphate-Footed Cavitein <b>42</b> ("N1/Bzl/PO <sub>3</sub> H <sub>2</sub> ") .....	204
Scheme 5.5.	Synthesis of Phosphate-Footed Cavitein <b>44</b> ("C/Bzl/PO <sub>3</sub> H <sub>2</sub> ") .....	205

## List of Figures

Figure 1.1.	(a) An <i>L</i> -Amino Acid, (b) A Polypeptide Chain of <i>L</i> -Amino Acids Showing the Various Dihedral Angles .....	8
Figure 1.2.	Three-Dimensional Representation of the Right-Handed $\alpha$ -Helix .....	10
Figure 1.3.	Two Helices Shown for Simplicity Viewed Down the Helical Axes Packing as in a (A) Coiled-Coil Bundle Illustrating the Left-Handed Superhelical Twist and a (B) Square Bundle .....	12
Figure 1.4.	Simple Illustration Showing Side-Chain Packing in Native-Like and Molten Globule States .....	14
Figure 1.5.	Schematic Representation of (a) Thermodynamic and (b) Kinetic Control of Protein Folding .....	15
Figure 1.6.	Schematic Representation of the "Funnel" Folding Process .....	17
Figure 1.7.	A Consensus Rank Order of Helix Propensities for the 20 Commonly Occurring Amino Acids .....	25
Figure 1.8.	Illustration of the Non-Hydrogen Bonded Atoms in the $\alpha$ -Helix Amide Backbone .....	29
Figure 1.9.	Simplified Two-Dimensional Illustration of (A) Crick's "Knobs-Into-Holes" Packing and (B) Chothia's "Ridges-Into-Grooves" Packing Between Two $\alpha$ -Helices. Side-Chains of Each Helix are Depicted as Open and Closed Circles. Only One Side of the Helices are Shown for Clarity .....	34
Figure 1.10.	Schematic Representation of Parallel and Antiparallel Four-Helix Bundles .....	36
Figure 1.11.	Helical Wheel Diagram of DeGrado's $\alpha_4$ De Novo Four-Helix Bundle Looking Down the Helical Axes. Adjacent Helices Are Drawn Antiparallel With Respect to One Another. Helix Terminating Gly Residues and Loop Regions Are Omitted for Clarity. Amino Acids Are Numbered 1-14 in a N- to C-terminal Direction .....	39
Figure 1.12.	Illustration of DeGrado's Iterative Design Process for $\alpha_4$ .....	42
Figure 1.13.	Three Potential Topologies for DeGrado's Four-Helix Bundle $\alpha_2D$ .....	44
Figure 1.14.	Schematic Representation of Template Assembly in De Novo Protein Design .....	51
Figure 1.15.	Representative Example of Mutter's Strategy in Designing De Novo Four-Helix Bundle TASP's .....	54
Figure 1.16.	Ghadiri's Metal-Templated De Novo Four-Helix Bundle .....	55
Figure 1.17.	Sasaki and Kaiser's Porphyrin-Based De Novo Four-Helix Bundle "Helichrome" .....	57
Figure 1.18.	General Structure of Tetraphenyl Porphyrins Substituted at Different Positions X, Y and Z .....	58
Figure 2.1.	Schematic Illustration of the Interconversion Between the Two ccc Isomers: Boat (left) and Crown (right). $R_2$ Groups Have Been Omitted for Clarity .....	71
Figure 3.1.	Schematic Illustration of the Ligation of Unprotected Peptides to Mutter-Type Peptide Templates by (a) Dawson and Kent,	

	(b) Futaki et. al., and (c) Tuchscherer .....	108
Figure 3.2.	Schematic Illustration of Futaki's Cross-Linked De Novo Four-Helix Bundles Using (a) All Disulfide Bonds and (b) Both Disulfide and Thioether Bonds .....	110
Figure 3.3.	Potential Hydrogen Bonding Interactions in Compound <b>25a</b> .....	120
Figure 4.1.	Primary Sequence of Peptide <b>26</b> Used in the Design of Cavitein <b>27</b> .....	138
Figure 4.2.	Helical Wheel Diagram of Four Molecules of Peptide <b>26</b> with Their Helices Oriented in Parallel and Where the Reader is Looking Down the Helical Axes From C- to N-Termini .....	139
Figure 4.3.	Schematic Representation of the Syntheses of Peptides <b>26</b> and <b>28</b> .....	141
Figure 4.4.	The Synthesis of Cavitein <b>27</b> as Monitored by Analytical Reversed-Phase (C <sub>18</sub> ) HPLC Using a Gradient of 35 - 75% Acetonitrile (with 0.05% TFA) in Water (with 0.1% TFA) Over 40 Minutes: (a) Crude Reaction Mixture After 16 h; (b) Purified Cavitein <b>27</b> .....	143
Figure 4.5.	CD Spectra of Separate 5 $\mu$ M Solutions of Cavitein <b>27</b> (Solid Line) and Peptide <b>28</b> (Dashed Line) in pH 7.0 Phosphate Buffer at 25 $^{\circ}$ C.....	145
Figure 4.6.	Effect of TFE on the Helicity ( $[\theta]_{222}$ ) of Cavitein <b>27</b> in pH 7.0 Phosphate Buffer at 25 $^{\circ}$ C .....	146
Figure 4.7.	Concentration Dependence of $[\theta]_{222}$ for Cavitein <b>27</b> (Circles) and Peptide <b>28</b> (X's) in pH 7.0 Phosphate Buffer at 25 $^{\circ}$ C .....	147
Figure 4.8.	Effect of GuHCl on the Helicity ( $[\theta]_{222}$ ) of Cavitein <b>27</b> at 3 $\mu$ M (Circles) and 175 $\mu$ M (Squares) and Peptide <b>28</b> at 12 $\mu$ M (Diamonds) and 200 $\mu$ M (X's) in pH 7.0 Phosphate Buffer at 25 $^{\circ}$ C .....	149
Figure 4.9.	Hypothetical Representation of Self-Associating Caviteins Without Altering the Structure or Stability of the Helical Bundles .....	150
Figure 4.10.	Sedimentation Equilibrium Concentration Distributions of Cavitein <b>27</b> (Circles) at 30 000 rpm in 50 mM Phosphate Buffer, pH 7, 25 $^{\circ}$ C at (A) 30 $\mu$ M and (B) 300 $\mu$ M. In the Lower Panels, The Solid Lines Represent Theoretical Fits to a Monomer-Dimer Equilibrium. The Upper Panels in Each Figure Represent the Residuals for the Fit .....	152
Figure 4.11.	Effect of Urea on the Helicity ( $[\theta]_{222}$ ) of Cavitein <b>27</b> in pH 7.0 Phosphate Buffer at 25 $^{\circ}$ C .....	154
Figure 4.12.	Sedimentation Equilibrium Analysis of Cavitein <b>27</b> in the Presence of 6.0 M GuHCl at 25 $^{\circ}$ C. In the Lower Panel, The Solid Line Represents the Theoretical Fit to a Single Non-Interacting Species. The Upper Panel Represents the Residuals for the Fit .....	156
Figure 4.13.	Effect of pH on the Helicity ( $[\theta]_{222}$ ) of Cavitein <b>27</b> at 25 $^{\circ}$ C .....	159
Figure 4.14.	Sedimentation Equilibrium Analysis of Cavitein <b>27</b> in the Presence of 2 M NaCl. In the Lower Panel, The Solid Line Represents the Theoretical Fit to a Single Non-Interacting Species. The Upper Panel Represents the Residuals for the Fit. Deviations in the Residuals are the Result of Window Imperfections (as Interpreted by the Kelly Research Group) .....	160
Figure 4.15.	Sedimentation Equilibrium Analysis of Cavitein <b>27</b> in the Presence of 10% Methanol. In the Lower Panel, The Solid Line Represents the Theoretical Fit to a Single Non-Interacting Species. The Upper Panel	



	Represents the Residuals for the Fit .....	161
Figure 4.16.	Effect of Temperature on the Helicity ( $[\theta]_{222}$ ) of Cavitein <b>27</b> in pH 7.0 Phosphate Buffer .....	163
Figure 4.17.	Near-UV CD Spectra of Cavitein <b>27</b> and Peptide <b>28</b> in pH 7.0 Phosphate Buffer at 25 °C .....	164
Figure 4.18.	Full 500 MHz $^1\text{H}$ NMR Spectrum of Cavitein <b>27</b> in 10% $\text{D}_2\text{O}$ , 45 mM Phosphate Buffer, pH 7.0, 25 °C .....	166
Figure 4.19.	Expansions of the Amide Regions of 500 MHz $^1\text{H}$ NMR Spectra of Cavitein <b>27</b> (~ 0.2 mM) and Peptide <b>28</b> at Different Concentrations at 25 °C, in Phosphate Buffer at pH 7.0 in the Presence of 10% $\text{D}_2\text{O}$ .....	167
Figure 4.20.	Expansions of the Aliphatic Regions of 500 MHz $^1\text{H}$ NMR Spectra of Cavitein <b>27</b> (~ 0.2 mM) and Peptide <b>28</b> at Different Concentrations at 25 °C, in Phosphate Buffer at pH 7.0 in the Presence of 10% $\text{D}_2\text{O}$ .....	168
Figure 4.21.	Stack Plot of 400 MHz $^1\text{H}$ NMR Spectra Illustrating the Time Dependent Amide H/D Exchange of Cavitein <b>27</b> in 50 mM pD 5.0 $\text{CD}_3\text{COOD/NaOD}$ Buffer at 25 °C .....	171
Figure 4.22.	Fluorescence Emission Spectra of 1 $\mu\text{M}$ ANS in the Presence of 95% Ethanol, 100% Methanol, 50 $\mu\text{M}$ Cavitein <b>27</b> , 50 $\mu\text{M}$ and 200 $\mu\text{M}$ (Indistinguishable) Peptide <b>28</b> and ANS Alone at 25 °C in pH 7.0, 50 mM Phosphate Buffer .....	173
Figure 4.23.	Schematic Representation of the <i>FastMoc</i> <sup>TM</sup> Protocol on the ABI 431A Peptide Synthesizer .....	178
Figure 5.1.	Structures of Three Different Templates Used by Fairlie in Evaluating the Effect of Different Templates on the Structures and Stabilities of De Novo Four-Helix Bundles. Positions Labeled “X” Represent Peptide Attachment Sites .....	193
Figure 5.2.	Schematic Representation of the Syntheses of Activated Peptides With Two (“N2Br”) and Four (“N4Br”) Methylene Linker Units .....	201
Figure 5.3.	Far-UV CD Spectra of Caviteins N1/Ar/Me, N1G/Ar/Me, N1GG/Ar/Me, N1GGG/Ar/Me in pH 7.0 Phosphate Buffer at 25 °C. Each Cavitein is at a Concentration of ~ 30 $\mu\text{M}$ .....	207
Figure 5.4.	Guanidine Hydrochloride-Induced Denaturation Curve for N1G/Ar/Me at 3 and 30 $\mu\text{M}$ at pH 7 and 25 °C .....	212
Figure 5.5.	Guanidine Hydrochloride-Induced Denaturation Curve for N1GG/Ar/Me at 3 and 30 $\mu\text{M}$ at pH 7 and 25 °C .....	213
Figure 5.6.	Guanidine Hydrochloride-Induced Denaturation Curve for N1GGG/Ar/Me at 3 and 30 $\mu\text{M}$ at pH 7 and 25 °C .....	214
Figure 5.7.	Guanidine Hydrochloride-Induced Denaturation Curves for N1/Ar/Me, N1G/Ar/Me, N1GG/Ar/Me and N1GGG/Ar/Me at 30 $\mu\text{M}$ in pH 7.0 Phosphate Buffer at 25 °C. Error Bars Have Been Omitted For Clarity .....	216
Figure 5.8.	Near-UV CD Spectra of Gly-Cavitein Variants in pH 7.0 Phosphate Buffer at 25 °C .....	220
Figure 5.9.	Fluorescence Emission Spectra of 1 $\mu\text{M}$ ANS in the Presence of 95% Ethanol, 100% Methanol and 50 $\mu\text{M}$ solutions of N1/Ar/Me, N1G/Ar/Me, N1GG/Ar/Me and N1GGG/Ar/Me at 25 °C in pH 7.0 Phosphate Buffer .....	222

Figure 5.10.	500 MHz $^1\text{H}$ NMR Spectrum of N1G/Ar/Me at $\sim 300\ \mu\text{M}$ in 10% $\text{D}_2\text{O}$ , 45 mM Phosphate Buffer, pH 7.0, 25 $^\circ\text{C}$ .....	223
Figure 5.11.	500 MHz $^1\text{H}$ NMR Spectrum of N1GG/Ar/Me at $\sim 300\ \mu\text{M}$ in 10% $\text{D}_2\text{O}$ , 45 mM Phosphate Buffer, pH 7.0, 25 $^\circ\text{C}$ .....	224
Figure 5.12.	500 MHz $^1\text{H}$ NMR Spectrum of N1GGG/Ar/Me at $\sim 300\ \mu\text{M}$ in 10% $\text{D}_2\text{O}$ , 45 mM Phosphate Buffer, pH 7.0, 25 $^\circ\text{C}$ .....	225
Figure 5.13.	Stacked Expansions of the Amide Regions of 500 MHz $^1\text{H}$ NMR Spectra of the Gly Variants ( $\sim 0.3\ \text{mM}$ ) at 25 $^\circ\text{C}$ , in 45 mM Phosphate Buffer at pH 7.0 in the Presence of 10% $\text{D}_2\text{O}$ .....	226
Figure 5.14.	Stacked Expansions of the Aliphatic Regions of 500 MHz $^1\text{H}$ NMR Spectra of the Gly Variants ( $\sim 0.3\ \text{mM}$ ) at 25 $^\circ\text{C}$ , in 45 mM Phosphate Buffer at pH 7.0 in the Presence of 10% $\text{D}_2\text{O}$ .....	227
Figure 5.15.	Stack Plot of 500 MHz $^1\text{H}$ NMR Spectra Illustrating the Time Dependent Amide H/D Exchange of N1G/Ar/Me in 50 mM pD 5.0 $\text{CD}_3\text{COOD/NaOD}$ Buffer at 25 $^\circ\text{C}$ .....	230
Figure 5.16.	Far-UV CD Spectra of Caviteins N1/Ar/Me, N2/Ar/Me, N4/Ar/Me and N1/Bzl/Me in pH 7.0 Phosphate Buffer at 25 $^\circ\text{C}$ . Each Cavitein is at a Concentration of $\sim 30\ \mu\text{M}$ .....	233
Figure 5.17.	Guanidine Hydrochloride-Induced Denaturation Curves for N1/Bzl/Me at 3 and 30 $\mu\text{M}$ at pH 7 and 25 $^\circ\text{C}$ .....	235
Figure 5.18.	Guanidine Hydrochloride-Induced Denaturation Curves for 30 $\mu\text{M}$ Solutions of N1/Ar/Me, N2/Ar/Me, N4/Ar/Me and N1/Bzl/Me at pH 7 and 25 $^\circ\text{C}$ . Error Bars Have Been Omitted For Clarity .....	236
Figure 5.19.	Illustration of the MM2-Predicted Conformations of S-Methyl Derivatives of (A) Cavitand <b>3</b> and (B) Cavitand <b>40</b> . Part (C) Illustrates that the Peptides are Likely Positioned Differently When Attached to Cavitands <b>3</b> and <b>40</b> .....	239
Figure 5.20.	Near-UV CD Spectra of the Methylene Cavitein Variants at pH 7.0 in 50 mM Phosphate Buffer and 25 $^\circ\text{C}$ .....	240
Figure 5.21.	Fluorescence Emission Spectra of 1 $\mu\text{M}$ ANS in the Presence of 95% Ethanol, 100% Methanol and 50 $\mu\text{M}$ solutions of N1/Ar/Me, N2/Ar/Me, N4/Ar/Me and N1/Bzl/Me at 25 $^\circ\text{C}$ in pH 7.0 Phosphate Buffer .....	241
Figure 5.22.	500 MHz $^1\text{H}$ NMR Spectrum of N1/Bzl/Me at $\sim 300\ \mu\text{M}$ in 10% $\text{D}_2\text{O}$ , 45 mM Phosphate Buffer, pH 7.0, 25 $^\circ\text{C}$ .....	243
Figure 5.23.	Expansions of the (A) Amide and (B) Aliphatic Regions of 500 MHz $^1\text{H}$ NMR Spectra of N1/Ar/Me and N1/Bzl/Me ( $\sim 0.3\ \text{mM}$ ) at 25 $^\circ\text{C}$ , in 45 mM Phosphate Buffer at pH 7.0 in the Presence of 10% $\text{D}_2\text{O}$ .....	244
Figure 5.24.	Illustration of the Different Linker Groups in N1G/Ar/Me and N4/Ar/Me .....	246
Figure 5.25.	Far-UV CD Spectra of 30 $\mu\text{M}$ Solutions of N1/Bzl/Me and N1/Bzl/ $\text{PO}_3\text{H}_2$ in 50 mM Phosphate Buffer at pH 7 and 25 $^\circ\text{C}$ .....	249
Figure 5.26.	Near-UV CD Spectra of 30 $\mu\text{M}$ Solutions N1/Bzl/Me and N1/Bzl/ $\text{PO}_3\text{H}_2$ in 50 mM Phosphate Buffer at pH 7 and 25 $^\circ\text{C}$ .....	250
Figure 5.27.	Guanidine Hydrochloride-Induced Denaturation Curves for 30 $\mu\text{M}$ Solutions of N1/Bzl/Me and N1/Bzl/ $\text{PO}_3\text{H}_2$ at pH 7 and 25 $^\circ\text{C}$ .....	251
Figure 5.28.	Sedimentation Equilibrium Analysis of N1G/Ar/Me at 3 $\mu\text{M}$ . In	

	the Lower Panel, The Solid Line Represents the Theoretical Fit to a Single Non-Interacting Species. The Upper Panel Represents the Residuals for the Fit .....	271
Figure 5.29.	Sedimentation Equilibrium Analysis of N1G/Ar/Me at 30 $\mu$ M. In the Lower Panel, The Solid Line Represents the Theoretical Fit to a Single Non-Interacting Species. The Upper Panel Represents the Residuals for the Fit .....	272
Figure 5.30.	Sedimentation Equilibrium Analysis of N1GG/Ar/Me at 30 $\mu$ M. In the Lower Panel, The Solid Line Represents the Theoretical Fit to a Single Non-Interacting Species. The Upper Panel Represents the Residuals for the Fit .....	273
Figure 5.31.	Sedimentation Equilibrium Analysis of N1GG/Ar/Me at 30 $\mu$ M in the Presence of 7.2 M GuHCl. In the Lower Panel, The Solid Line Represents the Theoretical Fit to a Single Non-Interacting Species. The Upper Panel Represents the Residuals for the Fit .....	274
Figure 5.32.	Sedimentation Equilibrium Analysis of N1GGG/Ar/Me at 30 $\mu$ M. In the Lower Panel, The Solid Line Represents the Theoretical Fit to a Single Non-Interacting Species. The Upper Panel Represents the Residuals for the Fit .....	275
Figure 5.33.	Sedimentation Equilibrium Analysis of N1GGG/Ar/Me at 300 $\mu$ M. In the Lower Panel, The Solid Line Represents the Theoretical Fit to a Single Non-Interacting Species. The Upper Panel Represents the Residuals for the Fit.....	276
Figure 5.34.	Sedimentation Equilibrium Analysis of N2/Ar/Me at 30 $\mu$ M. In the Lower Panel, The Solid Line Represents the Theoretical Fit to a Single Non-Interacting Species. The Upper Panel Represents the Residuals for the Fit .....	277
Figure 5.35.	Sedimentation Equilibrium Analysis of N4/Ar/Me at 30 $\mu$ M. In the Lower Panel, The Solid Line Represents the Theoretical Fit to a Single Non-Interacting Species. The Upper Panel Represents the Residuals for the Fit .....	278
Figure 5.36.	Sedimentation Equilibrium Analysis of N1/Bzl/Me at 30 $\mu$ M. In the Lower Panel, The Solid Line Represents the Theoretical Fit to a Single Non-Interacting Species. The Upper Panel Represents the Residuals for the Fit .....	279
Figure 5.37.	Sedimentation Equilibrium Analysis of N1/Bzl/Me at 300 $\mu$ M. In the Lower Panel, The Solid Line Represents the Theoretical Fit to a Single Non-Interacting Species. The Upper Panel Represents the Residuals for the Fit .....	280
Figure 5.38.	Analysis of N1/Bzl/PO <sub>3</sub> H <sub>2</sub> by Sedimentation Equilibrium at 30 $\mu$ M. In the Lower Panel, The Solid Line Represents the Theoretical Fit to a Single Non-Interacting Species. The Upper Panel Represents the Residuals for the Fit .....	281

## List of Tables

Table 1.1.	DeGrado's Series of De Novo Designed Four-Helix Bundle- Forming Peptides .....	41
Table 3.1.	Amide N-H $^1\text{H}$ NMR Chemical Shifts and IR Stretching Frequencies for De Novo Protein Control and Model Compounds .....	118
Table 5.1.	Nomenclature Used in Describing the Nine Caviteins Presented in Chapter Five .....	197
Table 5.2.	Experimentally Estimated Molecular Weights Determined by Sedimentation Equilibrium For All Cavitein Variants at 25 °C in 50 mM pH 7 Phosphate Buffer, Unless Otherwise Noted .....	210
Table 5.3.	Guanidine Hydrochloride-Induced Denaturation Data Calculated for the Gly-Cavitein Variants .....	217
Table 5.4.	Tabulated Data From the Amide H/D Exchange Experiment on N1G/Ar/Me in 50 mM pD 5.0 $\text{CD}_3\text{COOD/NaOD}$ Buffer at 25 °C .....	231
Table 5.5.	Guanidine Hydrochloride-Induced Denaturation Data Calculated for the Methylene Cavitein Variants .....	237
Table 5.6.	Guanidine Hydrochloride-Induced Denaturation Data Calculated for the Cavitand Foot Cavitein Variants .....	251
Table 5.7.	Experimental Details for Sedimentation Equilibrium Experiments Reported in Chapter Five .....	270

## List of Abbreviations

Anal.	microanalysis
ANS	1-anilinonaphthalene-8-sulfonate
Bn	benzyl [C <sub>6</sub> H <sub>5</sub> CH <sub>2</sub> -]
calcd	calculated
CD	circular dichroism
COSY	correlation (NMR) spectroscopy
$\Delta\delta$	change in chemical shift
$\delta$	chemical shift
d	day(s)
DBU	1,8-diazabicyclo[5.4.0]undec-7-ene
DCI	desorption chemical impact ( <i>mass spectrometry or spectrum</i> )
DCM	dichloromethane
DDP	di- <i>t</i> -butyl <i>N,N</i> -diethylphosphoramidite
DHB	2,5-dihydroxybenzoic acid
DIEA	diisopropylethylamine
DMF	<i>N,N</i> -dimethylformamide
DMA	<i>N,N</i> -dimethylacetamide
DMSO	dimethyl sulfoxide
dec	decomposed
equiv.	equivalents
ESI	electrospray ionization ( <i>mass spectrometry or spectrum</i> )
ESMS	electrospray ionization mass spectrometry or spectrum
EtOAc	ethyl acetate
EtOH	ethanol
FID	free induction decay
GuHCl	guanidine hydrochloride
h	hour(s)
HBTU	2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HPLC	high performance/pressure liquid chromatography
HOBt	1-hydroxybenzotriazole
HRMS	high resolution mass spectrometry or spectrum
IR	infrared (spectroscopy)
<i>J</i>	coupling constant
LSIMS	liquid secondary ionization mass spectrometry or spectrum
<i>m/z</i>	mass-to-charge ratio
M	parent mass ( <i>mass spectra</i> ) or moles per litre ( <i>concentration</i> )
MALDI	matrix laser desorption ionization mass spectrometry or spectrum
Me	methyl [CH <sub>3</sub> -]
MeOH	methanol
min	minutes
mp	melting point
MS	mass spectrometry or spectrum

3-NBA	3-nitrobenzylalcohol
<i>n</i> -BuLi	<i>n</i> -butyllithium
NBS	<i>N</i> -bromosuccinimide
NMP	<i>N</i> -methylpyrrolidone
NMR	nuclear magnetic resonance (spectroscopy)
1D	one-dimensional
<i>p</i>	para
Ph	phenyl (-C <sub>6</sub> H <sub>5</sub> )
ppm	parts per million
rel intensity	relative intensity
rpm	revolutions per minute
rt	room temperature
[ $\theta$ ]	molar ellipticity per residue
<i>t</i>	tertiary
<i>t</i> <sub>1/2</sub>	half-life
TBDPS	<i>tert</i> -butyl diphenylsilyl
TFA	trifluoroacetic acid
TFE	trifluoroethanol
THF	tetrahydrofuran

In addition, all one-letter and three-letter codes for the twenty commonly occurring amino acids can be found in Appendix A.

## Acknowledgments

I would first like to thank my supervisor, John Sherman, for his guidance throughout my graduate studies at UBC. He has taught me a great deal about chemistry and scientific thinking in general. On that note, I would also like to thank my undergraduate supervisor, Bob Lemieux at Queen's University, for getting me "excited" about scientific research. His excitement for science triggered my excitement which eventually prompted me to pursue graduate studies.

Thanks to my family for their limitless and unconditional support and encouragement over the last 5+ years. Also thanks to Kami for her support and dealing with me when I get into "science/thesis-mode."

I would like to thank both present and former Sherman group members, especially Ashley, Bob, Nav, Doug and Ayub, for keeping some of my endeavors at UBC "non-scientific."

This thesis was proof-read by many, and I then I would also like to thank them for taking time out of their schedule (at most inopportune moments, I'm sure) to read it: Ashley Causton, Diana Wallhorn, and Ayub Jasat.

This thesis incorporated quite a diversity of biophysical techniques which required expertise which were beyond myself and our research group. I would therefore like to thank G. Connolly and L. McIntosh for acquiring and processing the 1D NMR dispersion spectra, L. Burtnick for help with his fluorimeter, F. Cossell for help with their CD temperature apparatus, and Jeff Kelly, Ed Koepf and Mike Petrowski (Scripps Research Institute) for acquiring and interpreting all of the analytical ultracentrifuge data under very short notice.

# **Chapter One: Introduction**

## **A. General Introduction**

### **i. Why Study Proteins?**

Proteins are essential for life on earth as they affect nearly all the processes found in both animal and plant life. In humans, for example, proteins are vital for the transport of oxygen to the lungs, elimination of infectious agents via the immune system and the production of our skin. In addition, many of the biochemical processes found in the body are catalyzed by enzymes, the vast majority of which are also proteins, without which these processes would be too slow to sustain life. Proteins are so important to life itself that a better understanding of them would aid in solving many of the biochemical and medical problems that exist today.

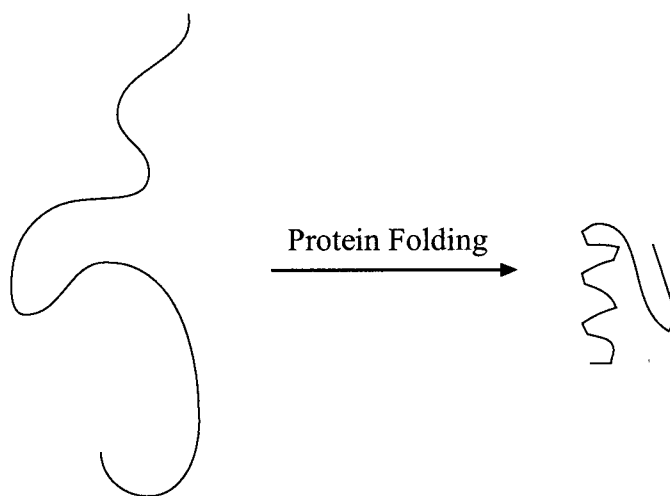
The three-dimensional structure of a protein determines its function whether it be an antibody recognizing an antigen or collagen forming skin and bones. Although proteins are an extremely versatile and diverse set of molecules, they are simply biopolymers composed of a linear chain of amino acids. This linear sequence is responsible for the three-dimensional structures of proteins.



## ii. The Protein Folding Problem

Several decades ago, Anfinsen demonstrated the reversible folding and unfolding of ribonuclease and concluded that all the information required to fold any given protein is “coded” in its amino acid sequence.<sup>1</sup> This led to the notion that one could predict the three-dimensional structure of a protein solely from its amino acid sequence. Scientists have yet to “crack the code” and make such predictions with any reliability. This is one of the major unsolved problems in biochemistry and has come to be known as the *protein folding problem*. This is illustrated schematically in Scheme 1.1 where a flexible linear chain folds into a compact structure.

**Scheme 1.1.** General Illustration of the Protein Folding Problem.



The folding of proteins is a truly remarkable event considering the astronomical number of allowable conformations in the polypeptide chain. To put this in perspective, consider a polypeptide chain consisting of 100 amino acids: using a conservative estimate of 10 possible conformations per residue, the polypeptide chain has  $10^{100}$  total possible conformations. If the polypeptide chain converts between conformations in the shortest possible time,  $10^{-13}$  seconds for example for bond rotation, it would take  $10^{77}$  years to sample all the possible conformations. Yet natural proteins fold into a well-defined three-dimensional structure usually on the order of  $10^{-1}$  to  $10^3$  seconds.<sup>2</sup> This phenomenon has come to be known as “Levinthal’s paradox,” an argument first made in the 1960s.<sup>3</sup> It became evident that protein folding does not proceed randomly but by some directed process derived from its amino acid sequence. A better understanding of the mechanism by which proteins fold would give scientists the tools to modify existing proteins or, more ambitiously, to design new proteins with novel functions.

### iii. Solving the Protein Folding Problem with De Novo Proteins

The study of the protein folding problem can be approached in many different ways. For example, computer modeling can attempt to quantify all of the interactions present in protein folding and make structural predictions solely from amino acid sequences. Although a few recent models have found some success,<sup>4</sup> no single computer program has solved the problem.

The advent of site-directed mutagenesis has brought forth another method of studying the protein folding problem.<sup>5</sup> This technique allows scientists to readily modify select amino acids within a natural protein sequence in order to study its effect on protein folding. However, naturally occurring proteins typically possess hundreds of amino acids and thus assigning

contributions from individual residues toward the overall fold of the protein is quite difficult. Alternatively, some researchers study isolated protein fragments or peptides in order to better understand the forces involved in governing protein folding. However, the folding of isolated model peptides represents only a small step in the folding of a natural protein.

An alternative approach to the study of protein folding is to design, synthesize and study *de novo* proteins; that is, proteins designed from scratch.<sup>6</sup> Unlike naturally occurring proteins, *de novo* proteins can be designed to be much smaller than natural proteins yet maintain many of the features responsible for the folding of natural proteins. Due to their relatively small size, *de novo* proteins may allow for a simpler elucidation of specific interactions contributing to protein folding and thus represent simple model systems for the study of protein folding.

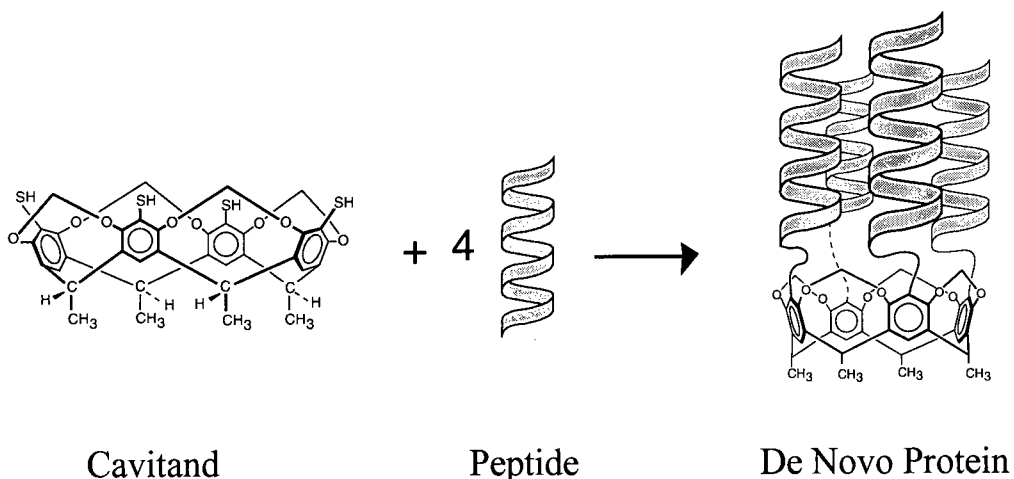
It has been shown that proteins can have very similar structures with fewer than 10% of the same residues at similar positions along the amino acid sequence.<sup>7</sup> Such a degeneracy in the folding “code” suggests that there are fundamental forces acting upon the amino acid sequence that remain constant over large sequence variations.<sup>8</sup> Testing our knowledge of these fundamental forces by designing proteins from first principles is a viable route to the study of the forces involved in protein folding.

In addition, *de novo* proteins have the potential to perform novel enzyme-like functions such as the catalysis of chemical reactions.<sup>9,10,11</sup> A thorough understanding of the forces controlling protein folding would enable “protein designers” to accomplish such a goal.

iv. Thesis Goal

In this thesis, protein folding will be explored with the design, synthesis and characterization of de novo four-helix bundles. Our model will consist of four  $\alpha$ -helical peptide strands that are covalently linked to a rigid organic macrocycle called a *cavitand* (Scheme 1.2).

**Scheme 1.2.** General Synthesis of a Cavitand-Based De Novo Four-Helix Bundle.



Isolated  $\alpha$ -helices have been well-studied and many of the factors controlling  $\alpha$ -helix formation have been determined (Section D.ii of this chapter). However, the factors controlling tertiary structure or helical bundle formation are not as well understood. We are interested in this aspect of protein structure and to this end will model de novo four-helix bundles. As will be described later in Section B, four-helix bundles are a common tertiary structure found in nature and also represent one of the simplest known tertiary structures. Four-helix bundles thus represent an attractive target in a first step towards de novo protein design.

As mentioned above, we intend to model the four-helix bundle by linking peptide strands to a cavitand template. Cavitands are defined as rigid organic macrocycles possessing an enforced cavity. By linking peptide fragments to this rigid macrocycle, large entropic barriers to folding may be bypassed giving rise to a template-stabilized four-helix bundle. We plan to design, synthesize and study the properties of various de novo four-helix bundles which may aid in solving the protein folding problem. We have named this new family of de novo proteins “*caviteins*” (*cavitand* + *protein*).

## v. Thesis Overview

Chapter One will provide some background information on protein folding, introduce some of the factors controlling protein folding and mention some examples of de novo protein design. In addition, the concept of *template assembly* will be introduced including recent examples of template assembly in de novo protein design. Chapter Two will discuss the preparation of cavitands suitable for incorporation into de novo proteins. Chapter Three will discuss some model reactions linking single amino acids to a cavitand template. These model studies demonstrate the synthetic methodologies used to incorporate the peptide strands onto the cavitands. Chapter Four will discuss the design, synthesis and characterization of the first de novo four-helix bundle based on a cavitand macrocycle. And finally, Chapter Five will present various four-helix bundle analogues of the prototype presented in Chapter Four. These studies illustrate the dramatic effect of the cavitand-peptide linkage on four-helix bundle structure and stability and lead to several conclusions regarding de novo protein design and protein folding.

Below, the basic elements of protein structure are presented, followed by a discussion of protein folding.

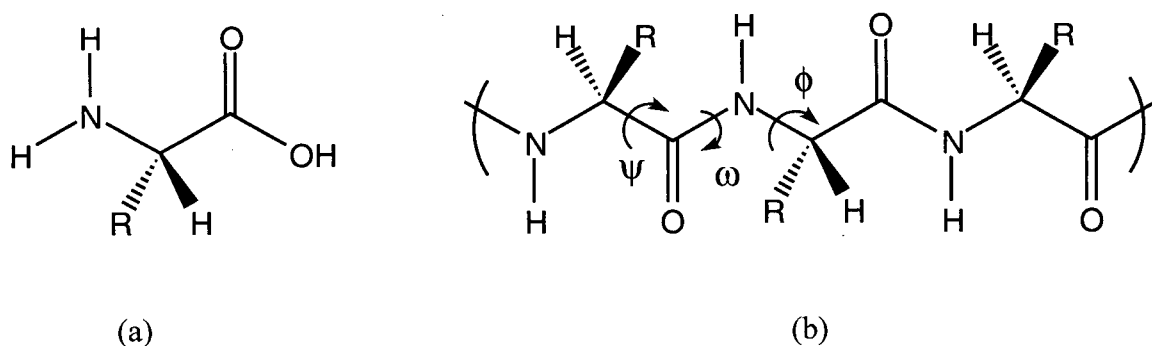
## B. Protein Structure

### i. General Aspects

A protein's linear sequence of amino acids is termed its *primary structure*. Evident from Figure 1.1 is that an amide bond is formed between each amino acid where the NH-CO dihedral angle,  $\omega$ , is typically planar with a value of  $180^\circ$ .<sup>12</sup> This provides some structural rigidity to an otherwise flexible chain of amino acids. As well, each amino acid possesses a hydrogen bonding donor (amide NH) and acceptor (amide CO) which aid in defining the overall structure of the amino acid sequence. Helpful in describing the conformations of the polypeptide backbone are the backbone dihedral angles  $\phi$  and  $\psi$  corresponding to the HN- $C_\alpha$  and  $C_\alpha$ -CO dihedral angles, respectively. The dihedral angles of side-chains are denoted by  $\chi_1$ ,  $\chi_2$ ,  $\chi_3$  and so on, starting from the atom nearest the amide backbone. The side-chain atoms are identified as follows: the carbon within the backbone is denoted as  $C_\alpha$  and others extending away from the chain are denoted as  $C_\beta$ ,  $C_\gamma$ ,  $C_\delta$  and so on. The side-chains ('R' groups) play pivotal roles in protein folding and determining the final three-dimensional structure. From a flexible glycine residue ( $R = H$ ) to a sterically more demanding valine residue ( $R = \text{isopropyl}$ ), proteins can use any of the twenty commonly occurring amino acid side-chains to influence the overall three-dimensional structure.

The twenty natural amino acids can be written in short-form using either one-letter or three-letter codes. Both of these codes will be used interchangeably throughout this thesis to save space. For reference, the one-letter and three-letter codes along with the side-chains of all twenty commonly occurring amino acids are listed in Appendix A.

**Figure 1.1.** (a) An *L*-Amino Acid, (b) A Polypeptide Chain of *L*-Amino Acids Showing the Various Dihedral Angles .



Amino acids are typically written left to right in an “amino to carboxylate” fashion as depicted above in Figure 1.1. Likewise, peptide and protein chemists typically refer to a protein’s amino acid sequence from its N- to C-terminus.

The polypeptide chain can fold into a regular pattern of conformations forming structures such as  $\alpha$ -helices or  $\beta$ -sheets. Such higher levels of structure are termed *secondary structure*. The folding of various forms of secondary structure forms a protein’s overall topology and is termed its *tertiary structure*. Aggregation of different polypeptide tertiary structures results in *quaternary structure*. The next section introduces  $\alpha$ -helices and will be the only type of secondary structure discussed in detail due to its relevance to this thesis.

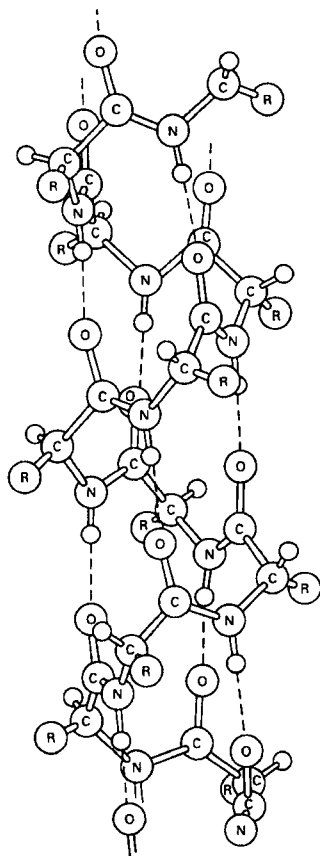
## ii. The Alpha-Helix

The  $\alpha$ -helix is one of the most common forms of secondary structure although helices themselves may only account for about one-third of all residues found in proteins.<sup>13</sup> The  $\alpha$ -helix was initially proposed by Pauling and coworkers<sup>14</sup> and later confirmed by X-ray crystallography.<sup>15</sup> The  $\alpha$ -helix is a right-handed twist of the polypeptide chain and consists of 3.6 residues per turn. Such a twist translates into a pitch of 5.41 Å with typical  $\phi$  and  $\psi$  angles of about  $-60^\circ$  and  $-40^\circ$ , respectively (Figure 1.2).

Also characteristic of the  $\alpha$ -helix is the hydrogen bonding pattern of the backbone amide bonds. The carbonyl of residue number  $i$  forms a hydrogen bond with the NH of residue number  $i + 4$ . These hydrogen bonds are near-linear, parallel to the helical axis and have a typical N-O and H $\cdots$ O distances of 2.86 Å and 1.93 Å, respectively.<sup>2</sup> One consequence of these highly directional hydrogen bonds is that the helix can be considered a macrodipole.<sup>16</sup> The N- and C-termini have approximate charges of +0.5 and -0.5, respectively. The consequence of this effect towards helix stability will be discussed later (Section D.ii.b). Although other types of helices exist (i.e.,  $3_{10}$ -helix), the term “helix” will be used interchangeably with “ $\alpha$ -helix” throughout this thesis, and is intended to refer to right-handed  $\alpha$ -helices as depicted in Figure 1.2.



**Figure 1.2.** Three-Dimensional Representation of the Right-Handed  $\alpha$ -Helix.



One might initially attribute the frequency with which helices occur in proteins to this near-ideal  $i \cdots i + 4$  hydrogen bonding pattern. Indeed, this hydrogen bonding most definitely stabilizes  $\alpha$ -helical structure. However, one consequence of the typical  $\alpha$ -helical  $i \cdots i + 4$  hydrogen bonding pattern is that the first and last four residues in the helix do not participate in this type of hydrogen bonding and consist of “unsatisfied” amide NHs and COs. Furthermore, if we consider that the average length of  $\alpha$ -helices found in proteins is only 12 amino acids in length,<sup>17</sup> one realizes that only one-third of the residues in an average helix fully participate in this type of hydrogen bonding. Hence, there must be many more factors responsible for the stability of  $\alpha$ -helices and these factors will be discussed in Section D.ii.

### iii. $\alpha$ -Helical Motifs

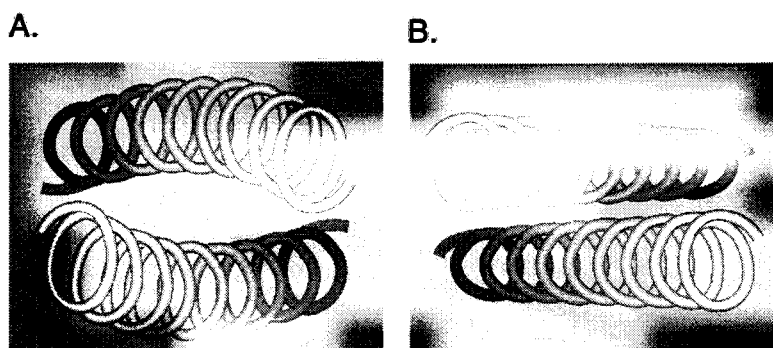
$\alpha$ -Helices are known to form a variety of structures in nature.<sup>18</sup> Such structures include coiled-coils, DNA binding motifs (excluding coiled-coil DNA binding motifs), zinc fingers, helix-loop-helix calcium binding motifs, and four-helix bundles. The zinc finger motifs incorporate a mixture of  $\alpha$  and  $\beta$  structures so as to include a zinc binding site. The DNA binding motifs typically contain  $\alpha$ -helices packed at angles such that a significant groove exists for DNA binding. Helix-loop-helix motifs typically bind calcium in the loop region while flanked by several helices.<sup>18</sup>

Helical bundles comprise a small portion of possible helical motifs and are found naturally with great diversity. For the purposes of this thesis, only four-helix bundles will be discussed in detail. Four-helix bundles fall into one of two main classes.<sup>19</sup> The first is typified by near-parallel helix-helix interactions while the second contains a mixture of parallel and perpendicular helix interactions. The parallel class can be further divided into two subclasses depending on the bend of each helix.<sup>20</sup> These two sub-classes of four-helix bundles are termed *coiled-coils* and *square bundles*.

a. Coiled-Coils

Two-, three-, and four-stranded coiled-coils were first proposed by Crick<sup>21</sup> and are characterized by right-handed  $\alpha$ -helices wrapping around one another in a left-handed superhelix with a crossing angle of about  $20^\circ$  between helices.<sup>22</sup> Such a superhelix allows the hydrophobic surfaces of each helix to be in continuous contact with an adjacent helix (Figure 1.3). In addition, the left-handed superhelix gives rise to a repeat of 3.5 residues per helical turn (as opposed to 3.6 residues per turn for  $\alpha$ -helices in square bundles). As a result, coiled-coils can be characterized by a *heptad repeat* [**a****b****c****d****e****f****g**]<sub>n</sub> in their amino acid sequence such that every seventh amino acid is in an identical position with respect to the helical axis. Typically, residues **a** and **d** are hydrophobic within coiled-coils. Coiled-coils, predominantly consisting of two parallel strands, are important motifs found in nature for fibrous proteins, DNA binding, and countless other functions.<sup>18</sup> One example of a naturally occurring four-helix bundle with a coiled-coil geometry is the protein ROP.<sup>23</sup>

**Figure 1.3.** Two Helices Shown for Simplicity Viewed Down the Helical Axes Packing as in a (A) Coiled-Coil Bundle Illustrating the Left-Handed Superhelical Twist and a (B) Square Bundle.



## b. Square Bundles

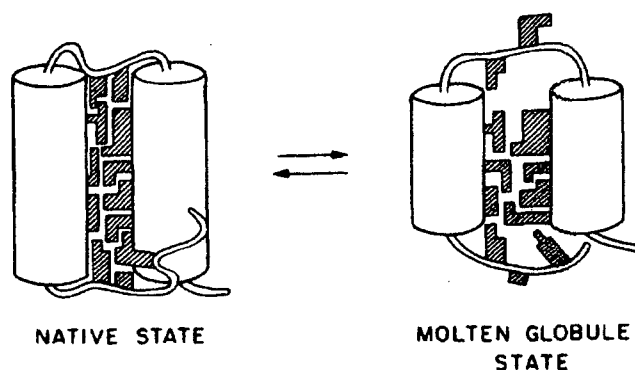
Square four-helix bundles also consist of four helices tilted at 20° with respect to one another. However, unlike coiled-coils, the helices do not bend significantly which results in helices that are splayed out from their point of closest contact (Figure 1.3). This divergence gives rise to a binding pocket between the helices, typically occupied by a prosthetic group. For example, the binding pocket of cytochrome b562 is occupied by a heme group. For the remainder of this thesis, the term 'four-helix bundle' will refer to the square bundle type. Instances of four-helix bundle coiled-coils will be specifically noted.

## iv. Molten Globules vs. Native Structures: Conformational Mobility Within Proteins

A protein's unique three-dimensional structure is termed its *native* structure. This structure is well packed and does not interconvert between multiple low energy conformations. In contrast, some proteins may exist in a partially folded state termed the *molten globule* state.<sup>24</sup> Structures in the molten globule state usually possess all the characteristics of tertiary structure in the native state, but lack the specific contacts, such as side-chain packing, needed for a single unique native state. This is shown schematically in Figure 1.4. In addition, due to the fluctuating structure, molten globules have an increased hydrodynamic radius of about 15 % relative to the native state.<sup>25</sup> Presence of a molten globule state can be probed in various ways and will be described in detail in Chapter 4. With respect to the de novo design of proteins, one prevalent problem up until the mid 1990s was the design of amino acid sequences able to adopt

native-like folds. Section F of this introduction will highlight some examples of native-like de novo four-helix bundles.

**Figure 1.4.** Simple Illustration Showing Side-Chain Packing in Native-Like and Molten Globule States (Figure Used With Permission from Ref. 25).



## C. Current Views on the Mechanism of Protein Folding

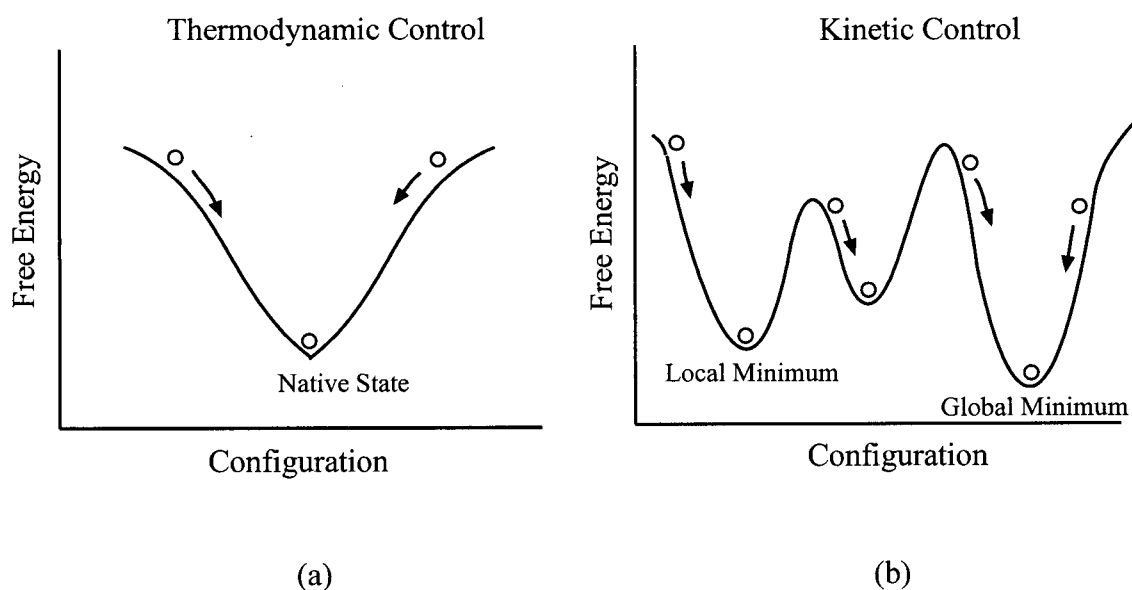
### i. Thermodynamic and Kinetic Control

The reversible folding and unfolding of proteins demonstrated by Anfinsen has led to the thermodynamic hypothesis of protein folding.<sup>1</sup> That is, native conformations of proteins are at a global free energy minima (Figure 1.5). However, Levinthal's paradox led to the view that the unfolded polypeptide chain does not go through a random search but via a specific folding pathway leading to the native structure. In this way, only a fraction of the possible conformations may be accessible and thus the polypeptide chain never encounters the problem of Levinthal's paradox. Therefore, it is argued that the native state is the lowest energy state

within the realm of all kinetically accessible states. A recent review article<sup>26</sup> stated that: “it is currently accepted that the three-dimensional structure of most proteins is under thermodynamic control and is reached through partially structured intermediates under kinetic control.”

States of lower energy may exist but may not be accessible and thus the thermodynamic hypothesis may not be easily disproved. In contrast, many instances of kinetic control have been reported.<sup>27</sup> In such cases, kinetic control was found to direct the folding process because the final conformation depended on the initial refolding conditions.

**Figure 1.5.** Schematic Representation of (a) Thermodynamic and (b) Kinetic Control of Protein Folding.

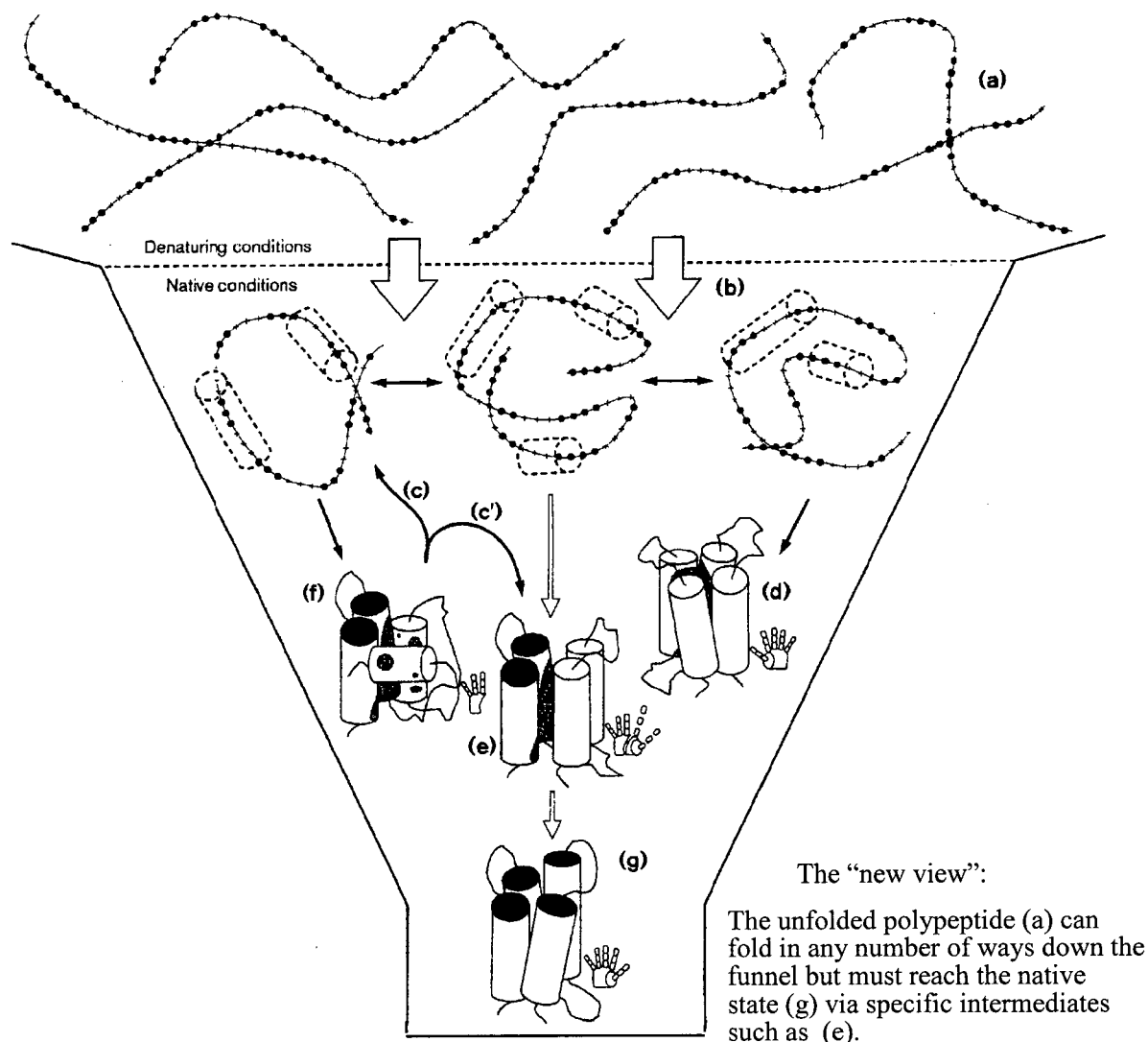


## ii. Folding Pathways and Funnels

The classical view of protein folding, held for several decades, was developed to overcome Levinthal's paradox. It states that the folding of the linear peptide chain follows *specific* folding pathways.<sup>3</sup> This theory led to the study and isolation of protein folding intermediates in the hopes of better understanding these pathways. Various theories have been proposed to model such pathways and some examples include the nucleation-propagation model,<sup>28</sup> the diffusion-collision model,<sup>29</sup> a sequential and hierarchical model,<sup>30</sup> a modular model,<sup>31</sup> the framework model,<sup>32</sup> and the hydrophobic collapse model.<sup>33</sup> Common to all of these theories is that there exists a single folding pathway unique to every protein.

Most recently, a "new view" of protein folding has emerged and is very similar to the previously proposed jigsaw puzzle model.<sup>34</sup> In the new view, there exists many possible folding pathways each travelling through an assembly of intermediates eventually leading to the native structure.<sup>35</sup> Rather than a single pathway, funnels can be used to illustrate the concept of parallel folding events leading to the same native structure (Figure 1.6).<sup>36</sup> As an analogy, Dill compares the classical and new views to water trickling down a mountainside into a lake.<sup>37</sup> In the classical view, water trickles down a single gully leading to the lake. In the new view, water can trickle down any part of the mountain, simultaneously, and reach the same lake.

**Figure 1.6.** Schematic Representation of the “Funnel” Folding Process. Figure Adapted From Ref. 38.



In the new view, multiple pathways may be followed down the funnel-like energy landscape to reach specific folding intermediates. Some argue these partially folded states may be thought of as kinetic traps corresponding to *misfolded* intermediates.<sup>39</sup> Nevertheless, in the new view the protein sample passes through a heterogeneous mixture of partially folded states,



each leading to the same final structure. As a result, isolation of partially folded intermediates may or may not give clues as to the mechanism of protein folding. Of course, this new view is not universally accepted and some still argue that folding processes can be described by simply extending the classical “single pathway” view.<sup>40</sup>

### iii. Chaperones and Assisted Folding

Although it is accepted that many proteins can fold and unfold reversibly *in vitro*, it became apparent that some more complex (i.e., oligomeric or multidomain) proteins require additional proteins in order to reach their native state.<sup>41</sup> These additional proteins have been named *molecular chaperones*.<sup>42</sup> Chaperones can act throughout the folding process by binding to partially folded intermediates and assisting the folding process. However, the role of chaperones is not fully understood as some proteins that have been studied both *in vivo* and *in vitro* have been found to proceed through the same respective folding intermediates. Presently, it is accepted that protein folding occurring at the high concentrations present in an intracellular environment requires chaperones to prevent spontaneous aggregation of partially folded intermediates.

Carbohydrates can also play a key role in the folding of certain proteins.<sup>43</sup> It has been found that *N*-linked oligosaccharides can facilitate protein folding. Although some glycoproteins have been found to fold properly without their *N*-linked glycans, others have been found to misfold, aggregate or degrade *in vivo*.<sup>44</sup>

As well, many disulfide bonds have been found to exist in a population of intermediates not present in the native structure.<sup>43</sup> This finding suggests that these disulfide bonds aid and direct the folding process in some manner.

There is no consensus on the mechanism for protein folding and hence there is ample opportunity for further study. Although this thesis does not enter into a discussion of folding mechanisms, the various theories were introduced as background information. As well, this thesis does not specifically address factors assisting folding such as chaperones, however, their existence should be acknowledged. Nevertheless, it is important to note that in many cases, amino acid sequence alone is enough to dictate a final protein fold and it is this aspect of protein folding that will be explored in this thesis.

## **D. Factors Controlling Protein Folding**

The native state of proteins is generally only 5 – 15 kcal/mol more stable than the unfolded state.<sup>45</sup> This small difference in energy is a result of many favorable and opposing forces common to all proteins such as conformational entropy, hydrophobic and electrostatic interactions, hydrogen bonding, and van der Waals interactions. In addition, there are several factors special to the folding of  $\alpha$ -helices and four-helix bundles. First, general factors controlling protein folding will be introduced (Section D.i) followed by a more detailed discussion of the factors controlling  $\alpha$ -helical (Section D.ii) and four-helix bundle stability (Section D.iii).

i. General Aspects Controlling Protein Folding

a. Conformational Entropy

The main energetic factor opposing protein folding is the loss of conformational entropy upon folding. An unfolded polypeptide chain is able to sample a vast number of conformations. The restriction of such an inherently flexible polypeptide chain is entropically unfavorable and other stabilizing factors must overcome this barrier for protein folding to occur.<sup>46</sup>

b. Electrostatic Interactions

Electrostatic interactions may be described by Coulomb's law: the energy change is proportional to the partial unit charges and inversely proportional to the distance between them and the dielectric constant. Opposite charges attract each other while like charges repel one another.

The side-chains of certain amino acids have the potential to form charges which may affect the overall stability of the protein. For example, *specific* charge interactions such as ion pairing (i.e.,  $\text{Glu}^-$ - $\text{Lys}^+$ ) serve to stabilize the folded state and many of these interactions have been observed in X-ray crystal structures of natural proteins. The magnitude of this stabilizing interaction has been estimated at 0.5-3 kcal/mol each.

In addition, the protein as a whole may have an overall net charge as a result of the various ionizable groups. For example, if a protein contains twelve lysine residues ( $\text{R-NH}_3^+$ ) and ten glutamic acid residues ( $\text{R-CO}_2^-$ ), the protein would possess an overall net charge of +2

at pH 7. The net charge density of the folded protein is greater than that of the unfolded protein and may result in electrostatic repulsions between like charges. This type of *classical* electrostatic interaction will depend on the square of the net charge and thus no net effect is expected near the isoelectric point (defined as the pH at which the protein has no net charge) of the protein. Hence, extremes of pH destabilize a protein and may result in protein unfolding.

Quantifying the importance of these interactions can be investigated in many different ways. Presently, experimental evidence suggests that electrostatic interactions can contribute to protein stability, but is not a dominant force in protein folding.<sup>48</sup>

#### c. Van der Waals Interactions

Weak van der Waals interactions occur between all atoms as a result of induced localized dipoles. At any given moment, an atom may possess an asymmetric distribution of electron density resulting in a weak dipole. Such a dipole may induce a complementary dipole on an adjacent atom or molecule thus creating a stabilizing interaction between them. The magnitude of this interaction between two atoms varies as the sixth power of the distance between them. This effect is especially important when considering the packing of various side-chains within the core of a protein. A more complementary packing arrangement will result in more van der Waals contacts and serve to stabilize the overall structure. Although this interaction is quite weak (ca.  $-0.3$  kcal/mol), the fact that it exists between all atoms makes it an important force in protein folding.

#### d. Hydrogen Bonding

First proposed in the early 1920s,<sup>47</sup> hydrogen bonding may be defined as an interaction between an electronegative atom with a covalently bound hydrogen (donor) and another electronegative atom (acceptor). The strength of a typical hydrogen bond can range from 2 – 10 kcal/mol.<sup>48</sup> The hydrogen bond is considered primarily an electrostatic interaction and as a result, its strength is largely dependent on the electronegativities of the donor and acceptor atoms.

The most important hydrogen bond donors and acceptors in proteins are the backbone amide NH's and CO's, respectively. This type of hydrogen bond makes up 68% of all hydrogen bonds in proteins, primarily found in secondary structures.<sup>49</sup> In particular, hydrogen bonding plays a key role in the stability of  $\alpha$ -helices as previously mentioned. The importance of the hydrogen bond to protein stability has been recently discussed in the context of protein mutant studies<sup>49</sup> which placed hydrogen bonding and hydrophobic interactions as equally important determinants of protein folding.

#### e. Hydrophobic Effects

The burial of hydrophobic surfaces within the core of a protein has generally been accepted as the dominant force in protein folding in aqueous media.<sup>48</sup> However, as mentioned above, more recent work on mutant proteins has introduced the notion that hydrogen bonding may be equally as important.<sup>49</sup>

The mechanism of this “hydrophobic effect” is still under debate.<sup>48</sup> One commonly held view is that water molecules are well ordered around non-polar groups and burial of hydrophobic surfaces releases these water molecules to the bulk solvent in an entropically favored process. In another sense, the magnitudes of interactions between water and non-polar molecules are less than the van der Waals interactions between apolar molecules and hydrogen bonding interactions between water molecules.<sup>2</sup> Other studies stress the unfavorable interaction between water and apolar groups rather than the favorable interactions between apolar groups.<sup>50</sup> Nevertheless, since the side-chains of many amino acids possess hydrophobic character (i.e., Phe, Leu, Ile, and Val), their specific burial into a protein core helps dictate a particular protein fold and drive the folding process.<sup>51</sup>

In summary, protein folding is governed mainly by the four above-mentioned forces. The relative contributions of each factor is not fully understood and is most likely different for different proteins. As a result, each factor must be considered when studying protein structure.

## ii. Factors Controlling $\alpha$ -Helical Structure

The above discussion applied to protein folding in general. More specifically suited to this thesis are the factors controlling  $\alpha$ -helix (secondary structure) and four-helix bundle (tertiary structure) formation. As mentioned in the introduction to  $\alpha$ -helices, the characteristic “ $i - i + 4$ ” hydrogen bonding pattern found in  $\alpha$ -helices undoubtedly stabilizes their structure. However, there are many more factors that contribute to their stability. The following is a brief review of the factors controlling  $\alpha$ -helix stability. First, the propensities of amino acids to form

$\alpha$ -helices will be discussed followed by a discussion of the factors responsible for these differing propensities. The discussion will then include other factors responsible for  $\alpha$ -helix stability such as helix macrodipole effects, helix capping effects, side-chain to side-chain interactions and helix length.

a. Helix Propensities and Factors Influencing Helix Propensities

The characterization of many helices within natural proteins has led to the discovery that some amino acid residues are found more frequently in protein helices than other residues. This led to the notion that some residues, and hence their side-chains, possess intrinsically higher helical propensities than others. Many different methods have been used to estimate the intrinsic helical propensities of the 20 amino acids. Chou and Fasman were the first to take a statistical survey of proteins of known structure and found certain amino acids had a very high probability of occurring in  $\alpha$ -helices while others were more likely to be found in  $\beta$ -sheets.<sup>52</sup> This led others to develop model systems in which they could measure the intrinsic helical propensities of each amino acid. Some model systems have included monomeric alanine-based peptides,<sup>53,54</sup> monomeric salt-bridged stabilized peptides<sup>55,56</sup> and a combination of the two methods.<sup>57,58</sup> In addition, helical propensities have been measured in two small single-domain proteins<sup>59,60</sup> as well as in a coiled-coils of *de novo* design.<sup>61</sup> Comparison of the helix propensities derived from the various experimental methods has shown a high level of correlation between them. One notable exception was the minor differences found between propensities derived from protein and peptide systems.<sup>62</sup> In order to further explore this

discrepancy, Myers et. al. recently measured the helix propensities in the  $\alpha$ -helix of ribonuclease T<sub>1</sub> and a helical peptide of identical sequence.<sup>63</sup> They found the peptide/protein propensities to be in good agreement with each other but found that the alanine-based model peptide systems overestimated the magnitude of some helical propensities by a factor of 2 to 3. Nevertheless, Figure 1.7 shows a consensus rank order of helix propensities from the above-mentioned studies.

**Figure 1.7.** A Consensus Rank Order of Helix Propensities for the 20 Commonly Occurring Amino Acids.

(best) Ala>Arg>Leu>Lys>Met>Gln>Glu>Ile>Phe>Trp>

Ser>Tyr>His>Asp>Cys>Asn>Val>Thr>Gly>Pro (worst)

In general, alanine is considered to be the only amino acid to contribute favorably to helix stability.<sup>53</sup> Alanine is followed by amino acids with long hydrophobic side-chains and by side-chains with polar groups several carbons removed from the backbone.  $\beta$ -branched amino acids and amino acids with polar groups only one carbon from the backbone have low helix propensities. These residues are followed by glycine and proline which have the lowest helical propensities. Interestingly, apart from proline, the differences in helical propensities as measured by the above methods are less than 1 kcal/mol. These small differences manifest themselves into large barriers to helix formation along the length of an average helix.

The specific factors contributing to an amino acid's helix propensity vary between the different amino acids. Proline has the lowest helical propensity mainly due to two factors. The rigid backbone of Pro induces strain into the helix and its lack of NH hydrogen bond donor



disrupts the hydrogen bonding pattern of the helix. The inherent flexibility of Gly favors a random coil conformation and its lack of  $\beta$ -carbon precludes the formation of a stabilizing interaction with the polypeptide backbone as discussed below.

Additional factors contributing to the helical propensities include: (1) loss of side chain entropy on helix formation, (2) hydrophobic effects between the side-chain and the peptide backbone and (3) formation of side-chain hydrogen bonds to the backbone. The  $\chi_1$  side-chain dihedral angle of  $\beta$ -branched amino acids (Val, Ile, Thr) is greatly limited in  $\alpha$ -helices due to conformational restrictions imposed by the polypeptide backbone. As a result, these amino acids undergo a greater loss of side-chain entropy on helix formation when compared to other non- $\beta$ -branched residues, and this is thought to be reflected in their poor intrinsic helical propensities.<sup>64</sup> A similar argument can be made for the aromatic residues Phe, Trp and Tyr.

While loss of side-chain entropy is unfavorable towards helix formation, burial of non-polar surfaces of side-chains is favorable.<sup>59</sup> Residues such as Leu, Met, Gln, Glu, Arg and Lys all possess long alkyl side-chains which upon packing against the helix can exclude solvent from the backbone and hence stabilize the helix. This effect was also thought to explain the relatively high helix propensity of Ile when compared to the similarly  $\beta$ -branched Val.<sup>59</sup>

Polar side-chains have the possibility of hydrogen bonding to the backbone or solvent and may contribute significantly to helical propensities.<sup>59</sup> Hydrogen bonding of a side-chain preferentially to the backbone may destabilize the helix and hence be reflected in a low helix propensity. Such an argument has been used in rationalizing the higher helical propensity of Gln over Asn. It is thought that as a result of one extra methylene group, Gln may be better solvated and have less of an interaction with the backbone hydrogen bonds than would Asn.<sup>62</sup>

Of course, the extra methylene group also buries more of a hydrophobic surface and thus may be a contributing factor as well.

The above discussion addressed merely the intrinsic preference for any amino acid to adopt the  $\alpha$ -helical conformation. However, within the context of an  $\alpha$ -helical chain, there are many other factors that can contribute to helix stability. As example of this importance of context, Minor and Kim recently showed that an 11 residue peptide formed an  $\alpha$ -helix when inserted into one location but also folded into a  $\beta$ -strand when inserted into another location within the same protein.<sup>65</sup>

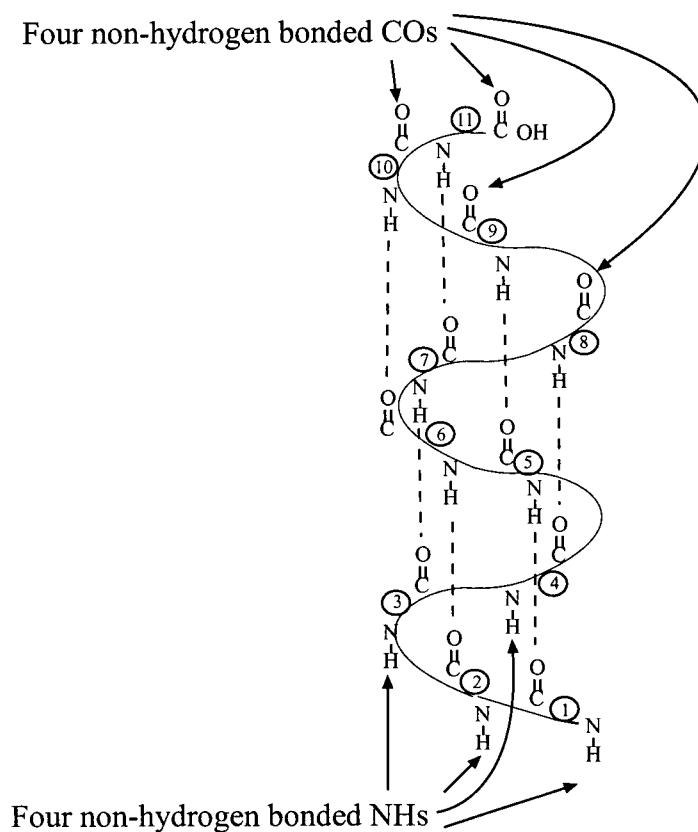
#### b. Helix Macro-dipole Effects

As mentioned in the introduction to  $\alpha$ -helices (Section B.ii), one consequence of the highly ordered hydrogen bonds that span the constituent amide backbone is the helix macro-dipole.<sup>16</sup> This dipole is thought to effectively place half a unit positive and negative charge at the N- and C-termini, respectively. This electrostatic effect has been verified experimentally in many different ways.<sup>66</sup> As a result of this effect, amino acids with negatively charged side-chains (i.e., Asp and Glu) stabilize  $\alpha$ -helices when placed near the N-terminus while amino acids with positively charged side-chains (i.e., Lys and Arg) are favored at the C-terminus.<sup>62</sup>

c. Helix-Capping Effects

One consequence of the amide  $i \cdots i + 4$  hydrogen bonding pattern is that the first and last four residues in the helix do not participate in this stabilizing interaction (Figure 1.8). As a result, it has been found that certain residues occur much more frequently at the N- and C-terminal positions.<sup>67</sup> It was postulated that these residues can fulfill these unsatisfied hydrogen bonding sites and, in turn, stabilize the resulting helix. These residues that begin or end a helix were termed N- and C-capping residues, respectively. N-caps are more frequently observed in nature occurring in nearly 70% of protein helices while C-caps occur in only 30% of protein helices.<sup>67</sup>

**Figure 1.8.** Illustration of the Non-Hydrogen Bonded Atoms in the  $\alpha$ -Helix Amide Backbone.



### 1. N-Capping

N-capping interactions have been studied in both peptides<sup>68</sup> and proteins<sup>69</sup> and have been found to contribute significantly to helix stability up to 2 kcal/mol relative to the unfolded state. N-cap propensities can be rationalized in four ways: (1) side-chain to backbone hydrogen bonding, (2) solvation of non-hydrogen bonded backbone sites, (3) burial of hydrophobic surface, and (4) electrostatic interaction with the helix macrodipole as mentioned above. For

example, amino acids with small hydrophilic side-chains (Asn, Asp, Ser, Thr) are generally good N-caps because their side-chains can hydrogen bond to the backbone amide groups. Glycine is thought to be a good N-cap because its small size does not restrict the solvation of non-hydrogen bonded peptide groups in the backbone. In addition, the size of the side chain correlates well with the N-cap propensities in a model peptide system<sup>68d</sup> which suggests that burial of hydrophobic surfaces also serves to stabilize the helix. However, others argue that larger hydrophobic side-chains make poorer N-caps because their size precludes solvent molecules from hydrogen bonding to backbone amides in their place.<sup>69a-c</sup> Gln is thought to be one of the poorer N-caps because its relatively long hydrophilic side chain can stabilize peptide groups in non-helical conformations.

One recurrent N-capping motif called the “capping box” was discovered by Harper and Rose.<sup>70</sup> It consists of a reciprocal hydrogen bonding pattern where the side chain of the N-cap residue forms a hydrogen bond to the backbone N-cap + 3 residue, while the side chain of the N-cap + 3 residue forms a hydrogen bond back to the N-cap residue. Another motif termed the “hydrophobic staple” involves a hydrophobic interaction between residues near the N-cap.<sup>71</sup>

## 2. C-Capping

C-capping interactions occur less frequently and have been found to have far less of an effect on  $\alpha$ -helix stability. It has been found that positively charged residues such as His, Lys and Arg stabilize helices when in the C-cap position by only 0.1 – 0.4 kcal/mol;<sup>68e</sup> an interaction which may simply be a charge-dipole effect as discussed previously. Some argue that the relatively small stabilization energies are partly due to the large loss in conformational entropy

which would go along with such a long side-chain to backbone interaction.<sup>72</sup> Glycine was found to occur commonly in natural proteins adopting the so-called "Schellman motif" whereby a flexible Gly residue can allow the amino acid *next* in the sequence to hydrogen bond back to the amide backbone.<sup>73</sup> As such, Gly is thought to be more important in helix termination than in stabilization.

In a recent study, Prieto and Serrano found proline to stabilize the C-termini of helices when placed one residue past the C-capping position.<sup>74</sup> It was reasoned that the rigid proline residue could restrict the allowable conformations of the C-cap and, in turn, stabilize the helix. In another study, Schneider and DeGrado explored various alkyldiamine C-capping auxiliaries in monomeric peptides and found *D*-Arg caboxamide to stabilize helices by 1.2 kcal/mol.<sup>72</sup> They attribute this stabilization to the C-cap's ability to both screen the helix macrodipole *and* hydrogen bond to non-hydrogen bonded backbone amide bonds.

#### d. Side-Chain to Side-Chain Interactions Within the Helix

Although amino acid side chains have been shown to play important roles in helix stabilization by interacting with the backbone, side-chain to side-chain interactions can also contribute to helix stability. Many studies have been done and have shown that regardless of the position within the helix, stabilizing interactions are possible. Since it is known that  $\alpha$ -helices consist of 3.6 residues per turn, it is possible to position potential stabilizing interactions on the same side of the helix by spacing amino acids three ( $i - i + 3$ ) and four ( $i - i + 4$ ) residues apart in the sequence.

Salt-bridge interactions have been quite well studied. For example, the strength of a Glu-Lys interaction has been estimated at  $-0.5$  kcal/mol in model Gln-Ala based peptides.<sup>75</sup> Some examples of other studies include Phe-His amino-aromatic interactions,<sup>76</sup> Tyr-Leu/Val non-polar interactions,<sup>77</sup> Asp-Gln hydrogen bond interactions<sup>62,78</sup> as well as His-Asp ion-pair interactions.<sup>79</sup>

#### e. Helix Chain Length

It is well known that the length of the  $\alpha$ -helix has a dramatic effect on the stability of the helix. For example, studies on two stranded coiled-coils showed a significant increase in helicity and stability when the helices were lengthened from 8 to 36 residues.<sup>80</sup> In another study on four-helix coiled-coils, extending the length of the helices from 21 to 35 residue increased the stability by 9 kcal/mol per helix.<sup>81</sup> Typically, the minimum number of amino acids required to form an isolated  $\alpha$ -helix is between 8 and 12 residues and generally many more are required for a stable  $\alpha$ -helix. In proteins, however, many other tertiary interactions are possible and as a result, shorter helices have been observed.

In summary, many different subtle interactions are at play in determining the overall stability of  $\alpha$ -helices. In order to design  $\alpha$ -helices effectively, all of these factors must be considered. Of course, the above discussion merely introduced factors controlling secondary structure formation. Many more factors must be considered when studying helical tertiary structures.

### iii. Factors Determining the Stabilities of Four-Helix Bundles

The stabilities of four-helix bundles are largely dependent on the stabilities of their constituent helices. However, with the context of the helical bundle there are other forces at play which include side-chain packing, other side-chain to side-chain interactions such as salt bridges, and helix orientation (i.e., parallel or antiparallel helices).

#### a. Inter-Helical Side-Chain Interactions

##### 1. Packing

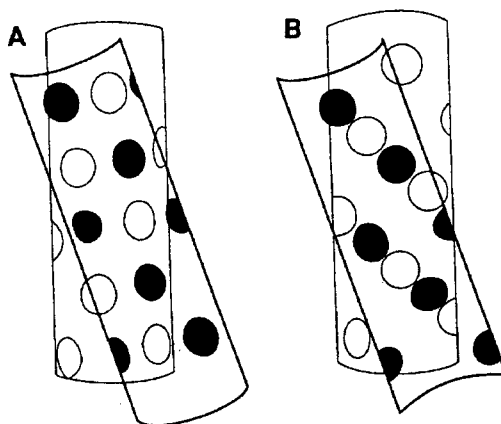
As mentioned previously,  $\alpha$ -helices in square four-helix bundles pack against one another at an angle of about  $20^\circ$  with respect to the helical axis. This was originally predicted by Crick who reasoned that helices could preferentially interact at this angle by way of a “knobs-into-holes” model.<sup>21,82</sup> In this model, a side-chain on one helix packs between residues at positions  $i$ ,  $i + 3$ ,  $i + 4$ , and  $i + 7$  of an adjacent helix (Figure 1.9). However, as structures of more and more proteins became available, it was clear that the “knobs-into-holes” model could not explain the common occurrence of near perpendicular helical packing arrangements.

In 1981, the “ridges-into-grooves” model was developed<sup>83,31</sup> to help explain the observed wide distribution of helix packing geometries. In this model, the side-chains of each  $\alpha$ -helix forms a ridge separated by grooves. Helices pack together by the ridges of one helix packing into the grooves of another (Figure 1.9). Ridges are usually formed by side-chains three ( $i - 3$ ,  $i$ ,



$i + 3$  ... etc.) and four ( $i - 4, i, i + 4$  ....) apart. For example, for the observed  $20^\circ$  crossing angle found in four-helical coiled-coil and square bundles, the  $i + 3$  ridges of one helix intercalate with  $i + 4$  the ridges and grooves of another. The use of the various ridges depends largely on the side-chains of each individual helix.

**Figure 1.9.** Simplified Two-Dimensional Illustration of (A) Crick's "Knobs-Into-Holes" Packing and (B) Chothia's "Ridges-Into-Grooves" Packing Between Two  $\alpha$ -Helices. Side-Chains of Each Helix are Depicted as Open and Closed Circles. Only One Side of Each Helix is Shown for Clarity.



More recently, Bowie reviewed packing angle preferences in helices and suggested that the steric models proposed above do not adequately describe helix packing.<sup>84</sup> He cited, for example, the fact that helices are not flat and thus are not well-described by two-dimensional models described above. As well, the use of a  $C_\alpha$  or  $C_\beta$  atom position as a "knob" is an inadequate description of flexible side-chains. Bowie proposed that statistics alone could account for the observed inter-helical angles and that acceptable packing may be achieved at a variety of inter-helical angles. Moreover, he states that: "it would seem that packing can usually adjust to fit the inter-helical angle rather than the inter-helical angle adjusting to fit packing."

The manner in which helices pack against one another is still controversial and no single model has been universally accepted.

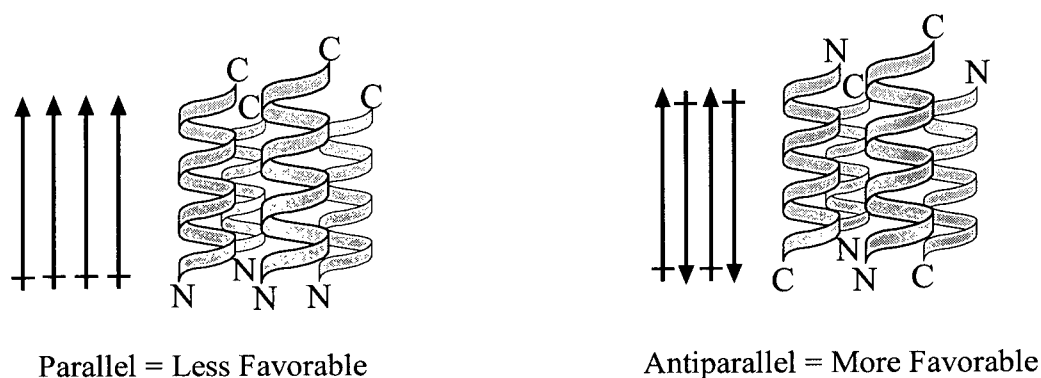
## 2. Other Side-Chain to Side Chain Interactions Within Four-Helix Bundles

Section D.i.d of this chapter described *intra*-helical side-chain interactions and their importance to helical stability. These types of interactions (i.e., salt bridges, hydrogen bonds, etc.) can also occur between  $\alpha$ -helices (*inter*-helical) and may aid in determining the stabilities and/or structures of  $\alpha$ -helical bundles. For example, Hodges and coworkers' recent studies on two-stranded coiled coils have shown that inter-helical ion pairs (Glu<sup>-</sup>Lys<sup>+</sup>) can contribute up to 0.5 kcal/mol of stability<sup>85</sup> and aid in controlling helix orientation (i.e., parallel vs. antiparallel, see below).<sup>86</sup>

### b. Helix Orientation

As a result of the helix macrodipole, four-helix bundles are more stable when their constituent helices are in an antiparallel arrangement. In such a geometry, the helices are arranged such that two of the helices point in one direction while the remaining two helices point in the opposite direction (Figure 1.10). This minimizes the unfavorable dipole-dipole interactions between helices.

**Figure 1.10.** Schematic Representation of Parallel and Antiparallel Four-Helix Bundles.



This effect, however, has been found to be quite small. In one study, the magnitude of this stabilization was estimated to be only 1 kcal/mol.<sup>87</sup> In another study, backbone dipole interactions were estimated to be one order of magnitude less than backbone-charge interactions (i.e., side-chain to backbone hydrogen bond).<sup>66d</sup>

In summary to Section D, there are many factors influencing the overall structures of proteins and four-helix bundles. Some factors may be more important than others, but nevertheless, all of these factors must be considered when (1) assigning contributions from individual amino acids and (2) designing a sequence aimed at adopting a specific structure. In the next section, Section E, various examples of de novo four-helix bundles will be discussed, highlighting important aspects of each design.

## **E. Examples of De Novo Protein Design**

De novo protein design gives the protein chemist an opportunity to test their understanding of the rules of protein folding. Although there are many examples in the literature of all  $\beta$ -sheet proteins,<sup>88</sup> mixed  $\alpha$  and  $\beta$  structures,<sup>89</sup> and helical bundles composed of two<sup>90</sup> or three strands,<sup>91</sup> for the purposes of this thesis only examples involving four helices will be discussed in detail. Section E.i will discuss select examples of four-helix bundles and coiled-coils designed from scratch, while the Section E.ii will discuss a few interesting examples of *redesign* of naturally occurring four-helix bundles and coiled-coils.

One common theme that emerges with the following examples of de novo design is the use of varied hydrophobic residues to restrict the conformational mobility of the hydrophobic core within the bundle. For example, isoleucine may possess less rotameric states than leucine and as a result may impart rigidity to the core of the four-helix bundle. As the reader will notice, this design methodology results in much more specific and native-like structures. Note that the specific amino acid sequences for those not listed in this section can be found in Appendix B.

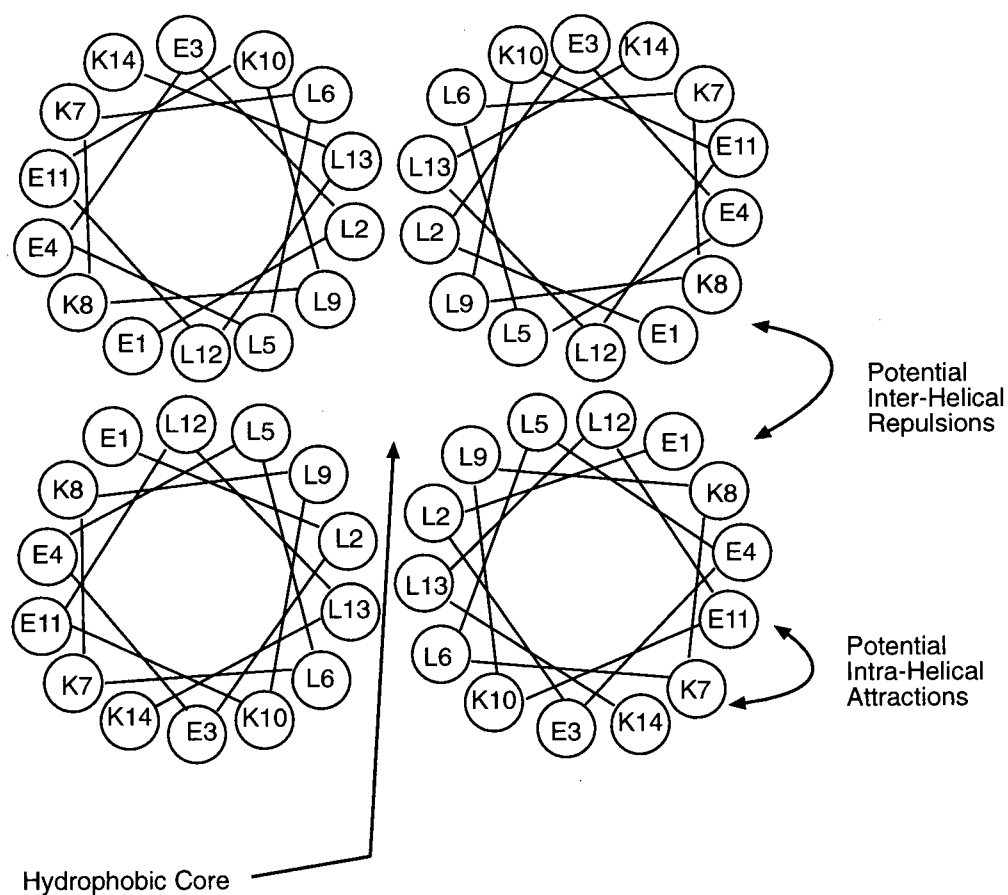
i. De Novo Designed Four-Helix Bundles

a. DeGrado

William DeGrado is one of the pioneers in de novo design. Over a decade ago, he and his coworkers embarked on a program aimed at the design of four-helix bundle motifs. The peptides were designed in a minimalistic manner such that they would contain minimum variations in residues in order to breakdown the folding process into its fundamental forces.<sup>8</sup> The next paragraph discusses in detail the design principles that were incorporated into the first generation of de novo designed four-helix bundles. Note that throughout this thesis, many of these design features will be repeated as they are incorporated into many de novo designs.

In accord with the minimalistic approach, only a few different types of amino acids of high helical propensity were incorporated into the early designs (Table 1.1). This included the use of Leu as the sole hydrophobic residue and Glu and Lys as the sole hydrophilic residues. The peptides were designed to be amphiphilic; that is, when folded into an  $\alpha$ -helix, the peptides would possess hydrophobic (leucines) and hydrophilic (lysines and glutamic acids) faces. Figure 1.11 shows a helical wheel diagram of the designed peptide sequence showing the arrangement of hydrophobic and hydrophilic residues when the helices are packed in an antiparallel arrangement. Note that the hydrophobic faces point into the core of the bundle, thereby minimizing their exposure to solvent. This hydrophobic effect was expected to be the main driving force for bundle formation.

**Figure 1.11.** Helical Wheel Diagram of DeGrado's  $\alpha_4$  De Novo Four-Helix Bundle Looking Down the Helical Axes. Adjacent Helices Are Drawn Antiparallel With Respect to One Another. Helix Terminating Gly Residues and Loop Regions Are Omitted For Clarity. Amino Acids Are Numbered 1-14 in a N- to C-terminal Direction.



The peptides were also designed to have the potential to form helix-stabilizing *intra*-strand salt-bridges between Glu and Lys residues by spacing these residues four apart in the amino acid sequence (Figure 1.11). However, note that in this antiparallel orientation, the helices possess like charges at the interface between the helices. This potential *inter*-strand repulsion may destabilize the bundle.

To avoid unfavorable interactions with the helix macrodipole<sup>16</sup> two additional design features were added: (1) the N- and C-termini were capped with acetyls and amides, respectively, and (2) Glu and Lys residues were incorporated near the N- and C-termini, respectively. As well, glycine was used as both N- and C-terminal capping residues to aid in helix termination as Gly is commonly found at the ends of naturally occurring helices.<sup>67</sup>

Using an iterative design process and after an initial effort at peptide design,<sup>92</sup> a single-stranded peptide, named  $\alpha_1B$ , was designed, synthesized and found to aggregate into homotetramers thereby forming a de novo four-helix bundle (Table 1.1).<sup>93</sup> Later, a loop region connected two of these peptides such that in solution the peptides,  $\alpha_2B$ , would self-associate to form four-helix bundle homodimers.<sup>93</sup> And finally, in the last stage of the initial design process, two  $\alpha_2B$  peptides were joined together by another loop region to form a single stranded polypeptide, named  $\alpha_4$ , which folded into a four-helix bundle in solution (Figure 1.12).<sup>94</sup> This was the first example of a de novo designed protein adopting a stable, folded structure in aqueous solution.

**Table 1.1.** DeGrado's Series of De Novo Designed Four-Helix Bundle-Forming Peptides.

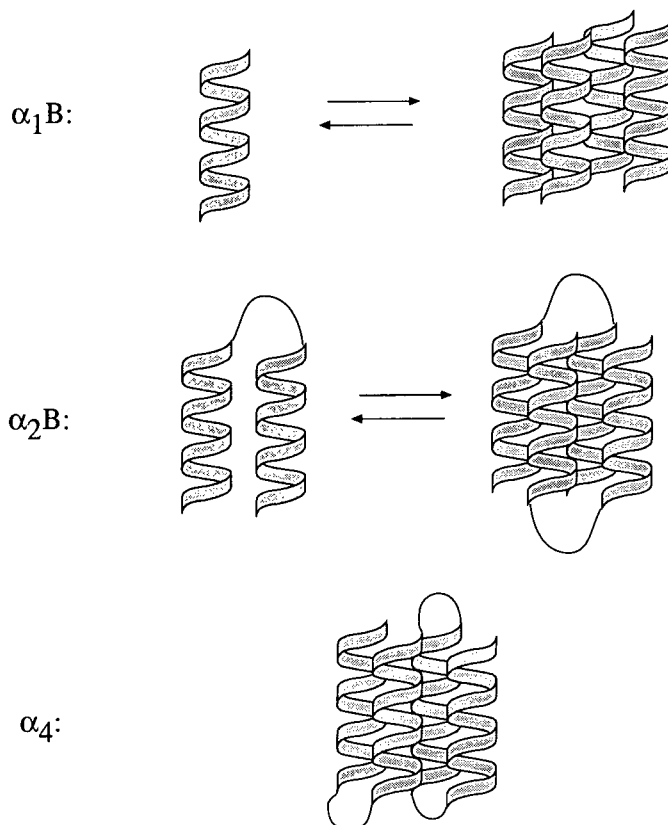
Sequence Name	Amino Acid Sequence in One-Letter Code <sup>a</sup>
Helix:	[GELEELLKCLKELLKG]
Loop:	[PRR]
$\alpha_1$ B:	Ac-[Helix]-CONH <sub>2</sub>
$\alpha_2$ B:	Ac-[Helix]-[Loop]-[Helix]-CONH <sub>2</sub>
$\alpha_4$ :	M-[Helix]-[Loop]-[Helix]-[Loop]-[Helix]-[Loop]-[Helix]-COOH
$\alpha_2$ C:	Ac-[GEVEELLKKFKEL <b>W</b> KG]-[Loop]-[GEIEEL <b>F</b> KK <b>F</b> KELIKG]-CONH <sub>2</sub>
$\alpha_2$ D:	Ac-[GEVEELEKK <b>F</b> HEL <b>W</b> KG]-[Loop]-[GEIEEL <b>H</b> KK <b>F</b> KELIKG]-CONH <sub>2</sub>

<sup>a</sup> Letters in bold illustrate changes from previous generation.

$\alpha_4$  was found to be monomeric in solution, highly helical and highly stable towards chemical denaturants. In addition, the stabilities of the four-helix bundles were found to increase as successive loop regions were added in between the helical peptide strands. Unfortunatley, both  $\alpha_2$ B<sup>95</sup> and  $\alpha_4$ <sup>96</sup> possessed some of the properties associated with molten globule-like structures most likely due to imperfect packing of the hydrophobic core of the protein.



**Figure 1.12.** Illustration of DeGrado's Iterative Design Process for  $\alpha_4$ .

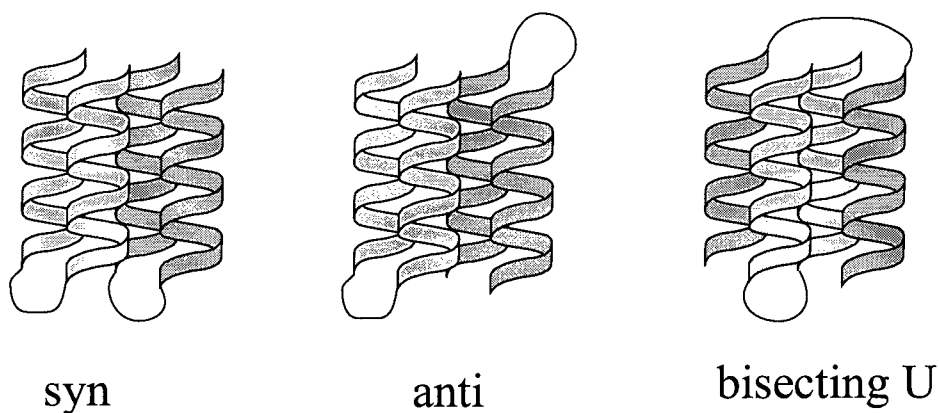


This early work led to the notion that to effectively design polypeptides to adopt certain folds, not only should interactions be included to stabilize the desired fold, but interactions should also be included to *destabilize* potential alternative folds. In an attempt to restrict the hydrophobic packing of the helices,  $\alpha_2C$  was synthesized, incorporating more structurally diverse amino acids into the core of the bundle replacing several leucines for valine, isoleucine, phenylalanine and tryptophan (Table 1.1).<sup>97</sup> An improvement was seen in that the resulting four-helix bundle was found to undergo a thermally induced transition from a native-like to a molten globule state at room temperature.

One additional approach was taken to further define the overall structure. Metal binding sites were introduced into the core of the bundle hoping that such binding would restrict the hydrophobic core of the bundle and form a more native-like structure.<sup>98</sup> By simply incorporating two His residues and an additional Glu residue into  $\alpha_2C$ , the resulting peptide,  $\alpha_2D$ , formed a four-helix bundle with native-like properties, even in the absence of metal ions.<sup>99</sup> This native-like structure was attributed to the careful positioning of the His and Glu residues such that alternative folds were destabilized.

Although  $\alpha_2D$  was known to form dimers, the overall topology of the helices was not known until recently.<sup>100</sup> Its NMR solution structure was determined and found an unexpected overall topology. The helices were tilted with respect to one another by  $19^\circ$ , as may be expected for four-helix bundles. Interestingly, however, rather than associating as *syn* or *anti* antiparallel dimers,  $\alpha_2D$  self-associated as a *bisecting U* whereby the two loop regions reach across the bundle and connect diagonally opposite helices (Figure 1.13). This result demonstrates the difficulty in predicting the folding and aggregation properties of designed peptides.

**Figure 1.13.** Three Potential Topologies for DeGrado's Four-Helix Bundle  $\alpha_2D$ .



Another interesting approach to aid in structure stabilization was the incorporation of cysteine-containing loop regions into  $\alpha_2B$  and  $\alpha_4$  which allowed for subsequent disulfide formation between the loop regions.<sup>101</sup> Although detailed structural analyses were not reported for the various de novo proteins, each disulfide-linked four-helix bundle was stabilized relative to the non-crosslinked structures. As well, not all possible disulfide links were formed which suggests that the backbone has limited flexibility towards accommodating certain geometries. Further studies on the dynamics of the structures were not reported.

DeGrado has also designed several four-stranded coiled-coils, some of which were found to possess native-like structure. The “coil-XL” series of peptides were designed to aggregate into tetrameric four-stranded coiled-coils,<sup>102</sup> while the “RLP-X” series of peptides were designed to form dimeric four-stranded coiled-coils.<sup>103</sup> As mentioned earlier, coiled-coils differ from four-helix bundles by the superhelical twisting of their helices. Many of the effects controlling helical bundle formation are similar to coiled-coil formation as only subtle differences exist between their primary sequences. This work emphasized the importance of

introducing a variety of residues into an all leucine-packed hydrophobic core in order to optimize the core packing. Core residues such as Val or Ala were incorporated in addition to Leu and, depending on their locations within the sequence, resulted in less stable but native-like structures. Such a strategy introduces *specificity* into the structure at the expense of global *stability*. More examples of this strategy will be presented in the next section.

b. Hecht

While DeGrado used a minimalist approach to designing helical bundle structures, Hecht and coworkers designed their four-helix bundles by patterning non-polar and polar residues without explicitly assigning the identities of those residues. They hoped that by simply patterning hydrophobic and hydrophilic residues, a global fold could be designed. Using the characteristic 3.6 residues per turn of the  $\alpha$ -helix, they used a combinatorial approach to place hydrophilic residues at some locations and hydrophobic residues at others.<sup>104</sup> The majority of the sequences generated by such a *binary code* were both  $\alpha$ -helical and water-soluble.<sup>105</sup> A few sequences that were characterized were found to be both monomeric and highly stable. More recently, Hecht and coworkers have characterized one of the best designed sequences from their combinatorial set, named “M-60”, and have found it to possess some native-like properties.<sup>106</sup> This work emphasizes the importance of burying non-polar residues in determining the overall fold of a protein.

c. Baltzer

Other researchers have also designed native-like four-helix bundles by optimizing core packing. Baltzer and coworkers synthesized a 43 amino acid peptide, named “GTD-43”, designed to adopt a four-stranded helix-loop-helix dimeric coiled-coil using many of the previously mentioned design strategies.<sup>107</sup> Unique to their design was the incorporation of three aromatic residues, 2 x Phe and Trp, into the hydrophobic core of the bundle. Their approach to generating a native-like structure was to introduce these aromatic residues into the core of the bundle in the hopes that interactions between the aromatic side-chains would be strong and specific thereby creating large barriers to conformational rearrangements. Indeed, their design generated a four-stranded coiled-coil structure with native-like properties.<sup>107</sup> The authors attributed much of this structure to specific contacts between aromatic residues in the core of the bundle. Note from Figure 1.7 that aromatic residues do not possess high helical propensities. This design thus incorporates residues less likely to form helices but then leads to more specific packing interactions resulting in a more native-like structure. This is in contrast to the general use of aliphatic residues in the cores of de novo four-helix bundles (i.e., those synthesized in early work by DeGrado).

d. Dutton

The Dutton group has been interested in incorporating functionality into de novo dimeric four-stranded coiled-coil structures.<sup>108</sup> Recently, isoleucine and phenylalanine substitutions were introduced within the core of their primarily leucine-packed bundle named

“H10H24”.<sup>109</sup> This alteration was found to greatly increase stability and enhance its native-like properties as was the case with DeGrado’s coiled-coils previously mentioned. Again, restriction of the hydrophobic core by introducing non-leucine hydrophobic residues led to increased structural specificity, and in this case, increased stability.

e. Farid

Farid and coworkers made a similar discovery in their four-helix bundle named “HYP-1”.<sup>110</sup> They synthesized two peptides, named “A” and “B”, and linked them by forming a disulfide bond between N- and C-terminal cysteine residues. This “two-stranded” peptide formed an antiparallel four-helix bundle. They found that their leucine-filled core was best packed when the core contained either a number of valines and isoleucine residues to restrict conformational mobility and specify packing within the hydrophobic core.

ii. Examples of Redesign of Naturally Occurring Sequences Forming Four-Helix Bundles

An alternative to de novo protein design is to redesign naturally occurring proteins. Much can be gleaned from the study and redesign of nature’s sequences optimized by evolution. Below are just a few interesting examples from various research groups modifying natural sequences to explore the factors controlling four-helix bundle tertiary structures.

a. Alber

One early example of a redesigned natural helix-forming sequence was reported by Alber and coworkers using the two-stranded dimeric leucine zipper peptide GCN4-p1.<sup>111</sup> They found that by simply altering the pattern of hydrophobic residues between leucine, isoleucine and valine, coiled-coils of two, three and four strands could be formed.<sup>112,113</sup> More specifically, the peptide named “GCN4-pLI” used alternating leucine and isoleucine residues within the hydrophobic core to form a parallel four-stranded coiled-coil. Interestingly, hydrophobic packing of the bundle was the only factor dictating the oligomeric state of the coil and hence its overall structure.

b. Regan

Regan and coworkers used the natural four-stranded coiled-coil protein ROP to study a variety of factors responsible for its structure. ROP is a homodimer of helix-loop-helix monomers.<sup>23</sup> In one set of studies, the hydrophobic core of ROP was redesigned to include an assortment of Ala and Leu residues.<sup>114</sup> It was found that, as in the above-mentioned examples, an entirely leucine packed core resulted in an “overpacked” core with loss of defined structure. In contrast, an entirely alanine packed core was significantly less stable than the wild type. Ideal in the design was a mixture of alanine and leucine residues which was found to retain native-like structure, biological activity as well as increased stability relative to the wild type.

Other studies by Regan investigated the structural requirements of the surface turn between the two helices in ROP. They found that amino acid substitutions at the turn position

produced an unusually large range of thermodynamic stabilities; the majority of which possessed enhanced stabilities as compared to the wild type.<sup>115</sup> They also found that as the turn or loop region was lengthened with glycine residues, their stabilities were progressively decreased.<sup>116</sup> Furthermore, all of these loop-lengthened/modified mutants retained wild-type structure which suggests that the loop region does not dictate the final protein fold.

Another interesting redesign to ROP involved connecting the constituent helix-loop-helix monomers using various glycine-rich linkers to generate a monomeric form of ROP.<sup>117</sup> They found that too short a linker resulted in misfolded proteins while proteins with longer linkers afforded the desired topology with similar stabilities. Again, these results demonstrate that provided long enough linkers are present, turn regions do not specify the overall fold of the protein.

#### c. Fairman

Fairman studied peptides based on the tetramerization domain of Lac repressor which self-associates to form homotetrameric four-helix bundles. In one study, the length of each peptide was found to have a dramatic effect on the overall stability of the bundle. Lengthening each peptide from 21 to 35 (named "Lac 21", "Lac 28" and "Lac 35") residues stabilized the helical bundle by  $\sim 9$  kcal/mol per helix.<sup>81</sup> Another study explored the effects of multiple electrostatic attractions and repulsions using various Lys and Glu substitutions into the sequence.<sup>118</sup> One Lac-based peptide named "Lac21E" was flanked with numerous Glu residues while another peptide named "Lac21K" was flanked with numerous Lys residues. It was found that each isolated peptide was mostly non-helical while a mixture of the two resulted in a stable



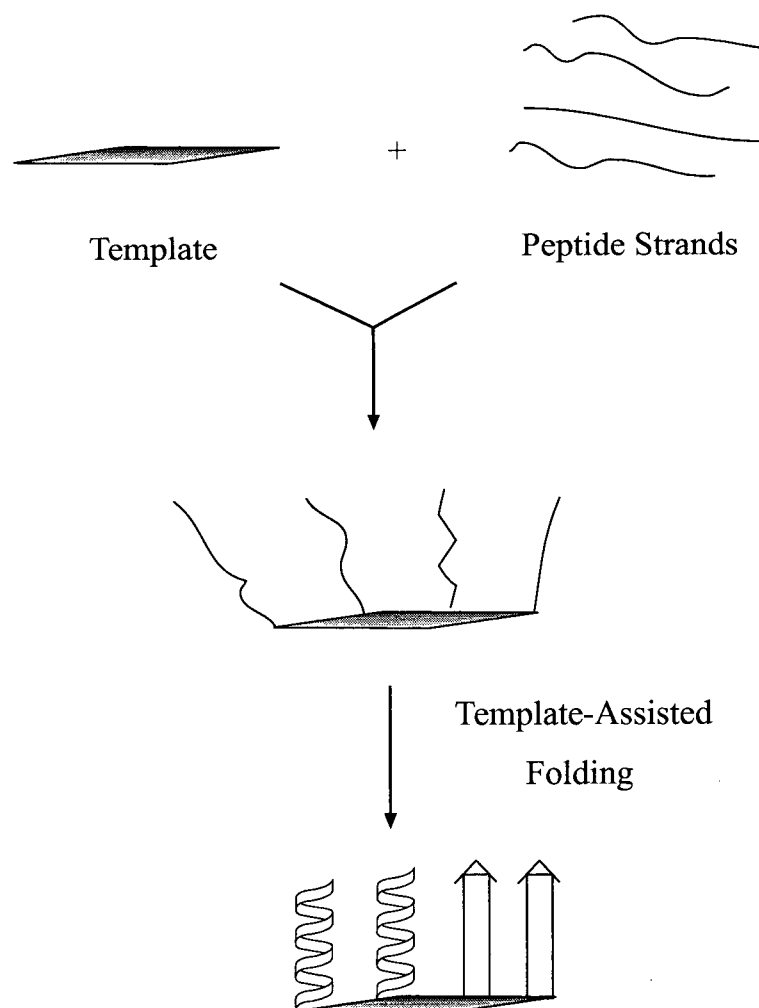
heterotetrameric coiled-coil; moreover, it was found to be 7.7 kcal/mol more stable than the parent Lac 21 peptide.<sup>118</sup> Although the authors did not investigate the presence of native-like structure, this work illustrates the importance of stabilizing *inter*-helical interactions in defining the overall structure.

## **F. Template Assembly in De Novo Protein Design**

### **i. What is Template Assembly?**

At the root of the protein folding problem is the large entropic cost of bringing inherently flexible peptide fragments close in space. As mentioned in Section D.i, nature overcomes this barrier using counterbalancing hydrophobic, electrostatic, van der Waals and hydrogen bonding interactions. Another way of overcoming this large entropic barrier is to first covalently link peptide strands to a template – effectively “prepaying” for the entropic cost of peptide association (Figure 1.14). The template serves to *preorganize* the peptide strands and hence aid in directing the folding process. This idea was first proposed by Mutter and has been termed *template assembly*.<sup>119</sup> Tertiary structures resulting from this type of peptide assembly have been named *template-assembled synthetic proteins* (TASPs).

**Figure 1.14.** Schematic Representation of Template Assembly in De Novo Protein Design.



It is interesting to consider the resulting template-assembled structures. A protein consisting of a single linear chain of amino acids links elements of secondary structure with loop regions. These elements of secondary structure aid in directing a folding process leading to the desired tertiary structure. By linking peptides to a template, TASP's possess a branched topology which allow it to bypass some of the folding processes of natural proteins which, in turn, may allow TASP's to adopt compact structures more directly.

The template molecule can explicitly direct the topology and the folding of the protein. Covalent attachment of peptides to a suitable template favors intramolecular interactions between peptide strands. In addition, the number, type and spatial arrangement of its functional groups can dictate the packing topology of the resulting three-dimensional structure. Such a template can thus be considered to be a stabilizing feature and an important determinant in the final three-dimensional structure of the TASP.

Templates have been used in the construction of both  $\alpha$ -helical and  $\beta$ -sheet proteins.<sup>120</sup> For example, Kelly and coworkers have used dibenzofurans<sup>121</sup> and copper coordinated bipyridine complexes<sup>122</sup> to template  $\beta$ -sheet structure in aqueous solution. As well, Nowick and coworkers have found oligoureas to be effective templates for  $\beta$ -sheet structure in organic solvents.<sup>123</sup> However, for the purposes of this thesis, only template-assembled  $\alpha$ -helical motifs will be discussed in detail.

## ii. Examples of Template-Assembled Four-Helix Bundles

This section discusses the main design features of each de novo protein. The synthetic strategies for a variety of de novo four-helix bundles will be discussed in Chapter Three. As well, the majority of the work presented in this section was completed between five and ten years ago and as a result, none of the template-assembled models enter into a discussion of native vs. molten globule-like states as discussed in Section E.i where much of the work mentioned is only a few years old.

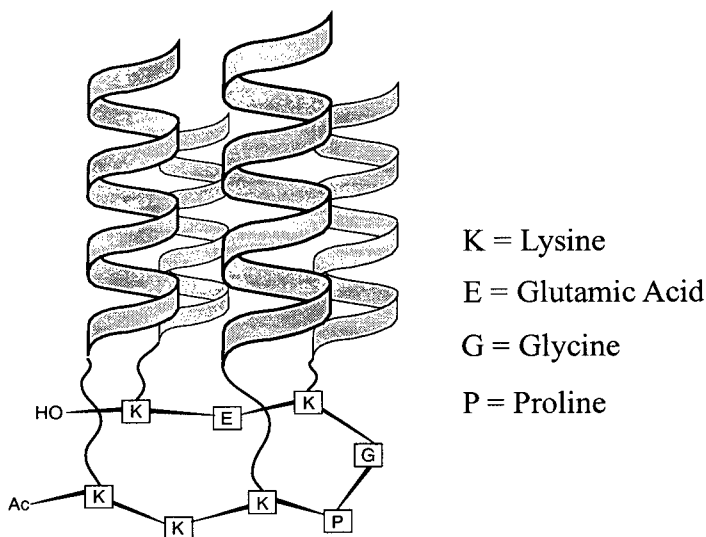
#### a. Peptide Templates

Mutter and coworkers were one of the first to use template assembly in the construction of folded polypeptide structures. Their approach was to use peptides themselves as templates for the assembly of parallel four-helix bundle motifs. Typically, the peptide templates would contain lysine residues such that the C-termini of the helix-forming peptides could be attached to the  $\epsilon$ -amino groups of Lys via a peptide bond (Figure 1.15).

Early work in Mutter's group concentrated on using acyclic templates which incorporated a -Pro-Gly- peptide fragment which was thought to induce a type II  $\beta$ -turn (template sequence = Ac-[KKKPGKEK]-CO<sub>2</sub>H).<sup>124</sup> Such a turn would aid in preorganizing the peptides for bundle formation. Further efforts to constrain the peptide templates and stabilize the structure led to the synthesis of cyclic peptide templates which incorporated disulfide bonds (template sequence = Ac-[CKAKPGKAKC]-CONH<sub>2</sub>)<sup>125</sup> and conformationally restricted nonnatural amino acids.<sup>126</sup> The helix-forming peptides were designed using established principles<sup>127</sup> including (1) appropriate spacing of hydrophobic and hydrophilic residues such that the sequence was amphiphilic, (2) amino acids with high helical propensities and (3) oppositely charged side-chains four residues apart in the sequence for possible salt bridge interactions (see Figure 1.15 for a representative amino acid sequence).

These studies demonstrated that parallel four-helix bundles were indeed formed and those based on cyclic templates were more stable than those based on acyclic templates.<sup>128</sup> As well, the template-assembled bundles were found to be considerably more stable and helical than non-templated control peptides. More detailed analyses of the dynamics of the structures were not reported.

**Figure 1.15.** Representative Example of Mutter's Strategy in Designing De Novo Four-Helix Bundle TASPs.



Representative Helix-Forming Peptide Sequence:<sup>125</sup>  $\alpha_{15}$  = EALEKALKEALAKLG

In an interesting example of the utility of TASP-type molecules, Mutter formed an ion-channel-forming de novo four-helix bundle TASP.<sup>129</sup> In comparison to a single-stranded control peptide, the TASP was found to be about 100 times more effective at ion-channel formation.

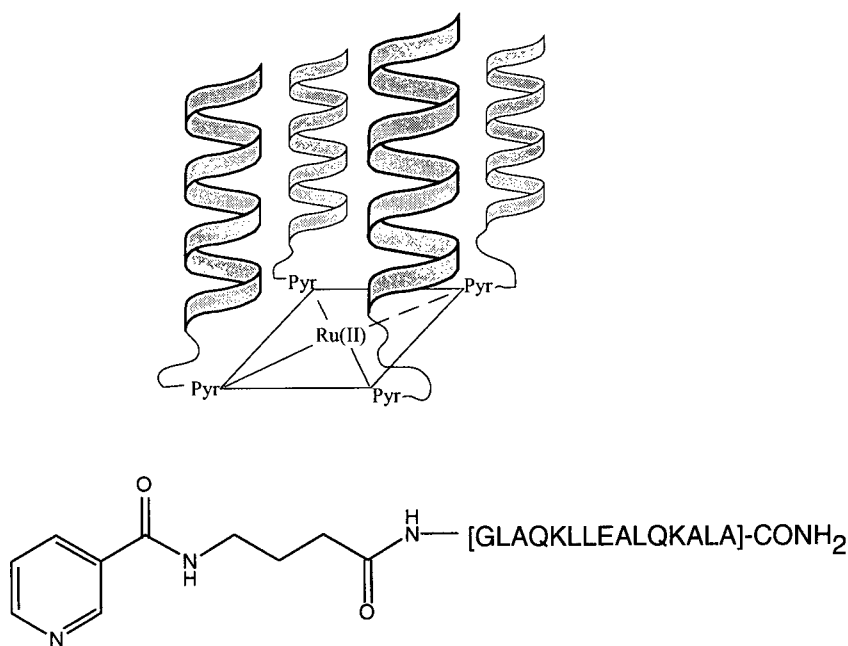
#### b. Pyridine- and Bipyridine-Based Templates

Ghadiri and coworkers have demonstrated the use of metal coordination in directing protein folding. They incorporated pyridyl moieties into the N-termini of their peptides and have shown that when bound to a Ru(II), a parallel four-helix bundle was formed (Figure

1.16).<sup>130</sup> The peptide sequence shown in Figure 1.16 was designed to be amphiphilic using appropriately spaced Leu and Ala residues. Some other design features include the spacing of Glu and Lys residues four residues apart in the sequence to allow for potential salt-bridge interactions. In addition, the C-terminus was amidated to offset unfavorable effects between the helix-dipole and a free C-terminal acid.

As with the peptide templates, the metal ion serves to orient four peptide strands close in space such that an intramolecular folding process can occur. Note the long linker group joining the peptide sequence to the pyridyl moiety to allow for bundle formation. This demonstrates that the ruthenium-based template may not be ideally preorganized for directing helical bundle assembly.

**Figure 1.16.** Ghadiri's Metal-Templated De Novo Four-Helix Bundle.

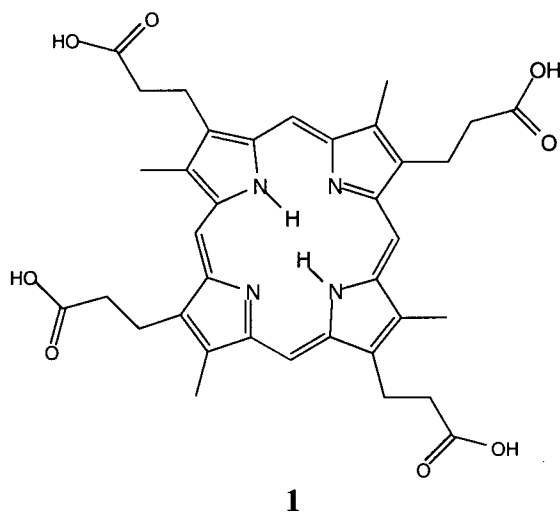


Three-helix bundles have also been synthesized with the aid of metal binding to 2,2'-bipyridyl moieties. This method was demonstrated independently by the Sasaki<sup>131</sup> and Ghadiri laboratories.<sup>132</sup>

As with the peptide templates, all of these metal-templated helical bundles possessed enhanced stability and helicity when compared to a non-templated peptide. Again, more detailed analyses investigating structural dynamics were not reported.

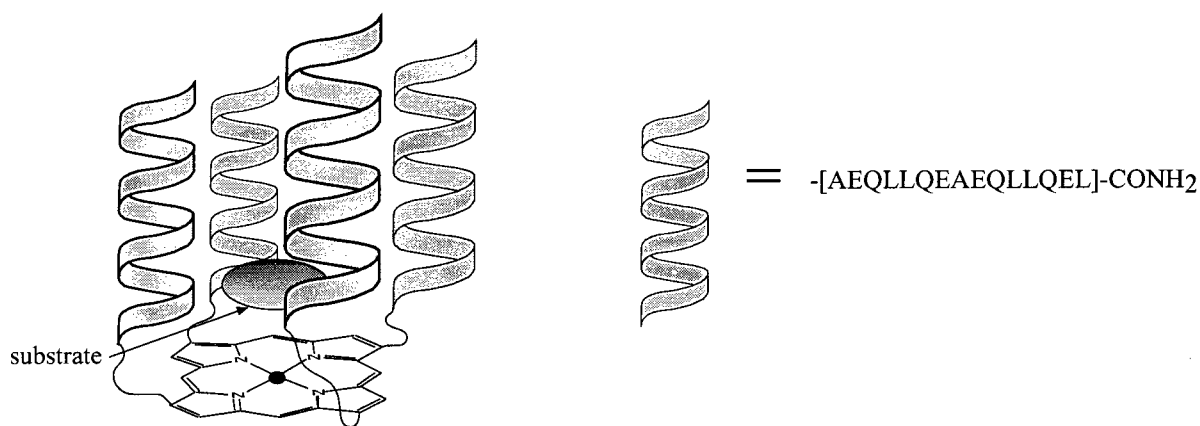
### c. Porphyrin Templates

Sasaki and Kaiser were the first to use porphyrin-based templates using coproporphyrin **1** and its four-fold symmetry to attach helical peptides to form a parallel four-helix bundle named "helichrome" (Figure 1.17).<sup>133</sup> The rigid nature of the porphyrin moiety was expected to have a pronounced effect on the stability of the bundle.



The peptide was designed to be amphiphilic but did not incorporate potential salt-bridge interactions. The C-terminus was amidated to avoid unfavorable helix-dipole interactions. In addition, helichrome was designed to mimic the hydroxylase activity of cytochrome P-450 where the active site is adjacent to several  $\alpha$ -helices which serve as a substrate binding pocket.

**Figure 1.17.** Sasaki and Kaiser's Porphyrin-Based De Novo Four-Helix Bundle "Helichrome".



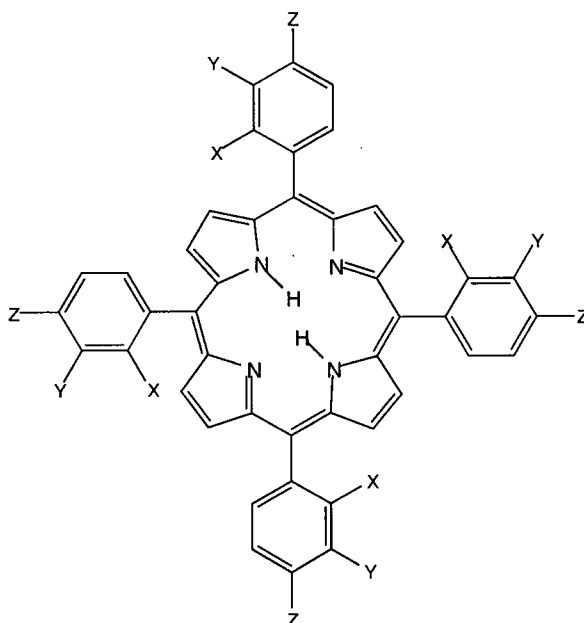
Biophysical studies on helichrome demonstrated enhanced helicity and stability when compared to isolated peptide fragments.<sup>134</sup> Kinetic studies on the catalytic activity of helichrome showed it to possess similar hydroxylase activity to that of native heme proteins.

DeGrado and coworkers have also explored the use of porphyrins as templates for four-helix bundle ion-channel formation.<sup>135</sup> They substituted the *meta* position of the phenyl groups in tetraphenylporphyrin (position 'Y' in Figure 1.18) with  $\alpha$ -helical proton-selective ion-channel-forming peptides to form so-called "tetraphilins." The added rigidity of the phenyl



substituent was hoped to play an added role in stability and peptide orientation. Although no stability studies were reported, it was found that these tetraphilins possessed a longer conductance state lifetime than that of unconstrained peptides of similar sequence. The authors attributed this increased lifetime to the stabilizing porphyrin template.

**Figure 1.18.** General Structure of Tetraphenyl Porphyrins Substituted at Different Positions X, Y and Z.



Nishino and coworkers also used tetraphenylporphyrins (Figure 1.18) for the construction of membrane-penetrating four-helix bundles.<sup>136</sup> They found that by substituting the phenyl group at the *ortho* position “X”, concentration independent CD spectra were observed which suggested that the desired four-helix bundle was formed. In contrast, substitution of the peptide groups at the *para* position “Z” led to concentration dependent CD spectra which

suggested that an intermolecular aggregate was formed. The authors state that this is likely the *para* substitution places the peptides too far from each other to form an intramolecular bundle.

Unfortunately, atropisomerism of the *ortho*-substituted porphyrin led to very low yields (3.8%) of this hybrid. This led to the synthesis of a new generation of *ortho*-substituted porphyrins incorporating an extra methoxy linker group.<sup>137</sup> This allowed for less hindered rotation about the phenyl-porphyrin bond and resulted in higher yields (35%). Although no stability studies were reported, the porphyrin-peptide hybrids were highly  $\alpha$ -helical and were capable of penetrating the egg yolk lecithin membrane as designed.

Much like the previously mentioned templated four-helix bundles, the structural dynamics of the above-mentioned porphyrin-based four-helix bundles were not studied in detail. Thus there exists a real lack of detailed knowledge as to the effect of templates in general on the dynamics of the tertiary structures in de novo four-helix bundles.

In addition, some of the templates mentioned here in Section F possess inherent disadvantages regarding their use in de novo four-helix bundle design. Briefly, Mutter's peptide templates may be too flexible to rigorously direct the peptides into a single orientation. Ghadiri's metal-templates require lengthy linker groups between the metal and the peptides to accommodate bundle formation; hence, the metal template may not be sufficiently preorganized to fully exploit the potential of template-stabilized structures. The rigidity of porphyrin templates show promise but there have been no reports explicitly studying the effect of porphyrin templates on the structures of de novo proteins – perhaps due to synthetic difficulties. This thesis will demonstrate the utility of cavitand templates (Scheme 1.2) in designing, synthesizing and studying de novo four-helix bundles.

## G. Conclusions

This introduction brought forward many different concepts regarding protein folding. In short, a multitude of factors are at play with a direct influence on the overall fold of any protein. It is thus not surprising that the protein folding problem remains elusive. The study of de novo proteins represents an alternative route to the study of protein folding by testing our understanding of the rules of protein folding. Although many successful de novo designs have been mentioned, our understanding of protein structure is only in its infancy. In addition, the “template assembly” approach bypasses some of the entropic costs of folding and thus represents a novel strategy in de novo design. The examples provided have illustrated the need for a more detailed analysis of the effect of templates on tertiary structures.

This thesis will explore the advantages and utility of cavitand macrocycles as templates in the design and synthesis of parallel de novo four-helix bundles. The extreme rigidity of cavitands and directionality of their reactive sites will be shown to have an enormous impact on the stabilities and structures of de novo designed four-helix bundles.

## H. References

1. Anfinsen, C. B. *Science* **1973**, *181*, 223-230.
2. Creighton, T. E. *Proteins: Structure and Molecular Properties*, 2<sup>nd</sup> Ed.; Freeman: New York, 1993.
3. Levinthal, C. *J. Chem. Phys.* **1968**, *65*, 44-45.
4. (a) Frishman, D.; Argros, P. *Proteins Struc. Funct. Genet.* **1997**, *27*, 329-335; (b) Banavar, J. R.; Cieplak, M.; Maritan, A.; Nadig, G.; Seno, F.; Vishveshwara, S. *Proteins Struc. Funct. Genet.* **1998**, *31*, 10-20; (c) Srinivasan, R.; Rose, G. D. *Proteins Struc. Funct. Genet.* **1995**, *22*, 81-99 and references cited therein.
5. Mathews, B. W. *Ann. Rev. Biochem.* **1993**, *62*, 139-160.
6. Bryson, J. W.; Betz, S. F.; Lu, H. S.; Suich, D. J.; Zhou, H. X.; O'Neil, K. T.; DeGrado, W. F. *Science* **1995**, *270*, 935-941.
7. (a) Lesk, A. M.; Chothia, C. *J. Mol. Biol.* **1980**, *136*, 225-270; (b) Doolittle, R. F. *Science* **1980**, *214*, 149-159.
8. DeGrado, W. F.; Wasserman, Z. R.; Lear, J. D. *Science* **1989**, *243*, 622-628.
9. (a) Lee, D. H.; Granja, J. R.; Martinez, J. A.; Severin, K.; Ghadiri, M. R. *Nature* **1996**, *382*, 525-528; (b) Severin, K.; Lee, D. H.; Kennan, A. J.; Ghadiri, M. R. *Nature* **1997**, *389*, 706-709; (c) Lee, D. H.; Severin, K.; Yokobayashi, Y.; Ghadiri, M. R. *Nature* **1997**, *390*, 591-594.
10. Johnsson, K.; Allemann, R. K.; Widmer, H.; Benner, S. A. *Nature* **1993**, *365*, 530-532.
11. (a) Broo, K. S.; Brive, L.; Ahlberg, P.; Baltzer, L. *J. Am. Chem. Soc.* **1997**, *119*, 11362-11372; (b) Broo, K. S.; Nilsson, H.; Nilsson, J.; Flodberg, A.; Baltzer, L. *J. Am. Chem. Soc.* **1998**, *120*, 4063-4068; (c) Broo, K. S.; Nilsson, H.; Nilsson, J.; Baltzer, L. *J. Am. Chem. Soc.* **1998**, *120*, 10287-10295.
12. Some deviations from planarity have been observed: MacArthur, M. W.; Thornton, J. M. *J. Mol. Biol.* **1996**, *264*, 1180-1195.
13. Barlow, D. J.; Thornton, J. M. *J. Mol. Biol.* **1988**, *201*, 601-619.
14. Pauling, L.; Corey, R. B.; Branson, H. R.; *Proc. Natl. Acad. Sci.* **1951**, *37*, 205-210.

15. Perutz, M. F. *Nature* **1951**, 167, 1053-1054.
16. (a) Wada, A.; *Adv. Biophys.* **1976**, 9, 1-63; (b) Hol, W. G. J.; van Duijnen, P. T.; Berendsen, H. J. C. *Nature* **1978**, 273, 443-446
17. Aurora, R.; Creamer, T. P.; Srinivasan, R.; Rose, G. D. *J. Biol. Chem.* **1997**, 272, 1413-1416.
18. Kohn, W. D.; Mant, C. T.; Hodges, R. S. *J. Biol. Chem.* **1997**, 272, 2583-2586.
19. Harris, N. L.; Presnell, S. R.; Cohen, F. E. *J. Mol. Biol.* **1994**, 236, 1356-1368.
20. Kamtekar, S.; Hecht, M. H. *FASEB J.* **1995**, 9, 1013-1022.
21. Crick, F. H. C. *Acta Crystallogr.* **1953**, 6, 689-697.
22. (a) Cohen, C.; Parry, D. A. D. *Trends Biochem. Sci.* **1986**, 11, 245-248; (b) Cohen, C.; Parry, D. A. D. *Proteins Struc. Func. Genet.* **1990**, 7, 1-15.
23. Banner, D. W.; Kokkinidis, M.; Tsernoglou, D. *J. Mol. Biol.* **1987**, 196, 657-675.
24. Kuwajima, K. *Proteins Struc. Funct. Genet.* **1989**, 6, 87-103.
25. Ptitsyn, O. B. *Adv. Prot. Chem.* **1995**, 47, 83-229.
26. Yon, J. M. *Cell. Mol. Life Sci.* **1997**, 53, 557-567.
27. Baker, D.; Agard, D. A. *Biochemistry* **1994**, 33, 7505-7509.
28. (a) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, 31, 526-535; (b) Lifson, S.; Roig, A. *J. Chem. Phys.* **1961**, 34, 1963-1974.
29. (a) Karplus, M.; Weaver, D. L. *Nature*, 260, 404-406; Karplus, M.; Weaver, D. L. *Protein Sci.* **1994**, 3, 650-658.
30. Kim, P.; Baldwin, R. L. *Ann. Rev. Biochem.* **1990**, 59, 631-660.
31. Chothia, C. *Ann. Rev. Biochem.* **1984**, 53, 537-572.
32. Ptitsyn, O. B. *FEBS Lett.* **1991**, 285, 176-181.
33. Dill, K. A. *Biochemistry* **1985**, 24, 1501-1509.
34. Harrison, S. C.; Durbin, R. *Proc. Natl. Acad. Sci.* **1985**, 82, 4028-4030.

35. Baldwin, R. L. *Nature* **1994**, 369, 183.
36. (a) Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. *Science* **1995**, 267, 1619-1620; (b) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins Struc. Funct. Genet.* **1995**, 21, 167-195.
37. Dill, K. A.; Chan, H. S. *Nature Struc. Biol.* **1997**, 4, 10-19.
38. Miranker, A. D.; Dobson, C. M. *Curr. Opin. Struc. Biol.* **1996**, 6, 31-42.
39. Baldwin, R. L. *J. Biomolec. NMR* **1995**, 5, 103-109.
40. Pande, V. S.; Grosberg, A. Y.; Tanaka, T.; Rokhsar, D. S. *Curr. Opin. Struc. Biol.* **1998**, 8, 68-79.
41. Jaenicke, R. *Prog. Biophys. Mol. Biol.* **1987**, 49, 117-237.
42. Buchner, J. *FASEB J.* **1996**, 10, 10-19.
43. Ruddon, R. W.; Bedows, E. *J. Biol. Chem.* **1997**, 272, 3125-3128.
44. Helenius, A. *Mol. Biol. Cell* **1994**, 5, 253-265.
45. Pace, C. N. *Trends. Biochem. Sci.* **1990**, 15, 14-17.
46. Doig, A. J.; Sternberg, M. J. E. *Protein Sci.* **1995**, 4, 2247-2251.
47. Huggins, M. L. *Angew. Chem Int. Ed. Engl.* **1971**, 10, 147-208.
48. Dill, K. A. *Biochemistry* **1990**, 29, 7133-7155.
49. Pace, C. N.; Shirley, B. A.; McNutt, M.; Gajiwala, K. *FASEB J.* **1996**, 10, 75-83.
50. Lins, L.; Brasseur, R. *FASEB J.* **1995**, 9, 535-540.
51. Privalov, P. L.; Gill, S. J. *Adv. Prot. Chem.* **1988**, 39, 191-234.
52. Chou, P. Y.; Fasman, G. *Annu. Rev. Biochem.* **1978**, 47, 251-276.
53. Chakrabartty, A.; Kortemme, T.; Baldwin, R. L. *Protein Sci.* **1994**, 3, 843-852.
54. Rohl, C. A.; Chakrabartty, A.; Baldwin, R. L. *Protein Sci* **1996**, 5, 2623-2637.
55. Lyu, P. C.; Liff, M. I.; Marky, L. A.; Kallenbach, N. R. *Science* **1990**, 669-673.

56. Gans, P. J.; Lyu, P. C.; Manning, M. C.; Woody, R. W.; Kallenbach, N. R. *Biopolymers* **1991**, *31*, 1605-1614.
57. Park, S. H.; Shalongo, W.; Stellwagen, E. *Biochemistry* **1993**, *32*, 7048-7053.
58. Park, S. H.; Shalongo, W.; Stellwagen, E. *Biochemistry* **1993**, *32*, 12901-12905.
59. (a) Blaber, M.; Zhang, X. J.; Mathews, B. W. *Science* **1993**, *260*, 1637-1640; (b) Blaber, M.; Zhang, X. J.; Lindstrom, J. D.; Pepiot, S. D.; Baase, W.; Mathews, B. W. *J. Mol. Biol.* **1994**, *235*, 600-624.
60. Horovitz, A.; Mathews, J. M.; Fersht, A. R. *J. Mol. Biol.* **1992**, *227*, 560-568.
61. (a) O'Neil, K. T.; DeGrado, W. F. *Science* **1990**, *250*, 646-651; (b) Zhou, et. al. *Protein and Peptide Letters* **1994**, *1*, 114-119; (c) Monera, et. al. *J. Peptide Science* **1995**, *1*, 319-329.
62. Chakrabarty, A.; Baldwin, R. L. *Adv. Protein Chem.* **1995**, *46*, 141-176.
63. Myers, J. K.; Pace, C. N.; Scholtz, J. M. *Biochemistry* **1997**, *36*, 10923-10929.
64. Creamer, T. P.; Rose, G. D. *Proc. Natl. Acad. Sci.* **1992**, *89*, 5937-5941.
65. Minor, D. L., Jr.; Kim, P. S. *Nature* **1996**, *380*, 730-734. Similar context dependence was also demonstrated by Kwok, S. C.; Tripet, B.; Man, J. H.; Chana, M. S. Lavigne, P.; Mant, C. T.; Hodges, R. S. *Biopolymers* **1998**, *47*, 101-123.
66. (a) Ihara, S.; Ooi, T.; Takahashi, S. *Biopolymers* **1982**, *21*, 131-145; (b) Shoemaker, K. R.; Kim, P. S.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Nature* **1987**, *326*, 563-567; (c) Huyghues-Despointes, B. M. P.; Scholtz, J. M.; Baldwin, R. L. *Protein Sci.* **1993**, *2*, 1604-1611; (d) Lockhart, D. S.; Kim, P. S. *Science* **1993**, *260*, 198-202; (e) Kohn, et. al. *J. Peptide Science* **1997**, *3*, 209-233.
67. (a) Aurora, R.; Rose, G. D. *Protein Sci.* **1998**, *7*, 21-38; (b) Presta, L. G.; Rose, G. D. *Science* **1988**, *240*, 1632-1641; (c) Richardson, J. S.; Richardson, D. C. *Science* **1988**, *240*, 1648-1652.
68. (a) Bruch, M. D.; Dhingra, M. M.; Gierasch, L. M. *Proteins* **1991**, *10*, 130-139; (b) Lyu, P. C.; Wemmer, D. E.; Zhou, H. Z.; Pinker, R. J.; Kallenbach, N. R. *Biochemistry* **1993**, *32*, 421-425; (c) Forood, B.; Feliciano, E. J.; Nambiar, K. P. *Proc. Natl. Acad. Sci.* **1993**, *90*, 838-842; (d) Chakrabarty, A.; Doig, A. J.; Baldwin, R. L. *Proc. Natl. Acad. Sci.* **1993**, *90*, 11332-11336; (e) Doig, A. J.; Baldwin, R. L. *Protein Sci.* **1995**, *4*, 1325-1336; (f) Petukhov, M.; Yumoto, N.; Murase, S.; Onmura, R.; Yoshikawa, S. *Biochemistry* **1996**, *35*, 387-397.

69. (a) Serrano, L.; Fersht, A. R. *Nature* **1989**, *342*, 296-299; (b) Serrano, L.; Neira, J. L.; Sancho, J.; Fersht, A. R. *Nature* **1992**, *356*, 453-455; (c) Serrano, L.; Sancho, J.; Hirshberg, M.; Fersht, A. R. *J. Mol. Biol.* **1992**, *227*, 544-599; (d) Bell, J. A.; Becktel, W. J.; Sauer, U.; Baase, W. A.; Mathews, B. W. *Biochemistry* **1992**, *31*, 3590-3596.
70. Harper, E. T.; Rose, G. D. *Biochemistry* **1993**, *32*, 7605-7609.
71. Munoz, V. Blanco, F. J.; Serrano, L. *Nature Struc. Biol.* **1995**, *2*, 380.
72. Schneider, J. P.; DeGrado, W. F. *J. Am. Chem. Soc.* **1998**, *120*, 2764-2767.
73. (a) Aurora, R.; Srinivasan, R.; Rose, G. D. *Science* **1994**, *264*, 1126-1130; (b) Viguera, A. R.; Serrano, L. *J. Mol. Biol.* **1995**, *251*, 150-160; (c) Datta, S.; Shamala, N.; Banerjee, A.; Pramanik, A.; Bhattacharjya, S.; Balaram, P. *J. Am. Chem. Soc.* **1997**, *119*, 9246-9251.
74. Prieto, J.; Serrano, L. *J. Mol. Biol.* **1997**, *274*, 276-288.
75. (a) Sholtz, J. M.; Qian, H.; Robbins, V. H.; Baldwin, R. L. *Biochemistry* **1993**, *32*, 9668-9678; (b) Smith, J. S.; Sholtz, J. M. *Biochemistry* **1998**, *37*, 33-40.
76. Armstrong, K. M.; Fairman, R.; Baldwin, R. L. *J. Mol. Biol.* **1993**, *230*, 284-291.
77. Padmanabhan, S.; Baldwin, R. L. *J. Mol. Biol.* **1994**, *241*, 706-713.
78. Huyghues-Despointes, B. M. P.; Klinger, T. M.; Baldwin, R. L. *Biochemistry* **1995**, *41*, 13268-13271.
79. Huyghues, B. M. P.; Baldwin, R. L. *Biochemistry* **1997**, *36*, 1965-1970.
80. (a) Lau, S. Y. M.; Taneja, A. K.; Hodges, R. S. *J. Biol. Chem.* **1984**, *259*, 13253-13261; (b) Su, J. Y.; Hodges, R. S.; Kay, C. M. *Biochemistry* **1994**, *33*, 15501-15510.
81. Fairman, R.; Chao, H.-G.; Mueller, L.; Lavoie, T. B.; Shen, L.; Novotny, J.; Matsueda, G. R. *Protein Sci.* **1995**, *4*, 1457-1469.
82. Walther, D.; Eisenhaber, F.; Argos, P. *J. Mol. Biol.* **1996**, *255*, 536-553.
83. Chothia, C.; Levitt, M.; Richardson, D. *J. Mol. Biol.* **1981**, *145*, 215-250.
84. Bowie, J. U. *Nature Struc. Biol.* **1997**, *4*, 915-919.
85. (a) Zhou, N. E.; Kay, C. M.; Hodges, R. S. *Protein Eng.* **1994**, *7*, 1365-1372. For inter-helical electrostatic repulsions see: (b) Kohn, W. D.; Monera, O. D.; Kay, C. M.;



- Hodges, R. S. *J. Biol. Chem.* **1995**, *270*, 25495-25506; (c) Kohn, W. D.; Kay, C. M.; Hodges, R. S. *Protein Sci.* **1995**, *4*, 237-250.
86. Monera, O. D.; Kay, C. M.; Hodges, R. S. *Biochemistry* **1994**, *33*, 3862-3871.
  87. Robinson, C. R.; Sligar, S. G. *Protein Sci.* **1993**, *2*, 826-837.
  88. (a) Quinn, T. P.; Tweedy, N. B.; Williams, R. W.; Richardson, J. S.; Richardson, D. C. *Proc. Natl. Acad. Sci.* **1994**, *91*, 8747-8751; (b) Yan, Y.; Erickson, B. W. *Protein Sci.* **1994**, *3*, 1069-1073; (c) Blondelle, S. E.; Forood, B.; Houghten, R. A.; Perez-Paya, E. *Biochemistry* **1997**, *36*, 8393-8400; (d) Schenck, H. L.; Gellman, S. H. *J. Am. Chem. Soc.* **1998**, *120*, 4869-4870; (e) Sharman, G. J.; Searle, M. S.; *J. Am. Chem. Soc.* **1998**, *120*, 5291-5300; (f) Kortemme, T.; Ramirez-Alvarado, M.; Serrano, L. *Science* **1998**, *281*, 253-256; (g) Smith, C. K.; Regan, L. *Acc. Chem. Res.* **1997**, *30*, 153-161.
  89. (a) Struthers, M. D.; Cheng, R. P.; Imperiali, B. *Science* **1996**, *271*, 342-345; (b) Struthers, M. D.; Cheng, R. P.; Imperiali, B. *J. Am. Chem. Soc.* **1996**, *118*, 3073-3081; (c) Dahiyat, B. I.; Mayo, S. L. *Science* **1997**, *278*, 82-87; (d) Dahiyat, B. I.; Sarisky, C. A.; Mayo, S. L. *J. Mol. Biol.* **1997**, *273*, 789; (e) Malakauskas, S. M.; Mayo, S. L. *Nature Struc. Biol.* **1998**, *5*, 470-475.
  90. For a recent example, see: Kohn, W. D.; Kay, C. M.; Sykes, B. D.; Hodges, R. S. *J. Am. Chem. Soc.* **1998**, *120*, 1124-1132 and references cited therein.
  91. (a) Schneider, J. P.; Lombardi, A.; DeGrado, W. F. *Folding & Design* **1998**, *3*, 29-40; (b) Lombardi, A.; Bryson, J. W.; DeGrado, W. F. *Biopolymers* **1997**, *40*, 495-504.
  92. Eisenberg, D.; Wilcox, W.; Eshita, S. M.; Pryciak, P. M.; Ho, S. P.; DeGrado, W. F. *Proteins Struc. Funct. Genet.* **1986**, *1*, 16-22.
  93. Ho, S. P.; DeGrado, W. F. *J. Am. Chem. Soc.* **1987**, *109*, 6751-6758.
  94. Regan, L.; DeGrado, W. F. *Science* **1988**, *241*, 976-978.
  95. Osterhout, J. J.; Handel, T.; Na, G.; Toumadje, A.; Long, R. C.; Connolly, P. J.; Hoch, J. C.; Johnson Jr., W. C.; Live, D.; DeGrado, W. F. *J. Am. Chem. Soc.* **1992**, *114*, 331-337.
  96. Handel, T. M.; Williams, S. A.; DeGrado, W. F. *Science* **1993**, *261*, 879-885.
  97. Raleigh, D. P.; DeGrado, W. F. *J. Am. Chem. Soc.* **1992**, *114*, 10079-10081.
  98. Handel, T.; DeGrado, W. F. *J. Am. Chem. Soc.* **1990**, *112*, 6710-6711.
  99. Raleigh, D. P.; Betz, S. F.; DeGrado, W. F. *J. Am. Chem. Soc.* **1995**, *117*, 7558-7559.

100. Hill, R. B.; DeGrado, W. F. *J. Am. Chem. Soc.* **1998**, *120*, 1138-1145.
101. Regan, L.; Rockwell, A.; Wasserman, Z.; DeGrado, W. F. *Protein Sci.* **1994**, *3*, 2419-2427.
102. Betz, S. F.; DeGrado, W. F. *Biochemistry* **1996**, *35*, 6955-6962.
103. Betz, S. F.; Liebman, P. A.; DeGrado, W. F. *Biochemistry* **1997**, *36*, 2450-2458.
104. Hecht, M. H.; Richardson, J. S.; Richardson, D. C.; Ogden, R. C. *Science* **1990**, *249*, 884-891.
105. Kamtekar, S.; Schiffer, J. M.; Xiong, H.; Babik, J. M.; Hecht, M. H. *Science* **1993**, *262*, 1680-1685.
106. Roy, S.; Ratnaswamy, G.; Boice, J. A.; Fairman, R.; McLendon, G.; Hecht, M. H. *J. Am. Chem. Soc.* **1997**, *119*, 5302-5306.
107. (a) Brive, L.; Dolphin, G. T.; Baltzer, L. *J. Am. Chem. Soc.* **1997**, *119*, 8598-8607; (b) Dolphin, G. T.; Brive, L.; Johansson, G.; Baltzer, L. *J. Am. Chem. Soc.* **1996**, *118*, 11297-11298.
108. (a) Johansson, J. S.; Gibney, B. R.; Rabanal, F.; Reddy, K. S.; Dutton, P. L. *Biochemistry* **1998**, *37*, 1421-1429; (b) Gibney, B. R.; Rabanal, F.; Reddy, K. S.; Dutton, P. L. *Biochemistry* **1998**, *37*, 4635-4643; (c) Sharp, R. E.; Diers, J. R.; Bocian, D. F.; Dutton, P. L. *J. Am. Chem. Soc.* **1998**, *120*, 7103-7104; (d) Robertson, D. E.; Farid, R. S.; Moser, C. C.; Mullholland, S. E.; Pidikiti, R.; Lear, J. D.; Wand, A. J.; DeGrado, W. F.; Dutton, P. L. *Nature* **1994**, *368*, 425.
109. Gibney, B. R.; Rabanal, F.; Skalicky, J. J.; Wand, A. J.; Dutton, P. L. *J. Am. Chem. Soc.* **1997**, *119*, 2323-2324.
110. Jiang, X.; Bishop, E. J.; Farid, R. S. *J. Am. Chem. Soc.* **1997**, *119*, 838-839.
111. O'Shea, E. K.; Rutkowski, R.; Kim, P. S. *Science* **1989**, *243*, 538-542.
112. Harbury, P. B.; Zhang, T.; Kim, P. S.; Alber, T. *Science* **1993**, *262*, 1401-1407.
113. A similar article reported that the relative positions of alanine in the hydrophobic core could control the formation of two- and four-stranded coiled-coils. Monera, O. D.; Sonnichsen, F. D.; Hicks, L.; Kay, C. M.; Hodges, R. S. *Protein Eng.* **1996**, *9*, 353-363.

114. (a) Munson, M.; O'Brien, R.; Sturtevant, J. M.; Regan, L. *Protein Sci.* **1994**, *3*, 2015-2022. (b) Munson, M.; Balasubramanian, S.; Fleming, K. G.; Nagi, A. D.; O'Brien, R.; Sturtevant, J. M.; Regan, L. *Protein Sci.* **1996**, *5*, 1584-1593.
115. Predki, P. F.; Agrawal, V.; Brunger, A. T.; Regan, L. *Nature Struct. Biol.* **1996**, *3*, 54-58.
116. Nagi, A. D.; Regan, L. *Folding & Design* **1997**, *2*, 67-75.
117. Predki, P. F.; Regan, L. *Biochemistry* **1995**, *34*, 9834-9839.
118. Fairman, R.; Chao, H-G.; Lavoie, T. B.; Villafranca, J. J.; Matsueda, G. R.; Novotny, J. *Biochemistry* **1996**, *35*, 2824-2829.
119. Mutter, M.; Vuilleumeir, S. *Angew. Chem. Int. Ed. Engl.* **1989**, *28*, 535-554.
120. Schneider, J. P.; Kelly, J. W. *Chem. Rev.* **1995**, *95*, 2169-2187.
121. (a) Tsang, K. Y.; Diaz, H.; Graciani, N.; Kelly, J. W. *J. Am. Chem. Soc.* **1994**, *116*, 2988-4005; (b) Choo, D. W.; Schneider, J. P.; Graciani, N. R.; Kelly, J. W. *Macromolecules* **1996**, *29*, 355-366.
122. Schneider, J. P.; Kelly, J. W. *J. Am. Chem. Soc.* **1995**, *117*, 2533-2546.
123. Holmes, D. L.; Smith, E. M.; Nowick, J. S. *J. Am. Chem. Soc.* **1997**, *119*, 7665-7669; Nowick, J. S.; Smith, E. M.; Pairish, M. *Chem. Soc. Rev.* **1996**, *25*, 401.
124. (a) Mutter, M.; Tuchscherer, G. *Macromol. Chem. Rapid. Commun.* **1988**, *9*, 437-443; (b) Mutter, M.; Altmann, E.; Altmann, K.; Hersperger, R.; Koziej, P.; Nebel, K.; Tuchscherer, G.; Vuilleumier, S. *Helv. Chim. Acta* **1988**, *71*, 835-847.
125. Mutter, M.; Tuchscherer, G. G.; Miller, C.; Altmann, K.; Carey, R. I.; Wyss, D. F.; Labhardt, A. M.; Rivier, J. E. *J. Am. Chem. Soc.* **1992**, *114*, 1463-1470.
126. Tuchscherer, G.; Dömer, B.; Sila, U.; Kamber, B.; Mutter, M. *Tetrahedron* **1993**, *49*, 3559-3575.
127. DeGrado, W. F. *Adv. Prot. Chem.* **1988**, *39*, 51-124.
128. Vuilleumeir, S.; Mutter, M. *Biopolymers* **1993**, *33*, 389.
129. Pawlak, M.; Meseth, U.; Dhanpal, B.; Mutter, M.; Vogel, H. *Protein Sci.* **1994**, *3*, 1788-1805.
130. Ghadiri, M. R.; Soares, C.; Choi, C. *J. Am. Chem. Soc.* **1992**, *114*, 4000-4002.

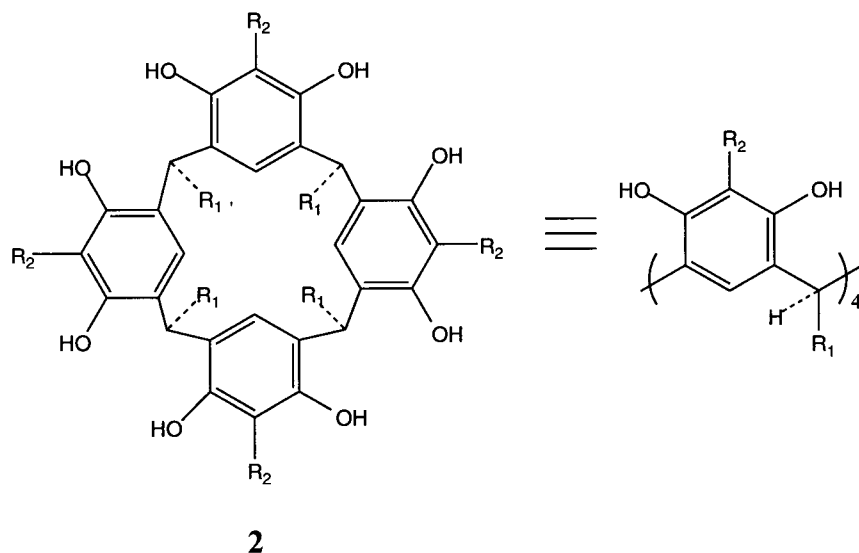
131. Lieberman, M.; Sasaki, T. *J. Am. Chem. Soc.* **1991**, *113*, 1470-1471.
132. (a) Ghadiri, M. R.; Soares, C.; Chio, C. *J. Am. Chem. Soc.* **1992**, *114*, 825-831; (b) Ghadiri, M. R.; Case, M. A. *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 1594-1597.
133. Sasaki, T.; Kaiser, E. T. *J. Am. Chem. Soc.* **1989**, *111*, 380-381.
134. Sasaki, T.; Kaiser, E. T. *Biopolymers* **1990**, *29*, 79-88.
135. Akerfeldt, K. S.; Kim, R. M.; Camac, D.; Groves, J. T.; Lear, J. D.; DeGrado, W. F. *J. Am. Chem. Soc.* **1992**, *114*, 9656-9657.
136. Mihara, H.; Nishino, N.; Hasegawa, R.; Fujimoto, T. *Chem. Lett.* **1992**, 1805-1808.
137. Arai, T.; Kobata, K.; Mihara, H.; Fujimoto, T.; Nishino, N. *Bull. Chem. Soc. Jpn.* **1995**, *68*, 1989-1998.

## Chapter Two: Synthesis of Cavitands Suitable For Incorporation Into De Novo Proteins

### A. Introduction to Cavitands

#### i. Resorcin[4]arenes

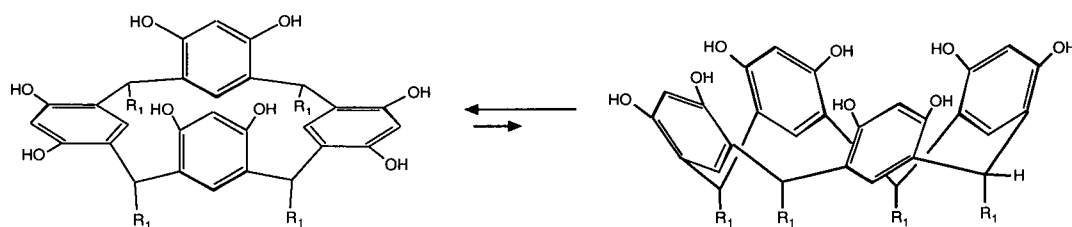
Over 125 years ago, Adolf von Bayer reported that the addition of sulfuric acid to a mixture of benzaldehyde and resorcinol followed by heating gave an unknown crystalline compound.<sup>1</sup> In 1940, the cyclic tetrameric macrocycle **2** was proposed as a possible structure for this material based on molecular weight determinations.<sup>2</sup> However, it was not until 1968 that the three-dimensional structure of macrocycle **2** was confirmed by X-ray diffraction.<sup>3</sup> Although several trivial names for macrocycle **2** have been used, one commonly accepted name is *resorcinarene*, or *resorcin[4]arene*, which will be used here.



Resorcinarenes can readily be synthesized in a one-step process involving an acid-catalyzed condensation between an aliphatic or aromatic aldehyde and resorcinol.<sup>1c</sup> Many different resorcinarene derivatives have been synthesized with a variety of  $R_1$  and  $R_2$  groups.<sup>1c</sup>

Due to the nature of the reaction, the macrocycle may be present in different stereoisomeric forms, which can interconvert under the reaction conditions. The most useful of these isomers is the *cis-cis-cis* (ccc) isomer in which the all four  $R_1$  groups are cis with respect to one another. Typically, the ccc isomer is the thermodynamically most stable and least soluble isomer and hence can be isolated with relative ease. In addition, the ccc isomer is known to interconvert via bond rotation between the  $C_{2v}$  symmetric “boat” isomer and the  $C_{4v}$  symmetric “crown” isomer (Figure 2.1).<sup>4</sup>

**Figure 2.1.** Schematic Illustration of the Interconversion Between the Two ccc Isomers: Boat (left) and Crown (right).  $R_2$  Groups Have Been Omitted for Clarity.



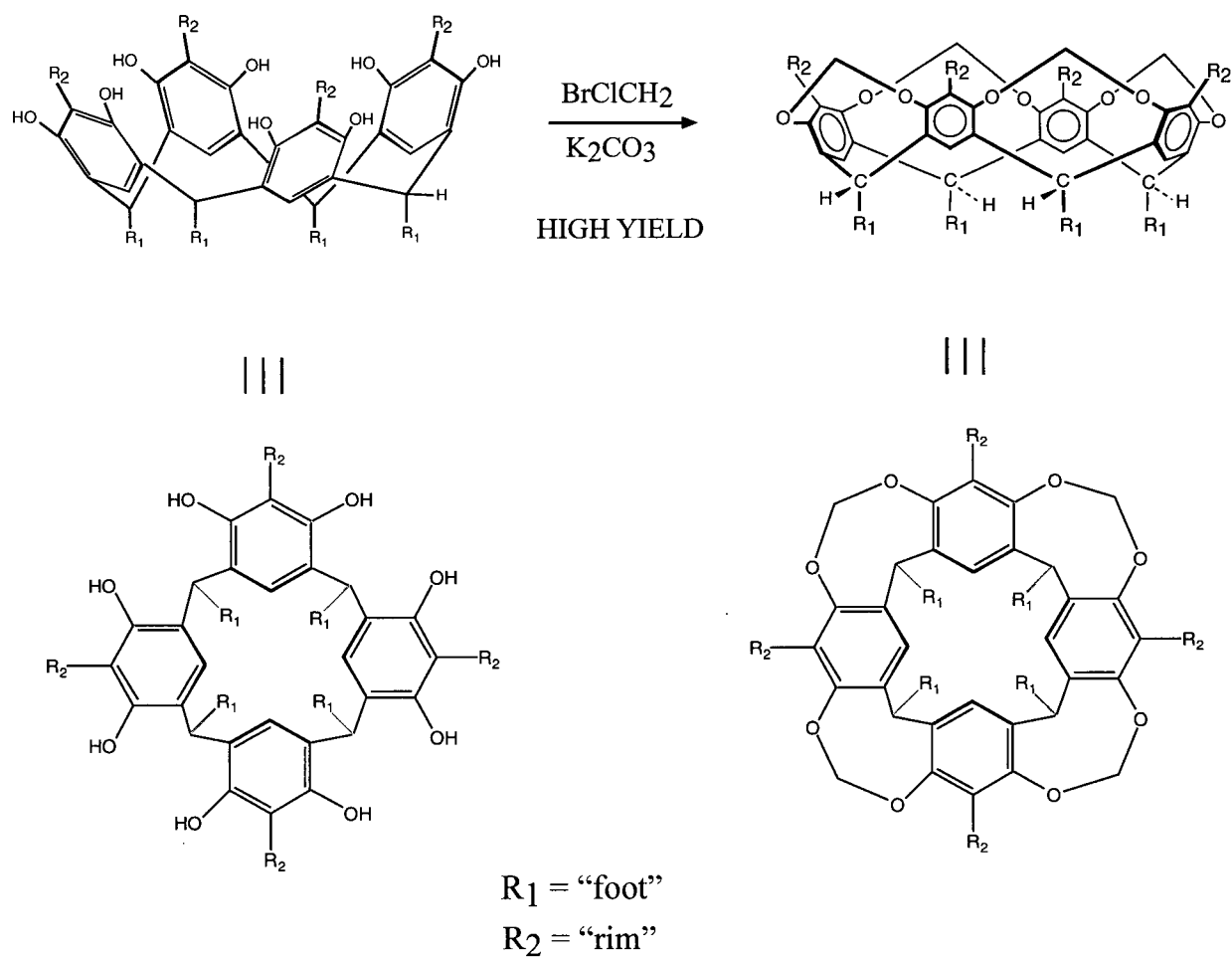
With respect to de novo protein synthesis, one desirable aspect of a potential template molecule would be rigidity. As briefly mentioned in Chapter One, a rigid template would decrease the conformational flexibility of the peptide structure and may lead to a more stable and native-like structure. Hence, upon consideration of the dynamics of the ccc isomer, such a resorcinarene may be too flexible to impart the desired rigidity into the de novo design.

## ii. Methylene-Bridged Resorcin[4]arenes

The term *cavitand* was coined by Cram in 1982 to describe molecules that contain an enforced cavity large enough to contain small molecules or ions.<sup>5,6</sup> Cram was the first to impart rigidity to resorcinarenes by adding methylene bridging groups between adjacent phenolic groups of the ccc resorcinarene isomer, which effectively locks the resorcinarene into a crown-like conformation providing a permanent enforced cavity within the molecule (Scheme 2.1).<sup>5</sup>

Methylene-bridged resorcinarenes, or cavitands as they will be referred to for the remainder of the thesis, are extremely rigid molecules. In efforts to exploit this rigidity, a variety of cavitands have been synthesized and have found uses in (1) carceplex synthesis, where the chemical and physical properties of molecules can be altered by encapsulation,<sup>7</sup> (2) binding neutral guest molecules,<sup>8</sup> useful, for example in the removal of toxins from waste water,<sup>8f</sup> and (3) monolayer formation, which have potential as molecular sensors.<sup>9</sup> We plan to exploit the rigidity of these cavitands in the design of de novo four-helix bundles. This chapter discusses the syntheses of various cavitands while Chapter Four will discuss in detail the design rationale for use of these cavitands in de novo protein synthesis.

**Scheme 2.1.** General Synthesis of Methylene-Bridged Resorcin[4]arenes or Cavitands.



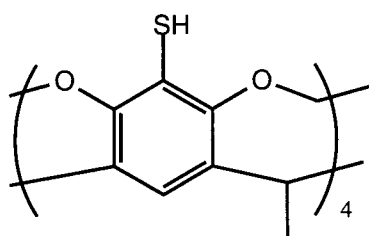


## B. Results and Discussion

### i. Considerations and Requirements for Peptide Incorporation Onto Cavitands

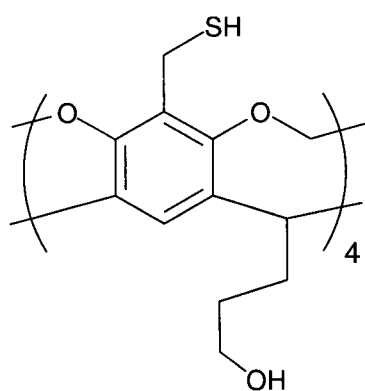
Cavitands possess four sites at their “rims” (position  $R_2$ ) that may be transformed into useful functional groups. We sought to incorporate a highly nucleophilic functional group into this site such that it may react with a complementary electrophilic site on the peptide to be incorporated. To this end, a highly nucleophilic thiophenol was initially incorporated into a cavitand.

Since the de novo proteins presented in this thesis are intended to be studied in aqueous solution, it is important to minimize the hydrophobicity of the cavitand template in order to prevent intermolecular aggregation via the hydrophobic effect. Typically, long alkyl chains have been incorporated into the “foot” position ( $R_1$ ) of cavitands to aid their solubility in organic solvents.<sup>10</sup> However, for better aqueous solubility, methyl groups were initially incorporated into this “foot” position of thiol cavitands. A recent study in our group<sup>11</sup> has shown that water-soluble cavitands with methyl feet show concentration independent  $^1\text{H}$  NMR spectra in  $\text{D}_2\text{O}$ . In contrast, an analogue with more hydrophobic phenyl-ethyl feet displayed concentration dependent  $^1\text{H}$  NMR spectra  $\text{D}_2\text{O}$ . This suggests that water-soluble cavitands with methyl feet remain monomeric in aqueous solution while those with more hydrophobic feet form aggregates. To this end, methyl-footed tetrathiol cavitand **3** was an initial synthetic target for de novo protein synthesis.

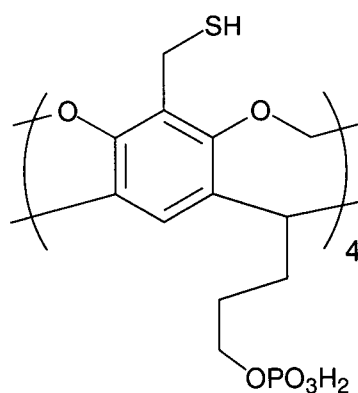


3

In addition to methyl feet, alkyl-hydroxyls and alkyl-phosphates were also incorporated into the “feet” (position R<sub>1</sub>) of thiol cavitands to further explore the effect of hydrophilic/hydrophobic feet on the structure of the de novo protein. For these compounds, benzylthiol cavitands were targeted as an initial starting point since their synthesis was expected to be much more straightforward than that of the above-mentioned aryl thiol cavitand. We, therefore, set out to synthesize both hydroxyl-footed benzylthiol cavitand **4** and phosphate-footed benzylthiol cavitand **5**.



4



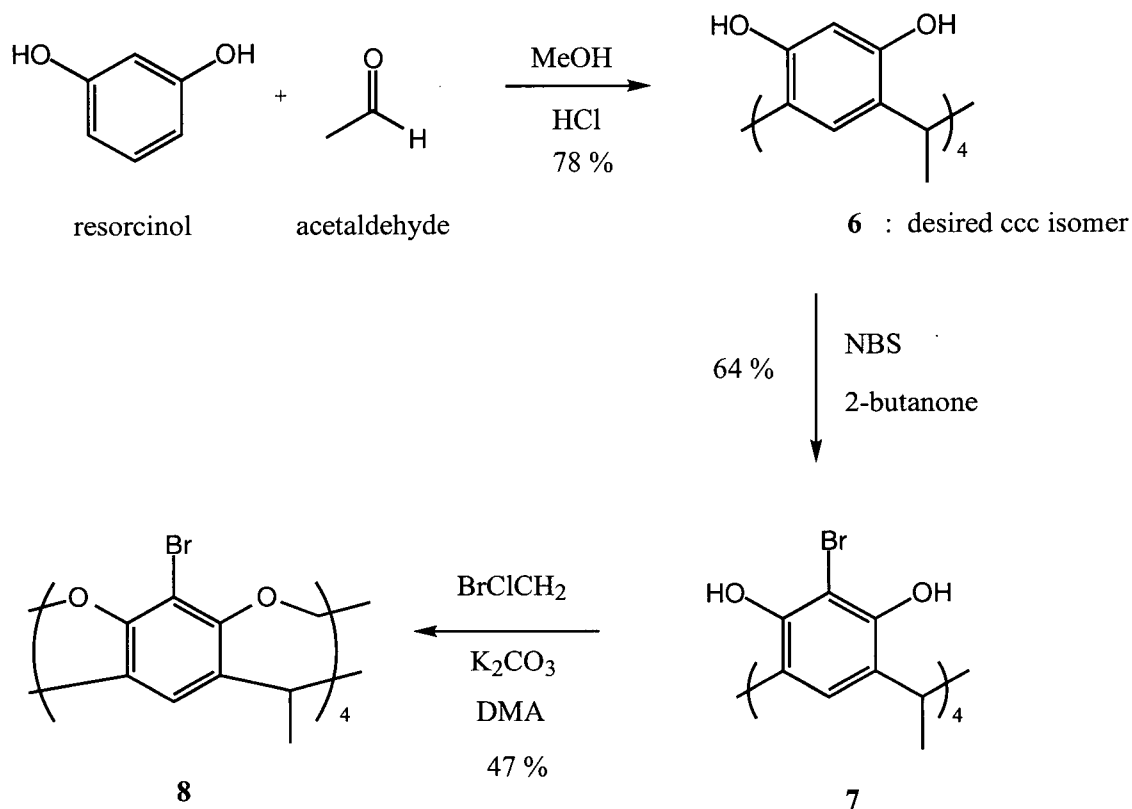
5

Interestingly, the phosphate-footed cavitands presented in this chapter represent the first example of cavitands deriving their aqueous solubility from their feet. This method of introducing water-solubility leaves the rim positions available for further functionalization, a feature which may be important and useful for developing more complex water-soluble host molecules.

ii. Synthesis of Methyl-Footed Tetrathiol **3**

Bromocavitand **8** was synthesized as described in literature procedures.<sup>5</sup> Briefly, resorcinol and acetaldehyde were stirred in methanol and HCl. Over the course of one week, resorcinarene **6** precipitated out from the reaction mixture (78%, Scheme 2.2). Treatment of macrocycle **6** with *N*-bromosuccinimide (NBS) in 2-butanone afforded tetrabromoresorcinarene **7** (64%), which was then bridged with bromochloromethane in the presence of potassium carbonate to yield bromocavitand **8** (47%).

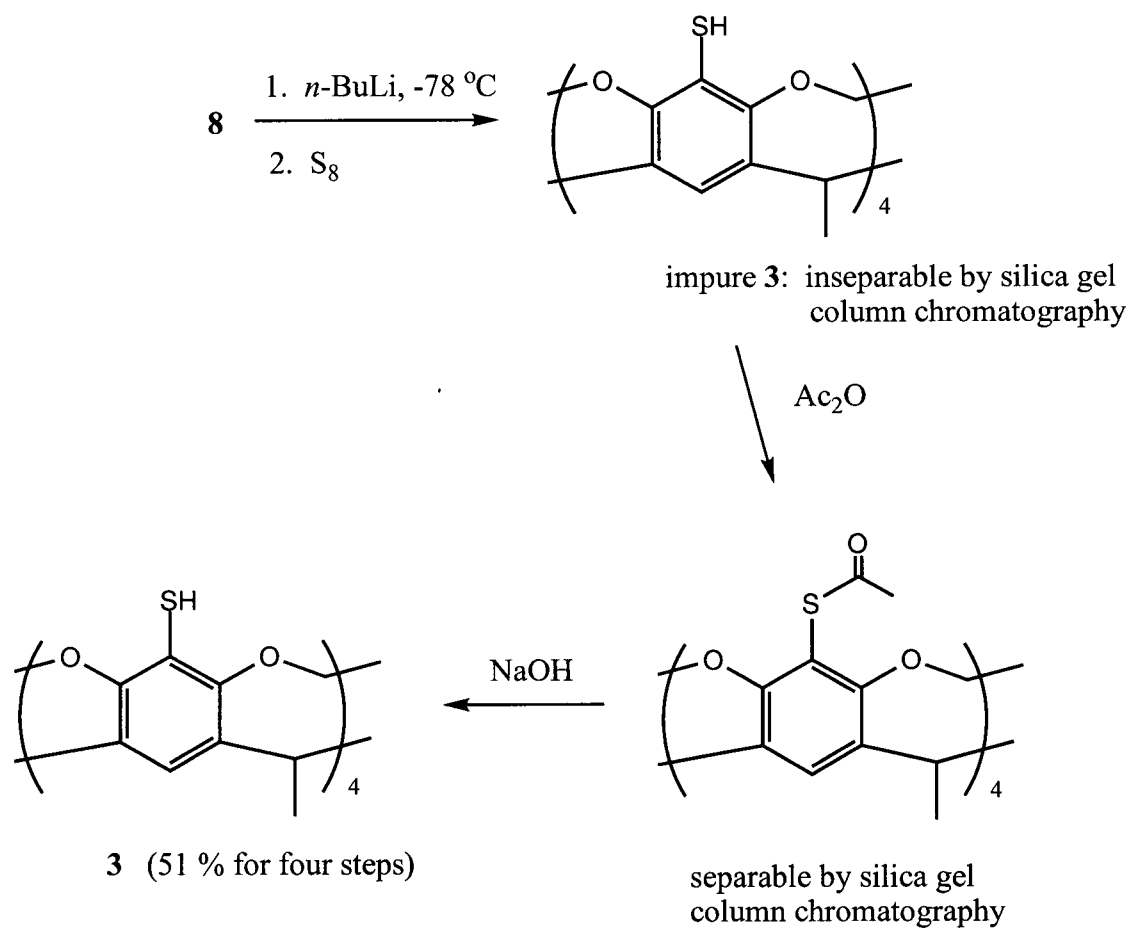
**Scheme 2.2.** Synthesis of Bromocavitand **8**.



Treatment of tetrabromocavitand **8** with *n*-butyllithium at  $-78\text{ }^\circ\text{C}$  followed by quenching with  $\text{S}_8$  led to a mixture of thiol cavitand products (Scheme 2.3). This mixture was inseparable by silica gel column chromatography and consisted of tetrathiol cavitand **3** and ca. 25% of the corresponding tris-thiol. However, acetylation of the thiol groups with acetic anhydride in pyridine afforded the *S*-acetyl analogues which were readily separable by column chromatography. Subsequent removal of the acetyl groups with sodium hydroxide afforded

pure cavitand **3** (51% for 4 steps) suitable for de novo protein synthesis.<sup>12,13</sup> Note that all of the results in this Chapter have been published and are referenced accordingly.

**Scheme 2.3.** Synthesis of Tetrathiol Cavitand **3**.



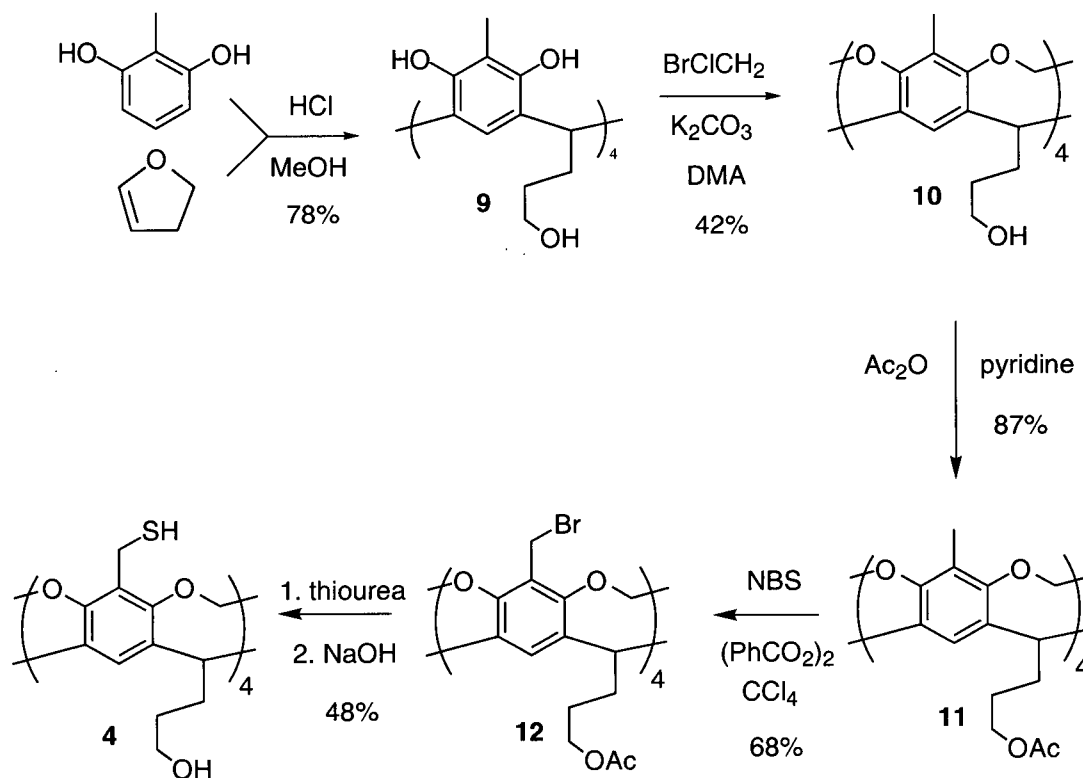
### iii. Synthesis of Hydroxyl- and Phosphate-Footed Tetrathiols **4** and **5**

In 1989, Cram introduced a method to incorporate hydroxyl groups into the feet ( $R_1$  in structure **2**) of cavitands during the initial condensation step to form butanol-footed resorcinarenes.<sup>14</sup> More recently, our group developed a method to selectively bridge the phenolic hydroxyls to afford the first hydroxyl-footed methylene-bridged cavitands.<sup>15</sup> We used this methodology in conjunction with a series of orthogonal protection and deprotection steps to incorporate hydroxyl and phosphate feet in our set of tetrathiol cavitands. The syntheses of compounds **4** – **21** have been recently published.<sup>16</sup>

In the syntheses of the following hydrophilic cavitands, we chose an alternative approach to tetrathiol cavitands which involved the synthesis of benzylthiol moieties. In a recent communication, Sorrell and coworkers have demonstrated that methyl groups in the rim position ( $R_2$ ) can lead to a variety of different functional groups.<sup>17</sup> This method was expected to be a milder and cleaner reaction which, in turn, may allow for a much simpler purification as compared to the aryl thiol system mentioned above.

In order to minimize the hydrophobicity of the pendent hydroxyl feet, 2,3-dihydrofuran was used as the latent aldehyde to introduce short propanol feet. In addition, 2-methyl resorcinol was used to introduce the desired aryl methyl group. The resulting acid-catalyzed condensation reaction of 2-methyl resorcinol and 2,3-dihydrofuran proceeded in 78% yield to give hydroxyl-footed resorcin[4]arene **9**. Selective bridging of the phenolic hydroxyl groups with  $\text{BrClCH}_2$  afforded rigid methylene-bridged 2-methyl cavitand **10** in 42% yield.

**Scheme 2.4.** Synthesis of Hydroxyl-Footed Tetrathiol **4**.



The first target cavitand was hydroxyl-footed benzylthiol **4** accessible via Sorrell's bromination method.<sup>17</sup> Initial attempts to directly brominate hydroxyl-footed cavitand **10** failed due to its low solubility in the desired solvent  $\text{CCl}_4$ . It was found that acetylation of the pendent hydroxyl groups with acetic anhydride in pyridine afforded cavitand **11** (87%) which was soluble in  $\text{CCl}_4$  and stable to the subsequent radical bromination conditions. Accordingly, radical bromination with NBS afforded acetyl-protected benzylbromide **12** (68%). Treatment of cavitand **12** with thiourea followed by base hydrolysis produced benzylthiols and regenerated the free hydroxyls in one step to form hydroxyl-footed benzylthiol **4** (48%). Interestingly, this

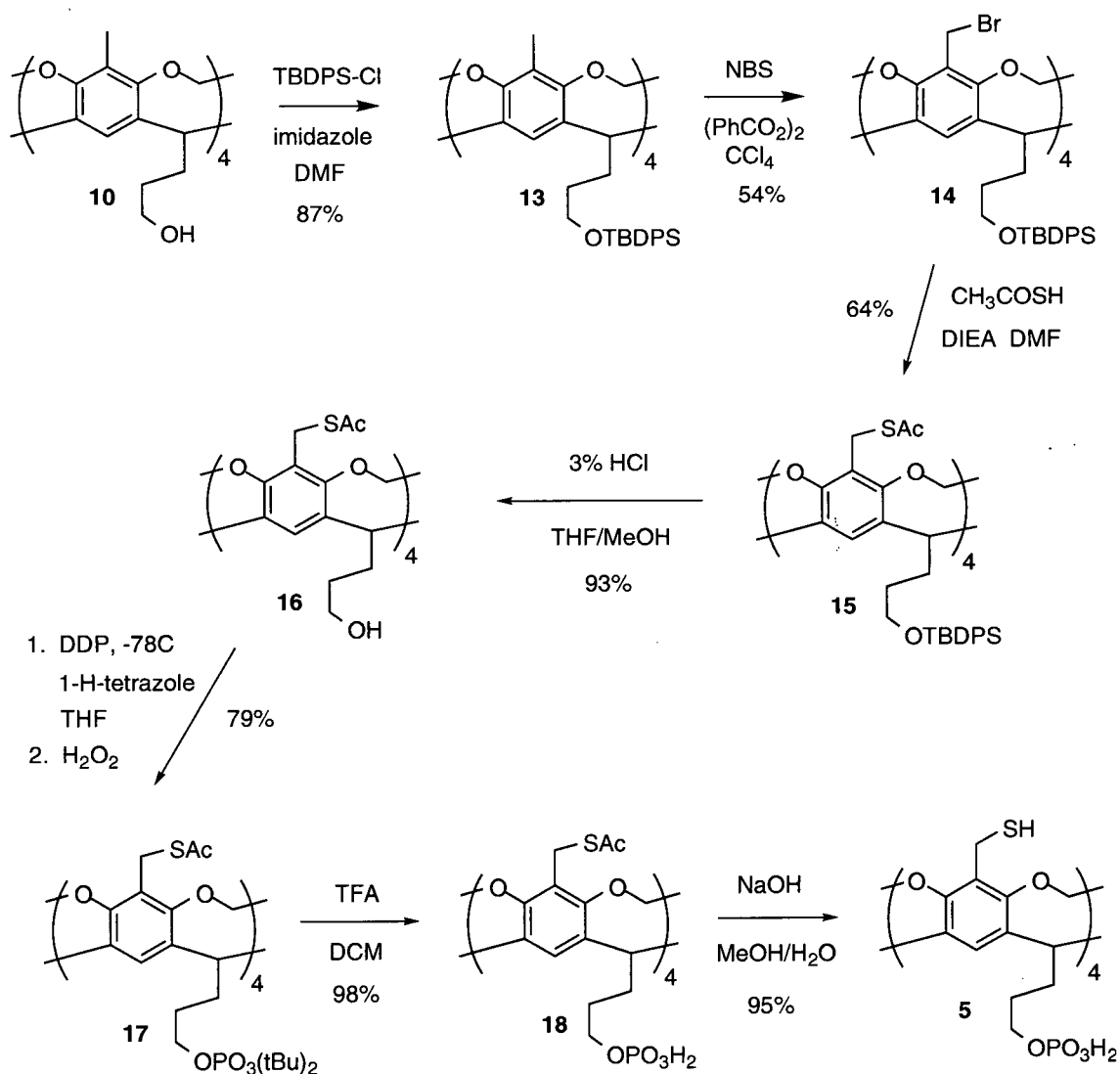
route to a benzylthiol cavitand was found to be more efficient than one previously developed by Cram who used additional benzylester and benzylalcohol intermediates.<sup>18</sup>

Although efficient in the synthesis of benzylthiol **4**, the use of acetyl protecting groups in the R<sub>1</sub> position does not allow for selectively addressable sites. Therefore, in the planning for the synthesis of the desired phosphate-footed tetrathiol cavitand **5**, it was evident that positions R<sub>1</sub> and R<sub>2</sub> ("feet" and "rims", respectively) should possess orthogonal reactivities. As base-sensitive benzylbromides were required for this method at the rim position R<sub>2</sub>, we chose to use base-sensitive protecting groups in rim position R<sub>2</sub> and acid-sensitive protecting groups for the feet (R<sub>1</sub>). As a starting point, the chosen acid-sensitive protecting group for the hydroxyl feet must be stable to NBS yet be removable without affecting the acetal bridges of the cavitand. Thus, cavitand **10** was treated with *t*-butyldiphenylsilyl chloride<sup>19</sup> (TBDPS-Cl) and imidazole to form TBDPS protected cavitand **13** (87%). Subsequent selective bromination with NBS and benzoyl peroxide afforded benzylbromide **14** (54%).

At this point in the synthesis of cavitand **5**, it was necessary to introduce a protected thiol onto TBDPS-protected benzylbromide **14**, such that removal of the TBDPS group could allow for selective phosphorylation. Such an *S*-protecting group must be stable to the acidic conditions required to remove the TBDPS group (HCl) as well as the conditions required for the phosphorylation outlined below. To this end, the *S*-acetyl protecting group was chosen and introduced by treatment of benzylbromide **14** with thioacetic acid and diisopropylethylamine (DIEA) to afford *S*-acetyl cavitand **15** (64%). Removal of the TBDPS groups with HCl afforded hydroxyl-footed *S*-acetyl cavitand **16** in 93% yield.



**Scheme 2.5.** Synthesis of Phosphate-Footed Tetrathiol **5**.



We phosphorylated the hydroxyl feet with di-*t*-butyl *N,N*-diethylphosphoramidite<sup>20</sup> (DDP) followed by oxidation with hydrogen peroxide. The use of DDP has four main advantages: DDP (1) eliminates the need to generate alkoxides in base sensitive compounds such as **16**, (2) provides acid-labile phosphate protecting groups consistent with our protection

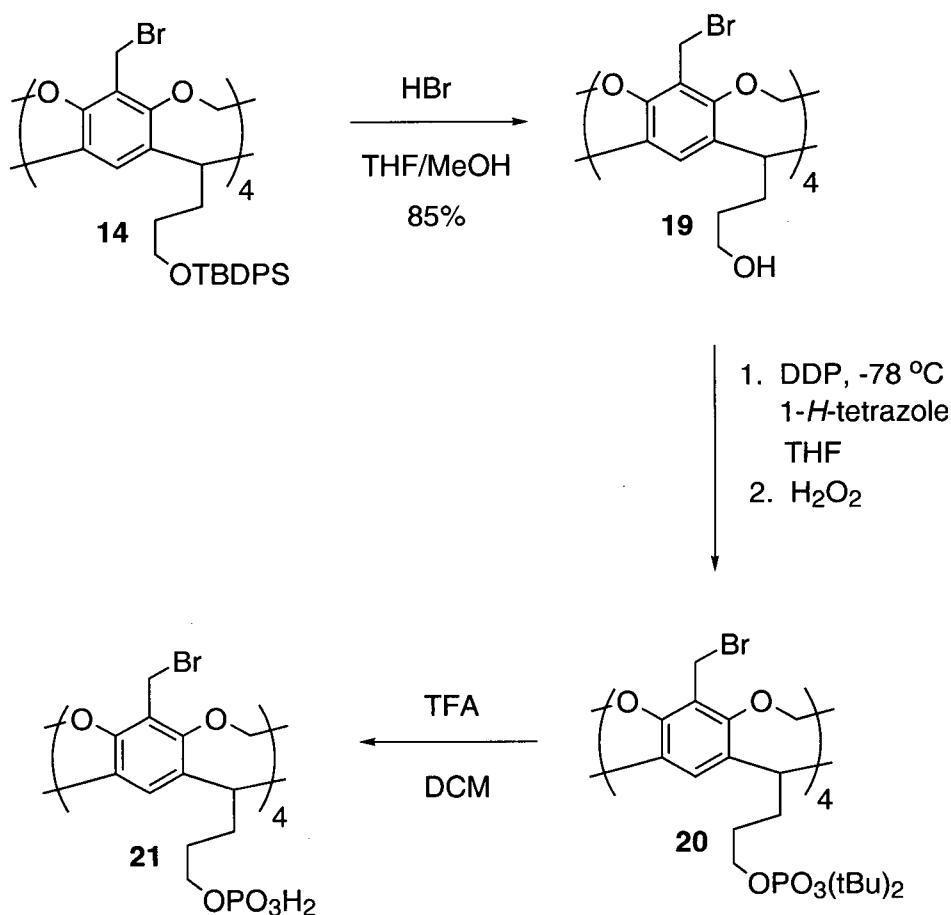
strategy, (3) generates phosphite esters in high yields, and (4) is quite stable, thus allowing for convenient storage. Phosphorylation of cavitand **16** with DDP followed by oxidation with hydrogen peroxide at  $-78\text{ }^{\circ}\text{C}$  afforded *t*-butyl-protected phosphate **17** (79%). Subsequent removal of the *t*-butyl groups with TFA followed by removal of the acetyl groups with NaOH afforded phosphate-footed tetrathiol **5** (93% for two steps). Cavitand **5** is a highly rigid, water soluble (pH 8) template, suitably functionalized with benzylthiol groups for further derivatization with peptide strands.

#### iv. Synthesis of Hydroxyl- and Phosphate-Footed Benzylbromides **19** and **21**

Over the course of the synthesis of cavitand **5**, we realized that it may be possible to isolate water-soluble cavitands bearing electrophilic benzylbromide groups at their rim positions. These compounds may be useful in the synthesis of de novo proteins via an alternative synthetic strategy: rather than nucleophiles at the rim position  $R_2$  reacting with electrophilic sites on the peptides, this “reverse” strategy would entail using electrophiles at the rim position  $R_2$  reacting with nucleophilic sites on the peptides.

We thus treated benzylbromide **14** with HBr and generated hydroxyl-footed benzylbromide cavitand **19** (85%, Scheme 2.6). HBr was found to eliminate the complication of a halogen exchange reaction that was observed with the use of HCl.

**Scheme 2.6.** Synthesis of Hydroxyl- and Phosphate-Footed Cavitands **19** and **21**.



Phosphorylation of cavitand **19** with DDP followed by oxidation with hydrogen peroxide at -78 °C afforded *t*-butyl-protected phosphate **20** (Scheme 2.6). Removal of the *t*-butyl groups with TFA afforded phosphate-footed benzylbromide **21**. Unfortunately, the phosphorylation did not proceed to completion and resulted in residual amounts of the corresponding tris-phosphate impurity, which were evident in the <sup>1</sup>H NMR and LSIMS mass spectra. Attempts to further purify the compound before and after the removal of the *t*-butyl groups were unsuccessful and

often led to decomposition of the reactive benzylbromide sites. We disregarded this setback and hoped that subsequent derivatization would facilitate purification. Indeed, in the pursuit of de novo proteins, it will be shown in Chapter Five that reaction of impure benzylbromide **21** with a suitable peptide gave a mixture of products that was separable by reversed-phase HPLC.

## C. Conclusion

We have developed synthetic pathways for a series of cavitands possessing either reactive thiol or benzylbromide functionalities at their rim positions suitable for further manipulation. As an initial starting point, tetrathiol cavitand **3** was synthesized with small hydrophobic methyl feet in the hopes of minimizing the hydrophobicity of the cavitand.

In efforts to incorporate hydrophilic groups into the cavitand while retaining suitable reactive rim functionalities, hydroxyl- and phosphate-footed cavitands **4**, **5**, **19**, and **21** were synthesized. The effect of the cavitand feet on the resulting de novo proteins will be presented in Chapter Five.

Interestingly, although not the focus of this research, this new family of cavitands may open the door to new areas of aqueous host-guest chemistry. The phosphate-footed cavitands are the first water-soluble cavitands that derive their water-solubility from their feet. This leaves the rim position near the enforced cavity available for further manipulation. For example, Cram and coworkers have shown that methyl-footed cavitands bind molecules such as CH<sub>3</sub>CN in a 1:1 stoichiometry in organic solvents.<sup>8a</sup> Moreover, our group has demonstrated binding of neutral guest molecules in a 2:1 cavitand:guest stoichiometry.<sup>21</sup> Such complexation

may indeed be far stronger in aqueous solution. Furthermore, another focus of research in our group is the synthesis of large molecular vessels consisting of a cyclic arrangement of cavitands.<sup>22</sup> The series of cavitands presented here may lead to large *water-soluble* molecular vessels which may, one day, lead to useful drug-delivery systems.

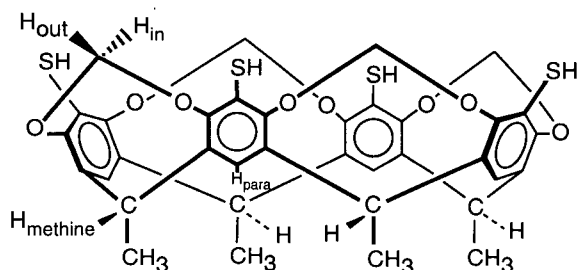
## D. Experimental

### i. General

All chemicals were reagent grade (Aldrich) except di-*tert*-butyl *N,N*-diethylphosphoramidite (Toronto Research Chemicals) which was technical grade and used without further purification. THF was distilled under N<sub>2</sub> from sodium benzophenone ketyl. DMF and *N,N*-dimethylacetamide (DMA) were dried over 4 Å molecular sieves. Desorption chemical ionization (DCI) and liquid secondary ionization mass spectra (LSIMS) were recorded on a Kratos Concept II HQ using various matrices as noted. Matrix-assisted laser desorption ionization (MALDI) mass spectra were recorded on a Bruker Reflex in reflectron mode using 2,5-dihydroxybenzoic acid (DHB) as the matrix. Melting points (uncorrected) were determined on a Thomas Hoover Unimelt capillary melting point apparatus. Microanalyses were performed by Mr. P. Borda of the UBC Microanalytical Laboratory on a Carlo-Erba CHN elemental analyzer, model 1106 or a Fisons CHN-O elemental analyzer, model 1108. All <sup>1</sup>H NMR spectra were recorded on a Bruker WH-400 spectrometer at ambient temperature using residual <sup>1</sup>H signals from deuterated solvents as a reference (CDCl<sub>3</sub>, 7.24; MeOD, 3.30; DMSO-*d*<sub>6</sub>, 2.49 ppm). <sup>31</sup>P NMR spectra were recorded on a Bruker AC-200 spectrometer referenced to external H<sub>3</sub>PO<sub>4</sub> (0.00 ppm). Silica gel (230-400 mesh, BDH) was used for column chromatography and silica gel glass-backed analytical plates (0.2 mm, Aldrich) were used for t.l.c. with UV detection. Size exclusion chromatography was performed using Sephadex LH-20. All products were dried overnight at rt and 0.1 Torr. Compounds **4** – **21** have been recently published.<sup>16</sup>

ii. Synthetic Procedures

**Methyl-Footed Tetrathiol 3.**<sup>12</sup>



A solution of cavitand **8**<sup>5</sup> (1.0 g, 1.1 mmol) in THF (400 mL) was cooled to -78 °C and *n*-BuLi (8.9 mL of a 1.6 M solution in hexanes, 14.3 mmol) was added. The reaction was stirred for 2 min at which point a room temperature solution of sulfur (0.49 g, 15.4 mmol) in THF (30 mL) was cannulated slowly into the reaction mixture over a period of 10 min. The cooling bath was then removed and the reaction was allowed to warm to rt over 1.5 h. Degassed 2 M HCl was added until acidic and then the reaction mixture was evaporated *in vacuo*. Water (50 mL) was added to the residue and then extracted with EtOAc (4 x 150 mL). The combined organic extracts were washed with brine (2 x 100 mL), dried (MgSO<sub>4</sub>), concentrated *in vacuo*. The residue was then dissolved in CHCl<sub>3</sub> (5 mL) and precipitated with hexanes. The precipitate was collected by filtration and recrystallized twice from CHCl<sub>3</sub>/hexanes to afford 0.42 g of tetrathiol **3** with ca. 10% of the corresponding trithiol. Removal of residual trithiol was achieved via acetylation (pyridine and acetic anhydride), column chromatography (3:2, hexanes:acetone),





precipitate was filtered, washed rigorously with H<sub>2</sub>O and dried *in vacuo* for 24 h. The solid was then suspended in THF, sonicated and filtered to afford dodecol **9** as an off-white solid (6.1 g, 78%); mp > 250°C.

**<sup>1</sup>H NMR (DMSO-d<sub>6</sub>, 400 MHz):** δ 8.62 (s, 8H, OH<sub>aromatic</sub>), 7.27 (s, 4H, <sub>para</sub>H), 4.19 (t, *J* = 7.8 Hz, 4H, H<sub>methine</sub>), 4.10 (bs, 4H, OH<sub>alkyl</sub>), 3.44 (t, *J* = 6.7 Hz, 8H, H<sub>c</sub>), 2.24 (m, 8H, H<sub>b</sub>), 1.94 (s, 12H, CH<sub>3</sub>), 1.35 (m, 8H, H<sub>a</sub>) ppm.

**MS (LSIMS<sup>+</sup>, thioglycerol + DMSO) *m/z* (rel intensity):** 776 ((M + H)<sup>+</sup>; 90).

**Anal.** Calcd for C<sub>44</sub>H<sub>56</sub>O<sub>12</sub>•H<sub>2</sub>O: C, 66.48; H, 7.35. Found: C, 66.71; H, 7.28.

## 2-Methyl Hydroxyl-Footed Cavitand **10**.

Dodecol **9** (20.0 g, 25.8 mmol) was dissolved in degassed DMA (100 mL) and added via a syringe pump over 48 h under N<sub>2</sub> at rt to a mixture of DMA (700 mL), bromochloromethane (7.50 mL, 113 mmol) and potassium carbonate (46.3 g, 335 mmol). After an additional 24 h at rt, bromochloromethane (7.50 mL, 113 mmol) was added and the reaction mixture heated to 45 °C. An additional aliquot of bromochloromethane (7.50 mL, 113 mL) was added the next day and the reaction mixture heated to 65 °C. After another 24 h at 65 °C, the reaction mixture was evaporated *in vacuo* followed by careful neutralization of the carbonate salts with 2 M HCl. The crude precipitate was filtered, washed with water until the filtrate was neutral, dissolved in THF, dried over MgSO<sub>4</sub> and evaporated. The solid was purified by column chromatography (CHCl<sub>3</sub>:MeOH, 9:1) to afford 2-methyl cavitand **10** as an off-white solid (8.84 g, 42%); mp > 250 °C;

**<sup>1</sup>H NMR (DMSO-d<sub>6</sub>, 400 MHz):** δ 7.43 (s, 4H, H<sub>e</sub>), 5.86 (d, *J* = 7.4 Hz, 4H, H<sub>i</sub>), 4.59 (t, *J* = 6.1 Hz, 4H, H<sub>e</sub>), 4.43 (t, *J* = 5.1 Hz, 4H, OH), 4.19 (d, *J* = 7.4 Hz, 4H, H<sub>h</sub>), 3.48 (m, 8H, H<sub>c</sub>), 2.35 (m, 8H, H<sub>b</sub>), 1.88 (s, 12H, CH<sub>3</sub>), 1.42 (m, 8H, H<sub>a</sub>) ppm.

**MS (LSIMS<sup>+</sup>, thioglycerol) *m/z* (rel intensity):** 825 ((M + H)<sup>+</sup>; 100).

**Anal.** Calcd for C<sub>48</sub>H<sub>56</sub>O<sub>12</sub>•H<sub>2</sub>O: C, 68.39; H, 6.93. Found: C, 68.07; H, 6.82.

### Acetylated 2-Methyl Cavitand 11.

2-Methyl bowl **10** (0.72 g, 0.87 mmol) was dissolved in pyridine (8 mL) and acetic anhydride (8 mL) and stirred overnight under N<sub>2</sub>. The solvent was removed in vacuo and purified by column chromatography (CHCl<sub>3</sub>:MeOH, 96:4) to afford acetylated 2-methyl cavitand **11** as a white solid (0.76 g, 87%): mp 131 °C (dec);

**<sup>1</sup>H NMR (CDCl<sub>3</sub>, 400 MHz):** δ 6.93 (s, 4H, <sub>para</sub>H), 5.86 (d, *J* = 6.9 Hz, 4H, H<sub>out</sub>), 4.82 (t, *J* = 9.1 Hz, 4H, H<sub>methine</sub>), 4.23 (d, *J* = 6.9 Hz, 4H, H<sub>in</sub>), 4.15 (t, *J* = 6.5 Hz, 8H, H<sub>c</sub>), 2.26 (m, 8H, H<sub>b</sub>), 2.03 (s, 12H, CH<sub>3</sub> acetyl), 1.95 (s, 12H, CH<sub>3</sub> aromatic), 1.69 (m, 8H, H<sub>a</sub>) ppm.

**MS (LSIMS<sup>+</sup>, thioglycerol) *m/z* (rel intensity):** 993 ((M + H)<sup>+</sup>; 100).

**Anal.** Calcd for C<sub>56</sub>H<sub>64</sub>O<sub>16</sub>: C, 67.73; H, 6.50. Found: C, 67.65; H, 6.30.

### Acetylated Benzylbromide Cavitand 12.

A solution of 2-methyl acetylated cavitand **11** (0.74 g, 0.75 mmol), NBS (0.58 g, 3.3 mmol) and benzoyl peroxide (10 mg) in  $\text{CCl}_4$  (50 mL) was refluxed for 16 h. The reaction mixture was then cooled to rt, filtered, and evaporated *in vacuo*. The crude product was purified by column chromatography (EtOAc:hexanes, 1:1) to afford benzyl bromide **12** as a white solid (0.66 g, 68%): mp 136 °C (dec);

**$^1\text{H}$  NMR ( $\text{CDCl}_3$ , 400 MHz)**  $\delta$  7.09 (s, 4H,  $\text{H}_{\text{para}}$ ), 6.01 (d,  $J = 6.7$  Hz, 4H,  $\text{H}_{\text{out}}$ ), 4.82 (t,  $J = 8.1$  Hz,  $\text{H}_{\text{methine}}$ ), 4.55 (d,  $J = 6.7$ , 4H,  $\text{H}_{\text{in}}$ ), 4.39 (s, 8H,  $\text{CH}_2\text{Br}$ ), 4.15 (t,  $J = 6.5$  Hz, 8H,  $\text{H}_c$ ), 2.28 (m, 8H,  $\text{H}_b$ ), 2.05 (s, 12H,  $\text{CH}_3$ ), 1.67 (m, 8H,  $\text{H}_a$ ) ppm.

**MS (LSIMS<sup>+</sup>, thioglycerol)  $m/z$  (rel intensity):** 1229 ( $(\text{M} - \text{Br} + \text{H})^+$ ; 100), 1309 ( $(\text{M} + \text{H})^+$ , 80).

**Anal.** Calcd for  $\text{C}_{56}\text{H}_{60}\text{O}_{16}\text{Br}_4$ : C, 51.40; H, 4.62; Found: C, 51.60; H, 4.42.

### Hydroxyl-Footed Tetrathiol Cavitand 4.

Thiourea (5.1 mg, 0.072 mmol) was added to a solution of bromocavitand **12** (20 mg, 0.015 mmol) in degassed DMF (2 mL) and stirred for 2 h. The reaction mixture was poured onto degassed 2M NaOH (2 mL) and stirred for 1 h. The reaction mixture was evaporated *in vacuo*, dissolved in water and acidified with 5% acetic acid. The precipitate was filtered, thoroughly washed with water, dissolved in  $\text{CHCl}_3$ :MeOH (9:1), dried over  $\text{MgSO}_4$  and evaporated. The

crude product was purified by column chromatography (9:1, CHCl<sub>3</sub>:MeOH) to afford tetrathiol **4** as a white solid (7.0 mg, 48%): mp > 250 °C;

**<sup>1</sup>H NMR (DMSO-d<sub>6</sub>, 400 MHz):** δ 7.56 (s, 4H, <sub>para</sub>H), 5.95 (d, *J* = 7.5 Hz, 4H, H<sub>out</sub>), 4.59 (t, *J* = 7.9 Hz, 4H, H<sub>methine</sub>), 4.43 (m, 8H, OH, H<sub>in</sub>), 3.47 (m, 16H, CH<sub>2</sub>SH, H<sub>c</sub>), 2.71 (t, *J* = 7.8 Hz, 4H, SH), 2.38 (m, 8H, H<sub>b</sub>), 1.42 (m, 8H, H<sub>a</sub>) ppm.

**MS (LSIMS<sup>+</sup>, 3-NBA) *m/z* (rel intensity):** 917 (M - SH - H)<sup>+</sup>, 100), 951 ((M - H)<sup>+</sup>, 50).

**Anal.** Calcd for C<sub>48</sub>H<sub>56</sub>O<sub>12</sub>S<sub>4</sub>•H<sub>2</sub>O: C, 59.36; H, 6.02. Found: C, 59.43; H, 5.73.

### TBDPS-protected 2-Methyl Cavitand **13**.

TBDPS-Cl (6.06 mL, 23.3 mmol) was added to a solution of cavitand **10** (2.40 g, 2.91 mmol) and imidazole (3.16 g, 46.6 mmol) in DMF (20 mL) and stirred overnight under N<sub>2</sub> at rt. The DMF was then evaporated *in vacuo*, CHCl<sub>3</sub> (20 mL) was added and the organic layer washed with water (3 x 10 mL), dried over MgSO<sub>4</sub> and evaporated. The crude product was purified by column chromatography (9:1, hexanes:EtOAc) to afford cavitand **13** as a white foam (4.6 g, 87%):

**<sup>1</sup>H NMR (CDCl<sub>3</sub>, 400 MHz):** δ 7.60 (m, 16H, TBDPS aromatic H), 7.28 (m, 24H, TBDPS aromatic H), 6.91 (s, 4H, <sub>para</sub>H), 5.85 (d, *J* = 6.9 Hz, 4H, H<sub>out</sub>), 4.77 (t, *J* = 8.2 Hz, 4H, H<sub>methine</sub>), 4.23 (d, *J* = 6.9 Hz, 4H, H<sub>in</sub>), 3.66 (t, *J* = 6.4 Hz, 8H, H<sub>c</sub>), 2.22 (m, 8H, H<sub>b</sub>), 1.95 (s, 12H, CH<sub>3</sub>), 1.59 (m, 8H, H<sub>c</sub>), 0.99 (s, 36H, TBDPS alkyl H) ppm.

**MS (LSIMS<sup>-</sup>, 3-NBA + thioglycerol + CHCl<sub>3</sub>) *m/z* (rel intensity):** 1824 ((M - H)<sup>-</sup>; 40), 1931 ((M•thioglycerol - H)<sup>-</sup>; 100).

**Anal.** Calcd for C<sub>112</sub>H<sub>128</sub>O<sub>12</sub>Si<sub>4</sub>: C, 75.64; H, 7.25. Found: C, 75.49; H, 7.12.

#### **TBDPS-protected Benzylbromide 14.**

NBS (2.93 g, 16.5 mmol) and benzoyl peroxide (0.10 g) were added to a solution of cavitand **13** (6.83 g, 3.74 mmol) in CCl<sub>4</sub>. The solution was refluxed for 18 h under N<sub>2</sub> at which point the reaction mixture was cooled to rt and filtered. The filtrate was evaporated *in vacuo* and the crude residue purified by column chromatography (hexanes:EtOAc, 95:5; followed by 92:8) to afford benzyl bromide **14** as a white foam (2.41 g, 54%):

**<sup>1</sup>H NMR (CDCl<sub>3</sub>, 400 MHz):** δ 7.59 (m, 16H, TBDPS aromatic H), 7.30 (m, 24H, TBDPS aromatic H), 7.06 (s, 4H, <sub>para</sub>H), 6.00 (d, *J* = 6.2 Hz, 4H, H<sub>out</sub>), 4.78 (t, *J* = 7.9 Hz, 4H, H<sub>methine</sub>), 4.54 (d, *J* = 6.2 Hz, 4H, H<sub>in</sub>), 4.39 (s, 8H, CH<sub>2</sub>Br), 3.65 (t, *J* = 6.3 Hz, 8H, H<sub>c</sub>), 2.22 (m, 8H, H<sub>b</sub>), 1.56 (m, 8H, H<sub>a</sub>), 0.99 (s, 36H, TBDPS alkyl H) ppm.

**MS (LSIMS<sup>-</sup>, 3-NBA + CHCl<sub>3</sub>) *m/z* (rel intensity):** 2093 ((M - H)<sup>-</sup>; 100).

**Anal.** Calcd for C<sub>112</sub>H<sub>124</sub>O<sub>12</sub>Si<sub>4</sub>Br<sub>4</sub>: C, 64.24; H, 5.97. Found: C, 64.64; H, 5.96.

### ***S*-Acetyl TBDPS Cavitand 15.**

Thioacetic acid (219  $\mu$ L, 3.08 mmol) was added to a solution of benzylbromide **14** (1.50 g, 0.701 mmol) and diisopropylethylamine (535  $\mu$ L, 3.08 mmol) in DMF. The reaction mixture was stirred for 16 h, evaporated *in vacuo* and purified by column chromatography (hexanes:EtOAc, 3:1) to afford cavitand **15** as a white foam (930 mg, 64%):

**$^1\text{H}$  NMR (CDCl<sub>3</sub>, 400 MHz):**  $\delta$  7.59 (m, 16H, TBDPS aromatic H), 7.30 (m, 24H, TBDPS aromatic H), 6.96 (s, 4H, *para*H), 5.88 (d,  $J$  = 7.2 Hz, 4H,  $H_{\text{out}}$ ), 4.73 (t,  $J$  = 8.1 Hz, 4H,  $H_{\text{methine}}$ ), 4.26 (d,  $J$  = 7.2 Hz, 4H,  $H_{\text{in}}$ ), 4.01 (s, 8H, CH<sub>2</sub>S), 3.64 (t,  $J$  = 6.3 Hz, 8H,  $H_{\text{c}}$ ), 2.31 (s, 12H, COCH<sub>3</sub>), 2.19 (m, 8H,  $H_{\text{b}}$ ), 1.55 (m, 8H,  $H_{\text{a}}$ ), 0.99 (s, 36H, TBDPS alkyl H) ppm.

**MS (LSIMS<sup>+</sup>, 3-NBA)  $m/z$  (rel intensity):** 2226 ((M•3-NBA)<sup>+</sup>; 100).

**Anal.** Calcd for C<sub>120</sub>H<sub>136</sub>O<sub>16</sub>Si<sub>4</sub>S<sub>4</sub>: C, 69.46; H, 6.61. Found: C, 69.60; H, 6.51.

### ***S*-Acetyl Hydroxyl-Footed Cavitand 16.**

Concentrated HCl (150  $\mu$ L, 1.82 mmol) was added to a solution of *S*-Ac TBDPS cavitand **15** (100 mg, 0.0472 mmol) in THF:MeOH (4:1, 5 mL). The reaction mixture was stirred for 4 h, neutralized with saturated NaHCO<sub>3</sub> and evaporated *in vacuo*. The crude product was dissolved in CHCl<sub>3</sub>:MeOH (9:1, 10 mL), dried over MgSO<sub>4</sub>, evaporated *in vacuo* and purified by column

chromatography (9:1, CHCl<sub>3</sub>:MeOH) to afford cavitand **16** a white solid (49 mg, 93%): mp > 250 °C;

**<sup>1</sup>H NMR (DMSO-d<sub>6</sub>, 400 MHz):** δ 7.58 (s, 4H, *para*H), 5.79 (d, *J* = 7.7 Hz, 4H, H<sub>out</sub>), 4.55 (t, *J* = 8.1 Hz, 4H, H<sub>methine</sub>), 4.44 (br s, 4H, OH), 4.27 (d, *J* = 7.7 Hz, 4H, H<sub>in</sub>), 3.94 (s, 8H, CH<sub>2</sub>S), 3.47 (t, *J* = 6.5 Hz, 8H, H<sub>c</sub>), 2.37 (m, 8H, H<sub>b</sub>), 2.29 (s, 12H, COCH<sub>3</sub>), 1.40 (s, 8H, H<sub>a</sub>) ppm.

**MS (LSIMS<sup>+</sup>, 3-NBA + CHCl<sub>3</sub>/MeOH) *m/z* (rel intensity):** 1273 ((M+3-NBA)<sup>+</sup>; 100).

**Anal.** Calcd for C<sub>120</sub>H<sub>136</sub>O<sub>16</sub>S<sub>4</sub>•H<sub>2</sub>O: C, 59.03; H, 5.84. Found: C, 59.42; H, 5.72.

#### ***t*-Butyl Phosphorylated S-Acetyl Cavitand 17.**

1-*H*-Tetrazole (0.19 g, 2.7 mmol) was added to a THF solution of hydroxyl-footed S-Ac cavitand **16** (0.10 g, 0.089 mmol) and di-*tert*-butyl *N,N*-diethylphosphoramidite (0.25 mL, 0.89 mmol) in THF and stirred for 10 min under N<sub>2</sub>. The reaction mixture was then cooled to -78 °C at which point H<sub>2</sub>O<sub>2</sub> (0.12 mL, 1.1 mmol) was added and the reaction mixture allowed to warm to rt over 30 min. The reaction mixture was then poured onto H<sub>2</sub>O and extracted three times with CHCl<sub>3</sub>. The combined extracts were then washed with 10% NaHSO<sub>3</sub>, H<sub>2</sub>O, dried over MgSO<sub>4</sub> and evaporated. The residue was purified by size exclusion chromatography (EtOAc:MeOH:H<sub>2</sub>O; 40:10:4) followed by column chromatography (CHCl<sub>3</sub>:MeOH; 96:4) to afford *t*-butyl protected S-Ac phosphate **17** as a white solid (0.13 g, 79%): mp 100 °C (dec);

**<sup>1</sup>H NMR (CDCl<sub>3</sub>, 400 MHz)** δ 7.04 (s, 4H<sub>para</sub>H), 5.82 (d, *J* = 7.3 Hz, 4H, H<sub>out</sub>), 4.76 (t, *J* = 8.1 Hz, 4H, H<sub>methine</sub>), 4.25 (d, *J* = 7.3 Hz, 4H, H<sub>in</sub>), 4.00 (m, 16H, ArCH<sub>2</sub>S, H<sub>c</sub>), 2.29 (bm, 20H, H<sub>b</sub>, SCOCH<sub>3</sub>), 1.65 (m, 8H, H<sub>a</sub>), 1.44 (s, 72H, C(CH<sub>3</sub>)<sub>3</sub>) ppm.

**<sup>31</sup>P NMR (CDCl<sub>3</sub>, 81 MHz):** δ -9.96 (s, 4P) ppm.

**MS (LSIMS<sup>+</sup>, 3-NBA + CHCl<sub>3</sub>) *m/z* (rel intensity):** 1441 (M - (C(CH<sub>3</sub>)<sub>3</sub>)<sub>8</sub> + 9H)<sup>+</sup>, 100).

**Anal.** Calc for C<sub>88</sub>H<sub>132</sub>O<sub>28</sub>P<sub>4</sub>S<sub>4</sub>: C, 55.92; H, 7.04. Found: C, 56.10; H, 7.00.

### **S-Acetyl Phosphate-Footed Cavitand 18.**

TFA (0.40 mL, 5.2 mmol) was added to a solution of cavitand **17** (46 mg, 0.024 mmol) in DCM and stirred for 10 min. The reaction mixture was evaporated *in vacuo* to afford cavitand **18** as a white solid (34 mg, 98%): mp 100 °C (dec);

**<sup>1</sup>H NMR (methanol-d<sub>4</sub>, 400 MHz):** δ 7.35 (s, 4H, <sub>para</sub>H), 5.82 (d, *J* = 7.3 Hz, 4H, H<sub>out</sub>), 4.75 (t, *J* = 8.0 Hz, 4H, H<sub>methine</sub>), 4.32 (d, *J* = 7.3 Hz, 4H, H<sub>in</sub>), 4.10 (m, 16H, H<sub>c</sub>, CH<sub>2</sub>SCO), 2.44 (m, 8H, H<sub>b</sub>), 2.05 (br s, 12H, SCOCH<sub>3</sub>), 1.60 (m, 8H, H<sub>a</sub>) ppm.

**<sup>31</sup>P NMR (methanol-d<sub>4</sub>, 81 MHz):** δ 0.00 (s, 4P) ppm.

**MS (LSIMS<sup>-</sup>, 3-NBA + methanol) *m/z* (rel intensity):** 1439 ((M - H)<sup>-</sup>, 100).

**Anal.** Calcd for C<sub>56</sub>H<sub>68</sub>O<sub>28</sub>P<sub>4</sub>S<sub>4</sub>•3H<sub>2</sub>O: C, 44.98; H, 4.99. Found: C, 45.06; H, 4.72.



### Phosphate-Footed Benzylthiol Cavitand 5.

A 0.190 M degassed solution of NaOH (1.39 mL, 0.264 mmol) was added to a degassed suspension of *S*-acetylated cavitand **18** (19 mg, 0.0132 mmol) in MeOH and H<sub>2</sub>O (1:1, 7 mL). The reaction mixture was stirred for 30 min under N<sub>2</sub> at which point pre-washed Amberlite ion-exchange resin (H<sup>+</sup>) was added via a sidearm until the solution was no longer basic. The mixture was stirred for 2 min then cannulated to another flask so as to leave the resin behind. The solution was evaporated *in vacuo* to afford tetrathiol **5** as a white solid (16 mg, 95%): mp 100 °C (dec);

**<sup>1</sup>H NMR (methanol-d<sub>4</sub>, 400 MHz):** δ 7.32 (s, 4H, <sub>para</sub>H), 5.92 (d, *J* = 7.3 Hz, 4H, H<sub>out</sub>), 4.77 (t, *J* = 8.3 Hz, 4H, H<sub>methine</sub>), 4.60 (d, *J* = 7.3 Hz, 4H, H<sub>in</sub>), 4.03 (dt, *J* = 5.8 Hz, *J* = 5.8 Hz, 8H, H<sub>c</sub>), 3.61 (s, 8H, CH<sub>2</sub>S), 2.45 (m, 8H, H<sub>b</sub>), 1.64 (m, 8H, H<sub>a</sub>) ppm.

**<sup>31</sup>P NMR (methanol-d<sub>4</sub>, 81 MHz):** δ 3.24 (s, 4P) ppm.

**MS (LSIMS<sup>+</sup>, thioglycerol + methanol) *m/z* (rel intensity):** 1271 (M - H)<sup>+</sup>, 100).

**Anal.** Calcd for C<sub>48</sub>H<sub>60</sub>O<sub>24</sub>P<sub>4</sub>S<sub>4</sub>•4H<sub>2</sub>O: C, 42.86; H, 5.10. Found: C, 42.53; H, 4.73.

### Hydroxyl-Footed Benzylbromide 19.

48% HBr (420 μL, 2.49 mmol) was added to a solution of tetrabromocavitand **14** (80 mg, 0.038 mmol) in THF:MeOH (4:1, 4 mL) and stirred for 4 h. The reaction mixture was then evaporated

*in vacuo* and purified by column chromatography (9:1, CHCl<sub>3</sub>:MeOH) to afford hydroxyl-footed cavitand **19** as a white solid (37 mg, 85%): mp > 250 °C;

**<sup>1</sup>H NMR (acetone-d<sub>6</sub>, 400 MHz):** δ 8.02 (s, 4H, <sub>para</sub>H), 6.03 (d, *J* = 7.5 Hz, 4H, H<sub>out</sub>), 4.73 (t, *J* = 8.2 Hz, 4H, H<sub>methine</sub>), 4.55 (s, 8H, CH<sub>2</sub>Br), 4.49 (d, *J* = 7.5 Hz, 4H, H<sub>in</sub>), 3.96 (t, *J* = 5.9 Hz, 4H, OH), 3.64 (dt, *J* = 5.9 Hz, *J* = 5.9 Hz, 8H, H<sub>c</sub>), 2.70 (m, 8H, H<sub>b</sub>), 1.51 (m, 8H, H<sub>a</sub>) ppm.

**MS (LSIMS<sup>+</sup>, thioglycerol + CHCl<sub>3</sub>) *m/z* (rel intensity):** 1140 ((M)<sup>+</sup>; 60), 1219 ((M•Br)<sup>+</sup>; 100).

**Anal.** Calcd for C<sub>48</sub>H<sub>52</sub>O<sub>12</sub>Br<sub>4</sub>•H<sub>2</sub>O: C, 49.91; H, 4.72. Found: C, 50.17; H, 4.62.

#### ***t*-Butyl Phosphorylated Benzylbromide **20**.**

1-*H*-Tetrazole (46 mg, 0.66 mmol) was added to a THF (10 mL) solution of hydroxyl-footed benzylbromide cavitand **19** (25 mg, 0.022 mmol) and di-*tert*-butyl *N,N*-diethylphosphoramidite (61 μL, 0.22 mmol) and stirred for 15 min at rt. The reaction mixture was then cooled to -78 °C, H<sub>2</sub>O<sub>2</sub> (30 μL, 0.26 mmol, 12 equiv) was added and the reaction mixture allowed to warm to rt over 30 min. The reaction mixture was poured onto H<sub>2</sub>O, extracted three times with CHCl<sub>3</sub>, dried over MgSO<sub>4</sub> and evaporated. The crude product was purified with size exclusion chromatography (EtOAc:MeOH:H<sub>2</sub>O, 40:10:4) followed by column chromatography (CHCl<sub>3</sub>:MeOH, 96:4) to afford a white solid which was a mixture of tetraphosphate **20** and ca. 15% of the corresponding tris-phosphate derivative; tetraphosphate **20** was not purified further.

**$^1\text{H}$  NMR ( $\text{CDCl}_3$ , 400 MHz):**  $\delta$  7.15 (bs, 4H,  $\text{paraH}$ ), 5.99 (d,  $J = 6.5$  Hz, 4H,  $\text{H}_{\text{out}}$ ), 4.82 (t,  $J = 7.5$  Hz, 4H,  $\text{H}_{\text{methine}}$ ), 4.54 (d,  $J = 6.5$  Hz, 4H,  $\text{H}_{\text{in}}$ ), 4.38 (s, 8H,  $\text{CH}_2\text{Br}$ ), 4.00 (m, 8H,  $\text{H}_c$ ), 2.35 (bm, 8H,  $\text{H}_b$ ), 1.67 (bm, 8H,  $\text{H}_a$ ), 1.45 (s, 72H,  $\text{C}(\text{CH}_3)_3$ ) ppm.

**$^{31}\text{P}$  NMR ( $\text{CDCl}_3$ , 81 MHz):**  $\delta$  -9.97 (s, 4P) ppm.

**MS (LSIMS $^+$ , 3-NBA)  $m/z$  (rel intensity):** 1461 ((M - 8( $\text{C}(\text{CH}_3)_3$ ) + 9H) $^+$ ; 100), 1798 (M - 2( $\text{C}(\text{CH}_3)_3$ ) + 9H) $^+$ ; 20), 1855 (M -  $\text{C}(\text{CH}_3)_3$  + 9H) $^+$ ; 10).

**HRMS (LSIMS $^+$ , 3-NBA):** Calcd for (M - 8( $\text{C}(\text{CH}_3)_3$ ) + 9H): 1460.8910; Found: 1460.8924.

#### Phosphate-Footed Benzylbromide 21.

TFA (0.32 mL, 4.2 mmol) was added to a  $\text{CH}_2\text{Cl}_2$  (4 mL) solution of impure tetraphosphate **20** (25 mg of the mixture described above, ca. 0.01 mmol) and stirred for 10 min. The reaction mixture was evaporated *in vacuo* to afford a white solid, which was a mixture of tetraphosphate-footed benzylbromide **21** and ca. 15% of the corresponding tris-phosphate derivative (19 mg); tetraphosphate **21** was not purified further.

**$^1\text{H}$  NMR ( $\text{CDCl}_3$  + 20% acetone- $\text{d}_6$ , 400 MHz):**  $\delta$  7.36 (bs, 4H,  $\text{paraH}$ ), 5.95 (d,  $J = 7.3$  Hz, 4H,  $\text{H}_{\text{out}}$ ), 4.69 (bt,  $J = 7.6$  Hz, 4H,  $\text{H}_{\text{methine}}$ ), 4.43 (m, 12H,  $\text{CH}_2\text{Br}$ ,  $\text{H}_{\text{in}}$ ), 4.15 (bm, 8H,  $\text{H}_c$ ), 2.42 (bm, 8H,  $\text{H}_b$ ), 1.59 (bm, 8H,  $\text{H}_a$ ) ppm.

**$^{31}\text{P}$  NMR ( $\text{CDCl}_3$  + 20% acetone- $\text{d}_6$ , 81 MHz):**  $\delta$  -0.53 (s, 4P) ppm.

**MS (LSIMS $^-$ , 3-NBA +  $\text{CHCl}_3$  + MeOH)  $m/z$  (rel intensity):** 1379 ((M -  $\text{PO}_3\text{H}_2$  - H) $^-$ ; 60), 1459 ((M - H) $^-$ ; 100).

**HRMS (LSIMS<sup>+</sup>, 3-NBA + CHCl<sub>3</sub> + MeOH):** Calcd for (M - H)<sup>+</sup>: 1458.8754; Found:  
1458.8766.

## E. References

1. (a) Baeyer, A. *Ber. Dtsch. Chem. Ges.* **1872**, 5, 25; (b) Baeyer, A. *ibid.* **1872**, 5, 280.  
These two references were taken from a recent review of resorcinarenes: (c) Timmerman, P.; Verboom, W.; Reinhoudt, D. N. *Tetrahedron* **1996**, 52, 2663-2704.
2. Niederl, J. B.; Vogel, H. J. *J. Am. Chem. Soc.* **1940**, 62, 2512.
3. Erdtman, H.; Högberg, S.; Abrahamsson, S.; Nilsson, B. *Tetrahedron Lett.* **1968**, 1679.
4. Abis, L.; Dalcanale, E.; Du Vosel, A.; Spera, S. *J. Org. Chem.* **1988**, 53, 5475.
5. Moran, J. R.; Karbach, S.; Cram, D. J. *J. Am. Chem. Soc.* **1982**, 104, 5826-5828.
6. Cram, D. J.; Cram, J. M. *Container Molecules and Their Guests*; Royal Society of Chemistry: Cambridge, 1994.
7. (a) Sherman, J. C. *Tetrahedron* **1995**, 51, 3395-3422; (b) Sherman, J. C. In *Large Ring Molecules*; Semlyen, J. A., Ed.; J. Wiley & Sons: West Sussex, England, 1996; (c) Chapman, R. G.; Sherman, J. C. *J. Org. Chem.* **1998**, 63, 4103-4110.
8. (a) Tucker, J. A.; Knobler, C. B.; Trueblood, K. N.; Cram, D. J. *J. Am. Chem. Soc.* **1989**, 111, 3688-3699; (b) Rudkevich, D. M.; Hilmersson, G.; Rebek, J. *J. Am. Chem. Soc.* **1997**, 119, 9911-9912; (c) Careri, M.; Dalcanale, E.; Mangia, A.; Ruffini, M. *Anal. Comm.* **1997**, 34, 13-15; (d) Vincetti, M.; Dalcanale, E. *J. Chem. Soc. Perkin Trans. 2* **1995**, 1069-1076; (e) Boerrigter, H.; Verboom, W.; Reinhoudt, D. N. *J. Org. Chem.* **1997**, 62, 7148-7155; (f) Dalcanale, E.; Costantini, G.; Soncini, P. *J. Inclusion Phenom.* **1992**, 13, 87-92.
9. (a) Schierbaum, K. D.; Weiss, T.; Thoden van Velzen, E. U.; Engbersen, J. F. J.; Reinhoudt, D. N.; Göpel, W. *Science* **1994**, 265, 1413-1415. (b) Thoden van Velzen, E. U.; Engbersen, J. F. J.; de Lange, P. J.; Mahy, J. W. G.; Reinhoudt, D. N. *J. Am. Chem. Soc.* **1995**, 117, 6853-6862.
10. Sherman, J. C.; Knobler, C. B.; Cram, D. J. *J. Am. Chem. Soc.* **1991**, 113, 2194-2204.
11. Fraser, J. R.; Borecka, B.; Trotter, J.; Sherman, J. C. *J. Org. Chem.* **1995**, 60, 1207-1213.
12. Gibb, B. C.; Mezo, A. R.; Causton, A. S.; Fraser, J. R.; Tsai, F. C. S.; Sherman, J. C. *Tetrahedron* **1995**, 51, 8719-8732.

13. A synthesis of tetrathiol **3** with pentyl feet ( $R_1 = (CH_2)_4CH_3$ ) was reported in the same year: Helgeson, R. C.; Knobler, C. B.; Cram, D. J. *J. Chem. Soc. Chem. Commun.* **1995**, 307-308.
14. Tunstad, L. M.; Tucker, J. A.; Dalcanale, E.; Weiser, J.; Bryant, J. A.; Sherman, J. C.; Helgeson, R. C.; Knobler, C. B.; Cram, D. J. *J. Org. Chem.* **1989**, *54*, 1305-1312.
15. Gibb, B. C.; Chapman, R. G.; Sherman, J. C. *J. Org. Chem.* **1996**, *61*, 1505-1509.
16. Mezo, A. R.; Sherman, J. C. *J. Org. Chem.* **1998**, *63*, 6824-6829.
17. Sorrell, T. N.; Pigge, F. C. *J. Org. Chem.* **1993**, *58*, 784-785.
18. Cram, D. J.; Karbach, S.; Kim, Y. H.; Baczynskyj, L.; Marti, K.; Sampson, R. M.; Kallemeyn, G. W. *J. Am. Chem. Soc.* **1988**, *110*, 2554-2560.
19. Hanessian, S.; Lavalley, P. *Can. J. Chem.* **1975**, *53*, 2975-2977.
20. Perich, J. W.; Johns, R. B. *Synth. Commun.* **1988**, 142-144.
21. (a) Chapman, R. G.; Sherman, J. C. *J. Am. Chem. Soc.* **1995**, *117*, 9081-9082; (b) Chapman, R. G.; Olovsson, G.; Trotter, J.; Sherman, J. C. *J. Am. Chem. Soc.* **1998**, *120*, 6252-6260; (c) Chapman, R. G.; Sherman, J. C. *J. Am. Chem. Soc.* **1998**, *120*, 9818-9826.
22. Chopra, N. C.; Sherman, J. C. *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 1727-1729.

## Chapter Three: Model Syntheses for De Novo Proteins

### A. Introduction

#### i. Synthesis Using Protected Peptides

One major consideration in making new covalent bonds for template-assembled de novo protein synthesis is the large number of possible side reactions that may occur due to different reactive sites on amino acid side-chains. As a result, early attempts at linking peptide segments to various templates have logically used side-chain protected peptides. Intuitively, by reducing the number of reactive sites on a peptide segment, more site-selective reactions may take place. Early work in de novo protein synthesis removed side chain protecting groups only in the final stage of their synthesis.

For example, Sasaki and Kaiser in their porphyrin-based four-helix bundle first linked their protected peptides to the template via amide bonds.<sup>1</sup> Subsequent deprotection of the peptides afforded the porphyrin-based de novo protein. In another example, Mutter and coworkers used stepwise solid-phase peptide synthesis (SPPS) to couple their protected peptides with an amide bond to their lysine-rich peptide template.<sup>2</sup> Again, subsequent deprotection afforded the desired de novo protein.

Unfortunately, covalent peptide assembly using protected peptides has some fundamental limitations. Firstly, protected peptides are notoriously insoluble and thus synthesizing de novo proteins of appreciable size (> 50 residues) using protected peptides may be limited. Secondly, purification of elaborate peptides/proteins generated by protected peptide

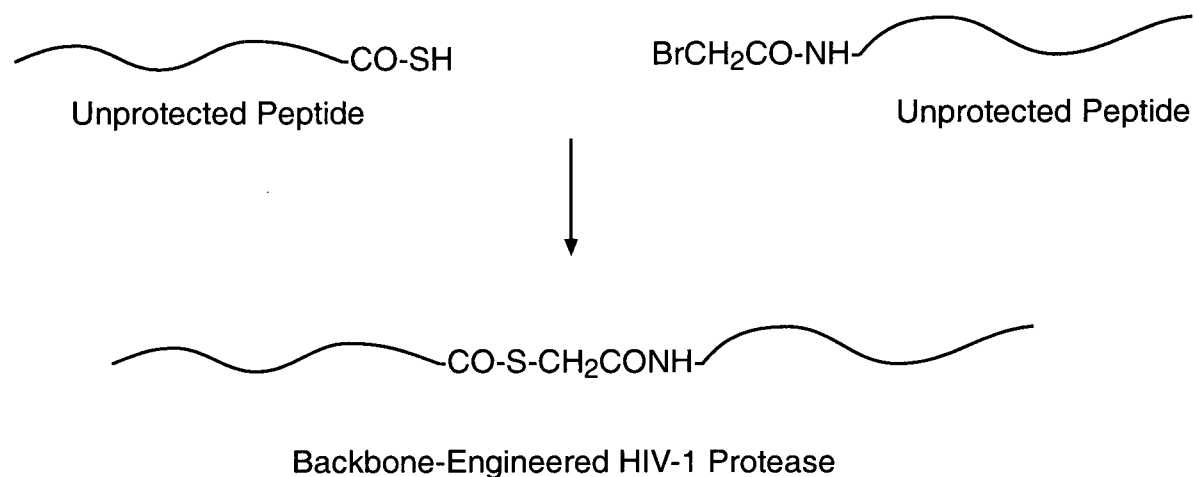
assembly has been found to be very difficult.<sup>2</sup> In addition, coupling of protected peptide segments have been found to be, in some cases, quite slow.<sup>3</sup> Hence, in the pursuit of a general method for the chemical synthesis of proteins, various methods were developed to link *unprotected* peptide segments to other unprotected peptides or templates.<sup>4</sup>

## ii. Chemoselective Ligation

A key requirement for the ligation of unprotected peptides is *chemoselectively*. As mentioned above, many side-chains possess reactive functional groups and thus an effective synthetic strategy must incorporate two functional groups exclusively tailored to react with one another. One of the first examples of such a strategy appeared in 1992 by Schnölzer and Kent.<sup>5</sup> They coupled two unprotected ~ 50-residue peptides to give a backbone-engineered analogue of the 99-residue protein HIV-1 protease. This was accomplished by incorporating a nucleophilic thioacid group at the C-terminus of one segment and a bromoacetyl group at the N-terminus of the other (Scheme 3.1).



**Scheme 3.1.** Illustration of the Chemoselective Synthetic Methodology Employed by Schnölzer and Kent.<sup>5</sup>



These two functional groups were found to react chemoselectively to form a single thioester-linked product. This methodology was later used in the synthesis of a de novo four-helix bundle based on Mutter's peptide template,<sup>6</sup> as well as other large proteins<sup>7</sup> and was given the general term *chemoselective ligation*. Unfortunately, the resulting thioester linkage is somewhat unstable to basic conditions and thus its utility may be limited.

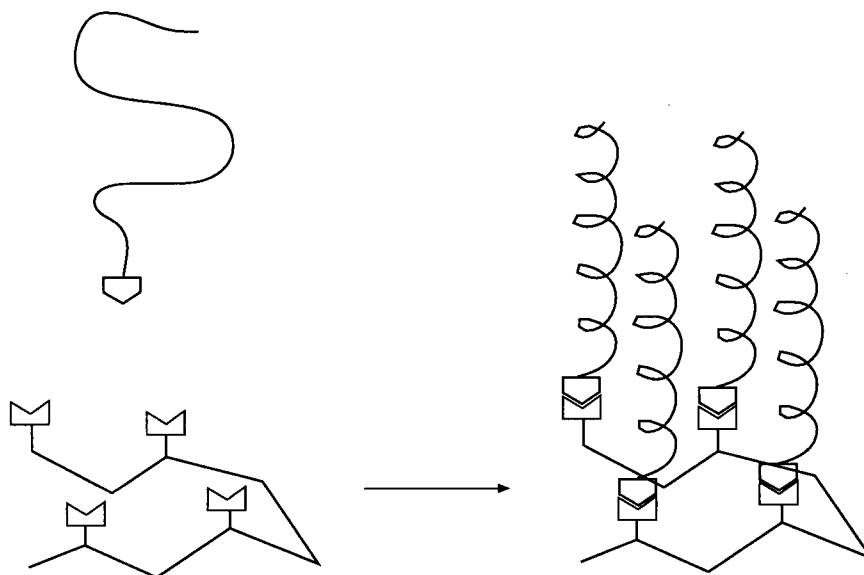
Others followed in the development of methods for ligating unprotected peptide segments to other peptides or templates. Some of the following methods were derived from established methods for chemically modifying or cross-linking natural proteins.<sup>8</sup> Some examples include the use of disulfide,<sup>9</sup> oxime<sup>10</sup> and thioether<sup>11</sup> bonds as well as hydrazone,<sup>12</sup> thiazolidine<sup>13</sup> and maleimide<sup>14</sup> rings.




Some researchers were also interested in the synthesis of natural proteins and thus the incorporation of the unnatural functionalities above was an undesirable artifact of the coupling

procedure. This led to the development of several coupling procedures able to link suitably activated unprotected peptide segments with a "natural" amide bond. Examples have been reported independently by both Kent<sup>15</sup> and Tam.<sup>16</sup>

With respect to template-assembled de novo four-helix bundle synthesis, chemoselective ligation is now the most common method of linking unprotected peptides to templates. For example, Dawson and Kent attached bromoacetyl groups to the lysines on Mutter's peptide template (Figure 3.1).<sup>6</sup> Subsequent reaction with unprotected peptides possessing C-terminal thioacid groups resulted in a chemoselective reaction in the presence of other functional groups such as amines, alcohols and carboxylic acids. Tuchscherer used an oxime bond to link four unprotected peptides to Mutter's lysine-rich template.<sup>10a</sup> The lysine groups on the template were functionalized with aldehydes while aminooxyacetyl groups were attached to the peptides. Subsequent reaction to form the oxime bond proceeded quantitatively (Figure 3.1). In another approach, Futaki and coworkers used a cysteine-rich peptide template in conjunction with activated chloroacetylated peptides. Again, this reaction proceeded chemoselectively to form a thioether bond via the cysteine (Figure 3.1).<sup>9b</sup>

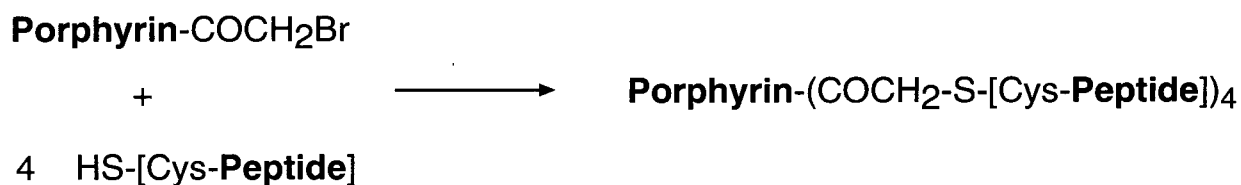
**Figure 3.1.** Schematic Illustration of the Ligation of Unprotected Peptides to Mutter-Type Peptide Templates by (a) Dawson and Kent<sup>6</sup> and (b) Futaki et. al.<sup>9b</sup> and (c) Tuchscherer.<sup>10</sup>



	(a) Dawson and Kent:	(b) Futaki et. al.:	(c) Tuchscherer:
	peptide-CO-SH	ClCH <sub>2</sub> -CO-NH-peptide	peptide-CO-CH <sub>2</sub> ONH <sub>2</sub>
+	+	+	+
	BrCH <sub>2</sub> CO-template	template-Cys-SH	HCOCO-template
↓	↓	↓	↓
	peptide-CO-S-CH <sub>2</sub> CO-template	template-Cys-S-CH <sub>2</sub> CO-NH-peptide	peptide-CO-CH <sub>2</sub> ON=CHCO-template

DeGrado and coworkers developed an similar approach to Futaki and coupled unprotected peptides to a porphyrin template.<sup>11a</sup> Reaction of a cysteine-containing peptide with a bromoacetyl group on the porphyrin afforded the desired four-helix bundle via thioether bonds (Scheme 3.2). This synthetic strategy was recently used in the synthesis of a variety of template-assembled four-helix bundles.<sup>17</sup>

**Scheme 3.2.** DeGrado's Ligation of Unprotected Peptides to a Porphyrin Template Using a Bromoacetyl Group and a Cysteine Residue.<sup>11a</sup>

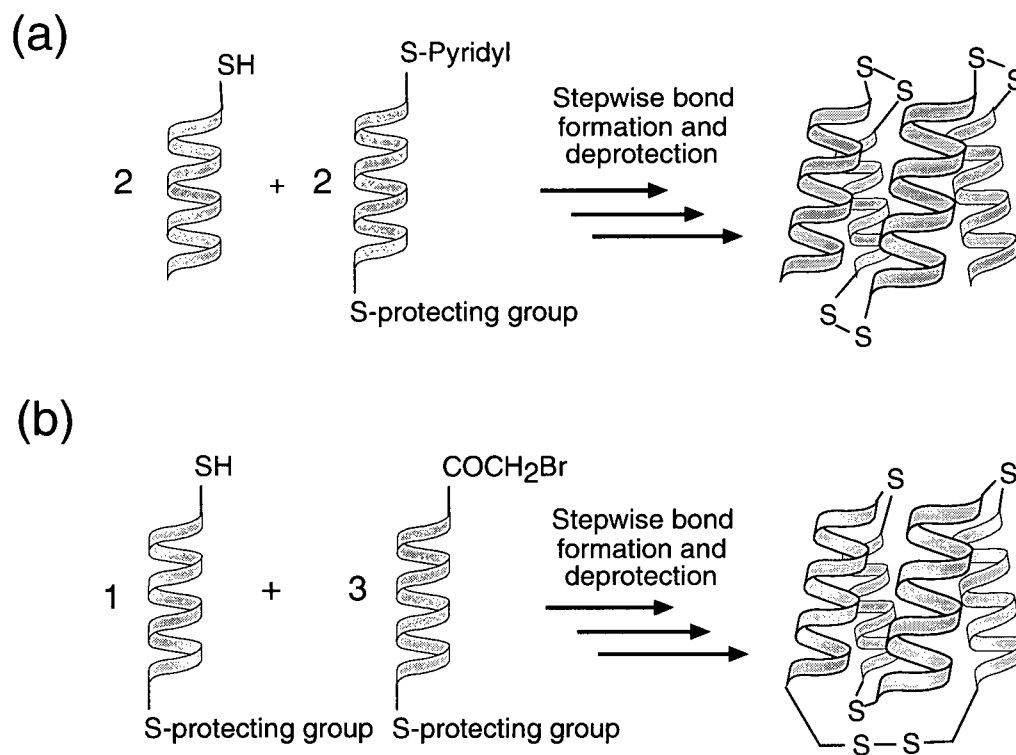


As mentioned above, various methods of protein cross-linking are known<sup>8</sup> and may be applied to de novo protein design. For example, Futaki and coworkers use such a strategy and linked several unprotected peptides via disulfide bonds.<sup>9</sup> Reaction of a thiol-bearing cysteine residue with another cysteine activated with an *S*-pyridyl moiety resulted in disulfide-linked helices in a de novo four-helix bundle (Figure 3.2).

In similar work, Futaki and coworkers used a combination of the above-mentioned disulfide strategy and a thioether strategy similar to that of DeGrado (Figure 3.2).<sup>11c</sup> The N-terminus of one peptide was activated with a bromoacetyl group and subsequently reacted with

the thiol of another cysteine-containing peptide. The resulting de novo four-helix bundle possessed both thioether and disulfide cross-links.

**Figure 3.2.** Schematic Illustration of Futaki's Cross-Linked De Novo Four-Helix Bundles Using (a) All Disulfide Bonds<sup>9</sup> and (b) Both Disulfide and Thioether Bonds.<sup>11c</sup>



In summary, there are many ways to chemoselectively link unprotected peptide strands to other peptides or templates. De novo design using template-assembly can exploit these methods to synthesize a variety of analogues with different linkages. In this chapter, a synthetic strategy for linking unprotected peptides to tetrathiol cavitand **3** (Chapter Two) will be modeled using activated single amino acids protected as their ethyl esters. In addition, the hydrogen

bonding patterns of these model compounds will be studied to aid in the design of more complex four-helix bundles.

## **B. Results and Discussion**

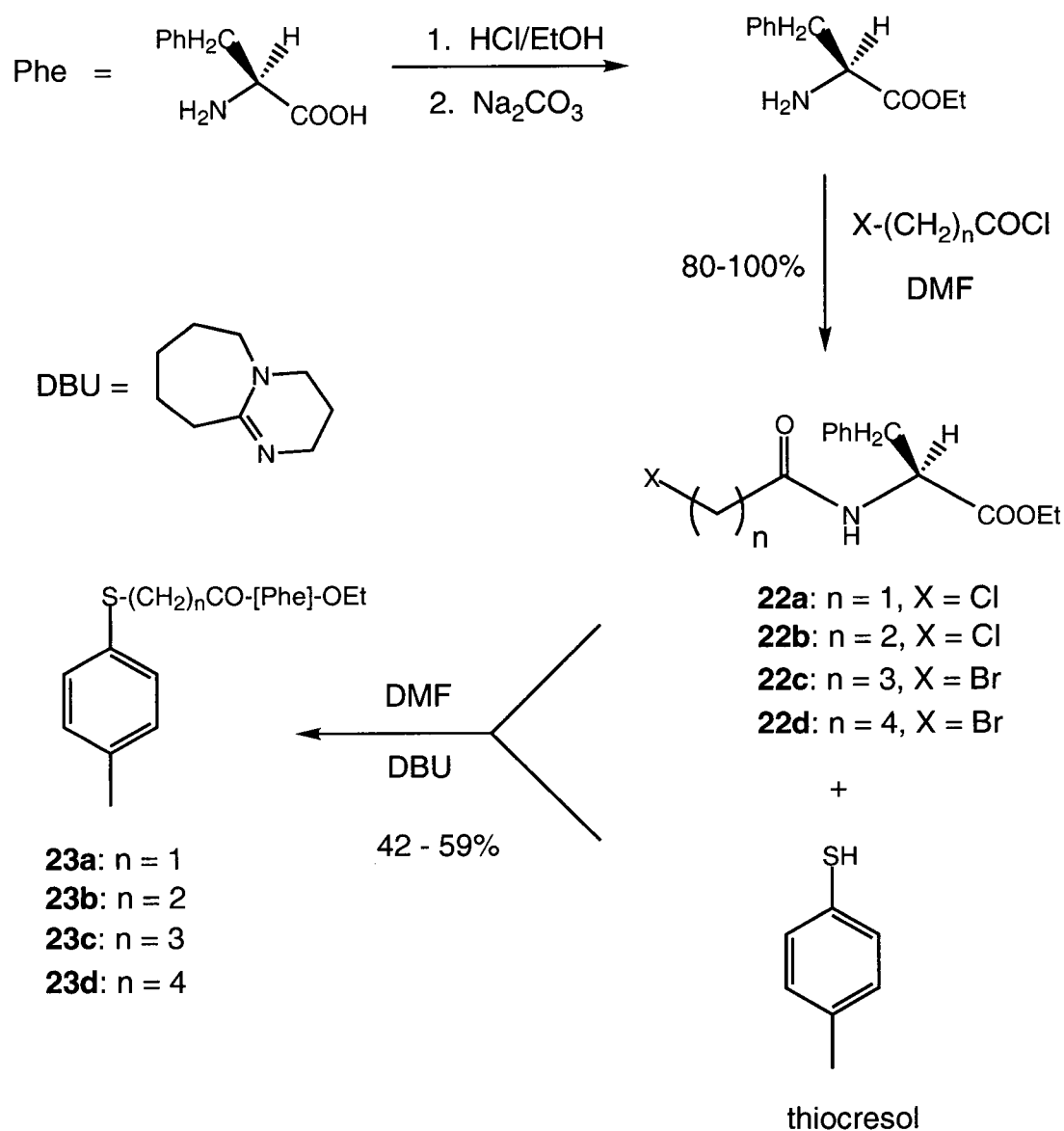
### **i. Synthesis**

Chapter Two described the synthesis of a variety of cavitands possessing functional groups that hold promise for incorporating peptide strands. We hoped that as a starting point, thiol groups would possess sufficient nucleophilicity to chemoselectively react with unprotected peptides. As a synthetic model to test the suitability of such a reaction, we first chose to couple single amino acids to a single thiol site on thiocresol, then to the four thiol sites on methyl-footed tetrathiol cavitand **3**. More specifically, we chose the amino acid phenylalanine protected as its ethyl ester to function as such a model. Note that all of the results presented in this Chapter have been published.<sup>18</sup>

Following the synthetic strategy proposed in Chapter Two, nucleophilic sites (thiols) on the cavitand were to react with electrophilic sites on the peptide. We thus needed to incorporate a suitable electrophilic functionality into the peptide strand. We chose the chloroacetyl group for two main reasons: (1) chloroacetyl groups should provide a suitable electrophile for selective reaction with thiols and (2) chloroacetyl groups should be synthetically simple to incorporate into the N-terminus of peptide strands.

In synthesizing the model amino acids, phenylalanine was first protected as its ethyl ester by treatment with HCl in ethanol to form the HCl salt of phenylalanine ethyl ester or HCl·Phe-OEt (Scheme 3.3).<sup>19</sup> The free base, Phe-OEt, was subsequently isolated by treatment of HCl·Phe-OEt with aqueous Na<sub>2</sub>CO<sub>3</sub>.<sup>20</sup> Treatment of Phe-OEt with chloroacetyl chloride in DMF afforded activated chloroacetylated phenylalanine ethylester (ClCH<sub>2</sub>CO-[Phe]-OEt) **22a** quantitatively. Subsequent reaction with thiocresol in DMF in the presence of the base 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) afforded the synthetic model compound **23a** in 42 % yield.

**Scheme 3.3.** Synthesis of Thiocresol-Based Control Compounds **23a-d**.



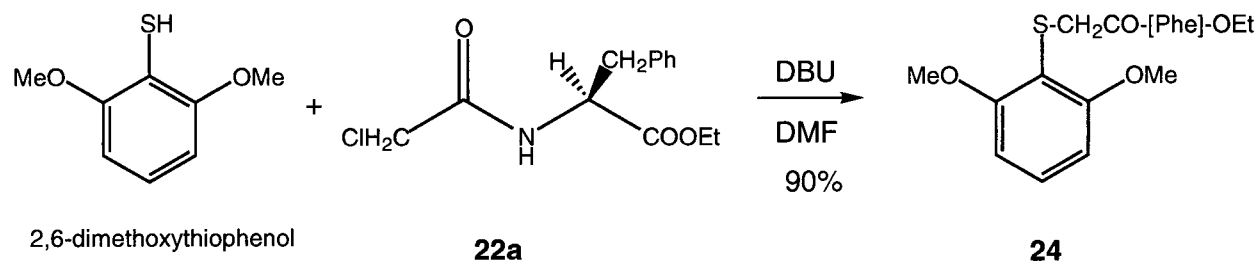
One of the goals of this thesis is to study the effect of the cavitand on the resulting structure of the de novo four-helix bundle. The extreme rigidity of the cavitand is expected to have a dramatic structural effect depending on the proximity of the peptides to the cavitand. We thus hoped to extend the linker group between the cavitand and the peptides in order to study



this effect. In efforts to model such a synthesis, we synthesized thiocresol-based model compounds **23b**, **23c** and **23d**, where the linkages include two, three and four methylenes, respectively (Scheme 3.3). These compounds were synthesized by alkylation of thiocresol with the corresponding acid chlorides **22b-d**:  $X-(CH)_nCO-Phe$  (where  $n = 2, 3$  and  $4$ , respectively). Note that for  $n = 3$  and  $4$ , the chlorine atom was replaced with a bromine. It was found that as the number of methylenes increased to 3 and 4, the electrophilicity of the carbon adjacent to the halogen leaving group was greatly reduced the further its distance from the carbonyl.

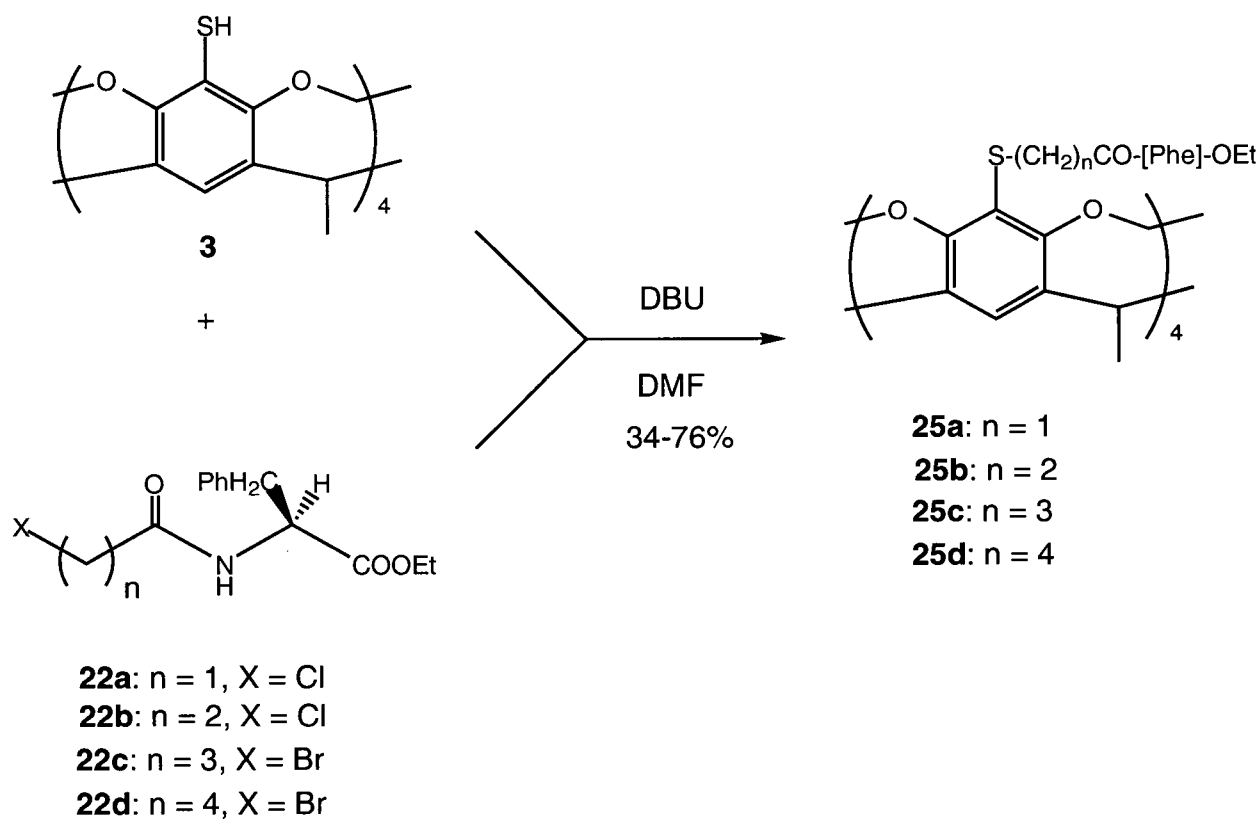
As an additional control compound for the cavitand-based models, 2,6-dimethoxythiophenol<sup>21</sup> was reacted with chloroacetyl **22a** to afford control compound **24** in 90% yield (Scheme 3.4). This compound is relevant to the discussion on hydrogen bonding below.

**Scheme 3.4.** Synthesis of 2,6-Dimethoxythiocresol-Based Control Compound **24**.



We next extended this synthetic methodology to the cavitand system, which serves as the ultimate synthetic model. Coupling of methyl-footed tetrathiol cavitand **3** to the corresponding activated amino acids **22a-d** afforded model de novo proteins **25a-d** in yields of 34 - 76% (Scheme 3.5).

**Scheme 3.5.** Synthesis of Model De Novo Proteins **25a-d**.



The syntheses presented in this chapter demonstrate that between one and four methylene units can be incorporated between single activated amino acids and tetrathiol cavitand **3** with high efficiency. Although the number of reactive sites were limited (in comparison to unprotected peptides) in these model syntheses, the results nevertheless bode well for coupling unprotected peptides to tetrathiol **3**. This will be explored further in Chapters Four and Five.

## ii. Hydrogen Bonding

An understanding of the hydrogen bonding patterns observed in these model single amino acid-cavitand systems would provide useful information for the study of more complex peptide-cavitand systems which require a definite set of hydrogen bonding networks (i.e.,  $i \cdots i + 4$  pattern found in  $\alpha$ -helices) to stabilize their structures. We thus examined the hydrogen bonding interactions of the amide NH protons using  $^1\text{H}$  NMR and IR spectroscopy.

$^1\text{H}$  NMR chemical shifts of hydrogen bonded NH groups are typically observed further downfield from those not hydrogen bonded. In addition, NH groups are typically hydrogen bonded to polar hydrogen bonding solvent molecules like DMSO- $d_6$  but not to solvents like  $\text{CDCl}_3$ . As a result, the difference between the chemical shifts in the two solvents,  $\Delta\delta$ , ( $\Delta\delta = \delta_{\text{DMSO}} - \delta_{\text{CDCl}_3}$ ) is indicative of the extent of hydrogen bonding of the NH group in  $\text{CDCl}_3$ .

Infrared (IR) spectroscopy provides an alternative approach to studying hydrogen bonding. Stretching frequencies for hydrogen bonded protons are typically shifted to lower wavenumbers as compared to non-hydrogen bonded protons.<sup>22</sup> In addition, if a hydrogen exists in both a non-hydrogen bonded and hydrogen bonded state, two bands will be observed due to the fast IR timescale. In contrast, due to the slower  $^1\text{H}$  NMR timescale,  $^1\text{H}$  NMR spectra usually show a time-averaged signal. Hence, this provides additional information as to the hydrogen bonded state of any given proton.

We recorded the chemical shifts of the NH protons in compounds **23a-d**, **24**, and **25a-d** in both DMSO- $d_6$  and  $\text{CDCl}_3$  (Table 3.1). We also investigated the concentration dependence of the chemical shifts in  $\text{CDCl}_3$  and found them to be largely independent of concentration: the  $\Delta\delta$  ( $\delta_{50\text{ mM}} - \delta_{1\text{ mM}}$ ) were +0.09 ppm for thiocresol-based **23d**, +0.06 ppm for thiocresol-based **23c**

and  $< 0.04$  ppm for control compounds **23a-b** and **24**. The  $\Delta\delta$  ( $\delta_{20\text{ mM}} - \delta_{1\text{ mM}}$ ) were  $< 0.04$  ppm for cavitand models **25a-d**. These data suggest that any observed hydrogen bonding is largely due to an intramolecular interaction. IR data for compounds **23a-d**, **24**, and **25a-d** in  $\text{CDCl}_3$  are also shown in Table 3.1.

Evident from Table 3.1 is that compounds **24** and **25a** exhibit significant hydrogen bonding of their NH groups. Their  $^1\text{H}$  NMR  $\Delta\delta$  values are quite small (0.16 and 0.5 ppm) and their IR spectra show predominantly the hydrogen bonded form displayed at lower wavenumbers (ca.  $3340\text{ cm}^{-1}$ ). Thiocresol based control **23a** shows a weak hydrogen bonding interaction with an intermediary  $\Delta\delta$  value ( $\sim 1.5$  ppm) and IR stretching frequency ( $3373\text{ cm}^{-1}$ ).

Compounds with two or more methylene linker groups show little or no hydrogen bonding as indicated by their high  $\Delta\delta$  values (ca. 2.0 -2.5 ppm) and IR spectra displaying predominantly the non-hydrogen bonded form (ca.  $3420\text{ cm}^{-1}$ ).

**Table 3.1.** Amide N-H  $^1\text{H}$  NMR Chemical Shifts and IR Stretching Frequencies for De Novo Protein Control and Model Compounds.

Compound	Abbreviated Formula	$\delta_{\text{NH}}$ (ppm) in $\text{CDCl}_3$	$\delta_{\text{NH}}$ (ppm) in $\text{DMSO-}d_6$	$\Delta\delta$ (ppm) <sup>a</sup>	IR Stretching Frequency ( $\text{cm}^{-1}$ ) in $\text{CDCl}_3$ <sup>b</sup>
<b>23a</b>	thiocresol- ( $\text{CH}_2$ )CO-[Phe]-OEt	7 <sup>c</sup>	8.53	~1.5	3373
<b>23b</b>	thiocresol- ( $\text{CH}_2$ ) <sub>2</sub> CO-[Phe]-OEt	5.97	8.38	2.41	3425
<b>23c</b>	thiocresol- ( $\text{CH}_2$ ) <sub>3</sub> CO-[Phe]-OEt	5.85	8.32	2.47	3427
<b>23d</b>	thiocresol- ( $\text{CH}_2$ ) <sub>4</sub> CO-[Phe]-OEt	5.80	8.26	2.46	3428
<b>25a</b>	tetrathiol- ( $\text{CH}_2$ CO-[Phe]-OEt) <sub>4</sub>	7.87	8.37	0.50	3418 (w) 3346 (s)
<b>25b</b>	tetrathiol- (( $\text{CH}_2$ ) <sub>2</sub> CO-[Phe]-OEt) <sub>4</sub>	6.37	8.35	1.97	3422 (s) 3359 (w)
<b>25c</b>	tetrathiol- (( $\text{CH}_2$ ) <sub>3</sub> CO-[Phe]-OEt) <sub>4</sub>	6.10	8.25	2.15	3426 (s) 3372 (w)
<b>25d</b>	tetrathiol- (( $\text{CH}_2$ ) <sub>4</sub> CO-[Phe]-OEt) <sub>4</sub>	5.86	8.23	2.37	3428 (s) 3350 (vw)
<b>24</b>	2,6-dimethoxythiophenol- $\text{CH}_2$ CO-[Phe]-OEt	8.10	8.26	0.16	3336

<sup>a</sup>  $\Delta\delta = \delta_{\text{DMSO}} - \delta_{\text{CDCl}_3}$

<sup>b</sup> s = strong; w = weak; vw = very weak.

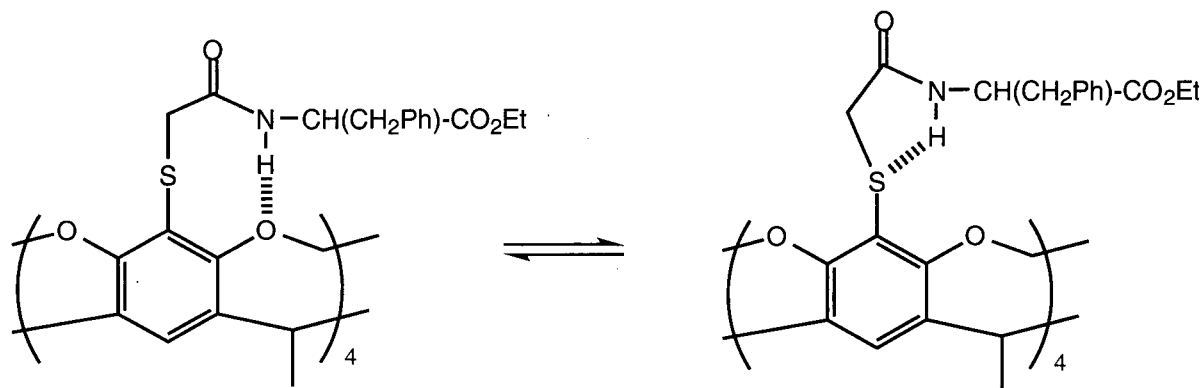
<sup>c</sup> Coincident with aromatic peaks; estimated from COSY spectrum.

We attribute part of the observed hydrogen bonding in the cavitand-based systems to an interaction between the amide NH and the bridge acetal oxygens. Support for this interaction comes from the increased hydrogen bonding observed in 2,6-dimethoxythiophenol-based control **24** when compared to thiocresol-based control **23a**. The simple addition of two methoxy groups substituted *ortho* to the amino acid effectively mimics the effect of the bridge oxygens in the cavitand system and results in increased hydrogen bonding when compared to the thiocresol-based control **23a**. Upon consideration of the compounds possessing extra methylene linker units, it appears that such an eight-membered ring hydrogen bond is favored over 9-, 10- and 11-membered ring hydrogen bonds. This type of "NH-bridge oxygen" hydrogen bonding interaction has been previously observed in a cavitand-based system.<sup>23</sup>

Interestingly, thiocresol-based control **23a** also shows a weak hydrogen bonding interaction. This is mostly likely due to an interaction between the amide NH proton and the aryl sulfide. Sulfur, although not commonly associated with hydrogen bonding, can indeed act as a hydrogen bond acceptor.<sup>24</sup> A survey of high resolution protein structures has also shown that this type of interaction is not uncommon.<sup>25</sup> In compound **23a**, the hydrogen bond can enclose a five-membered ring and appears to be favored over 6-, 7- and 8-membered ring hydrogen bonds.

Considering both types of hydrogen bonding, a unified picture emerges with compounds **25a** and **24** in an equilibrium process between two different but indistinguishable hydrogen bonded states: (1) hydrogen bonded to the bridge oxygen and (2) hydrogen bonded to the aryl sulfide. This is depicted below in Figure 3.3 for compound **25a**.

**Figure 3.3.** Potential Hydrogen Bonding Interactions in Compound **25a**.



The strong hydrogen bonding observed in compound **25a** is interesting and may be an asset in the design of more complex de novo four-helix bundles. As mentioned in Chapter One, the first four amide NH's in an  $\alpha$ -helix do not have complementary carbonyl hydrogen bonding acceptors. Hence, a relatively strong interaction between the first NH and the bridge oxygen and/or aryl sulfide may act as an *N-cap* and serve to influence and/or stabilize the helical bundle.

## C. Conclusion

In summary, a synthetic strategy for de novo protein synthesis was modeled using both thiocresol- and cavitand-based systems. It was found that between one and four methylene moieties could be incorporated efficiently between the amino acid and methyl-footed tetrathiol cavitand **3**. Hence, their syntheses hold promise for coupling of unprotected peptide fragments in pursuit of de novo proteins. In addition, it was found that compounds with a single methylene linker unit possessed enhanced hydrogen bonding capabilities, which may be useful and important in the design of cavitand-based four-helix bundles.



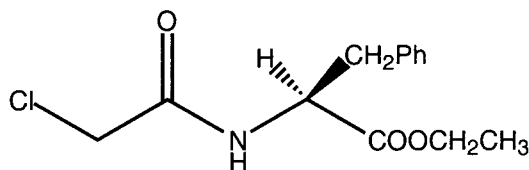
## D. Experimental

### i. General

All commercially available reagents were used without further purification. LSIMS mass spectra, elemental analyses, melting points and column chromatography were performed as described in Chapter Two.  $^1\text{H}$  NMR spectra were run on a Bruker AC-200E (200 MHz) using residual  $\text{CHCl}_3$  as a reference ( $\delta = 7.24$  ppm). Infrared spectra were recorded on a ATI Mattson Genesis Series FTIR spectrometer using NaCl cells of path length 0.516 mm. Samples for IR were solutions of  $\text{CDCl}_3$  and peaks were referenced to the  $1600\text{ cm}^{-1}$  peak of polystyrene. Synthetic procedures for all compounds reported in the next section have been published.<sup>18</sup>

### ii. Synthetic Procedures

#### $\text{ClCH}_2\text{CO-}[\text{Phe}]\text{-OEt 22a.}$



**Procedure "A":** A mixture of Phe-OEt (50 mg, 0.26 mmol) and chloroacetyl chloride (22  $\mu\text{L}$ , 0.29 mmol) in DMF was stirred for 1 h at rt. The DMF was removed *in vacuo* and the residue

was purified by column chromatography (9:1, EtOAc:hexanes) to yield 70 mg of **22a** (100%) as a white solid: mp 61-64 °C;

**<sup>1</sup>H NMR (200 MHz, CDCl<sub>3</sub>):** δ 7.08 - 7.29 (m, 5H, Ph), 6.98 (d, *J* = 7.1 Hz, 1H, NH), 4.82 (m, 1H, NCH), 4.16 (q, *J* = 7.1 Hz, 2H, CH<sub>2</sub>CH<sub>3</sub>), 4.00 (s, 2H, ClCH<sub>2</sub>), 3.13 (d, *J* = 5.9 Hz, 2H, CH<sub>2</sub>Ph), 1.22 (t, *J* = 7.1 Hz, 3H, CH<sub>2</sub>CH<sub>3</sub>) ppm.

**HRMS (LSIMS<sup>+</sup>, thioglycerol):** Calcd for C<sub>13</sub>H<sub>17</sub>ClNO<sub>3</sub> (M + H)<sup>+</sup>: 270.0897; Found 270.0897.

**Cl(CH<sub>2</sub>)<sub>2</sub>CO-[Phe]-OEt 22b.**

Procedure "A" was employed using Phe-OEt (500 mg, 2.59 mmol) and 3-chloropropionyl chloride (247 μL, 2.59 mmol) to yield 586 mg of **22b** (80%) as a yellow oil;

**<sup>1</sup>H NMR (200 MHz, CDCl<sub>3</sub>)** δ 7.07 - 7.29 (m, 5H, Ph), 6.10 (d, *J* = 7.3 Hz, 1H, NH), 4.82 - 4.92 (m, 1H, NCH), 4.16 (q, *J* = 7.1 Hz, 2H, CH<sub>2</sub>CH<sub>3</sub>), 3.69 - 3.80 (m, 2H, ClCH<sub>2</sub>CH<sub>2</sub>), 3.12 (d, *J* = 5.6 Hz, 2H, CH<sub>2</sub>Ph), 2.58 - 2.66 (m, 2H, ClCH<sub>2</sub>CH<sub>2</sub>), 1.24 (t, *J* = 7.1 Hz, 3H, CH<sub>2</sub>CH<sub>3</sub>,) ppm.

**HRMS (LSIMS<sup>+</sup>, thioglycerol):** Calcd for C<sub>14</sub>H<sub>19</sub>ClNO<sub>3</sub> (M + H)<sup>+</sup>: 284.1053; Found 284.1050.

**Br(CH<sub>2</sub>)<sub>3</sub>CO-[Phe]-OEt 22c.**

Procedure "A" was employed using Phe-OEt (520 mg, 2.70 mmol) and 4-bromobutyryl chloride (312  $\mu$ L, 2.70 mmol) to yield 767 mg of **22c** (83%) as a yellow oil;

**<sup>1</sup>H NMR (200 MHz, CDCl<sub>3</sub>)**  $\delta$  7.06 - 7.31 (m, 5H, Ph), 5.95 (d,  $J$  = 7.3 Hz, 1H, NH), 4.80 - 4.90 (m, 1H, NCH), 4.11 (q, 2H, CH<sub>2</sub>CH<sub>3</sub>,  $J$  = 7.1 Hz), 3.34 - 3.60 (m, 2H, BrCH<sub>2</sub>), 3.06 - 3.12 (m, 2H, CH<sub>2</sub>Ph), 2.30 - 2.38 (m, 2H, BrCH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>), 2.01 - 2.15 (m, 2H, BrCH<sub>2</sub>CH<sub>2</sub>), 1.23 (t,  $J$  = 7.1 Hz, 3H, CH<sub>2</sub>CH<sub>3</sub>,) ppm.

**HRMS (LSIMS<sup>+</sup>, thioglycerol):** Calcd for C<sub>15</sub>H<sub>21</sub>BrNO<sub>3</sub> (M + H)<sup>+</sup>: 342.0705; Found 342.0701.

**Br(CH<sub>2</sub>)<sub>4</sub>CO-[Phe]-OEt 22d.**

Procedure "A" was employed using Phe-OEt (496 mg, 2.57 mmol) and 5-bromovaleryl chloride (344  $\mu$ L, 2.57 mmol) to yield 900 mg of **22d** (98%) as a yellow oil:

**<sup>1</sup>H NMR (200 MHz, CDCl<sub>3</sub>)**  $\delta$  7.05 - 7.29 (m, 5H, ArH), 5.93 (d,  $J$  = 7.2 Hz, 1H, NH), 4.80 - 4.90 (m, 1H, NCH), 4.15 (q,  $J$  = 7.1 Hz, 2H, CH<sub>2</sub>CH<sub>3</sub>,), 3.36 (t,  $J$  = 7.1 Hz, 2H, BrCH<sub>2</sub>,), 3.06 - 3.12 (m, 2H, CH<sub>2</sub>Ph), 2.15 - 2.22 (m, 2H, CH<sub>2</sub>CON), 1.67 - 1.85 (m, 4H, BrCH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>) ppm.

**HRMS (LSIMS<sup>+</sup>, thioglycerol):** Calcd for C<sub>16</sub>H<sub>23</sub>BrNO<sub>3</sub> (M + H)<sup>+</sup>: 356.0861; Found 356.0856.

**4-CH<sub>3</sub>-C<sub>6</sub>H<sub>4</sub>SCH<sub>2</sub>CO-[Phe]-OEt 23a.**

**Procedure "B":** A mixture of thiocresol (50 mg, 0.19 mmol), chloroacetyl **22a** (23 mg, 0.19 mmol) and DBU (31  $\mu$ L, 0.20 mmol) in dry DMF (1 ml) was stirred overnight at rt under N<sub>2</sub>. The DMF was removed *in vacuo* and the residue was purified by column chromatography (9:1 hexanes:EtOAc, then 4:1 hexanes:EtOAc) to yield 28 mg of **23a** (42%) as a white solid: mp 66-67 °C;

**IR (CDCl<sub>3</sub>)** 3373 (NH), 1737 (ester CO), 1668 (amide CO) cm<sup>-1</sup>.

**<sup>1</sup>H NMR (200 MHz, CDCl<sub>3</sub>)**  $\delta$  6.95 - 7.26 (m, 10H, ArH, CH<sub>2</sub>Ph, NH), 4.79 (dd, *J* = 6.0, 7.9 Hz, 1H, NCH), 4.11 (q, *J* = 7.1 Hz, 2H, CH<sub>2</sub>CH<sub>3</sub>), 3.55 (s, 2H, SCH<sub>2</sub>), 3.03 (d, *J* = 6.0 Hz, 2H, CH<sub>2</sub>Ph), 2.29 (s, 3H, ArCH<sub>3</sub>), 1.18 (t, *J* = 7.1 Hz, 3H, CH<sub>2</sub>CH<sub>3</sub>) ppm.

**HRMS (LSIMS<sup>+</sup>, thioglycerol):** Calcd for C<sub>20</sub>H<sub>24</sub>NO<sub>3</sub>S (M + H)<sup>+</sup>: 358.1477; Found: 358.1478.

**4-CH<sub>3</sub>-C<sub>6</sub>H<sub>4</sub>S(CH<sub>2</sub>)<sub>2</sub>CO-[Phe]-OEt 23b.**

Procedure "B" was employed using thiocresol (65 mg, 0.52 mmol), chloropropionyl **22b** (74 mg, 0.26 mmol) and DBU (43  $\mu$ L, 0.28 mmol) to yield 57 mg of **23b** (59%) as a yellow solid: mp 44 °C;

**IR** (CDCl<sub>3</sub>) 3425 (NH), 1735 (ester CO), 1673 (amide CO) cm<sup>-1</sup>.

**<sup>1</sup>H NMR** (200 MHz, CDCl<sub>3</sub>)  $\delta$  7.06 - 7.22 (m, 9H, ArH, CH<sub>2</sub>Ph), 6.01 (d,  $J$  = 7.2 Hz, 1H, NH), 4.79 - 4.87 (m, 1H, NCH), 4.15 (q, 2H, CH<sub>2</sub>CH<sub>3</sub>,  $J$  = 7.1 Hz), 3.06 - 3.13 (m, 4H, CH<sub>2</sub>CHH<sub>2</sub>CO, CH<sub>2</sub>Ph), 2.41 (t,  $J$  = 7.4 Hz, 2H, SCH<sub>2</sub>), 2.30 (s, 3H, ArCH<sub>3</sub>), 1.23 (t,  $J$  = 7.1 Hz, 3H, CH<sub>2</sub>CH<sub>3</sub>) ppm.

**HRMS** (LSIMS<sup>+</sup>, thioglycerol): Calcd for C<sub>21</sub>H<sub>26</sub>NO<sub>3</sub>S (M + H)<sup>+</sup>: 372.1633; Found: 372.1634.

**4-CH<sub>3</sub>-C<sub>6</sub>H<sub>4</sub>S(CH<sub>2</sub>)<sub>3</sub>CO-[Phe]-OEt 23c.**

Procedure "B" was employed using thiocresol (37 mg, 0.30 mmol) bromobutyl **22c** (52 mg, 0.15 mmol) and DBU (25  $\mu$ L, 0.17 mmol) to yield 35 mg of **23c** (59%) as a yellow oil:

**IR** (CDCl<sub>3</sub>) 3427 (NH), 1735 (ester CO), 1673 (amide CO) cm<sup>-1</sup>.

**<sup>1</sup>H NMR** (200 MHz, CDCl<sub>3</sub>)  $\delta$  7.05 - 7.27 (m, 9H, ArH, CH<sub>2</sub>Ph), 5.89 (d,  $J$  = 7.4 Hz, 1H, NH), 4.79 - 4.89 (m, 1H, NCH), 4.15 (q,  $J$  = 7.1 Hz, 2H, CH<sub>2</sub>CH<sub>3</sub>), 3.08 (two

overlapping ABX dd,  $J = 5.2, 5.8, 13.3$ ,  $\Delta\nu = 18.0$  Hz, 2H,  $\text{HCH}_2\text{Ph}$ ), 2.77-2.91 (m, 2H,  $\text{CH}_2\text{CH}_2\text{CO}$ ), 2.26 - 2.33 (m, 2H,  $\text{SCH}_2\text{CH}_2$ ), 2.29 (s, 3H,  $\text{ArCH}_3$ ), 1.80 - 1.95 (m, 2H,  $\text{SCH}_2$ ), 1.22 (t,  $J = 7.1$  Hz, 3H,  $\text{CH}_2\text{CH}_3$ ) ppm.

**HRMS (LSIMS<sup>+</sup>, thioglycerol):** Calcd for  $\text{C}_{22}\text{H}_{28}\text{NO}_3\text{S}$  ( $\text{M} + \text{H}$ )<sup>+</sup>: 386.1790; Found: 386.1783.

**4-CH<sub>3</sub>-C<sub>6</sub>H<sub>4</sub>S(CH<sub>2</sub>)<sub>4</sub>CO-[Phe]-OEt 23d.**

Procedure "B" was employed using thiocresol (17 mg, 0.14 mmol), bromovaleryl **22d** (50 mg, 0.14 mmol) and DBU (23  $\mu\text{L}$ , 0.15 mmol) to yield 25 mg of **23d** (45%) as a yellow oil:

**IR (CDCl<sub>3</sub>)** 3428 (NH), 1734 (ester CO), 1671 (amide CO)  $\text{cm}^{-1}$ .

**<sup>1</sup>H NMR (200 MHz, CDCl<sub>3</sub>)**  $\delta$  7.04 - 7.31 (m, 9H, ArH,  $\text{CH}_2\text{Ph}$ ), 5.89 (d,  $J = 7.5$  Hz, 1H, NH), 4.79 - 4.89 (m, 1H, NCH), 4.15 (q,  $J = 7.1$  Hz, 2H,  $\text{CH}_2\text{CH}_3$ ), 3.07 (two overlapping ABX dd,  $J = 5.9, 5.9, 13.7$ ,  $\Delta\nu = 14.8$  Hz, 2H,  $\text{CH}_2\text{Ph}$ ), 2.84 (t,  $J = 7.0$  Hz, 2H,  $\text{CH}_2\text{CH}_2\text{CO}$ ), 2.29 (s, 3H,  $\text{ArCH}_3$ ), 2.16 (t,  $J = 7.0$  Hz, 2H,  $\text{SCH}_2$ ), 1.51 - 1.77 (m, 4H,  $\text{SCH}_2\text{CH}_2\text{CH}_2$ ), 1.23 (t,  $J = 7.1$  Hz, 3H,  $\text{CH}_2\text{CH}_3$ ) ppm.

**HRMS (LSIMS<sup>+</sup>, thioglycerol):** Calcd for  $\text{C}_{23}\text{H}_{30}\text{NO}_3\text{S}$  ( $\text{M} + \text{H}$ )<sup>+</sup>: 400.1946; Found: 400.1951.

**2,6-(CH<sub>3</sub>O)<sub>2</sub>-C<sub>6</sub>H<sub>4</sub>SCH<sub>2</sub>CO-[Phe]-OEt **24**.**

A mixture of 2,6-dimethoxythiophenol<sup>21</sup> (71 mg, 0.26 mmol), chloroacetyl **22a** (45 mg, 0.26 mmol) and DBU (43  $\mu$ L, 0.29 mmol) in dry DMF (1 ml) was stirred overnight at rt under N<sub>2</sub>. The DMF was removed *in vacuo* and the residue was purified by column chromatography (7:3, hexanes:ethyl acetate) to yield 96 mg of **24** (90%) as a colorless oil:

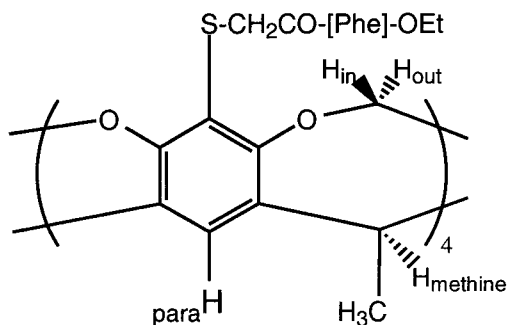
**IR** (CDCl<sub>3</sub>) 3336 (NH), 1735 (ester CO), 1661 (amide CO) cm<sup>-1</sup>.

**<sup>1</sup>H NMR** (200 MHz, CDCl<sub>3</sub>)  $\delta$  8.10 (d, 1H, NH,  $J$  = 7.8 Hz), 7.00 - 7.30 (m, 6H, *para* ArH, CH<sub>2</sub>Ph), 6.53 (d, 2H, *meta* ArH,  $J$  = 8.4 Hz), 4.76 (ddd,  $J$  = 6.5, 6.5, 8.0 Hz, 1H, NCH), 4.06 (q,  $J$  = 7.1 Hz, 2H, CH<sub>2</sub>CH<sub>3</sub>), 3.77 (s, 6H, OCH<sub>3</sub>), 3.53 (AB quartet,  $J$  = 17.0 Hz,  $\Delta\nu$  = 21.1 Hz, 2H, SCH<sub>2</sub>), 2.98 (two overlapping ABX dd,  $J$  = 6.5, 6.5, 13.9 Hz,  $\Delta\nu$  = 20.8 Hz, 2H, CH<sub>2</sub>Ph), 1.13 (t, 3H, CH<sub>2</sub>CH<sub>3</sub>,  $J$  = 7.1 Hz) ppm.

**MS** (LSIMS<sup>+</sup>, thioglycerol)  $m/z$  (rel intensity): 404 ((M + H)<sup>+</sup>; 100).

**Anal.** Calcd for C<sub>21</sub>H<sub>25</sub>NO<sub>5</sub>S: C, 62.51; H, 6.25; N, 3.47. Found: C, 62.13; H, 5.93; N, 3.50.

**Tetrathiol-(CH<sub>2</sub>CO-[Phe]-OEt)<sub>4</sub> 25a.**



**Procedure "C":** A mixture of tetrathiol **3** (36 mg, 0.050 mmol), chloroacetyl **22a** (69 mg, 0.22 mmol) and DBU (33  $\mu$ L, 0.22 mmol) in DMF (1 mL) was stirred overnight at rt under N<sub>2</sub>. The DMF was removed *in vacuo* and the residue was purified by column chromatography (3:2, hexanes:acetone) to yield 63 mg of **25a** (76%) as a white solid: mp 94 °C (dec);

**IR (CDCl<sub>3</sub>):** 3418 (NH), 3346 (NH), 1735 (ester CO), 1664 (amide CO) cm<sup>-1</sup>.

**<sup>1</sup>H NMR (200 MHz, CDCl<sub>3</sub>)**  $\delta$  7.87 (d,  $J$  = 8.2 Hz, 4H, NH), 7.00 - 7.21 (m, 24H, CH<sub>2</sub>Ph, <sub>para</sub>H), 5.96 (d,  $J$  = 7.4 Hz, 4H, H<sub>out</sub>), 5.00 (q,  $J$  = 7.4 Hz, 4H, H<sub>methine</sub>), 4.71 - 4.75 (m, 4H, NCH), 4.32 (d,  $J$  = 7.4 Hz, 4H, H<sub>in</sub>), 4.06 (q,  $J$  = 7.1 Hz, 8H, CH<sub>2</sub>CH<sub>3</sub>), 3.46 (AB quartet,  $J$  = 17.2 Hz,  $\Delta v$  = 21.4 Hz, 8H, SCH<sub>2</sub>), 2.99 (two overlapping ABX dd,  $J$  = 6.3, 7.2, 13.9 Hz,  $\Delta v$  = 28.6 Hz, 8H, CH<sub>2</sub>Ph), 1.72 (d,  $J$  = 7.4 Hz, 12H, CHCH<sub>3</sub>), 1.11 (t,  $J$  = 7.1 Hz, 12H, CH<sub>2</sub>CH<sub>3</sub>) ppm.

**MS (LSIMS<sup>+</sup>, thioglycerol)  $m/z$  (rel intensity):** 1654 ((M + H)<sup>+</sup>; 100).

**Anal.** Calcd for C<sub>88</sub>H<sub>92</sub>N<sub>4</sub>O<sub>20</sub>S<sub>4</sub>: C, 63.91; H, 5.61; N, 3.39. Found: C, 63.55; H, 5.61; N, 3.41.



**Tetrathiol-((CH<sub>2</sub>)<sub>2</sub>CO-[Phe]-OEt)<sub>4</sub> **25b**.**

Procedure "C" was employed using tetrathiol **3** (78 mg, 0.11 mmol), chloropropionyl **22b** (135 mg, 0.477 mmol) and DBU (71  $\mu$ L, 0.48 mmol) to yield 87 mg of **25b** (47%) as a white solid: mp 133 °C (dec);

**IR** (CDCl<sub>3</sub>) 3422 (NH), 3359 (NH), 1734 (ester CO), 1668 (amide CO) cm<sup>-1</sup>.

**<sup>1</sup>H NMR** (200 MHz, CDCl<sub>3</sub>)  $\delta$  7.10 - 7.30 (m, 24H, CH<sub>2</sub>Ph, <sub>para</sub>H), 6.37 (d,  $J$  = 7.6 Hz, 4H, NH), 5.68 (d,  $J$  = 7.5 Hz, 4H, H<sub>in</sub>), 4.98 (q,  $J$  = 7.3 Hz, 4H, CHCH<sub>3</sub>), 4.80 - 4.90 (m, 4H, NCH), 4.11 - 4.22 (m, 12H, H<sub>in</sub>, CH<sub>2</sub>CH<sub>3</sub>), 2.94 - 3.16 (m, 16H, CH<sub>2</sub>Ph, CH<sub>2</sub>CH<sub>2</sub>CO), 2.24 - 2.31 (m, 8H, SCH<sub>2</sub>), 1.74 (d,  $J$  = 7.4 Hz, 12H, CHCH<sub>3</sub>), 1.22 (t,  $J$  = 7.1 Hz, 12H, CH<sub>2</sub>CH<sub>3</sub>) ppm.

**MS** (LSIMS<sup>+</sup>, thioglycerol)  $m/z$  (rel intensity): 1710 ((M + H)<sup>+</sup>; 100).

**Anal.** Calcd for C<sub>92</sub>H<sub>100</sub>N<sub>4</sub>O<sub>20</sub>S<sub>4</sub>·H<sub>2</sub>O: C, 63.94; H, 5.95; N, 3.24. Found: C, 63.80; H, 5.96; N, 3.10.

**Tetrathiol-((CH<sub>2</sub>)<sub>3</sub>CO-[Phe]-OEt)<sub>4</sub> **25c**.**

Procedure "C" was employed using tetrathiol **6** (100 mg, 0.139 mmol), bromobutyryl **22c** (0.21 g, 0.62 mmol) and DBU (92  $\mu$ L, 0.62 mmol) to yield 106 mg of **25c** (43%) as a white solid: mp 52 °C (dec);

**IR (CDCl<sub>3</sub>):** 3426 (NH), 3372 (NH), 1732 (ester CO), 1667 (amide CO) cm<sup>-1</sup>.

**<sup>1</sup>H NMR (200 MHz, CDCl<sub>3</sub>)**  $\delta$  7.04 - 7.24 (m, 24H, CH<sub>2</sub>Ph, <sub>para</sub>H), 6.11 (d,  $J$  = 8.0 Hz, 4H, NH), 5.90 (d,  $J$  = 7.5 Hz, 4H, H<sub>out</sub>), 5.00 (q,  $J$  = 7.4 Hz, 4H, CHCH<sub>3</sub>), 4.78 - 4.88 (m, 4H, NCH), 4.30 (d,  $J$  = 7.5 Hz, 4H, H<sub>in</sub>), 4.13 (q,  $J$  = 7.0 Hz, 8H, CH<sub>2</sub>CH<sub>3</sub>), 3.08 (two overlapping ABX dd,  $J$  = 6.2, 6.8, 14.2 Hz,  $\Delta\nu$  = 21.1 Hz, 8H, CH<sub>2</sub>Ph), 2.72 - 2.89 (m, 8H, CH<sub>2</sub>CH<sub>2</sub>CO), 2.30 - 2.37 (m, 8H, SCH<sub>2</sub>), 1.64 - 1.80 (m, 20H, SCH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>, CHCH<sub>3</sub>), 1.20 (t,  $J$  = 7.0 Hz, 12H, CH<sub>2</sub>CH<sub>3</sub>) ppm.

**MS (LSIMS<sup>+</sup>, thioglycerol)  $m/z$  (rel intensity):** 1766 ((M + H)<sup>+</sup>; 100).

**Anal.** Calcd for C<sub>96</sub>H<sub>108</sub>N<sub>4</sub>O<sub>20</sub>S<sub>4</sub>·H<sub>2</sub>O: C, 64.63; H, 6.21; N, 3.14. Found: C, 64.58; H, 6.13; N, 2.98.

**Tetrathiol-((CH<sub>2</sub>)<sub>4</sub>CO-[Phe]-OEt)<sub>4</sub> **25d**.**

Procedure "C" was employed using tetrathiol **3** (50 mg, 0.069 mmol), bromovaleryl **22d** (0.11 g, 0.31 mmol) and DBU (46  $\mu$ L, 0.31 mmol) to yield 54 mg of **25d** (34%) as a white solid: mp 99 °C (dec);

**IR (CDCl<sub>3</sub>)** 3428 (NH), 1732 (ester CO), 1670 (amide CO) cm<sup>-1</sup>.

**<sup>1</sup>H NMR (200MHz, CDCl<sub>3</sub>)**  $\delta$  7.05 - 7.26 (m, 24H, CH<sub>2</sub>Ph, <sub>para</sub>H), 5.88 (m, 8H, NH, H<sub>out</sub>), 4.98 (q,  $J$  = 7.2 Hz, 4H, CHCH<sub>3</sub>), 4.79 - 4.89 (m, 4H, NCH), 4.27 (d,  $J$  = 7.3 Hz, 4H, H<sub>in</sub>), 4.14 (q,  $J$  = 7.1 Hz, 8H, CH<sub>2</sub>CH<sub>3</sub>), 3.09 (two overlapping ABX dd,  $J$  = 5.7, 5.8, 14.2 Hz,  $\Delta\nu$  = 23.3 Hz,

8H,  $\text{CH}_2\text{Ph}$ ), 2.75 - 2.82 (m, 8H,  $\text{CH}_2\text{CH}_2\text{CO}$ ), 2.10 - 2.18 (m, 8H,  $\text{SCH}_2$ ), 1.45 - 1.74 (m, 28H,  $\text{SCH}_2\text{CH}_2\text{CH}_2$ ,  $\text{CHCH}_3$ ), 1.21 (t,  $J = 7.1$  Hz, 12H,  $\text{CH}_2\text{CH}_3$ ) ppm.

**MS (LSIMS<sup>+</sup>, thioglycerol)  $m/z$  (rel intensity):** 1822 ( $(\text{M} + \text{H})^+$ ; 100).

**Anal.** Calcd for  $\text{C}_{100}\text{H}_{116}\text{N}_4\text{O}_{20}\text{S}_4 \cdot 2\text{H}_2\text{O}$  : C, 64.63; H, 6.51; N, 2.74. Found: C, 64.43; H, 6.29; N, 2.80.

## E. References

1. Sasaki, T.; Kaiser, E. T. *J. Am. Chem. Soc.* **1989**, *111*, 380-381.
2. Mutter, M.; Tuchscherer, G. G.; Miller, C.; Altmann, K.-H.; Carey, R. I.; Wyss, D. F.; Labhardt, A. M.; Rivier, J. E. *J. Am. Chem. Soc.* **1992**, *114*, 1463-1470.
3. (a) Kiyama, S.; Fujii, H.; Yajima, H.; Moriga, M.; Takagi, A. *Int. J. Pept. Protein Res.* **1984**, *23*, 174-186; (b) Sasaki, T.; Findeis, M. A.; Kaiser, E. T. *J. Org. Chem.* **1991**, *56*, 3159-3168.
4. Walker, M. A. *Angew. Chem., Int. Ed. Engl.* **1997**, *36*, 1069-1071.
5. Schnölzer, M.; Kent, S. B. H. *Science* **1992**, *256*, 221-225.
6. Dawson, P. E.; Kent, S. B. H. *J. Am. Chem. Soc.* **1993**, *115*, 7263-7266.
7. (a) Canne, L. E.; Ferre-D'Amare, A. R.; Burley, S. K.; Kent, S. B. H. *J. Am. Chem. Soc.* **1995**, *117*, 2998-3007; (b) Baca, M.; Muir, T. W.; Schnolzer, M.; Kent, S. B. H. *J. Am. Chem. Soc.* **1995**, *117*, 1881-1887.
8. (a) Means, G. E.; Feeney, R. E. *Bioconjugate Chem.* **1990**, *1*, 2-12; (b) Brinkley, M. *Bioconjugate Chem.* **1992**, *3*, 2-13.
9. (a) Futaki, S.; Kitagawa, K. *Tetrahedron Lett.* **1994**, *35*, 1267-1270; (b) Futaki, S.; Ishikawa, T.; Niwa, M.; Yagami, T.; Kitagawa, K. *Tetrahedron Lett.* **1995**, *36*, 201-204; (c) Jiang, X.; Bishop, E. J.; Farid, R. S. *J. Am. Chem. Soc.* **1997**, *119*, 838-839.
10. (a) Tuchscherer, G. *Tetrahedron Lett.* **1993**, *34*, 8419-8422; Rose, K. *J. Am. Chem. Soc.* **1994**, *116*, 30-33.
11. (a) Choma, C. T.; Kaestle, K.; Akerfeldt, K. S.; Kim, R. M.; Groves, J. T.; DeGrado, W. F. *Tetrahedron Lett.* **1994**, *35*, 6191-6194; (b) Liu, C.-F.; Tam, J. P. *J. Chem. Soc. Chem. Commun.* **1997**, 1619-1620; (c) Futaki, S.; Ishikawa, T.; Niwa, M.; Kitagawa, K.; Yagami, T. *Tetrahedron Lett.* **1995**, *36*, 5203-5206; (d) Futaki, S.; Futaki, S.; Ishikawa, T.; Aoki, M.; Kondo, F.; Niwa, M.; Kitagawa, K.; Yagami, T. in *Peptides: Chemistry, Structure and Biology*; Kaumaya, P.T.P. and Hodges, R. S. (Eds); Mayflower Scientific Ltd.: England, 1996; pp 579-580.
12. Shao, J.; Tam, J. P. *J. Am. Chem. Soc.* **1995**, *117*, 3893-3899.
13. Botti, P.; Pallin, T. D.; Tam, J. P. *J. Am. Chem. Soc.* **1996**, *118*, 10018-10024.

14. Nefzi, A.; Sun, X.; Mutter, M. *Tetrahedron Lett.* **1995**, 36, 229-230.
15. (a) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. *Science* **1994**, 266, 776-779; (b) Dawson, P. E.; Churchill, M. J.; Ghadiri, M. R.; Kent, S. B. H. *J. Am. Chem. Soc.* **1997**, 119, 4325-4329.
16. (a) Tam, J. P.; Lu, Y.-A.; Liu, C.-F.; Shao, J. *Proc. Natl. Acad. Sci. USA* **1995**, 37, 933; (b) Liu, C.-F.; Rao, C.; Tam, J. P. *J. Am. Chem. Soc.* **1996**, 118, 307-312; (c) Liu, C.-F.; Tam, J. P. *Proc. Natl. Acad. Sci. USA* **1994**, 91, 6584-6588; (d) Zhang, L.; Tam, J. P. *Tetrahedron Lett.* **1997**, 38, 3-6.
17. (a) Wong, A. K.; Jacobsen, M. P.; Wnizor, D. J.; Fairlie, D. P. *J. Am. Chem. Soc.* **1998**, 120, 3836-3841; (b) Rau, H. K.; Haehnel, W. *Ber. Bunsenges. Phys. Chem.* **1996**, 100, 2052-2056.
18. Gibb, B. C.; Mezo, A. R.; Causton, A. S.; Fraser, J. R.; Tsai, F. C. S.; Sherman, J. C. *Tetrahedron* **1995**, 51, 8719-8732.
19. Bodanszky, M.; Bodanszky, A. *The Practice of Peptide Synthesis*; Springer-Verlag: New York, 1984.
20. Katakya, R.; Parker, D.; Teasdale, A.; Hutchinson, J. P.; Buschmann, H.-J. *J. Chem. Soc., Perkin Trans. 2* **1992**, 1347-1351.
21. Trost, B. M.; Jungheim, L. N. *J. Am. Chem. Soc.* **1980**, 102, 7910-7925.
22. Gallo, E. A.; Gellman, S. H. *J. Am. Chem. Soc.* **1993**, 115, 9774-9788.
23. Timmerman, P.; Verboom, W.; van Veggel, F. C. J. M.; van Hoorn, W. P.; Reinhoudt, D. N. *Angew. Chem., Int. Ed. Engl.* **1994**, 33, 1292-1295.
24. (a) Adman, E.; Watenpaugh, K. D.; Jensen, L. H. *Proc. Natl. Acad. Sci. USA* **1975**, 72, 4864-4858; (b) Krueger, P. J. *Tetrahedron* **1970**, 26, 4753-4764; (c) Zuika, I. V.; Bankovskii, Y. A. *Russ. Chem. Rev.* **1973**, 42, 22-36; (d) Kissinger, C. R.; Adman, E. T.; Sieker, L. C.; Jensen, L. H. *J. Am. Chem. Soc.* **1988**, 110, 8721-8723.
25. Gregoret, L. M.; Rader, S. D.; Fletterick, R. J.; Cohen, F. E. *Proteins* **1991**, 9, 99-107.

## **Chapter Four: Design, Synthesis and Characterization of the First Cavitein: A De Novo Designed Four-Helix Bundle Templated by a Cavitand Macrocycle**

### **A. Introduction**

As discussed in detail in Chapter One, the protein folding problem can be studied by designing, synthesizing and studying de novo proteins. De novo proteins give researchers the opportunity to isolate and study specific protein-stabilizing interactions; such interactions may not be amenable to study in more complex natural proteins as a result of the multitude of stabilizing and destabilizing effects found in natural systems. This chapter will describe the design, synthesis and characterization of the first cavitein (*cavitand* + *protein*): a de novo four-helix bundle protein using a cavitand macrocycle as a template.

Section i of this chapter will describe various features in the design of the first cavitein using many of the design principles already mentioned in Chapter One. Section ii will use results from both Chapter Two (cavitand synthesis) and Chapter Three (testing a synthetic methodology for de novo protein synthesis) to describe the synthesis of the first cavitein. And finally, Section iii of this chapter will provide a structural analysis of the first cavitein using analytical ultracentrifugation as well as circular dichroism, NMR and fluorescence spectroscopies.

## B. Results and Discussion

### i. Design of Cavitein 27

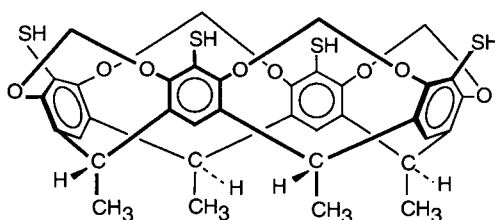
There are two main aspects to the design of the first cavitein: (1) choice of the template and (2) choice of the peptide sequence.

#### a. Tetrathiol Cavitands

From Chapter One, “template assembly” can be a useful tool in de novo protein design. By covalently linking individual peptide strands to a template, large entropic barriers to folding (e.g., bringing flexible peptide strands close in space) can be bypassed, thus giving rise to template-stabilized protein-like structures. An effective template should explicitly direct peptide strands into a desired folding pattern and be simple to incorporate into the design.

Cavitands bode well for use in the study of template assembled four-helix bundles. As mentioned in Chapter Two, cavitands are extremely rigid molecules and hence their incorporation into a de novo protein is expected limit the degrees of freedom afforded to the constituent peptides and thus have a dramatic influence over the resulting structure. In addition, adjacent rim functionalities on cavitands are  $\sim 7 \text{ \AA}$  apart<sup>1</sup> and are thus well-spaced for positioning helices for a four-helix bundle where inter-helix distances range between  $7 - 14 \text{ \AA}$ .<sup>2</sup> This feature allows for a peptide-cavitand linkage of minimal length, which will in turn take full advantage of the cavitand’s rigidity to influence and, ideally, stabilize the peptide structure. In contrast, previously designed templates<sup>3</sup> for use with four-helix bundles have been limited by

their inherent flexibility (e.g., Mutter's cyclic peptides) and/or their long and flexible peptide-template linkages (e.g., Ghadiri's metal templates) which diminish the effect of the template on the resulting structure.



Cavitand 3

More specifically, tetrathiol cavitand **3** is well-suited for de novo protein synthesis because it possesses nucleophilic thiol functional groups that have been shown to react efficiently with various activated phenylalanine derivatives (Chapter Three). Coupling of four suitably designed and activated peptides with tetrathiol **3** was envisioned to proceed selectively and efficiently to afford a cavitand-based four-helix bundle.

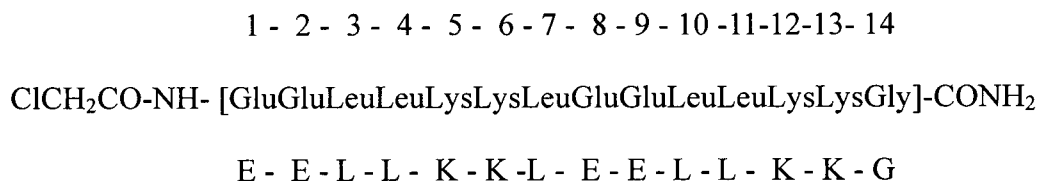
One interesting feature of all cavitands is that they possess an enforced cavity which has been shown to bind small organic molecules.<sup>4</sup> Eventually, this hydrophobic cavity could be used as a binding site en route to enzyme-like caviteins with catalytic activity.



b. Peptide Sequence

The peptide sequence was designed using many of the design features described in Section E.i.a of Chapter One. Firstly, a minimalistic approach<sup>5</sup> was taken in that a minimum number of different residues were included in the design (apart from glycine, only three different types of amino acids). This was done to simplify interpretation of the results and allow for more meaningful designs in the future.

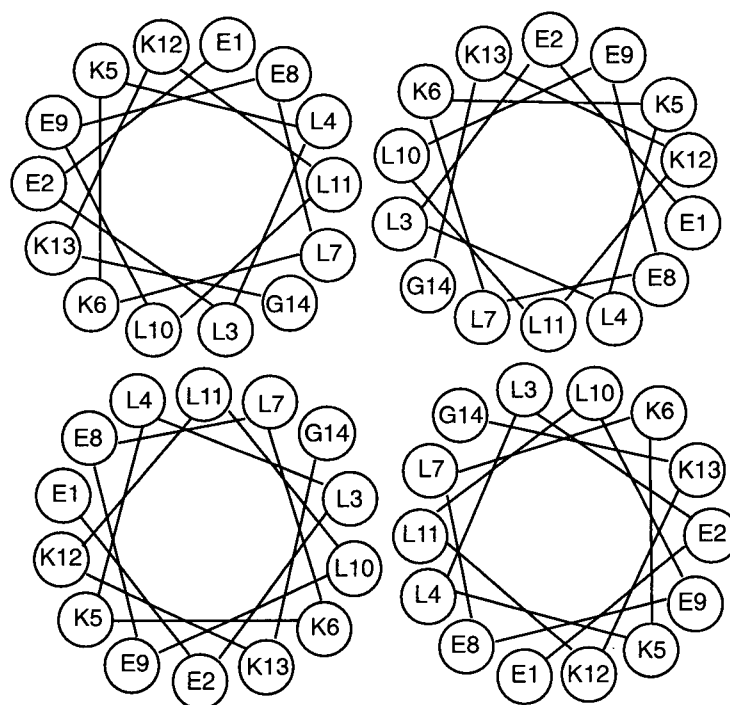
**Figure 4.1.** Primary Sequence of Peptide **26** Used in the Design of Cavitein **27**.



The primary sequence of the designed peptide (**26**) to be coupled with cavitand **3** is shown in Figure 4.1. Peptide **26** was designed to be amphiphilic such that when folded into an  $\alpha$ -helix, the peptide would possess a hydrophobic and hydrophilic face. It is known that  $\alpha$ -helices are composed of 3.6 residues per turn and thus placement of various residues three and four apart in the amino acid sequence will position those residues above one another on the same side of the helix. Since the cavitein will be studied in an aqueous environment, hydrophobic bundling of the helices was expected to be the primary driving force for folding. The amphiphilicity of the peptides is best illustrated with a helical wheel diagram as shown in

Figure 4.2. It is evident that once folded into an  $\alpha$ -helical bundle, hydrophobic Leu residues are sequestered within the core of the bundle while hydrophilic Glu and Lys residues remain on the exterior of the bundle thus aiding in water-solubility.

**Figure 4.2.** Helical Wheel Diagram of Four Molecules of Peptide **26** with Their Helices Oriented in Parallel and Where the Reader is Looking Down the Helical Axes From C- to N-Termini.



The sequence was also designed to be fourteen amino acids in length in order obtain a relatively simple  $\alpha$ -helical bundle yet remain well above the minimum chain length required for  $\alpha$ -helix formation ( $\sim 8$ -12 residues).<sup>6</sup> Apart from glycine, only amino acids with high  $\alpha$ -helical propensities (Leu, Glu, Lys) were used in the design. Glu and Lys residues were placed three and four apart in the sequence to allow for possible stabilizing inter- and intra-helical salt

bridges. The C-terminus was amidated to minimize both charge-charge repulsions between  $\alpha$ -helices and, as a result of the helix macrodipole, charge-dipole repulsions within each  $\alpha$ -helix. Glu and Lys residues were incorporated close to the N- and C-termini, respectively, to interact favorably with the helix macrodipole. In addition, glycine was incorporated as a C-cap since it has been found to occur frequently at the ends of  $\alpha$ -helices found in nature.<sup>7</sup> It is thought that the inherent flexibility of Gly may allow it to fold back onto the  $\alpha$ -helix and hydrogen bond to unsatisfied hydrogen bonding sites.

And lastly, from a synthetic standpoint, the N-terminus of the peptide **26** was activated with a chloroacetyl moiety for coupling with tetrathiol **3** in accord with the synthetic strategy modeled in Chapter Three.

In summary, all of the above design features were expected to give rise to a peptide capable of reacting with tetrathiol **3** and forming a stable four-helix bundle. Note that the design and synthesis of peptide **26** and cavitein **27** have been published.<sup>8</sup>

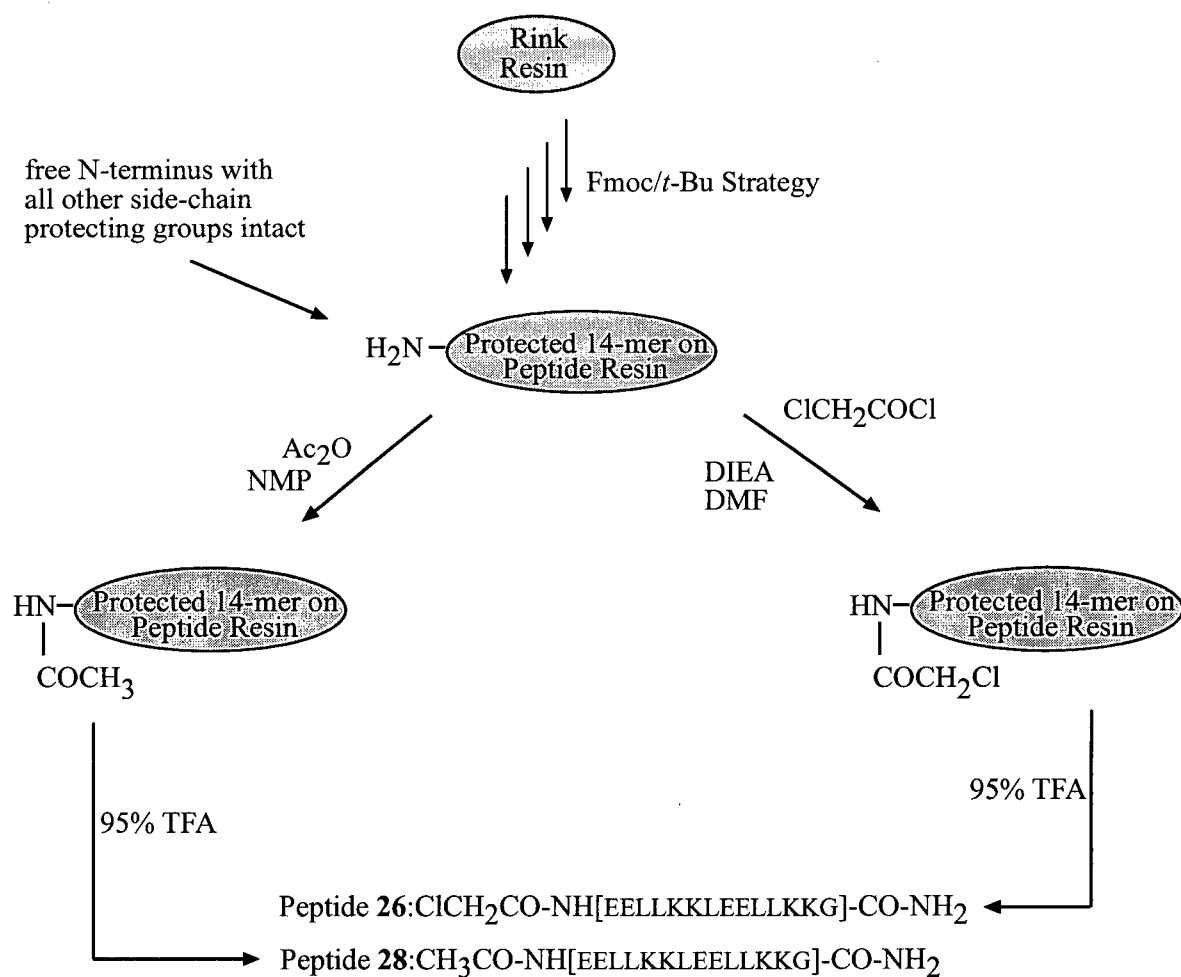
## ii. Synthesis of Cavitein **27**

The synthesis of cavitein **27** involves the coupling of two synthetically complex starting materials: activated peptide **26** and cavitand **3**. The synthesis of cavitand **3** is described in Chapter Two, Section B.ii. The synthesis of peptide **26** is described below.

Peptide **26** was synthesized using standard solid-phase techniques using an automated peptide synthesizer (see Experimental Section). In accord with the synthetic strategies discussed in Chapter Three, the last cycle in the synthesis entailed coupling of chloroacetyl

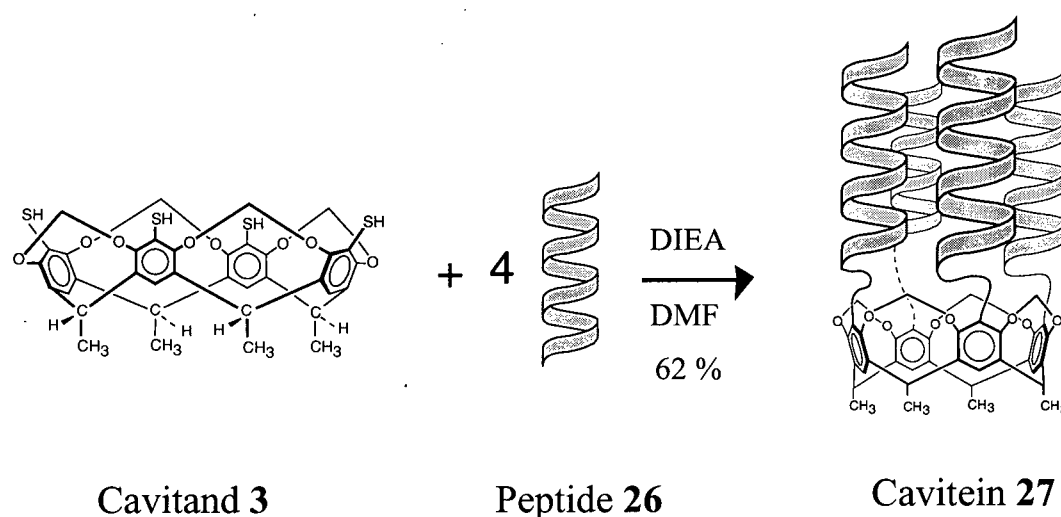
chloride with the free N-terminus of the peptide to afford resin-bound, side-chain-protected, activated analogue of peptide **26**. Subsequent treatment with 95% trifluoroacetic acid (TFA) cleaved the peptide from the resin and removed the side-chain protecting groups in one step to afford peptide **26** (Figure 4.3). A resin developed by Rink was used to afford a C-terminal amide upon cleavage from the resin.<sup>9</sup>

**Figure 4.3.** Schematic Representation of the Syntheses of Peptides **26** and **28**.



Single-stranded peptide **28** was also synthesized for some structural and stability comparisons. This fourteen residue peptide was synthesized as described for peptide **26** but the last cycle entailed acetylation of the N-terminus with acetic anhydride. Cleavage from the resin with TFA afforded single-stranded peptide **28** (Figure 4.3).

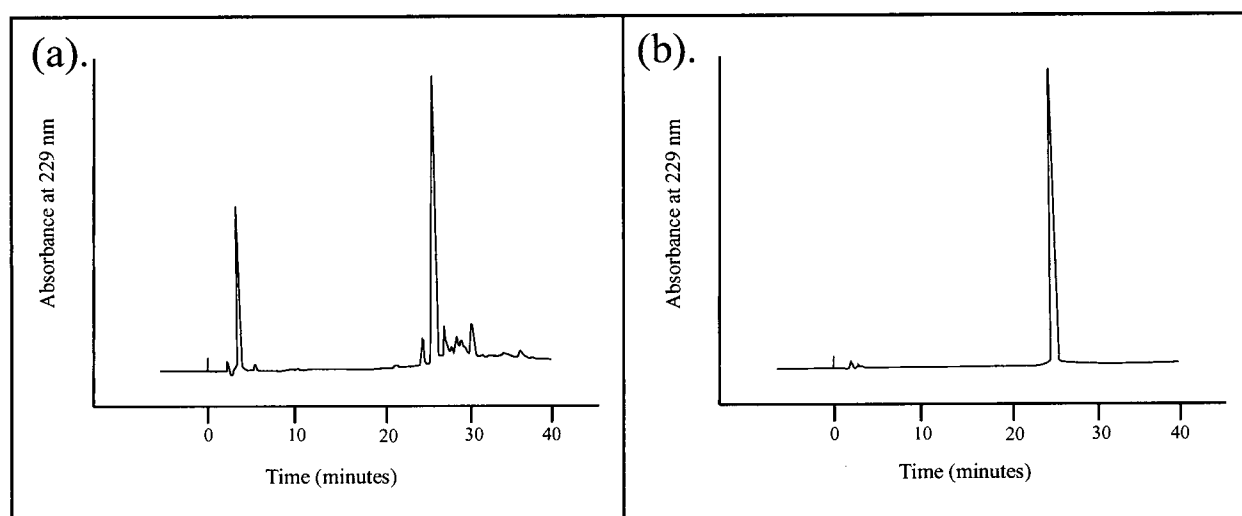
**Scheme 4.1.** Synthesis of Cavitein **27**. The Linkage Between the Cavitand and the Peptides is “cavitand-(S-CH<sub>2</sub>CO-NH-peptide)<sub>4</sub>”.



Cavitein **27** was synthesized by reacting 4.4 equivalents of unprotected peptide **26** with one equivalent of tetrathiol cavitand **3** in the presence of diisopropylethylamine (DIEA) in dimethylformamide (DMF) for 16 h at 25 °C. The reaction proceeded very efficiently as monitored by reversed-phase HPLC (Figure 4.4) to afford cavitein **27** in 62% yield. Evidently, the synthetic strategy described in Chapter Three works well for chemoselectively linking unprotected peptide **26** to tetrathiol **3**. The purity of cavitein **27** was assessed by analytical

reversed-phase HPLC (one peak) and its identity was confirmed by amino acid analysis and electrospray mass spectrometry (molecular mass = 7557 Da).

**Figure 4.4.** The Synthesis of Cavitein **27** as Monitored by Analytical Reversed-Phase (C<sub>18</sub>) HPLC Using a Gradient of 35 - 75% Acetonitrile (with 0.05% TFA) in Water (with 0.1% TFA) Over 40 Minutes: (a) Crude Reaction Mixture After 16 h; (b) Purified Cavitein **27**.



### iii. Structural Characterization of Cavitein **27**

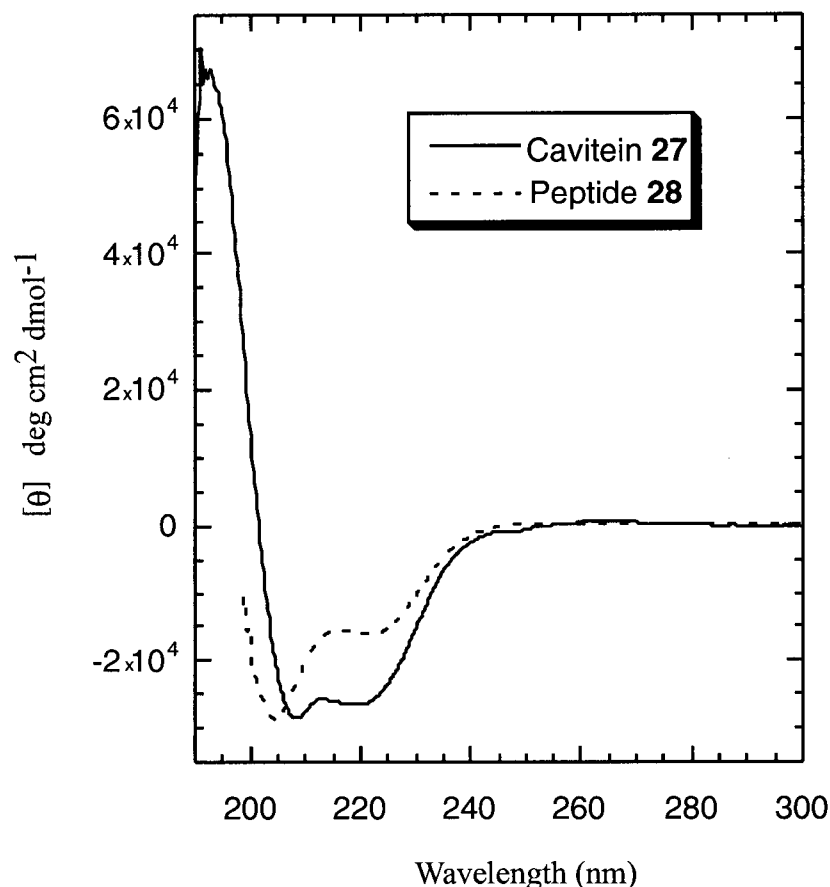
The structures of helical bundles can be characterized in many different ways. We used analytical ultracentrifugation, circular dichroism (CD) spectroscopy, one-dimensional (1D) <sup>1</sup>H NMR spectroscopy and the binding of a hydrophobic dye monitored by fluorescence spectroscopy to assess the structure of cavitein **27**. Analysis of single-stranded analogue peptide **28** was not explored in detail since peptides similar to **28** have been studied extensively.<sup>10</sup>

Nevertheless, some elements of the structure and stability of **28** (e.g., GuHCl denaturations) were studied for comparative purposes.

a. Far-UV CD Spectrum of Cavitein **27**

Circular dichroism (CD) spectroscopy in the far-UV region can be used to quantify amounts of secondary structure present in peptide and protein structures.<sup>11</sup> In particular, CD spectroscopy is especially useful in studying  $\alpha$ -helical structure since two negative bands near 222 and 208 nm and one positive band near 195 nm are typically observed for  $\alpha$ -helices. The CD spectrum of cavitein **27** showed minima at 222 nm and 208 nm and a maximum at 195 nm which is characteristic of  $\alpha$ -helical structure (Figure 4.5). Typically, the molar ellipticity at 222 nm ( $[\theta]_{222}$ ) is used as an indicator of  $\alpha$ -helical structure. For cavitein **27** at pH 7.0 and 25 °C,  $[\theta]_{222}$  was found to be -24 000 deg cm<sup>2</sup> dmol<sup>-1</sup>. An estimate of a maximum value for  $[\theta]_{222}$  was calculated to be -32 000 deg cm<sup>2</sup> dmol<sup>-1</sup> using the chain-length-dependent equation developed by Chen et. al.<sup>12</sup> Hence, the percentage of residues in an  $\alpha$ -helical conformation was estimated to be ~75% and thus the amino acids comprising cavitein **27** are highly  $\alpha$ -helical (Note: this value of  $[\theta]_{222}$  does not take into account possible aromatic contributions to the far-UV region by the cavitand template. This effect will be discussed further in Chapter Five). The CD spectrum of single-stranded peptide **28** was found to be much less helical than those of cavitein **27** at 5  $\mu$ M (Figure 4.5). The effect of the cavitand template is already apparent: the cavitand template increases  $\alpha$ -helical structure relative to the single-stranded peptide at low concentrations.

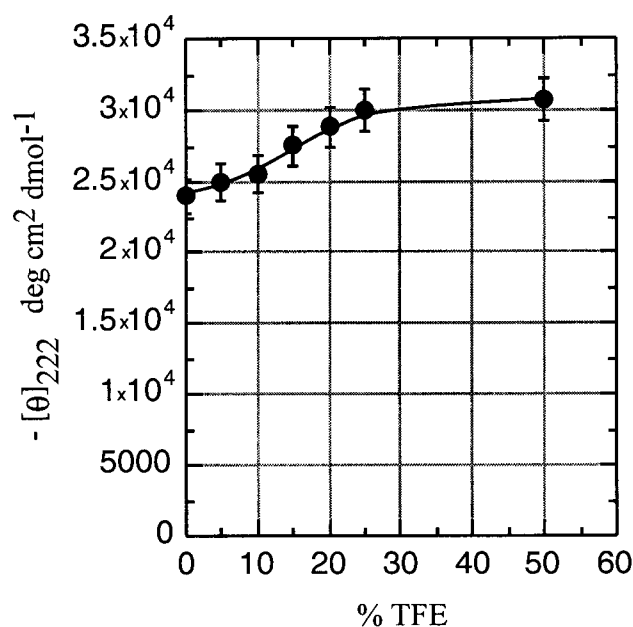
**Figure 4.5.** CD Spectra of Separate 5  $\mu$ M Solutions of Cavitein **27** (Solid Line) and Peptide **28** (Dashed Line) in pH 7.0 Phosphate Buffer at 25  $^{\circ}$ C.



Trifluoroethanol (TFE) can be used to induce  $\alpha$ -helical conformations<sup>13</sup> although its mechanism of action is not yet clear.<sup>14</sup> With respect to  $\alpha$ -helical bundles, TFE is known to disrupt bundling and promote TFE-stabilized isolated  $\alpha$ -helices.<sup>15</sup> Hence, titrations of TFE into solutions of partially helical peptides can provide a qualitative estimate of a maximal value for their helicity. Addition of TFE to cavitein **27** showed only a  $\sim 20\%$  increase in helicity as determined by monitoring  $[\theta]_{222}$  (Figure 4.6). This result reaffirms the conclusion that the amino acids in cavitein **27** are highly helical.



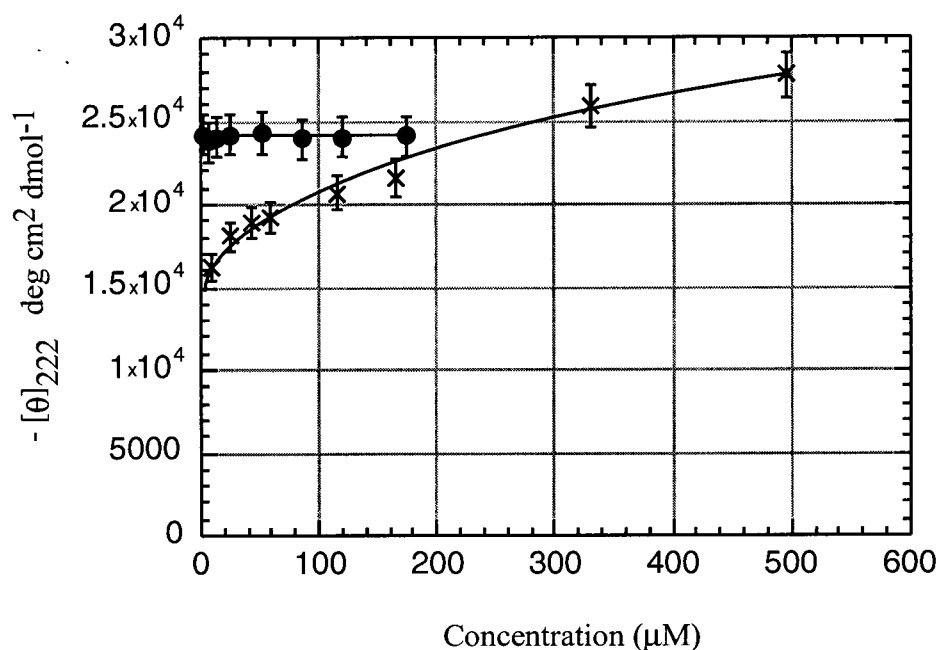
**Figure 4.6.** Effect of TFE on the Helicity ( $[\theta]_{222}$ ) of Cavitein **27** in pH 7.0 Phosphate Buffer at 25 °C.



b. Oligomeric State of Cavitein **27**

Cavitein **27** was designed to sequester the hydrophobic residues of its constituent peptides within the core of its four-helix bundle while relegating its hydrophilic residues to the exterior of the bundle to afford a monomeric unit (Figure 4.2). In contrast, single-stranded peptide **28** possesses exposed hydrophobic faces and is thus expected to form an aggregate. The oligomeric states of cavitein **27** and peptide **28** were studied by monitoring the concentration dependence of their CD spectra. Figure 4.7 shows a plot of  $[\theta]_{222}$  vs. different concentrations for both cavitein **27** and peptide **28**.

**Figure 4.7.** Concentration Dependence of  $[\theta]_{222}$  for Cavitein **27** (Circles) and Peptide **28** (X's) in pH 7.0 Phosphate Buffer at 25 °C.



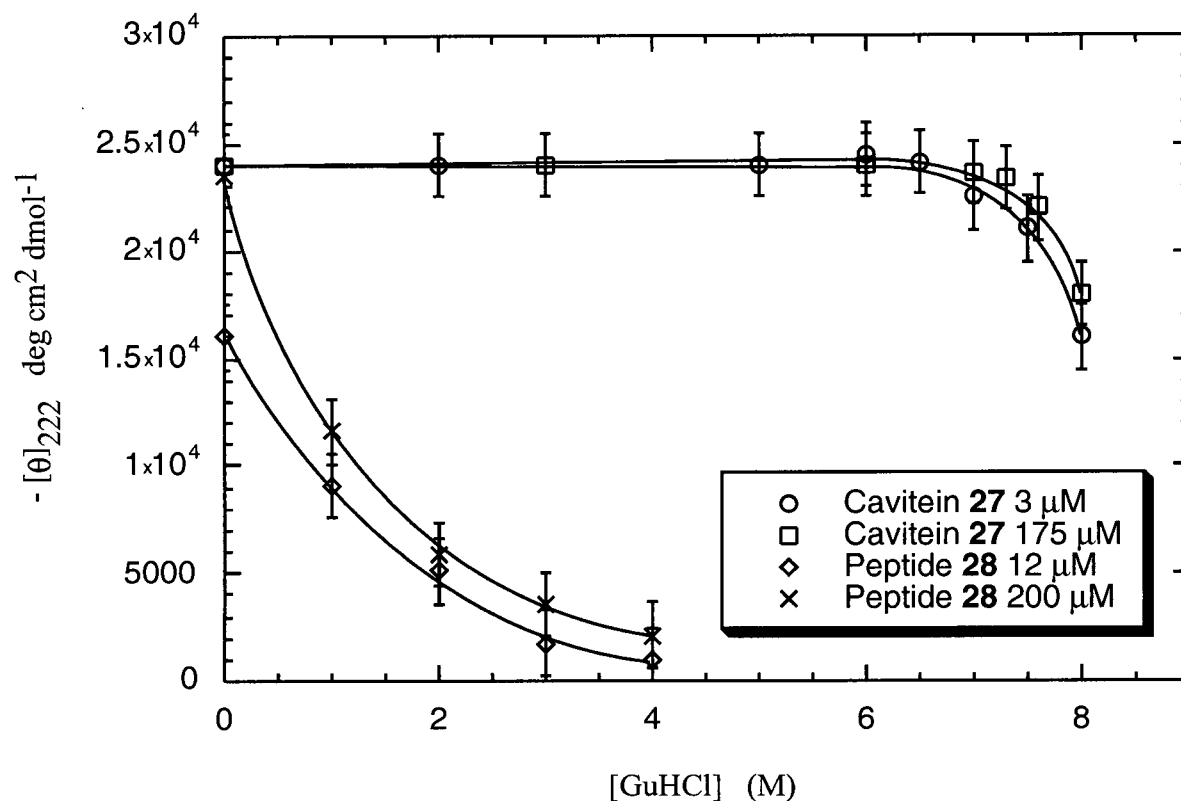
The CD spectra of cavitein **27** were concentration independent from 1 – 175 μM, which is consistent with the presence of a monomeric species in aqueous solution under these conditions. However, this argument relies on the assumption that self-association of cavitein **27** would be accompanied by an increase in helicity. Indeed, for peptides that are marginally helical as monomers, self-association of individual helices typically results in an increase in helicity. This behavior was observed for single-stranded peptide **28**:  $[\theta]_{222}$  increased with increasing peptide concentration (12 – 200 μM) which suggests that peptide **28** exists as a concentration-dependent aggregate. Note that at high concentrations of peptide **28**,  $[\theta]_{222}$  is greater than that of cavitein **27**. This suggests that the template may prevent the peptides from

adopting a fully  $\alpha$ -helical structure (as inferred previously with the calculation of ~75% helicity).

However, for helical bundles that are already highly helical and near their maximum helicity, reversible self-association may *not* be accompanied by a detectable increase in helicity. Thus, the observation of concentration independent CD spectra for cavitein **27** does not necessarily prove its monomeric state.

Further evidence can be gathered by conducting denaturation experiments at low and high concentrations of protein. The stabilities of oligomeric proteins are expected to vary with protein concentration, in contrast to monomeric proteins which should give rise to concentration independent stabilities and hence identical denaturation curves. Underlying this argument is the assumption that self-association would be accompanied by an increase in structural stability. We thus conducted denaturation experiments using the chemical denaturant guanidine hydrochloride (GuHCl) at high and low concentrations of peptide **28** and cavitein **27** (Figure 4.8, also see Section B.iii.c for further discussion of this data).

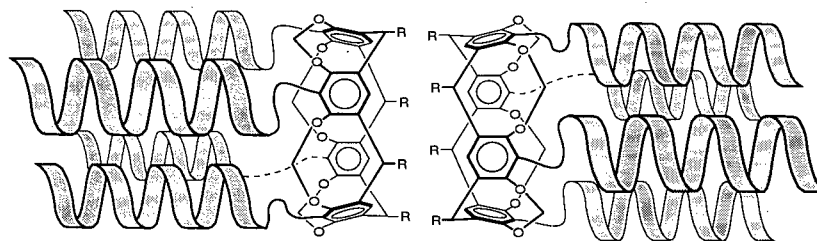
**Figure 4.8.** Effect of GuHCl on the Helicity ( $[\theta]_{222}$ ) of Cavitein **27** at 3  $\mu\text{M}$  (Circles) and 175  $\mu\text{M}$  (Squares) and Peptide **28** at 12  $\mu\text{M}$  (Diamonds) and 200  $\mu\text{M}$  (X's) in pH 7.0 Phosphate Buffer at 25 °C.



Peptide **28** displayed different stabilities at different concentrations as expected by aggregation of its hydrophobic faces to form an oligomer. In contrast, the stability of cavitein **27** varied only slightly from 3 to 175  $\mu\text{M}$  – the difference between the two curves lying within the error of the experiment. These results suggest cavitein **27** is monomeric under the conditions described; however, since a full unfolding transition was not observed, it is difficult to interpret the significance of the difference between the two curves.

Although no substantial gain in stability was observed at 175  $\mu\text{M}$ , it still remains possible that self-association could occur *without* adding significant stability to the structure. In other words, it is possible that cavitein **27** could fold into a four-helix bundle, but self-associate *without* altering the structure of the helical bundle. One example of this may be hydrophobic association via the cavitand template (Figure 4.9).

**Figure 4.9.** Hypothetical Representation of Self-Associating Caviteins Without Altering the Structure or Stability of the Helical Bundles.



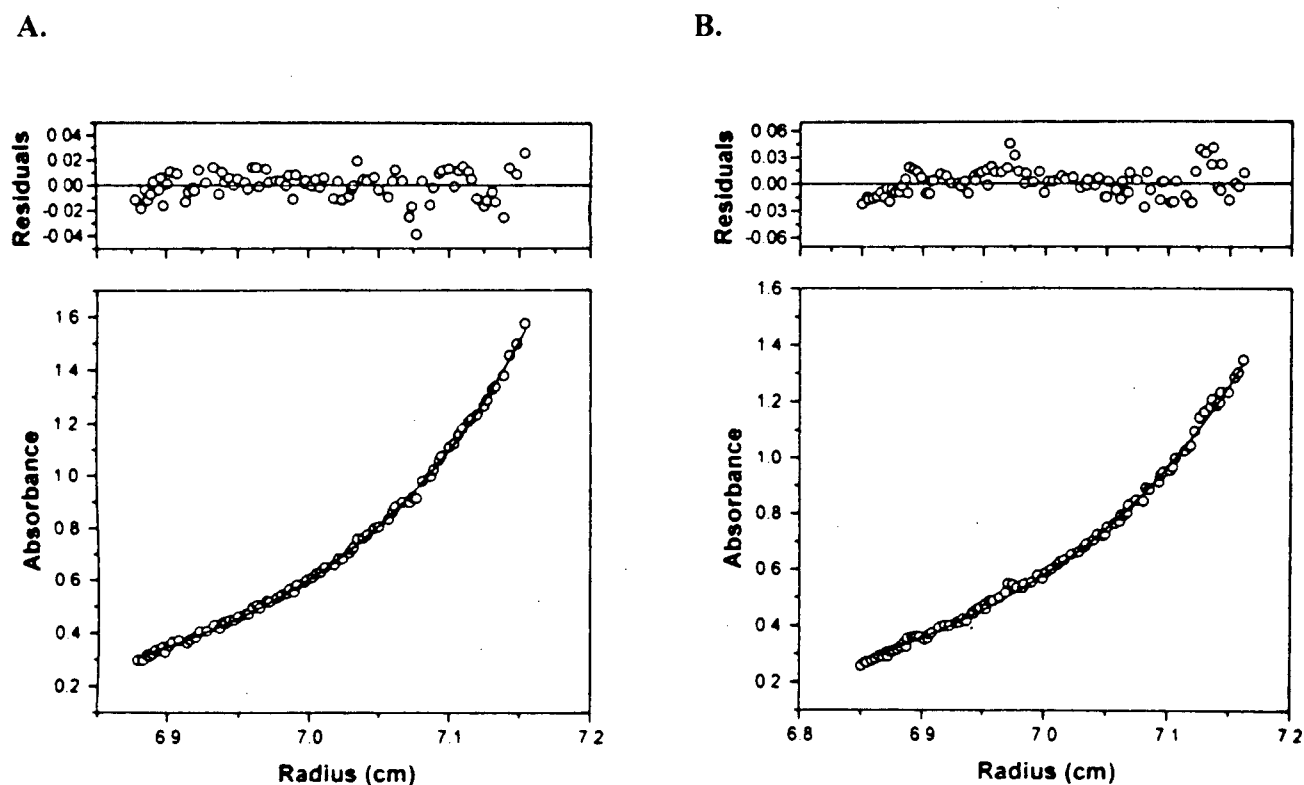
In light of the ambiguities associated with the slight change in GuHCl denaturation curves for cavitein **27**, we decided to further study its oligomeric state. Sedimentation equilibrium experiments using an analytical ultracentrifuge can be used to estimate the molecular weight of macromolecules in solution.<sup>16</sup> Briefly, if the centrifugal force imposed on a macromolecule by centrifugation is small enough to allow for diffusion to oppose sedimentation, an equilibrium concentration distribution of that macromolecule will be obtained throughout the cell. For ideal, non-interacting single species, this distribution can be described by an exponential function of its molecular mass (see Experimental Section). In addition, the

concentration distribution can be fit to equations describing self-association systems such as monomer-dimer, monomer-trimer, monomer-dimer-trimer, etc., in order to identify possible aggregation. Modern analytical ultracentrifuges are equipped with data-analyzing software such that the acquired data can be described by any of the equations described above to obtain the best fit.

The University of British Columbia is not equipped with a modern analytical ultracentrifuge. As a result, all of the ultracentrifugation data presented in this thesis were acquired elsewhere; namely at the Scripps Research Institute in the research group of Jeff Kelly. In addition to running the experiments, several Kelly group members aided in the interpretation of the data.

The sedimentation equilibrium behaviour of cavitein **27** was studied at both 30 and 300  $\mu\text{M}$  at 25  $^{\circ}\text{C}$  in pH 7.0 phosphate buffer (Figure 4.10) which corresponds to the same conditions used in the CD experiments mentioned above. It was found that the observed concentration distribution was best described as a monomer-dimer equilibrium mixture at both concentrations.

**Figure 4.10.** Sedimentation Equilibrium Concentration Distributions of Cavitein **27** (Circles) at 30 000 rpm in 50 mM Phosphate Buffer, pH 7, 25 °C at (A) 30  $\mu$ M and (B) 300  $\mu$ M. In the Lower Panels, The Solid Lines Represent Theoretical Fits to a Monomer-Dimer Equilibrium. The Upper Panels in Each Figure Represent the Residuals for the Fit.



These results came as a surprise considering the CD data which was consistent with the presence of a monomeric species. As proposed above, it is quite likely that cavitein **27** is near its maximum helicity and thus self-association does not alter its helicity. From the near-concentration independent GuHCl stability data, it is likely that the dimerization of cavitein **27** does not result in a change in the structure or stability of the four-helix bundle. The nature of this dimerization is explored in the next few sections.

In a general sense, the ultracentrifugation results demonstrate that concentration independent helicity and stability are *not necessarily* indicators of monomeric species. Other complementary techniques are required to firmly establish the monomeric nature of these caviteins.

It should be noted that gel filtration chromatography also was attempted<sup>17</sup> on four different caviteins presented in this thesis (including cavitein **27**). The results were inconclusive<sup>18</sup> and as a result, this technique for assessing the oligomeric state of the caviteins was abandoned in favour of sedimentation equilibrium. In addition, characterization of all the caviteins presented in this thesis by electrospray mass spectrometry showed no evidence of oligomerization.<sup>19</sup>

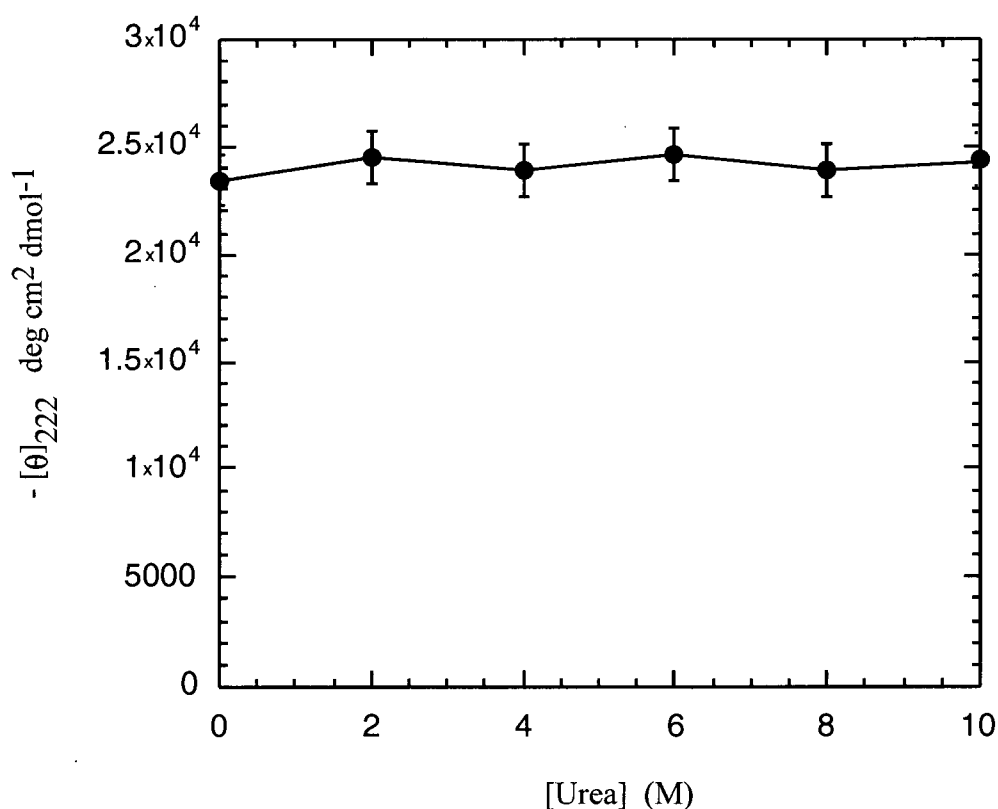
#### c. Effect of Guanidine Hydrochloride and Urea

The stability of cavitein **27** was studied using the chemical denaturants guanidine hydrochloride (GuHCl), as mentined above, and urea.<sup>20</sup> Although the exact mode of action is still not known, strong evidence exists that both GuHCl and urea preferentially bind to protein surfaces to disrupt both hydrogen bonding interactions and hydrophobic packing.<sup>21</sup> However, GuHCl and urea are known to give slightly different estimates of protein stability.<sup>22</sup> Since GuHCl is a salt, it is able to screen electrostatic interactions at low concentrations of GuHCl as well as hydrogen bonding and hydrophobic packing interactions. In general, urea does not affect electrostatic interactions but can still interfere with both hydrophobic packing and hydrogen bonding interactions, as mentioned above.<sup>23</sup> As a result, GuHCl is typically considered a more effective denaturant.<sup>24</sup>



Upon treating cavitein **27** with urea at 25 °C and pH 7.0, its helicity changed less than 5% on going from 0 – 10 M urea (Figure 4.11), thereby illustrating the extreme stability of cavitein **27**. Since no unfolding transition was observed, urea was not used in any further denaturation experiments in this thesis.

**Figure 4.11.** Effect of Urea on the Helicity ( $[\theta]_{222}$ ) of Cavitein **27** in pH 7.0 Phosphate Buffer at 25 °C.



Cavitein **27** also displayed extreme stability towards GuHCl (Figure 4.8). At 25 °C and pH 7.0, GuHCl was unable to fully denature the cavitein even at 8 M GuHCl (the approximate solubility limit of GuHCl). Helicity, as measured by monitoring  $[\theta]_{222}$ , decreased by only ~

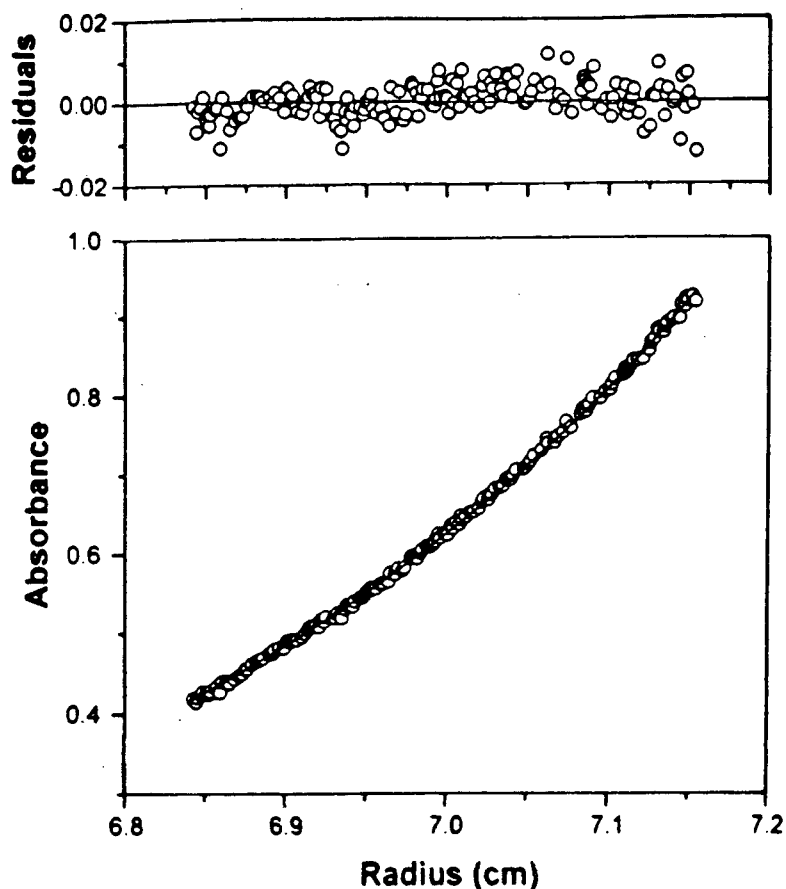
30% in 8 M GuHCl at both 3 and 175  $\mu$ M cavitein **27**. In contrast, single-stranded peptide **28** was half-unfolded at  $\sim 1$  M GuHCl at both 12 and 200  $\mu$ M despite the enhanced helicity observed for peptide **28** at 200  $\mu$ M (Figure 4.8). In comparison to peptide **28**, the protein moiety of cavitein **27** is greatly stabilized by the cavitand template.

Consideration of the greater effectiveness of GuHCl over urea towards unfolding suggests that there may be a considerable electrostatic component to the stability of cavitein **27**. Recall that since GuHCl is a salt, this denaturant is more able to screen potential stabilizing electrostatic interactions (e.g., salt-bridges) and therefore destabilize the structure. Since urea is unable to screen such interactions as effectively, electrostatics may play an important role in the stability of cavitein **27**. Such electrostatic effects are explored in more detail in the next section.

One question that remains is how does the monomer-dimer equilibrium of cavitein **27** affect its structure and stability? Or, put another way, are four-helix bundles being formed? These questions were answered by performing a sedimentation equilibrium experiment in the presence of 6.0 M GuHCl. At this concentration of denaturant, cavitein **27** is still fully helical as observed by CD spectroscopy. However, the concentration distribution of cavitein **27** in the sedimentation equilibrium experiment (Figure 4.12) was best described by a single non-interacting ideal species of molecular weight  $12\,850 \pm 50$  Da (molecular weight of monomer is 7557 Da). Although this molecular weight is closer to that of a dimer than that of a monomer, there exists strong evidence that suggests that this single non-interacting species is predominantly monomeric: (1) GuHCl is a denaturant and is thus expected to disrupt structural aggregates, not enhance them, and (2) experiments described in the next chapter will demonstrate that a similar cavitein which is fully unfolded in the presence of 7.2 M GuHCl has an estimated mass of near 11 000 Da or about 4 000 Da higher than expected. This 11 kDa

molecular weight most likely represents a monomer. Indeed there are significant errors associated with this technique of molecular weight estimation, but what are its possible sources?

**Figure 4.12.** Sedimentation Equilibrium Analysis of Cavitein **27** in the presence of 6.0 M GuHCl at 25 °C. In the Lower Panel, The Solid Line Represents the Theoretical Fit to a Single Non-Interacting Species. The Upper Panel Represents the Residuals for the Fit.



The solution molecular weight of any given protein as determined by sedimentation equilibrium is, among other constants, proportional to the solution density ( $\rho$ ) and the partial specific volume ( $\bar{v}$ ) of the solute (see Experimental Section). Accurate estimations of each parameter is essential for an accurate molecular weight estimation. Solution densities may be

estimated quite well from the buffer composition.<sup>16a</sup> However, estimation of  $\bar{v}$  is much more difficult and small errors in  $\bar{v}$  manifest themselves into large errors in the molecular weight.<sup>16a</sup> The partial specific volume of a protein may be defined as the change in volume of the solution per gram of solute. This value is essentially the partial volume of the protein corrected for hydration, solute binding and electrostriction effects.<sup>25</sup> Typical estimates of  $\bar{v}$  are derived from the amino acid composition of the protein. However, significant errors in this method may arise from: (1) proteins that exhibit preferential hydration, (2) proteins that have a non-globular structure and (3) proteins with significant non-amino acid components to its composition. With respect to the caviteins, the cavitand component is not accounted for and may contribute significantly to the observed error. Moreover, in the case of solutions containing GuHCl, preferential binding of GuHCl to a protein will affect the partial specific volume and these effects are sometimes difficult to predict.<sup>26</sup> This is especially the case with proteins containing substantial non-amino acid constituents as in caviteins where the cavitand component comprises ~ 10% of the cavitein's molecular weight. The effect of GuHCl on the cavitand component of cavitein **27** is not known. One last possible source of error arises from the possibility that up to 10% of an aggregate may go undetected when analyzed by single-species analysis, despite giving rise to the best fit.<sup>27</sup> The presence of a small amount of higher order aggregates would effectively skew the average molecular weight to higher values.

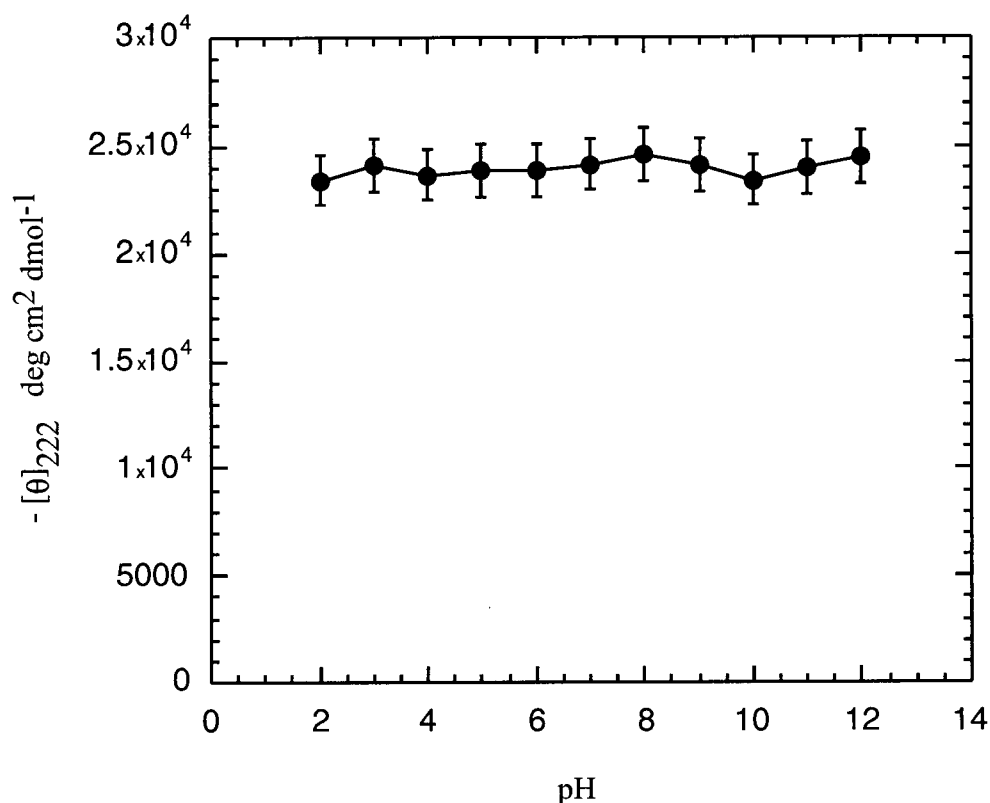
Despite all of the above-mentioned errors associated with sedimentation equilibrium, useful information can still be gleaned from the data when taken together with other sedimentation experiments and data acquired from other experimental techniques. With respect to cavitein **27** and as mentioned above, interpretation of the sedimentation equilibrium data suggests that in the presence of 6 M GuHCl, its structure is monomeric. This result is significant because in the presence of 6 M GuHCl, cavitein **27** is still fully helical as monitored

by CD spectroscopy. Therefore, under non-denaturing conditions (e.g., 50 mM phosphate buffer, pH 7), the observed monomer-dimer equilibrium does *not* appear to significantly alter the structure of the cavitein and hence the structural stability towards GuHCl is likely to be the result of a *monomeric four-helix bundle* as designed.

#### d. Effect of Salt and Methanol on Cavitein **27**

Electrostatic contributions to the stability of cavitein **27** were studied by varying both the pH and salt concentration (KCl) of solutions containing cavitein **27**. Extremes of pH may neutralize ionizable side-chains and thus remove potentially stabilizing salt-bridge effects. The addition of KCl may screen electrostatic interactions which may have an effect on stability. The pH of a solution containing cavitein **27** was varied between 2 – 12 and resulted in less than a 5% change in helicity as monitored by  $[\theta]_{222}$  (Figure 4.13). Furthermore, up to 3 M KCl was also added to cavitein **27**, again with less than a 5% change in helicity (data not shown). Together, these results suggests that electrostatic effects are not the major driving force for  $\alpha$ -helical bundle formation/stabilization. Rather, hydrophobic bundling of the leucine residues is most likely the dominant force stabilizing cavitein **27**.

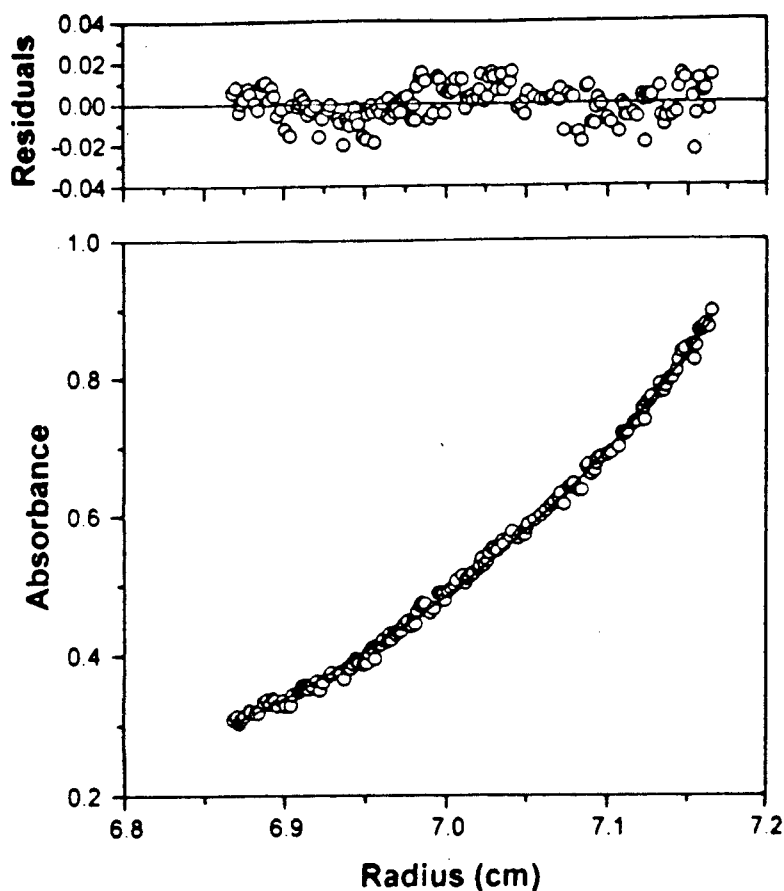
**Figure 4.13.** Effect of pH on the Helicity ( $[\theta]_{222}$ ) of Cavitein **27** at 25 °C.



One unanswered question thus far is the cause of the observed monomer-dimer equilibrium of cavitein **27**. Is there an electrostatic component contributing to the observed dimerization? Unfortunately, we are unable to monitor the monomer-dimer equilibrium with CD spectroscopy and therefore, although  $[\theta]_{222}$  did not change in the above experiments, it is still possible that aggregation was inhibited with the change in pH and added salt. We therefore conducted a sedimentation equilibrium experiment in the presence of 2 M NaCl (Figure 4.14). The observed concentration distribution was best described by a single non-interacting species of molecular weight  $8200 \pm 50$  Da, a value near that of the calculated monomeric molecular

weight of 7557 Da. This result suggests that there is an electrostatic component to the observed aggregation as a result of the multitude of charges on cavitein **27**.

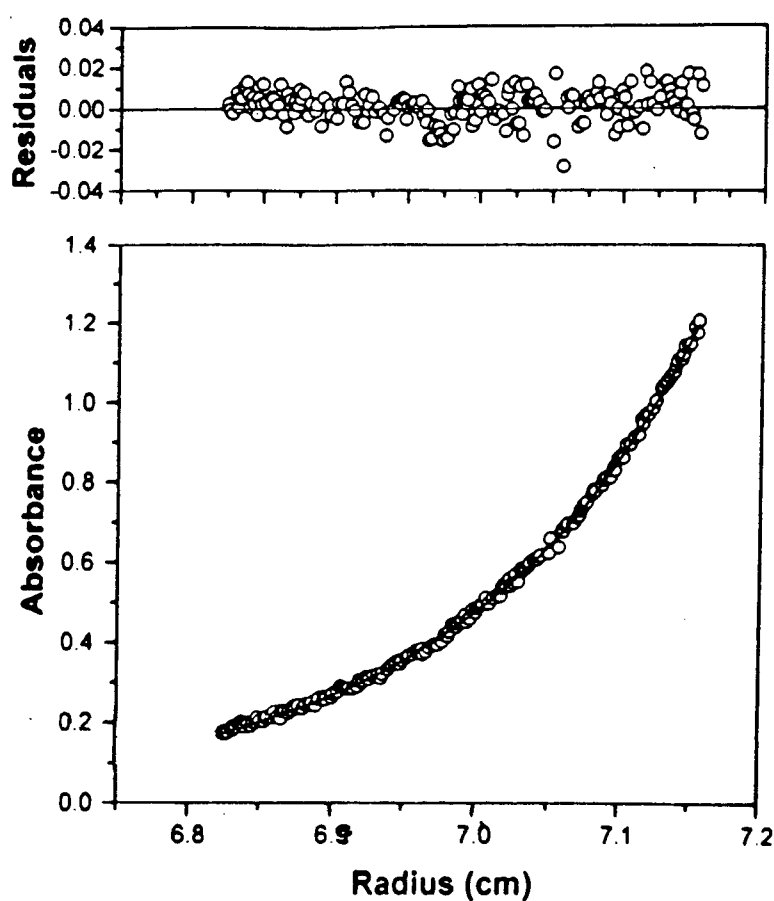
**Figure 4.14.** Sedimentation Equilibrium Analysis of Cavitein **27** in the Presence of 2 M NaCl. In the Lower Panel, The Solid Line Represents the Theoretical Fit to a Single Non-Interacting Species. The Upper Panel Represents the Residuals for the Fit. Deviations in the Residuals are the Result of Window Imperfections (as Interpreted by the Kelly Research Group).



As a test for a possible hydrophobic contribution to the observed aggregation, we also conducted a sedimentation equilibrium experiment in the presence of 10% methanol. Note that in the presence of 10% methanol, the CD spectrum of cavitein **27** had a near-identical shape to that under standard buffer conditions (data not shown). In addition,  $[\theta]_{222}$  changed less than 5%

in the presence of 10% methanol. Together, these results suggest that 10% methanol does not significantly interact with the structure of the four-helix bundle. Interestingly, the concentration distribution of cavitein **27** in the presence of 10% methanol was also best described as a single non-interacting species of molecular weight  $9800 \pm 50$  Da (Figure 4.15). In consideration of the errors mentioned above, this species is likely a monomer.

**Figure 4.15.** Sedimentation Equilibrium Analysis of Cavitein **27** in the Presence of 10% Methanol. In the Lower Panel, The Solid Line Represents the Theoretical Fit to a Single Non-Interacting Species. The Upper Panel Represents the Residuals for the Fit.





The effect of 2 M NaCl and 10% methanol on the monomer-dimer equilibrium suggests that both hydrophobic and electrostatic interactions are responsible for the observed aggregation. Is this a contradiction? Perhaps not. Due to the short cavitand-peptide linker, it is possible that the rigid cavitand skews the helices in some way that does not allow the helices to fully bury their hydrophobic leucine residues within the core of the bundle. This would effectively expose the hydrophobic side-chains to the solvent (water) and promote hydrophobic-induced aggregation. The addition of methanol may be sufficient to inhibit such intermolecular aggregation. An additional consequence of the helical skewing may be that the designed salt bridges are not ideally aligned for intramolecular  $i - i + 4$  interactions and hence render the side-chains available for an intermolecular interaction. Addition of 2 M NaCl may be sufficient to disrupt these intermolecular interactions to provide a monomeric structure.

The fact that the addition of either methanol *or* salt (i.e., both are not required) disrupt the aggregation suggests that both processes are occurring cooperatively and that removal of one of these interactions is sufficient to inhibit the dimerization.

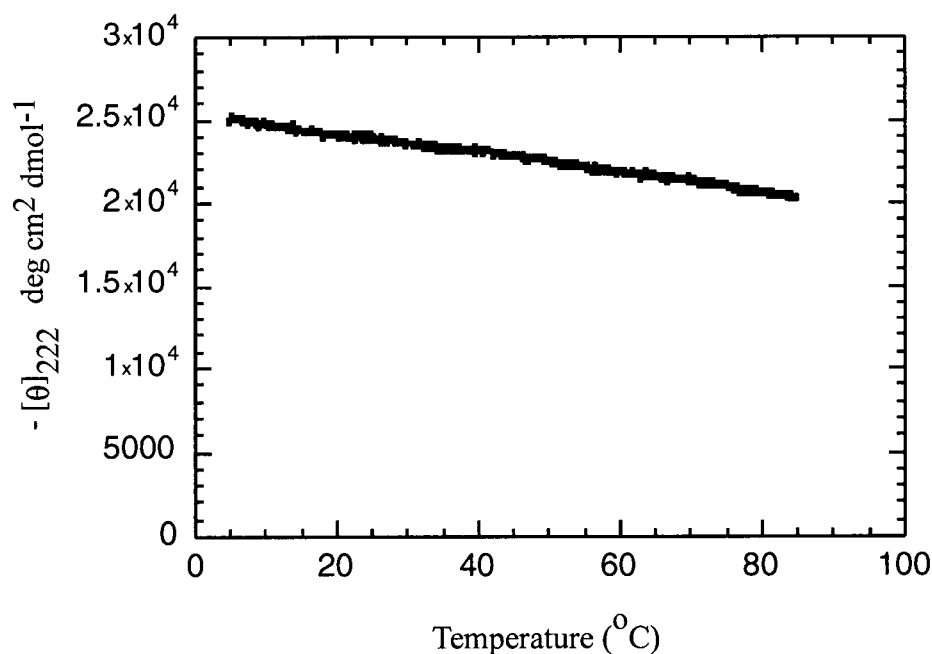
Possible aggregation via the methyl foot of the cavitand is possible based on the results presented in this chapter. However, it is shown in Chapter Five that a monomeric cavitein with an identical core amino sequence is successfully designed with methyl cavitand feet. This effectively rules out aggregation via the hydrophobic cavitand.

#### e. Effect of Temperature on Cavitein **27**

The stability of cavitein **27** was also monitored as a function of temperature. Helicity, as monitored by  $[\theta]_{222}$ , decreased linearly by  $\sim 20\%$  from 5 to 85 °C (Figure 4.16) demonstrating

substantial thermal stability. The experiment was subsequently repeated in the presence of 6 M GuHCl and the data were found to be *superimposable* on those shown in Figure 4.16. Since no unfolding transitions were observed, further analyses were not performed.<sup>28</sup> Since cavitein **27** was found to be monomeric in the presence of 6 M GuHCl, the observed aggregation does not affect its thermal stability which is consistent with aggregation not contributing significantly to the structure or stability of the bundle in general.

**Figure 4.16.** Effect of Temperature on the Helicity ( $[\theta]_{222}$ ) of Cavitein **27** in pH 7.0 Phosphate Buffer.

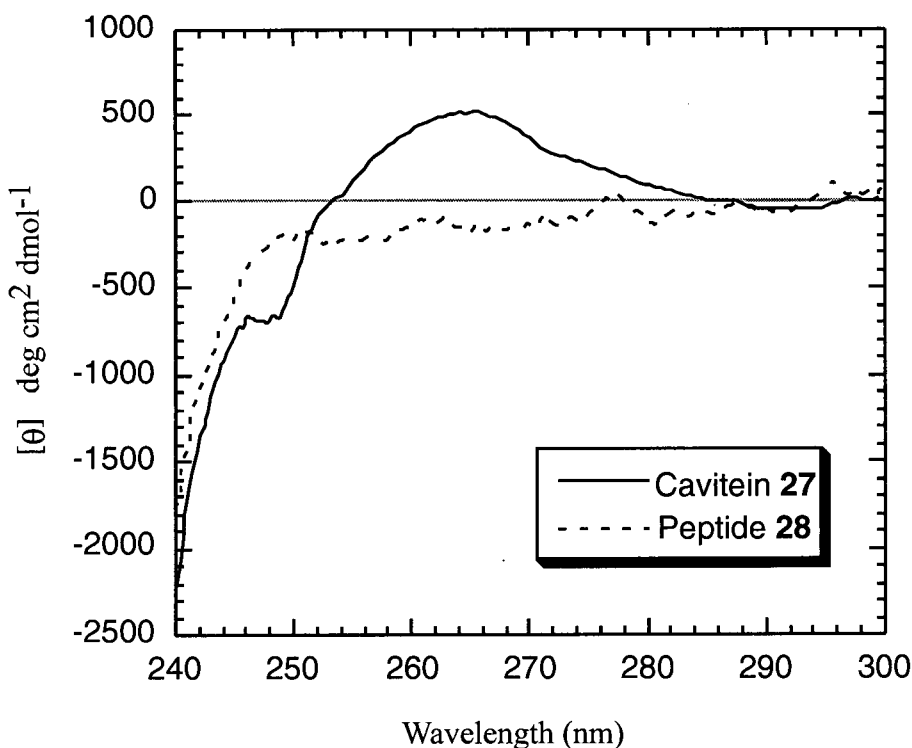


f. Near-UV CD Spectrum of Cavitein **27**

The structural characterization of cavitein **27** thus far has been in reference to helicity or secondary structure. As mentioned in the introductory chapter, molten globules and native

states contain nearly identical amounts of secondary structure but molten globules lack the specific packing interactions required for a unique native state. As a result, molten globules and native structures are indistinguishable by their far-UV spectra. However, the presence of aromatic absorptions in near-UV CD spectra suggest the existence of non-averaged structural elements near that aromatic chromophore.<sup>29,11b-c</sup> Molten globules typically possess a reduction or absence of near-UV absorptions due to the time-averaged fluctuational environment of their aromatic residues.<sup>30</sup> In contrast, native states typically possess strong near-UV absorptions due to the asymmetric environments of these chromophores.

**Figure 4.17.** Near-UV CD Spectra of Cavitein **27** and Peptide **28** in pH 7.0 Phosphate Buffer at 25 °C.

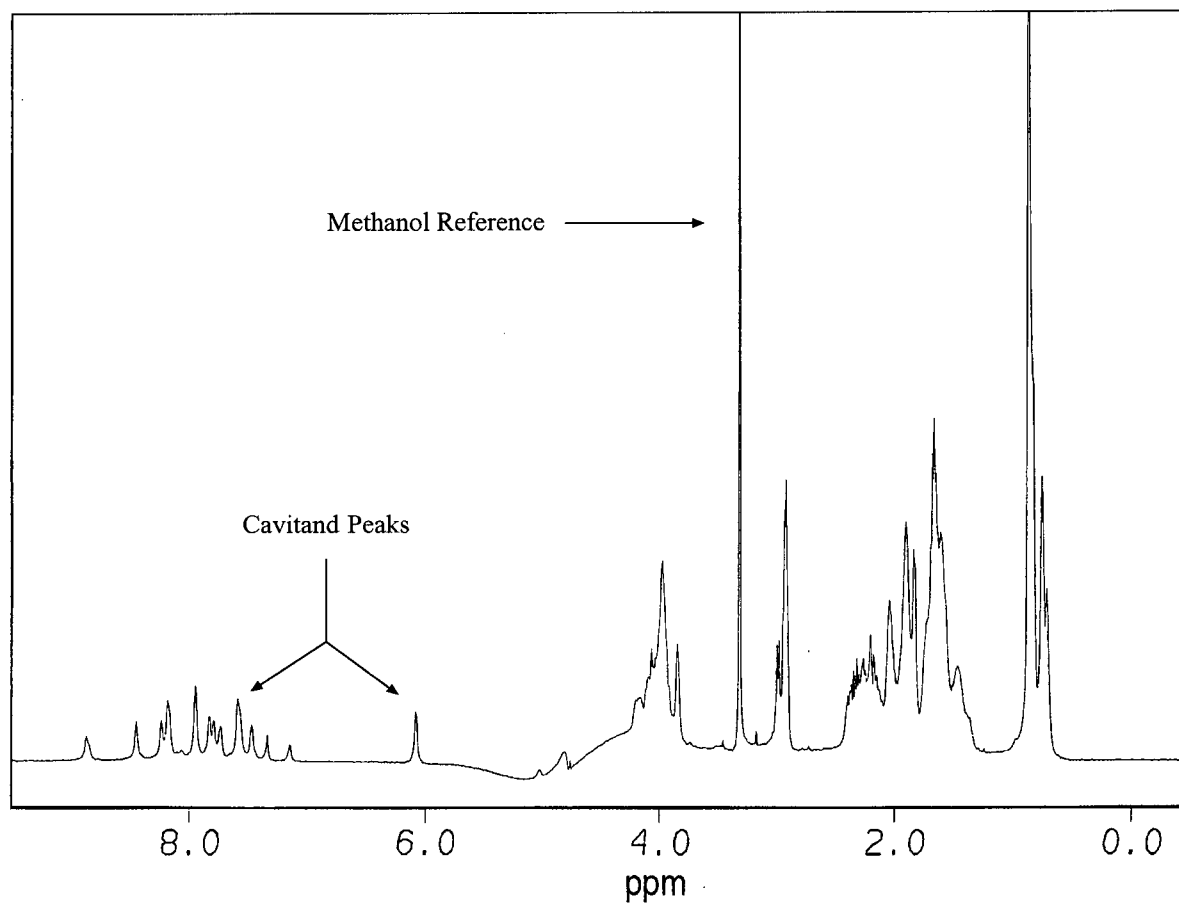


The near-UV spectrum of cavitein **27** (Figure 4.17) shows a positive absorption near 265 nm and a negative absorption near 246 nm. The only aromatic chromophores present in cavitein **27** lie in the four constituent aryl rings of the cavitand. This suggests that the cavitand template exists in an asymmetric environment which is indicative of specific non-averaged structural elements (i.e., native-like structure) near the cavitand-peptide linkage.

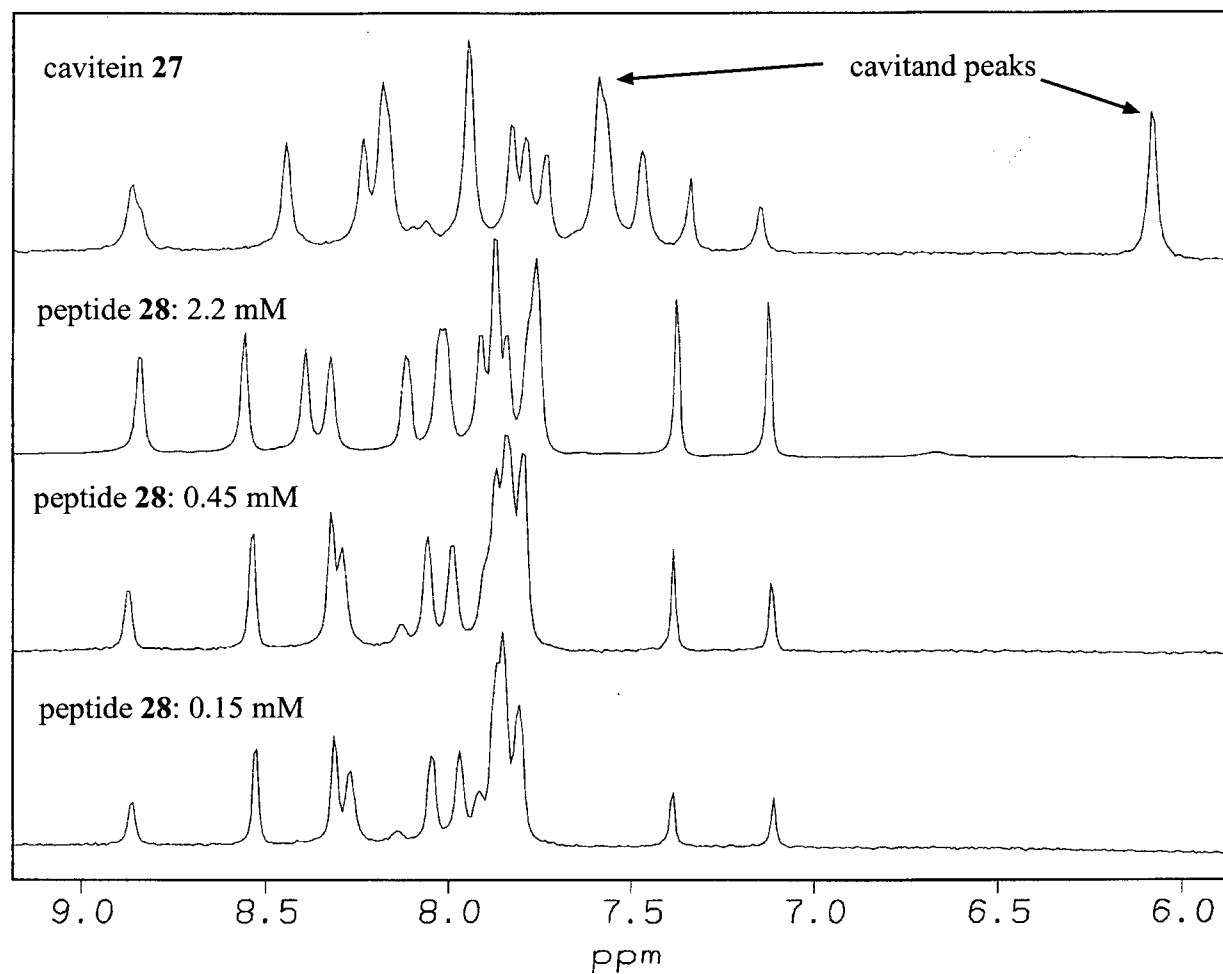
g. One-Dimensional  $^1\text{H}$  NMR Chemical Shift Dispersion

One-dimensional (1D)  $^1\text{H}$  NMR can be used to further study the structures of de novo proteins. Typically, considerable chemical shift dispersion is observed in natural proteins and is indicative of native-like structure.<sup>31</sup> In contrast, due to the looser packing arrangement of molten globules, signals arising from conformationally mobile groups are averaged and result in more broad, less disperse signals.<sup>32</sup> The 1D  $^1\text{H}$  NMR spectrum of cavitein **27** is shown in Figure 4.18. In addition, Figures 4.19 and 4.20 show expansions of the amide and aliphatic regions of this spectrum together with expansions of the 1D  $^1\text{H}$  NMR spectra of single-stranded peptide **28** at different concentrations.

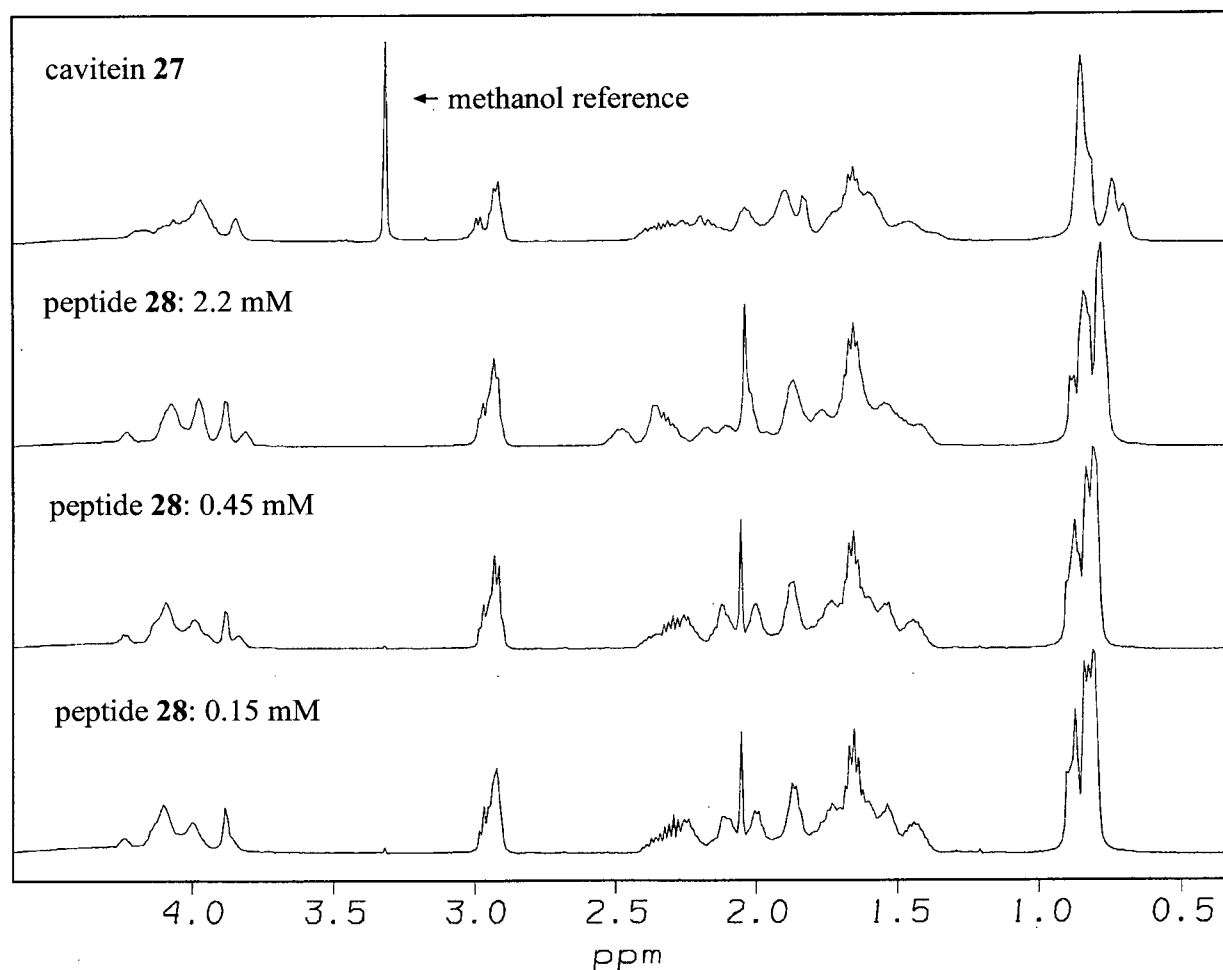
**Figure 4.18.** Full 500 MHz  $^1\text{H}$  NMR Spectrum of Cavitein **27** in 10%  $\text{D}_2\text{O}$ , 45 mM Phosphate Buffer, pH 7.0, 25  $^\circ\text{C}$ .



**Figure 4.19.** Expansions of the Amide Regions of 500 MHz  $^1\text{H}$  NMR Spectra of Cavitein **27** ( $\sim 0.2$  mM) and Peptide **28** at Different Concentrations at 25  $^\circ\text{C}$ , in Phosphate Buffer at pH 7.0 in the Presence of 10%  $\text{D}_2\text{O}$ .



**Figure 4.20.** Expansions of the Aliphatic Regions of 500 MHz  $^1\text{H}$  NMR Spectra of Cavitein **27** ( $\sim 0.2$  mM) and Peptide **28** at Different Concentrations at 25  $^\circ\text{C}$ , in Phosphate Buffer at pH 7.0 in the Presence of 10%  $\text{D}_2\text{O}$ .



The spectrum of cavitein **27** shows  $\sim 11$  relatively sharp and disperse amide peaks indicative of a well-defined amide backbone with a high content of secondary structure. This result complements the conclusions made from the CD studies with regard to the high helical content of cavitein **27**. The presence of just 11 distinct amide peaks suggests that the four helices are in degenerate environments likely due to the four-fold symmetry of the cavitein. However, note that despite the degeneracy in the primary sequence of the constituent peptides

(only four different amino acids), the well-defined amide backbone is still able to resolve ~ 11 different amide signals. Single-stranded peptide **28** also shows similar dispersion suggesting it too forms an aggregate with a high content of secondary structure.

The presence of well-defined tertiary structure can be probed by examining the aliphatic regions of  $^1\text{H}$  NMR spectra. Specific side-chain packing interactions typically give rise to sharp and disperse peaks. Relatively sharp signals are observed for cavitein **27** near 2.2 ppm which is likely indicative of specific side-chain interactions. Signals arising from the side-chain packing of the leucine core residues can be observed near 0.8 ppm. This region shows relatively poor dispersion perhaps indicative of mobility within the core of the bundle. In comparing the leucine peaks of **27** and **28** (Figure 4.20), two methyl peaks in cavitein **27** are shifted slightly upfield (near 0.7 ppm) from those of single-stranded peptide **28** which suggests a more tightly packed core in cavitein **27**.

The overall dispersion of the peaks in the 1D  $^1\text{H}$  NMR spectrum of cavitein **27** suggest that its structure possesses some native-like tertiary interactions. However, the relatively poor dispersion in the methyl region near 0.8 ppm suggests that its structure remains somewhat dynamic and thus somewhat characteristic of molten globule-like states.

One interesting observation was that even at concentrations ranging from 0.15 – 2.2 mM, single-stranded peptide **28** retains much of the chemical shift dispersion observed in monomeric cavitein **27**. This suggests that peptide **28** self-associates into an aggregate with structural interactions similar to cavitein **27**. Further studies into the nature of its structure were not performed but the observed spectra do suggest that regardless of the cavitand template and its significant effect on cavitein stability, the designed peptide sequence itself is well-suited for inducing specific structural components in  $\alpha$ -helical bundles.



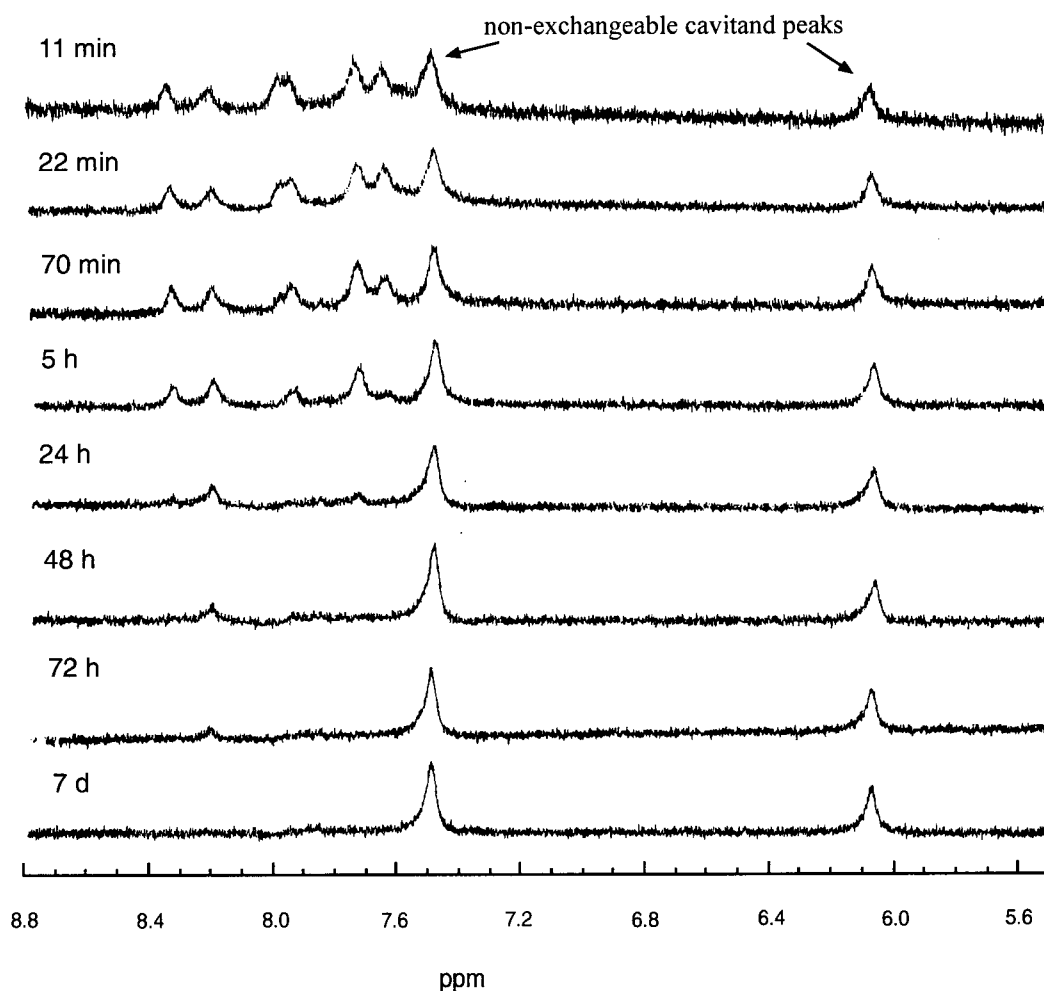
#### h. Hydrogen/Deuterium Amide Exchange

Hydrogen/deuterium (H/D) exchange can also be a useful diagnostic tool in the study of protein tertiary structures.<sup>33</sup> Due to the mobility of the molten globule state, H/D exchange rates of amide protons have been found to be much faster in the molten globule state when compared to the native state.<sup>34</sup> The observed rate of exchange of an amide proton in a structured environment ( $k_{\text{obs}}$ ) can be divided by a calculated intrinsic rate of exchange for an amide in an unstructured environment (i.e., “unprotected”) at any given temperature and pH ( $k_{\text{in}}$ ) to give a number called a “protection factor.” Larger protection factors indicate that an amide is slower to exchange with the solvent and is buried or “protected” from exchange. Backbone amides of native proteins typically possess protection factors in the range of  $10^5 - 10^8$ .<sup>34</sup> Molten globules typically exhibit protection factors in the range of  $10 - 10^3$ .<sup>34</sup>

In addition to temperature and pH, it is also known that the rate of amide exchange is dependent on the peptide sequence<sup>35</sup> and isotope effects.<sup>36</sup> In this H/D exchange experiment, the effect of different side-chains and neighboring groups were not taken into account since specific amide signals were not assigned to specific amino acids in the peptide sequence. In addition, kinetic isotope effects have been found<sup>36</sup> to be relatively small and are present in all H/D exchange experiments; hence, this effect is not considered here. To avoid these complications, a context-independent equation was used<sup>33a</sup> to calculate the intrinsic rate of exchange for an “unprotected” amide under the conditions stated below. As a result, the protection factors stated in the next paragraph are *not* sequence-corrected and may not be directly comparable to some literature values. However, a recently characterized de novo four-helix bundle<sup>37</sup> used this same method of calculating protection factors and thus comparison of

non-sequence-corrected protection factors is useful as a rough estimate of amide exchange when studying de novo proteins.

**Figure 4.21.** Stack Plot of 400 MHz  $^1\text{H}$  NMR Spectra Illustrating the Time Dependent Amide H/D Exchange of Cavitein **27** in 50 mM pD 5.0  $\text{CD}_3\text{COOD}/\text{NaOD}$  Buffer at 25  $^\circ\text{C}$ .



The rate of amide proton exchange in cavitein **27** was studied at pD 5.0 and 25  $^\circ\text{C}$  over the course of one week (Figure 4.21). It was found that many amides exchanged before the first scan could be acquired (11 min) and are typical of “unprotected” or molten globule-like amide

protons where protection factors can be  $< 20$ .<sup>34</sup> Several amides were visible for up to 24 hours and correspond to protection factors of  $\sim 4 \times 10^3$  or first-order rate constants of  $\sim 0.1 \text{ h}^{-1}$ . These exchange rates may correspond to some of the more stable amide protons found in molten globule-like structures or amides in an intermediary state between a molten globule and a native state.<sup>34</sup> One backbone amide proton ( $\sim 8.4 \text{ ppm}$ ) was visible for several days and corresponds to a protection factor of  $\sim 2 \pm 0.5 \times 10^4$  or a first-order rate constant of  $\sim 2 \pm 0.5 \times 10^{-2} \text{ h}^{-1}$ . This proton exhibits significant protection from exchange and its rate constant is consistent with an amide intermediary between that of a molten globule and a native state.<sup>34</sup>

Although some amide protons exchange rapidly, several amide protons exhibit significant protection from exchange. Therefore, this H/D exchange data suggests that cavitein **27** likely possesses properties of both molten globule and native-like states.

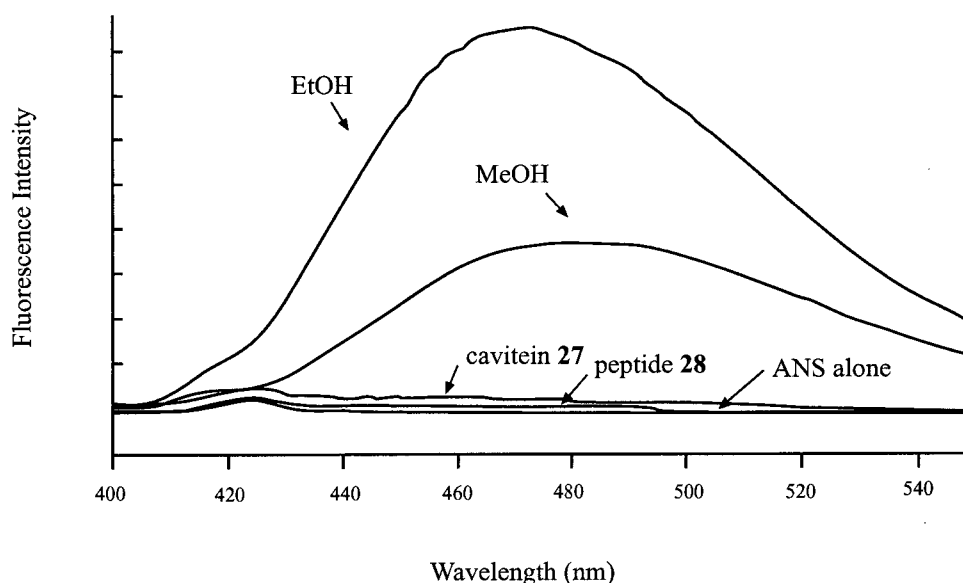
#### i. ANS Binding

The distinction between molten globule and native-like states can also be studied using the hydrophobic fluorescent dye 1-anilinonaphthalene-8-sulfonate (ANS).<sup>30</sup> Fluctuating molten globule-like structures typically expose more hydrophobic surfaces and consequently bind non-polar molecules much more strongly than native-like states. This effect can be investigated using ANS which has been found to preferentially bind to the molten globule state.<sup>38</sup> ANS binding can be conveniently monitored with fluorescence spectroscopy: ANS fluorescence intensity increases greatly and shifts to lower wavelengths when ANS is in a non-polar

environment (i.e., binding to the exposed hydrophobic surfaces of the dynamic molten globule state).

The binding of ANS to cavitein **27** and peptide **28** was studied with fluorescence spectroscopy and is shown in Figure 4.22. It is clear that under typical conditions of the experiment<sup>37,41a</sup> (see Figure caption), no significant binding of ANS to cavitein **27** was observed. This suggests that the structure of cavitein **27** is less dynamic than molten globule structures and possesses native-like characteristics to its tertiary structure.

**Figure 4.22.** Fluorescence Emission Spectra of 1  $\mu$ M ANS in the Presence of 95% Ethanol, 100% Methanol, 50  $\mu$ M Cavitein **27**, 50  $\mu$ M and 200  $\mu$ M (Indistinguishable) Peptide **28** and ANS Alone at 25 °C in pH 7.0, 50 mM Phosphate Buffer.



Interestingly, no binding of ANS was also observed for peptide **28** suggesting it too possesses native-like tertiary structure. As with its  $^1\text{H}$  NMR chemical shift dispersion, peptide **28** retains much of the tertiary interactions observed in monomeric cavitein **27**. Again, the

idealized peptide sequence may be sufficient to dictate tertiary interactions required for native-like structure. This will be discussed in more detail in the following concluding section to this chapter.

## C. Conclusion

This chapter presented the design, synthesis and characterization of the first cavitand-based de novo four-helix bundle: cavitein **27**. The peptide sequence was designed using a minimalist approach and previously established design principles. The coupling of the four unprotected peptide strands to tetrathiol cavitand **3** proceeded efficiently to afford cavitein **27**. Structural characterization of cavitein **27** by far-UV CD spectroscopy suggested its structure indeed possessed the desired parallel four-helix bundle topology. The rigid cavitand template renders cavitein **27** extremely stable to chemical denaturants when compared to peptide **28**. At pH 7 and 25 °C, cavitein **27** was found to exist as a mixture of monomers and dimers. Although the cause of the dimerization is not known for certain, evidence suggests that dimerization is caused by a combination of hydrophobic and electrostatic forces possibly from skewed helices. However, considerable evidence also suggested that dimerization does not significantly affect the structure and stability of the four-helix bundle.

Further investigations into the dynamics and packing of the structure were carried out using near-UV CD spectroscopy, 1D-<sup>1</sup>H NMR chemical shift dispersion, H/D amide proton exchange and ANS binding as monitored by fluorescence spectroscopy. Cavitein **27** appears to be in a state less dynamic than typical molten globules yet still lacks many of the characteristics

of native-like structure (i.e., NMR data). Taking into account all of the data presented in this chapter, cavitein **27** most likely lies in an intermediary state between that of a molten globule and a native-like state.

In the context of recent work in the area of de novo designed four-helix bundles, the structural properties of cavitein **27** are quite interesting. Typically, four-helix bundles containing all leucine-packed hydrophobic cores have led to molten globule structures.<sup>39</sup> Various research groups have thus incorporated aromatic residues,<sup>40</sup> potential metal-binding sites,<sup>41</sup> or alternating hydrophobic residues<sup>42</sup> in order to restrict the dynamics of the hydrophobic core and produce de novo four-helix bundles with native-like properties. The hydrophobic core of cavitein **27** contains only leucine residues yet possesses some of the properties of native-like structure. Does the cavitand aid in inducing native-like structure? Or, is the amino acid sequence sufficient to promote the observed native-like characteristics?

Indeed, the limited analysis of single-stranded peptide **28** suggested that it too possessed some characteristics of native-like structure. Therefore the peptide sequence alone may be sufficient to promote native-like structure. However, the structure of peptide **28** is not an ideal comparison to cavitein **27** since the unrestricted peptides of peptide **28** may fold into an antiparallel and/or non-tetrameric topology; a topology not available to the conformationally restricted peptides in cavitein **27**. We therefore conclude that although the cavitand template serves to greatly stabilize the helices, the effect of the cavitand on the structural dynamics of the four-helix bundle is uncertain. This effect is addressed in Chapter Five by varying the linkage between the cavitand and the peptides.

## D. Experimental

### i. General

All reagents were reagent grade except for the following: GuHCl and urea were electrophoresis grade; TFE was NMR grade. pH was determined using a Fisher Scientific Accumet pH meter 915 calibrated with two purchased buffered standards (pH 4.0 and 10.0). The synthesis of cavitand **3** can be found in Chapter Two. Peptides **26** and **28** and cavitein **27** were purified by reversed-phase HPLC using the Phenomenex Selectosil line of C<sub>18</sub> reversed-phase HPLC columns (analytical: 250 mm x 4.6 mm, 5  $\mu$ M particle size, 100 Å pore size; semi-preparative: 250 mm x 10 mm, 10  $\mu$ M, 300 Å; preparative: 250 mm x 22.5 mm, 10  $\mu$ M, 300 Å) and linear gradients of helium-sparged HPLC-grade acetonitrile (with 0.05% TFA) in de-ionized, helium-sparged filtered water (with 0.1% TFA). All samples were filtered with 0.45  $\mu$ M Nylon™ syringe filters (Phenomenex) before injection onto either (1) a Perkin-Elmer Biocompatible Pump 250 equipped with a PE LC90 BIO Spectrophotometric UV detector and a KIPP & ZONEN chart recorder or a (2) Varian 9010 Pump equipped with a Varian 9050 UV detector and a Varian 4290 Integrator. UV detection was set at 229 nm for the amide chromophore. In some cases, observation of a cavitand component to the chromatogram was studied by monitoring 280 nm. Analytical samples were run at 1 mL/min, semi-preparative samples at 5 mL/min and preparative samples at 10 mL/min. Purified samples were evaporated *in vacuo* and lyophilized. Sample concentrations were determined by amino acid analysis and errors in this method are typically  $\pm 10\%$ . LSIMS mass spectra for peptides **26** and **28** were acquired as described in Chapter Two. The electrospray mass spectrum for cavitein **27** was

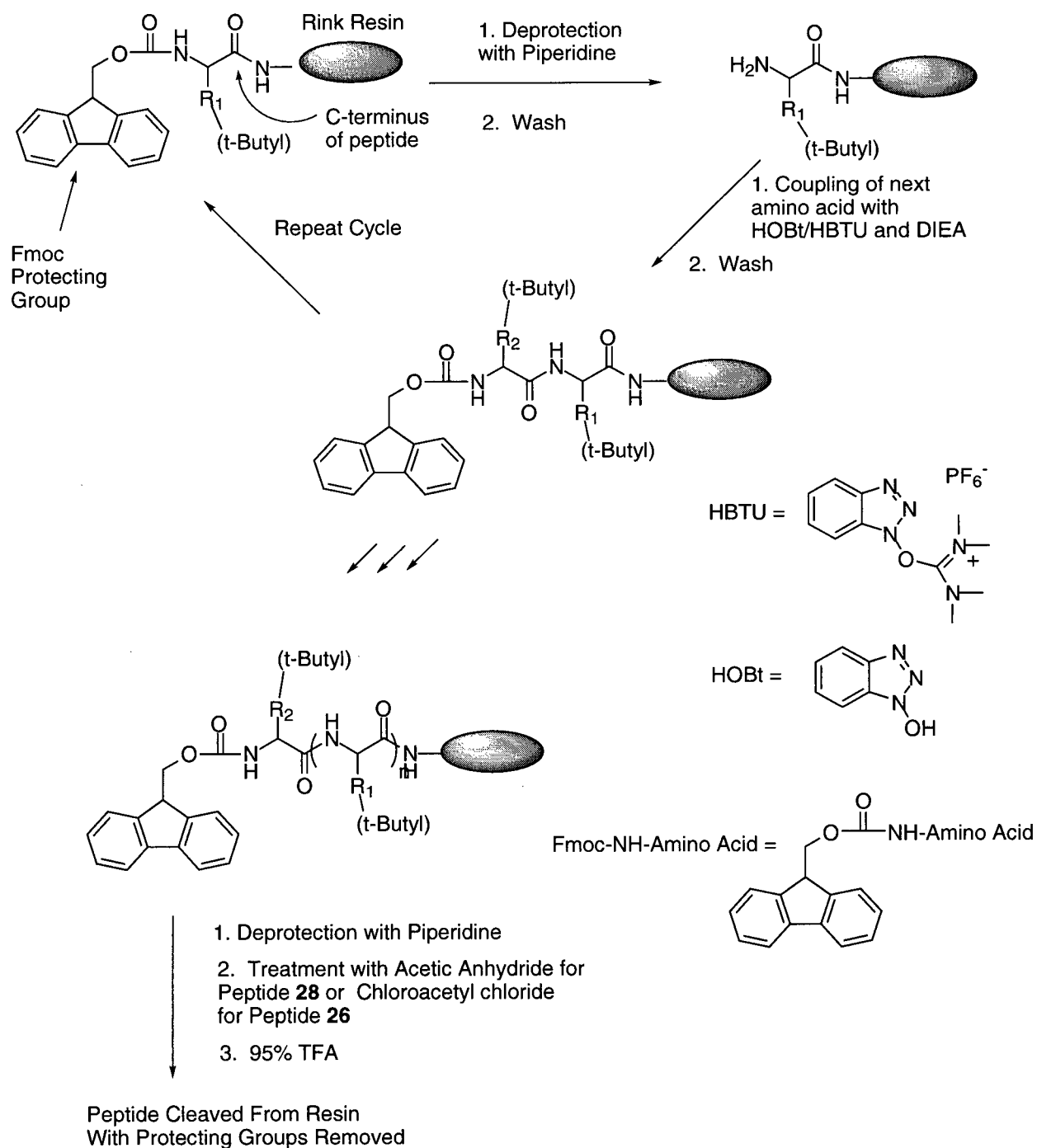
recorded using a PE-Sciex API III mass spectrometer. The UV spectrum of cavitein **27** was recorded on a PE Lambda 2 UV/Vis Spectrometer background-corrected with 50 mM pH 7.0 phosphate buffer at 25 °C. Note that the design and synthesis of peptide **26** and cavitein **27** have been published.<sup>8</sup>

## ii. Synthesis of Peptides **26** and **28**

All Fmoc protected amino acids, coupling reagents and solvents were purchased from Advanced Chemtech (Louisville, KY, USA). Peptides **26** and **28** were synthesized on an Applied Biosystems (ABI) 431A Automated Peptide Synthesizer coupled to an Apple MacIntosh IIsi computer. Peptides were synthesized on a 0.25 mmol scale using the *FastMoc*<sup>TM</sup> protocols shipped with the instrument. This entailed use of the following side-chain protected Fmoc amino acids: Fmoc-Glu(*t*-Bu)-COOH; Fmoc-Lys(CO<sub>2</sub>-*t*-Bu)-COOH; Fmoc-Gly-COOH; Fmoc-Leu-COOH. Rink resin was used to afford C-terminal amides.<sup>9</sup> The synthesis was accomplished using the solvent *N*-methylpyrrolidone (NMP) where a typical cycle entailed a: (1) 13 min Fmoc deprotection step using piperidine, (2) 6 min wash step using NMP, (3) 30 min coupling step to 1.0 mmol of the next Fmoc amino acid using HBTU (2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate) and HOBt (1-hydroxybenzotriazole) activation of the amino acid (started with DIEA), and (4) 6 min wash with NMP for a total of 55 min per cycle (Figure 4.23). See ABI 431A manual for more details.



**Figure 4.23.** Schematic Representation of the *FastMoc*<sup>TM</sup> Protocol on the ABI 431A Peptide Synthesizer.



For peptide **26**, the last cycle in the synthesis entailed chloroacetylation of the N-terminus by treatment of the resin (300 mg resin, ~ 100 mg peptide, ~ 0.06 mmol) with  $\text{ClCH}_2\text{COCl}$  (25  $\mu\text{L}$ , 0.32 mmol, 5 equiv.) and DIEA (55  $\mu\text{L}$ , 0.32 mmol, 5 equiv.) in DMF for 1 h at rt. For peptide **28**, the last cycle in the synthesis entailed acetylation of the N-terminus by treatment of the resin (150 mg) with 5 mL of 10% acetic anhydride in NMP for 1 h at rt. In each case, the reaction mixtures were filtered and the resin washed with DCM.

Both peptides **26** and **28** were cleaved from the resin and protecting groups removed with a 2 h treatment of 95% TFA. Typically, the resin was removed by filtration through a coarse frit while washing with  $\text{CH}_2\text{Cl}_2$ . The TFA/ $\text{CH}_2\text{Cl}_2$  mixture was evaporated to a few mLs *in vacuo* and the crude peptide was precipitated with ice-cold ether, filtered on a medium frit, purified by reversed-phase HPLC and lyophilized. Peptides **26** and **28** were characterized for purity by the observation of a single peak (at ~ 45% acetonitrile) by analytical HPLC (> 95% pure), and identified by amino acid analysis and LSIMS mass spectrometry. For peptide **26**:  $m/z$  1745 ( $(\text{M} + \text{H})^+$ , 100). For peptide **28**:  $m/z$  1711 ( $(\text{M} + \text{H})^+$ , 100).

### iii. Synthesis of Cavitein **27**

DIEA (2.6  $\mu\text{L}$ , 15  $\mu\text{mol}$ , 10 equiv.) was added to a solution of cavitand **3** (1.1 mg, 1.5  $\mu\text{mol}$ , 1 equiv.) and peptide **26** (11.9 mg, 6.8  $\mu\text{mol}$ , 4.4 equiv.) in degassed DMF and stirred at 25 °C for 16 h. The crude reaction mixture was evaporated *in vacuo* and purified by reversed-phase HPLC to afford cavitein **27** as a white solid (7.0 mg, 62%) after lyophilization. Purity of cavitein **27** was assessed by observation of one peak by analytical HPLC. Its identity was

confirmed by amino acid analysis and electrospray mass spectrometry to give a mass of 7556.0  $\pm$  0.9 Da [calcd 7557.1 Da (average isotope composition)].

#### iv. CD Studies

All CD spectra were recorded on a JASCO J-710 spectropolarimeter except for the temperature-dependence study of cavitein **27** which was recorded on a JASCO J-720 equipped with a "Windows-95" computer-controlled water bath (lab of G. Mauk in the biochemistry department at UBC). The J-710 was equipped with circulating water bath set at 25 °C and an IBM-compatible PC (80286) for processing of the spectra. Each spectrum was an average of three scans with the solvent baseline recorded separately and subtracted. All data points were performed in triplicate unless otherwise noted. Some of the parameter settings include: scanning speed = 50 nm/min; step resolution = 0.2 nm; 2 nm bandwidth. In addition, two parameter settings were used depending on the concentration of the sample: (1) response = 2.0 sec; sensitivity = 20 mdeg or (2) response = 0.5 sec; sensitivity = 200 mdeg. Calibration of the instrument was performed routinely with *d*-10-(+)-camphorsulfonic acid.<sup>43</sup> Error bars represent an average of the standard deviations for each point and found to be approximately  $\pm$ 5%.

Raw spectra were normalized to mean residue ellipticity  $[\theta]$  using the equation:

$$[\theta] = \theta_{\text{obs}} / 10 l c n$$

where  $\theta_{obs}$  is the observed ellipticity measured in millidegrees,  $l$  in the pathlength of the cell in centimeters,  $c$  is the peptide concentration in mol/L and  $n$  is the number of residues in the de novo protein.

Estimates of the maximum value of the molar ellipticity at 222 nm ( $[\theta]_{222-MAX}$ ) were calculated using the equation below developed by Chen et. al.:<sup>12</sup>

$$[\theta]_{222-MAX} = 39\,500 (1 - (2.57 / n))$$

where  $n$  is the number of residues per helix. Note that 39 500 and 2.57 are wavelength dependent constants and thus apply only to 222 nm.

Quartz cuvettes were used with either 1 mm or 1 cm path lengths depending on the peptide concentrations used. In studies where both pathlengths were used (i.e., concentration dependence studies), at least two overlapping points were acquired for each path length.

Guanidine hydrochloride denaturation experiments were performed using an 8.0 M solution of GuHCl buffered at pH 7.0 with 50 mM phosphate. The exact concentrations of various 8 M stock solutions were determined by refractometry by the method of Pace<sup>20</sup> and found to be within 0.5% of 8.00 M GuHCl. Each data point was performed in duplicate and acquired with a separate sample of cavitein **27** in pH 7.0, 50 mM phosphate buffer. To attain the high concentrations of GuHCl, cavitein **27** was dissolved directly in the 8 M GuHCl stock solution, and diluted accordingly to give the appropriate concentrations – note that this also demonstrates that the unfolding of cavitein **27** is a reversible process. Samples were equilibrated for 5 min only since it was found that any effect of GuHCl was immediate as the

mean residue ellipticity did not change from 5 min to 24 h at 25 °C. The urea experiment was performed in a similar manner.

As mentioned above, peptide and cavitein concentrations were determined by amino acid analysis. For cavitein **27**, its UV spectrum was recorded and the molar extinction coefficient was found to be  $\epsilon = 16\,500 \pm 500 \text{ M}^{-1}\text{cm}^{-1}$  for  $\lambda_{\text{max}} = 270 \text{ nm}$  in 50 mM phosphate buffer at pH 7.0 at 25 °C. Under these conditions, this extinction coefficient was used in subsequent concentration determinations. Despite the presence of dimers, this extinction coefficient was found to be concentration independent in the range of 13 – 100  $\mu\text{M}$ .

#### v. NMR Studies

$1\text{D-}^1\text{H}$  NMR spectra used to evaluate chemical shift dispersion were run at 500 MHz on a Varian Unity (acquired and processed by G. Connelly in the laboratory of L. McIntosh, Department of Biochemistry, UBC) equipped with shimming hardware. Each sample was run at 25 °C and dissolved in 45 mM phosphate buffer (90:10,  $\text{H}_2\text{O}:\text{D}_2\text{O}$ ) at pH 7.0 using a sweep width of 8000.00 Hz, a data size of 4096 complex points and 128 scans. A relaxation delay of 1.0 s was used, during which time the water signal was saturated using a frequency-selective low-power decoupling pulse. Spectra were processed using Felix v2.30 (Biosym, San Diego, CA, USA). A convolution function was applied to remove the residual water signal. No other line-broadening functions were used. A residual amount of methanol was used as a reference (3.31 ppm). The concentration of cavitein **27** was  $\sim 0.2 \text{ mM}$ ; concentrations of peptide **28** were  $\sim 0.15 \text{ mM}$ ,  $0.45 \text{ mM}$  and  $2.2 \text{ mM}$ .

The H/D exchange experiment was run at 400 MHz on a Bruker WH-400 at 25 °C using presaturation of the water signal. The experiment was initiated by dissolving cavitein **27** directly into 50 mM acetic acid- $d_4$  buffer in  $D_2O$  at pD 5.0 and transferred quickly to the NMR tube. pD was corrected for isotope effects using the equation:

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

where  $pH_{\text{read}}$  is the reading of the pH electrode.<sup>44</sup> The first scan was acquired immediately at 11 min followed by others at 22 min, 70 min, 5 h, 24 h, 48 h, 72 h, and 7 d after the start of the experiment. The peak heights were integrated and normalized with the non-exchangeable proton  $H_{\text{out}}$  (near 6 ppm) from the cavitand. The first-order rate constant for the longest-lived amide proton at ca. 8.4 ppm was calculated to be  $2 \pm 0.5 \times 10^{-2} \text{ h}^{-1}$ . This value was calculated using the first-order rate equation:

$$\ln ([H_0] / [H_t]) = k t$$

where  $k$  is the first order rate constant,  $t$  is the time at which a scan was taken,  $[H_0]$  is the integration of an amide proton at time zero, and  $[H_t]$  is the integration of the same proton at time  $t$ . From this equation, the half-life ( $t_{1/2}$ ) of the proton at ca. 8.4 ppm was calculated to be  $37 \pm 9$  h at pD 5.0 and 25 °C using the equation:

$$t_{1/2} = 0.693 / k$$

Its protection factor was then calculated to be  $2 \pm 0.5 \times 10^4$  where  $t_{1/2}$  for an “unprotected” amide proton at pH 5.0 at 25 °C is calculated to be 0.11 min using the following equation:<sup>33a</sup>

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

where  $t_{1/2 - \text{intrinsic}}$  is the intrinsic half-life for an unprotected proton, and  $T$  is the temperature in °C. Errors represent one standard deviation from four rate constant estimates.

#### vi. ANS Binding Studies

ANS fluorescence measurements were determined on a Perkin-Elmer LS-5B Luminescence Spectrophotometer using a 1 cm path length. All runs contained 1 μM ANS in 50 mM phosphate buffer pH 7.0 at 25 °C with either 50 μM cavitein **27**, 50 μM control **28**, 200 μM control **28**, 95% ethanol or 100% methanol. Excitation was 370 nm and emission was recorded between 400 and 550 nm.

#### vii. Analytical Ultracentrifugation Studies

Sedimentation equilibrium studies were performed in the laboratory of Jeff Kelly by Kelly group members at the Scripps Research Institute and thus the following experimental procedure was written with the help of members of the Kelly research group.

Sedimentation equilibrium measurements were performed on a temperature-controlled Beckman XL-I analytical ultracentrifuge equipped with an An60Ti rotor and photoelectric scanner. A double sector cell, equipped with a 12 mm epon centrepiece and sapphire windows loaded with 120-140  $\mu\text{L}$  of sample using a blunt-end microsyringe. Data were collected at 25  $^{\circ}\text{C}$  at a rotor speed of 30 000 rpm until equilibrium was established across the cell. Samples were allowed to equilibrate for 24-32 h, and duplicate scans 3 h apart were overlaid to determine that equilibrium had been reached. Scans were acquired by UV detection. The partial specific volumes of the cavitein was calculated<sup>16a</sup> based on its amino acid composition only (refer to Section B.iii.c in the chapter for an account of the errors involved in this estimation). The solution densities were calculated based on the buffer composition.<sup>16a</sup> Specific experimental details for the experiments described in this chapter are listed in Table 4.1.

**Table 4.1.** Experimental Parameters For Sedimentation Equilibrium Studies on Cavitein **27** Run at 25  $^{\circ}\text{C}$  at pH 7.0.

Conditions	Concentration of Cavitein <b>27</b> ( $\mu\text{M}$ )	Partial Specific Volume ( $\text{mL/g}$ )	Solution Density ( $\text{g/mL}$ )	Detection Wavelength ( $\text{nm}$ )
50 mM Phosphate	30	0.7910	1.00000	270
50 mM Phosphate	300	0.7910	1.00000	300
6.0 M GuHCl	30	0.8131	1.14227	270
2.0 M NaCl	30	0.7823	1.07705	270
10% MeOH	30	0.7977	0.98465	270



The data were analyzed by a nonlinear least-squares analysis using the Origin software provided by Beckman. The data were then fit to a single non-associating ideal species model using the following equation to determine the molecular weight:

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

where  $A_r$  is the absorbance at radius  $r$ ,  $A_o$  is the absorbance at a reference radius  $r_o$  (usually the meniscus),  $\bar{v}$  is the partial specific volume of the peptide,  $\rho$  is the density of the solvent (g/mL),  $\omega$  is the angular velocity of the rotor (radian/sec),  $E$  is the baseline error correction factor,  $M$  is the molecular weight,  $T$  is the absolute temperature and  $R$  is the universal gas constant.

In cases where the data did not fit well to the equation above, best fits were made to the equation below accounting for possible oligomeric species:

$$\begin{aligned} A_r = & A_{(\text{monomer}, r_o)} \exp [ ((1 - \bar{v}\rho)\omega^2/2RT) (M(r^2 - r_o^2))] \\ & + A_{(\text{monomer}, r_o)}^{n_2} K_{a,2} \exp [ ((1 - \bar{v}\rho)\omega^2/2RT) n_2(M(r^2 - r_o^2))] \\ & + A_{(\text{monomer}, r_o)}^{n_3} K_{a,3} \exp [ ((1 - \bar{v}\rho)\omega^2/2RT) n_3(M(r^2 - r_o^2))] \\ & + A_{(\text{monomer}, r_o)}^{n_4} K_{a,4} \exp [ ((1 - \bar{v}\rho)\omega^2/2RT) n_4(M(r^2 - r_o^2))] + E \end{aligned}$$

where  $A_r$  is the absorbance at radius  $r$ ,  $A_{(\text{monomer}, r_o)}$  is the absorbance of the monomer at the reference radius  $r_o$ ,  $M$  is the molecular weight,  $n_2$  is the stoichiometry for species 2,  $K_{a,2}$  is the association constant for the monomer- $n$ -mer equilibrium of species 2,  $n_3$  is the stoichiometry for species 3,  $K_{a,3}$  is the association constant for species 3,  $n_4$  is the stoichiometry for species 4,  $K_{a,4}$  is the association constant for species 4 and  $E$  is the baseline offset. The accuracy of the fit was

evaluated on the basis of the randomness and magnitude of the residuals, expressed as the difference between the theoretical curve and the experimental data and by checking the fit parameters for physical reasonability.

In the case of cavitein **27**, the best fit was achieved using an equation describing a monomer-dimer equilibrium mixture. In this determination, the molecular weight is fixed at 7557 Da (the actual molecular weight of cavitein **27** as determined from its chemical composition) and the association constant is varied until the best fit is achieved. Since there are large errors associated with estimating  $\bar{v}$ , the apparent solution molecular weight of cavitein **27** as determined by sedimentation equilibrium may not be near 7557 Da. In fact, the majority of caviteins in this thesis give apparent monomeric molecular weights of  $\sim 2\,000$  Da greater than their actual molecular weight likely the result of errors in  $\bar{v}$ . A consequence of the error associated with fixing the molecular weight is that this error would be reflected in the calculated association constant. For this reason, the association constant for the monomer-dimer equilibrium in cavitein **27** was not calculated.

## E. References

1. According to a crystal structure, the distance between adjacent bromines when Br's replace SH's of **3** are 6.8 Å and the distance between opposing bromines is 9.6 Å. See: Cram, D. J.; Karbach, S.; Kim, H.-E.; Knobler, C. B.; Maverick, E. F.; Ericson, J. L.; Helgeson, R. C. *J. Am. Chem. Soc.* **1988**, *110*, 2229-2237.
2. Reddy, B. V. B.; Blundell, T. L. *J. Mol. Biol.* **1993**, *233*, 464-479.
3. (a) Schneider, J. P.; Kelly, J. W. *Chem. Rev.* **1995**, *95*, 2169-2187; (b) Tuchscherer, G.; Mutter, M. *J. Biotechnol.* **1995**, *41*, 197-210.
4. Tucker, J. A.; Knobler, C. B.; Trueblood, K. N.; Cram, D. J. *J. Am. Chem. Soc.* **1989**, *111*, 3688-3699.
5. DeGrado, W. F.; Wasserman, Z. R.; Lear, J. D. *Science* **1989**, *243*, 622-628. The sequence is also similar to that of: Zhou, N. E.; Zhu, B.-Y.; Sykes, B. D.; Hodges, R. S. *J. Am. Chem. Soc.* **1992**, *114*, 4320-4326.
6. Chakrabarty, A.; Baldwin, R. L. *Adv. Prot. Chem.* **1995**, *46*, 141-176.
7. Aurora, R.; Srinivasan, R.; Rose, G. D. *Science* **1994**, *264*, 1126-1130.
8. Gibb, B. C.; Mezo, A. R.; Sherman, J. C. *Tetrahedron Lett.* **1995**, *36*, 7587-7590.
9. Rink, H. *Tetrahedron Lett.* **1987**, *28*, 3787.
10. (a) Eisenberg, D.; Wilcox, W.; Eshita, S. M.; Pryciak, P. M.; Ho, S. P.; DeGrado, W. F. *Proteins Struc. Funct. Genet.* **1986**, *1*, 16-22; (b) Ho, S. P.; DeGrado, W. F. *J. Am. Chem. Soc.* **1987**, *109*, 6751-6758; (c) Hill, C. P.; Anderson, D. H.; Wesson, L.; DeGrado, W. F.; Eisenberg, D. *Science* **1990**, *249*, 543-546; (d) Osterhout, J. J.; Handel, T.; Na, G.; Toumadje, A.; Long, R. C.; Connolly, P. J.; Hoch, J. C.; Johnson Jr., W. C.; Live, D.; DeGrado, W. F. *J. Am. Chem. Soc.* **1992**, *114*, 331-337; (e) Betz, S.; Fairman, R.; O'Neil, K.; Lear, J.; DeGrado, W. *Phil. Trans. R. Soc. Lond. B* **1995**, *348*, 81-88; (f) Fezoui, Y.; Connolly, P. J.; Osterhout, J. J. *Protein Sci.* **1997**, *6*, 1869-1877; (g) Chmielewski, J.; Lipton, M. *Int. J. Peptide Protein Res.* **1994**, *44*, 152-157; (h) Ciesla, D. J.; Gilbert, D. E.; Feignon, J. *J. Am. Chem. Soc.* **1991**, *113*, 3957-3961; (i) Olofsson, S.; Johansson, G.; Baltzer, L. *J. Chem. Soc. Perkin Trans. 2* **1995**, 2047-2056.
11. (a) Johnson Jr., W. C. *Proteins Struc. Func. Genet.* **1990**, *7*, 205-214; (b) Woody, R. W. *Methods Enzymol.* **1995**, *246*, 34-70; (c) Fasman, G. D. (Ed.) *Circular Dichroism and the Conformational Analysis of Biomolecules*; Plenum Press: New York, 1996.

12. Chen, Y.-H.; Yang, J. T.; Chau, K. H. *Biochemistry* **1974**, *13*, 3350-3359.
13. Nelson, J. W.; Kallenbach, N. R. *Proteins Struc. Funct. Genet.* **1986**, *1*, 211-217.
14. (a) Walgers, R.; Lee, T. C.; Cammers-Goodwin, A. *J. Am. Chem. Soc.* **1998**, *120*, 5073-5079; (b) Luo, P.; Baldwin, R. L. *Biochemistry* **1997**, *36*, 8413-8421; (c) Cammers-Goodwin, A.; Allen, T. J.; Oslick, S. L.; McClure, K. F.; Lee, J. H.; Kemp, D. S. *J. Am. Chem. Soc.* **1996**, *118*, 3082-3090; (d) Kentsis, A.; Sosnick, T. R. *Biochemistry* **1998**, *37*, 14613-14622.
15. Su, J. Y.; Hodges, R. S.; Kay, C. M.; *Biochemistry* **1994**, *33*, 15501-15510.
16. (a) Harding, S. E.; Rowe, A. J.; Horton, J. C. Eds. *Analytical Ultracentrifugation in Biochemistry and Polymer Science*; Royal Society of Chemistry: Oxford, 1992; (b) Hansen, J. C.; Lebowitz, J.; Demeler, B. *Biochemistry* **1994**, *33*, 13155-13163.
17. Gel filtration chromatography was performed using a Waters Protein Pak 125 column (MW range 2 000 – 80 000) using degassed 50 mM phosphate pH 7 buffer with and without 0.5 M NaCl at a flow rate of 0.5 mL/min using UV detection at 280 nm. The column was calibrated with known protein standards.
18. The following retention times were found using 50 mM phosphate buffer, pH 7.0 (no added salt): solvent front: 11 min; 25 kDa protein standard: 22 min; 13.7 kDa protein standard: 25 min; N1/Ar/Me injected at 150  $\mu$ M: 26 min (~ 10%), 27 min (~ 25%), 28 min (~ 65%); N1/Ar/Me injected at 15  $\mu$ M: same as 150  $\mu$ M run; N1GG/Ar/Me: 29 min (20%), 31 min (80%); N1GGG/Ar/Me: 44 min (20%), 47 min (80%); N1/Bzl/PO<sub>3</sub>H<sub>2</sub>: 17 min (100%).  
 The following retention times were found using 50 mM phosphate buffer, pH 7.0 and 0.5 M NaCl: 25 kDa protein standard: 19 min; 13.7 kDa protein standard: 26 min; N1/Ar/Me: 31 min (100%); N1GG/Ar/Me: 38 min (100%); N1GGG/Ar/Me: 53 min (100%); N1/Bzl/PO<sub>3</sub>H<sub>2</sub>: 18 min (100%).  
 Although the added salt appeared to resolve each cavitein to a single peak, the vastly varying retention times suggested that the caviteins were interacting in an anomalous manner with the stationary phase of the column.
19. However, a reversed-phase column was used prior to injection onto the ESI mass spectrometer in order to concentrate the cavitein and provide better signal:noise ratios in the resulting spectra. As a consequence, gradients of acetonitrile in water were used to elute the caviteins off the column. In consideration of the sedimentation equilibrium experiment conducted on cavitein **27** in the presence of 10% MeOH (described later in Chapter Four) which demonstrated that methanol inhibits dimerization, the acetonitrile required to elute the cavitein off the column (likely ~50% v/v) may be more than sufficient to inhibit dimerization/oligomerization prior to injection into the spectrometer. Since no evidence for oligomerization was observed by ESMS for any of the caviteins presented in this thesis, it is consistent with acetonitrile inhibiting oligomerization.

20. Pace, C. N. *Methods Enzymol.* **1986**, *131*, 266-280.
21. Creighton, T. E. *Proteins: Structures and Molecular Properties*; W. H. Freeman: New York, 1993; pp 293-296.
22. Monera, O. D.; Kay, C. M.; Hodges, R. S. *Protein Sci.* **1994**, *3*, 1984-1991.
23. Zou, Q.; Habermann-Rottinghaus, S. M.; Murphy, K. P. *Proteins Struc. Funct. Genet.* **1998**, *31*, 107-115.
24. Green, S. M.; Pace, C. N. *J. Biol. Chem.* **1974**, *249*, 5388-5393.
25. Perkins, S. J. *Eur. J. Biochem.* **1986**, *157*, 169-180.
26. Lee, J. C.; Timasheff, S. N. *Methods Enzymol.* **1979**, *61*, 49-57.
27. Condino, J. *Determination of Molecular Weights by Sedimentation Equilibrium* In *Technical Information DS820*; Palo Alto, California, Beckman Coulter Inc., 1992.
28. In the following reference, the addition of GuHCl was found to linearly reduce the melting temperatures of two-stranded coiled-coils allowing for an extrapolation to the  $T_m$  in the absence of denaturant: Su, J. Y.; Hodges, R. S.; Kay, C. M. *Biochemistry* **1994**, *33*, 15501-15510.
29. Kahn, P. C. *Methods Enzymol.* **1979**, *61*, 339-377.
30. Kuwajima, K. *Proteins Struc. Funct. Genet.* **1989**, *6*, 87-103.
31. (a) Cavanagh, J.; Fairbrother, W. J.; Palmer III, A. G.; Skelton, N. J. *Protein NMR Spectroscopy: Principles and Practice*; Academic Press: San Diego, 1996; (b) Roy, S.; Helmer, K. J.; Hecht, M. H. *Folding & Design* **1997**, *2*, 89-92.
32. Ptitsyn, O. B. *Adv. Prot. Chem.* **1995**, *47*, 83-229.
33. (a) Englander, S. W.; Downer, N. W.; Teitelbaum, H. *Annu. Rev. Biochem.* **1972**, *41*, 903-905; (b) Raschke, T. M.; Marqusee, S. *Curr. Opin. Biotech.* **1998**, *9*, 80-86; (c) Englander, S. W.; Kallenbach, N. R. *Q. Rev. Biophys.* **1984**, *16*, 521-655; (d) Englander, S. W.; Mayne, L.; Bai, Y.; Sosnick, T. R. *Protein Sci.* **1997**, *6*, 1101-1109.
34. (a) Hughson, F. M.; Wright, P. E.; Baldwin, R. L. *Science* **1990**, *249*, 1544-1548; (b) Jeng, M.-F.; Englander, S. W.; Elove, G. A.; Wand, A. J.; Roder, H. *Biochemistry* **1990**, *29*, 10433-10437.
35. Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. *Proteins Struc. Funct. Genet.* **1993**, *17*, 75-86.

36. Connolly, G. P.; Bai, Y.; Jeng, M.-F.; Englander, S. W. *Proteins Struct. Funct. Genet.* **1993**, *17*, 87-92.
37. Roy, S.; Ratnaswamy, G.; Boice, J. A.; Fairman, R.; McLendon, G.; Hecht, M. H. *J. Am. Chem. Soc.* **1997**, *119*, 5302-5306.
38. Semisotnov, G. V.; Rodionova, N. A.; Razgulyaev, O. I.; Uversky, V. N.; Gripas, A. F.; Gilmanshin, R. I. *Biopolymers* **1991**, *31*, 119-128.
39. Betz, S. F.; Raleigh, D. P.; DeGrado, W. F. *Curr. Opin. Struct. Biol.* **1993**, *3*, 601-610.
40. Brive, L.; Dolphin, G. T.; Baltzer, L. *J. Am. Chem. Soc.* **1997**, *119*, 8598-8607.
41. (a) Raleigh, D. P.; Betz, S. F.; DeGrado, W. F. *J. Am. Chem. Soc.* **1995**, *117*, 7558-7559; (b) Hill, R. B.; DeGrado, W. F. *J. Am. Chem. Soc.* **1998**, *120*, 1138-1145.
42. (a) Jiang, X.; Bishop, E. J.; Farid, R. S. *J. Am. Chem. Soc.* **1997**, *119*, 838-839; (b) Gibney, B. R.; Rabanal, F.; Skalicky, J. J.; Wand, A. J.; Dutton, P. L. *J. Am. Chem. Soc.* **1997**, *119*, 2323-2324; (c) Betz, S. F.; Liebman, P. A.; DeGrado, W. F. *Biochemistry* **1997**, *36*, 2450-2458.
43. Chen, G. C.; Yang, J. T. *Anal. Lett.* **1977**, *10*, 1195-1207.
44. Glasoe, P. K.; Long, F. A. *J. Phys. Chem.* **1960**, *64*, 188-190.

## Chapter Five: Effect of Peptide-Cavitand Linkage and Cavitand Foot on Cavitein Structure and Stability.

### A. Introduction

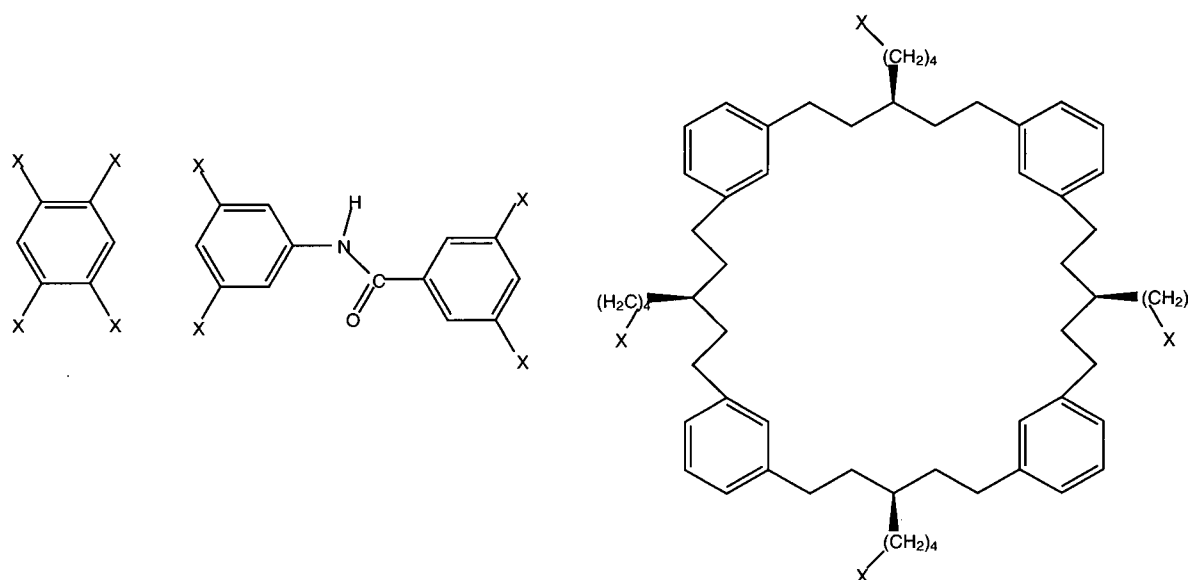
#### i. Effect of Cavitand-Peptide Linkages

The “template-assembly” approach in de novo protein design is known to greatly enhance the stability of designed protein structures.<sup>1</sup> Section F of Chapter One discussed various examples illustrating the use of templates in the de novo design of four-helix bundles. In each case, the template stabilized the resulting helical bundle relative to the single-stranded peptide and promoted an intramolecular folding process between the peptides. In these examples, however, the relationship between the structures of the four-helix bundle and the template was not explored. Hence, this raises a number of questions regarding the effect of the template on the resulting protein structures. Is all the information required to fold a template-assembled four-helix bundle found in the amino acid sequence of its constituent peptides? Do different templates and peptide-template linkages contribute differently to the final structure of the four-helix bundle?

Recently, Fairlie and coworkers partially addressed these questions by synthesizing de novo four-helix bundles using the same peptide sequence on three significantly different templates (Figure 5.1).<sup>2</sup> The authors found that: “for suitably long linkers, the *shape, size and directionality* of the template are not critical” for four-helix bundle formation. Each of the three de novo four-helix bundles presented used different templates, yet possessed similar stabilities and helicities. However, this article did not explore the structures of each four-helix bundle in

detail and thus the side-chain packing and structural dynamics of each bundle may have differed significantly. In particular, the authors did not account for the different slopes of their GuHCl denaturation transitions, which may indeed suggest significant differences between the structures of the three de novo proteins.

**Figure 5.1.** Structures of Three Different Templates Used by Fairlie<sup>2</sup> in Evaluating the Effect of Different Templates on the Structures and Stabilities of De Novo Four-Helix Bundles. Positions Labeled “X” Represent Peptide Attachment Sites.



We were interested in studying the effect of linkers of various lengths and characteristics on the structure and stability of the resulting four-helix bundles. Conceptually, one can visualize too short a linker in which portions of the peptides may be forced, by the rigid cavitation, to adopt non-ideal helical conformations. On the other hand, it is possible to incorporate sufficiently long peptide-cavitation linkers such that the added stabilizing effect of the cavitation's rigidity is lost when compared to other, more flexible templates. There should exist



an ideal linker where the conformations of the peptides are sufficiently restricted by the rigid cavitand to significantly contribute to structural stability, yet given enough conformational freedom to adopt ideal helical structures.

As mentioned in Chapter Four, Section B.i.a, the rigid cavitand template places fixed sites for peptide attachment approximately 7 Å apart. Conveniently, this distance is approximately the interhelical distance found in natural four-helix bundles. Therefore, the length of the peptide-cavitand linkage in cavitein **27** was first minimized in an attempt to exploit the rigidity of the template in an effort to impart great stability to cavitein **27**. To this end, we incorporated (Chapter Four) a short “CH<sub>2</sub>CO” linkage between the thiols of cavitand **3** and the N-termini of the peptides. Indeed, the prototype cavitein (**27**) was extremely stable and possessed some of the properties of native-like structure. In addition, it was clear that the cavitand template stabilized the helical bundle in comparison to single-stranded peptide **28**.

However, the relative stability advantage of using rigid cavitands over other templates remained unclear. For example, would it be possible to simply use the template of Mutter (Chapter One, Section F.ii.a) or Ghadiri (Chapter One, Section F.ii.b) and attain the same stability?

In addition, cavitein **27** was found to exist as a mixture of monomers and dimers. Is this a result of the cavitand template or the peptide sequence? Moreover, studies on single-stranded peptide **28** revealed that it possessed some of the properties of native-like structure – a feature derived solely from its primary sequence. Does the cavitand aid in inducing the native-like structure observed for cavitein **27** or is the native-like structure solely a function of the peptide sequence?

Indeed, there are many questions that need to be addressed. In this chapter, cavitein variants are presented in which the linkage between the cavitand and the peptides is varied in an

effort to study the effect of the cavitand on the oligomeric state, stability and structural dynamics of caviteins composed of the same core  $4 \times 14$ -amino acid sequence.

ii. Effect of Cavitand Feet

Cavitein **27** and the majority of the caviteins presented in this chapter are based on cavitand **3** possessing methyl "feet". We were also interested in studying caviteins with more hydrophilic feet to determine if the feet had an effect on the stability and/or oligomeric state of the bundle. For example, is the monomer-dimer equilibrium described for cavitein **27** a result of hydrophobic association via the methyl feet of the cavitand (Figure 4.9)? One way of answering this question is to incorporate more hydrophilic feet into the cavitand portion of the cavitein.

Chapter Two described the synthesis of phosphate-footed thiol cavitand **5** which is used in this chapter for cavitein synthesis, and subsequently compared with the methyl-footed equivalent. In addition, Chapter Two described the synthesis of phosphate-footed benzyl bromide **21**, inseparable from residual amounts of the corresponding tris-phosphate derivative. In this chapter, cavitand **21** is shown to react with a cysteine-containing peptide to afford a pure tetra-phosphate-footed cavitein easily separable from its tris-phosphate derivative.

### iii. Nomenclature

This chapter describes several variants of cavitein **27** presented in Chapter Four. Since each of these caviteins and their constituent peptides vary only slightly in composition, each is given a name that is indicative of its composition. These names are given below in Table 5.1 along with compound numbers, cavitand-peptide linkages, and cavitand foot.

Each cavitein name is made up of three parts separated by a slash (/). In the first part of the name, “N” refers to the number of methylene units in the linkage (“N4”, or four methylenes), while “G” refer to the number of glycine linker amino acids (e.g., “GGG” = three glycine residues). The first part of the cavitein name also corresponds to the name of the peptide used in its synthesis. The middle part of the name refers to the type of cavitand linkage (e.g., Ar = aryl thiol cavitand; Bzl = benzylthiol or benzyl bromide cavitand). The last part of the name refers to the type of foot present in the cavitand (e.g., Me = methyl foot;  $\text{PO}_3\text{H}_2$  = phosphate foot). The code-names of each cavitein will be used instead of their compound numbers for sake of clarity in the ensuing discussions. Note that each cavitein uses the same core 14-amino acid sequence used in peptide **26**, and all peptides are attached to the cavitand at their *N*-terminal ends.

**Table 5.1.** Nomenclature Used in Describing the Nine Caviteins Presented in Chapter Five.

Cavitein Name	Cavitein Number	Cavitand-Peptide Linkage <sup>a,b</sup>	Cavitand Foot
N1/Ar/Me	<b>27</b>	<i>cavitand-S-CH<sub>2</sub>CO-peptide</i>	CH <sub>3</sub>
N1G/Ar/Me	<b>30</b>	<i>cavitand-S-CH<sub>2</sub>CO-[Gly]-peptide</i>	CH <sub>3</sub>
N1GG/Ar/Me	<b>32</b>	<i>cavitand-S-CH<sub>2</sub>CO-[Gly-Gly]-peptide</i>	CH <sub>3</sub>
N1GGG/Ar/Me	<b>34</b>	<i>cavitand-S-CH<sub>2</sub>CO-[Gly-Gly-Gly]-peptide</i>	CH <sub>3</sub>
N2/Ar/Me	<b>36</b>	<i>cavitand-S-CH<sub>2</sub>CH<sub>2</sub>CO-peptide</i>	CH <sub>3</sub>
N4/Ar/Me	<b>38</b>	<i>cavitand-S-CH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>CO-peptide</i>	CH <sub>3</sub>
N1/Bzl/Me	<b>41</b>	<i>cavitand-CH<sub>2</sub>S-CH<sub>2</sub>CO-peptide</i>	CH <sub>3</sub>
N1/Bzl/PO <sub>3</sub> H <sub>2</sub>	<b>42</b>	<i>cavitand-CH<sub>2</sub>S-CH<sub>2</sub>CO-peptide</i>	(CH <sub>2</sub> ) <sub>3</sub> PO <sub>3</sub> H <sub>2</sub>
C/Bzl/ PO <sub>3</sub> H <sub>2</sub>	<b>44</b>	<i>cavitand-CH<sub>2</sub>-SCH<sub>2</sub>(Cys)-peptide</i>	(CH <sub>2</sub> ) <sub>3</sub> PO <sub>3</sub> H <sub>2</sub>

<sup>a</sup> Italicized atoms or words represent part of the cavitand or the core amino acid sequence.

<sup>b</sup> “*peptide*” refers to the amino acid sequence -NH-[EELLKKLEELLKKG]-CONH<sub>2</sub>

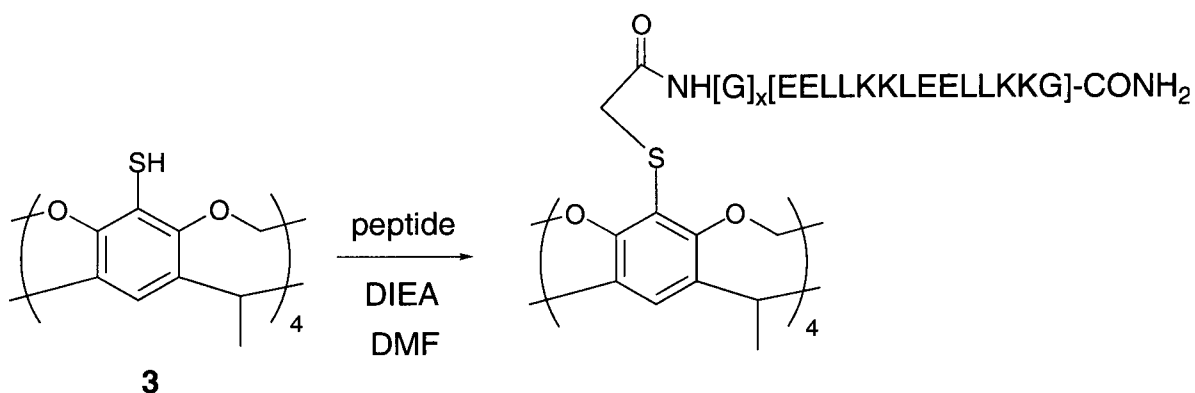
## B. Results and Discussion

### i. Synthesis

#### a. N-terminal Glycine Cavitein Variants

One of the simplest ways to extend the cavitand-peptide linkage is to use the automated peptide synthesizer and add flexible N-terminal glycine residues. Accordingly, peptides **29** ("N1G"), **31** ("N1GG") and **33** ("N1GGG") were synthesized incorporating one, two and three extra N-terminal glycine residues, respectively. In addition, the last step in each synthesis entailed coupling of the N-terminus with chloroacetyl chloride to afford activated chloroacetylated peptides. The syntheses of these peptides is analogous to that of peptide **26** described in Section B.ii of Chapter Four (Figure 4.3). Each of these peptides was then reacted with tetrathiol cavitand **3** in DMF in the presence of DIEA to afford the corresponding caviteins in high yields (Scheme 5.1).

**Scheme 5.1.** Syntheses of Cavitein Variants Including One ("N1G/Ar/Me"), Two ("N1GG/Ar/Me") and Three ("N1GGG/Ar/Me") N-terminal Glycine Residues.



cavitand **3** + peptide **26**  $\longrightarrow$  cavitein **27**: N1/Ar/Me  $x = 0$

cavitand **3** + peptide **29**  $\longrightarrow$  cavitein **30**: N1G/Ar/Me  $x = 1$

cavitand **3** + peptide **31**  $\longrightarrow$  cavitein **32**: N1GG/Ar/Me  $x = 2$

cavitand **3** + peptide **33**  $\longrightarrow$  cavitein **34**: N1GGG/Ar/Me  $x = 3$

N1 = peptide **26**:  $\text{ClCH}_2\text{CO-NH-[EELLKKLEELLKKG]-CONH}_2$

N1G = peptide **29**:  $\text{ClCH}_2\text{CO-NH-[G][EELLKKLEELLKKG]-CONH}_2$

N1GG = peptide **31**:  $\text{ClCH}_2\text{CO-NH-[GG][EELLKKLEELLKKG]-CONH}_2$

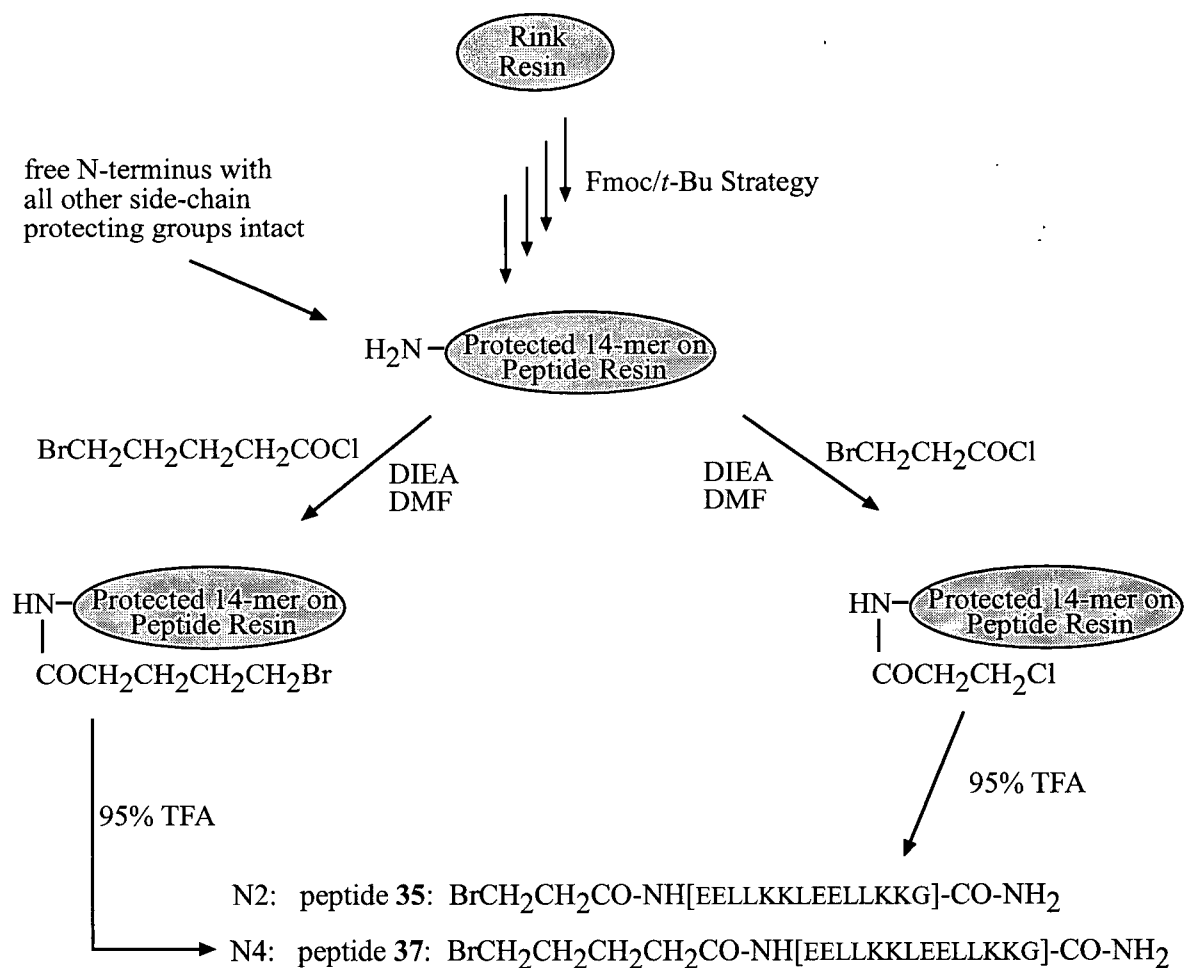
N1GGG = peptide **33**:  $\text{ClCH}_2\text{CO-NH-[GGG][EELLKKLEELLKKG]-CONH}_2$

#### b. N-terminal Methylene Cavitein Variants

The addition of N-terminal glycine residues extends the cavitand-peptide linkage and adds considerable flexibility to this linkage. However, each glycine also incorporates one peptide bond which (1) adds an element of rigidity due to the planarity of the peptide bond and (2) can act as a potential hydrogen bond donor (NH) and acceptor (CO). As a result, the addition of glycine linker units may, in fact, aid in stabilizing and defining the overall structure of the four-helix bundle. We thus incorporated N-terminal methylene linker units to avoid these effects in an effort to better understand the true effect of the template.

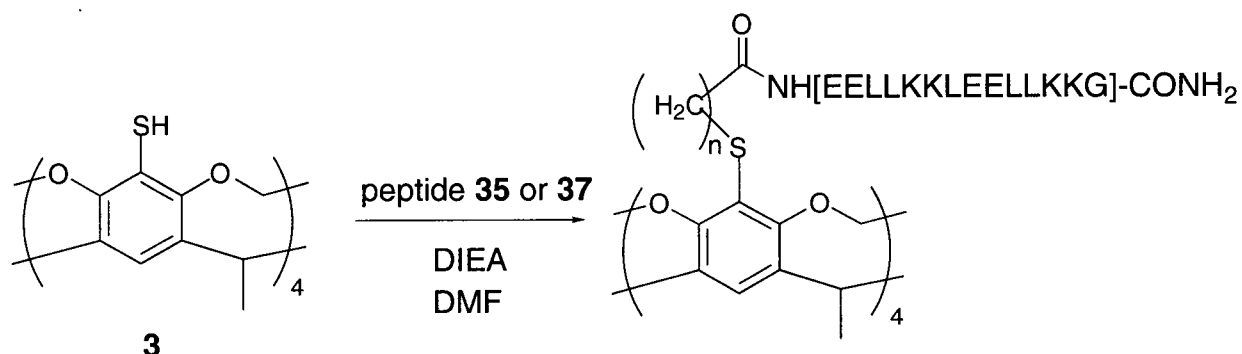
Peptides incorporating two ("N2Br," peptide **35**) and four ("N4Br," peptide **37**) methylene groups were synthesized using the synthetic methodologies discussed in Chapter Three. Each peptide was synthesized on the automated peptide synthesizer as in peptide **26**. However, rather than chloroacetylation as in peptide **26**, both peptides N2Br and N4Br were synthesized by reacting the free N-terminus of the peptide with different activating reagents. For peptide N2, this entailed reacting the N-terminus of the peptide with 3-bromopropionyl chloride in DMF in the presence of DIEA. For peptide N4, this entailed reacting the N-terminus of the peptide with 5-bromovaleryl chloride in DMF in the presence of DIEA (Figure 5.2). Peptides N2Br and N4Br were subsequently reacted with tetrathiol **3** (DMF, DIEA, 16 h, 25 °C) to afford caviteins "N2/Ar/Me" and "N4/Ar/Me," respectively (Scheme 5.2).

**Figure 5.2.** Schematic Representation of the Syntheses of Activated Peptides With Two (“N2Br”) and Four (“N4Br”) Methylene Linker Units.





**Scheme 5.2.** Syntheses of Cavitein Variants Including Two (“N2/Ar/Me”) and Four (“N4/Ar/Me”) Methylene Linker Units.

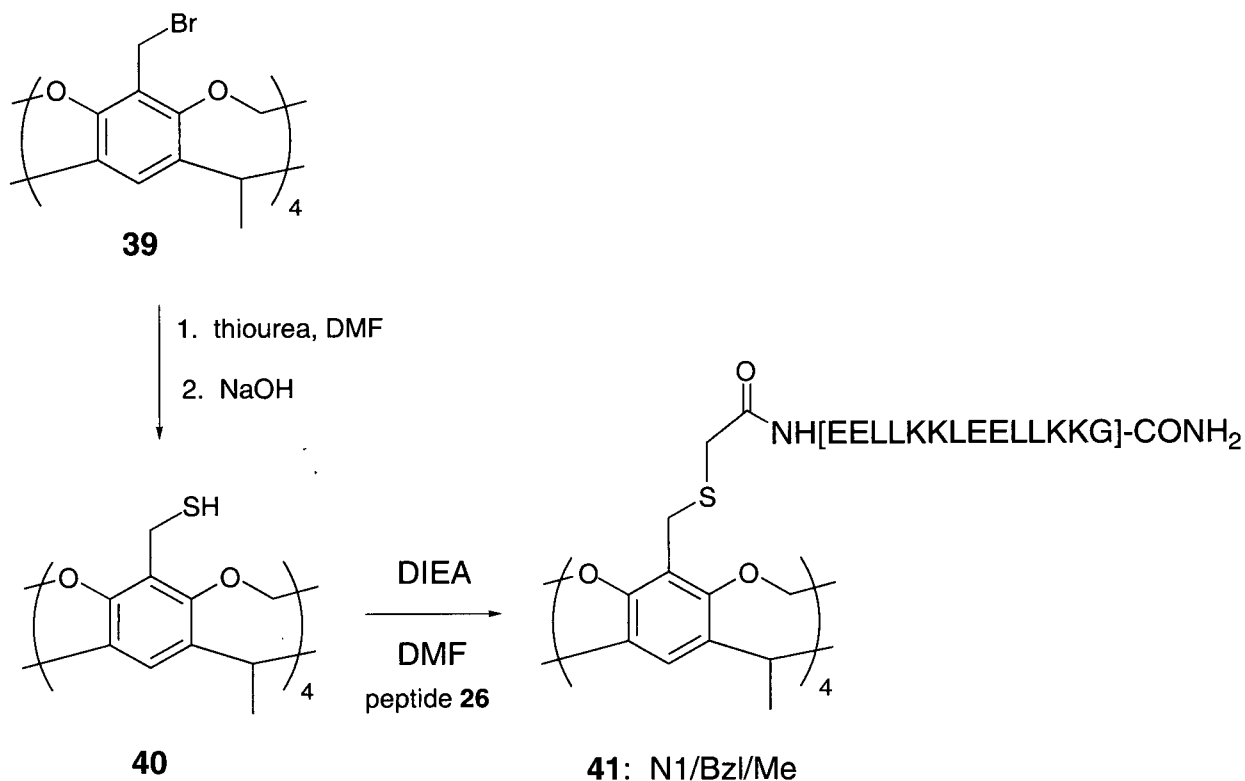


cavitand **3** + peptide **35** (N2)  $\longrightarrow$  cavitein **36**: N2/Ar/Me  $n = 2$

cavitand **3** + peptide **37** (N4)  $\longrightarrow$  cavitein **38**: N4/Ar/Me  $n = 4$

An alternative method to adding an extra methylene linker unit is to employ the known compound benzylthiol cavitand **40** for cavitein synthesis. Cram originally synthesized cavitand **40**,<sup>3</sup> however, a more direct synthetic route is accessible via Sorrell's direct bromination method to first generate the known compound benzylbromide **39**.<sup>4</sup> Subsequent treatment with thiourea (as in the synthesis of compound **4**, Chapter Two) and NaOH afforded benzylthiol **40** (Scheme 5.3).

**Scheme 5.3.** Schematic Representation of the Synthesis of Cavitein **41** (“N1/Bzl/Me”).

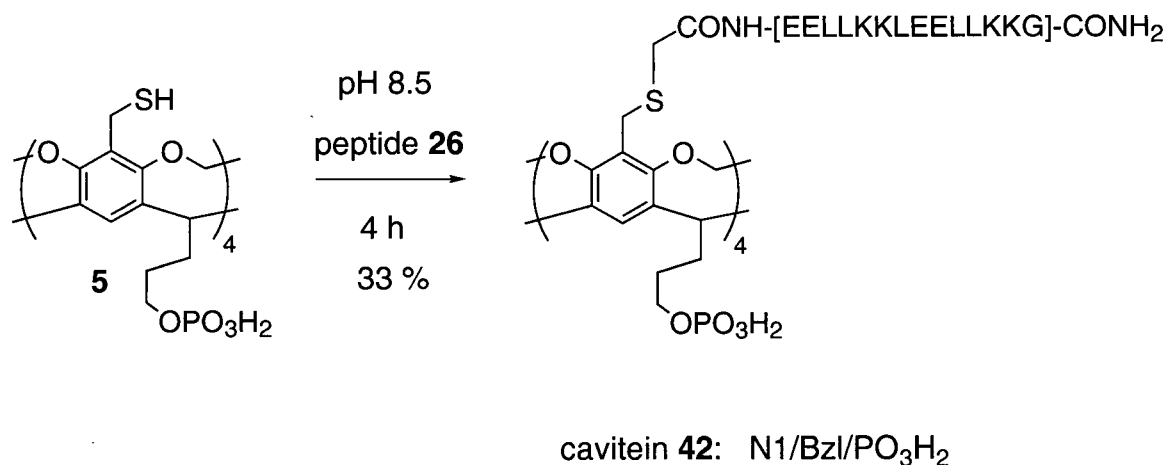


Coupling of peptide **26** (“N1”, Chapter Four) to benzylthiol cavitand **40** afforded cavitein **41** (“N1/Bzl/Me,” Scheme 5.3) in 35% yield. Note that this cavitein is a constitutional isomer of N2/Ar/Me. Interestingly, their structural properties are quite different and will be discussed later in Section B.ii.b of this chapter.

c. Phosphate-Footed Cavitein Variants

In order to evaluate the effect of the cavitand feet, we synthesized phosphate-footed caviteins via two different synthetic methods (Schemes 5.4 & 5.5). In the first method, phosphate-footed tetrathiol **5** (Chapter Two) was coupled with peptide **26** ("N1", Chapter Four) in aqueous buffer (pH 8.5, 4 h) to afford phosphate-footed cavitein **42** ("N1/Bzl/PO<sub>3</sub>H<sub>2</sub>") in 33% yield (Scheme 5.4).

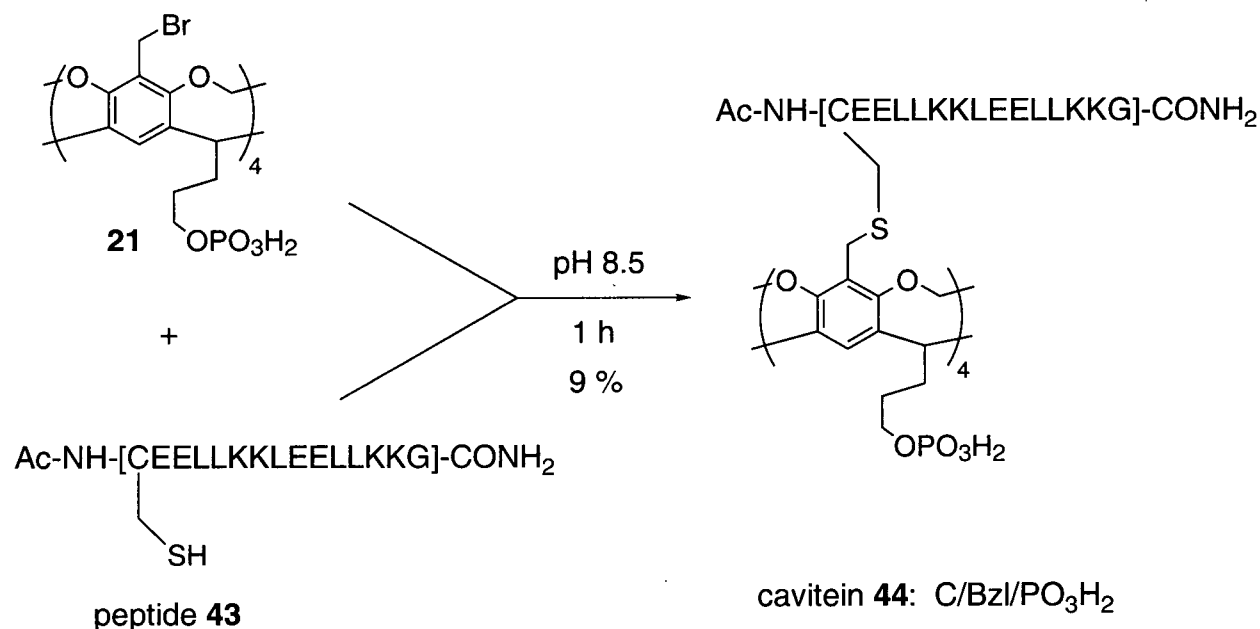
**Scheme 5.4.** Synthesis of Phosphate-Footed Cavitein **42** ("N1/Bzl/PO<sub>3</sub>H<sub>2</sub>").



In the second synthetic route to a phosphate-footed cavitein, electrophilic benzyl bromide cavitand **21** (Chapter Two) was coupled with peptide **43** which contains a highly nucleophilic *N*-terminal cysteine residue (the side-chain of cysteine possesses a reactive thiol moiety). This method had been used previously in our laboratory for coupling methyl-footed benzyl bromide **40** with a 15-residue cysteine-containing peptide.<sup>5</sup>

Peptide **43** was synthesized on the automated peptide synthesizer where the last cycle simply entailed acetylation of the N-terminus (as in peptide **28**, Chapter Four). In this case, cleavage from the solid support employed a mixture of TFA (95%), H<sub>2</sub>O (2.5%) and 1,2-ethanedithiol (2.5%) to “scavenge” for the trityl protecting group used for the cysteine residue.

**Scheme 5.5.** Synthesis of Phosphate-Footed Cavitein **44** (“C/Bzl/PO<sub>3</sub>H<sub>2</sub>”).



Reaction of peptide **43** with cavitand **21** in aqueous buffer (pH 8.5, 1 h) afforded phosphate-footed cavitein **44** (“C/Bzl/PO<sub>3</sub>H<sub>2</sub>”) in 9% yield (Scheme 5.5). Note that this cavitand-peptide linkage is fundamentally different than all of the other cavitand-peptide linkages presented in this thesis: each peptide in C/Bzl/PO<sub>3</sub>H<sub>2</sub> is linked to the cavitand via the side-chain of the *N*-terminal cysteine residue.

Recall that tetra-phosphate cavitand **21** was inseparable from its tris-phosphate derivative (Chapter Two, Section B.iv). In the synthesis of C/Bzl/PO<sub>3</sub>H<sub>2</sub>, the tetra-phosphate derivative was easily separable from the tris-phosphate derivative. The structural properties of cavitin C/Bzl/PO<sub>3</sub>H<sub>2</sub> will not be discussed in this thesis since the effect of this cavitand-peptide linkage is the focus of another member of our research group.<sup>5</sup> The synthesis of C/Bzl/PO<sub>3</sub>H<sub>2</sub> was presented to simply illustrate the utility cavitand **21** in cavitin synthesis. In addition, the synthesis of C/Bzl/PO<sub>3</sub>H<sub>2</sub> demonstrates that further derivatization of impure cavitand **21** can indeed facilitate separation of the tris- and tetra-phosphate derivatives. The effect of phosphate feet on cavitin structure will be investigated using N1/Bzl/PO<sub>3</sub>H<sub>2</sub>.

Note that the syntheses of peptide **43**, N1/Bzl/PO<sub>3</sub>H<sub>2</sub> and C/Bzl/PO<sub>3</sub>H<sub>2</sub> have been recently published.<sup>6</sup>

## ii. Structural Characterization

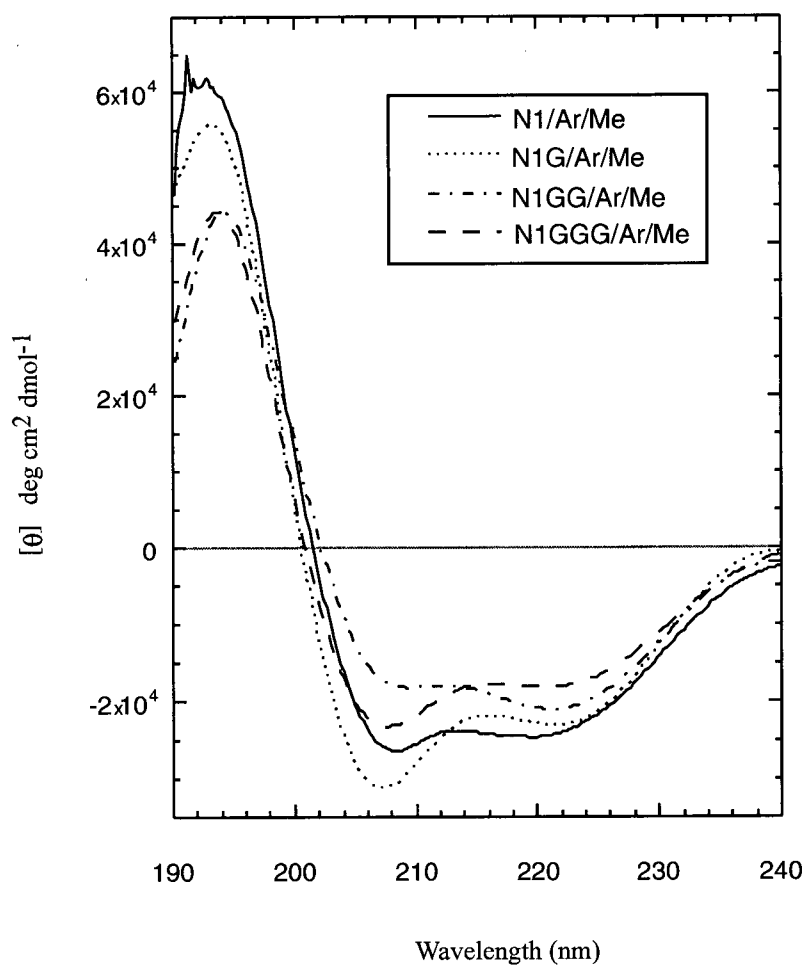
### a. Comparison of Glycine Variants

#### 1. Far-UV CD Spectroscopy

The far-UV CD spectra of cavitins containing one, two and three extra N-terminal glycine residues are shown in Figure 5.3. It is evident that each cavitin remains helical despite the addition of one to three N-terminal glycines to the amino acid sequence. The absolute helicity, as measured by  $[\theta]_{222}$ , decreased as the number of glycines increase from one to three. However, this is partly due to the helix length-dependent equation used in calculating molar

ellipticity *per residue*. In calculating  $[\theta]_{222}$ , each glycine is “counted” as an additional residue and therefore even if the same helical structure is present in the glycine-enriched caviteins, the value of  $[\theta]_{222}$  would still be reduced. If each glycine is not considered to be part of the  $\alpha$ -helical structure, the  $[\theta]_{222}$  values are within 3 000 deg cm<sup>2</sup>dmol<sup>-1</sup> of each other (not shown).

**Figure 5.3.** Far-UV CD Spectra of Caviteins N1/Ar/Me, N1G/Ar/Me, N1GG/Ar/Me, N1GGG/Ar/Me in pH 7.0 Phosphate Buffer at 25 °C. Each Cavitein is at a Concentration of ~ 30  $\mu$ M.



One interesting aspect of these far-UV CD spectra are their shapes. Note that N1GG/Ar/Me has a  $[\theta]_{222}/[\theta]_{208}$  ratio  $> 1.0$ , while the other caviteins possess ratios  $< 1.0$ . Some researchers interpret  $[\theta]_{222}/[\theta]_{208}$  ratios  $> 1.0$  to be indicative of coiled-coil-like structures, and ratios  $< 1.0$  to be indicative of non-interacting helices.<sup>7</sup> Is this indicative of coiled-coil-like structure in just this one case? If this is true, do all of the other caviteins with  $[\theta]_{222}/[\theta]_{208}$  ratios  $< 1.0$  possess non-interacting helices?

Firstly, the results presented in Chapter Four and later in this chapter clearly demonstrate that helices in all caviteins self-associate into compact, stable helical structures and therefore  $[\theta]_{222}/[\theta]_{208}$  ratios  $< 1.0$  are, in our case, not indicative of non-interacting helices. In the case of N1GG/Ar/Me, it is possible that it exists in a coiled-coil-like conformation; however, it is more likely that the aromatic cavitand template is contributing (positively or negatively) to the far-UV CD spectra of the caviteins and that this effect is perturbed with the added N-terminal glycines. Indeed, it is well-known that aromatic side-chains can contribute significantly to the far-UV CD spectrum of a proteins.<sup>8</sup> Recently, Baldwin and co-workers have shown that N-terminal tyrosines and tryptophans contribute positively ( $+5000 \text{ deg cm}^2/\text{dmol}$ ) and negatively ( $-3000 \text{ deg cm}^2/\text{dmol}$ ), respectively, to  $[\theta]_{222}$  values of single-stranded  $\alpha$ -helices.<sup>9</sup> In the same study, they found that inserting between one and three glycines diminished the effect of those residues on  $[\theta]_{222}$ .

With respect to the caviteins presented in this thesis, it is thus likely that the cavitand chromophore contributes significantly to the far-UV spectrum of each cavitein, and that this contribution may be different for each cavitein. For this reason, interpretation of absolute  $[\theta]_{222}$  values and  $[\theta]_{222}/[\theta]_{208}$  ratios will be kept to a minimum. Much more information regarding their structures can be gleaned from examination of guanidine hydrochloride denaturation,

sedimentation equilibrium and ANS binding experiments as well as their near-UV CD and NMR spectra.

## 2. Oligomeric State of the Gly Variants

The oligomeric state of each Gly variant was studied using sedimentation equilibrium and GuHCl denaturations at different cavitein concentrations. Although each cavitein presented herein possessed concentration independent CD spectra over at least a fifty-fold concentration range, this was not taken as an indication of monomeric states in light of the results presented in Chapter Four.

Sedimentation equilibrium studies were performed at 25 °C in 50 mM pH 7 phosphate buffer and the results are shown in Table 5.2. In addition, the fits of the raw data to the theoretical curves are shown in the Experimental Section in Figures 5.28 – 5.33 for various conditions and concentrations of N1G/Ar/Me, N1GG/Ar/Me and N1GGG/Ar/Me. In all cases, the best fit to the data was achieved using a model describing a single species in solution. Interestingly, however, experimental data on N1G/Ar/Me gave molecular weights near that of a dimer, while those on both N1GG/Ar/Me and N1GGG/Ar/Me gave molecular weights near that of a monomer.



**Table 5.2.** Experimentally Estimated Molecular Weights Determined by Sedimentation Equilibrium For All Cavitein Variants at 25 °C in 50 mM pH 7 Phosphate Buffer, Unless Otherwise Noted.

Cavitein	Concentration (μM)	Experimentally Estimated Molecular Weight (Da)	Calculated Molecular Weight (Da)	Predominant Species
N1G/Ar/Me	3	18 800 ± 1000	7788	Dimer
	30	18 700 ± 300		Dimer
N1GG/Ar/Me	30	10 000 ± 600	8019	Monomer
	30 (in 7.2 M GuHCl)	10 800 ± 1400		Monomer
N1GGG/Ar/Me	30	8 200 ± 400	8250	Monomer
	300	8 900 ± 200		Monomer
N2/Ar/Me	30	10 800 ± 100	7615	Monomer
N4/Ar/Me	30	22 100 ± 500	7725	Dimer
N1/Bzl/Me	30	19 300 ± 100	7615	Dimer
	300	19 400 ± 100		Dimer
N1/Bzl/PO <sub>3</sub> H <sub>2</sub>	30	11 600 ± 400	8112	Monomer

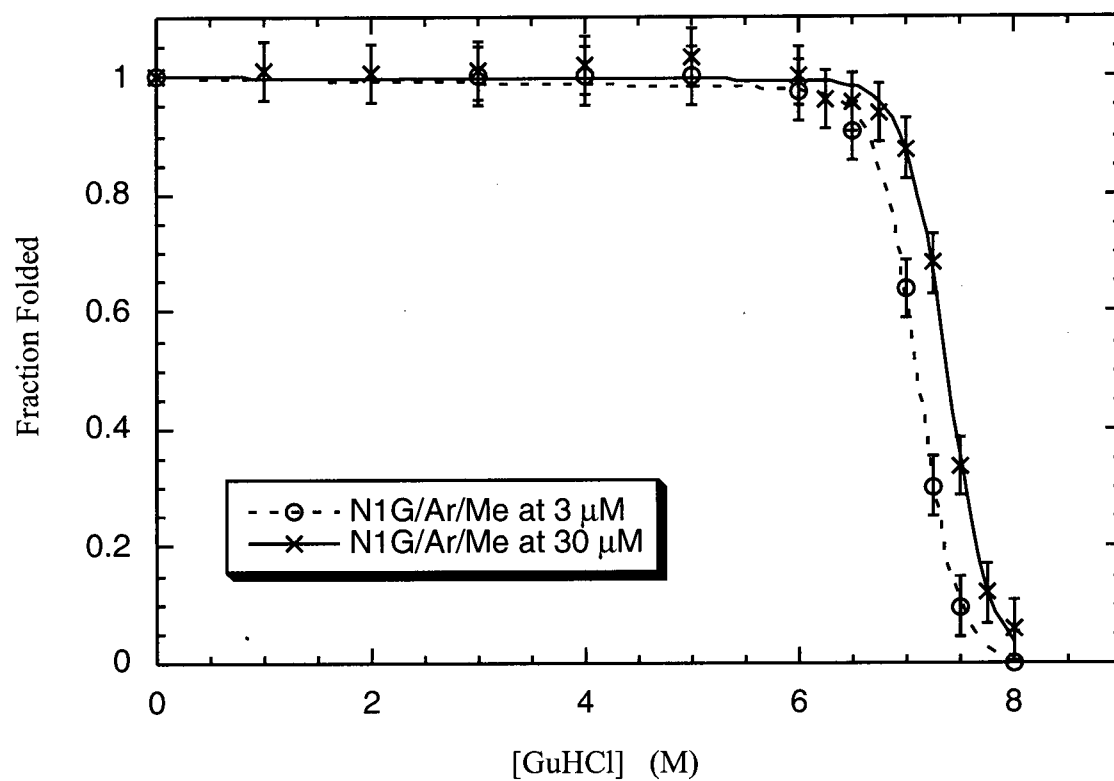
Note that the estimated molecular weights for each cavitein vary considerably when compared to the expected monomeric molecular weight. We therefore conducted an experiment using N1GG/Ar/Me in 7.2 M GuHCl – a concentration at which this cavitein is fully unfolded (see below) and is therefore expected to be monomeric. Under these denaturing conditions, the molecular weight of N1GG/Ar/Me was found to be 10 800 ± 1400 Da. Hence, this value can be considered a “baseline” value for the monomeric species which suggests that both N1GG/Ar/Me

and N1GGG/Ar/Me are monomeric under the conditions listed while N1G/Ar/Me is dimeric.

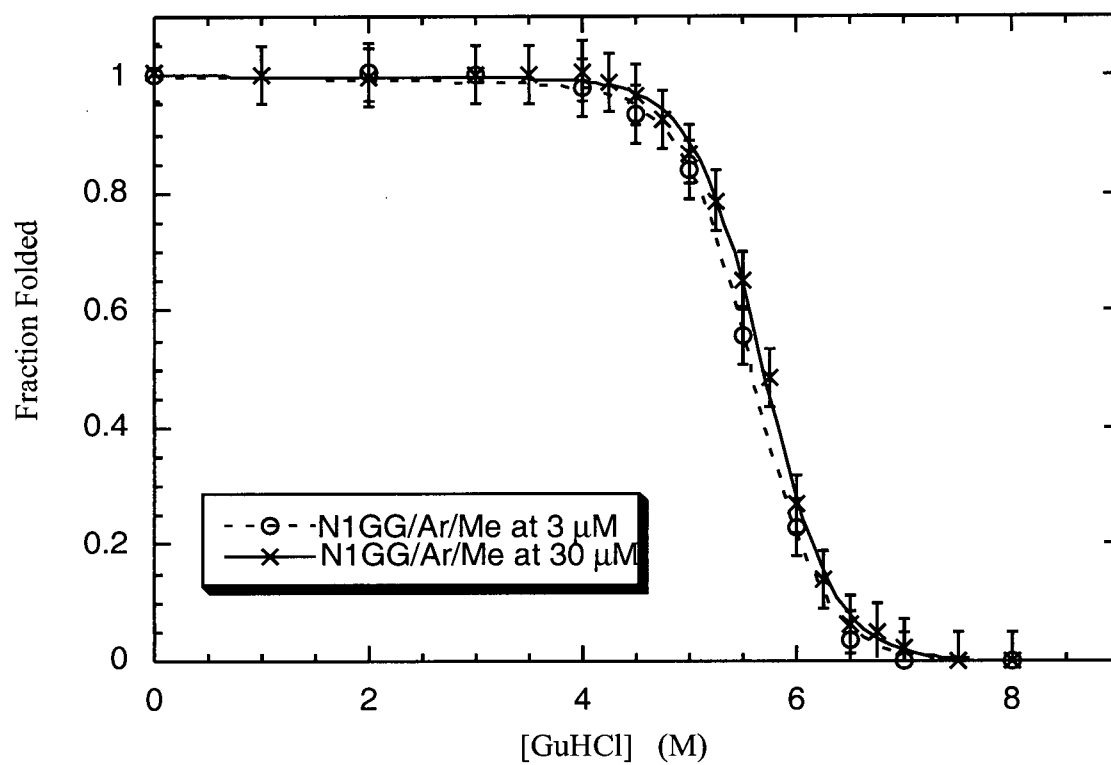
As mentioned in Chapter Four, many possible errors could account for the discrepancy between the estimated and calculated molecular weights. Refer back to Chapter Four, Section B.iii.c for this discussion.

Further insight into the nature of the oligomeric state can be gleaned from analysis of GuHCl-induced denaturation curves at different concentrations of cavitein. The stabilities of each Gly cavitein variant were studied using GuHCl at pH 7.0, 25 °C and monitored by CD spectroscopy at 222 nm at cavitein concentrations of 3 and 30  $\mu$ M. The denaturation curves for N1G/Ar/Me, N1GG/Ar/Me and N1GGG/Ar/Me are shown in Figures 5.4, 5.5 and 5.6 respectively. Since each of the caviteins possessed different  $[\theta]_{222}$  values under non-denaturing conditions, their ellipticities were normalized to one for comparison's sake.

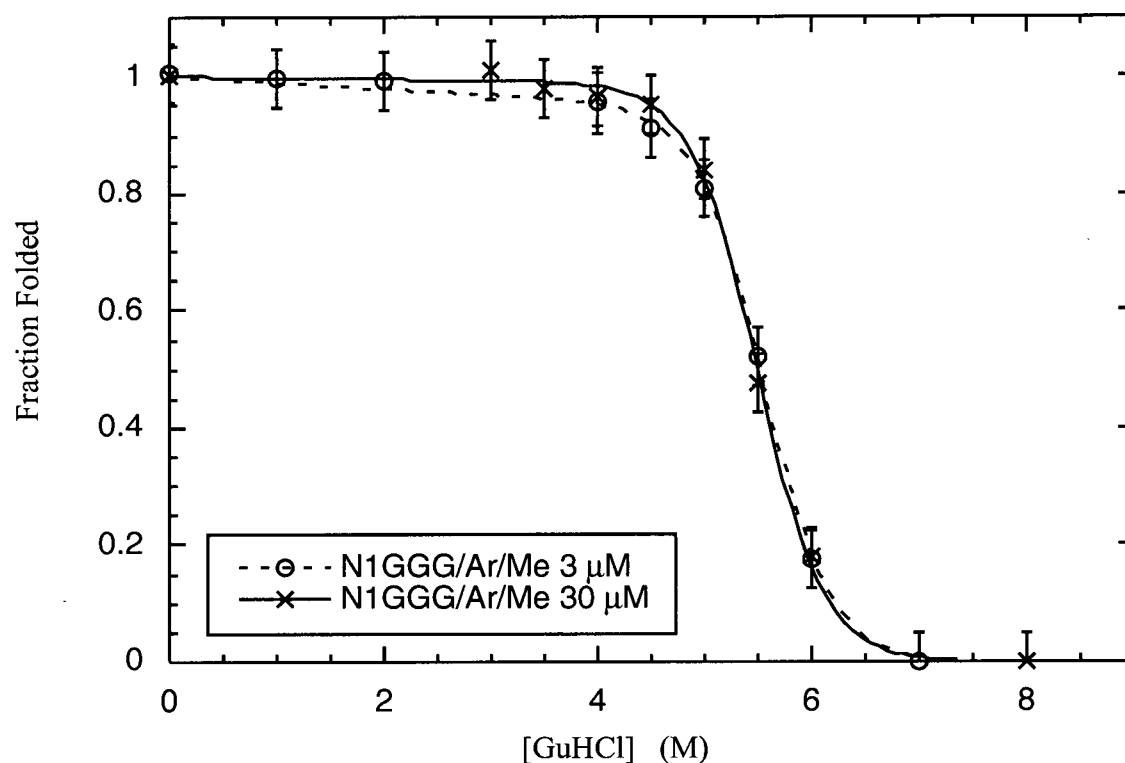
**Figure 5.4.** Guanidine Hydrochloride-Induced Denaturation Curve for N1G/Ar/Me at 3 and 30  $\mu\text{M}$  at pH 7 and 25  $^{\circ}\text{C}$ .



**Figure 5.5.** Guanidine Hydrochloride-Induced Denaturation Curve for N1GG/Ar/Me at 3 and 30  $\mu$ M at pH 7 and 25  $^{\circ}$ C.



**Figure 5.6.** Guanidine Hydrochloride-Induced Denaturation Curve for N1GGG/Ar/Me at 3 and 30  $\mu$ M at pH 7 and 25  $^{\circ}$ C.



It is evident that N1G/Ar/Me exhibits concentration dependent denaturation curves while both N1GG/Ar/Me and N1GGG/Ar/Me display concentration independent denaturation curves. These results are consistent with the sedimentation equilibrium data that suggest that, in the absence of GuHCl, N1G/Ar/Me is dimer while both N1GG/Ar/Me and N1GGG/Ar/Me are predominantly monomers. In addition, the concentration dependence observed in the GuHCl denaturation curves for N1G/Ar/Me suggests that dimerization alters the packing of the four-helix bundle and contributes significantly to its stability and structure. In other words, it is likely that the N1G/Ar/Me dimer is *not* the result of two, independently folded four-helix bundles (as was hypothesized for N1/Ar/Me). This notion is built upon throughout this chapter.

Note that despite characterization of each cavitein by electrospray mass spectrometry, no evidence for dimerization was observed.<sup>10</sup>

### 3. Stabilities Towards GuHCl

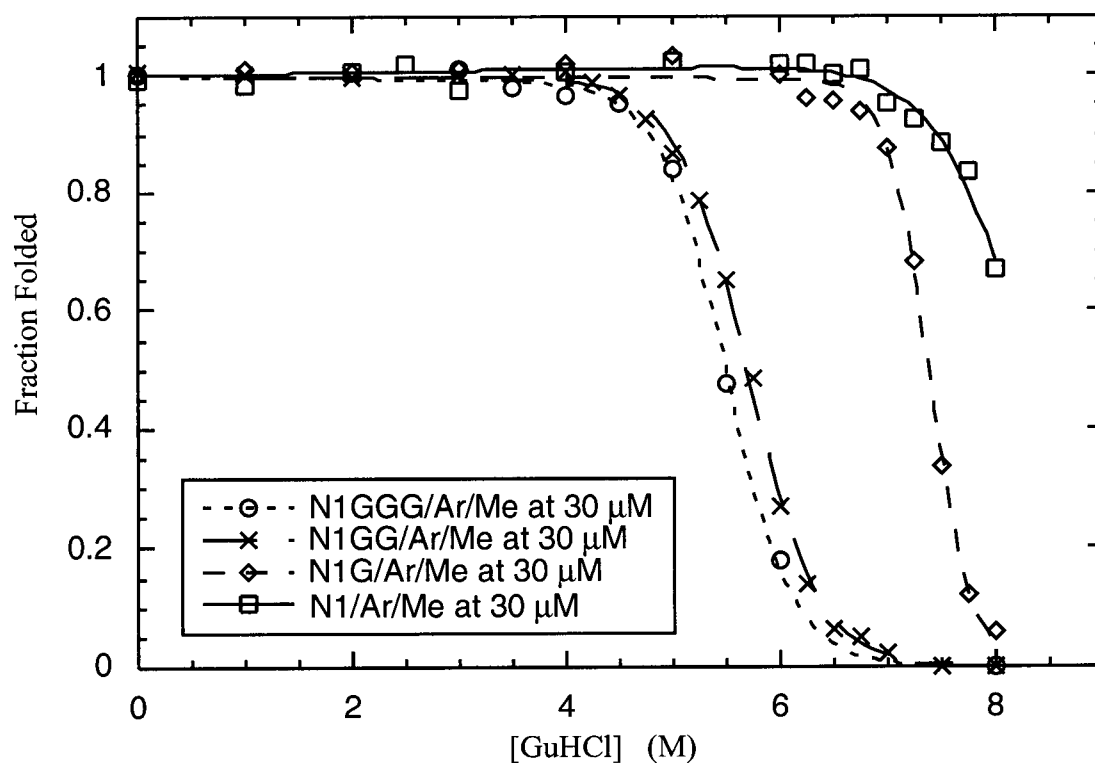
From Figures 5.4-5.7, it is evident that some of the Gly variants possess different stabilities. When evaluating the stabilities of very similar proteins towards GuHCl, it is typical to take the concentration of GuHCl required to half-unfold the protein ( $[\text{GuHCl}]_{1/2}$ ) as a crude measure of their stabilities. However, a more rigorous and common method of estimating the stabilities of closely related proteins is to use the “linear extrapolation method.”<sup>11</sup> In this method, the free energy of unfolding ( $\Delta G_{\text{obsd}}$ ) is assumed to vary linearly with denaturant concentration. Therefore, the free energy of unfolding can be extrapolated back to zero denaturant concentration to give an apparent free energy for unfolding in the absence of denaturant ( $\Delta G^{\circ}_{\text{H}_2\text{O}}$ ). This can be described by equation 5.1:

$$\Delta G_{\text{obsd}} = \Delta G^{\circ}_{\text{H}_2\text{O}} - m[\text{GuHCl}] \quad (5.1)$$

where  $m$  is the slope of the unfolding transition and provides a qualitative estimate of the cooperativity associated with the unfolding transition. We used the expanded method of Santoro and Bolen<sup>12</sup> in which, by assuming the unfolding transition is a reversible, two-state process, the data can be fit using non-linear least squares methods to derive  $\Delta G^{\circ}_{\text{H}_2\text{O}}$  (see Experimental Section). The calculated extrapolated values for  $\Delta G^{\circ}_{\text{H}_2\text{O}}$ ,  $m$ , and the denaturation

midpoint for each of the caviteins at different concentrations are listed in Table 5.3. Note that although no unfolding transition was observed for N1/Ar/Me or N1G/Ar/Me, their denaturation profiles were also analyzed by the same non-linear least squares analysis for comparative purposes. This data was not reported in Chapter Four for N1/Ar/Me due to the large errors associated with its calculation (see below).

**Figure 5.7.** Guanidine Hydrochloride-Induced Denaturation Curves for N1/Ar/Me, N1G/Ar/Me, N1GG/Ar/Me and N1GGG/Ar/Me at 30  $\mu$ M in pH 7.0 Phosphate Buffer at 25  $^{\circ}$ C. Error Bars Have Been Omitted For Clarity.



**Table 5.3.** Guanidine Hydrochloride-Induced Denaturation Data Calculated for the Gly-Cavitein Variants.

Cavitein	Cavitein Concentration ( $\mu\text{M}$ )	$[\text{GuHCl}]_{1/2}$ <sup>a</sup> (M)	$m$ kcal mol <sup>-1</sup> M <sup>-1</sup>	$\Delta G^{\circ}_{\text{H}_2\text{O}}$ kcal/mol
N1/Ar/Me	3	> 8.0	$-1.3 \pm 0.2^b$	$-10.4 \pm 1.1^b$
N1/Ar/Me	30	> 8.0	$-1.4 \pm 0.2^b$	$-11.9 \pm 1.5^b$
N1G/Ar/Me	3	$7.1 \pm 0.1$	$-3.2 \pm 0.3^b$	$-22.6 \pm 2.3^b$ (~40) <sup>c</sup>
N1G/Ar/Me	30	$7.4 \pm 0.1$	$-3.1 \pm 0.3^b$	$-22.9 \pm 1.9^b$ (~40) <sup>c</sup>
N1GG/Ar/Me	3	$5.6 \pm 0.1$	$-1.8 \pm 0.1$	$-10.0 \pm 0.5$
N1GG/Ar/Me	30	$5.7 \pm 0.1$	$-1.7 \pm 0.1$	$-9.9 \pm 0.3$
N1GGG/Ar/Me	3	$5.5 \pm 0.1$	$-1.8 \pm 0.1$	$-9.8 \pm 0.6$
N1GGG/Ar/Me	30	$5.5 \pm 0.1$	$-1.9 \pm 0.1$	$-10.8 \pm 0.5$

<sup>a</sup>  $[\text{GuHCl}]_{1/2}$  is the concentration of GuHCl required to unfold half of the helical structure.

<sup>b</sup> A full unfolding transition was not observed for N1/Ar/Me or N1G/Ar/Me. Therefore, these values likely contain larger errors than those calculated by the computer program and listed in this table (see Experimental Section).

<sup>c</sup> This value was generated from an equation taking into account the dimerization of N1G/Ar/Me (See Experimental Section).

From Table 5.3, it is evident that as the number of N-terminal glycine linker units are added between the cavitand and the peptide sequence, the concentration of GuHCl required to half-unfold the caviteins is successively reduced until N1GG/Ar/Me and N1GGG/Ar/Me which display similar GuHCl unfolding midpoints,  $m$  values and  $\Delta G^{\circ}_{\text{H}_2\text{O}}$  values. This suggests that the effect of the rigid template may be near a minimum when the number of flexible glycine



residues is two or greater. However, the NMR data presented later in this chapter prove this not to be the case.

One striking result listed in Table 5.3 is the greater apparent stability of N1G/Ar/Me. Its  $\Delta G^{\circ}_{\text{H}_2\text{O}}$  value is  $\sim 10$  kcal/mol more stable than the other three caviteins at both 3 and 30  $\mu\text{M}$ . Even when considering a large potential error (i.e., even 5 kcal/mol) associated with the N1/Ar/Me and N1G/Ar/Me data (see below), N1G/Ar/Me remains the most stable of the Gly-variants. The unfolding transition of N1G/Ar/Me possesses the greatest cooperativity in its unfolding transition and this is reflected in its higher  $m$  value and, ultimately, its  $\Delta G^{\circ}_{\text{H}_2\text{O}}$  value.

Note that denaturant  $m$  values have also been correlated well with solvent accessible surfaces areas; larger  $m$  values generally correspond to larger protein surfaces (i.e., protein size).<sup>13</sup> This is consistent with the above-mentioned data suggesting that N1G/Ar/Me is a dimer with a different topology to the other Gly-variants.

When interpreting  $\Delta G^{\circ}_{\text{H}_2\text{O}}$  values derived from extrapolation, errors increase drastically as the length of extrapolation (i.e., denaturation midpoint) is increased.<sup>12</sup> Small errors in  $m$  can be translated into large errors in  $\Delta G^{\circ}_{\text{H}_2\text{O}}$ . In addition, there is no theoretical basis for making the assumption that the free energy of unfolding varies linearly with denaturant concentration; this assumption may not hold true at high concentrations of denaturant thereby introducing errors. Moreover, there are additional errors associated with the analysis in cases where a full unfolding transition was not observed (see Experimental Section). In light of all the errors involved, we limit our interpretation of the above results to: (1) N1GG/Ar/Me and N1GGG/Ar/Me have similar stabilities, (2) N1G/Ar/Me is clearly the most stable of the four caviteins (although possibly a result of its dimeric structure, see below), and (3) the stability of N1/Ar/Me is uncertain but its stability is certainly less than that of N1G/Ar/Me.

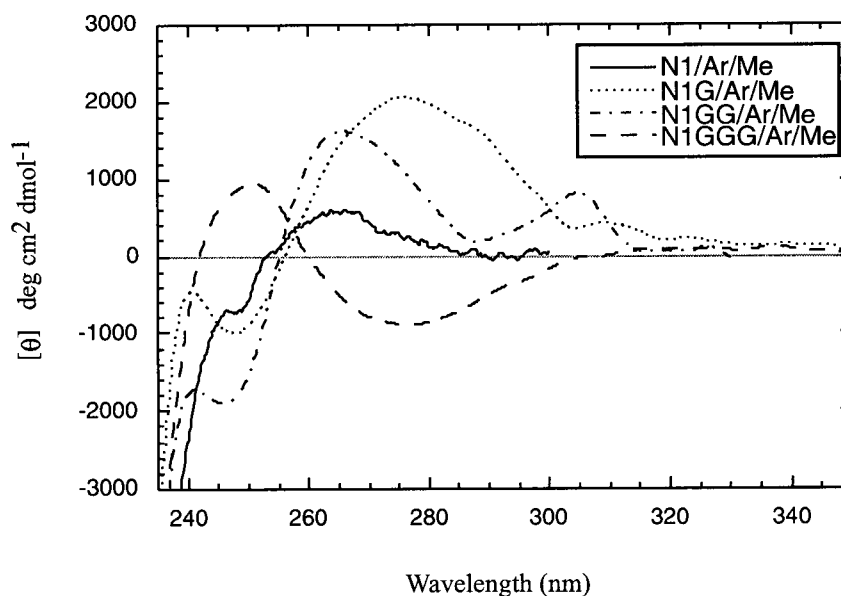
Recall that N1/Ar/Me was observed to exist as monomer-dimer equilibrium mixture under the conditions of the sedimentation equilibrium experiment. However, N1/Ar/Me exhibited concentration independent stabilities towards GuHCl (although a full unfolding transition was not observed) and therefore it appeared that dimerization was not affecting the structure of the four-helix bundle. Indeed, sedimentation equilibrium data in the presence of 6 M GuHCl, 2 M NaCl or 10% methanol (conditions under which N1/Ar/Me is fully helical) demonstrated that its structure is monomeric. This is in contrast to the concentration dependent denaturation curves observed for N1G/Ar/Me. The existence of non-overlapping denaturation curves suggests that its structure remains dimeric until unfolding occurs. Whereas N1/Ar/Me goes through a dimer-to-monomer-to-unfolded transition, N1G/Ar/Me likely goes through a more direct dimer-to-unfolded transition. In other words, the dimer is an inherent component of the structure of N1G/Ar/Me. Hence, it is likely that N1G/Ar/Me is *not* forming a four-helix bundle but a stable helical structure of unknown topology. More detailed structural information will be required to identify its structure and attempts to crystallize N1G/Ar/Me are underway for structural determination by X-ray crystallography.

#### 4. Near-UV CD Spectroscopy

As mentioned in Chapter Four (Section B.iii.f), near-UV CD absorptions suggest the existence of non-averaged structural elements (i.e., native-like structure) near that aromatic chromophore. The near-UV CD spectra of the Gly-cavitein variants are shown in Figure 5.8, along with the spectrum of N1/Ar/Me presented in Chapter Four.

It is clear that each of the Gly-cavitein variants display enhanced absorptions in comparison to N1/Ar/Me. This result suggests that in comparison to N1/Ar/Me, each Gly-cavitein variant possesses even more specific and less dynamic structural elements near the cavitand-peptide linkage. This is interesting considering the long three-Gly linker present in N1GGG/Ar/Me. Clearly, the peptides' proximity to the cavitand is not crucial in inducing specific structural elements near the cavitand.

**Figure 5.8.** Near-UV CD Spectra of Gly-Cavitein Variants in pH 7.0 Phosphate Buffer at 25 °C.



The strongest absorption was observed for N1G/Ar/Me suggesting this linker may be ideal in inducing non-averaged structures near the cavitand-peptide linkage (i.e., the cavitand chromophore is in an asymmetric environment). However, this enhanced absorption may also be due, in part, to the closer proximity of the asymmetric helices to the cavitand in comparison to the N1GG/Ar/Me and N1GGG/Ar/Me since some of these Gly linkers may not take part in

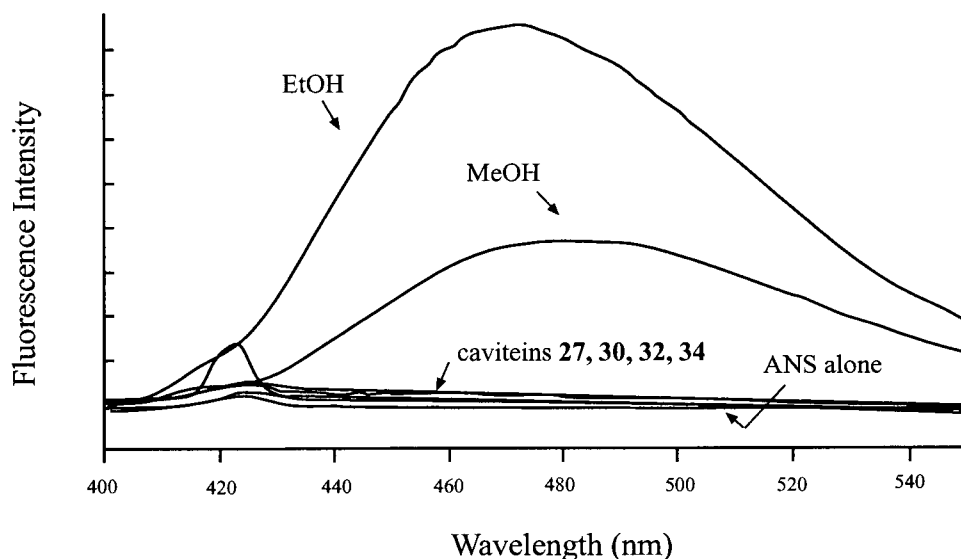
the  $\alpha$ -helix (i.e., glycine is a known helix-breaker). In addition, the enhanced absorption may be a result of the dimeric structure of N1G/Ar/Me inducing structural elements not available to other monomeric caviteins.

Interestingly, the near-UV CD spectrum of N1GGG/Ar/Me is of the opposite sign in comparison to the other three caviteins. Although interpretation of this result can only be speculative, the three-Gly linker may allow the helices to pack in an opposite sense to that of the caviteins with shorter linkers. For example, the helices may be packing at  $20^\circ$  with respect to one another in N1GGG/Ar/Me, but  $-20^\circ$  in the other three caviteins. More detailed structural information would be required to answer this question and is beyond the scope of this thesis.

## 5. ANS Binding

The hydrophobic dye ANS is a convenient probe for detecting molten globule-like states (e.g., see Chapter Four, Section B.iii.i). ANS has been found to bind exposed hydrophobic surfaces of dynamic molten globule states, and not native or unfolded states. No binding of ANS was detected for any of the Gly-cavitein variants under the experimental conditions (Figure 5.9). These results suggest that each cavitein is less dynamic than the molten globule state and possesses native-like characteristics to their tertiary structures. Interestingly, despite its long three-Gly linker, native-like characteristics are still retained in the structure of N1GGG/Ar/Me.

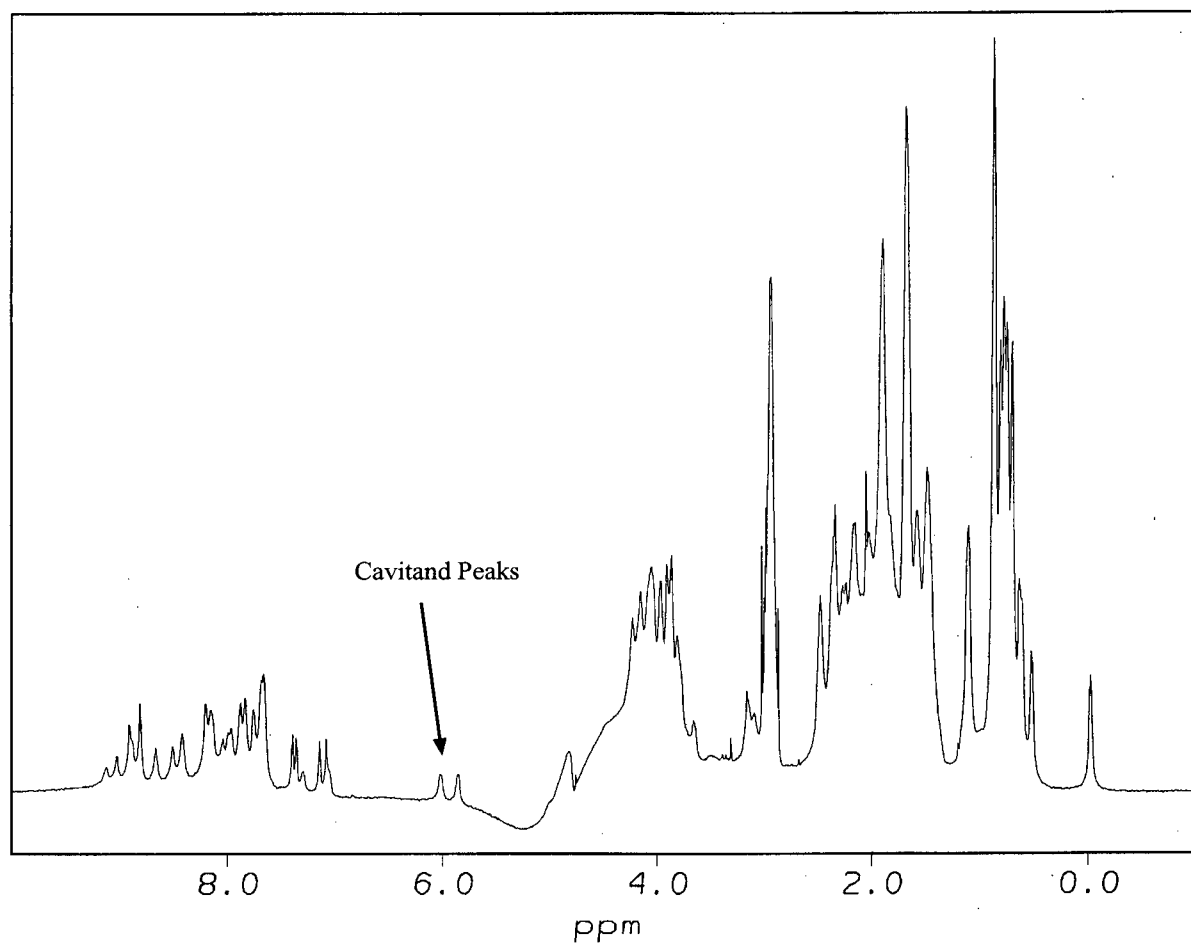
**Figure 5.9.** Fluorescence Emission Spectra of 1  $\mu$ M ANS in the Presence of 95% Ethanol, 100% Methanol and 50  $\mu$ M solutions of N1/Ar/Me, N1G/Ar/Me, N1GG/Ar/Me and N1GGG/Ar/Me at 25  $^{\circ}$ C in pH 7.0 Phosphate Buffer.



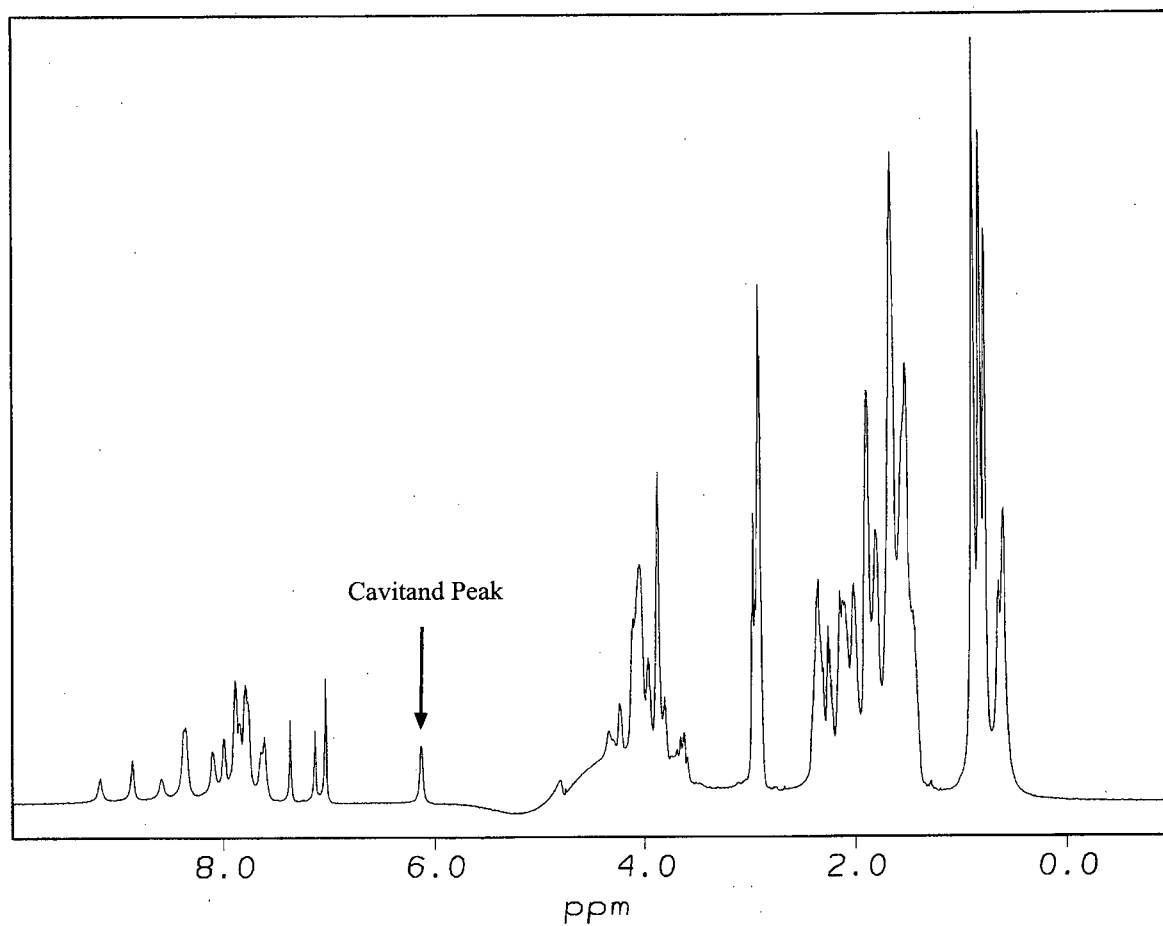
## 6. $^1\text{H}$ NMR Chemical Shift Dispersion

Recall that native proteins typically possess significant chemical shift dispersion associated with their  $^1\text{H}$  NMR spectra (Chapter Four, Section B.iii.g). We acquired 1D  $^1\text{H}$  NMR spectra for the Gly-variants and these spectra are shown in Figures 5.10 – 5.12 with stacked-plot expansions shown in Figures 5.13 and 5.14.

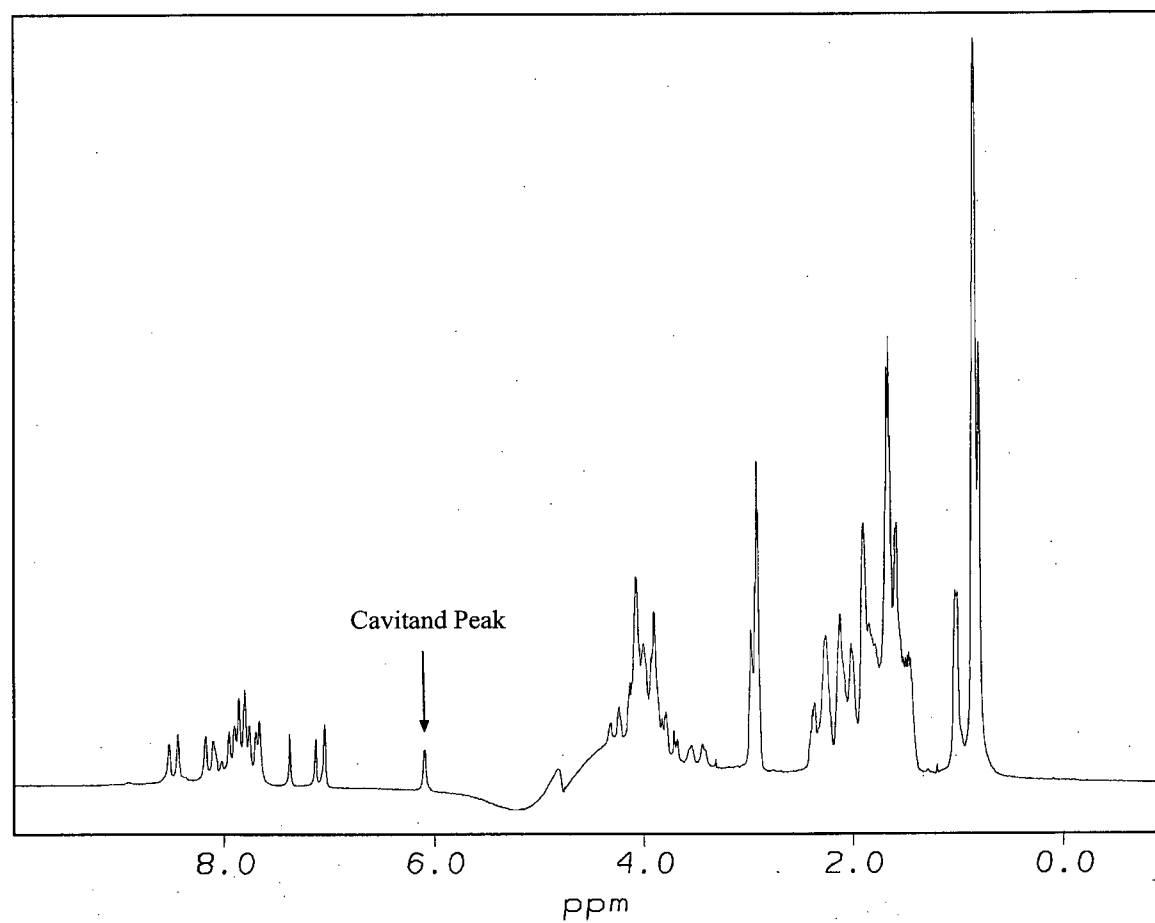
**Figure 5.10.** 500 MHz  $^1\text{H}$  NMR Spectrum of N1G/Ar/Me at  $\sim 300\ \mu\text{M}$  in 10%  $\text{D}_2\text{O}$ , 45 mM Phosphate Buffer, pH 7.0, 25  $^\circ\text{C}$ .



**Figure 5.11.** 500 MHz  $^1\text{H}$  NMR Spectrum of N1GG/Ar/Me at  $\sim 300\ \mu\text{M}$  in 10%  $\text{D}_2\text{O}$ , 45 mM Phosphate Buffer, pH 7.0, 25  $^\circ\text{C}$ .

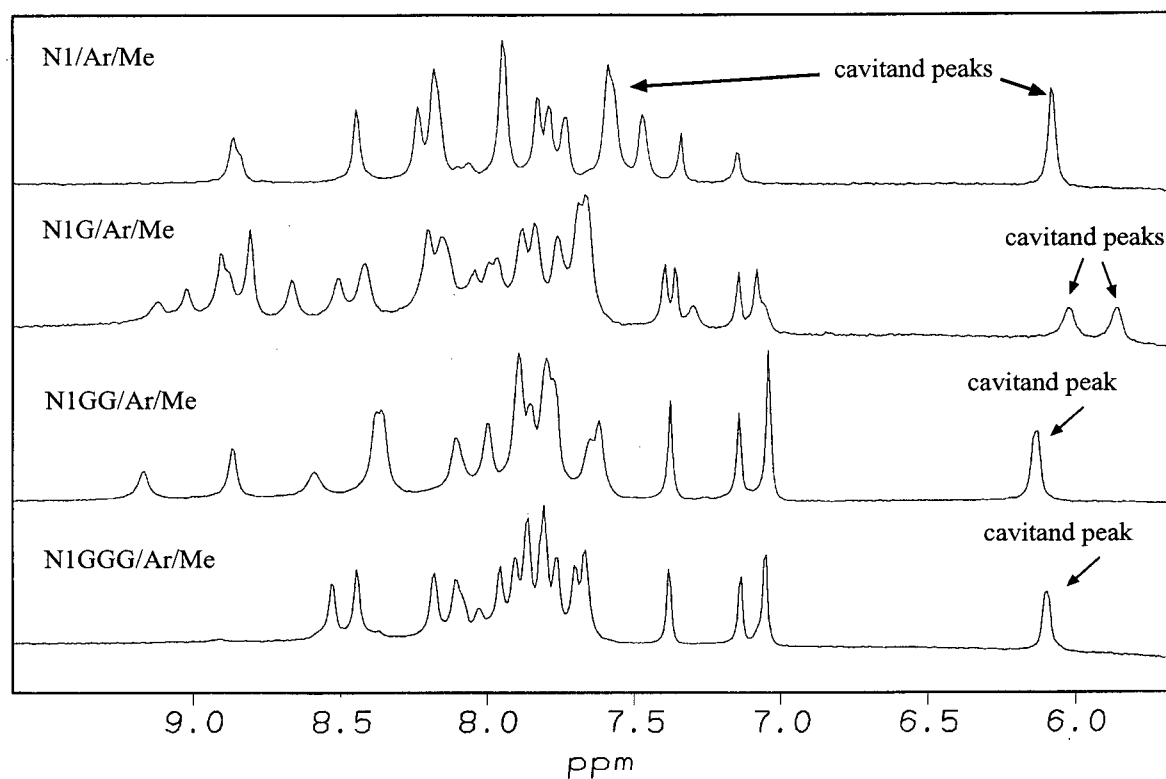


**Figure 5.12.** 500 MHz  $^1\text{H}$  NMR Spectrum of N1GGG/Ar/Me at  $\sim 300\ \mu\text{M}$  in 10%  $\text{D}_2\text{O}$ , 45 mM Phosphate Buffer, pH 7.0, 25  $^\circ\text{C}$ .

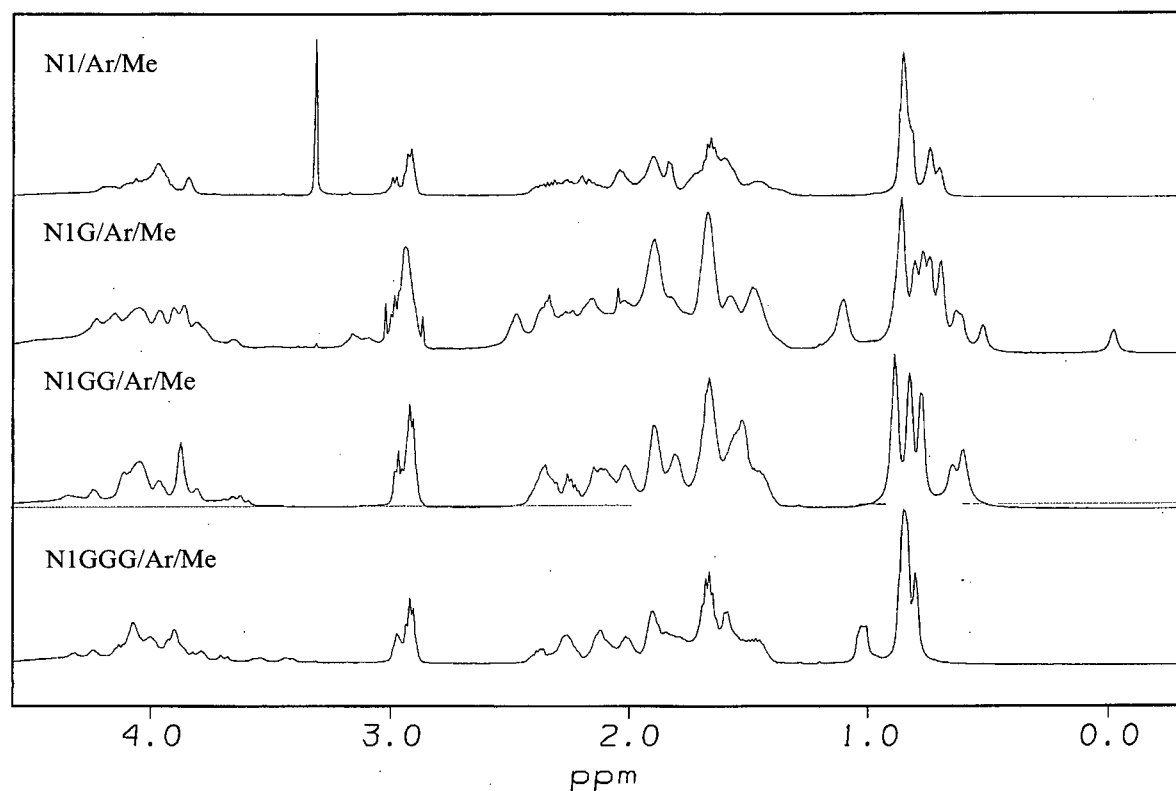




**Figure 5.13.** Stacked Expansions of the Amide Regions of 500 MHz  $^1\text{H}$  NMR Spectra of the Gly Variants ( $\sim 0.3$  mM) at 25  $^\circ\text{C}$ , in 45 mM Phosphate Buffer at pH 7.0 in the Presence of 10%  $\text{D}_2\text{O}$ .



**Figure 5.14.** Stacked Expansions of the Aliphatic Regions of 500 MHz  $^1\text{H}$  NMR Spectra of the Gly Variants ( $\sim 0.3$  mM) at 25  $^\circ\text{C}$ , in 45 mM Phosphate Buffer at pH 7.0 in the Presence of 10%  $\text{D}_2\text{O}$ .



The  $^1\text{H}$  NMR spectrum of N1G/Ar/Me displays the most significant chemical shift dispersion of all Gly-cavitein variants. Despite the sequence and helix degeneracies, approximately 20 amide protons are visible which suggests that the four helices are *not* in degenerate environments as in N1/Ar/Me (as well as N1GG/Ar/Me and N1GGG/Ar/Me, see below). This effect, however, is likely the result of the dimeric nature of N1G/Ar/Me. Also note that the cavitand peak near 6.0 ppm has been split into two peaks which also suggests asymmetry in the structure and may also be a result of the observed dimerization. In comparison to N1/Ar/Me, N1G/Ar/Me possesses significantly more disperse peaks in the  $\alpha$ -

carbon region (near 4 ppm) and more notably in the leucine-methyl region (near 1 ppm). The poor dispersion observed in the methyl region for N1/Ar/Me suggested a dynamically averaged hydrophobic core (Chapter Four). The significant dispersion displayed for N1G/Ar/Me near 1 ppm including the upfield shifted peak near 0 ppm is indicative of native-like packing of the leucine side chains.

The  $^1\text{H}$  NMR spectrum of N1GG/Ar/Me is slightly less disperse in comparison to N1G/Ar/Me. However, in comparison to N1/Ar/Me, the leucine methyl peaks of N1GG/Ar/Me are still more disperse and shifted slightly upfield. The helices in N1GG/Ar/Me appear to be in degenerate environments (as in N1/Ar/Me) since only  $\sim 14$  amide protons are visible and the cavitand peak near 6 ppm is a singlet once again. Thus, the addition of two glycines in N1GG/Ar/Me still appears to allow the helical bundle to adopt a more native-like structure than N1/Ar/Me.

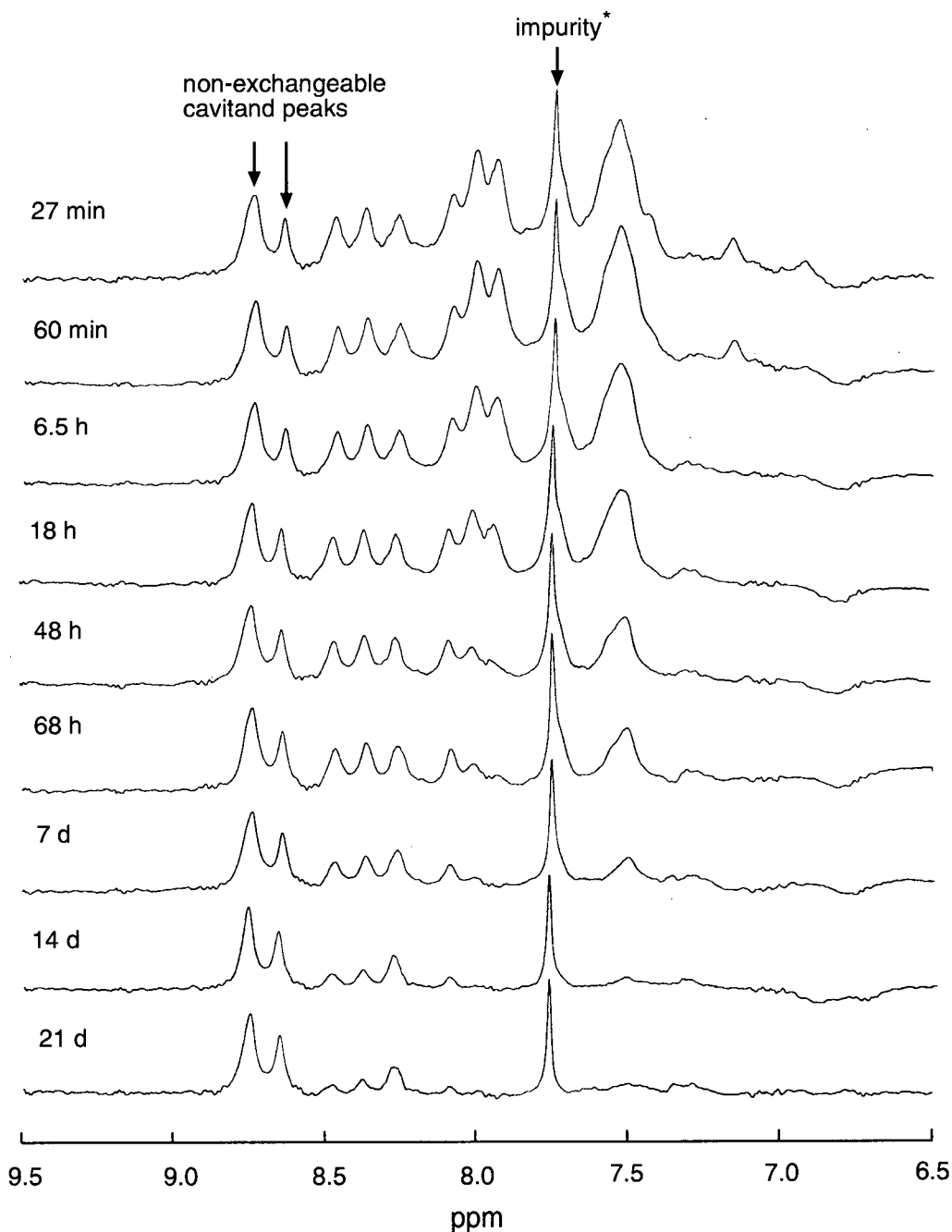
The  $^1\text{H}$  NMR spectrum of cavitein N1GGG/Ar/Me loses considerable chemical shift dispersion. Most notable is the poor dispersion of the leucine methyl peaks in the hydrophobic core. The upfield-shifted methyl peaks observed in N1G/Ar/Me and N1GG/Ar/Me are lost and the dispersion is similar to that of N1/Ar/Me. It appears that the addition of three glycines significantly reduces the native-like character of this cavitein and most likely allows the hydrophobic core to adopt a more dynamic, averaged structure. Note that despite the similar stabilities observed for N1GG/Ar/Me and N1GGG/Ar/Me, the former is considerably more native-like from its chemical shift dispersion. This is one example illustrating that structural stability does not necessarily induce native-like structure.

## 7. Hydrogen/Deuterium Exchange of N1G/Ar/Me

As mentioned in Chapter Four, Section B.iii.h, hydrogen/deuterium (H/D) exchange rates of amide protons have been found to be much faster in more dynamic, molten globule-like structures. From its amide H/D exchange rates, the structure of N1/Ar/Me was found to most likely lie between that of a molten globule and a native-like state. In comparison to N1/Ar/Me, N1G/Ar/Me has been shown to be more stable to GuHCl, display a stronger near-UV absorption and display significantly more dispersion in its 1D  $^1\text{H}$  NMR spectrum. We were therefore interested in measuring the H/D exchange rates of the amide protons in N1G/Ar/Me to determine if they are significantly more protected than those in N1/Ar/Me and if the exchange rates are typical of native-like structures.

The time-course of the H/D exchange experiment is shown in Figure 5.15. Under the identical conditions used for the same experiment on N1/Ar/Me, it is apparent that amide exchange in N1G/Ar/Me is *substantially* slowed. Although many amides exchange before the first scan could be acquired (27 min), at least three amide peaks are visible even after three weeks. Their rate constants, half-lives and protection factors are listed in Table 5.4.

**Figure 5.15.** Stack Plot of 500 MHz  $^1\text{H}$  NMR Spectra Illustrating the Time Dependent Amide H/D Exchange of N1G/Ar/Me in 50 mM pD 5.0  $\text{CD}_3\text{COOD}/\text{NaOD}$  Buffer at 25  $^\circ\text{C}$ .



\* The impurity, likely residual  $\text{CHCl}_3$  from a microsyringe, was not present when this experiment was first attempted on a 400 MHz spectrometer. The H/D exchange behaviour, during this prior 400 MHz experiment, was very similar and thus the impurity likely does not affect H/D exchange in the above 500 MHz experiment.

**Table 5.4.** Tabulated Data From the Amide H/D Exchange Experiment on N1G/Ar/Me in 50 mM pD 5.0 CD<sub>3</sub>COOD/NaOD Buffer at 25 °C.

Amide Proton Chemical Shift (ppm)	First-Order Rate Constant (h <sup>-1</sup> )	Half-Life (d)	Protection Factor <sup>a</sup>
8.5 <sup>b</sup>	$5 \pm 1 \times 10^{-3}$	$6 \pm 1$	$8 \pm 1 \times 10^4$
8.4 <sup>b</sup>	$4 \pm 1 \times 10^{-3}$	$8 \pm 1$	$1 \pm 0.2 \times 10^5$
8.3 <sup>c</sup>	$5 \pm 1 \times 10^{-4}$	$60 \pm 15$	$8 \pm 2 \times 10^5$

<sup>a</sup> these values are based on the half-life of an unprotected proton at 25 °C at pD 5.0 to be 0.11 min (see Experimental Section of Chapter Four)

<sup>b</sup> calculated data are the average of three estimates (7 d, 14 d, 21 d); listed errors represent the standard deviation of these three estimates

<sup>c</sup> calculated data are the average of two estimates (14 d and 21 d); listed errors represent the standard deviation between these two estimates

Although there are large errors associated with the listed protection factors, these amide protons, nevertheless, exhibit substantial protection from exchange and these protection factors are typical of native-like structures. This data augments the well-disperse <sup>1</sup>H NMR spectrum, strong near-UV CD absorption and negligible ANS binding for N1G/Ar/Me. Together, these results indicate that N1G/Ar/Me possesses native-like characteristics to its dimeric helical structure under the conditions described.

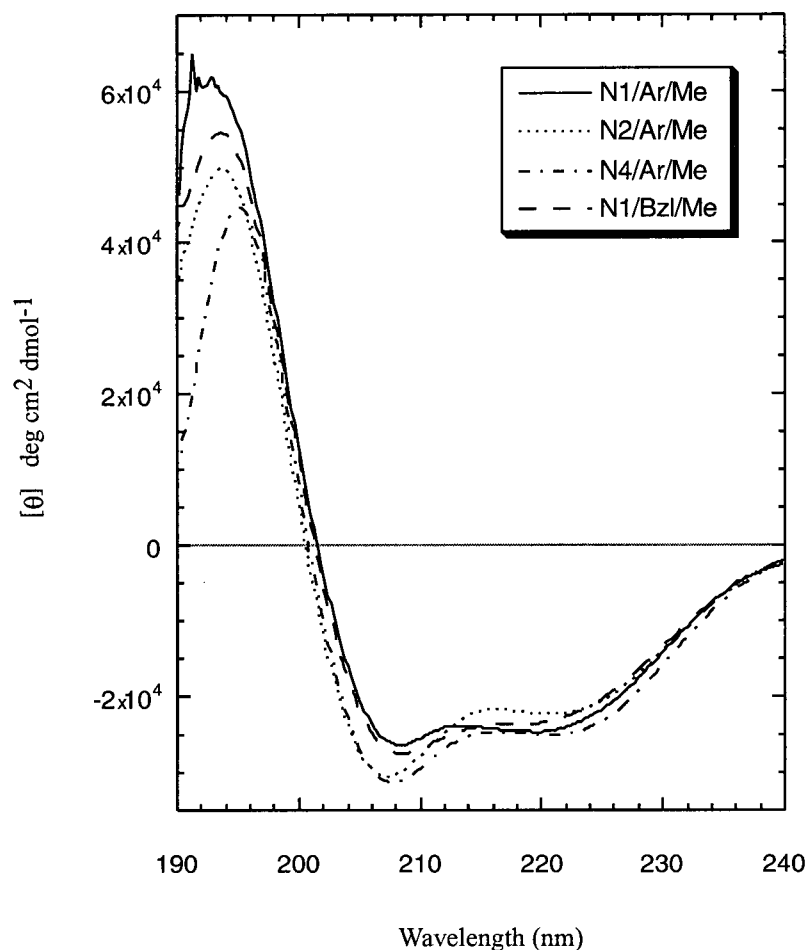
## b. Comparison of Methylene Variants

As mentioned previously, the addition of *N*-terminal glycine residues not only add flexibility to the linkage but also add (1) an element of rigidity due to the planarity of the peptide bond, and (2) potential hydrogen bond donors (NH) and acceptors (CO). Therefore, the additional glycines may have aided in generating some of the native-like structural characteristics mentioned above. We were interested in understanding the true effect of the template and thus synthesized caviteins possessing flexible and non-hydrogen bonding *N*-terminal methylene linker units (N2/Ar/Me, N4/Ar/Me and N1/Bzl/Me).

### 1. Far-UV CD Spectroscopy

The far-UV CD spectra of the methylene-linker cavitein variants are shown in Figure 5.16. From their spectra, it is clear that up to three extra methylene linker units can be added without significantly altering the helicity of the caviteins when compared to N1/Ar/Me. This result is to be expected since up to three glycine units have been added as linker groups without significantly altering the helicity. The small differences could be attributed to small effects of the cavitand template on the far-UV CD spectra.

**Figure 5.16.** Far-UV CD Spectra of Caviteins N1/Ar/Me, N2/Ar/Me, N4/Ar/Me and N1/Bzl/Me in pH 7.0 Phosphate Buffer at 25 °C. Each Cavitein is at a Concentration of  $\sim 30 \mu\text{M}$ .



## 2. Oligomeric State and Stability Towards Guanidine Hydrochloride

More relevant to the discussion of structure and topology is the effect of the linker on oligomeric state of the caviteins. Sedimentation equilibrium experiments were performed, as described previously, for caviteins N2/Ar/Me, N4/Ar/Me and N1/Bzl/Me and the results shown in Table 5.2. In addition, fits of the raw data to the theoretical curve are shown in the



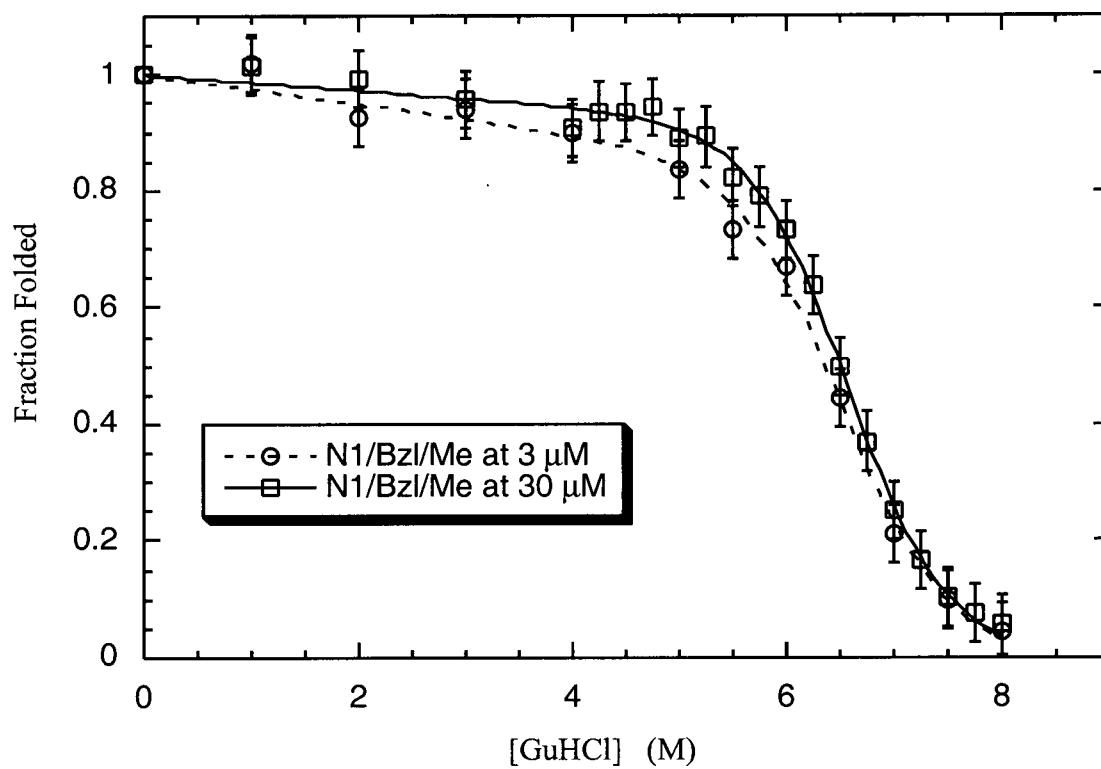
Experimental Section in Figures 5.34 – 5.37. Interestingly, N2/Ar/Me is predominantly monomeric while N4/Ar/Me and N1/Bzl/Me are predominantly dimeric.

Recall that N1/Ar/Me was found to be a mixture of monomers and dimers. This is quite significant as the addition of a single methylene linker unit can alter the oligomeric state of the caviteins. The cavitand linkage clearly has a pronounced effect on the structure and topology of each cavitein, although the exact nature of this effect is not fully understood.

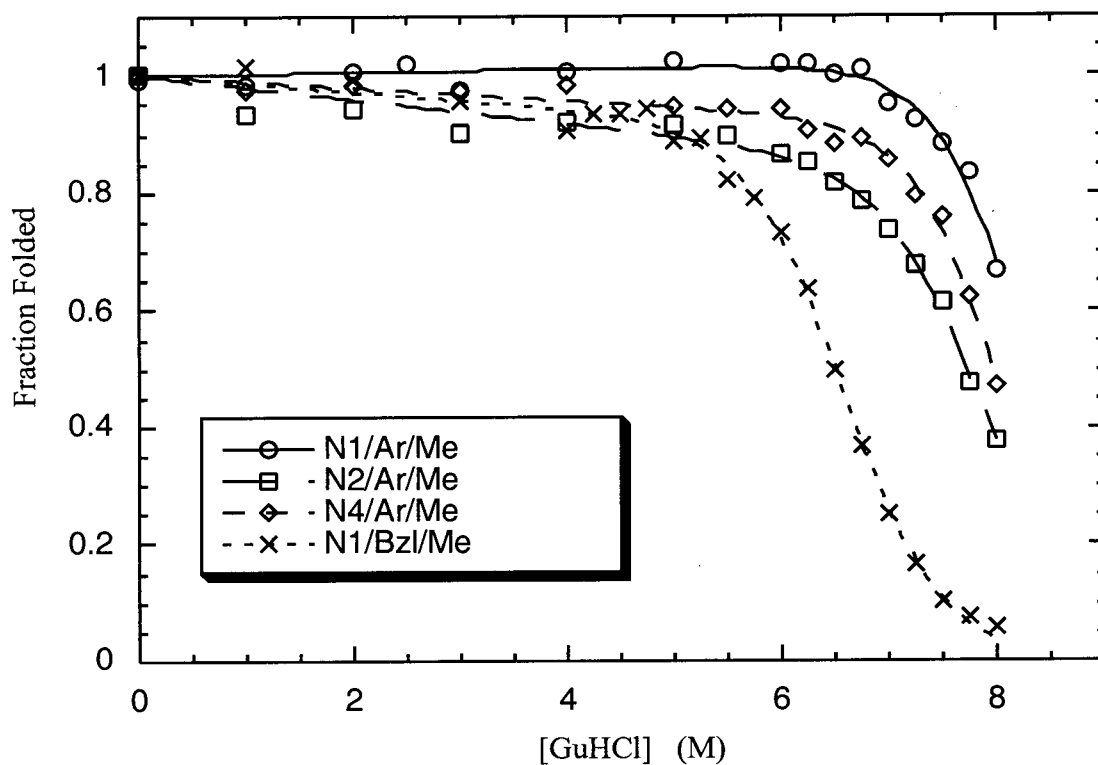
In addition to the sedimentation equilibrium experiments, GuHCl denaturation experiments were also done at 3 and 30  $\mu\text{M}$  for N1/Bzl/Me to further explore the effect of dimerization on the stability and structure of the cavitein (Figure 5.17). Very little change (within error) was observed between the denaturation curves at 3 and 30  $\mu\text{M}$  which suggests that dimerization does not affect side-chain packing in the core of the four-helix bundle. In other words, it is likely that N1/Bzl/Me goes through a dimer-to-monomer-to-unfolded transition in a similar manner to N1/Ar/Me. Note that this is in contrast to the concentration dependence of the N1G/Ar/Me denaturation curve which suggested that dimerization has a significant effect on the packing of its structure.

One interesting observation is the change in oligomeric state between N2/Ar/Me and N1/Bzl/Me. These two caviteins are constitutional isomers since only the order of cavitand-peptide linker atoms has changed (“*cavitand-S-CH<sub>2</sub>-CH<sub>2</sub>-peptide*” vs. “*cavitand-CH<sub>2</sub>-S-CH<sub>2</sub>-peptide*”). The oligomeric state has been altered by simply swapping the order of the atoms. How does this change affect the stability of the bundle? This question is addressed below.

**Figure 5.17.** Guanidine Hydrochloride-Induced Denaturation Curves for N1/Bzl/Me at 3 and 30  $\mu\text{M}$  at pH 7 and 25  $^{\circ}\text{C}$ .



**Figure 5.18.** Guanidine Hydrochloride-Induced Denaturation Curves for 30  $\mu$ M solutions of N1/Ar/Me, N2/Ar/Me, N4/Ar/Me and N1/Bzl/Me at pH 7 and 25  $^{\circ}$ C. Error Bars Have Been Omitted For Clarity.



**Table 5.5.** Guanidine Hydrochloride-Induced Denaturation Data Calculated for the Methylene Cavitein Variants.<sup>a</sup>

Cavitein	Cavitein Concentration ( $\mu\text{M}$ )	$[\text{GuHCl}]_{1/2}$ (M)	$m$ ( $\text{kcal mol}^{-1} \text{M}^{-1}$ )	$\Delta G^\circ_{\text{H}_2\text{O}}$ kcal/mol
N1/Ar/Me	3	$> 8.0$	$-1.3 \pm 0.2$	$-10.4 \pm 1.1$
N1/Ar/Me	30	$> 8.0$	$-1.4 \pm 0.2$	$-11.9 \pm 1.5$
N2/Ar/Me	30	$7.7 \pm 0.1$	$-1.2 \pm 0.1$	$-9.9 \pm 0.9$
N4/Ar/Me	30	$8.0 \pm 0.1$	$-1.5 \pm 0.1$	$-11.9 \pm 1.0$
N1/Bzl/Me	3	$6.5 \pm 0.1$	$-1.3 \pm 0.1$	$-8.7 \pm 0.8$
N1/Bzl/Me	30	$6.4 \pm 0.1$	$-1.3 \pm 0.1$	$-8.6 \pm 0.3$

<sup>a</sup> In all cases, a full unfolding transition was not observed which introduces additional errors to those listed (see Experimental Section).

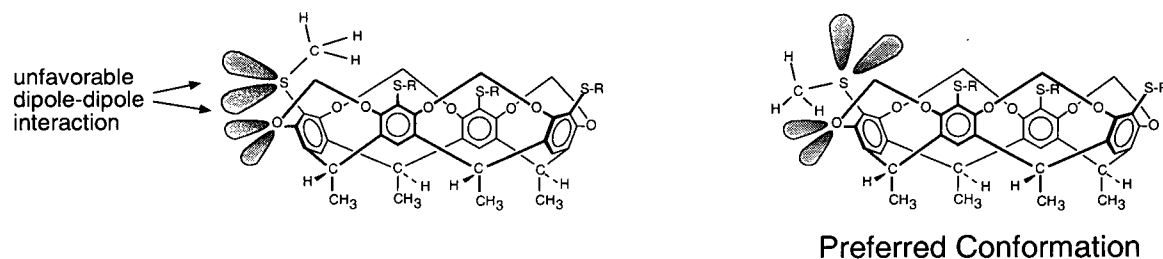
The stabilities of all the methylene variants at 30  $\mu\text{M}$  towards GuHCl are shown in Figure 5.18 and tabulated in Table 5.5. Their stabilities and  $m$  values listed in Table 5.5 were calculated by a nonlinear least-squares analysis of the data as mentioned previously. As with N1/Ar/Me, none of the methylene linker variants displayed a full unfolding transition and therefore the errors in their calculated stabilities are likely to be high (see Experimental Section) and hence difficult to interpret.

Interestingly, the concentration of GuHCl required to half-unfold N1/Bzl/Me is  $\sim 1.2 \text{ M}$  less than that of N2/Ar/Me – its constitutional isomer. Although they possess similar extrapolated stabilities, the differences in their unfolding curves suggest that their structures are considerably different. Why does the simple change in order of the linker atoms significantly alter its oligomeric state and its general structure?

One possible reason lies in the preferred conformation of the sulfur atom with respect to the cavitand. "MM2" molecular mechanics calculations predicted that the preferred conformation of sulfur is such that its lone pairs are directed *away* from the lone pairs of the cavitand bridge-oxygens in order to minimize dipole-dipole repulsions. For *S*-methyl derivatives of cavitands **3** and **40**, this conformation about the four sulfides was found to be ~ 5 kcal/mol more stable than that where the lone pairs are oriented in the same direction (Figure 5.19-A,B). One consequence of these preferred cavitand conformations is that the peptides may be positioned differently (Figure 5.19-C). This difference in positioning may be the cause for the observed difference in structural properties. Unfortunately, further interpretations regarding cavitand structure cannot be made with this data alone.

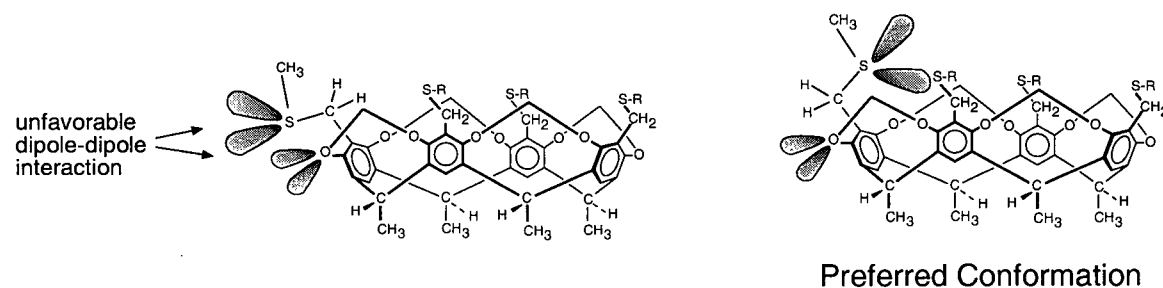
**Figure 5.19.** Illustration of the MM2-Predicted Conformations of *S*-Methyl Derivatives of (A) Cavitand **3** and (B) Cavitand **40**. Part (C) Illustrates that the Peptides are Likely Positioned Differently When Attached to Cavitands **3** and **40**.

**A. *S*-Methyl Cavitand **3****

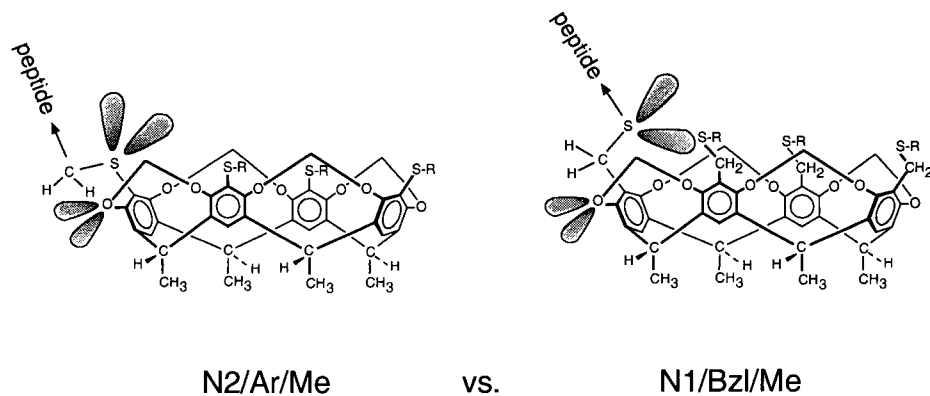


**B. *S*-Methyl Cavitand **40****

R = CH<sub>3</sub>



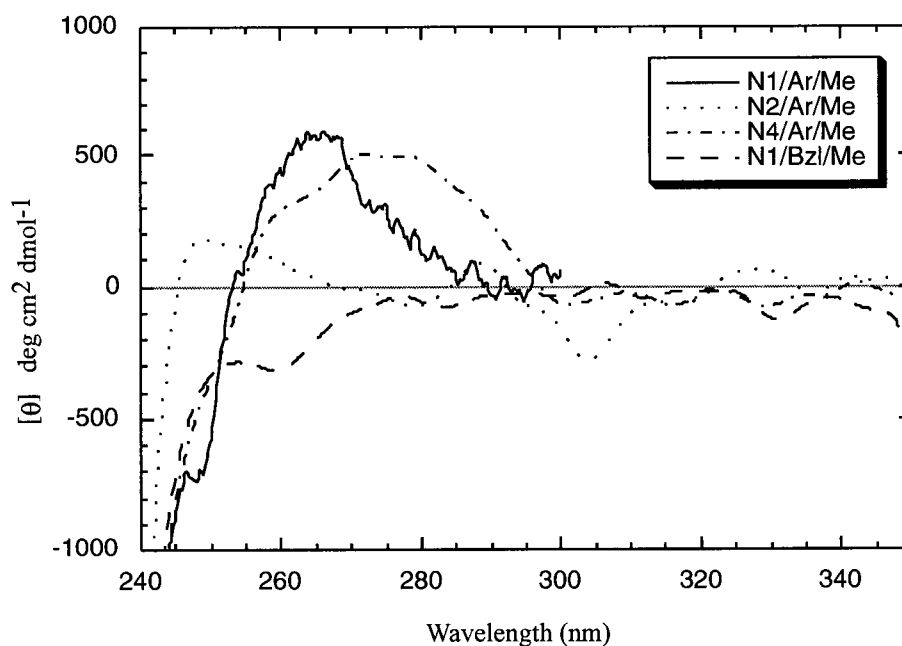
**C. Possible Consequence of Preferred Linker Conformations**



### 3. Near-UV CD Spectroscopy

The near-UV CD spectra of the methylene linker variants are shown in Figure 5.20. Interestingly, both N2/Ar/Me and N1/Bzl/Me possess drastically reduced absorptions in this region and possess the weakest near-UV absorption of all the caviteins. This suggests that the cavitand template is in an environment completely lacking defined structural elements and is in an environment more typical of molten globule-like structures. In general, all the methylene cavitein variants possess reduced near-UV CD absorptions when compared to the Gly cavitein variants (Figure 5.8).

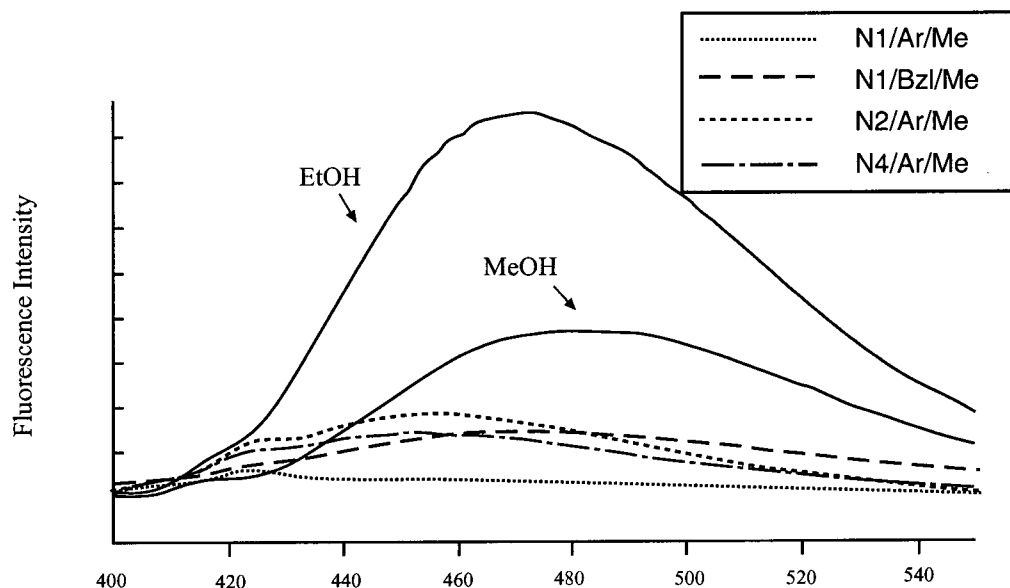
**Figure 5.20.** Near-UV CD Spectra of the Methylene Cavitein Variants at pH 7.0 in 50 mM Phosphate Buffer and 25 °C.



#### 4. ANS Binding

We further investigated the structural dynamics of the methylene variants using ANS binding as a fluorescent probe for the molten globule state. We found that all three methylene variants bound ANS weakly under typical conditions for the experiment (Figure 5.21). Note that these are the same conditions with which we found negligible ANS binding to peptide **28**, N1/Ar/Me and all the Gly-cavitein variants. These results suggest that the methylene variants are more typical of molten globule structures as previously suggested from the near-UV CD data. What is the source of the structural differences between the Gly and methylene cavitein variants? This question is addressed in Section B.iii.c of this Chapter.

**Figure 5.21.** Fluorescence Emission Spectra of 1  $\mu$ M ANS in the Presence of 95% Ethanol, 100% Methanol and 50  $\mu$ M solutions of N1/Ar/Me, N2/Ar/Me, N4/Ar/Me and N1/Bzl/Me at 25  $^{\circ}$ C in pH 7.0 Phosphate Buffer.

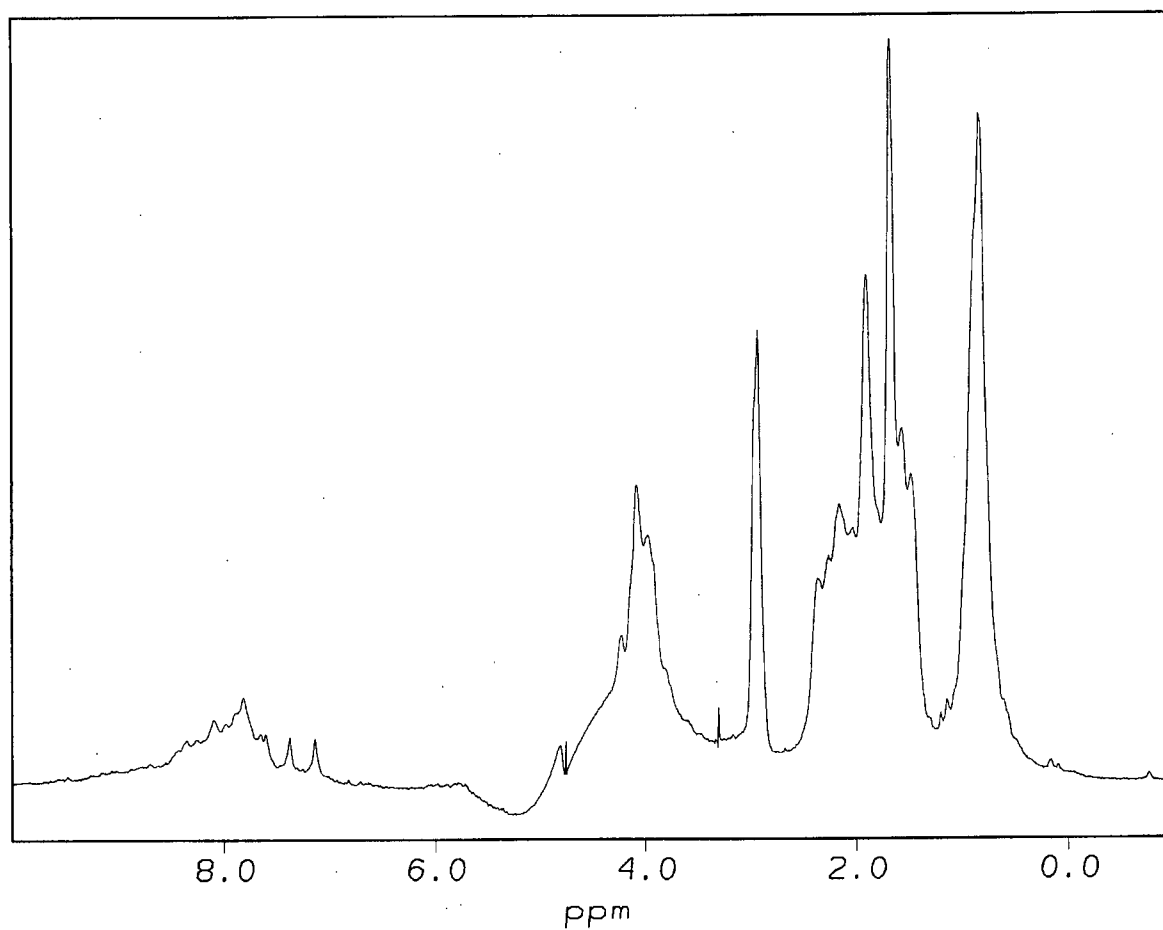




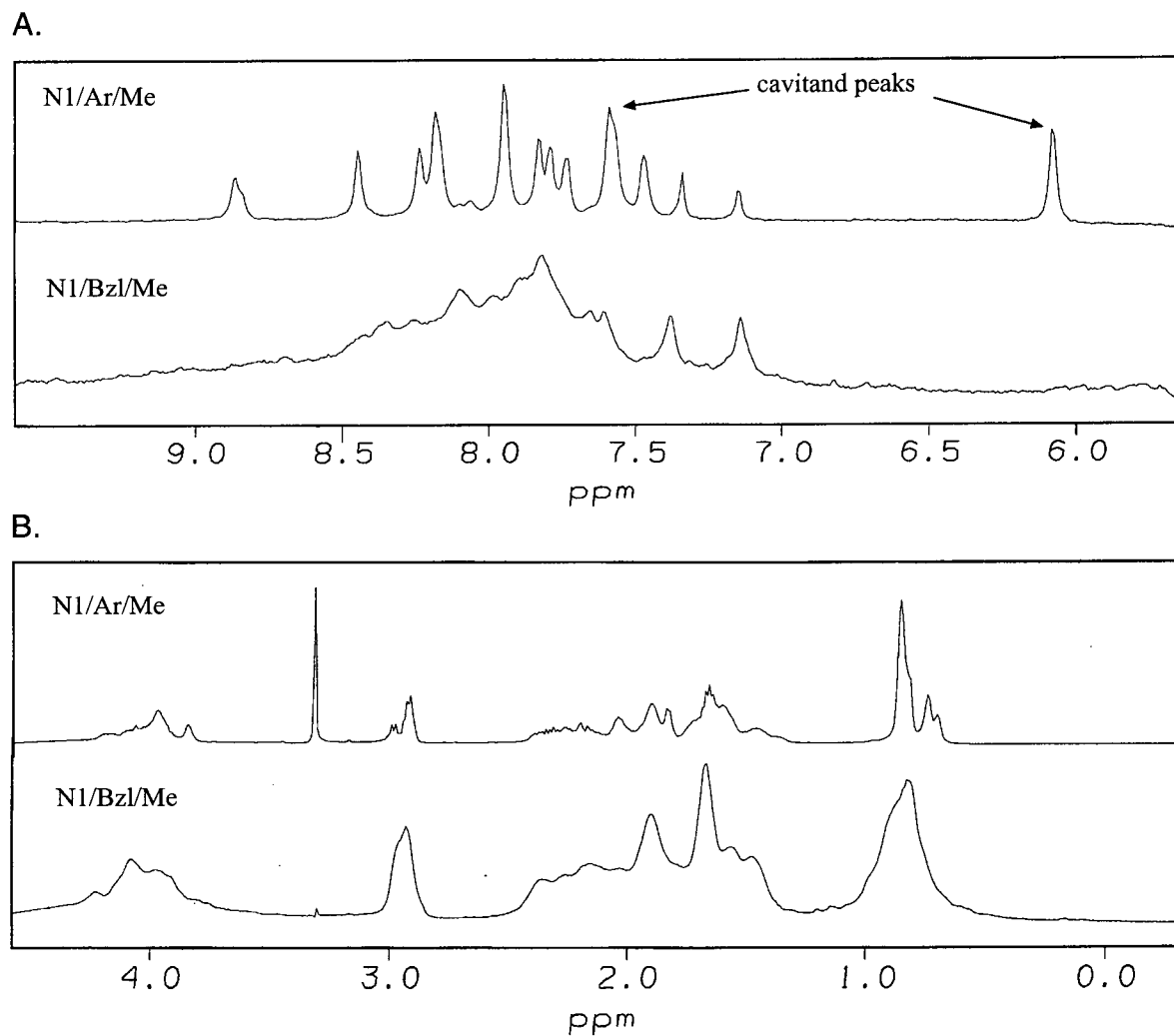
## 5. $^1\text{H}$ NMR Chemical Shift Dispersion of N1/Bzl/Me

In light of the near-UV CD and ANS binding results, we further investigated the presence of a molten globule state using N1/Bzl/Me as a representative example. Recall that native-like structures give rise to sharp and disperse peaks while fluctional molten globule structures give rise to peaks that are less disperse and more broad. The full  $^1\text{H}$  NMR spectrum of N1/Bzl/Me is shown in Figure 5.22 with expansions in Figure 5.23 along with expansions of the NMR spectrum of N1/Ar/Me for comparison.

**Figure 5.22.** 500 MHz  $^1\text{H}$  NMR Spectrum of N1/Bzl/Me at  $\sim 300\ \mu\text{M}$  in 10%  $\text{D}_2\text{O}$ , 45 mM Phosphate Buffer, pH 7.0, 25  $^\circ\text{C}$ .



**Figure 5.23.** Expansions of the (A) Amide and (B) Aliphatic Regions of 500 MHz  $^1\text{H}$  NMR Spectra of N1/Ar/Me and N1/Bzl/Me ( $\sim 0.3$  mM) at 25  $^\circ\text{C}$ , in 45 mM Phosphate Buffer at pH 7.0 in the Presence of 10%  $\text{D}_2\text{O}$ .



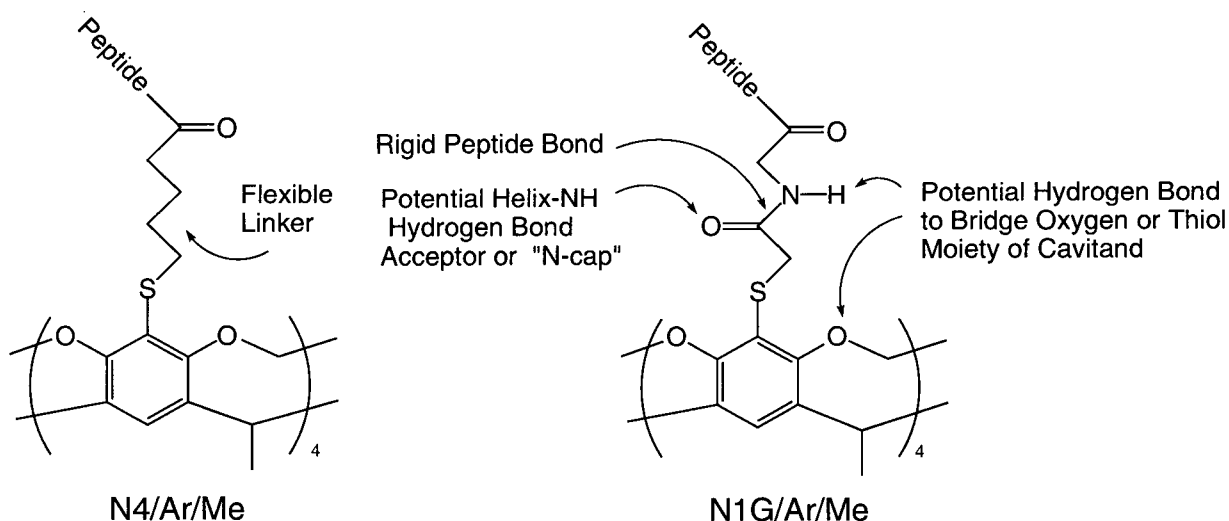
In the  $^1\text{H}$  NMR spectrum of N1/Bzl/Me, both the aliphatic and the amide regions have broadened substantially when compared to N1/Ar/Me. This result is consistent with a fluctuating or dynamic molten globule-like tertiary structure. The effect of dimerization of N1/Bzl/Me cannot be ruled out completely as a possible cause of the broadening; however, the

NMR spectra of both N1/Ar/Me (monomer-dimer mixture) and N1G/Ar/Me (dimer) possess significantly sharper and more disperse peaks.

c. Comparison of Glycine and Methylene Variants: Particularly N1G/Ar/Me and N4/Ar/Me

As mentioned previously, why, in general, are the methylene variants more molten globule-like and the glycine variants more native-like in structure? This question can be addressed by comparing N1G/Ar/Me and N4/Ar/Me. These caviteins possess vastly different structural properties: (1) the near-UV absorption of N1G/Ar/Me is significantly larger than that of N4/Ar/Me, (2) N1G/Ar/Me does not bind ANS while N4/Ar/Me does, and (3) N1G/Ar/Me and N4/Ar/Me possess considerably different GuHCl denaturation curves, extrapolated stabilities and cooperativity to their unfolding transitions. Put another way, N4/Ar/Me likely exists in a molten globule state while N1G/Ar/Me likely exists in a native-like state (although both appear to be dimeric). Yet, these caviteins are constitutionally quite similar because they possess the same number of atoms linking the cavitand template to the 14-amino acid peptide sequence: while N4/Ar/Me possesses four methylene groups, N1G/Ar/Me possesses a “-CH<sub>2</sub>-NH-CO-CH<sub>2</sub>-” linker group, (Figure 5.24) – effectively replacing two methylene groups with a peptide bond. As a result, arguments used to rationalize structural differences based on the peptides’ proximity to the template do not apply. From their structural differences, it is clear that, when attached to the cavitand, the peptide sequence alone is insufficient to induce native-like structure and that there must exist a significant “native-like-structure-inducing” linker effect.

**Figure 5.24.** Illustration of the Different Linker Groups in N1G/Ar/Me and N4/Ar/Me.



Three additional effects are likely associated with the added peptide bond in the linker of N1G/Ar/Me when compared to N4/Ar/Me. This first is the added rigidity of the linker due to the planarity of the peptide bond. Secondly, the added peptide bond adds an additional hydrogen bond acceptor (CO) which may act as a helix-stabilizing "N-cap" (Chapter One, Section D.ii.c.1) by hydrogen bonding to the unsatisfied NH group four amino acids up the peptide chain. And thirdly, the peptide bond possesses an NH group which has been shown to hydrogen bond well to the bridge-oxygen of the cavitand and the thioether linkage when placed in that position with respect to the cavitand (Chapter Three, Section B.ii). Each of these effects serve to increase the structural rigidity and specificity of the linker group and hence impart native-like structure to the overall helical bundle. These effects may be especially important in helices of only 14 amino acids in length which possess a maximum of 10 NH-CO backbone

hydrogen bonds per helix. Two additional hydrogen bonds per helix could indeed make a significant contribution to the overall structure by inducing more specific structural elements.

These added effects may be more generally applicable to all of the Gly variants which not only add a somewhat flexible linkage, but also potential hydrogen bond donors and acceptors. Additional hydrogen bond donors and acceptors likely stabilize the helix despite the known helix-breaking tendency of glycine. The fact that N1GGG/Ar/Me is more native-like in structure than any of the methylene variants suggests that the peptide's proximity to the cavitand is not as important as the addition of potential hydrogen bond donors and acceptors for inducing native-like structure.

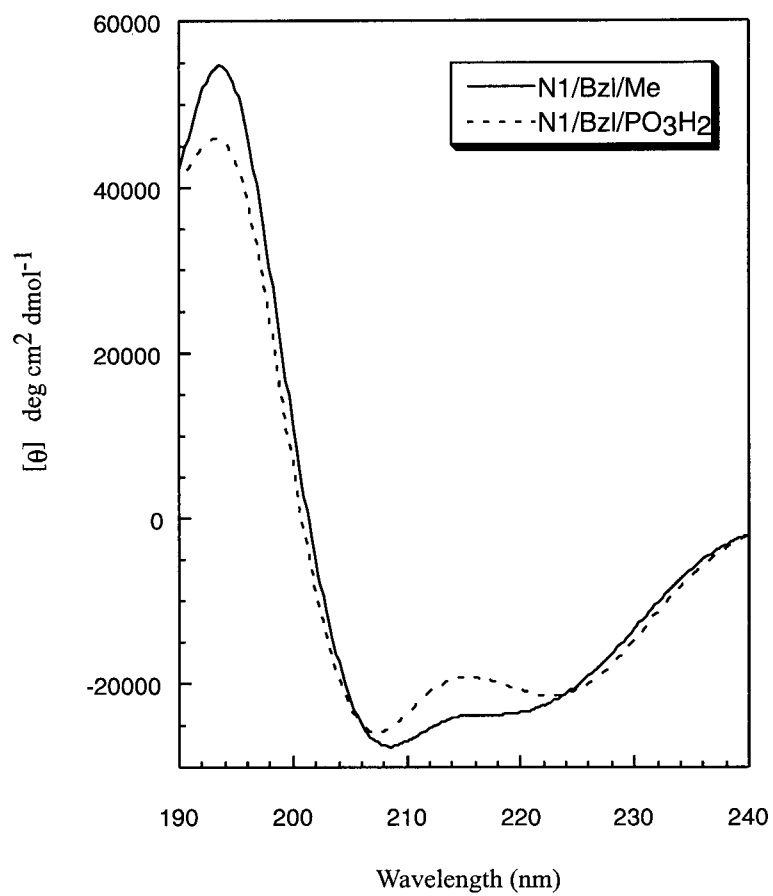
Note that in the series of 1, 2 and 4 methylene linker moieties, only N1/Ar/Me possesses some native-like properties. This is likely a combination of (1) the peptides' close proximity to the rigid template thereby restricting the flexibility of the peptides and/or (2) the position of the first NH in the amino acid sequence in relation to the cavitand which has been shown to promote hydrogen bonding to the cavitand-bridge oxygen and sulfur. The addition of extra methylene linker units do not promote such hydrogen bonding (Chapter Three, Section B.ii) which may result in less specific structural elements near the cavitand-peptide linkage and hence promote more molten globule-like structures. Note that this occurs despite their stability towards GuHCl. Structural stability is not necessarily accompanied by structural specificity; specific interactions within a protein, such as hydrogen bonding, are required to induce native-like structure. From all of the data thus far, the cavitand-peptide linkage and arguably the cavitand itself play important roles in inducing native-like structure.

d. Comparison of Phosphate- and Methyl-Footed Cavitein Variants

1. CD Spectroscopy

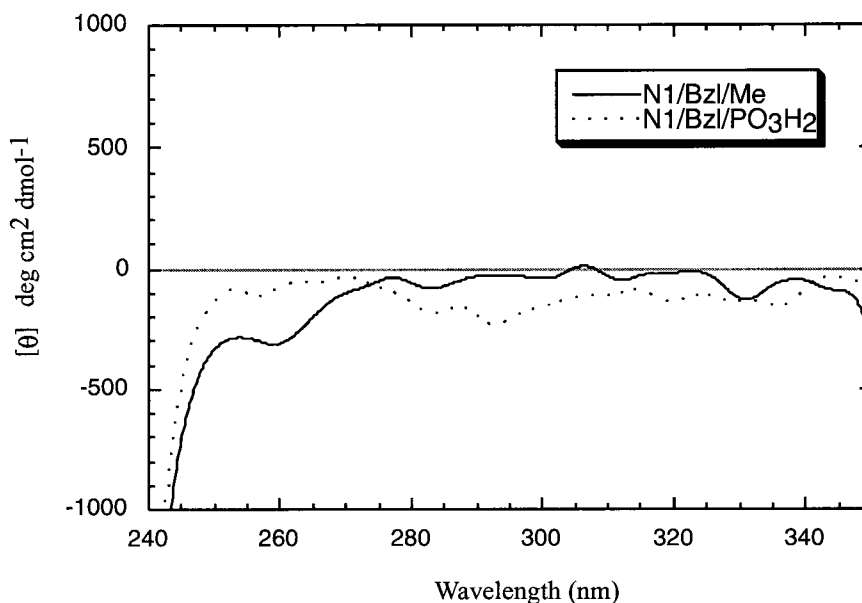
The effect of the cavitand foot on cavitein structure was studied by comparing the structural properties of N1/Bzl/Me and N1/Bzl/PO<sub>3</sub>H<sub>2</sub>. These two caviteins possess identical cavitand-peptide linkages and differ only by the foot of the cavitand (methyl vs. propyl-phosphate). Their far-UV CD spectra are both typical of  $\alpha$ -helical structure and only differ slightly which may be attributable to the different aromatic cavitand chromophores (Figure 5.25). Their near-UV CD spectra are quite similar, lacking significant aromatic absorptions indicative of averaged structural elements near the cavitand-peptide linkage (Figure 5.26). Therefore, by CD spectroscopy, the addition of the hydrophilic phosphate foot has done little to affect cavitein structure.

**Figure 5.25.** Far-UV CD Spectra of 30  $\mu\text{M}$  Solutions of N1/Bzl/Me and N1/Bzl/ $\text{PO}_3\text{H}_2$  in 50 mM Phosphate Buffer at pH 7 and 25  $^\circ\text{C}$ .





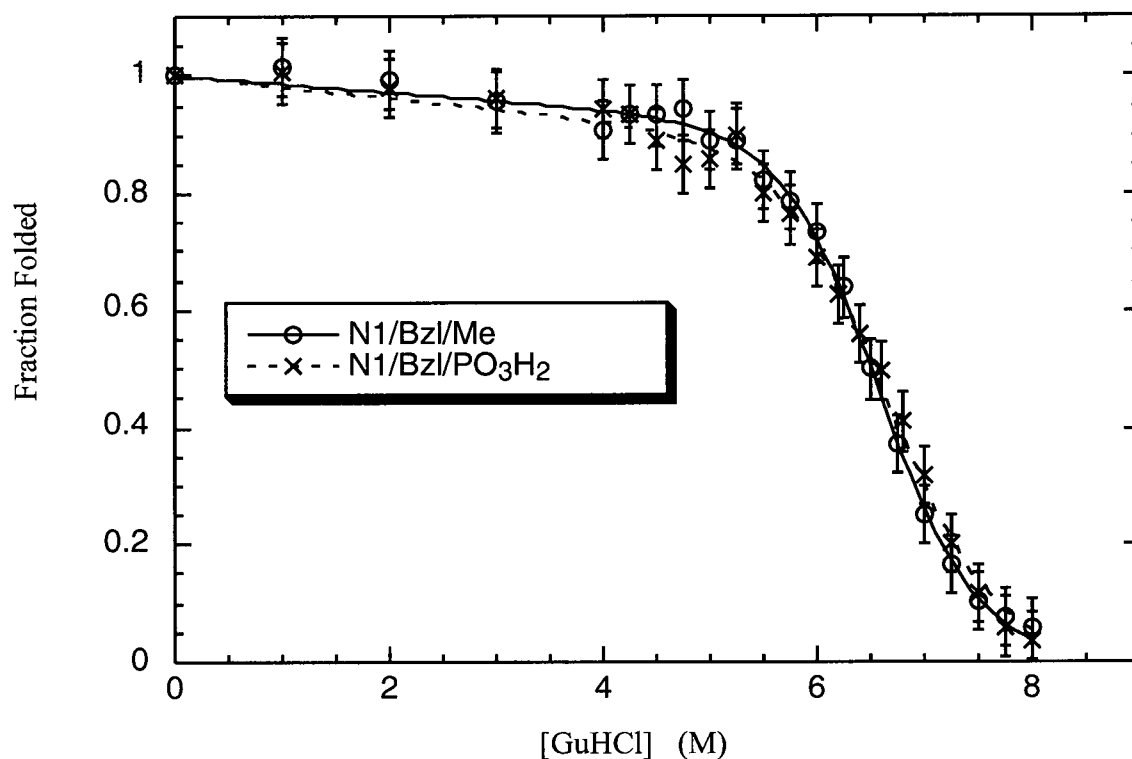
**Figure 5.26.** Near-UV CD Spectra of 30  $\mu\text{M}$  Solutions N1/Bzl/Me and N1/Bzl/ $\text{PO}_3\text{H}_2$  in 50 mM Phosphate Buffer at pH 7 and 25  $^\circ\text{C}$ .



## 2. Oligomeric State and Stability Towards Guanidine Hydrochloride

Additional information can be gleaned by studying the oligomeric state of N1/Bzl/ $\text{PO}_3\text{H}_2$ . Recall that N1/Bzl/Me was found to be predominantly dimeric by analytical ultracentrifugation. Analysis of N1/Bzl/ $\text{PO}_3\text{H}_2$  by the same method found that it was predominantly *monomeric* under the typical experimental conditions (Table 5.2 and Figure 5.38 in the Experimental Section). This is quite interesting considering their similar stabilities towards GuHCl and denaturation curves (Figure 5.27, Table 5.6).

**Figure 5.27.** Guanidine Hydrochloride-Induced Denaturation Curves for 30  $\mu\text{M}$  solutions of N1/Bzl/Me and N1/Bzl/ $\text{PO}_3\text{H}_2$  at pH 7 and 25  $^\circ\text{C}$ .



**Table 5.6.** Guanidine Hydrochloride-Induced Denaturation Data Calculated for the Cavitant and Foot Cavitein Variants.<sup>a</sup>

Cavitein	Cavitein Concentration ( $\mu\text{M}$ )	$[\text{GuHCl}]_{1/2}$ (M)	$m$ ( $\text{kcal mol}^{-1} \text{M}^{-1}$ )	$\Delta G^\circ_{\text{H}_2\text{O}}$ (kcal/mol)
N1/Bzl/Me	30	$6.4 \pm 0.1$	$-1.3 \pm 0.1$	$-8.6 \pm 0.3$
N1/Bzl/ $\text{PO}_3\text{H}_2$	30	$6.5 \pm 0.1$	$-1.2 \pm 0.1$	$-8.0 \pm 0.5$

<sup>a</sup> In each case, a full unfolding transition was not observed which introduces additional errors to those listed (see Experimental Section).

This result is quite significant and may be interpreted in several ways: (1) the caviteins are self-associating by their hydrophobic feet and thus addition of hydrophilic phosphate feet eliminates aggregation, or (2) the peptides are interacting significantly with the cavitand foot such that altering the foot can alter the overall structure and result in monomers or dimers, or (3) self-association is partially electrostatically-driven and is hampered by the introduction of the charged phosphate feet.

With respect to the first possibility, hydrophobic aggregation by the cavitand feet is unlikely since several monomeric caviteins have been presented in this thesis possessing methyl feet. The second possibility of a peptide-cavitand foot interaction cannot be ruled out with certainty; however, it is unlikely that the peptides could be in such dissimilar conformations and possess the observed near-identical GuHCl denaturation curves and stabilities. More likely is the third possibility that self-association is mediated, at least partially, by electrostatic interactions as a result of non-ideal helical conformations (recall that this was hypothesized as a potential cause of the monomer-dimer equilibrium in N1/Ar/Me). Introduction of an overall charge of at least  $-4$  (depending on the  $pK_a$  of the second ionizable phosphate proton which is near 7) to the cavitein via the phosphate feet likely inhibits dimerization by making self-association electrostatically unfavorable. This hypothesis may be tested by synthesizing a methyl-footed cavitein with an overall charge derived from its amino acid sequence. If successful, this may be a general method of inhibiting dimerization in future cavitein designs since phosphate-footed cavitands are somewhat labour-intensive to synthesize (Chapter Two). Note that the core amino acid sequence used in all of the caviteins presented in this thesis were originally designed to be electrostatically neutral to aid in bundle stability.

## C. Conclusions

This chapter explored the effect of the cavitand-peptide linkage on the structures and stabilities of the resulting de novo proteins. We found that the linkage had a dramatic effect on the structures and stabilities of the different caviteins, which suggests that the rigid cavitand template indeed greatly influences the protein structure. Interestingly, all of the the Gly-cavitein variants possess some native-like properties to their structures in contrast to the methylene-linker cavitein variants whose structures were more characteristic of molten globules. One explanation may be the added hydrogen bonding capabilities of the glycine variants. Not only do extra glycines add potential hydrogen bond donors and acceptors which can stabilize the helix, they also allow for a possible NH to bridge oxygen and sulfur hydrogen bond (Chapter Three). This hydrogen bond is not available with the addition of extra methylene linker groups. The loss or gain of any structure-inducing hydrogen bonds in a helix composed of only ten NH-CO backbone hydrogen bonds could very well be significant and discriminate between molten globule and native-like structures. Note that despite the long three-Gly linker in N1GGG/Ar/Me, its structure still retains some native-like properties likely due, in part, to such hydrogen bonding.

One of the most interesting and unexpected discoveries was the native-like dimeric structure of N1G/Ar/Me. This cavitein is constitutionally very similar to all of the other caviteins, yet possesses a completely different helical topology. It is difficult to speculate on the nature of its structure and further analyses will be required by X-ray crystallography to fully understand the source of the interactions at play.

Recall that single-stranded peptide **28** did not bind ANS which suggested that it too possessed some of the properties of native-like structure (Chapter Four). The fact that all of the methylene variants exhibit more molten globule-like structures in comparison to peptide **28** suggest that the peptide **28** does *not* self-associate forming the same parallel four-helix bundle topology. It is more likely that peptide **28** forms an anti-parallel and/or oligomeric bundle of unknown stiochiometry which may allow for different packing interactions and promote more native-like structures.

The effect of the cavitand foot on cavitein structure and stability was investigated using N1/Bzl/Me and its phosphate-footed analogue N1/Bzl/PO<sub>3</sub>H<sub>2</sub>. Although the oligomeric state of N1/Bzl/Me was altered with the addition of phosphate feet, it is likely that the phosphates do not directly affect the packing of the four-helix bundle. Rather, it is probable that the introduction of four formal charges inhibits dimerization in N1/Bzl/PO<sub>3</sub>H<sub>2</sub> via electrostatic effects.

In this thesis, the cavitand template has been shown to greatly influence the structures of attached peptide strands. The observed cavitein oligomerization, in many cases, is likely due to the helices adopting non-ideal geometries as a result of the rigid template. The addition of more flexible linkers likely allow the peptide strands to adopt more ideal helical conformations. This suggests that the cavitand template itself (without long linker groups) may not be an *ideal* template for four-helix bundle formation. Nevertheless, with the use of longer cavitand-peptide linkages (e.g., two glycines), the cavitand template can greatly enhance the stabilities of the attached helices and enforce a monomeric four-helix bundle topology useful for further studies.

In a more general perspective, this thesis presented a first-generation effort to design template assembled four-helix bundles in an effort to better understand protein structure, protein folding and the interactions at play. Although sources for many of the observed structural differences of the caviteins are not entirely clear, it is evident that subtle changes to the

composition of the cavitein have profound effects on its structure. The effect of a single added methylene or glycine linker have significant effects on the structure of the four-helix bundle and allows for a full appreciation of the complexity and subtlety of forces involved in determining the structures of naturally occurring proteins composed of hundreds of amino acids. Nevertheless, some of the caviteins presented herein (especially N1GG/Ar/Me, see below) represent ideal models for future studies aimed at elucidating the factors controlling four-helix bundle structure. Future cavitein variants should provide insight into the protein folding problem for years to come.

#### **D. Future Work**

One of the goals of this thesis was to generate de novo protein models for studying protein folding. Unfortunately, the structures of the caviteins possessing short cavitand-peptide linkages were very difficult to predict and varied extensively. In addition, modifications of the amino acid sequence may completely alter the specific structure-inducing properties of each cavitand-peptide linkage. However, one cavitein in particular was found to be a monomeric four-helix bundle possessing some of the properties of native-like structure: N1GG/Ar/Me. This cavitein possesses a relatively long cavitand-peptide linkage and therefore the added effect of the rigidity of the cavitand template is somewhat minimized. However, in light of the less chemical shift dispersion observed in the analogous three-gly linker N1GGG/Ar/Me, N1GG/Ar/Me still takes partial advantage of the rigid cavitand template to influence its

structure. In my opinion, N1GG/Ar/Me possesses an ideal balance between linker rigidity and flexibility which should enable it to act as a useful model for further study.

In the short-term, such study could include amino acid substitutions within the core 14-amino acid sequence in order to breakdown the forces and amino acids involved in contributing to stability and native-like structure. For example, several leucine residues could be substituted for alanine in order to determine their effect on cavitein structure, stability and packing within the hydrophobic core. Glutamic acid residues could be substituted with glutamine residues in order to truly evaluate the effect of the designed salt-bridges on cavitein structure and stability.

One severe limitation of the structural interpretations of the caviteins presented in this thesis is the lack of high-resolution structural data. Assuming X-ray quality crystals can be obtained, it may now be time to obtain crystal structures of important caviteins such as N1G/Ar/Me and N1GG/Ar/Me to better understand their folds.

Another common tool used in protein structure determination is multi-dimensional NMR. The relatively small size of these caviteins make them ideal candidates for such analysis except for some fundamental problems arising from the amino acid sequences. As part of the minimalist approach to the design of these first-generation caviteins, a minimum number of different residues were used in the core 14-amino acid sequence. As a result, the peptide sequence consists of four glutamic acids, four lysines, five leucines and one glycine. The assignment of various NMR peaks to individual residues in such a degenerate sequence would be extremely difficult. In addition, each helix is four-fold symmetric and thus gives rise to additional degeneracies which would be difficult to resolve in 2-D NMR experiments.

In the short-term, there are a couple of simple ways in which this problem can be overcome. Firstly, the amino acid sequence should be varied to include a variety of different hydrophobic and hydrophilic residues. This would effectively remove some of the degeneracies

in the NMR spectra. Somewhat more problematic is the four-fold symmetry of the four-helix bundle. One possible solution is to couple four different peptides to the cavitand, one at a time. This idea has considerable potential for success in light of some of the observations made over the course of synthesizing the caviteins presented in Chapter Five. During their syntheses, intermediates possessing two or three peptides attached to cavitand **3** were isolated and characterized by electrospray mass spectrometry. For the purposes of this thesis, either the temperature or the number of peptide equivalents (or both) was increased to alleviate the problem and afford the tetra-peptide cavitein. However, these “synthetic problems” demonstrate that sequential addition of *different* peptide helices to the cavitand is possible. In theory, four different peptides could be sequentially attached to cavitand **3** which would remove the four-fold symmetry and make structural characterization by 2-D NMR techniques much more feasible.

Over the long-term, once the factors controlling four-helix bundle stability and structure are better understood, caviteins could be “taylor-made” with specific structural elements and/or designed for specific applications. To date, several “functional” de novo proteins have been successfully designed to act as ion-channels,<sup>14</sup> peptide receptors,<sup>15</sup> metal-ion sensors<sup>16</sup> and catalysts.<sup>17</sup> With regards to the caviteins, the enforced hydrophobic cavity derived from the cavitand template may prove quite useful for such applications.



## E. Experimental

### i. General

All general experimental details (CD, NMR chemical shift dispersion, ANS fluorescence, sedimentation equilibrium) were performed as described in Chapter Four. Section v describes the analysis of the GuHCl denaturation curves. Section vi describes the H/D exchange experiment for N1G/Ar/Me. Section vii describes specific details for the sedimentation equilibrium experiments. Note that the syntheses of peptide **43**, cavitein N1/Bzl/PO<sub>3</sub>H<sub>2</sub> and cavitein C/Bzl/PO<sub>3</sub>H<sub>2</sub> have been recently published.<sup>6</sup>

### ii. Synthesis of Cavitand **40**

Benzylbromide **39** was prepared as described in the literature.<sup>4</sup> Cavitand **40** was synthesized as described previously<sup>18</sup> with the following two exceptions: (1) benzylbromide **39** was used in place of the analogous benzylchloride cavitand and (2) DMF was used as the solvent in place of DMSO. The following is a representative procedure:

Thiourea (18 mg, 0.24 mmol, 4.4 equiv.) was added to a solution of benzylbromide **39** (50 mg, 55  $\mu$ mol) in degassed DMF (5 mL). The reaction was stirred for 2 h at rt and ice-cold degassed 2 M NaOH (0.6 mL) was added. The reaction mixture was then stirred for an additional 1 h and the solvent was evaporated *in vacuo*. 5% Acetic acid was added until acidic and the precipitate

was collected by filtration over a fine frit. The crude product was then dissolved in  $\text{CHCl}_3$ , dried with  $\text{MgSO}_4$  and precipitated with hexanes to afford cavitand **40**.

The product gave an identical  $^1\text{H}$  NMR spectrum to that previously reported.<sup>18</sup>

### iii. Synthesis of Peptides

All of the peptides were synthesized by solid-phase methods as described in Chapter Four. For peptides **29** (N1G), **31** (N1GG), and **33** (N1GGG), additional glycines were added to the N-terminus using the automated synthesizer. In the case of peptide **42**, the additional N-terminal cysteine residue was also added using the automated synthesizer and an *S*-trityl protecting group. Each peptide was cleaved from the resin with a 2 h treatment with 95% TFA/ $\text{H}_2\text{O}$  except for peptide **43** for which a mixture of TFA (95%),  $\text{H}_2\text{O}$  (2.5%) and 1,2-ethanedithiol (2.5%) was used. Each of the peptides was purified by reversed-phase HPLC as described in Chapter Four and eluted at  $\sim 50\%$  acetonitrile. In addition, the purity of each peptide was assessed by the observation of a single peak by analytical reversed-phase HPLC and their identities confirmed by LSIMS mass spectrometry.

**Peptide 35** (N2Br):  $\text{Br}(\text{CH}_2)_2\text{CO-NH-[EELLKKLEELLKKG]-CONH}_2$ .

The last cycle entailed coupling the free N-terminus of the 14-amino acid peptide resin synthesized for peptide **26** (450 mg resin,  $\sim 150$  mg peptide,  $\sim 0.094$  mmol) with 3-bromopropionyl chloride (47  $\mu\text{L}$ , 0.47 mmol, 5 equiv.), DIEA (82  $\mu\text{L}$ , 0.47 mmol, 5 equiv.) and

NaBr (97 mg, 0.94 mmol, 10 equiv.) in dry DMF (5 mL) for 45 min at rt. Subsequent cleavage from the resin and purification by reversed-phase HPLC afforded peptide **35** as a white solid (36 mg, ~ 20%).

**MS:** (LSIMS, thioglycerol):  $m/z$  1805 ((M + H)<sup>+</sup>, 100).

**Peptide 37 (N4Br):** Br(CH<sub>2</sub>)<sub>4</sub>CO-NH-[EELLKKLEELLKKG]-CONH<sub>2</sub>

The last cycle entailed coupling free the N-terminus of the 14-amino acid peptide resin synthesized for peptide **26** (350 mg resin, ~ 120 mg peptide, ~ 0.071 mmol) with 5-bromovaleryl chloride (47  $\mu$ L, 0.36 mmol, 5 equiv.) and DIEA (63  $\mu$ L, 0.36 mmol, 5 equiv.) in dry DMF (5 mL) for 1 h at rt. Subsequent cleavage from the resin and purification by reversed-phase HPLC afforded peptide **37** as a white solid (49 mg, ~ 40%).

**MS:** (LSIMS, thioglycerol):  $m/z$  1833 ((M + H)<sup>+</sup>, 100).

**Peptide 29 (N1G):** ClCH<sub>2</sub>CO-NH-[GEELLKKLEELLKKG]-CONH<sub>2</sub>

After synthesizing the peptide sequence containing an additional N-terminal glycine, the 15-amino acid peptide resin (300 mg resin, ~ 150 mg peptide, ~ 0.089 mmol) was treated with chloroacetyl chloride (42  $\mu$ L, 0.53 mmol, 6 equiv.) and DIEA (92  $\mu$ L, 0.53, 6 equiv.) in dry

DMF (5 mL) for 1.5 h at rt. Subsequent cleavage from the resin and purification by reversed-phase HPLC afforded peptide **29** as a white solid (35 mg, ~ 25%).

**MS:** (LSIMS, thioglycerol):  $m/z$  1802 ((M + H)<sup>+</sup>, 100).

**Peptide 31 (N1GG):** ClCH<sub>2</sub>CO-NH-[GGEELLKKLEELLKKG]-CONH<sub>2</sub>

After synthesizing the peptide sequence containing two additional N-terminal glycine residues, the 16-amino acid peptide resin (300 mg resin, ~ 150 mg peptide, ~ 0.089 mmol) was treated with chloroacetyl chloride (42  $\mu$ L, 0.53 mmol, 6 equiv.) and DIEA (92  $\mu$ L, 0.53 mmol, 6 equiv.) in dry DMF (5 mL) for 1.5 h at rt. Subsequent cleavage from the resin and purification by reversed-phase HPLC afforded peptide **31** as a white solid (42 mg, ~ 25%).

**MS:** (LSIMS, thioglycerol):  $m/z$  1859 ((M + H)<sup>+</sup>, 100).

**Peptide 33 (N1GGG):** ClCH<sub>2</sub>CO-NH-[GGGEELLKKLEELLKKG]-CONH<sub>2</sub>

After synthesizing the peptide sequence containing three additional N-terminal glycine residues, the 17-amino acid peptide resin (300 mg resin, ~ 150 mg peptide, ~ 0.089 mmol) was treated with chloroacetyl chloride (42  $\mu$ L, 0.53 mmol, 6 equiv.) and DIEA (92  $\mu$ L, 0.53 mmol, 6 equiv.) in dry DMF (5 mL) for 1.5 h at rt. Subsequent cleavage from the resin and purification by reversed-phase HPLC afforded peptide **33** as a white solid (42 mg, ~ 25%).

equiv.) in dry DMF (5 mL) for 1.5 h at rt. Subsequent cleavage from the resin and purification by reversed-phase HPLC afforded peptide **33** as a white solid (45 mg, ~ 25%).

**MS:** (LSIMS, thioglycerol):  $m/z$  1916 ((M + H)<sup>+</sup>, 100).

**Peptide 43:**<sup>6</sup> Ac-[CEELLKKLEELLKKG]-CONH<sub>2</sub>

After synthesizing the peptide sequence containing an N-terminal cysteine residue, the 15-amino acid peptide resin (408 mg resin, ~ 150 mg peptide) was treated with 10% acetic anhydride in NMP for 1 h. Subsequent cleavage from the resin and purification by reversed-phase HPLC afforded peptide **43** as a white solid (23 mg, 15%)

**MS:** (LSIMS, thioglycerol):  $m/z$  1815 ((M + H)<sup>+</sup>, 100).

#### iv. Synthesis of Cavitein Variants

Each cavitein was assessed for purity by observation of a single peak by analytical reversed-phase HPLC. Their identities were characterized by electrospray mass spectrometry. The extinction coefficients listed were calculated based on cavitein concentrations determined by amino acid analysis and were recorded at 25 °C in pH 7 phosphate buffer. Subscripts next to  $\epsilon$  denote the wavelength (nm) at which a maximum absorption was observed for the aromatic

chromophore. Errors in the concentration determined by amino acid analysis can be as high as  $\pm 10\%$  and are thus reflected in  $\epsilon$  values.

**Cavitein 30 (N1G/Ar/Me).**

Procedure "D": A solution of cavitand **3** (1.4 mg, 1.9  $\mu\text{mol}$ ), peptide **29** (12 mg, 6.7  $\mu\text{mol}$ , 5 equiv.) and DIEA (2.3  $\mu\text{L}$ , 13  $\mu\text{mol}$ , 10 equiv.) in degassed DMF (2 mL) was incubated under  $\text{N}_2$  at 30  $^\circ\text{C}$  for 16 h. The reaction mixture was then evaporated *in vacuo*, dissolved in deionized water and purified by reversed-phase HPLC to afford cavitein **30** (6.0 mg, 58%). Cavitein **30** eluted at  $\sim 55\%$  acetonitrile.

**ESMS (deconvoluted spectrum):**  $7787.9 \pm 0.3$  [calcd: 7785.4 (average isotope composition)].

$$\epsilon_{278} = 14\,500 \text{ L mol}^{-1} \text{ cm}^{-1}$$

**Cavitein 32 (N1GG/Ar/Me).**

Procedure "D" was employed using cavitand **3** (1.4 mg, 1.9  $\mu\text{mol}$ ), peptide **31** (22 mg, 12  $\mu\text{mol}$ , 6 equiv.) and DIEA (3.3  $\mu\text{L}$ , 19  $\mu\text{mol}$ , 10 equiv.) to afford cavitein **32** (10 mg, 65%). Cavitein **32** eluted at  $\sim 50\%$  acetonitrile.

**ESMS (deconvoluted spectrum):**  $8015.1 \pm 0.8$  [calcd: 8013.6 (average isotope composition)].

$$\epsilon_{272} = 13\,500 \text{ L mol}^{-1} \text{ cm}^{-1}$$

**Cavitein 34** (N1GGG/Ar/Me).

Procedure "D" was employed using cavitand **3** (2.8 mg, 3.9  $\mu\text{mol}$ ), peptide **33** (33 mg, 17  $\mu\text{mol}$ , 4.4 equiv.) and DIEA (6.8  $\mu\text{L}$ , 39  $\mu\text{mol}$ , 10 equiv.) to afford cavitein **34** (6.8 mg, 21%). Cavitein **34** eluted at  $\sim 50\%$  acetonitrile.

**ESMS (deconvoluted spectrum):**  $8242.8 \pm 0.9$  [calcd: 8241.8 (average isotope composition)].

$$\epsilon_{278} = 17\,000 \text{ L mol}^{-1} \text{ cm}^{-1}$$

**Cavitein 36** (N2/Ar/Me).

Procedure "D" was employed using cavitand **3** (0.9 mg, 1.3  $\mu\text{mol}$ ), peptide **35** (10 mg, 5.5  $\mu\text{mol}$ , 4.4 equiv.) and DIEA (2.2  $\mu\text{L}$ , 13  $\mu\text{mol}$ , 10 equiv.) to afford cavitein **36** (0.9 mg, 9%). Cavitein **36** eluted at  $\sim 55\%$  acetonitrile.

**ESMS (deconvoluted spectrum):**  $7615.1 \pm 1.0$  [calcd: 7613.3 (average isotope composition)].

$$\epsilon_{272} = 17\,000 \text{ L mol}^{-1} \text{ cm}^{-1}$$

**Cavitein 38** (N4/Ar/Me).

Procedure "D" was employed using cavitand **3** (1.0 mg, 1.3  $\mu\text{mol}$ ), peptide **37** (11 mg, 5.8  $\mu\text{mol}$ , 4.4 equiv.) and DIEA (2.3  $\mu\text{L}$ , 13  $\mu\text{mol}$ , 10 equiv.) to afford cavitein **38** (1.2 mg, 12%). Cavitein **38** eluted at  $\sim 55\%$  acetonitrile.

**ESMS (deconvoluted spectrum):**  $7727.9 \pm 0.3$  [calcd: 7725.5 (average isotope composition)].

$$\epsilon_{279} = 20\,000 \text{ L mol}^{-1} \text{ cm}^{-1}$$

**Cavitein 41** (N1/Bzl/Me).

Procedure "D" was employed using cavitand **40** (1.0 mg, 1.4  $\mu\text{mol}$ ), peptide **26** (14 mg, 8.1  $\mu\text{mol}$ , 6 equiv.) and DIEA (2.4  $\mu\text{L}$ , 14  $\mu\text{mol}$ , 10 equiv.) to afford cavitein **41** (3.7 mg, 35%). Cavitein **41** eluted at  $\sim 50\%$  acetonitrile.

**ESMS (deconvoluted spectrum):**  $7614.4 \pm 0.6$  [calcd: 7613.3 (average isotope composition)].

$$\epsilon_{280} = 6\,000 \text{ L mol}^{-1} \text{ cm}^{-1}$$



**Cavitein 42** (N1/Bzl/PO<sub>3</sub>H<sub>2</sub>).<sup>6</sup>

Cavitand **5** (3.6 mg, 2.9  $\mu$ mol) and peptide **26** (20 mg, 12  $\mu$ mol, 4 equiv.) were stirred in a degassed 150 mM sodium phosphate pH 8.5 buffer (4 mL) for 4 h at rt under N<sub>2</sub>. The crude reaction mixture was purified twice by reversed-phase HPLC to afford cavitein **42** (7.8 mg, 33%). Cavitein **42** eluted at  $\sim$  45% acetonitrile.

**ESMS (deconvoluted spectrum):** 8110.6  $\pm$  1.2 [calcd: 8109.4 (average isotope composition)].

$$\epsilon_{279} = 6\,500 \text{ L mol}^{-1} \text{ cm}^{-1}$$

**Cavitein 44** (C/Bzl/PO<sub>3</sub>H<sub>2</sub>).<sup>6</sup>

Peptide **43** (20.4 mg, 11.2  $\mu$ mol) and impure cavitand **21** were stirred in a degassed 150 mM sodium phosphate pH 8.5 buffer (4 mL) for 1 h at rt under N<sub>2</sub>. The crude product was purified twice by reversed-phase HPLC to afford cavitein **44** (1.9 mg, 9%) separable from its tris-phosphate derivative. Cavitein **44** eluted at  $\sim$  45% acetonitrile.

**ESMS (deconvoluted spectrum):** 8394.6  $\pm$  1.2 [calcd: 8393.7 (average isotope composition)].

## v. Guanidine Hydrochloride Denaturation Analysis

Guanidine hydrochloride denaturations were performed as described in Chapter Four. Unfolding was analyzed using the linear extrapolation method of Santoro and Bolen.<sup>12</sup> This method assumes that unfolding is a reversible, two-state process and that the free energy of folding is a linear function of the GuHCl concentration. Accordingly, the data were fit to the following equation using nonlinear least-squares methods to fit the pretransitional baseline and, due to the extreme stability of the caviteins, approximates the post-transitional baseline:<sup>19</sup>

$$\theta_{\text{obs}} = \theta_{\text{N}} (f_n) (1 - a[\text{GuHCl}]) + \theta_{\text{U}} (1 - f_n)$$

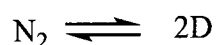
where  $\theta_{\text{obs}}$  is the ellipticity at 222 nm at a given concentration of GuHCl,  $\theta_{\text{N}}$  is the ellipticity of the folded state in the absence of GuHCl,  $\theta_{\text{U}}$  is the ellipticity of the unfolded state,  $a$  is a constant and  $f_n$  is the fraction of the cavitein in the folded state:

$$f_n = e^{-((\Delta G_{\text{H}_2\text{O}} - m [\text{GuHCl}])/RT)} / [1 + e^{-((\Delta G_{\text{H}_2\text{O}} - m [\text{GuHCl}])/RT)}]$$

where  $\Delta G_{\text{H}_2\text{O}}$  is the free energy of unfolding in the absence of denaturant,  $m$  represents the change in free energy with respect to the concentration of GuHCl,  $R$  is the universal gas constant and  $T$  is the temperature. The values of  $\Delta G_{\text{H}_2\text{O}}$ ,  $m$ , and  $a$  were determined by nonlinear least-squares analysis by the program KaleidaGraph v. 3.08d (Synergy Software). The value of  $\theta_{\text{N}}$  was normalized in all cases to one. The errors reported in this thesis were calculated by this analysis. In cases where a full unfolding transition was observed,  $\theta_{\text{N}}$  (and all

of the points on the curve) was corrected by subtracting the value of  $\theta_U$  such that  $\theta_U$  is set to zero. In cases where a full unfolding transition is not observed, this analysis was performed for comparison's sake but it should be noted that additional errors may be present. For example, one approximation that was made was that  $\theta_U$  was estimated to be zero. Note that positive or negative values for  $\theta_U$  (experimentally unobservable) will introduce additional errors to the values of  $m$  and  $\Delta G^\circ_{H_2O}$ . However, a sensitivity analysis revealed that the additional errors were relatively small ( $< 0.5$  kcal/mol effect on  $\Delta G^\circ_{H_2O}$ ) in most cases due to the relatively low  $m$  values. However, in the case of N1G/Ar/Me, small errors in  $\theta_U$  are manifested into larger errors in  $\Delta G^\circ_{H_2O}$  as a result of the higher  $m$  value. For example a 2% error in  $\theta_U$  (which represents a typical  $\theta_U$  offset),  $\Delta G^\circ_{H_2O}$  is affected by  $\sim 1$  kcal/mol. In addition, in all cases where a full unfolding transition was not observed, the  $m$  value is also approximated and thus errors in its value would also be reflected in  $\Delta G^\circ_{H_2O}$ . For these reasons, interpretation of  $\Delta G^\circ_{H_2O}$  and  $m$  values were kept to a minimum in this thesis.

In the case of N1G/Ar/Me, we observed a concentration dependent denaturation curve suggesting a dimer (folded) to monomer (unfolded) transition. Therefore, in addition to the above analysis, we analyzed the data according to:<sup>20</sup>



$$K_u = [U]^2 / [N_2] = 2 P_t (f_u^2 / (1 - f_u))$$

where  $[N_2]$  and  $[U]$  are the concentrations of the native dimer and the unfolded monomer, respectively,  $P_t$  is the total cavitein concentration in mol/L, and  $f_u$  is the fraction of cavitein unfolded. If we assume (as above) that the free energy of unfolding is a linear function of denaturant concentration, we find that:

$$f_u = k^2 / 2 P_t (1 + (1 + 8P_t / k^2)^{1/2})$$

where

$$k = \exp - ((\Delta G^\circ_{H_2O} - m[GuHCl])/RT)$$

The extrapolated value was calculated using a non-linear least squares curve fitting as above. Note that this equation assumes that the monomeric form of N1G/Ar/Me is entirely unstable which may not be valid.

#### vi. NMR H/D Exchange on N1G/Ar/Me

This experiment was performed on a Bruker AMX-500 spectrometer running at 500 MHz for protons. Spectra were recorded at 298 K with a sweep width of 6579 Hz, 32768 data points, 256 scans and a relaxation delay of 1.0 seconds. Spectra were processed with WIN-NMR software using an exponential line-broadening function: "LB" was set at 5.0 Hz. Sample concentration was ~ 0.3 mM. Other experimental details such as the calculation of rate constants were performed as described in Chapter Four.

vii. Analytical Ultracentrifugation Studies

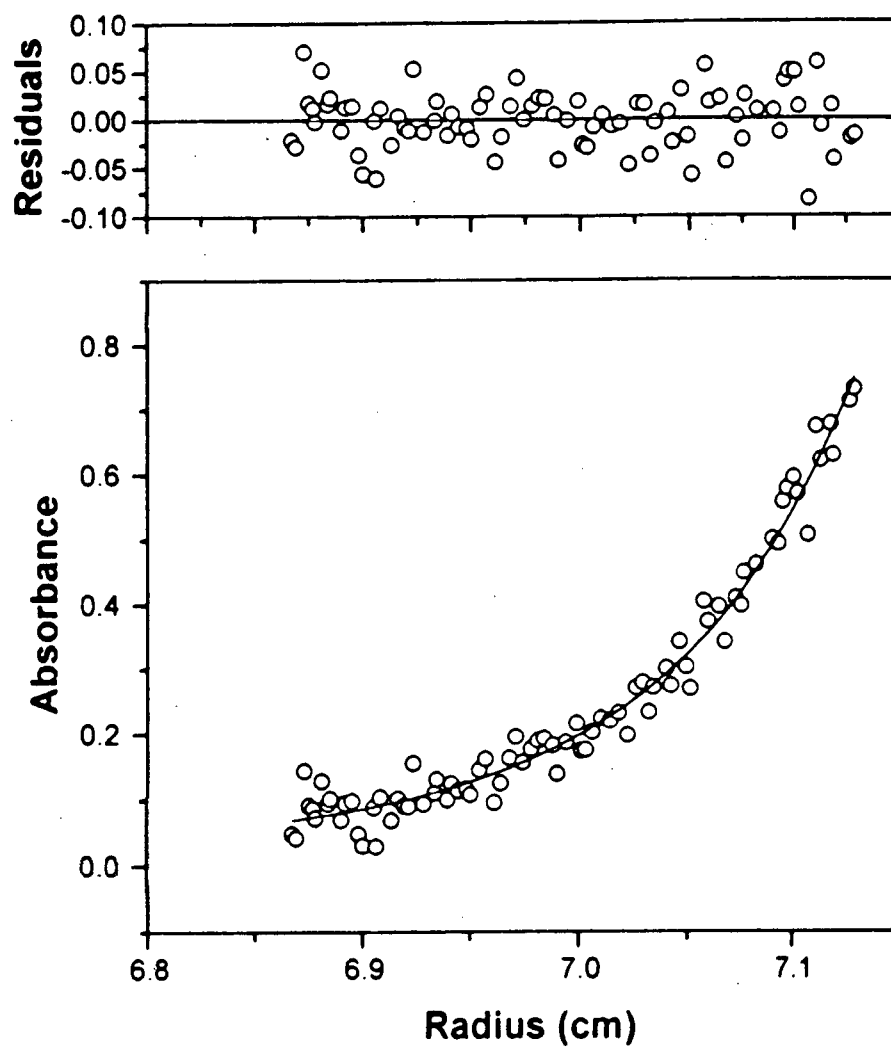
Sedimentation equilibrium experiments were performed as described in Chapter Four. Each sample was run at 30 000 rpm and in 50 mM phosphate buffer at pH 7.0 unless otherwise noted. Table 5.7 lists specific experimental details for each experiment.

**Table 5.7.** Experimental Details for Sedimentation Equilibrium Experiments Reported in Chapter Five.

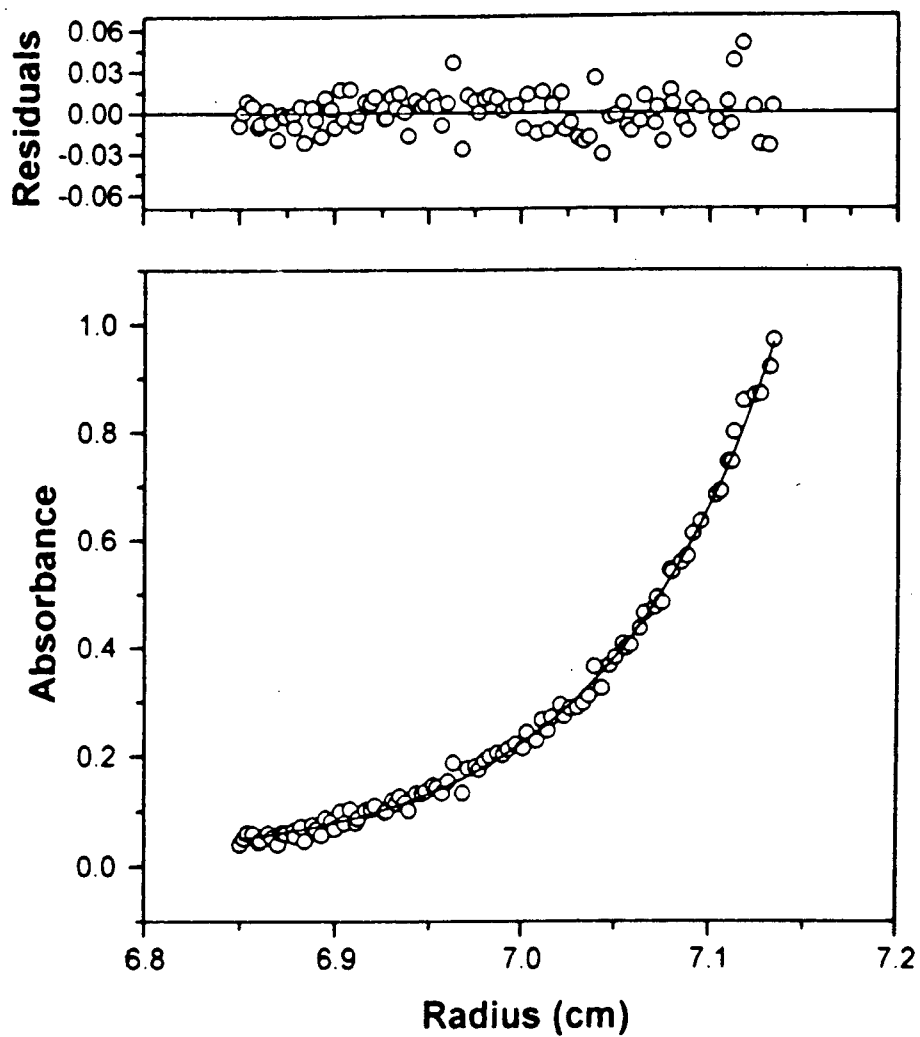
Cavitein	Concentration of Analyzed Sample ( $\mu\text{M}$ )	Partial Specific Volume ( $\text{mL/g}$ )	Wavelength Used For Detection ( $\text{nm}$ )
N1G/Ar/Me	3	0.7816	220
N1G/Ar/Me	30	0.7816	280
N1GG/Ar/Me	30	0.7814	272
N1GG/Ar/Me <sup>a</sup>	30 + 7.2 M GuHCl	0.7982	272
N1GGG/Ar/Me	30	0.7770	278
N1GGG/Ar/Me	300	0.7770	301
N2/Ar/Me	30	0.7900	280
N4/Ar/Me	30	0.7900	280
N1/Bzl/Me	30	0.7900	240
N1/Bzl/Me	300	0.7900	290
N1/Bzl/ $\text{PO}_3\text{H}_2$	30	0.7900	280

<sup>a</sup> due to the GuHCl, the density of the solvent was estimated to be  $\rho = 1.169 \text{ g/mL}$ . For all other runs,  $\rho$  was estimated to be  $1.000 \text{ g/mL}$ .

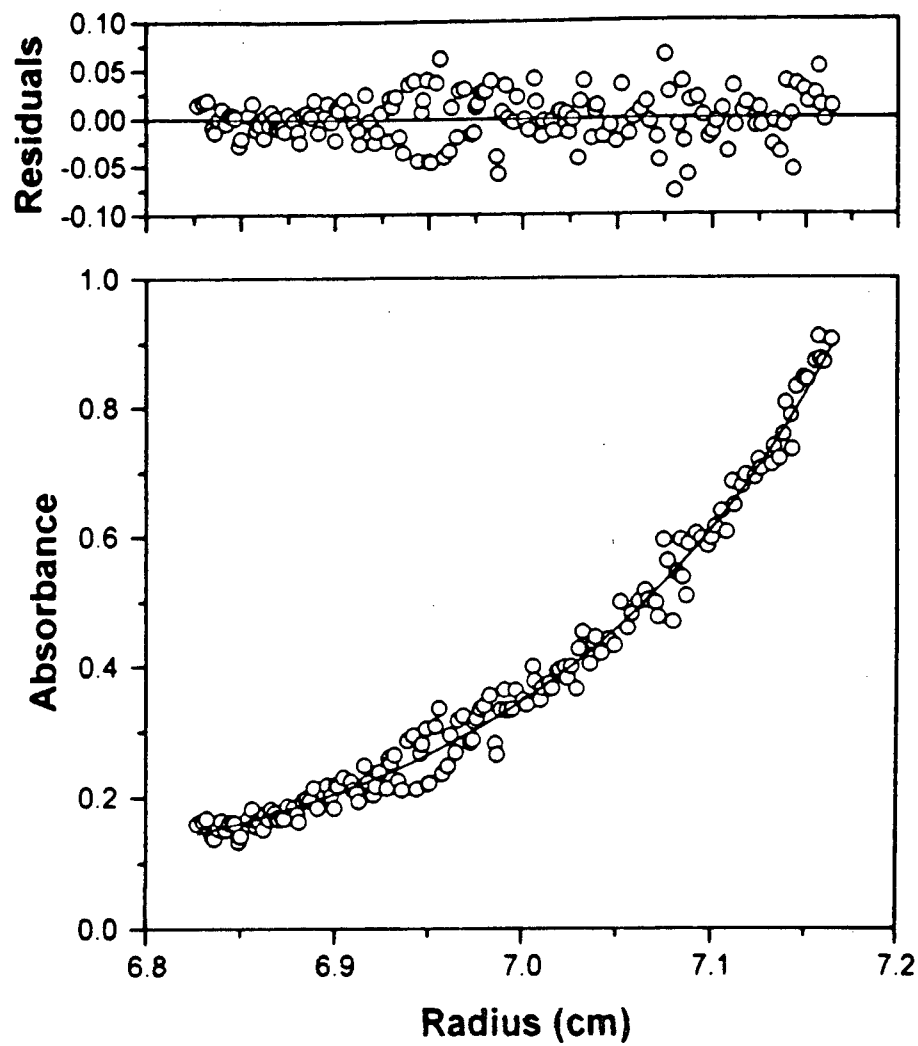
**Figure 5.28.** Sedimentation Equilibrium Analysis of N1G/Ar/Me at 3  $\mu$ M. In the Lower Panel, The Solid Line Represents the Theoretical Fit to a Single Non-Interacting Species. The Upper Panel Represents the Residuals for the Fit.



**Figure 5.29.** Sedimentation Equilibrium Analysis of N1G/Ar/Me at 30  $\mu$ M. In the Lower Panel, The Solid Line Represents the Theoretical Fit to a Single Non-Interacting Species. The Upper Panel Represents the Residuals for the Fit.

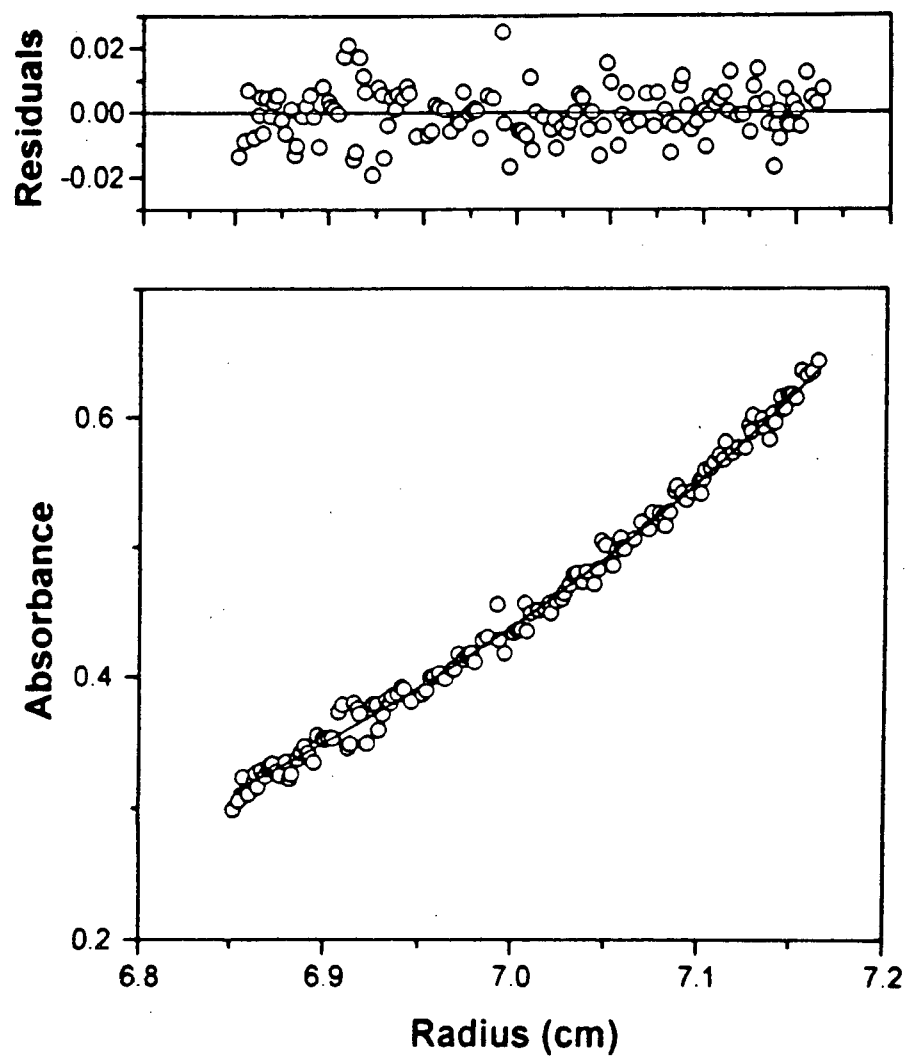


**Figure 5.30.** Sedimentation Equilibrium Analysis of N1GG/Ar/Me at 30  $\mu$ M. In the Lower Panel, The Solid Line Represents the Theoretical Fit to a Single Non-Interacting Species. The Upper Panel Represents the Residuals for the Fit.

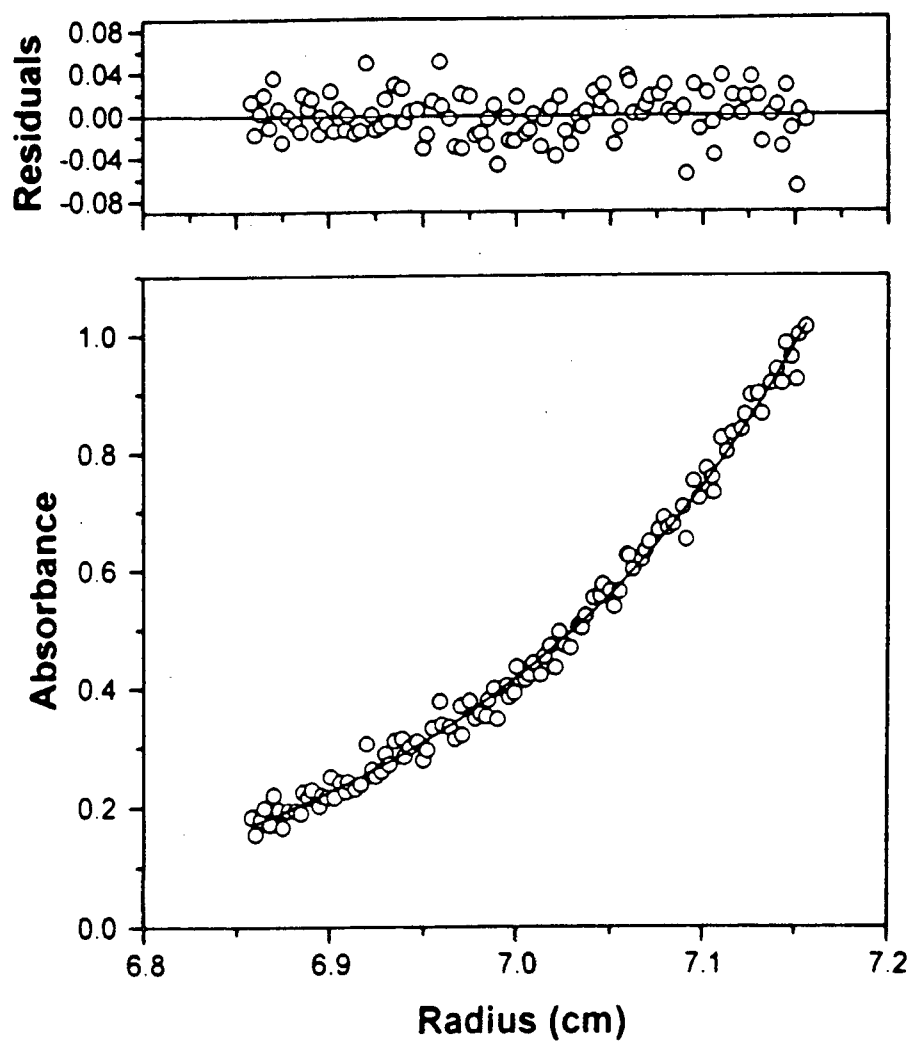




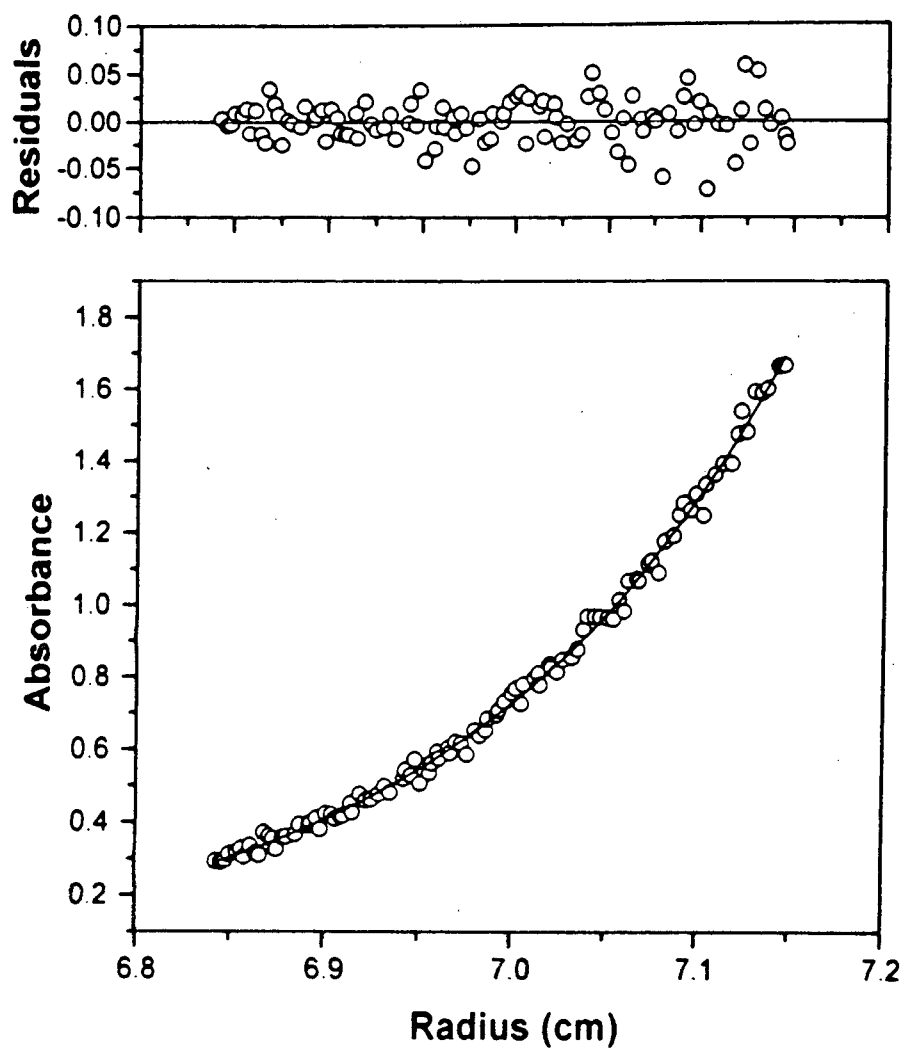
**Figure 5.31.** Sedimentation Equilibrium Analysis of N1GG/Ar/Me at 30  $\mu$ M in the Presence of 7.2 M GuHCl. In the Lower Panel, The Solid Line Represents the Theoretical Fit to a Single Non-Interacting Species. The Upper Panel Represents the Residuals for the Fit.



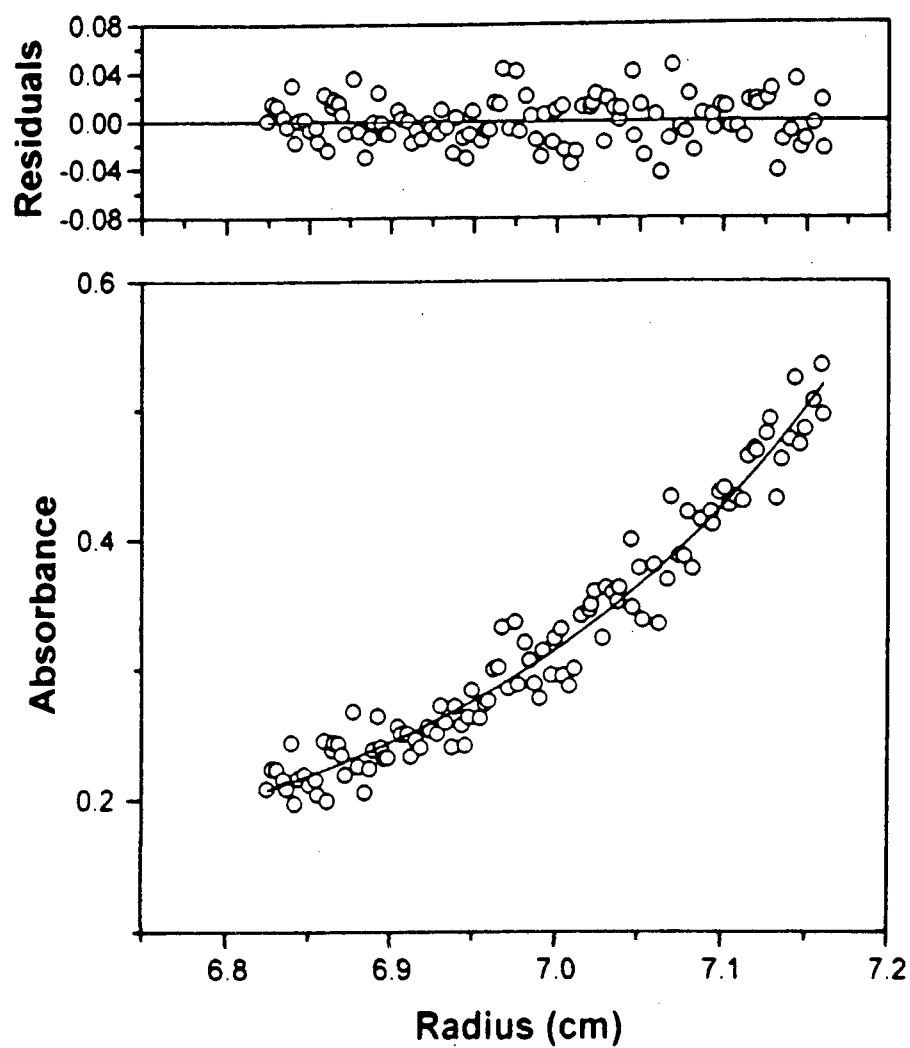
**Figure 5.32.** Sedimentation Equilibrium Analysis of N1GGG/Ar/Me at 30  $\mu$ M. In the Lower Panel, The Solid Line Represents the Theoretical Fit to a Single Non-Interacting Species. The Upper Panel Represents the Residuals for the Fit.



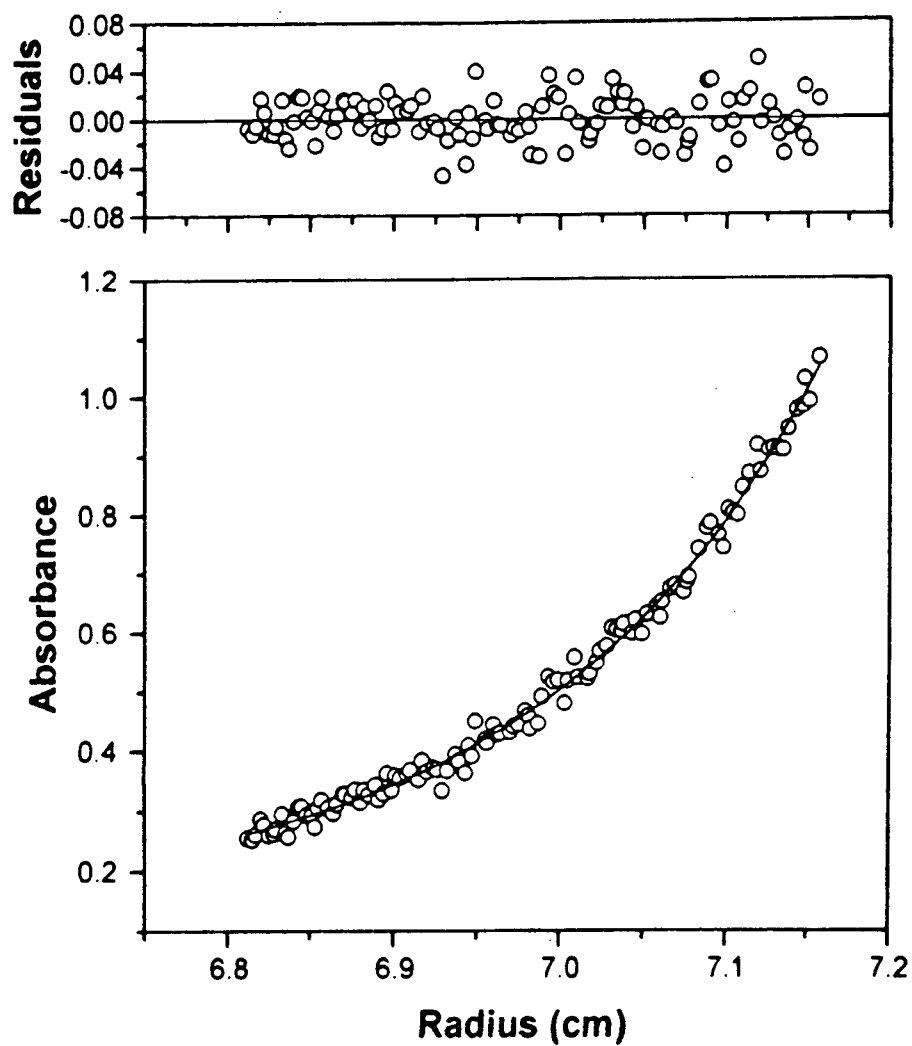
**Figure 5.33.** Sedimentation Equilibrium Analysis of N1GGG/Ar/Me at 300  $\mu$ M. In the Lower Panel, The Solid Line Represents the Theoretical Fit to a Single Non-Interacting Species. The Upper Panel Represents the Residuals for the Fit.



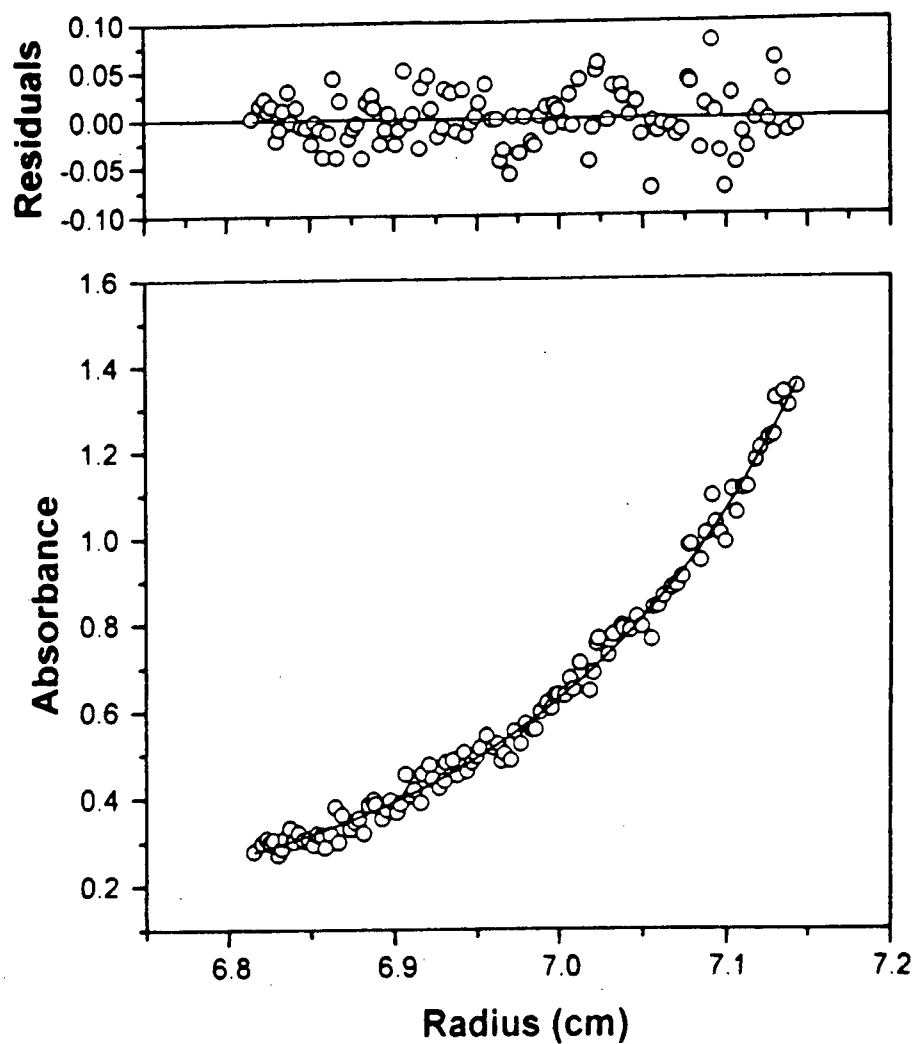
**Figure 5.34.** Sedimentation Equilibrium Analysis of N<sub>2</sub>/Ar/Me at 30  $\mu$ M. In the Lower Panel, The Solid Line Represents the Theoretical Fit to a Single Non-Interacting Species. The Upper Panel Represents the Residuals for the Fit.



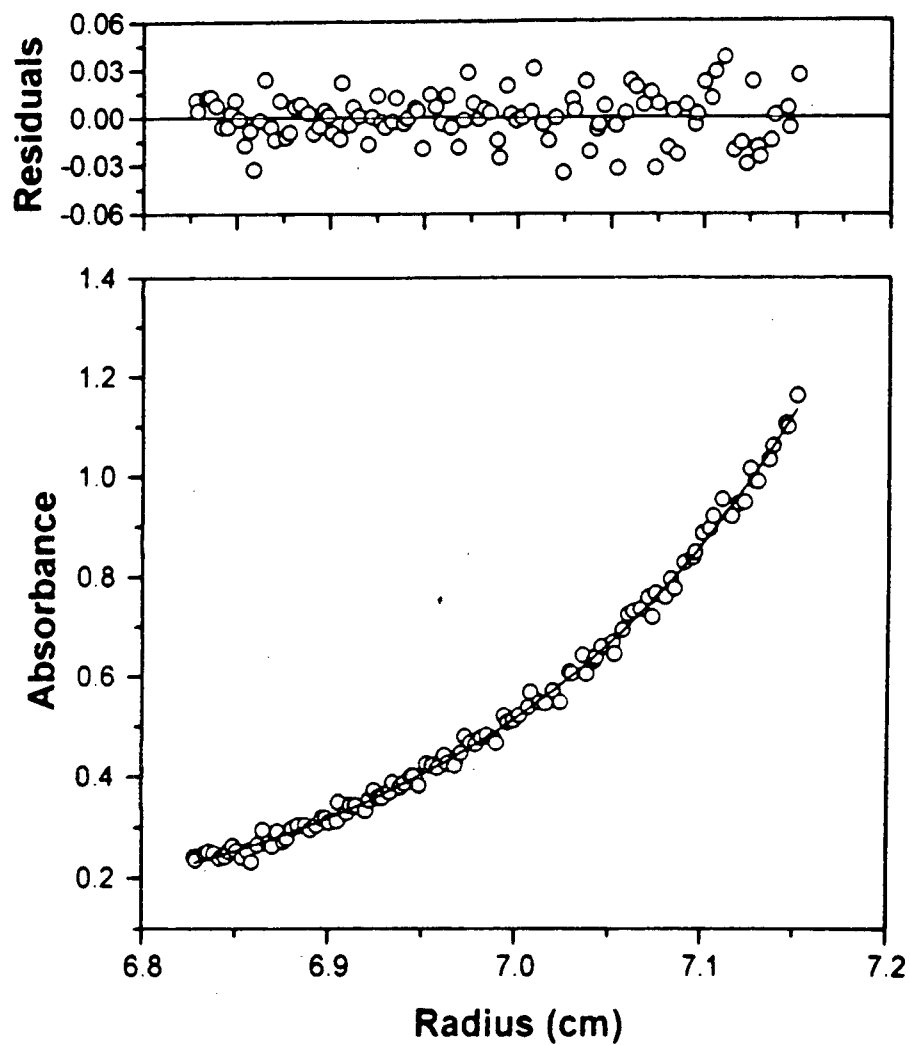
**Figure 5.35.** Sedimentation Equilibrium Analysis of N4/Ar/Me at 30  $\mu$ M. In the Lower Panel, The Solid Line Represents the Theoretical Fit to a Single Non-Interacting Species. The Upper Panel Represents the Residuals for the Fit.



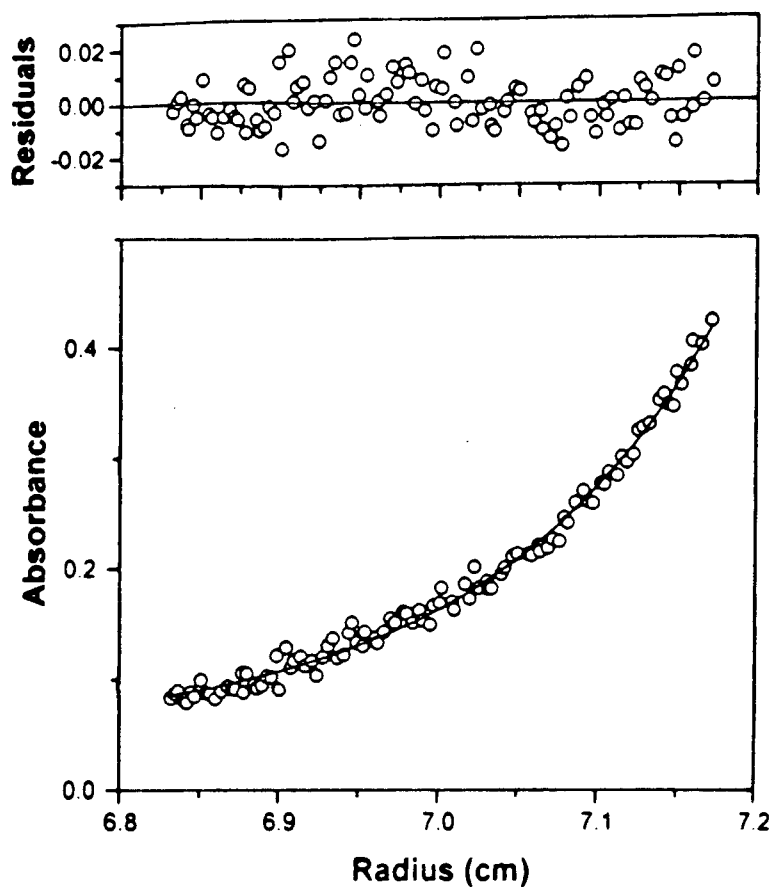
**Figure 5.36.** Sedimentation Equilibrium Analysis of N1/Bzl/Me at 30  $\mu$ M. In the Lower Panel, The Solid Line Represents the Theoretical Fit to a Single Non-Interacting Species. The Upper Panel Represents the Residuals for the Fit.



**Figure 5.37.** Sedimentation Equilibrium Analysis of N1/Bzl/Me at 300  $\mu$ M. In the Lower Panel, The Solid Line Represents the Theoretical Fit to a Single Non-Interacting Species. The Upper Panel Represents the Residuals for the Fit.



**Figure 5.38.** Analysis of N1/Bzl/PO<sub>3</sub>H<sub>2</sub> by Sedimentation Equilibrium at 30  $\mu$ M. In the Lower Panel, The Solid Line Represents the Theoretical Fit to a Single Non-Interacting Species. The Upper Panel Represents the Residuals for the Fit.



#### viii. Computer Modeling

Computer modeling of the preferred conformations of the *S*-methyl derivatives of cavitands **3** and **40** were performed with molecular mechanics calculations using the MM2 force field within the computer program "CS Chem3D Pro" v. 3.2 (CambridgeSoft Corporation) run on a Macintosh operating system. The majority of the observed 5 kcal/mol difference in energy between the preferred orientations of each cavitand was found to be the result of the dipole-dipole contribution to the force field.



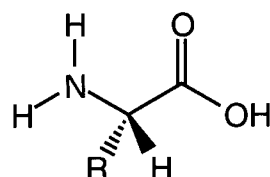
## F. References

1. Schneider, J. P.; Kelly, J. W. *Chem. Rev.* **1995**, *95*, 2169-2187.
2. Wong, A. K.; Jacobsen, M. P.; Winzor, D. J.; Fairlie, D. P. *J. Am. Chem. Soc.* **1998**, *120*, 3836-3841.
3. Cram, D. J.; Karbach, S.; Kim, Y. H.; Baczynskyj, L.; Marti, K.; Sampson, R. M.; Kallemeyn, G. W. *J. Am. Chem. Soc.* **1988**, *110*, 2554-2560.
4. Sorrell, T. N.; Pigge, F. C. *J. Org. Chem.* **1993**, *58*, 784-785.
5. Causton, A. S.; Sherman, J. C. *unpublished results*.
6. Mezo, A. R.; Sherman, J. C. *J. Org. Chem.* **1998**, *63*, 6824-6829.
7. (a) Gibney, B. R.; Rabanal, F.; Skalicky, J. J.; Wand, A. J.; Dutton, P. L. *J. Am. Chem. Soc.* **1997**, *119*, 2323-2324; (b) Kohn, W. D.; Kay, C. M.; Hodges, R. S. *Protein Sci.* **1995**, *4*, 237-250.
8. Woody, R. W. *Biopolymers* **1978**, *17*, 1451-1467.
9. Chakrabartty, A.; Kortemme, T.; Padmanabhan, S.; Baldwin, R. L. *Biochemistry* **1993**, *32*, 5560-5565.
10. However, a reversed-phase column was used prior to injection onto the ESI mass spectrometer in order to concentrate each cavitein and provide better signal:noise ratios in the resulting spectra. As a consequence, gradients of acetonitrile in water were used to elute the caviteins off the column. In consideration of the sedimentation equilibrium experiment conducted on cavitein **27** in the presence of 10% MeOH (Chapter Four) which demonstrated that methanol inhibits dimerization, the acetonitrile required to elute the cavitein off the column (likely ~50% v/v) may be more than sufficient to inhibit dimerization/oligomerization prior to injection into the spectrometer. Since no evidence for oligomerization was observed by ESMS for any of the caviteins presented in this thesis, it is consistent with high percentages of acetonitrile inhibiting oligomerization.
11. Pace, C. N. *Methods Enzymol.* **1986**, *131*, 266-280.
12. Santoro, M. M.; Bolen, D. W. *Biochemistry* **1988**, *27*, 8063-8068.
13. Myers, J. K.; Pace, C. N.; Scholtz, J. M. *Protein Sci.* **1995**, *4*, 2138-2148.
14. For example: Akerfeldt, K. S.; Kim, R. M.; Camac, D.; Groves, J. T.; Lear, J. D.; DeGrado, W. F. *J. Am. Chem. Soc.* **1992**, *114*, 9656-9657.

15. For example: Lombardi, A.; Bryson, J. W.; DeGrado, W. F. *J. Am. Chem. Soc.* **1997**, *119*, 12378-12379.
16. For example: Torrado, A.; Walkup, G. K.; Imperiali, B. *J. Am. Chem. Soc.* **1998**, *120*, 609-610.
17. For example: (a) Johnsson, K.; Allemann, R. K.; Widmer, H.; Benner, S. A. *Nature* **1993**, *365*, 530-532; (b) Severin, K.; Lee, H.; Kennan, A. J.; Ghadiri, M. R. *Nature* **1997**, *389*, 706-709; (c) Broo, K. S.; Brive, L.; Ahlberg, P.; Baltzer, L. *J. Am. Chem. Soc.* **1997**, *119*, 11362-11372; (d) Gibney, B. R.; Rabanal, F.; Reddy, K. S.; Dutton, P. L. *Biochemistry* **1998**, *37*, 4635-4643.
18. Cram, D. J.; Karbach, S.; Kim, Y. H.; Baczynskyj, L.; Marti, K.; Sampson, R. M.; Kallemeyn, G. W. *J. Am. Chem. Soc.* **1988**, *110*, 2554-2560.
19. Regan, L.; Rockwell, A.; Wasserman, Z.; DeGrado, W. F. *Protein Sci.* **1994**, *3*, 2419-2427.
20. (a) Neet, K. E.; Timm, D. E. *Protein Sci.* **1994**, *3*, 2167-2174; (b) Mok, Y.-K.; Gay, G. D.; Butler, P. J.; Bycroft, M. *Protein Sci.* **1996**, *5*, 310-319; (c) De Francesco, R. D.; Pastore, A.; Vecchio, G.; Cortese, R. *Biochemistry* **1991**, *30*, 143-147; (d) Bowie, J. U.; Sauer, R. T. *Biochemistry* **1989**, *28*, 7140-7143.

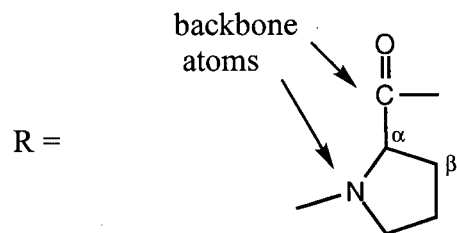
# Appendix A.

## The Twenty Commonly Occurring Amino Acids.

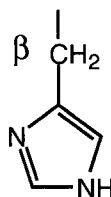


Amino Acid	One-Letter Code	Three-Letter Code	Side Chain: R =
Glycine	G	Gly	H
Alanine	A	Ala	CH <sub>3</sub>
Valine	V	Val	CH(CH <sub>3</sub> ) <sub>2</sub>
Leucine	L	Leu	CH <sub>2</sub> CH(CH <sub>3</sub> ) <sub>2</sub>
Isoleucine	I	Ile	isobutyl
Serine	S	Ser	CH <sub>2</sub> OH
Threonine	T	Thr	CH(OH)CH <sub>3</sub>
Cysteine	C	Cys	CH <sub>2</sub> SH
Methionine	M	Met	CH <sub>2</sub> CH <sub>2</sub> SCH <sub>3</sub>
Proline	P	Pro	see below
Aspartic Acid	D	Asp	CH <sub>2</sub> COOH
Asparagine	N	Asn	CH <sub>2</sub> CONH <sub>2</sub>
Glutamic Acid	E	Glu	CH <sub>2</sub> CH <sub>2</sub> COOH
Glutamine	Q	Gln	CH <sub>2</sub> CH <sub>2</sub> CONH <sub>2</sub>
Lysine	K	Lys	CH <sub>2</sub> (CH <sub>2</sub> ) <sub>3</sub> NH <sub>2</sub>
Arginine	R	Arg	(CH <sub>2</sub> ) <sub>3</sub> NHC(NH)(NH <sub>2</sub> )
Histidine	H	His	see below
Phenylalanine	F	Phe	CH <sub>2</sub> Ph
Tyrosine	Y	Tyr	CH <sub>2</sub> Ph- <sub>para</sub> OH
Tryptophan	W	Trp	see below

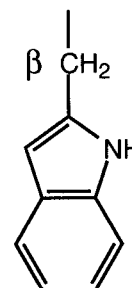
Proline



Histidine



Tryptophan



## Appendix B.

Amino Acid Sequences of Various Naturally Occurring and De Novo Designed Four-Helix Bundles and Coiled-Coils as Discussed in Chapter One, Section E. Reference Numbers Refer to Those In Chapter One.

Principal Author	Name of Sequence	Amino Acid Sequence in One-Letter Code <sup>a, b</sup>
DeGrado	Coil-VL <sup>102</sup>	GNADELYRMVDALREHVQSLRRKVRSG
DeGrado	Coil-AL <sup>102</sup>	GNADELYRMADALREHAQSLRRKARSG
DeGrado	Coil-TL <sup>102</sup>	GNADELYRMTDALREHTQSLRRKTRSG
DeGrado	Coil-LL <sup>102</sup>	GNADELYRMLDALREHLQSLRRKLRS
DeGrado	RLP-1 <sup>103</sup>	SAQELLKIARRLRKEAKELLKRAEHGGPELLKE AEELLEKKVDKLYKIAEHG
DeGrado	RLP-2 <sup>103</sup>	SAQELLKIARRLRKEAKELLKRAEHGGPELLKE VEELLEKKADKLYKIVEHG
DeGrado	RLP-3 <sup>103</sup>	SAQELLKIARRLRKEAKELLKRAEHGGPELLKE VEELLEKKVDKLYKIVEHG
Baltzer	GTD-43 <sup>107</sup>	SL[Aad]A[Nva]LQEAFAWLQYHAAKGTGPAQ DQEALRAFA-[Aad]-QL-[Nva]-AKIN <sup>c</sup>
Dutton	H10H24 <sup>108b</sup>	CGGGELWKLHEELLKKFEELLKLHEERLKKL
Dutton	H10H24-L6I <sup>109</sup>	CGGGEIWKLHEELLKKFEELLKLHEERLKKL
Dutton	H10H24-L13F <sup>109</sup>	CGGGELWKLHEEFLLKKFEELLKLHEERLKKL
Dutton	H10H24-L6I,L13F <sup>109</sup>	CGGGEIWKLHEEFLLKKFEELLKLHEERLKKL
Farid	HYP-1 <sup>110</sup>	Peptide A: NH <sub>2</sub> - KEEEILKKVEYLKKAVEELKKKVEEGC Peptide B: NH <sub>2</sub> - CGELKKKLEEVKKKLEEVKKKLWEKK
Alber	GCN4-pLI <sup>112</sup>	RMKQIEDKLEEILSKLYHIENELARIKKLLGER- CO <sub>2</sub> H
Fairman	Lac 21 <sup>81</sup>	MKQLADSLMQLARQVSRLESA
Fairman	Lac 28 <sup>81</sup>	LMQLARQMKQLADSLMQLARQVSRLESA
Fairman	Lac 35 <sup>81</sup>	LMQLARQLMQLARQMKQLADSLMQLARQVS RLESA
Fairman	Lac 21E <sup>118</sup>	MEELADSLEELARQVEELES
Fairman	Lac 21K <sup>118</sup>	MKKLADSLKKLARQVKKLES
Regan/ Banner	ROP <sup>23</sup>	NH <sub>2</sub> - GTKQEKTALNMARFIRSQTLTLEKLNELDAD EQADICESLHDHADEL YRSCLARFGDDGENL- CO <sub>2</sub> H

<sup>a</sup> Amino acids in bold were varied in the series.

<sup>b</sup> All N-termini were acetylated and C-termini amidated unless otherwise noted with NH<sub>2</sub> or CO<sub>2</sub>H.

<sup>c</sup> Aad =  $\alpha$ -amino adipic acid; Nva = norvaline.